Genomic structure of murine methylmalonyl-CoA mutase: evidence for genetic and epigenetic mechanisms determining enzyme activity

Michael F. WILKEMEYER, Elizabeth R. ANDREWS and Fred D. LEDLEY*
Departments of Cell Biology and Pediatrics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, U.S.A.

Methylmalonyl-CoA mutase (MCM) is a nuclear-encoded mitochondrial matrix enzyme. We have reported characterization of murine MCM and cloning of a murine MCM cDNA and now describe the murine *Mut* locus, its promoter and evidence for tissue-specific variation in MCM mRNA, enzyme and holoenzyme levels. The *Mut* locus spans 30 kb and contains 13 exons constituting a unique transcription unit. A B1 repeat element was found in the 3′ untranslated region (exon 13). The transcription initiation site was identified and upstream sequences were shown to direct expression of a reporter gene in cultured cells. The promoter contains sequence motifs characteristic of: (1) TATA-less housekeeping promoters; (2) enhancer elements purportedly involved in co-ordinating expression of nuclear-encoded mitochondrial proteins; and (3) regulatory elements including CCAAT boxes, cyclic AMP-response elements and potential AP-2-binding sites. Northern blots demonstrate a greater than 10-fold variation in steady-state mRNA levels, which correlate with tissue levels of enzyme activity. However, the ratio of holoenzyme to total enzyme varies among different tissues, and there is no correlation between steady-state mRNA levels and holoenzyme activity. These results suggest that, although there may be regulation of MCM activity at the level of mRNA, the significance of genetic regulation is unclear owing to the presence of epigenetic regulation of holoenzyme formation.

INTRODUCTION

Methylmalonyl-CoA mutase (MCM; EC 5.4.99.2) is a nuclear-encoded mitochondrial matrix enzyme which catalyses isomerization between methylmalonyl-CoA and succinyl-CoA. In man and higher eukaryotes, this enzyme is involved in the degradation of propionyl-CoA generated during catabolism of odd-chain fatty acids and certain branched-chain amino acids as well as metabolism of propionate absorbed from the gut [1–3]. MCM is commonly considered to be a housekeeping enzyme and is present in all tissues examined [4–6]. Genetic deficiency of MCM in man leads to an often fatal metabolic disorder known as *mut* methylmalonic acidaemia (McKusick no. 251000) [7].

Human MCM has been purified, characterized and cloned. Enzyme from human placenta and liver is a homodimer which binds 1 mol of adenosylcobalamin/subunit [8,9]. The monomer is synthesized as a 742-amino-acid (82283 Da) propeptide containing a 32-amino acid mitochondrial targetting sequence which is cleaved during the process of uptake into the mitochondria leaving a mature protein of 710 amino acids (78489 Da) [10–12]. The amino acid sequence of human MCM shows considerable identity with the homologous adenosylcobalamin-dependent MCM from *Propionibacterium shermanii* [13]. The locus encoding MCM (*mut*) has been mapped and cloned [14–16] and the recombinant enzyme has been overexpressed for molecular and enzymic characterization [17]. Mutations underlying various forms of *mut* methylmalonic acidaemia have been characterized and have identified critical functional domains within the enzyme [18–20].

We have undertaken the molecular characterization of murine MCM with the intent of using the mouse as a model for studying the role of MCM in metabolic homoeostasis and disease. Preliminary biochemical characterization of murine MCM demonstrated that it exhibited affinity constants for substrate and cofactor similar to its human homologue, and that the rate of metabolic flux from propionate through MCM in murine cell cultures was similar to that of human cells [21]. Murine MCM has been cloned and sequenced revealing a protein of 748 amino acids (82891 Da) with a 30-amino acid leader sequence and 718-amino acid (79284 Da) mature protein. The amino acid sequence exhibits 94% identity with that of homologous human MCM [21].

Southern blotting demonstrated that the murine MCM gene is encoded by a single genetic locus designated *Mut*. This locus was mapped by *in situ* hybridization [22] and by genetic linkage in somatic cell hybrids [23] to mouse chromosome 17C-D. There are no pseudogenes or other loci homologous to MCM in the murine genome.

The present work describes the genomic structure of the *Mut* locus including its promoter. The observation that this promoter contains sequence motifs characteristic of both ubiquitously expressed housekeeping genes and more highly regulated genes led us to consider the role of genetic regulation of enzyme levels in different murine tissues. We demonstrate a more than 10-fold variation in the steady-state levels of MCM mRNA and a log–linear correlation between the level of mRNA and total enzyme. There is, however, no correlation between the steady-state level of mRNA and enzyme activity because of tissue-specific variation in the ratio of holoenzyme/apoenzyme. This suggests that both genetic and epigenetic factors play an important role in determining tissue-specific levels of enzyme activity.

EXPERIMENTAL

Cloning and characterization of genomic locus

Overlapping phage clones carrying the murine MCM gene were isolated from a Lambda gem11 (Promega, Madison, WI, U.S.A.) genomic library constructed from an *Mbo*I partial digest of DNA from FVB mice (a gift from Dr. Paul Overbeek, Baylor College

Abbreviations used: MCM, methylmalonyl-CoA mutase; CRE, cyclic AMP-response element; NEMP, nuclear-encoded mitochondrial proteins; β-gal, *Escherichia coli* β-galactosidase.

* To whom correspondence should be addressed at: GENEMEDICINE Inc., 8080 North Stadium Drive, Houston, TX 77054, U.S.A.

Table 1 List of oligonucleotides used in this report

Location refers to 5′ end of oligonucleotide at the corresponding base pair in the cDNA sequence [21].

Number	Orientation	Sequence	Location
16	Sense	5′-AGAAATCACAGATGCCCT-3′	bp-1740; exon 10
17	Antisense	5′-CCTCCACACATGACAA-3′	bp-2145; exon 12
23	Sense	5′-CCGCTCGAGTTTCCATGACTATGA-3′	bp-624; exon 3
30	Antisense	5′-CCTCCCCAAGGATCAGCCAC-3′	bp-1337; exon 6
35	Sense	5′-TGATGAAGCTTTGGGTTTGCC-3′	bp-1211; exon 6
62	Antisense	5′-CATCTCCACGAACTCTGGGG-3′	bp-490; exon 3
63	Sense	5′-CCAGGGCAGATACTCTGG-3′	bp-261; exon 2
65	Antisense	5′-TGCTGCTGGTGTAGAAGGCG-3′	bp-131; exon 2
126	Sense	5′-GAAGTTGATATATGAAGTTGAAG-3′	bp-1368; exon 7
127	Antisense	5′-GCGAAGTTTAGGGATCCTTCAG-3′	bp-1419; exon 7
129	Antisense	5′-GAAGTGTTGTCAATGGCCAGG-3′	bp-1546; exon 8
131	Antisense	5′-TATTGCCATCTCCACTGGCAG-3′	bp-1662; exon 9
132	Antisense	5′-ATATGCTCCACTCACCATACG-3′	bp-1781; exon 10
137	Sense	5′-GAGAGTTAATAAATTCATGGAACG-3′	bp-1846; exon 11
139	Sense	5′-TTCTGAGGTAATTGTTGGAG-3′	bp-1486; exon 8
140	Antisense	5′-AAGAGGGCCTATGTCCACATC-3′	bp-1970; exon 11
141	Sense	5′-GCATGCTGTGGATGCAGATGTC-3′	bp-2017; exon 12
142	Sense	5′-AGCAGGGATCAAGCTTTGGCTG-3′	bp-1610; exon 9
143	Antisense	5′-CTGCTTCAGTCATAACTAAAGATC-3′	bp-2300; exon 13
144	Sense	5′-CCAAAATTTAATCCAATTTCG-3′	bp-800; exon 4
145	Antisense	5′-TACTCCAACCCATCTGCCATG-3′	bp-880; exon 4
146	Antisense	5′-ATTAAGTAGCCCACAGTCTTC-3′	bp-1017; exon 5
147	Sense	5′-AGAAGACTGTGGGCTCACTTA-3′	bp-1016; exon 5
164	Antisense	5′-AAGTCCAGAGTATCTGCCCTGG-3′	bp-261; exon 2
182	Antisense	5′-ACAGTGTTTGGGGATCACCCG-3′	bp-+20; exon 1
196	Sense	5′-ATTTGATTGGCTGCCGAGCTC-3′	bp-+40; exon 1
203	Sense	5′-CACATATTTAAAGAATCTGAG-3′	bp-2474; exon 13
205	Sense	5′-GATGCATTTGCTATAATGACAGTC-3′	bp-3126; exon 13
206	Sense	5′-CCACCTGCCTCTGCCTTCCTAG-3′	bp-2718; exon 13
207	Antisense	5′-CCATCTGGAAGCTGATGGAATGTTTAGCTGC-3′	bp-97; exon 2
T3		5′-GCATTAACCCTCACTAAAGGG-3′	
T7		5′-TAATACGACTCACTATAGGG-3′	
Sp6		5′-GATTTAGGTGACACTATAG-3′	

