Order Number 8820873

Transcriptional regulation of hepatitis B virus by sequence-specific DNA-binding proteins

Karpen, Saul Joseph, Ph.D.

City University of New York, 1988

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

TRANSCRIPTIONAL REGULATION OF HEPATITIS B VIRUS BY SEQUENCE-SPECIFIC DNA-BINDING PROTEINS

by

SAUL J. KARPEN

A dissertation submitted to the Graduate Faculty in Biomedical Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1988

This manuscript has been read and accepted for the Graduate Faculty in Biomedical Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

12/9/87

Date

George Acs, M.D., Ph.D.
Chair of Examining Committee

12/9/87

Date

Terry A. Krulwich, Ph.D.
Executive Officer

Supervisory Committee

Carter Bancroft, Ph.D.

James Bieker, Ph.D.

Robert Lazzarini, Ph.D.

Jay Unkeless, Ph.D.

The City University of New York

ABSTRACT

TRANSCRIPTIONAL REGULATION OF HEPATITIS B VIRUS BY SEQUENCE-SPECIFIC DNA-BINDING PROTEINS

by

Saul J. Karpen

Advisor: Dr. George Acs

Two major messenger RNAs [mRNAs] are transcribed from the Hepatitis B Virus [HBV] genome. Although both are involved in viral gene expression, one of these, the 3.5 kilobase [kb] mRNA, also plays a central role in HBV replication. In addition, published reports have repeatedly demonstrated that 3.5 kb mRNA expression is highly cell-specific: limited to cell lines derived from human hepatocytes [the natural host of HBV replication]. In this thesis, studies exploring the factors controlling the cell-specific expression of the 3.5 kb mRNA were undertaken, with a particular emphasis placed on the possible role of sequence-specific DNA-binding proteins.

An EcoRI-AluI restriction fragment [HBV subtype ayw; nucleotides (nt) 1-1878, including the start site of 3.5 kb mRNA transcription and ≈1.8 kb of 5' flanking sequences] was inserted 5' to the bacterial chloramphenicol acetyltransferase [CAT] reporter gene of pSVoCAT, to create the recombinant plasmid pEcoAluCAT. Transfection experiments showed that pEcoAluCAT-driven CAT expression paralleled the human hepatocyte-specific expression of the 3.5 kb mRNA. **In vivo** CAT competition assays [in the human hepatoblastoma-derived cell line HepG2] demonstrated that pEcoAluCAT expression was primarily dependent on two cis-acting DNA elements; within the previously defined HBV

enhancer [ENH; nt 1082-1378], and core promoter [CP; nt 1645-1888] regions. Furthermore, pEcoAluCAT expression was shown to involve the interaction of trans-acting HepG2 cellular factors, with sequences within these two DNA elements. These data were corroborated with **in vitro** binding experiments, using gel mobility shift assays. Appropriate DNA competition experiments demonstrated the presence of distinct HepG2 nuclear proteins, containing high affinity DNA-binding domains specific for sequences within ENH and CP elements. Assays performed with extracts from non-human and non-hepatocyte nuclei revealed shift patterns different from those obtained from HepG2 nuclear extract. Methylation interference assays detailed the protein binding regions within the CP [between nt 1655-1684], and ENH [nt 1089-1170] DNA elements. Both binding sites contained palindromes, and homologies were seen with a limited number of sequenced genes.

ACKNOWLEDGEMENTS

Although a dissertation is, by nature, an individual endeavor, I am happy to say that there were many who significantly helped this work progress, in spite of my inherent thick-headedness [and various other, probably autosomal dominant, family traits]. In no particular order: Ranjit Banerjee, whose insightful help and constructive discussions advanced the work in obvious leaps; Peter Price, who was always available for successful troubleshooting and sound scientific advice; Mary Ann Sells and Judy Christman for patient introductions to the techniques of molecular biology; Arthur Zelent, whose technical wizardry and scientific genius, although not readily apparent to those who knew him superficially, unselfishly helped when it was needed most; Felix Höppe-Seyler, [who also appreciated Arthur's gifts], for support and friendship; and the laboratories of Jay Unkeless, Carter Bancroft, Jim Roberts, Peter Palese, Jim Bieker. Special thanks to my Dean, Terry Ann Krulwich, whom I knew I could always count on. And I wish to thank George Acs, my mentor, whose advice, wisdom, experience, insight, thoughtfulness, and more, is hopefully apparent in this thesis.

