Multiple Protein-Binding Domains and Functional cis-Elements in the 5′-Flanking Region of the Human Pyruvate Dehydrogenase α-Subunit Gene†

Mei Chang,‡ Sharon Naik,§ Gary L. Johanning,‖ Lap Ho,§ and Mulchand S. Patel*,‡,§

Department of Biochemistry, Center for Inherited Disorders of Energy Metabolism, and Pew Center for Molecular Nutrition, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106

Received December 1, 1992; Revised Manuscript Received February 11, 1993

ABSTRACT: We have characterized the 5′-flanking region of the α-subunit gene of the human pyruvate dehydrogenase (E_1). DNase I footprinting with rat liver nuclear extracts identified 7 major protein-binding domains termed P1 through P7 in a 796 base pair DNA fragment (base pairs −763 to +33). P1 through P4 are clustered in the −221/+33 region. These protein-binding domains contain several known consensus sequences such as a TATA box, CAAT box, Sp1, and CRE, which all have previously been implicated in the constitutive transcription of several genes. Oligonucleotide competition studies indicate that oligonucleotides specific for CTF/NF-1 and Sp1 displaced the nuclear proteins bound to the CAAT box (within P_3) and an Sp1 site (within P_4), respectively. Several other well-characterized and purified transactivators (c-Fos, c-Jun, C/EBP, AP-2, and Sp1) have been shown to bind to the −221/+33 region. Other elements located upstream of the −221/+33 region, which includes nuclease protection domains P5–P7, are required for enhanced promoter activity of the 796 bp sequence. Promoter activity was measured by transient expression of a chloramphenicol acetyltransferase gene ligated to deletion fragments of the 5′-flanking region. Crucial element(s) for promoter activity and complex DNA–nuclear protein interactions were confined within a region spanning −221/+33. This region also retained more than 75% of the promoter activity of the 796 bp sequence. Additionally, this promoter region shows characteristics of both facultative and housekeeping gene promoters, suggesting complex transcriptional regulation.

Mammalian pyruvate dehydrogenase complex (PDC),[1] a mitochondrial multicomponent enzyme, plays a pivotal role in energy metabolism by catalyzing the oxidative decarboxylation of pyruvate resulting in conversion to acetyl-CoA. This complex consists of three catalytic components: pyruvate dehydrogenase (E_1), dihydrolipoamide acetyltransferase, and dihydrolipoamide dehydrogenase. Other proteins of this complex include two regulatory proteins, E_1 kinase and phospho-E_1 phosphatase, and protein X (Reed, 1974; Patel & Roche, 1990). E_1, the first catalytic component of the PDC, catalyzes the irreversible decarboxylation of pyruvate, which is the rate-limiting step in the overall reaction of the complex. E_1 is composed of two nonidentical subunits, α and β, with molecular masses of 41 000 and 36 000 Da, respectively, and has a tetrameric ($\alpha_2\beta_2$) structure (Reed, 1974). The human α gene is located on chromosome X (Brown et al., 1989) and the β gene on chromosome 3 (Olson et al., 1990). Additionally, another functional $E_1\alpha$-processed gene in humans has been identified on chromosome 4 (Dahl et al., 1990).

The activity of mammalian E_1 is regulated posttranslationally by the phosphorylation/dephosphorylation of the $E_1\alpha$ peptide (Reed, 1974). The level of $E_1\alpha$ mRNA is regulated by the glucose concentration in cultured pancreatic islets and by fasting in isolated islet preparations (MacDonald et al., 1991). Nevertheless, the potential role of cis-elements and trans-acting elements in the regulation of $E_1\alpha$ gene transcription has not been examined. The genomic structure and the 5′-flanking sequence of the human $E_1\alpha$ gene have been reported (Maragos et al., 1989). This information provides the basis for further investigation of the mechanisms underlying regulated expression of this gene. In this paper, we report that the proximal promoter of the human $E_1\alpha$ gene contains multiple regulatory elements which may coordinate its transcriptional regulation.

MATERIALS AND METHODS

Materials. DNA-modifying enzymes and poly[d(I·C)] were products of Boehringer Mannheim. The polymerase chain reaction reagents were from Perkin-Elmer/Cetus. [γ-^{32}P]ATP (6000 Ci/mmol) and [α-^{32}P]dNTP (6000 Ci/mmol) were purchased from DuPont—New England Nuclear. Purified recombinant rat c-Fos and c-Jun transcription factors were expressed using vectors pCMV-Fos and pCMV-Jun containing the entire open reading frame of either c-Fos or c-Jun downstream of the cytomegalovirus (CMV) promoter. These vectors were the gifts of Dr. Tom Curran (Department of Molecular Oncology and Virology, Roche Institute of Molecular Biology). Purified recombinant human AP-2, Sp1, and TFIID transcription factors were products of Promega. Recombinant C/EBP factor prepared by overexpression in *Escherichia coli* cells was the gift of Dr. Steven L. McKnight (Department of Embryology, Carnegie Institute of Washington). Oligonucleotide PCR primers were chemically synthesized using an Applied Biosystem 380A DNA synthesizer. AP-1, AP-2, CTF/NF-1, and Sp1 consensus oligomers were products of Promega.

Enzymatic Amplification of the $E_1\alpha$ Promoter-Regulatory Region by PCR. Genomic DNA isolated from cultured human skin fibroblasts (Wigler et al., 1979) was used for the enzymatic

† This work was supported in part by U.S. Public Health Service Grant DK20478, Center for Inherited Disorders of Energy Metabolism Grant MCJ-009122 (M.C.), Metabolism Training Grant AM07319 (G.L.J. and L.H.), and a Pew Nutrition grant (G.L.J.).

* Address correspondence to this author at the Department of Biochemistry, Case Western Reserve University School of Medicine, 10900 Euclid Ave., Cleveland, OH 44106-4935.

‡ Center for Inherited Disorders of Energy Metabolism.

§ Department of Biochemistry.

‖ Pew Center for Molecular Nutrition.

[1] Abbreviations: PDC, pyruvate dehydrogenase complex; E_1, pyruvate dehydrogenase.