of Medicine). Approx. 8×10^5 recombinants were screened by filter hybridization with segments of the murine MCM cDNA clones. Genomic clones were plaque-purified and phage was prepared using standard methods [24,25].

Overlapping exon-containing restriction endonuclease fragments from the phage clones were subcloned into pGEM7Zf(+), pGEM5Zf(+) (Promega) or pBSK+ (Stratagene, LaJolla, CA, U.S.A.). Double-stranded dideoxy sequencing was performed using T7 Sequencing (Pharmacia, Piscataway, NJ, U.S.A.) and Sequenase Version 2.0 (United States Biochemical, Cleveland, OH, U.S.A.). Oligonucleotide sequences are shown in Table 1.

Standard protocols for PCR, genomic DNA isolation, restriction endonuclease digestions, Southern-blot analysis and preparation of radioactive probes were used [24–27]. Genomic Southern-blot analysis was performed on DNA isolated from the FVB strain of mice or from the embryonic stem cell line AB1 passaged twice without the feeder layer [28]. AB1 cells were derived from a black agouti 129 strain of mice [29].

Analysis of MCM mRNA

Total RNA was isolated from various tissues of ICR mice and embryonic stem cells (AB1) using RNAzol B (Biotex Laboratories Inc., Houston, TX, U.S.A.). Samples of tissue from two or three different mice were combined for RNA isolation. Poly(A) RNA was isolated using the Micro-Fast Track kit (InVitrogen, San Diego, CA, U.S.A.). Northern-blot analysis was performed using standard procedures with 10–20 μg of total RNA [25]. Several different RNA isolations and Northern-blot analyses were performed for each tissue.

The amount of MCM mRNA identified by hybridization was normalized to the amount of rRNA in each lane of the Northern blot. Quantification of the rRNA loaded in each lane was performed using a Kodak BioImage 110 to scan a photographic negative of the ethidium bromide-stained RNA gel. The integrated intensity of the 28 S or 18 S RNA bands was determined and was used as a quantitative measure of mRNA loading. The amount of hybridization was determined using a Betascope 603 blot analyser to quantify specific hybridization with a ^{32}P-MCM probe. To compare results from different Northern-blot experiments in which the specific activity of the probe may vary, the amount of MCM mRNA in each experiment is expressed as a fraction of the level observed in a standard liver sample analysed simultaneously. Figure 8 combines data from three different Northern blots representing three to ten animals.

Primer extension was performed using 50 μg of kidney poly(A) RNA and MCM-specific antisense oligonucleotides (no. 207 and 65, Table 1) as primers, and AMV (avian myeloblastosis virus) reverse transcriptase (Promega) with annealing at 50 °C or 55 °C for 4–5 h [25]. The final products were separated on an 8% polyacrylamide sequencing gel containing 7 M urea.

The extreme 5′ end of the cDNA was cloned by PCR amplification of a murine muscle cDNA library in lambdaZap phage (Stratagene) (gift from Dr. Tom Caskey, Baylor College of

Medicine) [11]. PCR primers were designed to bind the murine MCM cDNA (no. 145) and the T3 primer site in the phage (Figure 3). PCR was performed using 1×10^4 plaque-forming units of phage lysates as a template.

Promoter activity in transfected cells

Promoter activity was assayed by subcloning fragments from the genomic clones into the promoter-less β-galactosidase (β-gal) vector, pNAss-β-gal (Figure 6). Constructs were electroporated into STO murine fibroblasts [30] in 300 ml of medium with 3×10^6 cells/ml and 10–20 μg of DNA using settings of 260 V and 960 μF on a GenePulser (Bio-Rad, Richmond, CA, U.S.A.) [21,31]. Promoter activity was assessed 48 h after electroporation by histochemical staining with 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal) (Boehringer Mannheim, Indianapolis, IN, U.S.A.) [32]. Promoter activity was assessed by counting the number of X-gal-stained cells in ten random high-power fields from triplicate samples. Electroporation efficiency did not vary significantly in replicate control experiments performed simultaneously with identical conditions [26,33]. The number of cells exhibiting visible X-gal staining reflects the fraction of cells with expression above the threshold of visual detection and is a valid screening test for the level of gene expression.

In vitro assay of MCM activity

MCM activity was assayed using the succinic thiokinase assay as described [34,35]. Briefly, tissue extracts were homogenized in 0.02 M sodium phosphate, pH 7.2, containing 5 mM EDTA, and centrifuged to remove insoluble material. The supernatant was sonicated, centrifuged (5000 *g*) and the remaining soluble material was used for assays. The assay mixture contained 50 μg of tissue extract, 50 mM Tris/HCl, pH 9.0, 0.21 mM D/L-methylmalonyl-CoA, 30000 d.p.m. D/L-[^{14}C]methylmalonyl-CoA, 20 μl of succinic thiokinase (5 units/ml) (Sigma, St. Louis, MO, U.S.A.), and 18 mM GDP in a volume of 450 μl. Experiments were performed without added adenosylcobalamin to measure holoenzyme activity and in the presence of 0.063 mM adenosylcobalamin (Sigma) to measure total enzyme activity in tissues. The reaction was incubated at 30 °C for 30 min and stopped by addition of 3.0 M HCl, on ice. Samples were extracted with ethyl acetate and a portion of the organic phase was counted for ^{14}C. Results were calculated as the mean of triplicate samples after subtraction of values from a boiled blank reaction and are expressed as nmol of succinate formed/h per mg of protein.

Quantitative analysis

Sequence analysis was performed using the Molecular Biology Information Resource Center at Baylor College of Medicine including the programs QUEST and BLAST. Statistical analysis of correlations between the steady-state level of mRNA and enzyme activity were performed using the Spearman rank test and regression analysis [36].

