

# The interplay of ubiquitous DNA-binding factors, availability of binding sites in the chromatin, and DNA methylation in the differential regulation of phosphoenolpyruvate carboxykinase gene expression

Steven Faber, Tony Ip<sup>+</sup>, Daryl Granner and Roger Chalkley\*

Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37232, USA

Received May 17, 1991; Revised and Accepted August 2, 1991

## ABSTRACT

We have identified DNA elements in the phosphoenolpyruvate carboxykinase (PEPCK) gene promoter which are bound *'in vivo'* by proteins under conditions of basal level gene expression and have evaluated several hypotheses to account for the tissue specific expression of the gene. *In vitro* DNase I footprinting demonstrated that factors which bind to basal expression elements of the PEPCK promoter, the BSE/CRE and NFI/CCAAT sites, are also present in HTC and XC cells which do not express the PEPCK gene. *'In vivo'* DNase I footprinting demonstrated that the BSE/CRE, NFI/CCAAT, and three additional sites are bound by protein in H4IIE cells which express the PEPCK gene but not in the HTC or XC cells. No evidence for a repressor protein or for phased nucleosome binding to the PEPCK promoter in HTC or XC cells could be detected. Genomic sequencing was used to determine if differential methylation of the PEPCK promoter could account for the lack of factor binding in HTC and XC nuclei. None of the 14 cytosine residues in CpG dinucleotides was methylated in H4IIE or rat liver DNA, all were methylated in rat sperm DNA, and 6 were methylated in HTC DNA; including the cytosine at position -90 within the BSE/CRE. Only one cytosine residue, at position -90, was methylated in XC DNA. Treatment of XC cells with 5-azacytidine resulted in loss of methylation at the -90 position yet this was insufficient to allow synthesis of a detectable amount of PEPCK mRNA

## INTRODUCTION

In any given cell type a subset of the genes are expressed, or are capable of being expressed, while the remainder are quiescent. It is clear that transcription of eukaryotic class II genes requires the interaction of multiple specific trans-acting factors with the DNA regulatory elements of the gene. It is thought that

developmental and tissue specific patterns of gene expression arise, in part, by variations in the presence, amount, and activity of specific trans-acting factors (1–3). While some trans-acting factors have a restricted tissue distribution, others appear to be ubiquitous. Fundamental to understanding the tissue specific regulation of gene expression is to determine whether the binding of trans-acting factors to their DNA elements of quiescent genes is regulated, and if so, how this is achieved. That these protein-DNA interactions are indeed regulated is suggested by the observation that stably transfected cells can frequently express a gene even while the endogenous gene is inert (4,5). It has been shown however in only a few instances that ubiquitous trans-acting factors interact with their binding sites in a cell type-specific manner (6,7). This differential binding of ubiquitous factors may lie at the heart of selective gene expression in differentiated tissues.

A largely overlooked aspect of the regulation of gene expression is the accessibility of factor binding sites in DNA packaged into chromatin (8). 145 bp of DNA is wrapped around two each of the core histones H2A, H2B, H3, and H4 to form a nucleosome. The nucleosomal array is compacted further by coiling into a solenoid-like structure. Transcribed genes possess an altered chromatin structure as indicated by the increased sensitivity of their DNA to cleavage by nucleases such as micrococcal nuclease or DNase I compared to bulk DNA (9,10). Transcribed genes also contain short regions of DNA, typically, but not exclusively, at their 5' ends which are hypersensitive to DNase I (10–12). These regions, which are usually from 50 to 500 bp in length, coincide with DNA elements required for gene expression. It is thought that the interaction of multiple transcription factors with the DNA and the concomitant disruption of chromatin structure results in the formation of hypersensitive sites. The pattern of these protein-DNA interactions on chromatin is established during ontogeny (11) and is faithfully maintained during replication of a differentiated cell type. It is unclear what role chromatin structure plays in influencing these specific protein-DNA interactions.

\* To whom correspondence should be addressed

<sup>+</sup> Present address: Department of Biology, University of California at San Diego, La Jolla, CA 92093, USA

The binding of a repressor(s), a positioned nucleosome(s), or covalent modification of the DNA are among the possibilities for regulating the interaction of trans-acting factors with chromatin. Cytosine residues in mammalian DNA can be modified by methylation at the 5 position of the pyrimidine ring usually when in the sequence 5' CpG 3' (13). It has been shown that a gene tends to be hypomethylated in tissues in which the gene is active (14). Conversely in tissues in which the gene is inactive, it is usually more extensively methylated. The correlation between gene activity and DNA methylation is most clearly seen in the 5' region of the gene in many cases (15). This has led to the suggestion that methylation might be a means of controlling gene expression, either by stabilizing chromatin structure (13,16) or by modulating the binding of trans acting factors (17–19). However most of the information on which this conclusion is based comes from the use of a limited set of restriction enzymes which differ in their endonuclease activity when their recognition site is methylated (14). Thus for most genes only a small subset of all possible methylation sites have been examined. Only since the development of genomic sequencing techniques is it possible to definitively determine if cytosine methylation plays a widespread or selective role in the tissue specific regulation of gene expression.

We are using the gene encoding phosphoenolpyruvate carboxykinase (PEPCK) as a model to investigate the tissue-specific regulation of gene expression as well as to identify the protein-DNA interactions required *in vivo* for PEPCK gene expression. The PEPCK gene is expressed primarily in liver, adipose tissue, and the kidney cortex (20). Transcription of the PEPCK gene in liver begins at birth (21,22). Activation of the gene is associated with sequential demethylation of the gene (23) and the development of DNase I hypersensitive sites (24). Transcription of the PEPCK gene in adult liver is increased from basal levels by glucocorticoids and glucagon via cAMP and inhibited dominantly by insulin (25–27). DNA elements required for basal level transcription (28) and for response to these hormones (26,28–30) have been identified. The interaction of numerous factors in liver nuclear extracts with the PEPCK promoter *in vitro* have been observed (31–33). We have previously shown that transcriptional activity of the PEPCK gene is reflected in disruption of the chromatin structure of the promoter/regulatory region (31). In the current investigation we have used *in vitro* DNase I footprinting to identify ubiquitous factors which interact with basal expression elements of the PEPCK promoter. '*In vivo*' DNase I footprinting was used to assess factor, or other protein, binding to the PEPCK promoter in nuclei of expressing and nonexpressing cell types. Genomic sequencing was used to measure methylation at the 14 CpG sites in the promoter to rigorously evaluate the role of cytosine methylation in regulating PEPCK gene expression.

