Characterization of the Promoter Regulatory Region of the Human Pyruvate Dehydrogenase β Gene†

Kunapuli T. Madhusudhan,‡,§,|| Sharon S. Naik,‡,|| and Mulchand S. Patel*,⊥

Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, and Department of Biochemistry, State University of New York at Buffalo, Buffalo, New York 14214

Received July 18, 1994; Revised Manuscript Received October 24, 1994®

ABSTRACT: A genomic clone (19 kb) harboring the intron−exon sequences and the promoter-regulatory region of the E1β gene of human pyruvate dehydrogenase complex was isolated by screening a placental genomic library. The nucleotide sequence of the promoter region (1245 bp) showed 18 differences (including mismatches, insertions, and deletions) as compared to the published sequence [Koike *et al.* (1990) *Proc. Natl. Acad. Sci. U.S.A. 87,* 5594−5597]. The E1β promoter lacked a TATA box homology but contained initiator sequences (two) and Sp1 sites (three) which are frequently found in TATA-less promoters. The DNase I footprinting pattern of the promoter region with crude rat liver nuclear extracts showed at least seven regions of protein binding and nuclease protection (P1−P7). The DNase I protected regions contained consensus nucleotide sequences recognized by GATA-1, Sp1, IgNF-A, Lva, bicoid Q9, NF-kB, HNF-5, H4TF-1, WAP5, and ADH transcription factors. Transient expression of chloramphenicol acetyltransferase (CAT) suggested the possible presence of negative elements located within the sequence from −2316 to −930, whereas deletion constructs containing −929 to +32 and −98 to +32 DNA sequences showed approximately 7- and 20-fold increases in CAT activity over the basal CAT activity. Additional studies indicated the presence of an orientation-dependent *cis* element (or elements) within the region from −282 to −397 that acts as an enhancer or a repressor upon a heterologous thymidine kinase promoter. The results show that the human E1β promoter represents an unusual variation in a housekeeping gene promoter with a unique combination of initiator sequences and Sp1 sites together with a protein-binding site (or sites) within the first 100 bp upstream of the transcription start site.

The mammalian mitochondrial pyruvate dehydrogenase complex (PDC)[1] is composed of multiple copies of three catalytic components. Pyruvate dehydrogenase (E1) decarboxylates pyruvate in the presence of thiamin pyrophosphate to form an acetyl adduct, which is transferred to a dihydrolipoamide acetyltransferase. This component catalyzes the transfer of an acetyl group from an acetyllipoyl moiety to coenzyme A to form acetyl-CoA. The dihydrolipoamide dehydrogenase component catalyzes the reoxidation of reduced lipoyl groups that are covalently bound to the dihydrolipoamide acetyltransferase component of the complex. The E1 component is composed of two nonidentical α and β subunits that form a heterotetramer ($\alpha_2\beta_2$) to catalyze the decarboxylation and dehydrogenation reactions. The activity of E1α is regulated by phosphorylation−dephosphorylation by a specific kinase and phosphatase (Patel & Roche, 1990). The cloning and cDNA sequences of the α and β subunits of human PDC have been reported (Ho *et al.*, 1989; Ho & Patel, 1990). The DNA sequences of the human E1α and E1β genes have also been reported (Maragos *et al.*, 1989; Koike *et al.*, 1990). The protein-binding regions and functional *cis* elements in the 5′-flanking region of the α subunit gene have been reported (Chang *et al.*, 1993), but such information is not available on the human E1β promoter-regulatory region.

The activity of "total" PDC in most tissues is maintained at relatively constant levels; however, the levels of "active" PDC (dephosphorylated form) in most tissues are influenced by nutritional and hormonal modifications (Patel & Roche, 1990). Since the E1 component is the major site of PDC regulation, it is of interest to investigate the gene expression and regulation of coordinate expression of the α and β subunits of the E1 component. Gene expression and regulation involve specific interactions of *trans*-acting nuclear proteins with *cis*-acting nucleotide sequences. To understand the expression and regulation of the E1β gene of human PDC, the nucleotide sequence, the functional protein-binding regions in the proximal promoter-regulatory region, and the effect of *cis*-acting regulatory elements on the transient expression of chloramphenicol acetyltransferase (CAT) activity were studied and are presented in this paper.

MATERIALS AND METHODS

All restriction and modification enzymes were purchased from Boehringer Mannheim Biochemicals, Bethesda Research Laboratories (BRL), and New England Biolabs. Radioisotopes, such as [α-^{32}P]dCTP, [α-^{32}P]dATP, [γ-^{32}P]-

† This work was supported by United States Public Health Service Grant DK20478.

* To whom correspondence should be addressed: Department of Biochemistry, State University of New York at Buffalo, 140 Farber Hall, 3435 Main Street, Buffalo, NY 14214.

‡ These two individuals have made equal contributions to this paper.

§ Present address: Department of Biochemistry and Molecular Biology, Health Sciences Center, University of Oklahoma, Oklahoma City, OK 73190.

|| Case Western Reserve University School of Medicine.

⊥ State University of New York at Buffalo.

® Abstract published in *Advance ACS Abstracts,* January 1, 1995.

[1] Abbreviations: PDC, pyruvate dehydrogenase complex; E1, pyruvate dehydrogenase; CAT, chloramphenicol acetyltransferase; TK, thymidine kinase; kb, kilobases; bp, base pairs; Inr, initiator sequence.

0006-2960/95/0434-1288$09.00/0

ATP, and [^{3}H]chloramphenicol, were from New England Nuclear or ICN Radiochemicals. All other chemicals were of analytical reagent grade. Two plasmid vectors, BSKCAT and PSV2CAT, were a gift from Dr. John Nilson, Department of Pharmacology, Case Western Reserve University, and RSVβgal was kindly provided by Dr. Richard W. Hanson of Case Western Reserve University. Thymidine kinase CAT (TKCAT) vector was a gift from Dr. Richard Eckert, Department of Physiology and Biophysics, Case Western Reserve University.