0006-2960/93/0432-4263$04.00/0

amplification in the presence of specific primers. The sense primer spanned bases –763 to –743 (5′-CCCCCTTAAAAT-GTGGAGCCC-3′) from the putative transcription start site of the human $E_1\alpha$ gene (Maragos et al., 1989). Antisense primer (5′-CTTCAGGTGCCCCAGCAGA-3′) was complementary to bases +15 to +33 (Maragos et al., 1989) of the human $E_1\alpha$ gene. The amplification mixture (100 μL) contained 2.5 units of *Thermus aquaticus* (Taq) polymerase, 100 pmol each of the sense and antisense primers, 200 nmol of each dNTP, and 3 μg of genomic DNA. Each round of amplification consisted of denaturation at 94 °C (2 min), annealing at 62 °C (3 min), and primer extension at 72 °C (4 min). Thirty rounds of amplification were performed using a DNA thermal cycler (Perkin-Elmer/Cetus). The sizes of the DNA fragments amplified were examined by electrophoresis through a 1% agarose gel.

Construction of Plasmids and DNA Sequencing. The construction of plasmids containing the 5′-flanking region of the $E_1\alpha$ gene is described under Results. The nucleotide sequence of this region was determined by dideoxynucleotide chain termination sequencing (Sambrook et al., 1989) with Sequenase kit 2.0 (U.S. Biochemical). The 796 base pair promoter region of $E_1\alpha$ was sequenced in both directions using at least 2 independent clones.

DNase I Footprinting. Liver nuclear extracts were prepared from male Sprague-Dawley rats (250 g) by the method of Gorski et al. (1986). Protein concentrations were determined according to Bradford (1976) using bovine serum albumin as the standard. The DNA probes used in DNase I footprinting analysis were prepared by end-labeling one strand of DNA using T_4 polynucleotide kinase and [γ-^{32}P]ATP or Klenow fragment and [α-^{32}P]dNTP (Sambrook et al., 1989). The radioactively labeled probes were isolated by electrophoresis through a 4% polyacrylamide gel. Binding assays were performed as described by Roesler et al. (1989) except that 1 μg of poly[d(I·C)] was used in 50 μL of reaction mixture. Ten to twenty femtomoles of radiolabeled DNA probe was used in each reaction. Five hundred to eight hundred milliunits of DNase I was used in each reaction mixture containing liver nuclear extract, while 50–60 milliunits of DNase I was used in control reactions and reaction mixtures containing purified transcription factors. Boundaries of the nucleotide sequences protected from DNase I digestion were deduced from Maxam–Gilbert "G" and "G+A" sequencing reactions analyzed in parallel with DNase I protection reactions (Sambrook et al., 1989).

Construction of Vectors for Transfection. A 796 bp DNA fragment (–763 to +33) containing the putative $E_1\alpha$ promoter was cloned into the *Eco*RV site of the Bluescript vector (BSK-CAT) containing the coding region of the *E. coli* chloramphenicol acetyltransferase (CAT) gene and the SV40 polyadenylation signal subcloned at the *Sma*I site of pBluescript (Kennedy et al., 1990). Four subsequent deletions from the 5′ end of the clone were generated from the above $E_1\alpha$ promoter–CAT construct (pαCAT1) using restriction sites at –506 bp (*Acc*I; pαCAT2), –333 bp (*Pst*I; pαCAT3), –221 bp (*Apa*I; pαCAT4), and –102 bp (*Stu*I; pαCAT5). The RSVβgal plasmid containing the β-galactosidase gene driven by the Rous sarcoma virus promoter/enhancer was cotransfected with the pαCAT constructs to allow correction for transfection efficiency. 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 the assays.

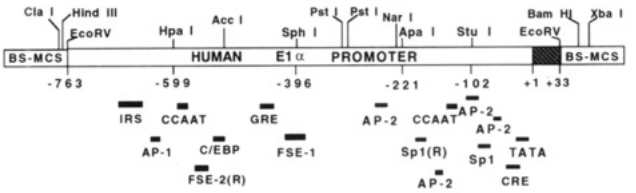

FIGURE 1: Partial restriction endonuclease map of the human $E_1\alpha$ promoter and polylinker region of Bluescript vector. The solid bars denote different cis-acting elements: TATA box, CAAT box, Sp1-binding sites (Sp1), AP-1-binding sites (AP-1), AP-2-binding sites (AP-2), glucocorticoid-responsive element (GRE), cAMP-responsive element (CRE), insulin-responsive sequence (IRS), fat-specific elements 1 and 2 (FSE-1 and FSE-2), and CCAAT/enhancer-binding protein (C/EBP). Reverse orientation of an element is identified by (R). +1 to +33 represents a part of the first exon. BS-MCS denotes Bluescript-multiple cloning sites. For the construction of vectors for transfection, the restriction endonuclease sites *Acc*I, *Pst*I, *Apa*I, and *Stu*I in the insert were used to create deletion vectors.

Cell Transfection and CAT Assays. HepG2 cells, a human hepatoma cell line, were transfected with DNA using the calcium phosphate precipitation procedure (Sambrook et al., 1989). Each transfection contained 5 μg of a pαCAT plasmid and 3 μg of RSVβgal. After 48 h, cells were harvested and lysed by freeze–thawing. CAT assays were performed by the phase-extraction method (Seed & Sheen, 1988) using [^{3}H]-chloramphenicol and butyryl-CoA as substrates. Cell extracts were incubated with the substrates for 2 h at 37 °C. The butyrylated chloramphenicol was extracted with xylenes and quantified by scintillation counting. CAT activities were normalized to β-galactosidase activity and expressed as fold changes in activity over that of the full-length (–763 bp) $E_1\alpha$ promoter. The optimum range for CAT assay is up to 50% conversion of the added labeled substrate (Seed & Sheen, 1988). The conversion of the radioactive substrate to its acyl derivative in the present study (prior to correction for transfection efficiency) ranged from 18 to 48%.