## MATERIALS and METHODS

### Materials

Restriction endonucleases, the Klenow fragment of E. coli DNA polymerase I, and RQ I DNase were from Promega Biotech. DNase I was from Worthington Diagnostics. Taq DNA polymerase was from United States Biochemicals. 5-azacytidine was from Boehringer Mannheim Biochemicals. Dimethyl sulfate, hydrazine, and piperidine were from Aldrich. Zetaprobe membrane was from Bio-Rad Laboratories. Chemicals were from Fisher Scientific Co. or Sigma. Tissue culture media was from Gibco.

### Cell culture

All cell lines were grown as monolayers in a 1:1 mixture of Dulbecco's modified Eagle medium and Ham's F-12 medium containing 10% supplemented calf serum (Hyclone), 10 U/ml penicillin G and 10  $\mu$ g/ml streptomycin sulfate. Cells were harvested for all procedures when approximately 90% confluent.

### '*In vivo*' DNase I footprinting procedure

Preparation of nuclei and treatment with the concentrations of DNase I indicated in the figure legends were performed as described (31). The DNA was purified by standard methods (34). All DNAs were digested to completion with Hind III before analysis.

### Genomic sequencing procedure

The genomic DNAs were digested with Hind III and then subjected to limited chemical cleavage at cytosine residues according to (35,36). The reaction was stopped and the modified DNA cleaved with piperidine according to (37).

### Sequencing gel markers

The plasmid pC103.B which contains a 7.2 kilo base-pair Bam HI fragment spanning –551 to +6650 of the rat PEPCK gene (38) was digested with Hind III and subjected to the cytosine specific chemical cleavage DNA sequencing procedures as described (36). Genomic DNA purified from tissue culture cells was subjected to limited DNase I digestion. 40  $\mu$ g of DNA was incubated in 50  $\mu$ l of digestion buffer lacking sucrose with 0.015 or 0.06 U/ml of DNase I for 5 minutes at 25°C. The reaction was stopped by adding EDTA to 15 mM and the DNA was purified by standard procedures (34).

### Analysis of DNase I or chemically cleaved genomic DNA

The genomic DNA cleavage pattern was visualized using linear PCR amplification as previously described (37) with slight modifications. The primers were synthesized using sets of oligonucleotides as indicated in Figure 1B. The boundaries of the oligonucleotide sets used in this investigation are (5' to 3') +68 to +100 and +94 to +86 for set #1, –129 to –97 and –102 to –110 for set #2, –297 to –265 and –270 to –278 for set #3, and –187 to –219 and –213 to –205 for set #4. 0.056  $\mu$ g of 9-mer and 0.2  $\mu$ g of 33-mer were used to prepare primer for 5 samples of genomic DNA. The 27-mer was cut from the preparative sequencing gel, eluted from the crushed gel slice with 0.3 M sodium acetate plus 5mM EDTA by diffusion, and precipitated with ethanol after the addition of Escherichia coli DNA to 30  $\mu$ g/ml. The primer was used for the linear amplification of the genomic DNA by polymerase chain reaction as described (37) with the reaction products purified using the cetyltrimethyl ammonium bromide precipitation procedure. 30  $\mu$ g of genomic DNA was used for each reaction with an annealing temperature of 65°C. Reactions for marker lanes on the gels contained 1 ng of chemically cleaved pC103.B DNA with 30  $\mu$ g of Hind III digested E. coli DNA or 30  $\mu$ g of purified genomic DNA treated with DNase I. The precipitated reaction products were electrophoresed at 95 W on 0.8 mm thick 6% polyacrylamide (acrylamide:bis-acrylamide, 30:1), 7 M urea gels set up and run in 50 mM Tris, 50 mM boric acid, and 1 mM EDTA. The gels were fixed for 30 minutes in 10% glacial acetic acid plus 10% methanol, dried, and exposed to X-ray film at –70°C with an intensifying screen.

### *In vitro* DNase I footprinting

The 664 bp Hind III to Bgl II DNA fragment which spans the PEPCK promoter was labeled on the 3' end at the Bgl II site with [ $\alpha$ - $^{32}$ P] dCTP and the Klenow fragment using standard procedures (34). Nuclei were prepared from the tissue culture cells and proteins extracted with 0.5 M NaCl as described (39). The DNase I footprinting analysis was performed as described (39) using 2  $\mu$ l of each extract.

## RESULTS

The PEPCK gene sequences required for liver-specific and hormonally-regulated expression are contained in the 560 bp upstream of the transcription start site (27,41,42). A map of this region of the PEPCK gene is presented in Figure 1. We document previously identified functional zones (26,28,43) and regions which show *in vitro* footprints (31-33) which are relevant to this investigation. The DNA sequences from -134 to +73, where +1 is the transcription start site of the PEPCK gene, are sufficient for basal level expression in transient expression assays (26). Of particular importance is the region from -125 to -60. This region contains a nuclear factor I (NFI)/CCAAT site required for basal level gene expression and a cAMP response element (CRE) centered at -110 and -84, respectively. Functional studies have documented 2 additional basal stimulatory elements (BSEs) (one is coincident with the CRE) and a basal inhibitory element (BIE) in this region. Figure 1 also indicates putative methylation sites and the positions of oligonucleotides used for *in vivo* footprinting and genomic sequencing.