Isolation and Characterization of an E1β Genomic Clone of Human PDC. A human placental genomic library constructed in EMBL3 (Clonetech Laboratories, Inc.) having an average insert size of 8−21 kilobases (kb) was screened for the isolation of an E1β genomic clone of PDC. The library was plated at a density of 1×10^5 plaques/150-mm plate, and about 15×10^5 plaques were screened. The filters were probed with a 500 base pair (bp) labeled cDNA fragment containing the leader sequence and the nucleotides coding up to 148 amino acids of the human E1β subunit (Ho & Patel, 1990). One positive plaque was obtained upon screening, and the DNA from the positive phage clone was characterized by standard restriction endonuclease mapping and Southern blot analysis (Sambrook *et al.*, 1989). The insert size of the positive clone was estimated to be approximately 19 kb. The presence of exons and introns in the clone was confirmed by dideoxy sequencing using oligonucleotide primers (18−29mers) that hybridize to the DNA in the exon regions.

Molecular Cloning. The insert from the positive phage clone was excised and digested with *Eco*RI. A 4-kb (*Eco*RI) DNA fragment containing the first exon and intron plus a part of the second exon (1.7 kb) together with the 2.3 kb of the 5′ promoter-regulatory region was cloned into the *Eco*RI site of pUC19 (pUC4E1β). To study the promoter region, subclones were constructed by digesting pUC4E1β containing the 4-kb insert with *NaeI*/*Sma*I. A 2.3-kb DNA fragment containing the E1β promoter and the 5′ noncoding region was cloned into the *Sma*I site of pUC19 (pMP6). Digestion of the pMP6 construct with *Pst*I released two insert fragments about 1.3 and 0.5 kb long, and the band corresponding to 1.3 kb was excised. The recovered DNA (1.3 kb) was further digested with *Sac*I to generate two bands of 0.8 and 0.5 kb length. The DNA corresponding to these bands was electroeluted and cloned into pUC19 to yield pMP7 (0.5-kb insert cloned at the *Pst*I site) and pMP8 (0.8-kb insert cloned into the *Sac*I/*Pst*I sites). To prepare deletion clones, pMP6 was digested with *Eco*RI/*Bam*HI, releasing the entire insert, and the insert was recloned into the *Eco*RI/*Bam*HI sites of pBluescript (SK$^+$; pMP14).

The 2348-bp DNA fragment (−2316 to +32) containing the promoter-regulatory region of the E1β gene (pMP6) and its deletion constructs were cloned at the *Eco*RV site in the BSKCAT vector. The BSKCAT vector (4850 bp) had the CAT gene and the SV40 polyadenylation signal subcloned at the *Sma*I site in the polylinker region of pBluescript (Kennedy *et al.*, 1990). The pSV2CAT vector containing the SV40 early promoter/enhancer to drive CAT expression (Gorman *et al.*, 1982) was used as a positive control in all experiments. The deletion clones of the β gene were also cloned into the thymidine kinase TKCAT vector in either forward or reverse orientation by using convenient restriction sites. The TKCAT vector had the TK promoter, the CAT structural gene, and the SV40 polyadenylation signal in pBR322 (Jacoby *et al.*, 1989).

DNA Sequencing. A partial DNA sequence of pMP14 containing the promoter-regulatory region and the 5′ untranslated region of the human E1β gene was determined. To facilitate the DNA sequencing, deletions were made using exonuclease III and S1 nuclease, yielding a set of ordered deletions (Henikoff, 1984). The nucleotide sequences of both strands of DNA were determined twice using a Sequenase kit (U.S. Biochemical Corporation), and the DNA sequences of the strands overlapped. To avoid band compressions, 7-deaza dGTP was used in place of dGTP as suggested by the manufacturer, and samples were subjected to denaturing gel electrophoresis in 7 M urea−6% acrylamide (acrylamide to bisacrylamide ratio, 19:1) gels in 89 mM Tris/89 mM boric acid/2.5 mM EDTA (pH 8.3).

To determine the exact bases protected in DNase I footprinting, the labeled DNA fragments were subjected to Maxam−Gilbert chemical sequencing (Sambrook *et al.*, 1989). The 3′ or 5′ end labeled DNA was subjected to methylation with dimethyl sulfate in aqueous solution at pH 8 for cleavage at G sites and partial depurination in piperidine formate buffer (pH 2.0), giving approximately equal cleavage at G and A sites. The DNA thus treated was subjected to hot aqueous piperidine treatment for base-specific cleavage (Sambrook *et al.*, 1989).

Nucleic Acid Preparations. Total RNA from human placenta extracted by the guanidium thiocyanate method (Sambrook *et al.*, 1989) followed by CsCl centrifugation was kindly provided by Dr. Judith Ilan, Department of Reproductive Biology, Case Western Reserve University. Plasmid and phage DNAs were prepared according to standard methods (Sambrook *et al.*, 1989). The cDNA was labeled by the random primer method for screening genomic libraries according to the manufacturer's instructions (U.S. Biochemical Corporation), and for footprint experiments 3′ recessed ends were labeled using Klenow polymerase and [α-^{32}P]-dCTP or [α-^{32}p]dATP (Sambrook *et al.*, 1989). For S1 nuclease mapping, the pUC4E1β construct (10 μg) was digested with *Eco*RI and *Sac*I, and the 920-bp fragment containing the DNA sequence corresponding to the promoter region, the 5′ untranslated region, and the first exon of the E1β gene was isolated. This fragment was treated with 200 units of exonuclease III at 40 °C for 45 min to degrade the coding strand. The single-stranded DNA was separated, and the 5′ end was dephosphorylated using calf intestinal alkaline phosphatase (Sambrook *et al.*, 1989) and was labeled using T4 polynucleotide kinase and [α-^{32}P]ATP (Sambrook *et al.*, 1989). The labeled probe was separated from the free nucleotides using a Qiagen column according to the manufacturer's instructions (Qiagen Inc.).

