Cloning and characterization of the promoter for the liver isoform of the rat carnitine palmitoyltransferase I (L-CPT I) gene

Edwards A. PARK^{*1}, Michelle L. STEFFEN^{*}, Shulan SONG^{*}, Vicki M. PARK[†] and George A. COOK^{*}

^{*}Department of Pharmacology, University of Tennessee, College of Medicine, 874 Union Avenue, Memphis, TN 38163, U.S.A., and [†]Department of Obstetrics and Gynecology, College of Medicine, University of Tennessee, Memphis, The Health Science Center, Memphis, TN 38163, U.S.A.

Carnitine palmitoyltransferase I (CPT I) catalyses the transfer of long chain fatty acids to carnitine for translocation across the mitochondrial inner membrane. The cDNAs of two isoforms of CPT I, termed the hepatic and muscle isoforms, have been cloned. Expression of the hepatic CPT I gene (L-CPT I) is subject to developmental, hormonal and tissue specific regulation. We have cloned the promoter of the L-CPT I gene from a rat genomic library. In the L-CPT I gene, there are two exons 5' to the exon containing the ATG that initiates translation. Exon 1 and the 5' end of exon 2 contain sequences that were not previously described in the rat L-CPT I cDNA. There is an

alternatively spliced form of the L-CPT I mRNA in which exon 2 is skipped. The proximal promoter of the L-CPT I gene is extremely GC rich and does not contain a TATA box. There are several putative Sp1 binding sites near the transcriptional start site. A 190 base pair fragment of the promoter can efficiently drive transcription of luciferase and CAT (chloramphenicol acetyltransferase) reporter genes transiently transfected into HepG2 cells. Sequences in both the first intron and the promoter contribute to basal expression. Our results provide the foundation for further studies into the regulation of L-CPT I gene expression.

INTRODUCTION

The rate controlling step in the pathway of fatty acid oxidation in the liver is the formation of acylcarnitines that are required for the transport of long chain fatty acids into the mitochondria [1]. The reaction catalysed by carnitine palmitoyltransferase I (CPT I) is the transfer of the acyl moiety from acyl-CoA to acylcarnitine in the outer mitochondrial membrane [1]. Acylcarnitine is translocated across the mitochondrial membrane in exchange for carnitine by carnitine acylcarnitine translocase and then re-esterified with CoA by the inner mitochondrial membrane carnitine palmitoyltransferase (CPT II) [1]. CPT I is inhibited by malonyl-CoA which is a substrate for fatty acid synthesis [2]. The sensitivity of CPT I to malonyl-CoA inhibition is increased by insulin but diminished in diabetes [3]. CPT II is not regulated by malonyl-CoA [4].

Based upon studies with inhibitors of the CPT I enzyme and by cDNA cloning, two isoforms of CPT I have been identified [5,6]. The cDNA for the liver isoform (L-CPT I) was cloned first from a rat liver cDNA library [5], and the muscle isoform (M-CPT I) was subsequently cloned from rat brown adipose tissue and rat heart libraries [6,7]. The cDNAs for the human isoforms of L-CPT I and M-CPT I have been described [8,9]. The liver and muscle isoforms share 63% identity at the amino acid level [1]. A partial human genomic clone for the liver isoform has been described and the human M-CPT I gene has been completely cloned [8,10]. The chromosomal locations of the human liver and muscle CPT I genes were found to be 11q13 and 22q13.3, respectively [11]. The cDNA for CPT II was originally cloned from a rat library and the CPT II gene has been characterized in

mouse and human [12–14]. There is considerable evidence for the presence of additional CPT activities in peroxisomes and microsomes based upon cDNA cloning, protein purification and antibody studies [15–17].

L-CPT I gene expression and enzyme activity increase dramatically in the liver at birth [18]. With weaning to a high fat diet, transcription of L-CPT I remains high while weaning to a high carbohydrate diet results in diminished L-CPT I expression [18]. These observations suggest an important role for hormones and diet in controlling L-CPT I expression. Transcription of the gene is inhibited by insulin but stimulated by cAMP, fatty acids and peroxisomal proliferators [19]. L-CPT I mRNA abundance is increased five-fold and 14-fold in hyperthyroidism and streptozotocin induced diabetes, respectively [3,20]. Cardiac muscle contains both CPT I isoforms, with the muscle isoform primarily expressed in the adult heart while L-CPT I is highly expressed before birth [7,21]. Following birth, L-CPT I activity in the heart declines while the abundance of the muscle isoform increases [21]. CPT II is expressed before birth and its activity and mRNA abundance are unchanged by fatty acids or insulin [18,19]. Only clofibrate is able to stimulate CPT II gene expression [19].

In order to understand the mechanisms underlying the changes in L-CPT I expression, we cloned the promoter of the rat L-CPT I gene. Here, we report the initial characterization of the rat L-CPT I gene and demonstrate that the L-CPT I promoter can efficiently drive expression of luciferase and CAT reporter genes in transient transfection assays. These studies will provide the basis for further studies of hormonal and developmental regulation of the L-CPT I gene.

Abbreviations used: CPT, carnitine palmitoyltransferase; CPT I, carnitine palmitoyltransferase I; CPT II, carnitine palmitoyltransferase II; L-CPT I, liver isoform of carnitine palmitoyltransferase I; M-CPT I, muscle isoform of carnitine palmitoyltransferase I; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription PCR; CAT, chloramphenicol acetyltransferase; Luc, luciferase; T3, 3,5,3'-tri-iodothyronine; PEPCK, phosphoenolpyruvate carboxykinase; PPAR, peroxisomal proliferator-activated receptor; DR4, direct repeat separated by four nucleotides; C/EBP, CCAAT enhancer binding protein; CRE, cAMP responsive element; CREB, cAMP responsive element binding protein; AP-2, activator protein-2; DTT, dithiothreitol.