We have utilized several different cell lines which differ in their ability to express the PEPCK gene. H4IIE cells are derived from a rat hepatoma, they express the PEPCK gene in a hormonally regulated manner (25,44). HTC cells are also derived from a rat hepatoma, however these cells have lost the ability to express the PEPCK gene (31). HTC cells have retained the ability to express certain hormonally responsive liver-specific

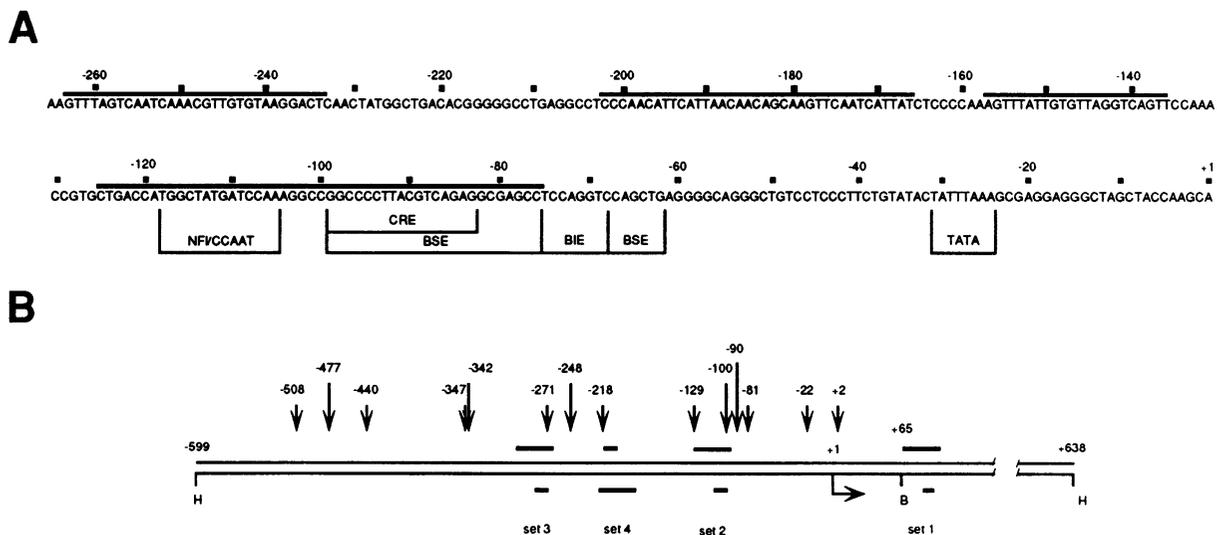
genes such as that encoding tyrosine aminotransferase (TAT) (45). XC cells are derived from rat fibroblasts and they do not express liver specific genes.

### Proteins binding to the BSE/CRE and NFI/CCAAT sites in the PEPCK promoter are found in nonexpressing cell types

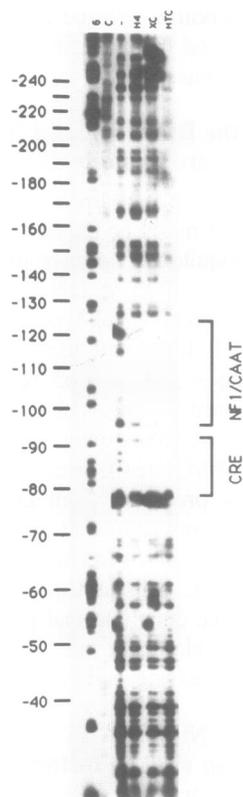
Since factors which interact with the basal expression elements of the PEPCK promoter might be expected to be widespread in distribution, if not ubiquitous, we have used *in vitro* DNase I footprinting to determine if these DNA binding activities are indeed present in non-expressing cell types. Nuclear extracts were prepared from the cell lines and used with a cloned DNA fragment spanning the -599 to +65 region in the DNase I footprinting assay. Figure 2 (lane 4) shows the results using the H4IIE extract. Two adjacent segments of DNA from positions -126 to -95, the NFI/CCAAT site, and -92 to -79, the BSE/CRE, are clearly protected from DNase I digestion, an observation in agreement with that of Quinn et al. (28). Interestingly, identical results were obtained using the nuclear extracts from XC or HTC cells as shown in Figure 2 (lanes 5 and 6, respectively). We conclude that proteins which bind to these basal regulatory elements are found in both PEPCK expressing and non-expressing cell types.

### The BSE/CRE and NFI/CCAAT sites in the PEPCK promoter are bound *in vivo* by factors in an expressing cell type but not in non-expressing cell types

A previous investigation of the PEPCK gene chromatin structure has demonstrated multiple nuclease hypersensitive sites in the 500 bp upstream of the transcription initiation site in H4IIE cells (31). These sites are entirely absent in HTC and XC cells. Although nuclease hypersensitive sites are indicative of an 'open' chromatin structure, presumably due to the interaction of transcription factors, it is not known if chromatin lacking such sites is necessarily incapable of binding any transcription factors. Since proteins which bind to the BSE/CRE and NFI/CCAAT



**Figure 1.** Diagram of the rat PEPCK gene promoter. Panel A shows the DNA sequence of a portion of the PEPCK promoter. DNase I footprints observed *in vitro* using rat liver nuclear extracts are indicated by the horizontal lines above the DNA sequence. Elements required for basal level expression *in vivo* are indicated by boxes below the DNA sequence. Panel B shows a schematic of the promoter DNA. The horizontal arrow at +1 indicates the transcription start site. The location of cytosine residues in CpG dinucleotides is indicated by vertical arrows. The sets of oligonucleotides used to synthesize the primers for this investigation are indicated by bold horizontal lines. The cleavage sites of only the relevant restriction endonucleases are also indicated: Hind III (H) and Bgl II (B). The DNA sequence is numbered according to Ip et al. (31).