RESULTS

Identification of cis-Elements in the 5′-Flanking Region of the Human $E_1\alpha$ Gene. A 796 bp DNA fragment containing the proximal promoter and 33 base pairs of 5′-untranslated region of the $E_1\alpha$ gene was generated by polymerase chain reaction amplification using isolated genomic DNA from cultured normal human skin fibroblasts. This fragment was cloned into the *Eco*RV site of Bluescript KS vector. The sequences from at least two individual clones originating from two independent polymerase chain reaction amplifications were found identical to the corresponding nucleotide sequence (–763 to +33) previously reported for the human $E_1\alpha$ gene (Maragos et al., 1989). The G+C content of the 763 bp 5′-flanking region is high, with an average of 57%, compared with 40% for the entire human genome. The G+C content was even higher (71%) for the 300 bp immediately 5′ to the putative transcription start site. Both strands of the nucleotide sequence were searched for consensus sequences that have been reported as potential transcriptional regulators. Several consensus sequences in the 5′-flanking region of the $E_1\alpha$ gene which have not been described previously (Maragos et al., 1989; Ohta et al., 1989; Tomura et al., 1990) have been identified. They include the following: one activator protein 1 (AP-1) like motif (–610 to –603) (Lee et al., 1987; Curran & Franza, 1988); one cAMP-responsive element (CRE) (–34 to –27) (Short et al., 1986); two activator protein 2 (AP-2) like motifs (–106 to –99 and –172 to –166) (Mitchell et al., 1987; Imagawa et al., 1987); one CCAAT/enhancer-binding protein (C/EBP) (–527 to –519); and one insulin-responsive sequence (IRS) (–646 to –637) (O'Brien et al., 1990) (Figure 1).

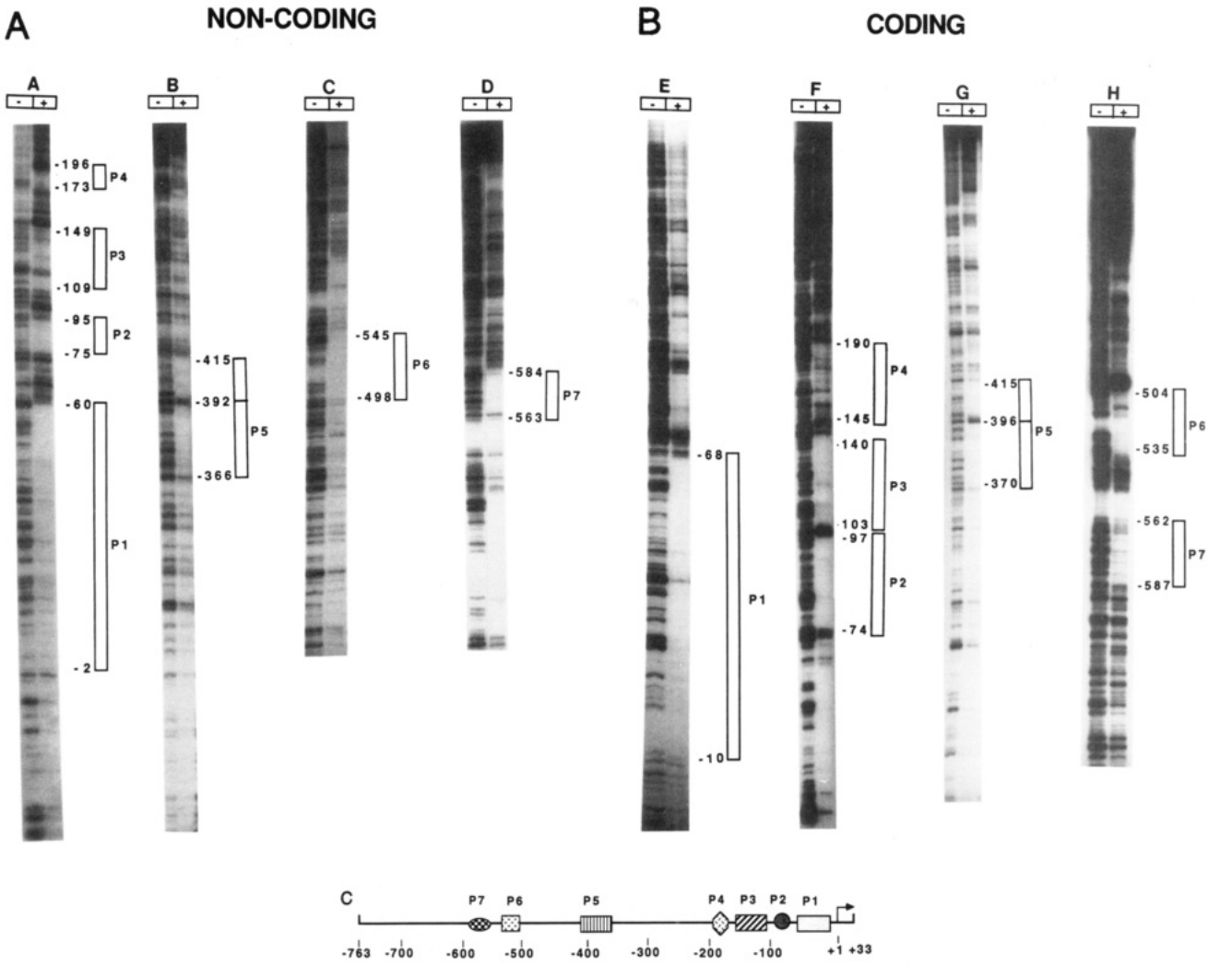

FIGURE 2: DNase I footprint analysis and schematic presentation of protein-binding domains of the $E_1\alpha$ promoter. DNase I footprinting was performed using 50 μg of rat liver nuclear extract. The *Xba*I/*Hin*dIII, *Cla*I/*Bam*HI, and *Nar*I/*Hin*dIII fragments are depicted in Figure 1. The *Xba*I and *Cla*I sites in the polylinker region and the *Nar*I site in the insert were end-labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP or Klenow fragment and [α-^{32}P]dNTP, respectively, in the footprinting analysis. (A) Noncoding strand; (B) coding strand; (C) schematic presentation. The protected regions are outlined by boxes and denoted as elements P1–P7. The positions of the binding sites relative to the initiation site of transcription are indicated.