¹ To whom correspondence should be addressed.

The sequence of the promoter, exon 1, intron 1 and exon 2 have been submitted to GenBank under the accession number AF020776.

EXPERIMENTAL

Screening of the rat λ genomic library

A rat λ genomic library in EMBL-3 cloning vector (Clontech, Palo Alto, CA, U.S.A.) was screened essentially as outlined in Ausubel et al. [22]. The *Escherichia coli* host strain was K802. The library was plated out on agar plates in NZYDT media and duplicate lifts were made with BA-S 85 nitrocellulose filters (Schleicher and Schuell, Keene, NH, U.S.A.). The filters were probed with a ^{32}P -labelled CPT cDNA fragment that contained 400 nucleotides from the initiating ATG to +400 in the coding sequence [5]. The probe was labelled with [α - ^{32}P]dCTP (6000 Ci/mmol) using a random priming kit (Stratagene, La Jolla, CA, U.S.A.). Plaques which hybridized with the probe were selected from the plates and subjected to two additional rounds of plating and purification. Lambda DNA was isolated from the purified clones using lambda DNA purification kits (Qiagen, Santa Clarita, CA, U.S.A.). The lambda DNA was analysed by Southern analysis for L-CPT I genomic sequences. Appropriate DNA fragments were subcloned into pBluescript for sequencing.

Analysis of the P1 rat genomic clones

Two P1 clones were obtained from Genome Systems (St. Louis, MO, U.S.A.). Two PCR primers were prepared based on a sequence obtained from sequencing the rat genomic L-CPT I clones λ 1A and λ 17B from the lambda library. The forward primer, gggcagaactgtagctgtgacc, hybridized with a sequence in intron 2, and the reverse primer, aactctctctccaggactgcaagc, annealed with exon 3. In a PCR reaction using rat genomic DNA as a template, these oligomers generated a 200 bp product. These oligomers were provided by Genome Systems. The P1 clones were grown in the *E. coli* host NS 3156 as recommended by the supplier. The DNA from the P1 clone was digested with various restriction enzymes and the resulting fragments were cloned into pBluescript KS+ for sequencing and Southern analysis.

Sequencing of genomic DNA

DNA was sequenced with the Sequenase 2.0 kit (Amersham/USB, Cleveland, OH, U.S.A.). In addition, sequencing reactions were conducted with the Dye terminator cycle sequencing system (Perkin-Elmer, Foster City, CA, U.S.A.) and analysed at the Molecular Resource Center of the University of Tennessee.

Primer extension analysis

Primer extension reactions were carried out as follows. Oligonucleotides were labelled using T4 polynucleotide kinase and [γ - ^{32}P]ATP (6000 Ci/mmol). The oligonucleotide primers contained nucleotides +94 to +67 (ctctgcttaggtcctcactgagctcgg) or +88 to +57 (tttagtctctcactgagctcggagtgaccga) along the bottom strand of exon 2. Labelled oligomer (400 000 c.p.m.) was mixed with 40 μg of total RNA or 4.0 μg of poly(A)⁺ RNA and denatured at 60 °C for 60 min. The primer and RNA were transferred to a 37 °C water bath and 0.25 mM each of dATP, dTTP, dCTP and dGTP, and Superscript RNaseH⁻ reverse transcriptase (Gibco-BRL, Grand Island, NY, U.S.A.) were added. The extension buffer contained a final concentration of 10 mM Tris (8.5), 10 mM MgCl₂ and 10 mM dithiothreitol (DTT). The reverse transcription reaction was allowed to continue for 60 min at 37 °C. The reaction mixture was extracted with phenol:chloroform (1:1) and the DNA from the resulting extract was precipitated by the addition of 3 volumes of ethanol.

The extension reaction products were resolved on a sequencing gel. Total RNA was isolated from rat liver using the RNA STAT-60 reagent (Tel Test Inc., Friendswood, TX, U.S.A.). Poly(A⁺) RNA was prepared from total RNA with the Oligotex mRNA kit (Qiagen, Chatsworth, CA, U.S.A.).

cDNA extension by RACE (rapid amplification of the 5' ends)

RACE of L-CPT I cDNA was conducted with the marathon cDNA amplification kit (Clontech). The reverse primer was called RP1 (cttggtgagcctctgccatcttgagtg) while the forward primer (AP-1) was provided by the manufacturer. The PCR products were subcloned into the pGEM T cloning vector (Promega, Madison, WI, U.S.A.) and sequenced using the Sequenase 2.0 system (Amersham/USB). The sequencing primers were either RP1, T7 or SP6.

Reverse transcription PCR (RT-PCR)

Total cellular RNA was treated with DNase I to remove contaminating genomic DNA and then used to prepare cDNA using random hexamers and Superscript II reverse transcriptase (Gibco-BRL, Gaithersburg, MD, U.S.A.). Subsequent PCR reactions contained 2 μl cDNA, 10 mM Tris/HCl, pH 8.3, 50 mM KCl, 2 mM MgCl₂, 0.2 mM each dATP, dTTP, dCTP and dGTP, 0.4 μM forward primer, 0.4 μM reverse primer, 2.5 units AmpliTaq Gold polymerase (Perkin-Elmer), and 0.02 units Vent polymerase (New England Biolabs, Beverly, MA, U.S.A.). Total reaction volume was 50 μl . PCR parameters were 35 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s. The initial cycle was preceded by 9 min at 94 °C to activate the AmpliTaq Gold in addition to denaturing the template, and the last cycle was followed by 5 min at 72 °C. RT-PCR products were resolved on 3% Nuseive agarose gels (FMC BioProducts, Rockland, ME, U.S.A.). Segments yielding positive results were sequenced by direct sequencing of PCR products using the reverse PCR primer as the sequencing primer or by subcloning the DNA fragments into pGEM-T for sequence analysis.