**Figure 2.** *In vitro* DNase I footprinting with nuclear extracts from PEPCK expressing and non-expressing cell types. The 664 bp Hind III/Bgl II fragment was labeled with  $^{32}\text{P}$  on the 3' end at the Bgl II site (at +65) and used for DNase I footprinting as described in Materials and Methods using an equal amount of nuclear extract from the indicated cell types. The three left lanes contain the DNA cleaved at guanine residues, cytosine residues, or by DNase I in the absence of nuclear extract, respectively. The DNA sequence is numbered with respect to the transcription initiation site. The regions of protection from DNase I digestion are indicated by brackets.

sites are present in HTC and XC cells we turned to approaches other than DNase I hypersensitive site determination to assay for the presence or absence of these factors on the PEPCK gene promoter in non-expressing cells. We used '*in vivo*' DNase I footprinting to determine if proteins bind to this region of the promoter in isolated nuclei. The nuclei were prepared from H4IIE, HTC, or XC cells and treated with various amounts of DNase I. The genomic DNA was then purified and analyzed using the methodology of Saluz and Jost (37). The DNase I cleavage ladder of the genomic DNA was visualized through a linear amplification using the polymerase chain reaction technique. A radiolabeled oligonucleotide was annealed to the 3' side of the region of interest and extended to positions of DNase I cleavage on the template strand. The reaction products were resolved on a sequencing gel and detected by autoradiography.

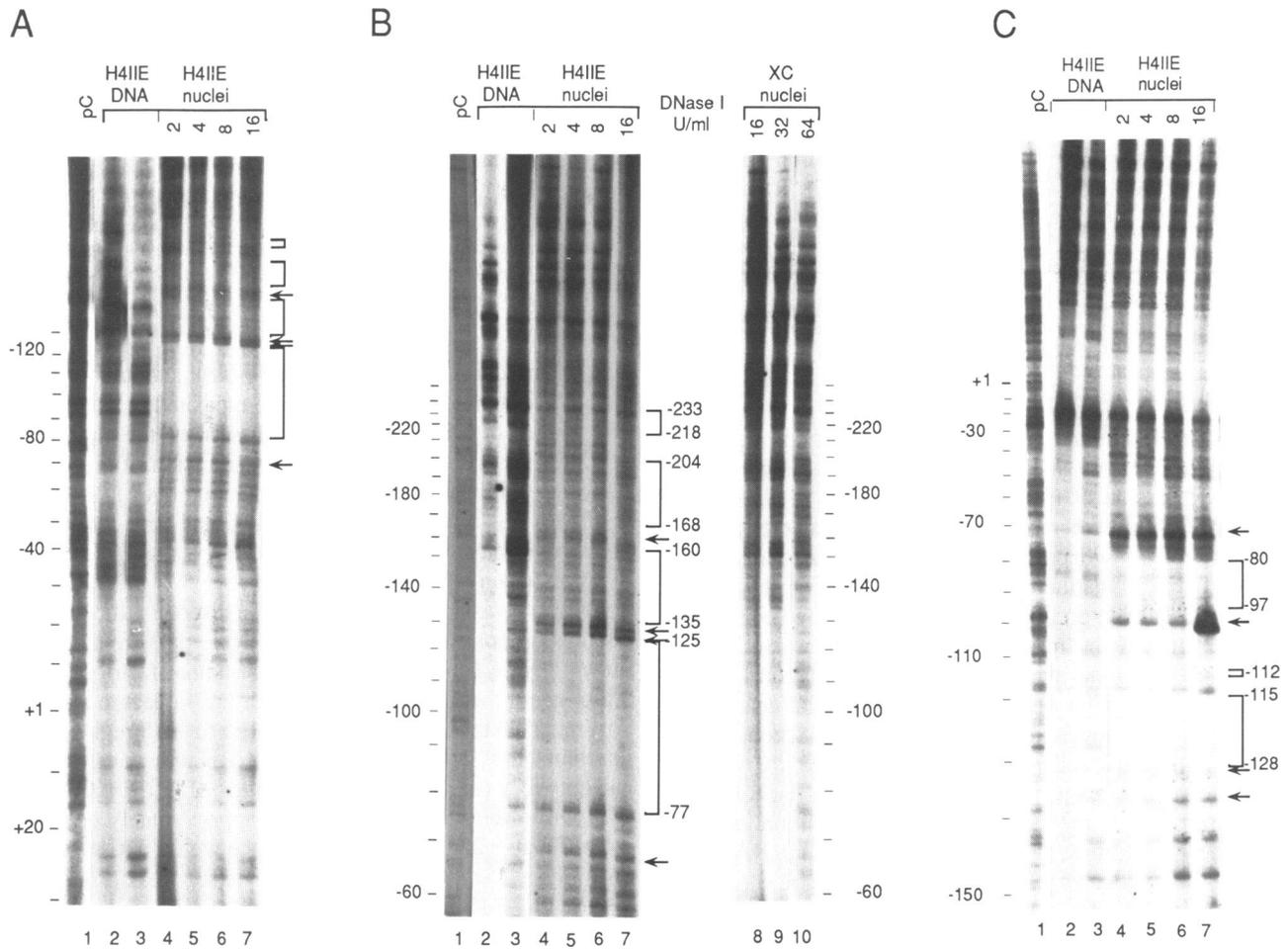
The results of such an analysis of the basal regulatory region of the PEPCK promoter are shown in Figure 3. The effect of increasing amounts of DNase I digestion of either purified H4IIE DNA or isolated nuclei are shown. It is apparent that in the H4IIE nuclei the DNA from  $-125$  to  $-75$  bp is protected from DNase I digestion (Fig. 3A and 3B). In addition, there is enhanced cutting at the upstream boundary of the footprint (see arrows). This region contains the NF1/CCAAT and BSE/CRE elements which contribute to basal expression of this promoter. The DNase