RESULTS

Isolation and characterization of the mouse MCM gene

We have previously demonstrated that murine MCM is encoded by a single gene at the *Mut* locus [22]. The *Mut* locus was isolated

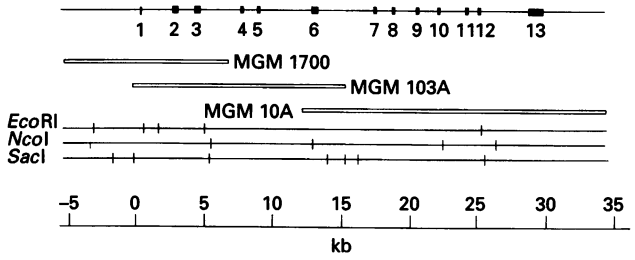

Figure 1 Genomic clones and restriction map of the murine MCM gene

Three overlapping genomic clones spanning the entire gene were isolated. The positions of the 13 exons are shown (boxes) as well as the positions of *Eco*RI, *Nco*I and *Sac*I restriction-endonuclease sites.

Table 2 Intron/exon boundaries of the murine MCM gene

The size of each exon and position relative to the cDNA sequence is shown with the sequence of the intron/exon boundaries and approximate sizes of each intron estimated by restriction mapping. The consensus signal sequences for splice donor and splice acceptor sites is $(T/C)_{>11}N(C/T)AG//G$---exon---$(C/A)AG//GT(A/G)GT$. A GT-rich sequence at the 3′ end of exon 13 representing the likely polyadenylation signal [37] is underlined.

Exon no.	Size (bp)	Position (start–stop)	Intron	(5′----exon----3′)	Intron	Intron size (kb)
1	59	+53– 6	...CCTGCGGCGAGCGGGTGTGA	CCCCG---1--- ACACT	GTGAGTAAGCCCGTGCGCT..	2.5
2	419	7– 425	...TGTTCTCTTTATTTTTCTAG	GACCG---2----TAAGG	GTAAGATTTTCATAGGAGA..	1.2
3	368	426– 793	...GTTTTCTTTATATCCTGTAG	CTGGT---3---CACAG	GTATGCTTGCTTTCTGCAG..	3.0
4	158	794– 951	...GTTTTTATTATTTTAAATAG	CACAT---4----CCAAG	GTCAGTAGATTAACTTATT..	1.1
5	172	952–1123	...TCTGGACTTTGTCCTTTTAG	GTTGT---5----AGCAG	GTATGGATGGAAAAATAGA..	4.0
6	249	1124–1372	...ATTTTGGTCAAATGTTTTAG	GATCC---6----TGAAG	GTCAGATTCTCTTCCGTCC..	4.4
7	112	1373–1484	...AGTTTATTTTTTTTCTTCAG	TTGAT---7----TTCTG	GTAAGAGAGAATAGAAAAA..	1.1
8	116	1485–1600	...TTCCTTTGGGAAATCAACAG	GTTCT---8----AGAAG	GTATTGATGGCGATTGTTT..	1.5
9	116	1601–1716	...TGTTCTCTTTTCTTTCTAAG	ATTAA---9----GCAAG	GTAAGCATGTGGGAGCTGG..	1.5
10	132	1717–1848	...ACCTCTTTTTAATATTGTAG	ATGTA---10---AAGAG	GTACTGTGTGAGCATAGTG..	2.0
11	148	1849–1996	...GCCTTTTAAAATGTTCCCAG	AGTTA---11---TTCAG	GTATGTCACATTTAGAAAA..	0.4
12	168	1997–2164	...TCTGCTTTCTGTGTGCTCAG	ACTCC---12----CACAG	GTATTTTTCCTTCTCTGTA..	4.0
13		2165–	...ATTCTATACTCTGTTAACAG	GATTA---13---AAAAT	GTCTGAAGAACCCTTGTAG TGTGAGTTACTTCACTCAT	

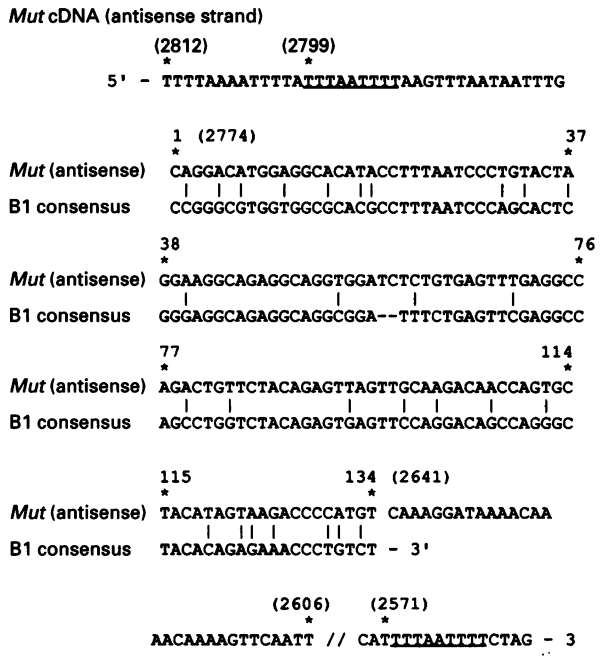

Figure 2 Alignment of B1 repeat element from 3′ untranslated region of MCM cDNA with the B1 consensus sequence

The B1 repeat element is present in an antisense orientation relative to the MCM cDNA. The B1 consensus [38] is shown with the antisense *Mut* sequence (inverse complement of cDNA from bases 2558 to 2812). Vertical lines indicate mismatched base pairs. The underlined sequences indicate direct repeats flanking the B1 element.

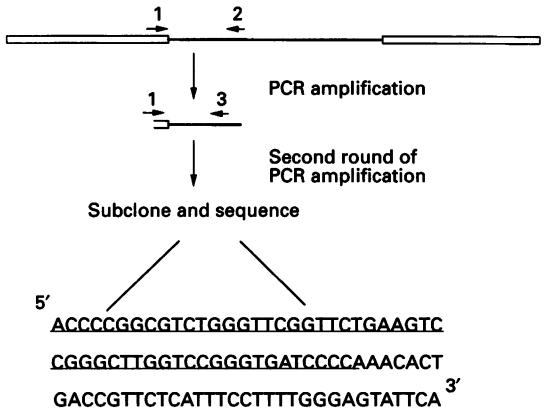

Figure 3 Identification of 5′ end of murine MCM cDNA

Cloning strategy for extreme 5′ end of the murine MCM cDNA. A muscle cDNA library, cloned into the lambda DASH phage vector (Stratagene), was screened by direct PCR amplification using primers T3 (no. 1) and MCM-specific oligonucleotides no. 145 (no. 2) and no. 165 (no. 3). The PCR amplification (primers nos. 1 and 2) produced a 900 bp fragment which was subsequently used for a second round of amplification with an internal MCM primer (no. 3). The resulting 450-base fragment was subcloned into pBSK and sequenced using oligonucleotides nos. 1 and 3. This sequence is shown below with previously unreported 5′ cDNA sequences underlined.