Northern analysis

Northern-blot analysis was conducted essentially as described previously [3]. The L-CPT I cDNA probe was labelled with [α - ^{32}P]dCTP and a Prime-it II random primer labelling kit (Stratagene). The multiple tissue Northern blot was obtained from Clontech (Palo Alto) and probed as described previously [3]. The probe was removed from the membrane for additional analysis with other probes by boiling the membrane in 0.1% SDS.

Construction of luciferase and CAT vectors

The Int 1 +28/+64 and Int 1 +28/+64 Rev CPT-luciferase (Luc) vectors were constructed by ligating a *Bam*HI-*Sal*I fragment of the L-CPT I gene in front of the luciferase reporter gene in the pGL3 basic vector from Promega (Madison, WI, U.S.A.). The -965/+64 CPT-Luc was prepared by ligating a *Sac*I to *Sal*I fragment into the appropriate restriction sites of pGL3 basic.

To prepare the -965/+19 and -193/+19 CPT-Luc vectors, the promoter regions were obtained by PCR amplification using forward and reverse primers that extended from -965 to -944 (ggaggtaccgaggcagacagactcttc) or -193 to -173 (ttaggtaccctgctcaggtctgtatgttc) and +19 to -1 (cgaagatctggacggcgccg-

cggaatgag), respectively. The forward primers contained a *KpnI* restriction site and the reverse primer contained a *BglII* restriction site for ligating into the pGL3 basic luciferase vector. The fragments with restriction sites were prepared by PCR amplification with Pfu polymerase (Stratagene) in 20 rounds of 95 °C for 1 min, 52° for 1 min and 72 °C for 2 min. The sequence of the fragments was confirmed by DNA sequencing.

The CAT reporter gene was placed in the *HindIII* to *PstI* sites of pTZ18R as has been described [23]. The CPT promoter fragments were removed from the luciferase constructs by digestion with *KpnI* and *BglII* and ligated in front of the CAT reporter gene. The -193/+19 Rev CPT-CAT was prepared by PCR amplification as described with the luciferase vectors using the forward primer ttaagatctctgtcctagtctgtagtc and the reverse primer cgaggtaccggacggcgccggcggaatgag. The resulting PCR product was cloned into the *KpnI* and *BglII* sites in front of the CAT reporter gene.

Cell transfections, luciferase assays and CAT assays

HepG2 cells were transfected by calcium phosphate precipitation as described previously [24]. For the CPT-Luc vectors, each transfection contained 3.0 µg of CPT-Luc and 1.0 µg of SV40 (simian virus 40)-β-galactosidase (SV40-βgal) reporter gene as a transfection control. For the CPT-CAT vectors, each transfection contained 5.0 µg of CPT-CAT vector and 2.5 µg of SV40-βgal.

Cells transfected with the luciferase vectors were harvested in reporter lysis buffer (Promega). After scraping cells from the plates, they were additionally lysed by one round of freeze-thawing. The cell debris was removed by centrifugation and the supernatant collected for further analysis. Luciferase assays were conducted with 10 µl of cell extract and 100 µl of luciferase assay reagent (Promega). The luciferase activity was measured in TD 20/20 luminometer (Turner Designs, Madison, WI, U.S.A.). The β-galactosidase activity was determined using 200 µl of cell extract and *o*-nitrophenyl β-D-galactopyranoside (ONPG) in a colorimetric assay. The protein content of the cell extract was measured with a colorimetric assay using the Bio-Rad protein assay dye (Bio-Rad, Hercules, CA, U.S.A.). The luciferase activity was corrected for protein concentration in the lysate and for β-galactosidase activity.

CAT assays were conducted with [³H]chloramphenicol and *n*-butyryl-CoA using the xylene phase extraction method [22,23]. The transfections of each vector were done in duplicate plates and repeated at least three times.

RESULTS AND DISCUSSION

Identification of the rat L-CPT I gene

Expression of the gene for L-CPT I is regulated during development and in response to dietary modifications [19]. As the first step towards investigating the transcriptional control of L-CPT I, our goal was to clone the promoter regulatory region of this gene. Initially, a rat genomic library was screened with a portion of the L-CPT I cDNA that contained 400 bp of CPT cDNA sequence 3' to the ATG that initiates translation of the L-CPT I protein [5]. Following three rounds of library screening and plaque purification, a genomic clone called λ1A was identified (Figure 1A). This clone contained eight L-CPT I exons including the 3' portion of exon 3 which encodes the ATG initiating translation. However, λ1A did not contain the most 5' sequences present in the published cDNA [5]. The library was rescreened and a second genomic clone, λ17B, was obtained which had the

Figure 1 Organization of the promoter and 5' exons of the rat L-CPT I gene

(a) The location of the exons and introns in the rat L-CPT I gene are shown. The exons are represented by the filled boxes and intron sequence is indicated by the solid line. The numbers below each exon indicate the number of the exon. The ATG in exon 3 indicates the start site of protein translation. The lines labelled λ1A and λ17B are the genomic sequences obtained through screening of rat lambda libraries. The more 5' region of the CPT I gene was found in the P1 clones. Restriction sites are indicated as follows: B, *BamHI*; S, *SalI*; E, *EcoRI*; X, *XhoI*. (b) The sequence of exon-intron boundaries of rat L-CPT I gene are given. The capital letters indicate nucleotides found in the exons while the lower case letters represent sequence in the introns. The numbering of the gene nucleotides in the larger font is based upon +1 as the start site of transcription while the numbering in parentheses is based upon the sequence from the cDNA clone in which the A of the ATG is +1 [5].