I footprint on the other strand is shown in Figure 3C. Clearly the proteins binding to the DNA in this region do not protect this strand as effectively as seen for the other strand. However regions of enhanced cutting at the boundaries of the two elements are well defined. We have also consistently observed weak protection from DNase I digestion over the regions from  $-230$  to  $-215$ ,  $-180$  to  $-165$ , and  $-160$  to  $-135$  (Fig. 3A and 3B). Analysis of the nuclease sensitivity revealed that the regions from  $-65$  to  $+1$  (Fig. 3A and 3C) and from  $-600$  to  $-235$  (data not shown) are unusually exposed in H4IIE nuclei and resemble naked DNA. On neither strand in the TATA box region do we detect a major footprint indicative of a permanent TFIID (or transcription complex) interaction although a pattern of DNase I cleavage which differs from naked DNA suggests that the DNA structure is altered in this region. Analysis of the  $-250$  to  $-60$  region of the PEPCK promoter in XC cells (in which the PEPCK gene is quiescent) is shown in Figure 3B (lanes 8–10). Identical results were obtained from the HTC cells (data not shown). The PEPCK promoter in these cell lines is resistant to DNase I digestion as compared to the H4IIE cells. At low amounts of DNase I the majority of the DNA is left in large fragments near the top of the sequencing gel. At higher amounts of enzyme the pattern of DNase I cleavage is identical to that of naked DNA (Fig. 3B, lane 10). Thus up to 32 fold more DNase I must be used to treat the HTC and XC cell lines to get digestion equivalent to that observed in H4IIE nuclei. The simplest interpretation for these observations is that in the non-expressing cell lines the promoter is randomly associated with nucleosomes. We find no evidence for specific binding of non-histone proteins from  $-600$  to  $+30$  in these cells (Fig. 3B and data not shown).

#### Methyl-cytosine content of the PEPCK promoter in expressing and non-expressing tissues

We wished to determine if there was a general correlation between the degree of methylation of all CpG sites in the promoter with gene expression. In addition, we have asked whether methylation within specific protein binding sites or functionally defined regions of DNA could account for tissue specific expression. The location of the CpGs in the PEPCK promoter is shown in Figure 1B (arrows). We used genomic sequencing techniques to measure the degree of methylation at each site. Purified restriction endonuclease digested DNA is cleaved at unmodified cytosines using the chemical cleavage sequencing method. In contrast, hydrazine reacts poorly with 5-methyl cytosine leaving these positions uncleaved (35). The DNA sequence of interest was amplified and analyzed as described above and in Materials and Methods. The band intensities in each lane were quantitated by densitometry with the degree of methylation reflected in a decreased signal intensity relative to other bands in the same lane.

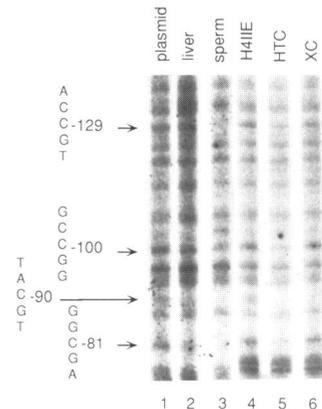
Figure 4 shows a typical result from the  $-130$  to  $-60$  region according to the numbering system of Figure 1. This region contains 4 CpG potential methylation sites at positions  $-129$ ,  $-100$ ,  $-90$ , and  $-81$ . The sequencing ladder produced from DNA of rat liver and H4IIE cells, both of which express the PEPCK gene, is shown in Figure 4 (lanes 2 and 4, respectively). The sequence ladder is identical in these lanes to the plasmid DNA, which is devoid of CpG methylation, indicating that none of the CpG sites is methylated. The sequence ladder from HTC DNA (Fig. 4, lane 5) shows distinct differences from the unmethylated DNAs. The band representing cytosine  $-90$  is absent from the sequence ladder indicating methylation at this



**Figure 3.** *In vivo* DNase I footprinting of the PEPCK promoter. DNase I footprinting and visualization of the genomic DNA cleavage ladder were performed as described in Materials and Methods using the primer synthesized from oligo set #1 for panels A and B or oligo set #4 for panel C. The DNA sequence is numbered with respect to the transcription initiation site of the PEPCK gene on the left of each panel. Regions of protection from DNase I digestion are indicated by brackets and sites of enhanced cleavage by DNase I are indicated by arrows. For all panels- Lane 1: a cloned piece of DNA containing the PEPCK promoter was cleaved at cytosine residues; lanes 2 and 3: purified H4IIE DNA treated with DNase I; lanes 4–7: DNA from H4IIE nuclei treated with 2, 4, 8, or 16 U/ml of DNase I, respectively; lanes 8–10: DNA from XC nuclei treated with 16, 32, or 64 U/ml of DNase I, respectively.

site in most cells. The intensity of the band representing cytosine –81 is reduced by 50% compared to other single bands in the lane indicating partial methylation. The bands representing cytosines at positions –100 and –129 are not resolved from adjacent cytosines. Densitometry indicates that the intensity of these unresolved bands is decreased by 50% relative to neighboring bands consistent with substantial methylation at the –100 and –129 positions. DNA from rat sperm has substantial methylation at all four sites (Fig. 4, lane 3). In contrast the DNA from XC cells is modified only at position –90 where the majority of the DNA at this position is methylated.

This strategy was applied to an analysis of the degree of methylation of the remaining 10 cytosine residues in CpG dinucleotides throughout the PEPCK promoter (data not shown). The results are given in Table 1. Transcriptional inactivity is correlated with methylation only at the –90 position. This site lies within a region from –99 to –76 containing a CRE (see Figure 1) which is capable of binding protein *in vitro* but not *in vivo* in HTC and XC nuclei. Methylation of the corresponding site in the CRE from the human glycoprotein hormone  $\alpha$  subunit gene inhibits factor binding *in vitro* and activity of the CRE in



**Figure 4.** Cytosine methylation of the PEPCK promoter in expressing and non-expressing cell types. The genomic sequencing technique described in Materials and Methods was used to visualize the cytosine-specific cleavage pattern of genomic DNA from the indicated cell types using the primer synthesized from oligo set #1. The location of cytosine residues in CpG dinucleotides is shown on the left of the figure. Lane 1 shows the cytosine specific cleavage pattern of plasmid DNA containing the PEPCK promoter.