Multiple Protein-Binding Domains in the 5′-Flanking Region of the Human $E_1\alpha$ Gene. To identify potential regulatory elements, we performed DNase I footprinting analysis of the human $E_1\alpha$ 5′-flanking region using rat liver nuclear extracts. Seven major protein-binding domains were identified in the $E_1\alpha$ 5′-flanking promoter-regulatory region (bp −763 to +33) (Figure 2). The boundaries of the domains on both strands are outlined with boxes and termed P1 through P7 (Figure 2A,B), and a schematic presentation of the protein-binding domains is shown in Figure 2C. The slight differences observed in the footprint patterns of the two strands at all binding sites indicated that the proteins may bind asymmetrically to the DNA. The footprint patterns of P5 on both the coding and noncoding strands were separated by a hypersensitive site, indicating the possibility of two distinct binding domains. Several domains showed sequence similarities to the following known consensus sequences: TATA box, CRE, and AP-2 consensus sequences in P1; Sp1 and CACCC in P2; CAAT box in P3; CACCC, Sp1, and AP-2 in P4.

Binding of Well-Characterized Transactivators to the 5′-Flanking Region of the $E_1\alpha$ Gene. To investigate whether the crude liver nuclear proteins which were shown to bind to the $E_1\alpha$ promoter by DNase I footprinting are related to some of the known transactivators, binding studies using well-characterized and purified regulatory proteins or footprint competition studies were conducted. These studies are summarized below:

(*A*) *CRE Site.* The P1 element, which footprints at positions −10 to −68 for the coding strand and at −2 to −60 for the noncoding strand, bound a protein which was relatively abundant in rat liver nuclear extract. When the amount of nuclear protein was limiting, this element was the first to be fully protected among P1–P4 elements (data not shown). This element contains a CRE-like motif, overlapping the TATA box. The CRE motif was reported previously to bind to Jun/Jun homodimers and Fos/Jun heterodimers as well as to the other members of the Jun and Fos families (Curran & Franza, 1988). DNase I footprinting assays were conducted using purified Fos and Jun (Abate et al., 1990). Although there is an AP-1-like motif in the $E_1\alpha$ promoter-regulatory region (Figure 1), the purified Fos/Jun bound to the CRE within the P1 element (Figure 3).

(*B*) *AP-2 Sites.* There are four AP-2-like elements present in the $E_1\alpha$ 5′-flanking promoter-regulatory region (−763 to +33 bp) (Figure 1). When purified AP-2 protein was used in the DNase I footprinting assays, two sites were protected from nuclease digestion (Figure 4). The proximal AP-2-binding site, which resembles an AP-2 consensus sequence of

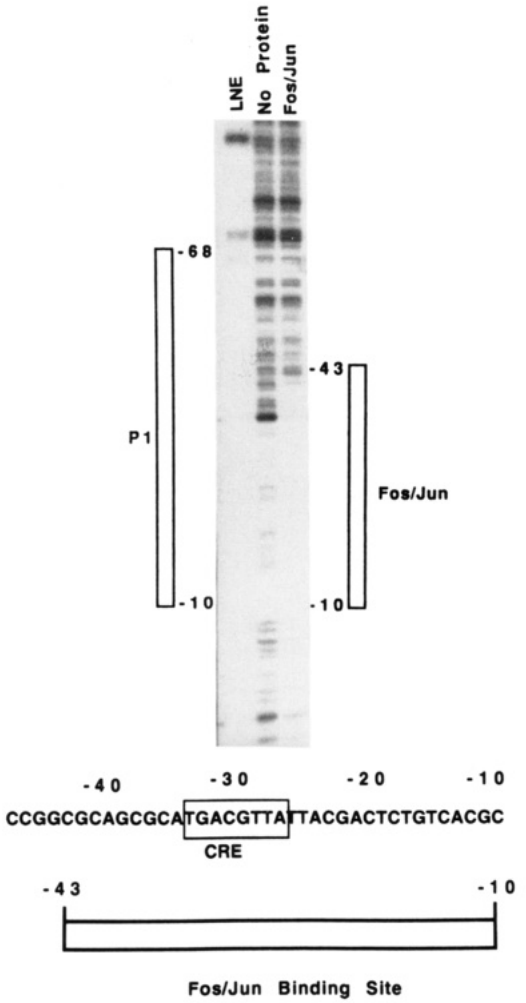

FIGURE 3: DNase I footprinting of the $E_1\alpha$ promoter with purified c-Fos and c-Jun heterodimer. (Top panel) DNase I footprinting of the $E_1\alpha$ promoter with purified c-Fos and c-Jun heterodimer. The *Xba*I/*Hin*dIII fragment described in Figure 1, end-labeled on the coding strand, was used as probe. This DNA probe was incubated either with purified c-Fos and c-Jun heterodimer (Abate et al., 1990) or with 50 μg of rat liver nuclear extract (LNE), as indicated, and then subjected to DNase I digestion. The position of the c-Fos- and c-Jun-binding sites relative to the P1 site is indicated. (Bottom panel) The position of the c-Fos- and c-Jun-binding sites relative to the consensus sequence of the CRE is indicated.