5' portion of exon 3 and an additional 5000 bp of upstream genomic sequence but still did not possess all of the cDNA sequence.

To acquire more 5' genomic sequence, we obtained two P1 clones (Genome Systems, St. Louis, MO, U.S.A.), each containing approx. 75–85 kb of genomic DNA. The genomic DNA within the P1 clones was cut with restriction enzymes and subcloned into pBluescript for further characterization. The P1 clones contained considerably more genomic DNA 5' to our λ clones, including all the known sequences in the 5' untranslated region of the L-CPT I mRNA. One P1 clone contained 14 kb of DNA upstream of exon 3. Through Southern analysis and DNA sequencing, it was determined that there is a 6 kb intron between exon 2 and exon 3 and a 1.2 kb intron between exon 1 and exon 2 (Figure 1A). Exon 1 and the initial 30 bp of exon 2 are comprised of novel sequence not identified in the initial cloning of the L-CPT I cDNA [5]. We have sequenced from the end of the second exon approx. 2000 bp 5' on both strands. This sequence includes all of intron 1 as well as 1000 bp of the promoter sequence.

In Figure 1(b), the sequence surrounding the intron/exon boundaries is shown. In this Figure, the exons are numbered using the start site of transcription as +1. In addition, the numbering system of Esser et al. [5] is shown to facilitate comparison with the previously described cDNA. The intron/exon junctions all contain consensus splice site sequences [25]. Within the λ clones, we have identified eight exons that encode L-CPT protein and the DNA sequence within these exons is identical to the cDNA sequence of Esser et al. [5]. These exons contain approx. 1000 bp of protein coding sequence. Together

CPT I cDNA 102 bp 5' to ATG

AGTCGGTCGA

5' end of RACE product: 142 bp 5' to ATG

CCACAGACTCGCAAGCATACATCACCACCACCTCCCTAGTCGGTCGA

Location of RACE product in genomic DNA

CCACAGACTCGgta cagCAAGCATACATCACCACCACCTCCCTAGTCGGTCGA
Exon 1 Intron 1 Exon 2

Figure 2 Identification of an additional 5' sequence in the L-CPT I cDNA by RACE

RACE reactions were conducted as described in the Materials and methods section. At the top of the Figure, the sequence underlined once indicates the nucleotides that are homologous with the 5' end of the previously identified cDNA [6]. In the centre, the additional 37 nucleotides that were found are shown. The sequence underlined twice identifies the 26 nucleotides from the RACE reaction that are contiguous with exon 2 of the genomic DNA. The nucleotides underlined with the hatched bar are in the genomic DNA but are separated from exon 2 by 1170 bp intron sequence. This sequence represents the most 3' end of exon 1. In the bottom panel, the splice junctions between exons 1 and 2 are shown.

exons 1, 2 and 3 contain 158 bp of untranslated sequences (27 bp, 118 bp and 13 bp for exons 1, 2 and 3, respectively).

Identification of additional 5' sequence in the L-CPT I cDNA

To identify the 5' end of the L-CPT I mRNA, a RACE system was employed. The reverse primer contained the sequence surrounding the ATG that initiates translation from +177 to +150 in the L-CPT I mRNA [5] while the forward primer hybridized with the adaptor primer on the 5' end of the cDNAs. Template cDNAs were prepared from rat liver RNA. The products from the PCR reactions were subcloned into the pGEM T cloning vector. Approx. 40 subclones were sequenced and these subclones gave products of varying length. The sequence was identical in all the subclones with the only variation being in the degree of 5' extension. The longest RACE products contained 37 bp of previously unidentified cDNA (Figure 2). The sequence of the additional 37 bp from the RACE reactions was identified within the genomic DNA. The first 26 bp had perfect homology with the genomic DNA at the 5' end of exon 2. The remaining 11 bp were found 1170 bp 5' to the previous exon. The sequences within the intron/exon boundaries encode excellent consensus splice sequences indicating that these 11 nucleotides represent the 3' end of a previously unknown exon. The longest products from the RACE reactions stopped at the identical C, suggesting that this region was very difficult to extend through. Indeed, when the sequence from the RACE reactions was aligned with the genomic sequence, it was found that this was the 3' end of an extremely GC rich region.

Identification of a transcriptional start site

Primer extension analysis was used to identify the transcriptional start site for the L-CPT I mRNA. Two primers extending from +94 to +67 (primer A) and +88 to +57 (primer B) in exon 2 of the gene were labelled on the 5' end with [γ - 32 P]ATP and used in primer extension analyses with poly(A)⁺ RNA or total RNA from rat liver or total RNA from rat quadriceps muscle. The liver RNA was obtained from animals which were starved for 2 days. Starvation increases the abundance of L-CPT I mRNA

Figure 3 Identification of the transcriptional start site of the L-CPT I mRNA

Primer extension analyses were conducted as described in the Experimental section. Poly(A)⁺ RNA from liver (Liver poly A) and total RNA from either liver (Liver total) or quadriceps muscle (Muscle total) was analysed. Two exon 2 specific primers called primer A (+94/+67) and primer B (+88/+57) were labelled with [γ - 32 P]ATP and used in the extension reaction. The products of the extension reaction were resolved on a denaturing acrylamide gel. The numbers adjacent to lane 1 indicate the length of the 32 P-labelled DNA markers. The arrows indicate the primary products of the primer extension reaction.

eight-fold [3]. The muscle RNA was obtained from fed animals. The hepatic CPT I isoform is not detected in the RNA from skeletal muscle of fed animals by Northern analysis (data not shown) so this RNA served as a negative control. With rat liver RNA, primary products of 98 and 93 bp were obtained from primer A and primer B, respectively (Figure 3). The extension products were 16 bp longer than the products which were observed in the RACE analysis. The data show that the major start site of transcription is 158 bp 5' to the initiating ATG. The primer extension reactions yielded more product with poly(A)⁺ RNA as the substrate. There are several weaker bands which are longer than the major product and most likely represent alternate start sites of transcription. Primer extension reactions with the RNA from muscle did not generate any detectable extension products even with extremely long exposures of the autoradiograms.