TABLE 1. Methyl-cytosine content of the PEPCK promoter<sup>3</sup>

Cell type	Location of cytosine in CpG														
	-506	-477	-440	-347	-342	-271	-245	-215	-129	-100	-90	-51	-22	+2	
liver	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
sperm	+	+	+/-	+	+	+	+	+	+	+	+	+	+	+/-	+/-
H4IIE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HTC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
XC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

<sup>3</sup> Less than 25% methylation (reduction of band intensity determined by densitometry) is indicated by (-), a (+/-) indicates 25% to 75% methylation, and (+) indicates greater than 75% methylation.

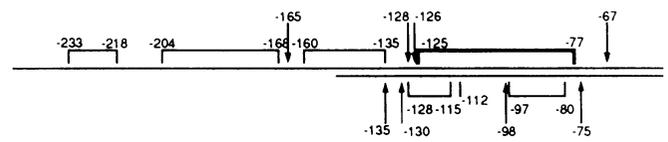
transient expression assays (18). Thus, methylation at the -90 site in such non-expressing cells could play a key role in preventing expression of the PEPCK gene.

As a final test of the role in PEPCK gene expression of methylation at the -90 position, XC cells were treated with 5-azacytidine. Such treatment has been shown to cause the loss of up to 85% of methyl groups on cytosine in cultured cells (46). We reasoned that if the modification plays a key role in the inactivation of this gene, then removal of methylation at -90 might allow transcription of the PEPCK gene. XC cells were treated with various levels of 5-azacytidine for 48 hours or 2 weeks after which the cells were allowed to grow for several passages. Genomic sequencing showed that two of the cell lines have lost essentially all methylation at the -90 position yet they fail to synthesize a detectable amount of PEPCK mRNA (data not shown). Taken together the results presented here show that there is no simple relationship between transcription of the PEPCK gene and methylation of cytosines in the promoter sequences.

## DISCUSSION

We have concentrated in this investigation on the proximal region of the PEPCK promoter which contains the cis-acting elements required for basal level expression of the gene (see Figure 1). We thought it likely that these elements would be bound by factors with a widespread tissue distribution. The results presented in Figure 2 demonstrate that factors capable of binding to the BSE/CRE and NFI/CCAAT sites in the PEPCK promoter are present in HTC and XC cells at concentrations comparable to that in H4IIE cells.

We have used genomic footprinting to identify interactions between the proteins in isolated nuclei and the endogenous PEPCK promoter, under conditions of basal level gene expression, on the coding strand from -600 to +30 and the noncoding strand from +1 to -150. The results obtained from H4IIE cells are summarized in Figure 5. The nearly complete suppression of DNase I cleavage from -125 to -75 indicates that the BSE/CRE (-99 to -76) and NFI/CCAAT (-117 to -103) sites are stably bound by protein in the PEPCK expressing H4IIE nuclei. Both of these elements are required *in vivo* for basal level expression of the PEPCK gene (28). In addition, the site at -99 to -83 contains a cAMP response element (28,47). The identity of the factors binding to these sites is unclear. The NFI/CCAAT site has potential binding sites for NFI/CTF, CBP, or possibly other CCAAT box-binding proteins and is bound *in vitro* by NFI/CTF (32). The BSE/CRE is bound *in vitro* by a



**Figure 5.** Summary of protein-DNA interactions in H4IIE nuclei. The DNA strands of the PEPCK promoter are indicated by the thin horizontal lines. Regions of strong protection from DNase I digestion are indicated by bold brackets and regions of weak protection by thin brackets. Positions of enhanced cleavage by DNase I are indicated by vertical arrows.

CREB protein (28,32) and by C/EBP (48), although it is not clear if both activities of this element are mediated by a single factor. Regardless of what factors bind to these sites, the *in vivo* DNase I footprints provide an important verification of those observed *in vitro*. We have not observed protection over the region from -75 to -62 which contains an additional BSE and a BIE (28) however protein interaction in this region is suggested by a DNase I hypersensitive site located at -67.

We have also observed that several discrete although less well defined regions of DNA from -230 to -130 are also protected from DNase I digestion. This suggests that factors are binding throughout this domain, but that they are subject to a more rapid off rate than those binding to the -125 to -75 region. A CREB protein binds *in vitro* to the BSE/CRE and with much lower affinity to a second site from -155 to -135 (32). C/EBP binds *in vitro* to the BSE/CRE and with approximately equal affinity to a site at -251 to -234 (48). Since we do not see a footprint over the -251 to -234 element but do see weak protection from -160 to -135 this suggests that the BSE/CRE is bound *in vivo* by a CREB protein. The region of weak protection from -180 to -165 corresponds in part to the element from -200 to -164 which is bound *in vitro* by HNF1 (48). It is unclear whether the difference in size of the footprint reflects an altered conformation of HNF1 *in vivo* or binding by a different factor. Protein binding to the region from -230 to -215 has not been detected using rat liver extracts for *in vitro* DNase I footprinting. There is no indication of surviving protein-DNA interactions other than those described above. Indeed, the remainder of the proximal PEPCK promoter is extremely sensitive to DNase I digestion. Only 2 units/ml are required to yield a digestion pattern identical to that of naked DNA. This suggests that, at least in isolated H4IIE nuclei, much of the promoter outside the -230 to -75 region is naked DNA.

The observation that some regions of the PEPCK promoter are tightly bound to proteins while other large regions (including the TATA box domain) are largely devoid of such interaction was intriguing to us. It may be that for these regions under the conditions of basal level gene expression used in this investigation that only a few cis-acting elements are stably bound by factors. Other elements may be bound by factors subject to a more rapid off rate, possibly reflecting a dynamic equilibrium between bound and free factor. Perhaps other elements are bound only transiently before initiation of transcription (and thus are present only in a fraction of the cell population). Alternatively, only very high affinity interactions may survive the isolation of nuclei while lower affinity interactions are lost during the procedure.