on three overlapping phage clones and characterized by restriction mapping, PCR and sequencing (Figure 1). The locus is composed of 13 exons and 12 introns spanning 30 kb. The exon sequences are co-linear with the cDNA sequence confirming that there is only a single transcription unit for MCM. Consensus

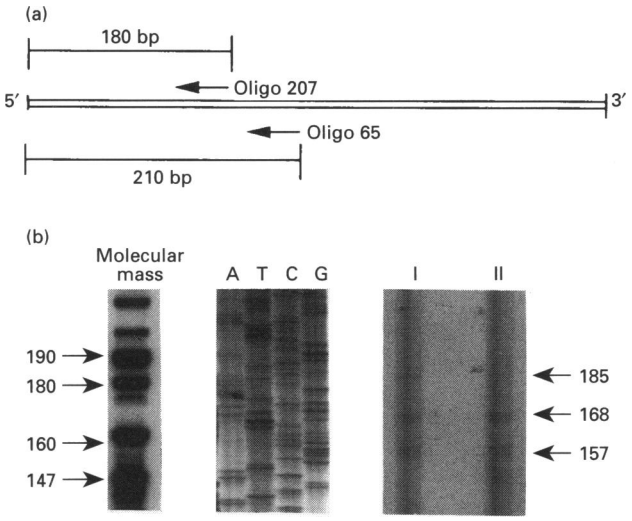

Figure 4 Transcription start sites of murine MCM mRNA

(**a**) Schematic diagram of the RNA primer extension strategy. Reverse transcription of murine RNA with oligonucleotide no. 207 should produce a fragment of approximately 180 bp corresponding to the longest identified cDNA clone. The position of an additional oligonucleotide, which gave consistent results, is also shown. (**b**) Primer extension analysis using oligonucleotide no. 207 and kidney poly(A) RNA. Two primer extension reactions are shown which used slightly different annealing temperatures (I, 50 °C; II, 55 °C). Also shown are molecular-mass markers (pBR322, *Msp*I digest) and a DNA sequencing ladder. Three main extension products were identified using oligonucleotide no. 207 as a primer and may represent alternative initiation sites from this TATA-less promoter.

signals for RNA processing are located adjacent to each exon except at the 5′ end of exon 1 and the 3′ end of exon 13 (Table 2). The position of introns 2–12 corresponded precisely to the intron positions of the homologous human locus [16]. Intron 1 lies within 5′ untranslated sequences which exhibit little sequence identity. Exon 2 is 22 bases longer in mouse than in human, indicating that there has been genetic drift in the position of intron 1.

Southern-blot analysis of mouse genomic DNA using the full-length murine cDNA probe resulted in a smear, suggesting possible repeat sequences [22]. When the probe was reduced to the coding region of the cDNA, only distinct bands corresponding to the MCM gene were identified. Sequence analysis of exon 13 containing the 3′ untranslated region identified a 134 bp B1 repeat element between base pairs 2641 and 2774 of the murine cDNA sequence [21] (Figure 2). The element is situated in an antisense orientation relative to the cDNA and exhibits approx. 79% similarity to the B1 consensus sequence [38]. At the 3′ end of the repeat there is a G/T-rich region (cDNA sequence 2619–2635), typical of other B1 repeats [39].

Identification of the transcription initiation site

The previously reported cDNA sequence contained only seven bases within the putative exon 1 which was insufficient to identify exon 1 in the genomic clones. To obtain probes for exon 1, additional cDNA clones were identified by direct PCR amplification of a cDNA library. The longest clone contained 54 bases of additional 5′ sequences (Figure 3).

To identify the transcription initiation site within the genomic clones, primer extension was performed using the scheme illustrated in Figure 4(a). Three major extension products were identified using an antisense oligonucleotide (no. 207) com-

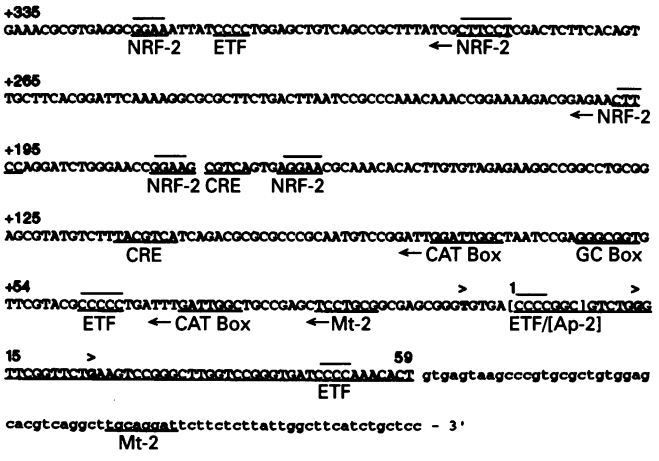

Figure 5 DNA sequence of 5′ flanking region of murine MCM gene

The sequence of the 5′ flanking region was determined from a genomic clone. The position of exon I, predicted from the cDNA sequence, is underlined. Sequences representing intron I are shown in lower case. Potential transcription regulation sites, as defined by sequence homology to consensus motifs (see text for details), are indicated and include: two motifs consistent with inverted CCAAT boxes (CAT Box), a potential Sp-1-binding site (GC Box), two sequences similar to consensus cyclic AMP-response elements (CRE), a cyclic AMP and phorbol ester response element (AP-2), sequence motifs often found in TATA-less genes (ETF) and elements conserved in the 5′ flanking region of several nuclear-encoded mitochondrial proteins (NRF-2 and Mt-2). Sequence elements with an arrow (←) are reverse complements of their consensus sequences; all others are in the same orientation. Putative transcription initiation sites based on primer extension data are indicated by >. The first base pair of exon 1, as determined from the longest cDNA clone, is marked as (1), and the upstream sequences are designated as +1-+335.

plementary to the murine cDNA as a primer (Figure 4b). The longest extension product of 185 bases corresponds to a transcript initiating five nucleotide residues upstream from the longest cDNA clone (Figure 3). The shorter extension products (168 and 157 bases) may represent alternative start sites for the TATA-less promoter, although these do not correspond to cDNAs isolated in our library or PCR cloning of the cDNA. An additional oligonucleotide was used (Figure 4a) to confirm the primer extension results and gave consistent results (not shown).

Sequence motifs within the 5′ upstream sequences and confirmation of promoter function

The extended cDNA clones were used to identify exon 1 in the genomic clones. Figure 5 shows the sequence of exon 1 (underlined) along with 335 bp of 5′ flanking DNA. Exon 1 is not preceded by a slice acceptor site, but rather a GC-rich region (60% G+C), exhibiting many sequence motifs characteristically associated with CpG islands and promoters of housekeeping genes [40–42]. The CpG/GpC ratio of this region is about 1.3 which is characteristic of CpG islands. By comparison, the CpG/GpC ratio of exon II (45% G+C) is 0.27 whereas the predicted genomic average in mammals is 0.2. The 5′ upstream sequences lack a consensus TATA box but contain at least one GC box (Figure 5) and two CCAAT boxes ('CAT Box', Figure 5) in the reverse complement orientation at 30 and 70 bp upstream of the first transcription start site. Several other consensus elements were identified by computer search including: (1) two sequences resembling cyclic AMP-response elements (CRE) [43], (ii) a putative Ap-2 element, the binding site for the cyclic AMP and phorbol ester responsive factor, Ap-2 [44] and (iii) multiple putative ETF consensus sequences [45].