S1 Nuclease Mapping. About 25 000−40 000 (Cerenkov) counts of the labeled probe and 75−100 μg of human placental total RNA were suspended in 30 μL of hybridization buffer [0.4 M NaCl, 0.62 M PIPES (1,4-piperazinediethanesulfonic acid) (pH 6.5), 5 mM EDTA, and 80% formamide] (Shambrook *et al.*, 1989). The solution was heated for 10 min at 75 °C and incubated at 40 °C overnight for hybridization of the DNA probe with E1β-specific mRNA. Unhybridized DNA was digested with 500 units of S1 nuclease in S1 buffer [0.25 M NaCl, 30 mM potassium acetate (pH 4.5), 1 mM $ZnSO_4$, and 5% glycerol] at 40 °C for 1 h. Nucleic acids were extracted with phenol and

chloroform and precipitated with ethanol, and the pellet was dissolved in deionized formamide containing tracking dyes. The solution was heated to denature the nucleic acids and then loaded on a sequencing gel along with dideoxy sequencing ladders for precise sizing.

Preparation of Nuclear Extracts. Crude liver nuclear extracts were prepared from male Sprague-Dawley rats, weighing about 250 g, according to the published procedure (Gorski *et al.*, 1986). The protein content of the extracts was determined by dye binding (Bradford, 1976) using bovine serum albumin as a standard. The extracts were stored at −70 °C until used. Repeated freezing and thawing of nuclear extracts was avoided (Hennighausen & Lubon, 1987).

Footprinting Analysis. Protein−DNA binding for DNase I footprinting was performed in a total volume of 50 μL containing rat liver nuclear extract (25−90 μg of protein), 2 ng of labeled (50 000 cpm) DNA probe, 1 μg of poly(dI-dC)·poly(dI-dC), 20 mM Hepes, pH 7.6, 0.1 mM EDTA, 1 mM dithiothreitol, 10% (v/v) glycerol, 2% (v/v) polyvinyl alcohol, and 50 mM NaCl. The reaction mixture was incubated on ice for 20 min and then at room temperature for 2 min. Five microliters of a 100 mM $MgCl_2$ and 30 mM $CaCl_2$ solution was added, followed by 2 μL of diluted DNase I. After 1 min, the digestion was stopped by the addition of 150 μl of stop solution (150 mM NaCl, 0.7% SDS, 15 mM EDTA, and 30 μg/mL tRNA). The samples were extracted with phenol and chloroform, ethanol precipitated, washed with 70% ethanol, dried, and dissolved in formamide−dye mix. The samples were heated at 90 °C for 1 min and loaded onto a 6% acrylamide gel containing 7 M urea (Roesler *et al.*, 1989).

Cell Culture, Transfection, and CAT Assays. The pβCAT or the pβTKCAT chimeric construct (5 μg) was transfected in human hepatoma (HepG2) cells, and the cells were maintained in a modified Eagle's medium supplemented with 10% fetal bovine serum. The RSVβgal plasmid (3 μg) containing the Rous sarcoma virus promoter in front of the bacterial β-galactosidase gene in pBluescript (SK^+) was cotransfected with pβCAT or pβTKCAT constructs so as to monitor, and subsequently correct for, the transfection efficiency. The cells were grown in 10-cm plates to 50−70% confluence, treated with trypsin, centrifuged, and resuspended in 1 mL of medium. Five to 10 μg of DNA was precipitated using the standard calcium phosphate method (Sambrook *et al.*, 1989), and the calcium phosphate−DNA precipitates were mixed thoroughly with 1 mL of cells. The DNA−cell mixture was left for 4−6 h, washed with phosphate-buffered saline, and transferred to plates containing 10 mL of complete medium that was incubated for about 48 h prior to harvesting. The cell-free extracts were used for the CAT activity assay by the phase-extraction method (Seed & Sheen, 1988). The reaction conditions were selected to allow the conversion of [^{3}H]chloramphenicol in the presence of butyryl-CoA to the product within the linear range (0.01−50%). The butyrylated products were extracted into the xylene phase and measured by scintillation counting. Percentage conversion represents the relative CAT activity as a fraction of the total [^{3}H]chloramphenicol added to the assay. The CAT activity was normalized to the β-galactosidase activity and expressed as fold changes in activity over that of the full-length (−2316 to +32) E1β construct. The protein content was estimated by standard methods (Bradford, 1976).

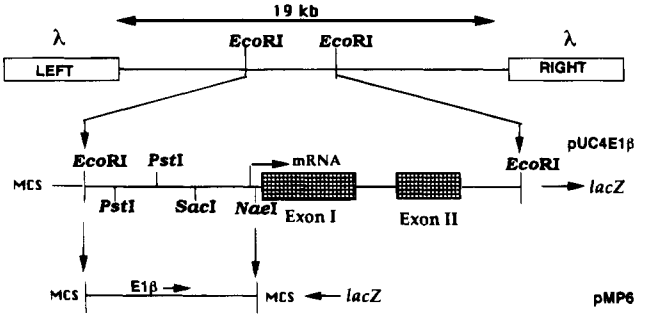

FIGURE 1: Physical map and subcloning of the promoter-regulatory region of the E1β gene of human PDC. The 19-kb insert in λEMBL3 was digested with *Eco*RI, and the 4-kb insert containing the promoter-regulatory region, exon 1, intron 1, and exon 2 was subcloned into pUC19 as detailed in Materials and Methods. Appropriate restriction sites of the multiple cloning site (MCS) of the vector and insert are shown. The insert cloned is shown with respect to the direction of *lacZ* transcription.