*In vivo* DNase I footprinting demonstrated that none of the specific protein-DNA interactions, described above occur in the non-expressing HTC or XC cells (Fig. 3). The lack of factor binding to the -125 to -75 region even though factors which

bind to this region are present in the cells indicates that the protein-DNA interaction is prevented *in vivo*. There is no footprint evidence whatsoever for any protein binding, transacting factor or repressor, other than histones. Although extremely slow in generation, in all cases, the digestion pattern was identical to that of naked DNA. Previous investigation has shown that the PEPCK gene and promoter are packaged into nucleosomes in HTC and XC cells (31). The lack of a 10 bp repeat from the DNase I digestion suggests that the nucleosomes are randomly associated with the PEPCK promoter sequences in these cells. Thus, specific positioning of nucleosomes over factor binding sites is apparently not the mechanism for preventing PEPCK gene expression in these cell lines.

We have established that the PEPCK promoter binds neither repressor proteins nor ubiquitous transcription factors in non-expressing cells. Of fundamental concern to understanding the regulation of PEPCK gene expression is how these protein-DNA interactions occur in a tissue specific manner. DNA methylation offers a possible means to regulate protein-DNA interactions on chromatin. We have used genomic sequencing techniques to identify positions of cytosine methylation in the PEPCK promoter. These results are summarized in Table 1. The results from all of the cell types, except XC, are consistent with the observation that regulatory sequences are in general deficient in methyl-cytosine in tissues where a gene is expressed (14).

The methylation pattern of XC DNA stands in sharp contrast to the results described above. The only site of methylation is at position -90 which is extensively, although not completely, modified. Thus, the inability of factors to bind to the PEPCK promoter is not due to an overall high degree of DNA methylation in these sequences. Since the XC cells are a fibroblast line which has never expressed the PEPCK gene these results do not support a cause and effect relationship between transcription of a gene and overall hypomethylation of its regulatory sequences. The site at -90 however lies within the BSE/CRE. This observation was intriguing since methylation of the corresponding site in the CRE from the human glycoprotein hormone  $\alpha$  subunit gene inhibits factor binding *in vitro* and activity of the CRE in transient expression assays (18). This prompted us to test whether methylation at this site plays a key role in regulating PEPCK gene expression. XC cells were treated with 5-azacytidine which causes loss of DNA methylation (46). Indeed we developed two cell lines which have lost substantial methylation at the -90 position yet do not detectably transcribe the PEPCK gene. We conclude that the binding of factors to the PEPCK promoter and activation of transcription is prevented in XC cells even in the complete absence of methylation of promoter sequences.

The exclusion of all protein interactions other than histones with a given promoter offers a mechanism for the maintenance of total inactivation of a gene in a specific cell type. While this aspect of gene regulation is just beginning to be addressed there may be several methods for achieving this result. It has been shown that positioning of a nucleosome on the mouse mammary tumor virus regulatory elements prevents factor binding and gene expression (49). Becker et al. have shown that ubiquitous factors bind to regulatory elements of the liver specific TAT gene only in cells in which the gene is expressed (6). In contrast to the results presented here cytosine methylation in the TAT regulatory sequences is correlated with the lack of factor binding and inactivity of the gene. Wolf et al. have shown that expression of the chicken lysozyme gene is not correlated with methylated cytosine in the promoter (50). It is not known however if factors

are prevented from binding to the chicken lysozyme gene promoter *in vivo* in non-expressing cell types. We feel that the regulation of factor-DNA interactions is a fundamental mechanism of gene regulation in the maintenance of the differentiated state and which does not necessarily require promoter specific methylation events, positioned nucleosomes, or repressor binding.

This should not be taken to imply that we are concluding that under some circumstances methylation of a specific site in a given promoter may not have dramatic negative effects on factor binding and transcriptional activity. Indeed in some specialized instances this seems likely. Thus treatment of cells with 5-azacytidine has led to gene activation which is correlated with decreased DNA methylation (46). It may be that DNA methylation is one of several means of limiting access of factors to DNA. Alternatively, for some genes the critical region where DNA methylation is important may be in regulatory sequences other than the proximal promoter sequences. The recently identified class of functional elements termed dominant control regions (DCRs) (51) may be such sites. These elements facilitate a tissue specific, chromatin position independent activation of transcription from homologous and heterologous promoters. DCRs generally possess a DNase I hypersensitive site suggesting interaction with non-histone protein. In the PEPCK gene there is an element located at -4800 which has some features of a DCR (39). This element, which is bound by a liver specific factor, has tissue-specific DNase I hypersensitive sites and can activate transcription from the PEPCK promoter. We are currently investigating the role of this putative DCR in the tissue specific regulation of PEPCK gene expression.

## ACKNOWLEDGEMENTS

This work was supported by Program Project Grant DK42502, Public Health Service grants from the National Institutes of Health and by the Vanderbilt Diabetes Research and Training Center. S.F. was supported by an individual National Research Service Award F32 CA 08465.

## REFERENCES

1. Johnson, P.F. and McKnight, S.L. (1989) *Ann. Rev. Biochem.*, **58**, 799-839.
2. Maniatis, T., Goodbourn, S. and Fischer, J.A. (1987) *Science*, **236**, 1237-1245.
3. Mitchell, P.J. and Tjian, R. (1989) *Science*, **245**, 371-378.
4. Busslinger, M., Hurst, J. and Flavell, R.A. (1983) *Cell*, **34**, 197-206.
5. Murray, E. and Grosfeld, F. (1987) *EMBO J.*, **5**, 2329-2335.
6. Becker, P.B., Ruppert, S. and Schutz, G. (1987) *Cell*, **51**, 435-443.
7. Mueller, P.R. and Wold, B. (1989) *Science*, **246**, 780-786.
8. McGhee, J.D. and Felsenfeld, G. (1980) *Annu. Rev. Biochem.*, **49**, 1115-1156.
9. Eissenberg, J.C., Cartwright, I.L., Thomas, G.H. and Elgin, S.C.R. (1985) *Annu. Rev. Genet.*, **19**, 485-536.
10. Elgin, S.C.R. (1988) *J. Biol. Chem.*, **263**, 19259-19262.
11. Gross, D.S. and Garrard, W.T. (1988) *Ann. Rev. Biochem.*, **57**, 159-197.
12. Pederson, D.S., Thoma, F. and Simpson, R.T. (1986) *Annu. Rev. Cell Biol.*, **2**, 117-147.
13. Selker, E.U. (1990) *Trends Biochem. Sci.*, 103-107.
14. Yisraeli, J. and Szyf, M. (1984) In Razin, A., Cedar, H. and Riggs, A.D. (ed.), DNA methylation: Biochemistry and biological significance. Springer-Verlag, New York, NY, pp.353-378.
15. Dorfler, W. (1983) *Ann. Rev. Biochem.*, **52**, 93-124.
16. Cedar, H. (1988) *Cell*, **53**, 3-4.
17. Hermann, R., Hoeveler, A. and Doerfler, W. (1989) *J. Mol. Biol.*, **210**, 411-415.
18. Iguchi-Arigo, S.M.M. and Schaffner, W. (1989) *Genes Dev.*, **3**, 612-619.