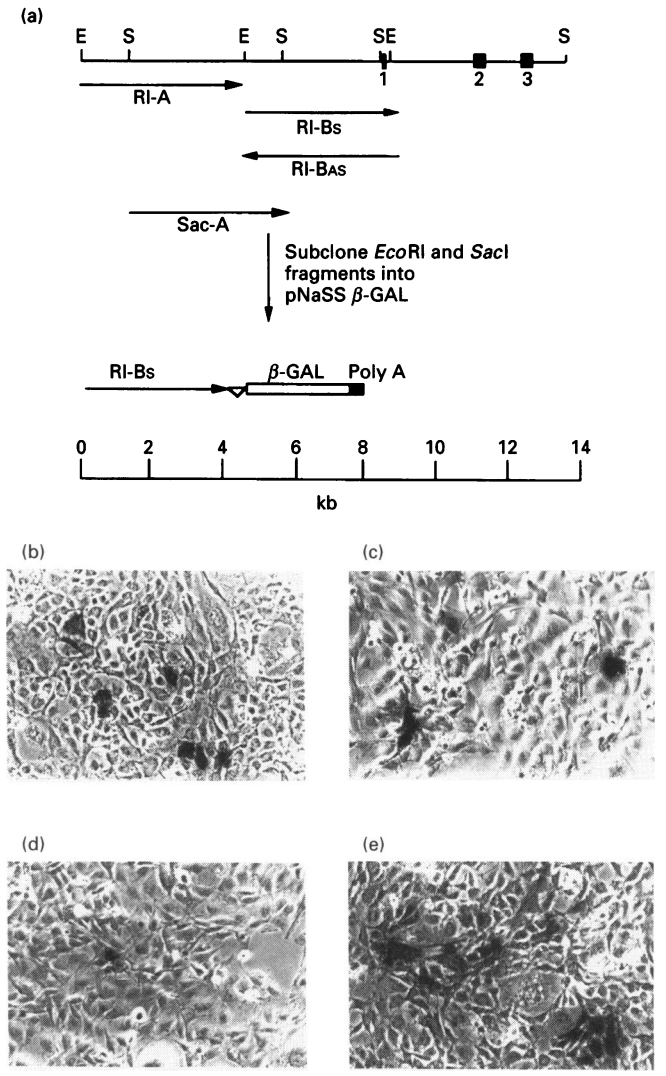

Figure 6 Promoter activity in 5′ flanking DNA of murine MCM gene

(**a**) Constructs used to assess promoter function of 5′ flanking regions. Restriction fragments from the genomic clone were subcloned into the promoterless reporter construct pNaSS β-gal which contains the SV40 splice donor/acceptor sequences 5′ to the *E. coli* β-gal followed by the SV40 polyadenylation signal. Constructs included: (i) the 5 kb *Eco*RI A fragment (RI-A, sense orientation), (ii) the 4 kb *Eco*-RI B fragment (RI-B_S, sense orientation and RI-B_{AS}, antisense orientation) and (iii) the 4.5 kb, *Sac*I A fragment (Sac-A, sense orientation). The RI-B fragment contains exon 1 and the putative promoter sequence. These constructs were electroporated into STO cells to assess the presence or absence of promoter function in each fragment. (**b**) Cells transfected with RI-B_S; (**c**) cells transfected with RI-B_{AS}; (**d**) cells transfected with empty expression vector; (**e**) cells transfected with positive control vector, pCMV β-gal.

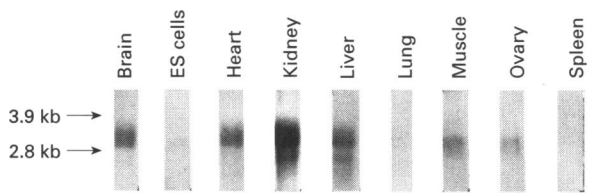

Figure 7 Steady-state levels of MCM mRNA in murine tissues

Northern blots were performed using total RNA from various tissues probed with a 750-base fragment (*Eco*RI–*Cla*I) from the 5′ end of the murine MCM cDNA. Equal RNA loading between the samples was verified by quantification of the 28 S and 18 S rRNA bands on the photographic negative of the agarose gel.

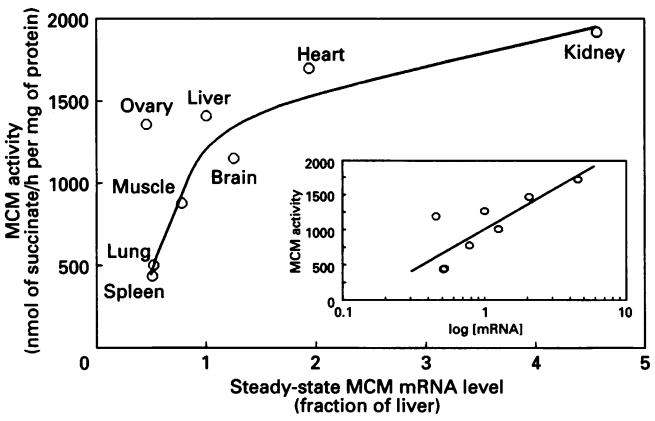

Figure 8 Relationship between steady-state level of MCM mRNA and total MCM enzyme activity in different murine tissues

MCM activity was measured in the presence of excess adenosylcobalamin and plotted against the steady-state levels of MCM mRNA (expressed as a fraction of the level found in liver). The steady-state level of MCM mRNA was determined by Northern-blot analysis and expressed as a fraction of the level in liver. There is a significant rank correlation between the steady-state level of MCM mRNA and total enzyme level ($P < 0.01$). A log–linear correlation is evident in the blot of log[mRNA] versus enzyme activity and is statistically significant ($r^2 = 0.92$, $P < 0.001$) (insert).

Table 3 Relationship between total enzyme and holoenzyme in different murine tissues

MCM activity (nmol of succinate/h per mg of protein) was measured in different tissues in the presence (total enzyme) or absence (holoenzyme) of excess adenosylcobalamin as described in the Experimental section. The results are means ± S.D. for three samples. The ratio of holoenzyme/total enzyme is indicated as the percentage holoenzyme. There was no correlation between holoenzyme and steady-state levels of MCM mRNA shown in Figure 8.

Tissue	Holoenzyme activity	Total enzyme activity	Holoenzyme (% of total)
Muscle	217 ± 20	788 ± 167	28
Ovary	415 ± 39	1354 ± 176	31
Heart	532 ± 43	1593 ± 296	33
Liver	510 ± 53	1408 ± 110	36
Kidney	764 ± 65	1871 ± 175	41
Lung	225 ± 39	468 ± 63	48
Spleen	285 ± 19	441 ± 111	65
Brain	1058 ± 150	1103 ± 148	96

The 5′ end of the mouse MCM gene also contains sequence motifs common to other nuclear-encoded mitochondrial genes. Four common motifs (Mt-1, Mt-2, Mt-3, Mt-4) have been identified in the 5′ region of chicken aminolaevulinate synthase, human cytochrome c_1 and ATP synthase β subunit [46–48]. Two sequences resembling Mt-2 sequence (Figure 5) were found in the 5′ end of the mouse MCM gene. One occurs as a reverse complement of the consensus (at nucleotide +15) and the other was found in the consensus orientation, 37 nucleotides downstream of the 3′ end of exon I. No Mt-1, Mt-3 or Mt-4 motifs are evident in this region. In addition, five copies of the sequence 5′-GGAA-3′, or its reverse complement, were found within 300 nucleotides of the transcription start site ('NRF-2' in Figure 5). This sequence motif has been identified as important in binding the nuclear factor NRF-2 in nuclear-encoded mitochondrial proteins (NEMP) including cytochrome *c* oxidase subunit IV, ATP synthase β subunit and cytochrome *c* oxidase subunit Vb [49,50].