RESULTS AND DISCUSSION

Screening the EMBL3 Genomic Library. A commercial human placental genomic library constructed in EMBL3 was screened with a 0.5-kb (*Eco*RI) fragment (containing the leader sequence and nucleotides encoding 148 amino acids) of a human E1β cDNA (Ho & Patel, 1990). One positive clone containing the 19-kb insert was identified and mapped with restriction enzyme digestion and Southern blotting using a 0.5-kb or a 90-bp (encoding the leader sequence) fragment derived from the human E1β cDNA (Ho & Patel, 1990). The phage insert contained four *Eco*RI fragments (only two *Eco*RI sites, which were used for cloning of the 4-kb fragment are shown in Figure 1). This 4-kb fragment harbored 2.5 kb of the promoter, the first exon, the first intron, the second exon, and a part of the second intron sequence (Figure 1). The intron−exon junctions were confirmed by DNA sequencing.

Nucleotide Sequence of the Promoter Region of the E1β Gene. The partial nucleotide sequence (1.2 kb) was determined in both directions (Figure 2). Six direct repeats and seven inverted repeats were identified with the minimum repeat lengths chosen as 6 and 7 nucleotides for inverted and direct repeats, respectively. The longest direct repeat found was 11 bp in the 5′ untranslated region. There was no clustering of these repeats in any part of the DNA sequence, and the G+C contents of the promoter region and the cDNA sequence were approximately 50% and 47%, respectively. When the 1245-bp nucleotide sequence of pMP14 was compared to the published sequence of the human E1β gene (Koike *et al.*, 1990), 18 differences in nucleotides (three additions, seven deletions, and eight substitutions) were observed including a deletion of a cytidine residue in the 5′ untranslated region (Figure 2). The variations between our sequence and that of Koike *et al.* (1990) may be attributed largely to the differences in methodologies and reagents (that would resolve GC compressions) used rather than to polymorphism.

The nucleotide sequence of the promoter region is important in terms of putative protein binding sites and *cis*-acting elements (Xiao *et al.*, 1991) that might be involved in the regulation of the E1β gene. Both strands of the nucleotide sequence of the E1β promoter region were

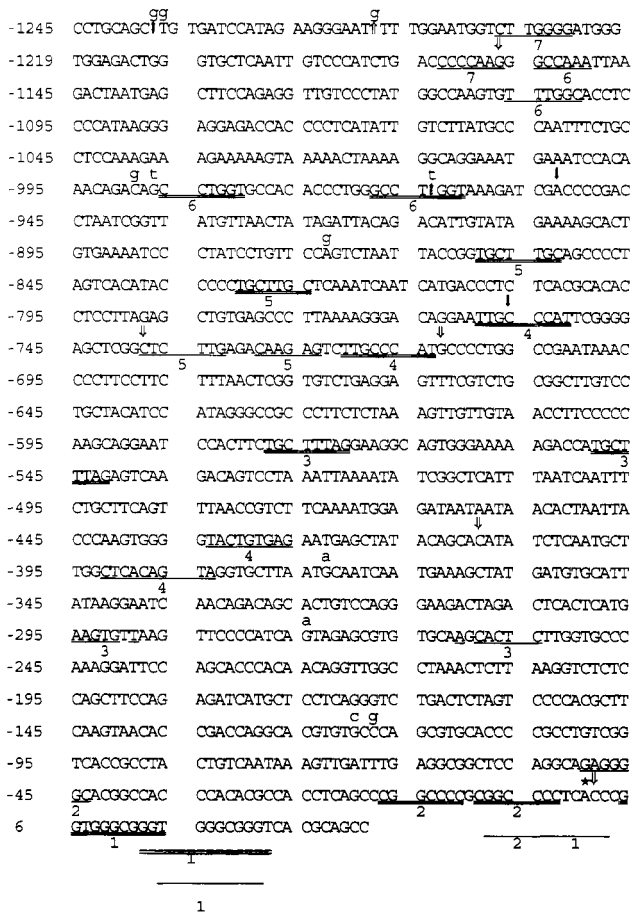

FIGURE 2: Nucleotide sequence of the promoter-regulatory region of the E1β gene of human PDC. The transcriptional start (+1) of the mRNA is shown by a star. The nucleotide numbering is relative to the transcriptional start site. The differences between our DNA sequence and the published sequence (Koike *et al.*, 1990) are also identified. The addition and deletion of bases are indicated by up and down arrows, respectively, compared to the published sequence (Koike *et al.*, 1990). Base substitutions are shown by lowercase letters, the top nucleotide being the base reported in the previous report (Koike *et al.*, 1990). Direct and inverted repeats in the DNA sequence are shown by double and single underlines, respectively.

analyzed for various *cis*-acting elements as well as for putative protein-binding sequences. The E1β promoter showed a CAAT box (−81 to −78), but no TATA box. The absence of a TATA box and the presence of multiple GC boxes have been reported for promoters of several housekeeping genes which encode enzymes involved in metabolic reactions in many cell types (Valerio *et al.*, 1985; Melton *et al.*, 1984; McGrogan *et al.*, 1985; Sigel-Sam *et al.*, 1984). The E1β promoter contained two initiator (Inr) sequences (TCAGCCCG/TCACCCGG) at −24 and −2 bp, respectively, and these sequences are 87% homologous to the consensus initiator sequence (Buchner, 1990). Inr sequences are reported to play an important role in transcription initiation in TATA-less promoters (Smale & Baltimore, 1989).

Transcriptional Start of the E1β Gene. The transcriptional start site of the E1β gene of human PDC was confirmed by S1 nuclease mapping. Treatment of the DNA/RNA hybrid with S1 endonuclease followed by denaturing gel electrophoresis showed a doublet, the components of which differed by a nucleotide in length, whereas S1 nuclease totally digested the radiolabeled probe (results not shown). The two bands of the DNA/RNA hybrid obtained for the sample corresponded to the cytidine and adenine (+2 and +1, respectively) residues. Koike *et al.* (1990) reported the position of the adenine residue as +1 from their transcript mapping using primer extension analysis. The appearance of doublets in S1 nuclease mapping has been reported for other promoters as well (Aldea *et al.*, 1988; Debarbouille & Raibaud, 1983).