Confirmation of exon 1 by RT-PCR

RT-PCR was used to demonstrate that the exon 1 sequence was present in mRNA from rat liver. cDNAs were made by reverse transcription of rat liver RNA with random primers. The cDNAs were subjected to PCR amplification using five different forward

Figure 4 Confirmation of exon 1 mRNA from rat liver by RT-PCR

(A) A map of the first three exons in the L-CPT I mRNA is shown. The locations of the forward primers (FP) and the reverse primers (RP) are shown. The ATG in exon 3 represents the start site of protein translation. (B) The nucleotides contained in the forward primers, labelled FP1 through FP5, are shown. The reverse primers, labelled RP1 and RP2, encode nucleotides +177 to +150 and +261 to +234 respectively. The length of product from each PCR reaction is indicated. (C) Gel electrophoresis of the RT-PCR products is shown. In lanes 1–6, RNA from the liver of a 48 h starved rat was used for the RT-PCR reactions along with the primers outlined below. In lane 7 RNA from the liver of a fed rat was used. Lane 8 is a negative control reaction containing no cDNA template in the PCR amplification. The following sets of primers were used: lane 1, FP3/RP2; lane 2, FP1/RP1; lane 3, FP2/RP1; lane 4, FP3/RP1; lane 5, FP4/RP1; lane 6, FP5/RP1; lane 7, FP2/RP1; lane 8, FP2/RP1. The molecular weight standards of 400, 200 and 100 base pairs (bp) are indicated.

primers (FP) which hybridized with different regions of the promoter, exon 1 or exon 2 and two reverse primers (RP) located in exon 3. The relative location of the primers is shown in Figure 4A, and the exact nucleotides encoded by the forward primers is given in Figure 4B. The products of the RT-PCR reactions are shown in Figure 4C. FP1 is a control primer which anneals with nucleotides in exon 2 from the previously described cDNA [5]. The length of the PCR product in lane 2 matched the anticipated mRNA size of 120 bp. FP2 hybridizes only with sequences in exon 1. The presence of a 171 bp product in lane 3 confirms that exon 1 is transcribed and contained in the L-CPT I mRNA. The primer FP3 (–16/+5) which overlaps the primary start site of transcription also amplified a PCR product of 193 bp (lane 4). This product confirms as suspected that there are additional transcriptional start sites 5' to the primary start site identified by the primer extension studies (Figure 3). By directly sequencing the PCR products generated with the primers FP2/RP1 and FP3/RP1, it was found that the sequence of the PCR products was 100% homologous with the nucleotide sequence determined by the RACE reactions, the published cDNA [5] and the genomic clone (data not shown). The nucleotide sequence of the RT-PCR products was also confirmed by subcloning the primary PCR

products from lanes 3 and 4 into the pGEM-T vector for sequencing (data not shown). In lanes 5 and 6, the promoter specific primers, FP4 and FP5, did not generate any products, indicating that there is no mRNA originating from these upstream sites. If these sequences were transcribed, these primers would have generated products of 255 and 394 bp, respectively. As a control, the FP3 and RP2 primers were used to produce a product of 277 bp, indicating that the RT-PCR reactions could amplify longer products in this size range (lane 1).

The results from these experiments indicate that the L-CPT I mRNA contains exon 1 and that we have identified the correct splice junctions for the L-CPT I gene. Because the RT-PCR reactions had 35 cycles of amplification, the bands do not reflect the quantitative amount of L-CPT I RNA in each sample. Using identical primers, the RNA from starved and fed liver generated approximately equal amounts of RT-PCR product (lanes 3 and 7). RNA from rat heart also gave an RT-PCR product that was the identical size and abundance as was obtained with liver RNA (data not shown). Therefore, exon 1 is also expressed in heart. Primers FP4 and FP5 did not generate products even when the annealing temperature was decreased from 55 to 52 °C (data not shown). Since in each RT-PCR reaction, genomic sequences contain at least one intron between the forward and reverse primers none of the observed products could have arisen from amplification of contaminating genomic DNA.

Identification of an alternatively spliced form of the L-CPT I mRNA

A less intense but faster migrating band was observed in several of the RT-PCR reactions (Figure 4C, lanes 3, 4 and 7). This DNA fragment was excised from the gel and subcloned into the pGEM-T vector for sequence identification. These RT-PCR products contained intact exons 1 and 3 but exon 2 was completely absent. Therefore in the L-CPT I mRNA, exon 2 can be removed by alternative splicing. The abundance of the alternatively spliced mRNA cannot be determined on the basis of the RT-PCR analyses. Although this splice variant does not change the protein coding region of the mRNA, it might alter the ability of the mRNA to initiate translation [26]. It raises the possibility that there may be other alternative splice sites within the protein coding region of the L-CPT I mRNA. Alternative splicing has been demonstrated in the human M-CPT I gene [27]. Finally, alternative splicing of the L-CPT I mRNA may represent an additional level of control of gene expression if splicing is shown to be affected by different nutritional or developmental conditions.

Transient transfections with L-CPT I promoter

The next experiments tested whether genomic fragments of the L-CPT I gene could drive transcription of luciferase and CAT reporter genes in transient transfection assays. Various lengths of the L-CPT I genomic sequence were ligated to the luciferase reporter gene (Figure 5A) and transiently transfected into HepG2 cells. The cells were harvested and luciferase activity determined. The activity from the L-CPT I luciferase vectors was compared with the promoterless luciferase construct, pGL3 basic, whose activity was given a relative value of 1.