To confirm whether the 5′ upstream sequences exhibited promoter activity, restriction fragments from this region were subcloned into the promoterless reporter vector pNaSS β-gal and electroporated into mouse fibroblasts (Figure 6a). The number of cells expressing β-gal during the transient phase of gene expression after electroporation was used as an indication of promoter activity in these various constructs. Representative fields are shown in Figures 6(b)–6(e). Transfection of cells with the construct containing the putative promoter and exon 1 in the sense orientation (RI-B_S, Figure 6a) produced significantly more X-gal-positive cells than the control constructs (RI-B_{AS}, RI-A and Sac-A; Figure 6a). This suggests that this region, in fact, exhibits orientation-dependent promoter activity similar to that observed previously with the human MCM promoter [16].

Tissue-specific patterns of MCM mRNA and enzyme activity

The steady-state level of mouse MCM mRNA in various tissues was assessed by Northern blotting (Figure 7). A single major species of mRNA of 3.3–3.4 kb was observed which is consistent with the length of the cDNA. A smaller band of 2.8 kb representing less than 10% of MCM hybridization is apparent in kidney and liver and to a lesser extent in heart. There was a more than 10-fold variation in the level of hybridizing MCM mRNA in different tissues expressed relative to the steady-state level of mRNA in the liver (Figure 8). Kidney consistently had levels of MCM mRNA 2–5 times higher than liver, with 5–10-fold lower levels evident in spleen and lung.

MCM enzyme activity was assayed in extracts from the same tissues as used for mRNA analysis. This assay was performed in the absence of exogenously added adenosylcobalamin to measure only holoenzyme and in the presence of excess adenosylcobalamin (0.063 mM) to measure total enzyme (apoenzyme + holoenzyme). Figure 8 shows a graph of enzyme activity (nmol of succinate formed/h per mg of protein) plotted against the steady-state levels of MCM RNA (calculated as the fraction of MCM mRNA in the liver). There was a rank-order correlation between the relative level of enzyme activity and mRNA in each tissue ($P < 0.01$), and a log–linear regression (Figure 8, insert) gave a highly significant coefficient of correlation ($r^2 = 0.92$). Control experiments (not shown) demonstrate that the enzyme assay was still linear in the range used for these assays.

Variations in the ratio of holoenzyme to apoenzyme in different tissues are shown in Table 3. Significantly, virtually all of the enzyme in brain was holoenzyme. In other tissues this fraction ranged from 0.3 to 0.6 which is consistent with previous observations in rodents and sheep [5,6,51]. Because of these variations in holoenzyme/total enzyme, there was no correlation between the steady-state level of mRNA and the fraction of active enzyme (holoenzyme) in different tissues.

DISCUSSION

This paper describes the cloning and characterization of the murine *Mut* locus including its promoter, as well as evidence for both genetic and epigenetic regulation of this enzyme's activity. These studies demonstrate an unexpected, more than 10-fold, variability in the steady-state level of mRNA in different murine tissues which parallels the level of MCM enzyme in these tissues. The relationship between the level of MCM mRNA, gene product and enzyme activity is complicated, however, by epigenetic regulation of holoenzyme formation. While there is a clear correlation between the steady-state levels of mRNA and total

enzyme in different tissues, there is no correlation between the steady-state level of MCM mRNA and the level of enzyme activity.

The cloning and characterization of murine MCM was undertaken to develop the mouse as a model for molecular genetic investigations of the role that MCM plays in homoeostasis and disease. Previous studies have demonstrated that murine MCM exhibits kinetic properties indistinguishable from its human homologue, that these proteins exhibit considerable sequence identity [21], and that the human and murine loci are syntenic [22]. The present studies extend these observations by demonstrating that the genomic structures of human [16] and murine MCM are similarly homologous, with all of the intron/exon boundaries within the open reading frame occurring at identical positions. Both the human and murine loci have a relatively large second exon (human 424 bases, murine 419 bases) containing 5′ untranslated sequences, the entire mitochondrial leader sequence and 96 amino acids of the mature peptide. In both the human and murine loci, there are introns within the 5′ untranslated regions, although there is some genetic drift of the position of this intron relative to the AUG in exon 2.

One difference between the mouse and human loci is that the murine exon 13 is twice as large as the human (murine 950 bases, human 440 bases). This difference is accounted for, in part, by the presence of the B1 repeat in the mouse 3′ untranslated region. B1 repeat elements are the most abundant (40000 to 80000 copies) repetitive sequence in the mouse genome and have been estimated to occur every 6–10 kb [52]. B1 is often found in the 3′ untranslated region of genes [38,53,54] and transcription of these sequences can affect transcriptional or post-transcriptional processing and the steady-state level of mRNA [55].

MCM has been considered a ubiquitous housekeeping enzyme similar to aspartate aminotransferase [56] and malate dehydrogenase [57], the expression of which is not regulated in a tissue-specific or temporal fashion. Housekeeping gene promoters contain predictable sequence motifs [42,58] evident in MCM including the absence of a TATA box, high G+C content and high CG/GC ratio. The murine MCM promoter exhibits several features which are somewhat atypical for housekeeping promoters including: two reverse complement CCAAT boxes, two sequences resembling CREs and a sequence motif homologous to an Ap-2 binding site. It is not uncommon, however, for TATA-less genes of NEMPs to contain similar regulatory elements. For example, the human cytochrome c_1 gene [46] contains several AP-1-responsive elements [59] as well as a sequence homologous to the cyclic AMP- and phorbol ester-inducible element of the proenkephalin gene [60] whereas the human ubiquinone-binding protein gene [47] contains four putative CAAT boxes and one sequence homologous to the AP-1-responsive element as well. The rat cytochrome *c* oxidase subunit IV [61] and mouse cytochrome *c* oxidase subunit Vb [62] promoters contain several similar regulatory motifs.

Four distinct sequence elements (Mt-1, 2, 3 and 4) have been identified which are common to the 5′ flanking region of NEMPs [46,47] and which may play a role in co-ordinating nuclear and mitochondrial gene expression. Similar sequences have been identified in other NEMPs (human acyl-CoA dehydrogenase and the bovine α subunit of ATP synthase gene) [63,64]. Two nuclear factors have been identified (NRF-1, NRF-2) whose binding to sequences in the 5′ end of several nuclear-encoded mitochondrial genes correlates with promoter activity [50,65–68]. The site recognized by NRF-2 is the consensus binding sequence for ETS-domain proteins which are a diverse family of transcriptional activators [49,69]. Finally, two binding sites (OXBOX and REBOX) for nuclear factors have been identified in the promoters of the ATP synthase β subunit and the muscle-specific adenine translocator (ANT1) genes [70,71] which may play a role in metabolic regulation of these genes. The MCM gene does not exhibit consensus sites for either OXBOX or REBOX. There are consensus elements for several putative mitochondrial co-ordinating elements (NRF-2 and Mt-2). The role that any of these sequences may play in regulating MCM expression remains to be investigated.