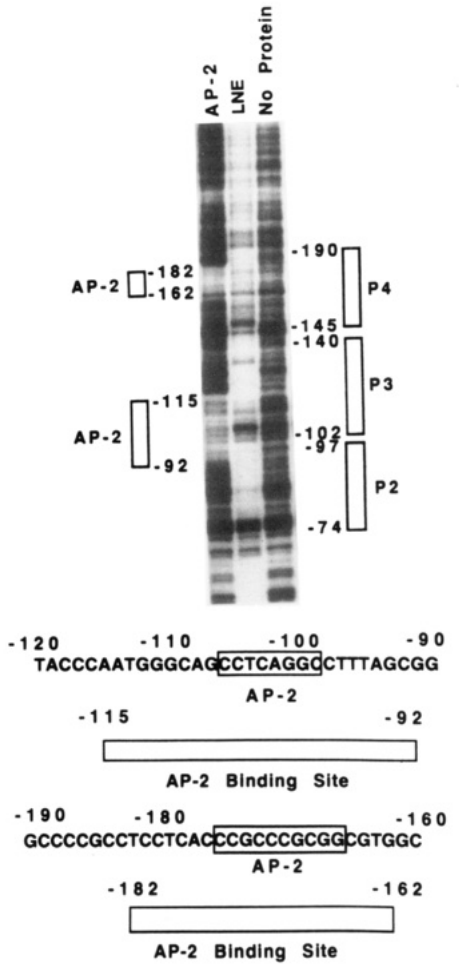

FIGURE 4: DNase I footprinting of the $E_1\alpha$ promoter with purified AP-2 protein. (Top panel) DNase I footprinting of the $E_1\alpha$ promoter with purified AP-2 protein. The *Xba*I/*Hin*dIII fragment described in Figure 1, end-labeled on the coding strand, was used as probe. This DNA probe was incubated with either purified AP-2 protein (1 footprinting unit) or 50 μg of rat liver nuclear extract (LNE), as indicated, and then subjected to DNase I digestion. The positions of the AP-2-binding sites relative to elements P2–P4 are indicated. (Bottom panel) The positions of the AP-2-binding sites relative to the consensus sequence of the AP-2 are indicated.

"CCCCAGGC" (Mitchell et al., 1987), partly overlapped P2 and P3 elements. Interestingly, although the $E_1\alpha$ promoter contains three such similar sequences (Figure 1), purified AP-2 protein interacted only with the distal site (–172 to –165) within the P4 element.

(*C*) *CCAAT Site.* The P3 element, which is localized at positions –103 to –140 for the coding strand and at –109 to –149 for the noncoding strand, contains a CCAAT element, which could be a recognition sequence for CCAAT-binding transcription factors, also called nuclear factor 1 (CTF/NF-1) (Santoro et al., 1988). To determine if the nuclear protein(s) from the crude nuclear extract which bound to the $E_1\alpha$ promoter possessed the DNA-binding specificity of CTF/NF-1 factor(s), an oligonucleotide containing the consensus recognition sequence for CTF/NF-1 (5′-CCTTTGGCATGCTGCCAATATG-3′) was used in the footprint competition studies. Competition for binding at P3 and a portion of P1 (Figure 5) was observed when a 64-fold or greater molar excess of CTF/NF-1 oligomer was used. An oligonucleotide not containing a CTF/NF-1 consensus sequence served as a negative control, and it did not alter the binding of nuclear proteins to any of the binding domains, even when a large molar excess (200-fold) of the oligomer was used (data not shown).

(*D*) *C/EBP Site.* C/EBP protein, which was overexpressed in *E. coli* cells, bound to a portion of P6 (Figure 6). The sequence within this region (–527, TTGGGAAAC, to –519) exhibits similarity to the C/EBP consensus binding sequence TGTGGAAAG (Johnson et al., 1987).

(*E*) *Sp1 Sites.* There are two Sp1 sites in the promoter-regulatory region of the $E_1\alpha$ gene (Figure 1). One is within the P2 element, and the other is within the P4 element. When purified Sp1 protein was used in the footprinting assays, both sites were protected from nuclease digestion, but the binding on the proximal Sp1 site was weak (Figure 7). This may be due to the fact that the P4 domain contains not one but two such Sp1 recognition sequences (antisense: –170 to –175 and –188 to –183) (Figure 7). To determine if the liver nuclear proteins bound are closely related or identical to the Sp1 factor, an oligonucleotide containing the Sp1 consensus sequence (5′-ATTCGATCGGGGCGGGGCGAGC-3′) was used in the footprint competition studies. The competition of binding was only observed at P4 and a portion of P1 (data not shown). In that portion of P1, liver nuclear protein binding was also competed away by CTF/NF-1 oligonucleotide (Figure 5).

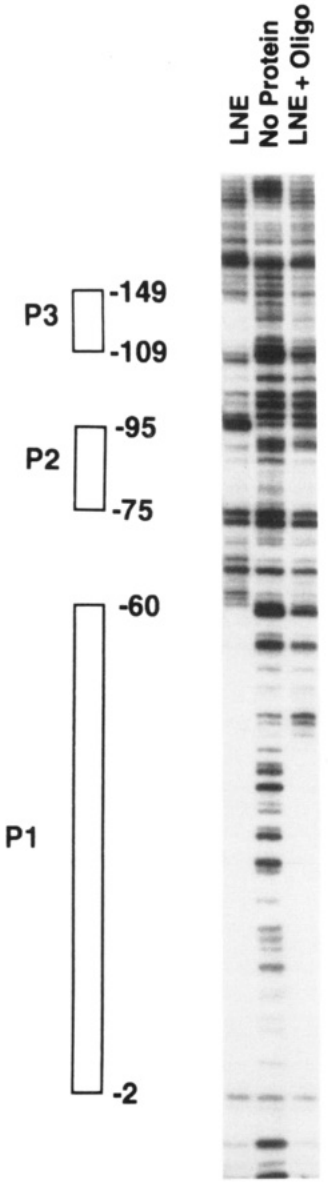

FIGURE 5: CTF/NF-1 oligonucleotide competition footprinting of the $E_1\alpha$ promoter. A 22 bp oligonucleotide (Oligo) which corresponds to the CTF/NF-1 recognition sequence was used in this analysis (see Results). The *Xba*I/*Hin*dIII fragment described in Figure 1, end-labeled on the noncoding strand, was used as probe. This DNA probe was incubated with 50 μg of rat liver nuclear extract (LNE) along with a molar excess of oligonucleotide competitor added to the binding reaction mixture. The competition occurred when a 64-fold molar excess of CTF/NF-1 oligomer was used (data not shown).