DNase I Footprinting. The protein-binding regions of the E1β 5'-flanking region from +32 to −1245 was mapped by DNase I footprinting as shown in Figure 3 . The amount of the carrier DNA (poly[d(I-C)]) was carefully titrated for each batch of nuclear extract prepared. No variation in footprint pattern was observed when the carrier DNA used was in amounts of 0.5−1 μg per 25−100 μg of protein in rat liver nuclear extracts (data not shown).

The entire promoter-regulatory region showed seven protein-binding sites termed P1−P7. The footprinting patterns of coding and noncoding strands of P1−P7 are shown in Figure 3. The regions protected by rat liver nuclear extracts appeared to be quite asymmetric for all the sites except P4. Sites P2 and P3 showed a gap in protein binding on both strands as compared to P1, which showed a gap of one nucleotide on the noncoding strand only. Regions P4 and P7 were >35 nt (Figure 3), indicating that several proteins might be involved in binding at this region. Alternatively, it is possible that there is no specific binding of proteins to the nucleotides in this region and that there may exist enhancer-specific core binding proteins.

A marked hypersensitive site was observed on the noncoding strand of P2 that spanned 3 nt (−384 to −382). Nuclease hypersensitivity is usually confined to discrete domains within the 5'- and 3'-flanking regions of the expressed genes (Wu, 1980; Elgin, 1981), and evidence from a variety of mammalian systems indicates that hypersensitivity is correlated with the binding of *trans*-acting factors to these regions (Weintraub, 1985; Piette *et al.*, 1985; Plumb *et al.*, 1985). In particular, it has been shown that the nuclease-hypersensitive domain of the chicken adult β-globin 5'-flanking DNA interacts both *in vivo* and *in vitro* with factors present in chicken embryo nuclei (Jackson & Felsenfeld, 1985; Emerson *et al.*, 1985).

The nucleotide sequences of the E1β promoter-regulatory region protected by rat liver nuclear extracts (on coding and noncoding strands) were analyzed for known transcription factor binding/recognition sites (Mitchell & Tjian, 1989) using the VAX/VMS computer program tfsites, version 7.30. The DNase I protected regions contained Sp1, IgNF-A, GATA-1, bicoid Q9, GH, ISGF3, NF-1, AP-1, AP-3, NF-κB, Lva, HNF-5, H4TF-1hist, WAP5, and ADH recognition sites. The promoter-regulatory region showed a CAAT box and an Sp1 site that were protected in P1 (Figure 4). The E1β promoter showed multiple GC boxes between +1 and −200 nt (Figure 2). Pugh and Tjian (1990) have demonstrated that the same general factors required for transcription of polymerase II dependent promoters containing a TATA box are also required for transcription of TATA-less promoters containing multiple GC boxes. Promoters that lack a TATA box but contain GC boxes require Sp1 in addition to other factors such as TAF110 and Sp1 (Hoey *et al.*, 1993). It has been noted that GC-rich regions of many promoters contain unusual numbers of GC dinucleotides, relative to the bulk of the genome, and that these are undermethylated. It is possible that the presence of a cellular protein such as