In the past few years, a growing number of so-called housekeeping, or more accurately TATA-less, genes have been identified with numerous regulatory promoter elements as well as patterns of RNA expression that vary considerably in a tissue-specific or developmental manner [46,47,63,64,72–75]. Like these genes, the murine MCM promoter may represent an intermediate between highly regulated TATA-containing genes and the classic TATA-less ubiquitously expressed housekeeping gene.

The presence of potential regulatory elements in the promoter and the tissue-specific variation in steady-state mRNA levels suggests that there could be significant genetic regulation of MCM. It is unclear, however, whether this genetic regulation plays a significant role in the capacity for metabolic flux through this enzyme. Variability in the level of MCM enzyme activity has been described previously. In sheep and rats, the highest levels of MCM activity were reported in the kidney and liver followed by brain, intestinal mucosa and skeletal muscle [4–6,51]. Our data confirm that MCM activity in different tissues varies and is significantly regulated by the ratio of holoenzyme/total enzyme. Our data further demonstrate that the level of holoenzyme in different tissues does not correlate with the steady-state level of mRNA. Furthermore, our previous studies have also demonstrated that, in cultured cells, MCM is not a rate-limiting enzyme in propionate flux, that constitution of low levels of MCM activity in genetically deficient cells by gene transfer restores levels of propionate flux to normal [21,26], and that over-expression of MCM in fibroblasts and hepatoma cells does not increase propionate flux above normal [33].

There are precedents for a correlation between the energy requirements of tissues and the steady-state RNA levels of critical rate-limiting enzymes as there is for enzymes involved in fatty acid oxidation [63,76]. The teleological speculation that high levels of MCM in kidney have evolved to provide a necessary enzyme activity is intriguing in light of evidence that renal failure is a significant complication of chronic MCM deficiency in man [77,78]. If MCM is not a rate-limiting enzyme, or epigenetic regulation of holoenzyme assembly affects the level of activity, then the significance of genetic regulation of total enzyme and the nature of the selective pressures that may have led to this pattern of tissue-specific regulation are not immediately clear. The observation that brain MCM is present entirely as holoenzyme whereas in other tissues it is only 20–40% holoenzyme may be significant with regard to the syndrome of B12-responsive neuropsychiatric disorders, disorders that do not correlate with any demonstrable systemic defect in MCM or methionine synthase activity [79].

The present study identifies several potential genetic and epigenetic mechanisms that may regulate MCM activity in a tissue-specific manner. Further studies will be required to understand the mechanisms of genetic and epigenetic regulation, the dynamic role of these mechanisms and their interactions, and the role of these regulatory events in maintaining homoeostasis and in the pathogenesis of disease.

We thank Dr. Randy Smith for assistance in sequence analysis, Dr. Paul Overbeek and C. Thomas Caskey for contributing cDNA and genomic libraries used in this work, and Eleanor Lo for participating in this work as a summer student. This work

was supported by grants HD-24186 and HD-24064 from the National Institutes of Health.