Functional Analysis of the $E_1\alpha$–CAT Fusion Constructs in HepG2 Cells. To assess the function of cis-acting elements of the $E_1\alpha$ promoter, a series of nested deletion mutants of the 5′-flanking region of the $E_1\alpha$ gene ligated to the reporter CAT gene were transfected transiently into HepG2 cells, and CAT activity was determined (Figure 8). It was observed that pαCAT1 containing the original 796 bp fragment (–763/+33) resulted in a very high CAT expression comparable to that derived from pSV2CAT (positive control). When the sequence between –763 and –333 was successively deleted from the 5′ end, there was a gradual decrease in CAT activity (pαCAT1 vs pαCAT2, 18% reduction, not significant; pαCAT2 vs pαCAT3, 38% reduction, $p = 0.05$). This suggests that a weak enhancer function may exist upstream of –333 bp. Noticeably, the –763 to –506 bp region deleted in pαCAT2 contained the P7 and P6 domains and a putative "enhancer" reported previously (Tomura et al., 1990). The –506 to –333 bp region further deleted in pαCAT3 contained the P5 domain. When the deletion was extended from –333 to –221 bp, there was a 53% increase in expression of pαCAT4, suggestive of the presence of a negative regulatory element within the deleted sequence. No protection of DNA sequences in this region from DNase footprint analysis was observed. Further deletion to –102 bp (pαCAT5) resulted in a decrease in CAT activity (pαCAT4 vs pαCAT5, 47% reduction, $p = 0.05$), suggesting a possible enhancer function in the P3 and P4 domains which are contained within –221 to –102 bp of the $E_1\alpha$ promoter region. The resulting plasmid pαCAT5, which included only 102 bp immediately 5′ of the transcription start site and 33 bp of exon 1 of the human $E_1\alpha$ gene, was still capable of efficiently promoting CAT expression. This implies that minimal sequences and factors required for constitutive transcription of the $E_1\alpha$ gene in HepG2 cells exist in this short region. This region contains the TATA box, CRE, Sp1, and AP-2 sites and is almost completely protected by liver nuclear proteins as observed from DNase I footprint analysis.

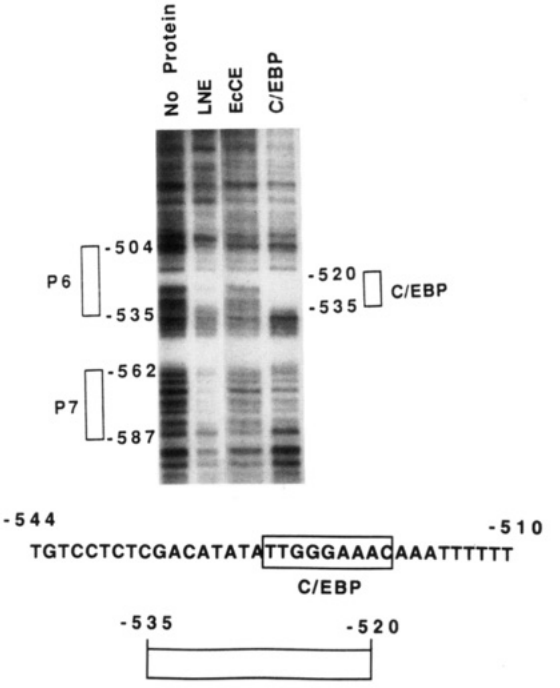

FIGURE 6: DNase I footprinting of the $E_1\alpha$ promoter with expressed C/EBP protein. The *Cla*I/*Bam*HI fragment described in Figure 1, end-labeled on the coding strand, was used as probe. The DNA probe was incubated with 4 μL of *E. coli* cell extract containing overexpressed C/EBP protein (C/EBP), 4 μL of *E. coli* cell extract (EcCE), or 50 μg of rat liver nuclear extract (LNE), as indicated, and then subjected to DNase I digestion. The position of the C/EBP-binding site relative to element P6 is indicated.

DISCUSSION

To understand how specific protein–DNA interactions regulate human $E_1\alpha$ gene expression and eventually to identify cis- and trans-acting elements, the 5′-flanking region of the $E_1\alpha$ gene was mapped by DNase I footprinting analysis and a transient expression assay. The complexity of this region is evident from the results of these mapping experiments, which identified seven distinct nuclear protein-binding domains as well as several cis-acting elements having the potential to enhance CAT expression in HepG2 cells.

Within the 763 bp 5′-flanking sequence of the $E_1\alpha$ gene, the first 221 bp 5′ to the putative transcription start site and 33 bp of the first exon (–221/+33) appear most critical in expression of the gene. This region retained more than 75%

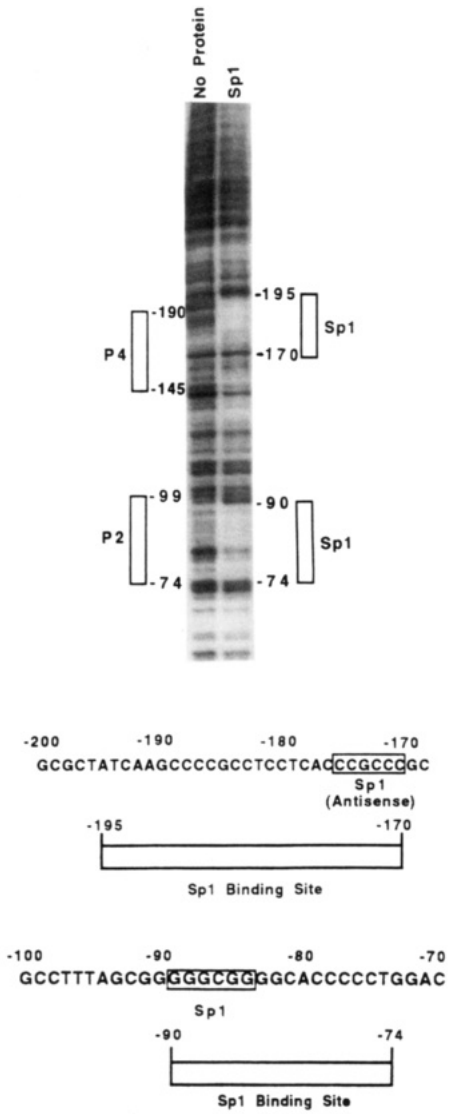

FIGURE 7: DNase I footprinting of the $E_1\alpha$ promoter with purified Sp1 protein. (Top panel) DNase I footprinting of the $E_1\alpha$ promoter with purified Sp1 protein. The *Xba*I/*Hin*dIII fragment described in Figure 1, end-labeled on the coding strand, was used as probe. This DNA probe was incubated with purified Sp1 protein (1 footprinting unit) as indicated and then subjected to DNase I digestion. The positions of the Sp1-binding sites relative to elements P2–P4 are indicated. (Bottom panel) The positions of the Sp1-binding sites relative to the Sp1 motifs are indicated.