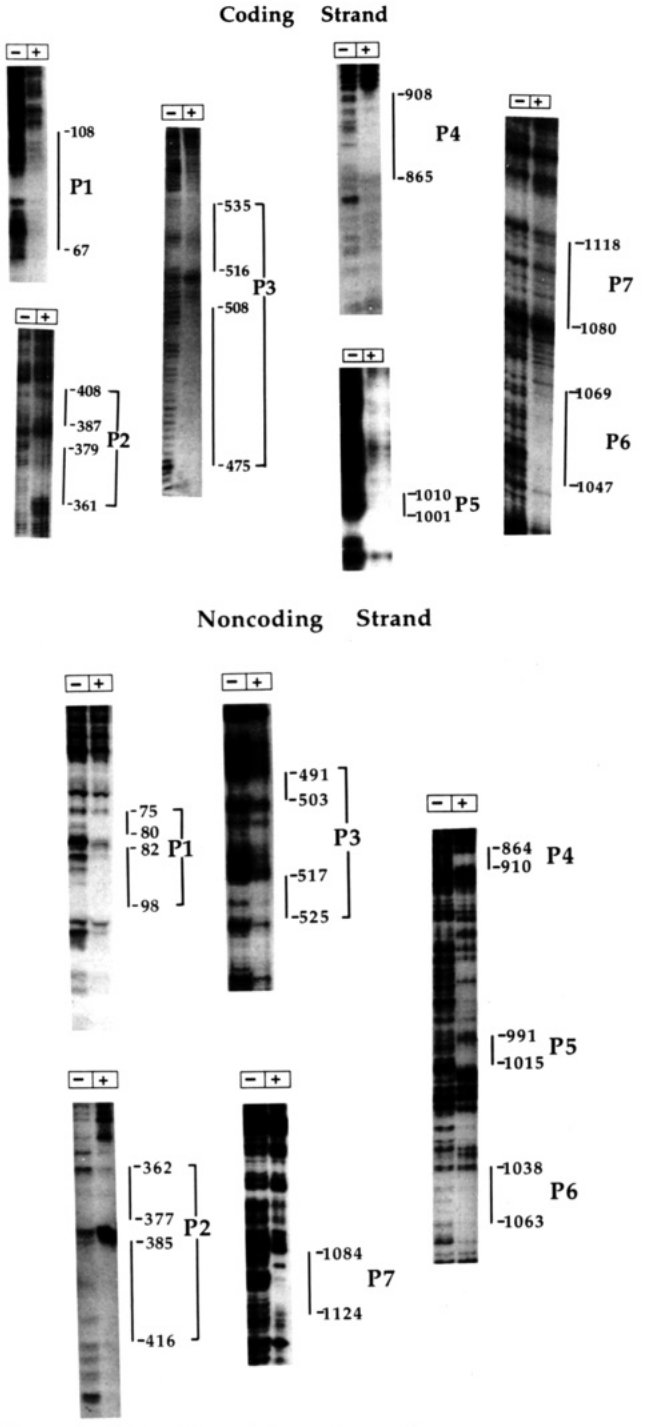

FIGURE 3: DNase I footprinting analyses of the promoter-regulatory region of the human E1β gene. DNase I footprinting was performed using 25−90 μg of rat liver nuclear extract (indicated by a plus sign). The DNA fragments 200−700 bp long were labeled at the 3′ end as detailed in Materials and Methods. The footprint patterns of coding (A) and noncoding (B) strands are shown separately. The protected regions are outlined and denoted as P1−P7. The nucleotides protected in footprints were identified by Maxam−Gilbert chemical sequencing. The nucleotides protected by DNase I are shown in the margin for each region, and the nucleotide number is relative to the transcriptional start site.

Sp1 which hinders methylation might fulfill this role in at least some instances (Dynan, 1986).

E1α and E1β are found in cells as a heterotetramer, and they are present in a 1:1 molar ratio. This raises the possibility that these two genes may be coordinately expressed. Comparing the DNase I footprinting pattern of the 5'-flanking region of the E1β gene of human PDC to that of the E1α gene (Chang *et al.*, 1993) reveals several differences. The promoter-regulatory region of the E1α gene of human PDC contained consensus sequences to TATA box, AP-1, AP-2, CAAT, and CACCC within 200 bp of the transcriptional start site. However, there was no obvious relationship between the two promoter regions which would indicate obvious sites for similar transcriptional regulation.

Functional Analysis of the E1βCAT and E1βTKCAT Constructs. To define the regulatory elements required for E1β gene expression, nested deletions of the E1β 5′-flanking region (+32 to −2316) were created, and these DNA fragments were inserted into the BSKCAT vector. The constructs were transfected into human hepatoma (HepG2) cells, and the transient expression of the CAT activity was measured after normalization for transfection efficiency with β-Gal. The full-length E1β (pβCAT1 and pβCAT1-R) constructs (−2316 to +32) gave a very low CAT activity in either orientation (Figure 5). However, upon deletion of the 5′-sequences (from −2316 to −930), the CAT activity of pβCAT2 (containing −929 to +32) increased approximately 7-fold compared to pβCAT1. As shown earlier, DNase I footprinting showed protein-binding sites (P5−P7) that had consensus sequences for several DNA-binding proteins (Figure 4), and these were deleted in the pβCAT2 construct. Compared to pβCAT2, there were no marked changes in the CAT activity of pβCAT3 and pβCAT4, which contained the DNA sequences from −564 to +32 and from −397 to +32, respectively. These constructs did not have P3 to P7 binding sites. However, when the deletion was extended to −282 nt (pβCAT5), CAT activity increased approximately 15-fold compared to pβCAT1. The plasmid construct pβCAT5 did not have the sequences of the protein-binding sites of the P2−P7 region. Further deletion to −98 nt (pβCAT6) resulted in a nearly 20-fold increase in CAT activity compared to pβCAT1. This maximum activity was achieved with a minimal promoter containing a potential Sp1 binding site (−102), a CAAT box (−78), and a P1 footprinted region (Figures 3 and 4). Two additional potential Sp1 binding sites (+8 and +16 bp) are present in the 5′ untranslated region of the E1β gene which may also contribute to the activity of pβCAT6 (Figure 5). These results strongly suggest the presence of one or more negative elements in the regions from −2316 to −929 and from −397 to −98 nt.

To further localize the repressor elements in the E1β promoter, DNA fragments containing the repressor were cloned in front of the TK promoter to test whether they affect expression of a minimal promoter in a heterologous construct. The DNA fragments were cloned in both orientations and are termed pβTKCAT1 to pβTKCAT6 as shown in Figure 6. Constructs pβTKCAT1 and pβTKCAT2 contained the sequence from −2316 to −929 nt in both orientations and repressed CAT activity from pTKCAT by 50% (Figure 6). The E1β promoter region from −397 to −98 nt was also cloned in both orientations, yielding four constructs (pβ-TKCAT3 to pβTKCAT6). Surprisingly, the CAT expression of pβTKCAT3 and pβTKCAT5 constructs was approximately 50% higher than that of pTKCAT. pβTKCAT4 and pβTKCAT6 were mildly inhibitory, exhibiting 40% and 74% of pTKCAT activity, respectively. This suggests the presence of an orientation-dependent *cis* element (or elements) that can act as an enhancer or a repressor on a heterologous