A portion of exon 2 and 1000 bp of the first intron were ligated in front of the luciferase reporter gene. When this vector (Int 1 +28/+64) was transiently transfected into HepG2 cells, only a two-fold increase in luciferase activity was observed (Figure 5A). When this segment was placed in the reverse orientation, it suppressed transcription of the luciferase gene. These observations suggested that this region did not contain the L-CPT I

Figure 5 Transient transfection of luciferase and CAT reporter genes driven by the L-CPT I promoter

HepG2 cells were transiently transfected with CPT-luciferase (Luc) and CPT-CAT vectors as outlined in the Experimental section. Each transfection contained a SV40- β gal vector as a transfection control. An outline of the L-CPT I promoter and exons 1 and 2 is provided at the top of the Figure. The vectors are numbered with the start site of transcription being +1. (A) The relative activity of the various CPT-Luc vectors is shown. The activity was normalized for comparison with the promoterless luciferase gene pGL3-basic. (B) Transfections of CPT-CAT vectors were conducted. The expression was normalized relative to the promoterless CAT vector p0-CAT.

promoter and confirmed the results of the RACE and RT-PCR experiments shown in Figures 2 and 4. When exon 1 and 965 bp of promoter were added, as in the -965/+64 CPT-Luc vector, expression increased nine-fold above the promoterless vector. Removal of intron 1 resulted in a striking increase in L-CPT I expression. The -193/+19 CPT-luciferase and -965/+19 CPT-luciferase expressed at 65- and 23-fold above the promoterless vector respectively. These observations confirm that transcription begins in the upstream exon. In addition, these experiments suggest that intron 1 contains sequences that decrease basal transcription. Future studies will indicate whether the intron or proximal promoter contain any hormone response elements.

Similar results were obtained using the CAT reporter gene (Figure 5B). In these experiments, the 210 bp fragment containing the proximal promoter and exon 1 was placed in the forward and reverse orientations relative to the CAT gene. The -193/+19 CPT-CAT construct with the promoter in the forward orientation gave 40-fold more activity than the promoterless vector while the -193/+19 Rev CPT-CAT construct with the promoter in the reverse orientation gave the same activity as the promoterless vector. The vector containing 965 bp of the promoter also expressed at a high level relative to the promoterless CAT gene. These experiments indicate that 193 bp of the promoter is

Figure 6 Sequence of L-CPT I promoter

The sequence of the proximal L-CPT I promoter and the first exon are shown. Putative protein binding sites for various transcription factors in the first 300 bp are indicated. The -1 indicates the first nucleotide 5' to the start site of transcription and the +1 indicates the first nucleotide of the mRNA. The underlined sequences indicate consensus binding sites for the transcription factors Sp1, AP-2 and the CCAAT enhancer binding protein (C/EBP). A CCAAT box is underlined.

sufficient for high levels of basal expression and that we have identified the functional promoter of the L-CPT I gene.

In these studies, we have provided three lines of evidence for the existence of exon 1. First the luciferase vector containing +28/+64 of exon 2 and 1000 bp of intron was unable to drive efficiently a luciferase reporter gene in transient transfection assays. Actually, it was this lack of activity in the transient transfection assays that lead us to conduct the RACE and RT-PCR experiments. Second, the RACE reactions revealed the presence of previously unidentified 5' sequences in the mRNA. When these sequences were identified in the genomic DNA, excellent consensus splice sites were found surrounding the most 5' intron. Finally, the RT-PCR experiments which utilized primers in exons 1 and 2 showed that exon 1 was contained in the L-CPT I mRNA.

Analysis of the sequence in L-CPT I promoter region

The sequence of the first 960 bp of the L-CPT I promoter and exon 1 is shown in Figure 6. The first base in the mRNA is labelled +1 and the first bp of the promoter is labelled -1. The transcriptional start site of L-CPT I conforms well with a consensus initiator sequence Py/Py/A (+1)/N/(A/T)/Py/Py in which the A is the first transcribed nucleotide [28]. The first 140 bp of the promoter are extremely GC rich with 77% of the nucleotides being G or C. There is no apparent TATA box but there are several consensus Sp1 binding sites at -21 to -27 and -100 to -107 [29,30]. Within this region, there are several AP-2 binding sites and binding sites for the CCAAT enhancer binding protein (C/EBP) [23,31]. In addition, there is a CCAAT box at nucleotides -62 to -67. CCAAT boxes which are generally located 50 to 80 bp upstream of the transcriptional start site bind the nuclear factor NF-Y [32]. In many genes, NF-Y is important in the recruitment of RNA polymerase II and

transcriptional initiation [32]. The rat L-CPT I promoter does not have any obvious homologies with the promoters of the mouse and human CPT II genes [13,14]. The human CPT II promoter contains a TATA box [14] but the mouse gene has no TATA box [13]. However, it would be anticipated that the promoters of the L-CPT I and CPT II genes should be different since the CPT II gene is expressed in all tissues and is not subject to hormonal or developmental control [18,20]. The human M-CPT I promoter also appears to be different as it utilizes two alternatively expressed first exons [10].

Transcriptional initiation from TATA-less promoters

TATA-less promoters are frequently described as housekeeping genes. However, L-CPT I is highly regulated, at least in the liver, by diet, hormones and developmental processes. TATA-less promoters often contain GC rich regions in the proximal promoter and utilize multiple start sites of transcription [33]. The promoter of the L-CPT I gene conforms with this model. In the absence of a TATA box, it has been postulated that an initiator element helps position the RNA polymerase II and locate a start site of transcription [34]. In addition to the initiator element where L-CPT I transcription starts, there are consensus initiator sites at -4 to -10 and -59 to -65. Since the L-CPT I isoform is widely expressed, it will be interesting to see if the same transcriptional start site is utilized in all tissues. TATA-less promoters still utilize TATA factors and several factors have been proposed to help initiate transcription from TATA-less genes [34]. Sp1 in particular has been shown to tether the TATA factors and participate in transcriptional initiation [33]. Since there are multiple members of the Sp family of transcription factors, others may also be involved in L-CPT I expression [29,30]. AP-2 has also been reported to be involved in the transcriptional regulation of TATA-less genes [31,35]. Future studies will clarify the role of these elements and factors in the expression of the L-CPT I gene.