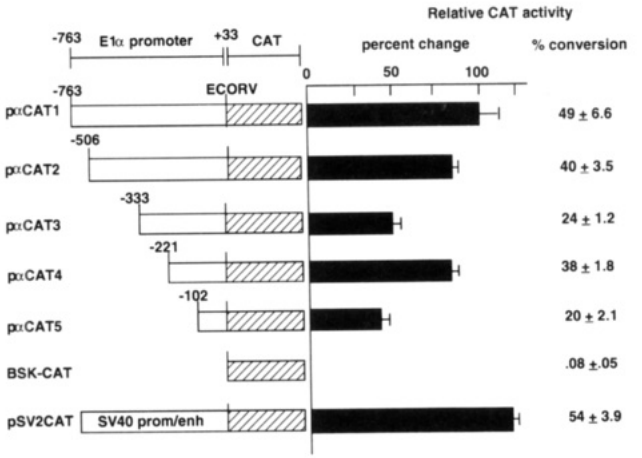

FIGURE 8: Transient expression of CAT activity by the human $E_1\alpha$ promoter in HepG2 cells. The diagram depicts the human $E_1\alpha$ promoter (from bp −763 to +33) ligated to the CAT structural gene (pαCAT1). Deletions were constructed from pαCAT1 using restriction sites at −506 bp (*Acc*I; pαCAT2), −333 bp (*Pst*I; pαCAT3), −221 bp (*Apa*I; pαCAT4), and −102 bp (*Stu*I; pαCAT5). The BSK-CAT construct was devoid of the $E_1\alpha$ promoter sequence. HepG2 cells were cotransfected with 5 μg of the $E_1\alpha$CAT constructs or pSV2CAT (as a positive control) and 3 μg of RSVβgal. The results shown are corrected for transfection efficiency using the β-galactosidase activity and expressed as fold change compared to pαCAT1. Percentage conversion represents the relative CAT activity as a fraction of the total [^{3}H]chloramphenicol present in the assay. The results are the means ± SE of eight independent transfections.

of the promoter activity of the 763 bp upstream sequence in our CAT assay (Figure 8). It was also almost completely protected by liver nuclear proteins in DNase I footprint assays. Four of the seven nuclear protein-binding domains contained in the 763 bp 5′-flanking region are clustered within this region (Figure 2). Compared to the remainder of the 5′-flanking sequence, it also contains an unusually high number of known consensus sequences. Three out of four well-characterized and purified transactivators, namely, AP-2, c-Fos/c-Jun, and Sp1, show the ability to bind autonomously to the corresponding motifs in this 221 bp region. This region may be further divided into two segments based on the results of footprint analysis and CAT assay. The −102/+33 region contains the P1 and P2 domains, which share sequence similarity to several cis elements. More than one protein bound at P1, indicating complex DNA–protein interactions. The minimal sequence and factors required for constitutive transcription of the $E_1\alpha$ gene in HepG2 cells very likely reside in this short region, although the precise boundaries of the sequence and the nature of the factor(s) have to be further defined. TATA box, Sp1, and CRE elements have all been previously implicated in the constitutive transcription of several genes (Imagawa et al., 1987; Breathnach & Chambon, 1981; Dynan & Tjian, 1983). They will be the candidates for our future investigation of this promoter region.

The second segment, the −221/−102 region, contains P3 and P4 domains. The deletion of this sequence resulted in approximately 50% reduction in the level of CAT activity driven by the −221/+33 region (Figure 8), implying an enhancer function in this region. Although we have yet to determine the precise location of this cis-element, one explanation for this decrease in the constitutive transcription might be the deletion of a "CAAT" homology present at positions −113 to −118 and/or the deletion of an "Sp1"-like motif at positions −170 to −175. Both elements have been found in a variety of eukaryotic as well as viral promoter-regulatory regions to play a critical role in directing efficient basal transcription (Dynan & Tjian, 1983; Myers et al., 1986). In the −221/−102 region, the "CAAT" box is a part of P3, and the Sp1 motif is a part of P4. As is evident from the results of footprint competition studies, the liver nuclear proteins bound at P3 and P4 indeed seem to be closely related to the Sp1 protein and to CTF/NF-1, which is the transactivator bound at the "CAAT" box (Figure 5). Moreover, well-characterized and purified human Sp1 protein bound at the Sp1 motif (Figure 7). Finally, there also appears to be a relationship between these two motifs and part of P1 as described under Results. Since both an oligonucleotide containing the CTF/NF-1 recognition sequence and an oligonucleotide containing the Sp1 recognition sequence competed for binding at the same region of P1, it is not likely that either CTF/NF-1 or Sp1 factor alone bound at this position of P1. Rather, nuclear protein binding at these sites may be interdependent and cooperative. These interactions of cis-elements and nuclear protein may play an important role in the efficient transcription of the $E_1\alpha$ gene.

The successive deletion of −763/−507 and −506/−333 regions resulted in a gradual decrease in CAT activity expressed by the $E_1\alpha$–CAT fusion gene. Although these deletions include several potential cis-acting regulatory elements, the most notable feature is the presence of three protein-binding domains in these regions. P7 and P6 are located in the −763/−507 region whereas P5 is in the −506/−333 region. The precise function of these binding domains, however, remains to be established. Interestingly, no protein binding was detected over the TGGTGTTTTGTCTCT region (−646 to −632), a putative enhancer element which has been found in three mitochondrial energy-transducing genes including the human $E_1\alpha$ gene (Tomura et al., 1990). This element (15 nucleotides) shares a complete identity (region −646 to −637) with a recently identified insulin response element (IRE) present in the promoter/regulatory region of several genes including phosphoenolpyruvate carboxykinase (O'Brien et al., 1990). Whether the absence of a footprint in these regions is due to the labile nature of the regulatory protein remains to be investigated.