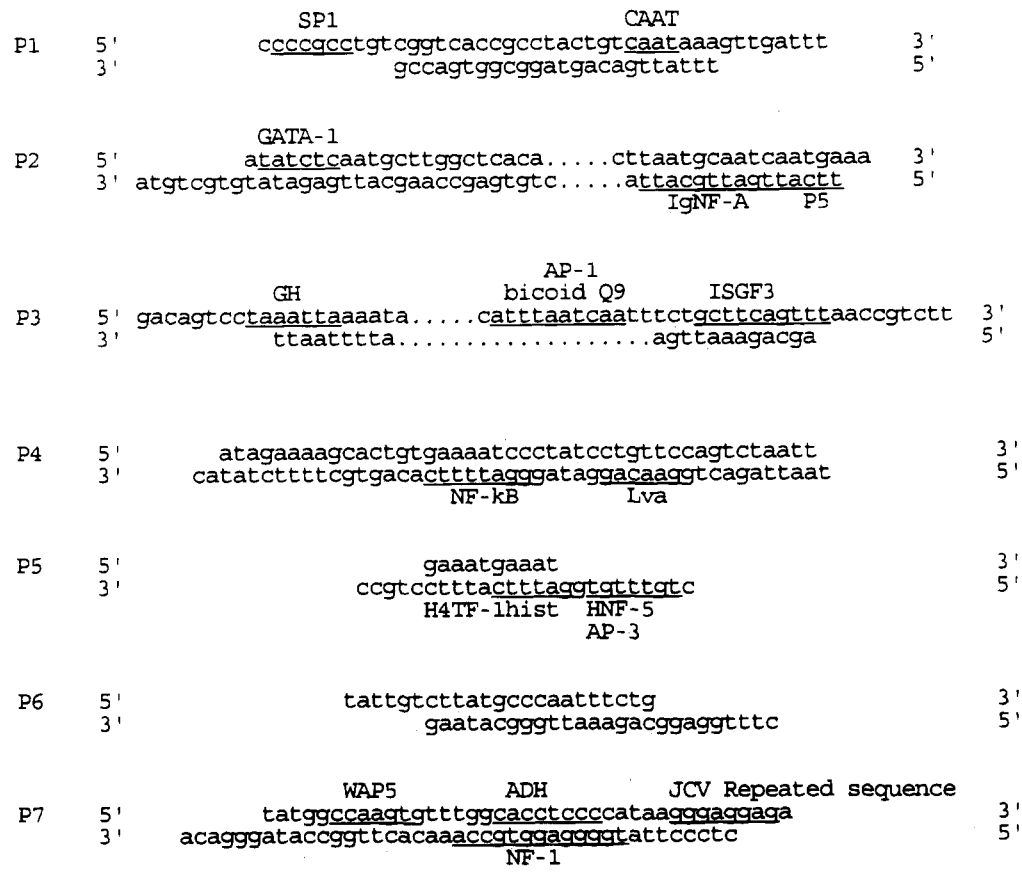

FIGURE 4: Protein-binding sites (P1−P7) of the promoter-regulatory region of the human E1β gene. The protected nucleotides of the coding and noncoding strands of each binding site are shown. The nucleotide sequences within the protected regions were compared to the consensus sequences of the known *trans*-acting factors, and those that showed complete nucleotide identity are shown.

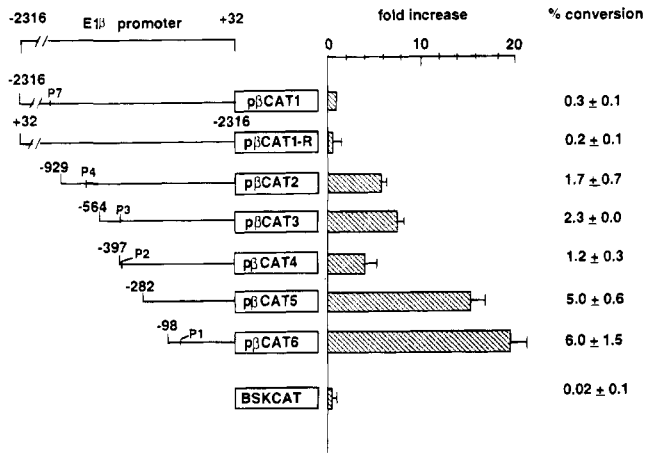

FIGURE 5: Transient expression of CAT activity by the human E1β promoter in HepG2 cells. The full-length E1β promoter (−2316 to +32 nt) or the deletion fragments were cloned into the BSKCAT vector, and the constructs (5 μg) were transfected into human hepatoma (HepG2) cells. The cells were cotransfected with RSVβgal, and the β-galactosidase activity of the cell-free extracts was used to correct for the transfection efficiency. The CAT activity (mean ± SE; $n = 6$) shown for each construct is the fold change compared to the activity of pβCAT1. Percentage conversion represents the relative CAT activity as a fraction of the total [^{3}H]-chloramphenicol added to the assay.

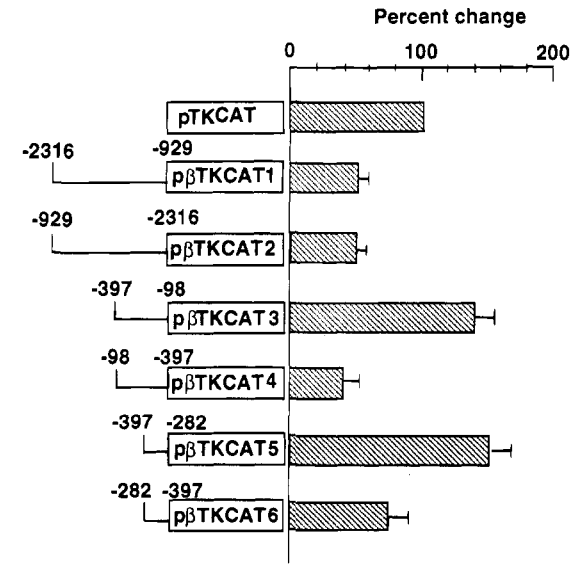

FIGURE 6: Effect of the human E1β promoter fragments on the transcriptional activity of the thymidine kinase (TK) promoter. The DNA fragments of the human E1β promoter were ligated in both orientations into the *Xba*I site at the 5′ end of the TK promoter in the TKCAT plasmid (pTKCAT). These constructs were assayed for CAT expression in HepG2 cells, and the CAT activity was expressed as a percentage of the activity of pTKCAT.

promoter. A deletion analysis combined with mutagenesis of the E1β promoter region should define more precisely the repressor and enhancer elements present in this region of the human E1β promoter.