Tissue-specific expression of the L-CPT I gene

Northern analysis was conducted to determine the tissue-specific distribution of the L-CPT I mRNA. A membrane containing poly(A)⁺ RNA from various rat tissues was probed with an L-CPT cDNA fragment (Figure 7). The liver isoform of L-CPT I is expressed in a variety of tissues besides the liver. L-CPT I mRNA is present in all the tissues on the blot except the skeletal muscle. The blot was reprobed with an actin cDNA which hybridized with a 2.0 kb mRNA except in the lanes containing heart and skeletal muscle RNA. These tissues contain two isoforms of actin mRNA which are 2.0 and 1.7 kb in length [36]. Expression in the kidney was quite high perhaps because the kidney is quite metabolically active.

Several observations suggest that the control of L-CPT I gene expression is a key feature in the regulation of fatty acid oxidation. Transcription of the L-CPT I gene is elevated in the livers of rats on high fat diets and decreased in the livers of rats on high carbohydrate diets [19]. In isolated hepatocytes, insulin inhibits transcription of the L-CPT I gene. Long-chain fatty acids and peroxisomal proliferator compounds such as clofibrate will also stimulate transcription of the gene. Transcription of the L-CPT I gene is increased in the liver at birth, providing the capacity to generate ketone bodies and to oxidize fatty acids from mothers' milk [19].

We have analysed 1000 bp of 5' flanking sequence and the first intron of the L-CPT I gene for potential hormone response elements. In this region, there are no insulin responsive elements similar to those identified in the gene for phosphoenolpyruvate

Figure 7 Tissue specific expression of L-CPT I mRNA

Northern analysis was conducted to determine the distribution of L-CPT I mRNA in various tissues. A membrane containing poly A⁺ RNA from a variety of tissues was probed with a ³²P-labelled L-CPT I cDNA probe. The membrane was exposed to an autoradiogram. Next, the L-CPT I probe was stripped from the membrane and the membrane was reprobed with actin. The L-CPT I and actin RNAs are indicated by arrows. The tissue source of the RNA is indicated at the top of the Figure.

carboxykinase (PEPCK) or insulin-like growth factor binding protein (IGFB) genes which are inhibited by insulin [37,38]. The transcriptional effects of long chain fatty acids and peroxisomal proliferators have been postulated to be mediated through peroxisomal proliferator-activated receptors (PPAR) [39,40]. PPARs bind to direct repeats of the AGGTCA motif separated by one nucleotide, but none of these motifs is apparent within the first kb of promoter or in intron 1. There are no consensus binding sites for the cAMP responsive protein (CREB) which has been shown to modulate the transcription induction by cAMP [41]. Several binding sites for AP-2 were also identified. AP-2 has been proposed to be a potential mediator of the cAMP induction of gene expression [42]. Cyclic AMP may increase L-CPT I expression through AP-2 sites rather than through CREB. We have failed to identify any thyroid hormone responsive elements within the first intron or in the proximal promoter [43]. In our transient transfection experiments with the L-CPT I promoter ligated to the luciferase reporter gene, T3 does not stimulate luciferase activity (data not shown). Therefore, the TRE and

possibly other hormone response elements are located further upstream in the promoter of the gene.

Within the first 1000 bp, we have found several consensus C/EBP protein binding sites (Figure 5) [24]. The C/EBP α isoform is believed to be an important regulator of genes encoding proteins involved in metabolism and to contribute to the induction of hepatic genes at birth [23,44]. Therefore, C/EBP proteins may contribute to the induction of L-CPT I expression in the liver at birth. C/EBP α has been shown to participate in the cAMP induction of the PEPCK gene [45]. However, L-CPT I is widely expressed and C/EBP α has been postulated to induce the expression of genes primarily in liver and adipose tissue. Little is known regarding the developmental or hormonal regulation of the L-CPT I isoform in non-hepatic tissues. A variety of regulators will be required to mediate expression of this gene in so many tissues. Future studies will be directed at defining more precisely the promoter elements and factors involved in hormone responsiveness.

This work was supported in part by grants from the National Institute of Health (DK-46399) and from the American Heart Association (96010180 to E.A.P. and 9650142N to G.A.C.). M.L.S. is supported by the NIH Lipid training grant (HL-07641).