19. Watt, F. and Molloy, P.L. (1988) *Genes Dev.*, **2**, 1136–1143.
20. Tilghman, S.M., Hanson, R.W. and Ballard, F.J. (1976) In Mehlman, M.A. (ed.), *Gluconeogenesis: its regulation in mammalian species*. John Wiley & Sons, Inc., New York, pp.47–91.
21. Philippidis, H., Hanson, R.W., Reshef, L., Hopgood, M.F. and Ballard, F.J. (1972) *Biochem. J.*, **126**, 1127–1134.
22. Mencher, D. and Reshef, L. (1979) *Eur. J. Biochem.*, **94**, 582–589.
23. Benvenisty, N., Mencher, D., Meyuhas, O., Razin, A. and Reshef, L. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 267–271.
24. Benvenisty, N. and Reshef, L. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1132–1136.
25. Sasaki, K., Cripe, T.P., Koch, S.R., Andreone, T.L., Petersen, D.D., Beale, E.G. and Granner, D.K. (1984) *J. Biol. Chem.*, **259**, 15242–15251.
26. Short, J.M., Wynshaw-Boris, A., Short, H.P. and Hanson, R.W. (1986) *J. Biol. Chem.*, **261**, 9721–9726.
27. Magnuson, M.A., Quinn, P.G. and Granner, D.K. (1987) *J. Biol. Chem.*, **262**, 14917–14920.
28. Quinn, P.G., Wong, T.W., Magnuson, M.A., Shabb, J.B. and Granner, D.K. (1988) *Mol. Cell. Biol.*, **8**, 3467–3475.
29. Imai, E., Stromstedt, P., Quinn, P.G., Carlstedt-Duke, J., Gustafsson, J. and Granner, D.K. (1990) *Mol. Cell. Biol.*, **10**, 4712–4719.
30. Forest, C.D., O'Brien, R.M., Lucas, P.C., Magnuson, M.A. and Granner, D.K. (1990) *Mol. Endocrinol.*, **4**, 1302–1310.
31. Ip, Y.T., Granner, D.K. and Chalkley, R. (1989) *Mol. Cell. Biol.*, **9**, 1289–1297.
32. Roesler, W.J., Vandembark, G.R. and Hanson, R.W. (1989) *J. Biol. Chem.*, **264**, 9657–9664.
33. Trus, M., Benvenisty, N., Cohen, H. and Reahef, L. (1990) *Mol. Cell. Biol.*, **10**, 2418–2422.
34. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
35. Church, G.M. and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
36. Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.*, **65**, 499–560.
37. Saluz, H. and Jost, J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2602–2606.
38. Beale, E.G., Chrapkiewicz, N.B., Scoble, H.A., Metz, R.J., Quick, D.P., Noble, R.L., Donelson, J.E., Biemann, K. and Granner, D.K. (1985) *J. Biol. Chem.*, **260**, 10748–10760.
39. Ip, Y.T., Poon, D., Stone, D., Granner, D.K. and Chalkley, R. (1990) *Mol. Cell. Biol.*, **10**, 3770–3781.
40. Ozer, J., Faber, M., Chalkley, R. and Sealy, L. (1990) *J. Biol. Chem.*, **265**, 22143–22152.
41. McGrane, M.M., deVente, J., Yun, J., Bloom, J., Park, E., Wynshaw-Boris, A., Wagner, T., Rottman, F.M. and Hanson, R.W. (1988) *J. Biol. Chem.*, **263**, 11443–11451.
42. Wynshaw-Boris, A., Short, J.M., Loose, D.S. and Hanson, R.W. (1986) *J. Biol. Chem.*, **261**, 9714–9720.
43. Klemm, D.J., Roesler, W.J., Liu, J., Park, E.A. and Hanson, R.A. (1990) *Mol. Cell. Biol.*, **10**, 480–485.
44. Sasaki, K. and Granner, D.K. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2954–2958.
45. Plesko, M.M., Hargrove, J.L., Granner, D.K. and Chalkley, R. (1983) *J. Biol. Chem.*, **258**, 13738–13744.
46. Jones, P.A. (1984) In Razin, A., Cedar, H. and Riggs, A.D. (ed.), *DNA methylation: Biochemistry and biological significance*. Springer-Verlag, New York, NY, pp.165–187.
47. Bokar, J.A., Roesler, W.J., Vandembark, G.R., Kaetzel, D.M., Hanson, R.W. and Nilson, J.H. (1988) *J. Biol. Chem.*, **263**, 19740–19747.
48. Park, E.A., Roesler, W.J., Liu, J., Klemm, D.J., Gurney, A.L., Thatcher, J.D., Shuman, J., Friedman, A. and Hanson, R.W. (1990) *Mol. Cell. Biol.*, **10**, 6264–6272.
49. Bresnick, E.H., John, S., Berard, D.S., LeFebvre, P. and Hager, G.L. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3977–3981.
50. Wolf, S., Schrader, M. and Wittig, B. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 271–275.
51. van Assendelft, G.B., Hanscombe, O., Grosveld, F. and Greaves, D.R. (1989) *Cell*, **56**, 969–977.