REFERENCES

1 Bergman, E. N. (1990) Phys. Rev. **70**, 567–590
2 Wolin, M. J. (1981) Science **213**, 1463–1468
3 Bremer, J. and Osmundsen, H. (1984) in Fatty Acid Metabolism and its Regulation (Numa, S., ed.), pp. 113–154, Elsevier Science Publishers, New York
4 Beck, W. S., Flavin, M. and Ochoa, S. (1957) J. Biol. Chem. **229**, 997–1010
5 Peters, J. P. and Elliot, J. M. (1984) J. Nutr. **114**, 660–670
6 Scott, J. S. D., Treston, A. M., Bowman, E. P. W., Owens, J. A. and Cooksley, W. G. E. (1984) Clin. Sci. **76**, 299–306
7 Rosenberg, L. E. and Fenton, W. A. (1989) in The Metabolic Basis of Inherited Disease (Scriver, C. R., Beudet, A. L., Sly, W. S. and Valle, D., eds.), 6th edn., pp. 822–844, New York, McGraw-Hill
8 Kolhouse, J. F., Utley, C. and Allen, R. H. (1980) J. Biol. Chem. **255**, 2708–2715
9 Fenton, W. A., Hack, A. M., Willard, H. F., Gertler, A. and Rosenberg, L. E. (1982) Arch. Biochem. Biophys. **214**, 815–818
10 Ledley, F. D., Lumetta, M., Nguyen, P. N., Kolhouse, J. F. and Allen, R. A. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 3518–3521
11 Jansen, R., Kalousek, F., Fenton, W. A., Rosenberg, L. E. and Ledley, F. D. (1989) Genomics **4**, 198–205
12 Isaya, G., Kalousek, F., Fenton, W. A. and Rosenberg, L. E. (1991) J. Cell Biol. **113**, 65–76
13 Ledley, F. D. and Leadley, P. F. (1990) in Cobalamin '88, Proceedings of the First International Symposium on Biomedicine and Physiology of Vitamin B12 (Linnell, J. C. and Bhatt, H. R., eds.), pp. 343–354, Children's Medical Charity, London
14 Ledley, F. D., Lumetta, M. L., Zoghbi, H. Y., Van Tuinen, P. and Ledbetter, D. (1988) Am. J. Hum. Genet. **42**, 839–846
15 Zoghbi, H., O'Brien, W. E. and Ledley, F. D. (1988) Genomics **3**, 396–398
16 Nham, S., Wilkemeyer, M. F. and Ledley, F. D. (1990) Genomics **8**, 710–716
17 Andrews, E., Jansen, R., Crane, A. M., Cholin, S., McDonnell, D. and Ledley, F. D. (1993) Biochem. Med. Metab. Biol. **50**, 135–144
18 Ledley, F. D. (1990) Bioessays **12**, 335–340
19 Jansen, R. and Ledley, F. D. (1990) Am. J. Human. Genet. **47**, 808–814
20 Crane, A. M., Martin, L., Valle, D. and Ledley, F. D. (1992) Hum. Genet. **89**, 259–264
21 Wilkemeyer, M. F., Crane, A. M. and Ledley, F. D. (1990) Biochem. J. **271**, 449–455
22 Threadgill, D. W., Wilkemeyer, M. F., Womack, J. E. and Ledley, F. D. (1990) Cytogenet. Cell Genet. **53**, 112–114
23 Sertic, J., Vineck, V., Ledley, F. D. and Klein, J. (1990) Genomics **6**, 560–566
24 Berger, S. and Kimmel, A. (1987) Methods Enzymol. **153**, 41–112
25 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
26 Wilkemeyer, M. F., Crane, A. M. and Ledley, F. D. (1991) J. Clin. Invest. **87**, 915–918
27 Innis, M., Gelfand, D., Sninsky, J. and White, T. (1990) PCR Protocols: A Guide to Methods and Applications, Academic Press, San Diego
28 Robertson, E. L. (1987) in Teratocarcinomas and Embryonic Stem Cells: A Practical Approach (Roberbson, E. J., ed.), pp. 71–112, IRL Press, Oxford
29 McMahon, A. P. and Bradley, A. (1990) Cell **62**, 1073–1085
30 Gunning, P., Ponte, P., Okayama, H., Engel, J., Blau, H. and Kedes, K. (1983) Mol. Cell. Biol. **3**, 787–795
31 Shigekawa, K. and Dower, W. J. (1988) Biotechniques **6**, 742–751
32 MacGregor, G. R., Nolan, G. P., Fiering, S., Roederer, M. and Herzenberg, L. A. (1991) Methods Mol. Biol. **7**, 217–235
33 Wilkemeyer, M. F., Stankovics, J., Foy, T. and Ledley, F. D. (1992) Somat. Cell Mol. Genet. **18**, 493–505
34 Kovachy, R. J., Stabler, S. P. and Allen, R. H. (1988) Methods Enzymol. **166**, 393–400
35 Kolhouse, J. F., Stabler, S. P. and Allen, R. H. (1988) Methods Enzymol. **166**, 407–414
36 Snedecor, G. W. and Cochran, W. G. (1972) in Statistical Methods, pp. 352–354, Iowa State University Press, Iowa
37 Proudfoot, N. (1991) Cell **64**, 671–674
38 Kalb, V. F., Glasser, S., King, D. and Lingrel, J. B. (1983) Nucleic Acids Res. **11**, 2177–2184
39 King, D., Snider, L. D. and Lingrel, J. B. (1986) Mol. Cell. Biol. **6**, 209–217
40 Bird, A. P. (1987) Trends Genet. **3**, 342–347
41 Dynan, W. S. (1986) Trends Genet. **2**, 196–197
42 Dynan, W. S., Sazre, S., Tijan, R. and Scmicke, R. T. (1986) Nature (London) **319**, 246–248
43 Montminy, M. R., Gonzalez, G. and Yamamoto, K. K. (1990) Trends Neurosci. **13**, 184–188
44 Roesler, W., Vandenbark, G. R. and Hansen, R. W. (1988) J. Biol. Chem. **263**, 9063–9066
45 Kageyama, R., Merlino, G. T. and Pastan, I. (1989) J. Biol. Chem. **264**, 15508–15514
46 Suzuki, H., Hosokawa, Y., Nishikimi, M. and Ozawa, T. (1989) J. Biol. Chem. **264**, 1368–1374
47 Suzuki, H., Hosokawa, Y., Toda, H., Nishikimi, M. and Ozawa, T. (1990) J. Biol. Chem. **265**, 8158–8163
48 Suzuki, H., Hosokawa, Y., Nishikimi, M. and Ozawa, T. (1991) J. Biol. Chem. **266**, 2333–2338
49 Virbasius, J. V. and Scarpulla, R. C. (1990) Nucleic Acids Res. **18**, 6581–6588
50 Virbasius, J. V. and Scarpulla, R. C. (1991) Mol. Cell. Biol. **11**, 5631–5638
51 Kennedy, D. G., Cannavan, A., Molloy, A., O'Harte, F., Taylor, S. M., Kennedy, S. and Blanchflower, W. J. (1990) Br. J. Nutr. **64**, 721–732
52 Krayev, A. S., Kramerov, D. A., Skryabin, K. G., Ryskov, A. P., Bayev, A. A. and Georgiev, G. P. (1980) Nucleic Acids Res. **8**, 1201–1215
53 Ryskov, A. P., Ivonov, P. L., Kramerov, D. A. and Georgiev, G. P. (1983) Nucleic Acids Res. **11**, 6541–6558
54 Zechner, R., Newman, T. C., Steiner, E. and Breslow, J. L. (1991) Genomics **11**, 62–76
55 Jagadeeswaran, P., Forget, B. G. and Weissman, S. M. (1981) Cell **26**, 141–142
56 Tsuzuki, T., Obaru, K., Setoyama, C. and Shimanda, K. (1987) J. Mol. Biol. **148**, 21–31
57 Takeshima, H., Joh, T., Tsuzuki, T., Shimanda, K. and Matsukado, Y. (1988) J. Mol. Biol. **200**, 1–11
58 Smale, S. T. and Baltimore, D. (1989) Cell **57**, 103–113
59 Bohmann, D., Bos, T. J., Admon, A., Nishimura, T., Vogt, P. K. and Tijan, R. (1987) Science **238**, 1386–1392
60 Comb, M., Birnberg, N. C., Seashotz, A., Herbert, E. and Goodman, H. M. (1986) Nature (London) **323**, 353–356
61 Yamada, M., Amuro, N., Goto, Y. and Okazaki, T. (1990) J. Biol. Chem. **265**, 7687–7692
62 Basu, A. and Avadhani, N. G. (1991) J. Biol. Chem. **266**, 15450–15456
63 Zhang, Z., Kely, D., Kim, J., Zhou, Y., Ogden, M., Whelan, A. and Strauss, A. (1992) Biochemistry **31**, 81–89
64 Pierce, D. J., Jordan, E. M. and Breen, G. A. M. (1992) Biochim. Biophys. Acta **1132**, 265–275
65 Evans, M. J. and Scarpulla, R. C. (1989) J. Biol. Chem. **264**, 14361–14368
66 Evans, M. J. and Scarpulla, R. C. (1990) Genes Dev. **4**, 1023–1034
67 Chau, C. M., Evans, M. J. and Scarpulla, R. C. (1992) J. Biol. Chem. **267**, 6999–7006
68 Virbasius, J. V., Virbasisu, C. A. and Scarpulla, R. C. (1993) Genes Dev. **7**, 380–392
69 Karim, F. D., Urness, L., Thummel, C., Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C., Maki, R., Gunther, C. V., Nye, J. A. and Graves, B. J. (1990) Genes Dev. **4**, 1451–1453
70 Li, K., Hodge, J. A. and Wallace, D. C. (1990) J. Biol. Chem. **265**, 20585–20588
71 Chung, A. C., Stepien, G., Haraguchi, Y., Li, K. and Wallace, D. C. (1992) J. Biol. Chem. **267**, 21154–21161
72 Brown, A. and Rechler, M. M. (1990) Mol. Endocrinol. **4**, 2039–2051
73 Werner, H., Woloschak, M., Adamo, M., Shen-Orr, Z., Roberts, C. T. and LeRoith, D. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 7451–7455
74 Werner, H., Stannard, B., Bach, M. A., LeRoith, D. and Roberts, C. T. (1990) Biochem. Biophys. Res. Commun. **169**, 1021–1027
75 Gross, B., Misrahi, M., Sar, S. and Milgrom, E. (1991) Biochem. Biophys. Res. Commun. **177**, 679–687
76 Kelly, D. P., Gordon, J. I., Alpers, R. and Strauss, A. W. (1989) J. Biol. Chem. **264**, 18921–18925
77 Molteni, K., Oberley, T. D., Wolff, J. A. and Friedman, A. L. (1991) Pediatr. Nephrol. **5**, 323–326
78 Walter, J. H., Michalski, A., Wilson, W. M., Leonard, J. V., Barrah, T. M. and Dillon, M. J. (1989) Eur. J. Pediatr. **148**, 344–348
79 Lindenbaum, J., Healton, E. B., Savage, D. G., Brust, J. C., Garrett, T. J., Podell, E. R., Marcell, P. D., Stabler, S. P. and Allen, R. H. (1988) N. Engl. J. Med. **318**, 1720–1728

Received 17 March 1993/21 July 1993; accepted 29 July 1993