REFERENCES

- McGarry, J. D. and Foster, D. W. (1997) *Eur. J. Biochem.* **244**, 1–14
- McGarry, J. D., Mannaerts, G. P. and Foster, D. W. (1977) *J. Clin. Invest.* **60**, 265–270
- Park, E. A., Mynatt, R. L., Cook, G. A. and Kahli, K. (1995) *Biochem. J.* **310**, 853–858
- Murthy, M. S. R. and Pande, S. V. (1987) *Biochem. J.* **248**, 727–733
- Esser, V., Britton, C. H., Weis, B. C., Foster, D. W. and McGarry, J. D. (1993) *J. Biol. Chem.* **266**, 5817–5822
- Yamazaki, N., Shinohara, Y., Shima, A. and Terada, H. (1995) *FEBS Lett.* **363**, 41–45
- Esser, V., Brown, N. F., Cowan, A. T., Foster, D. W. and McGarry, J. D. (1996) *J. Biol. Chem.* **271**, 6972–6977
- Britton, C. H., Schultz, R. A., Zhang, B., Esser, V., Foster, D. W. and McGarry, J. D. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1984–1988
- Yamazaki, N., Shinohara, Y., Shima, A., Yamanka, Y. and Terada, H. (1996) *Biochem. Biophys. Acta* **1307**, 157–161
- Yamazaki, N., Yamanaka, Y., Hashimoto, Y., Shinohara, Y., Shima, A. and Terada, H. (1997) *FEBS Lett.* **409**, 401–406
- Britton, C. H., Mackey, D. W., Esser, V., Foster, D. W., Yarnall, D. P., Froguel, P. and McGarry, J. D. (1997) *Genomics* **40**, 209–211
- Woeltje, K. F., Esser, V., Wies, B. C., Sen, A., Cox, W. F., McPhaul, M. J., Slaughter, C. A., Foster, D. W. and McGarry, J. D. (1990) *J. Biol. Chem.* **265**, 10720–10725
- Gelb, B. D. (1993) *Genomics* **18**, 651–655
- Montermini, L., Wang, H., Verderio, E., Taroni, F., DiDonato, S. and Finocchiaro, G. (1994) *Biochem. Biophys. Acta* **1219**, 237–240
- Choi, S. J., Oh, D. H., Song, C. S., Roy, A. K. and Chatterjee, B. (1995) *Biochem. Biophys. Acta* **1264**, 215–222
- Derrick, J. P. and Ramsey, R. R. (1989) *Biochem. J.* **262**, 801–806
- Murthy, M. S. R. and Pande, S. V. (1994) *J. Biol. Chem.* **269**, 18283–18286
- Thumelin, S., Esser, V., Charvy, D., Kolodziej, M., Zammit, V. A., McGarry, D., Girard, J. and Pegorier, J. P. (1994) *Biochem. J.* **300**, 583–587
- Chatelain, F., Kohl, C., Esser, V., McGarry, J. D., Girard, J. and Pegorier, J. P. (1996) *Eur. J. Biochem.* **235**, 789–798
- Mynatt, R. L., Park, E. A., Thorngate, F. E., Das, H. K. and Cook, G. A. (1994) *Biochem. Biophys. Res. Commun.* **201**, 932–937
- Brown, N. F., Weis, B. C., Husti, J. E., Foster, D. W. and McGarry, J. D. (1995) *J. Biol. Chem.* **270**, 8952–8957
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (eds) (1987) *Current Protocols in Molecular Biology*, John Wiley and Sons Publishers, New York
- Park, E. A., Gurney, A. L., Nizielski, S. L., Cao, Z., Hakimi, P. and Hanson, R. W. (1993) *J. Biol. Chem.* **268**, 613–619
- Park, E. A., Jerden, D. C. and Bahouth, S. W. (1995) *Biochem. J.* **309**, 913–919
- Ohshima, Y. and Gotoh, Y. (1987) *J. Mol. Biol.* **195**, 247–259
- Dietz, H. C. and Kendzior, R. J. (1994) *Nat. Genet.* **8**, 183–188
- van der Leij, F. R., Takens, J., van der Veen, A. Y., Terpstra, P. and Kuipers, J. R. G. (1997) *Biochem. Biophys. Acta* **1352**, 123–128
- Javahery, R., Khachi, A., Kiersten, L., Zenzie-Gregory, B. and Smale, S. T. (1994) *Mol. Cell. Biol.* **14**, 116–127
- Liang, Y., Robinson, D. F., Dennig, J., Suske, G. and Fahl, W. E. (1996) *J. Biol. Chem.* **271**, 11972–11977
- Hagen, G., Dennig, J., Preib, A., Beato, M. and Suske, G. (1995) *J. Biol. Chem.* **270**, 24989–24994
- Williams, T. and Tijan, R. (1991) *Genes Dev.* **5**, 673–680
- Linhoff, M. W., Wright, K. L. and Ting, J. P. Y. (1997) *Mol. Cell. Biol.* **17**, 4589–4596
- Pugh, F. B. and Tijan, R. (1991) *Genes Dev.* **5**, 1935–1945
- Smale, S. T. (1997) *Biochem. Biophys. Acta* **1351**, 73–88
- Chen, T. T., Wu, R. L., Castro-Munozledo, F. and Sun, T. T. (1997) *Mol. Cell. Biol.* **17**, 3056–3064
- Giovanna, P., Jardine, K. and McBurney, M. W. (1991) *Mol. Cell. Biol.* **11**, 4796–4803
- O'Brien, R. M. and Granner, D. K. (1996) *Physiol. Rev.* **76**, 1109–1161
- O'Brien, R. M., Noisin, E. L., Suwanishkul, A., Yammasaki, T., Lucas, P. L., Wang, J. C., Powell, D. R. and Granner, D. K. (1995) *Mol. Cell. Biol.* **15**, 1747–1758
- Issemann, I. and Green, S. (1990) *Nature (London)* **347**, 654–649
- Keller, H., Dreyer, C., Medin, J., Mahfoudi, A., Ozato, K. and Wahli, W. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 2160–2164
- Meyer, T. E. and Habner, J. F. (1993) *Endocrine Rev.* **14**, 269–290
- Imagawa, M., Chiu, R. and Karin, M. (1987) *Cell* **51**, 251–260
- Mangelsdorf, D. J. and Evans, R. M. (1995) *Cell* **83**, 841–850
- McKnight, S. L., Lane, M. D. and Gluecksohn-Waelsch, S. (1989) *Genes Dev.* **3**, 2021–2024
- Roesler, W. J., Crosson, S. M., Vinson, C. and McFie, P. J. (1996) *J. Biol. Chem.* **271**, 8068–8074