It seems that in some regard the $E_1\alpha$ promoter may represent a mixture of characteristics of both facultative and housekeeping gene promoters. The genes of the facultative group are usually induced in a tissue- or temporal-specific fashion, and contain a TATA box and an upstream control region (CAAT box) located, respectively, 20–30 and 40–110 residues upstream from the transcriptional start site. The absence of TATA and CAAT and the presence of multiple GC boxes have been reported as characteristics of promoters of housekeeping genes. They usually direct constitutive expression of enzymes which are widely distributed and perform essential metabolic functions, and are less stringently regulated and possess little tissue specificity (Dynan, 1986). The functional and structural differences in these two groups of promoters have prompted speculation that there may be fundamental differences in the way the cellular transcription machinery recognizes the two kinds of promoters (Smale & Baltimore, 1989).

Universal expression of PDC suggests a housekeeping role for its gene products. The 5′-flanking region of $E_1\alpha$, however, bears structural similarities to both facultative and housekeeping gene promoters. It closely resembles facultative promoters in having appropriately positioned "TATA" and "CAAT" boxes with some upstream enhancers. This region, however, is also housekeeping-like due to its high G+C content and two Sp1-binding sites as described above. The present study provides evidence of mixed features for the human $E_1\alpha$ promoter. For example, consistent with the housekeeping gene concept (Dynan, 1986), the relative amount of CAT expression was altered only modestly by progressively deleting the 5′-flanking sequence. However, the observed nuclear protein interaction with "TATA" and "CAAT" boxes seems to suggest that this promoter might be recognized by the cellular transcription machinery which usually recognizes the facultative gene promoter during expression. Therefore, the human $E_1\alpha$ promoter could be a very interesting system in elucidation of differences of the transcriptional regulation mechanism of these two groups of promoters.

In conclusion, the 5′-flanking sequence (−763 to +33) of the human $E_1\alpha$ gene is an active promoter whose minimal sequence and factor(s) required for constitutive transcription reside in the −102 to +33 region. In addition, it contains multiple binding domains for nuclear proteins and several functional cis-elements that have the potential to regulate transcription.

ACKNOWLEDGMENT

We are grateful to Drs. Tom Curran and Steven McKnight for the generous gifts of purified c-Fos and c-Jun proteins, and overexpressed C/EBP protein. We thank Drs. Austin L. Gurney and William J. Roesler for helpful advice in footprinting analysis, Ms. Beth Kleinhenz and Ms. Feng Wang for technical assistance, and Drs. David Samols and John Nilson for critical reading of the manuscript.

REFERENCES

Abate, C., Luk, D., Gentz, R., Rauscher, F. J., III, & Curran, T. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 1030–1036.
Bradford, M. M. (1976) *Anal. Biochem. 72*, 248–254.
Breathnach, R., & Chambon, P. (1981) *Annu. Rev. Biochem. 50*, 349–383.
Brown, R. M., Dahl, H.-H. M., & Brown, G. K. (1989) *Genomics 4*, 174–181.
Curran, T., & Franza, B. R., Jr. (1988) *Cell 55*, 395–397.
Dahl, H.-H. M., Brown, R. M., Hutchison, W. M., Maragos, C., & Brown, G. K. (1990) *Genomics 2*, 225–232.
Dynan, W. S. (1986) *Trends Genet. 2*, 196–197.
Dynan, W. S., & Tjian, R. (1983) *Cell 35*, 79–87.
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.
Imagawa, M., Chiu, R., & Karin, M. (1987) *Cell 51*, 251–260.
Johnson, P. F., Landschulz, W. H., Graves, B. J., & McKnight, S. L. (1987) *Genes Dev. 1*, 133–146.
Kennedy, G. C., Anderson, B., & Nilson, J. H. (1990) *J. Biol. Chem. 265*, 6279–6285.
Lee, W., Haslinger, A., Karin, M., & Tjian, R. (1987) *Nature 325*, 368–372.
MacDonald, M. J., Kaysen, J. H., Moran, S. M., & Pomije, C. E. (1991) *J. Biol. Chem. 266*, 22392–22397.
Maragos, C., Hutchison, M. M., Hayasaka, K., Brown, G., & Dahl, H.-H. M. (1989) *J. Biol. Chem. 264*, 12294–12298.
Mitchell, P. J., Wang, C., & Tjian, R. (1987) *Cell 50*, 847–861.
Myers, R. M., Tilly, K., & Maniatis, T. (1986) *Science 232*, 613–618.
O'Brien, R. M., Lucas, P. C., Forest, C. D., Magnuson, M. A., & Granner, D. K. (1990) *Science 249*, 533–537.
Ohta, S., Endo, H., Matsuda, K., & Kagawa, Y. (1989) *Ann. N.Y. Acad. Sci. 573*, 438–440.
Olson, S., Song, B. S., Huh, T.-L., Chi, Y.-T., Veech, R. L., & McBride, W. O. (1990) *Am. J. Hum. Genet. 46*, 340–349.
Patel, M. S., & Roche, T. E. (1990) *FASEB J. 4*, 3224–3233.
Reed, L. J. (1974) *Acc. Chem. Res. 7*, 40–46.
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., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Santoro, C., Mermod, N., Andrews, P. C., & Tjian, R. (1988) *Nature 334*, 218–224.
Seed, B., & Sheen, J. (1988) *Gene 67*, 271–277.
Short, J. M., Wynshaw-Boris, A., Short, H. P., & Hanson, R. W. (1986) *J. Biol. Chem. 261*, 9721–9726.
Smale, S. T., & Baltimore, D. (1989) *Cell 57*, 103–113.
Tomura, H., Endo, H., Kagawa, Y., & Ohta, S. (1990) *J. Biol. Chem. 265*, 6525–6527.
Wigler, M., Sweet, R., Sim, G. K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S., & Axel, R. (1979) *Cell 16*, 777–785.