The absence of a TATA box in the core promoter region and the presence of Sp1 and Inr elements in the E1β promoter suggest a role for Sp1 in E1β transcription. Sp1 has been shown to play a major role in transcription start site selection in TATA-less promoters as demonstrated in carbamoyl-phosphate synthase (glutamine-hydrolyzing)/aspartate carbamoyltransferase/dihydroorotase promoter (Kollmar *et al.*, 1994). Inr elements are believed to mediate the same

functions as TATA elements and play an important role in localizing the start of transcription, as the conserved nucleotides in the Inr are required for recognition by RNA polymerase II (Means & Farnham, 1990; Zawel & Reimberg, 1993). The human E1β promoter contains two initiator sequences (TCAGCCCG/TCACCCGG) at −24 and −2 bp, respectively, of the core promoter. These sequences are nearly 87% homologous to the consensus initiator sequence (KCABHYBY; K = G or T; B = C, G, or T; H = A, C, or T; Y = C or T; Buchner, 1990), which is significant because only the one at −2 bp is used (Koike *et al.*, 1990; this study). The reason for this specificity remains to be investigated. Inr elements function in cooperation with Sp1 elements, and almost every Inr that has been described functions with upstream Sp1 binding sites. The E1β promoter also contains two Sp1 sites (+8 and +16) in the promoter region which may allow for efficient CAT expression of pβCAT6 (Figure 5). Further mutational analysis of this start region (especially the two initiator sequences and the three Sp1 sites) should shed light on the minimal sequence requirement for sufficient and accurate initiation of transcription of the human E1β gene.

ACKNOWLEDGMENT

We thank Dr. David Samols of Case Western Reserve University and Dr. Alfred Ponticelli of State University of New York at Buffalo for their helpful discussions and critical reading of the manuscript. We also thank Ellen Stacey for help in the preparation of the manuscript.

REFERENCES

Aldea, M., Claverie-Martin, F., Diaz-Torres, M. R., & Kushner, S. R. (1988) *Gene 65*, 101−110.
Bradford, M. A. (1976) *Anal. Biochem. 72*, 2248−2254.
Buchner, P. (1990) *J. Mol. Biol. 212*, 563−578.
Chang, M., Naik, S., Johanning, G. L., Ho, L., & Patel, M. S. (1993) *Biochemistry 32*, 4263−4269.
Debarbouille, M., & Raibaud, O. (1983) *J. Bacteriol. 153*, 1221−1227.
Dynan, W. S. (1986) *Trends Genet. 2*, 196−197.
Elgin, S. C. R. (1981) *Cell 27*, 413−415.
Emerson, B. M., Lewis, C. D., & Felsenfeld, G. (1985) *Cell 41*, 21−30.
Gorman, C. M., Moffat, L. F., & Howard, B. H. (1982) *Mol. Cell. Biol. 2*, 1044−1051.
Gorski, K., Carneiro, M., & Schibler, U. (1986) *Cell 47*, 767−776.
Henikoff, S. (1984) *Gene 28*, 351−359.
Hennighausen, L., & Lubon, H. (1987) *Methods Enzymol. 152*, 721−735.
Ho, L., & Patel, M. S. (1990) *Gene 86*, 297−302.
Ho, L., Wexler, I. D., Liu, T.-C., Thekkumkara, T. J., & Patel, M. S. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5330−5334.
Hoey, T., Weinzieri, R. O., Gill, G., Chen, J. L., Dynlacht, B. D., & Tjian, R. (1993) *Cell 72*, 247−260.
Jackson, P. D., & Felsenfeld, G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2296−2300.
Jacoby, D. B., Zilz, N. D., & Towle, H. C. (1989) *J. Biol. Chem. 264*, 17623−17626.
Kennedy, G. C., Anderson, B., & Nilson, J. H. (1990) *J. Biol. Chem. 265*, 6279−6285.
Koike, K., Urata, Y., & Koike, M. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5594−5597.
Kollmar, R., Sukow, K. A., Sponagle, S. K., & Farnham, P. J. (1994) *J. Biol. Chem. 269*, 2252−2257.
Maragos, C., Hutchison, W. M., Hayasaka, K., Brown, G. K., & Dahl, H.-H. M. (1989) *J. Biol. Chem. 264*, 12294−12298.
McGrogan, N. M., Simonsen, C. C., Smouse, D. T., Farnham, P. J., & Schimke, R. T. (1985) *J. Biol. Chem. 260*, 2307−2314.
Means, A. L., & Farnham, P. J. (1990) *Mol. Cell. Biol. 10*, 653−661.
Melton, D. W., Konecki, D. S., Brennand, J., & Caskey, C. T. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2147−2151.
Mitchell, P. J., & Tjian, R. (1989) *Science 245*, 371−378.
Patel, M. S., & Roche, T. E. (1990) FASEB J. *4*, 3224−3233.
Piette, I., Kryszke, M. H., & Yanif, M. (1985) *EMBO J. 4*, 2675−2685.
Plumb, M. A., Nicholas, R. H., Wright, C. A., & Goodwin, G. H. (1985) *Nucleic Acids Res. 13*, 4047−4065.
Pugh, B. F., & Tjian, R. (1990) *Cell 61*, 1187−1197.
Roesler, W. J., Vandenbark, G. R., & Hanson, R. W. (1989) *J. Biol. Chem. 264*, 9657−9664.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Vols. 1−3, Cold Spring Harbor Laboratory Press, Cold Spring Habor, NY.
Seed, B., & Sheen, J. (1988) *Gene 67*, 271−277.
Sigel-Sam, H., Keith, S. H., Tani, K., Simmer, R. L., Shively, L., Lindsay, S., Yoshida, A., & Rigs, A. D. (1984) *Gene 32*, 409−417.
Smale, S. T., & Baltimore, D. (1989) *Cell 57*, 103−113.
Valerio, D., Duyvesteyn, M. G. C., Dekker, B. M. M., Weeda, G., Berkvens, T. M., Van der Voorn, L., Van Ormondt, H., & Van der Eb, A. J. (1985) *EMBO J. 4*, 437−443.
Weintraub, H. (1985) *Cell 42*, 705−711.
Wu, C. (1980) *Nature 286*, 854−860.
Xiao, J. H., Davidson, I., Matthes, H., Garnier, J.-M., & Chambon, P. (1991) *Cell 65*, 551−568.
Zawel, L., & Reimberg, D. (1993) *Prog. Nucleic Acid Res. Mol. Biol. 44*, 67−108.

BI941612X