I will thank my parents for more below, but first I thank them mainly for having me last, giving me two elder kin [Judy and Gary] to act as scientific trailblazers and true inspirations; thereby allowing me to greatly benefit from their recent experiences, and realize that my trials were far from unique. I hope they are satisfied that I am a "real" doctor [read Ph.D.]. And to my parents, I am deeply thankful for many things, including; support far above and beyond the expected, their availability, fun, humanity, Hayes Pond, and constant love. It is to them that I dedicate this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

It has been estimated that there are some 50,000-100,000 genes in the 3 x 10^9 base pairs comprising the human genome [Lewin, 1986]. Per cell, roughly 10,000 genes are active at any one time, of which about 1,000-2,000 genes are restricted in expression to specific tissues, or stages of development [classic examples include the liver-specific expression of albumin and α-fetoprotein, and the B-cell restricted expression of immunoglobulins]; while the approximately 8,000 others are collectively grouped as "housekeeping" genes [e.g., actin]. The factors regulating gene expression--especially tissue-specific gene expression--are essentially unknown.

To study the factors controlling tissue-specific gene expression, one must first have a well-defined and workable experimental system. The 3.5 kb mRNA transcript of the hepatitis B virus [HBV], exclusively expressed in human liver cells, is one such system. I have focused my studies on the factors governing its transcription. Although a viral transcript, it must, by nature utilize the cellular transcriptional machinery, and hence is a valid model for the study of tissue-specific cellular gene expression.

Characteristics of the hepatitis B virion

The infectious agent responsible for causing "long-incubation-period serum hepatitis" was first identified by Dane et al [1970]. Although definitely viral in nature, these "Dane particles" possessed characteristics unlike any virus known at that time. Along with its newly accepted nomenclature, the **hepatitis B virus** became the first member of a new family of viruses, the **hepadnaviridae**; coined for the **hepa**totropic scope of infection, and the **DNA** within virus particles.

This family now includes viruses that cause hepatitis in ducks, woodchucks and ground squirrels [Summers, 1981]. Known similarities among all four hepadnaviruses include: viral structure and antigens, genome organization and nucleotide sequence, and patterns of gene expression [see Tiollais et al, 1985; Standring and Rutter, 1986 for recent reviews]. The clinical course of infection differs widely among the animals infected with their homologous hepadnavirus.

Serum from infected individuals can contain up to 10^{10} Dane particles per milliliter, which provides ample material for various detailed evaluations of HBV's constituents. Information on identified viral structure and components is presented in recent reviews by Tiollais et al [1985] and Standring & Rutter [1986]. Ultrastructurally, hepatitis B virions are double-shelled, spherical particles, 42nm in diameter. The lipid-based envelope contains a mixture of variably glycosylated virally-encoded proteins, whose most abundant member is the major envelope protein, the hepatitis B virus surface antigen [HBsAg]. Incubation of virions with non-ionic detergents removes this outer layer to release 28nm nucleocapsid core particles. Hepatitis B virus nucleocapsids are comprised of the hepatitis B core antigen [HBcAg] enclosing the viral DNA, DNA polymerase activity, protein kinase activity, a protein covalently linked to the 5' end of one strand of the viral DNA, and a soluble protein antigenically related to the HBcAg known as the hepatitis B e antigen [HBeAg]. The DNA-linked protein and the protein kinase activity are of unknown origin, while the DNA polymerase activity is thought to be mediated by a protein encoded by the longest open reading frame [ORF] of the viral genome. There is amino acid homology between some residues of retroviral reverse transcriptases and residues within this large ORF, suggesting that this protein also has the reverse transcriptase activity implicated in viral replication [Miller & Robinson, 1986].

Characteristics of the viral genome

HBV DNA isolated from nucleocapsid cores is a partially double-stranded, circular molecule [Figure 1]. One strand is a full genome length linear molecule, approximately 3.2 kb in size, and is usually called the **long** strand [sometimes termed the minus (-) strand]. The **long** strand is terminally redundant, with a nine nucleotide overlap of its 5' and 3' ends occurring within **DR1**, one of two 11 nt direct repeats [Seeger et al, 1986; Will et al, 1987]. The covalently-linked protein is attached to the 5' end of this strand.

The strand complementary to the **long** strand [the **short**, or plus (+) strand] is usually not of full genome length. Its 5' end is approximately 230 nt downstream of the 5' end of the **long** strand, occurring within **DR2**. The 3' end of the **short** strand is extremely heterogenous, usually situated 1.5-2.5 kb downstream from its 5' end. By spanning the terminally redundant ends of the linear **long** strand, the **short** strand maintains the partially double-stranded molecule's loosely circular form. Upon entry into host cells, the endogenous DNA polymerase activity fills out the **short** strand to its full length with host cell deoxynucleotides, thus providing a completely double-stranded, [and hence **bona fide**], transcriptional template.

Coding capacity of the four conserved open reading frames

HBV genomes of various subtypes have been cloned and sequenced [Galibert et al, 1979; Valenzuela et al, 1980; Ono et al, 1983]. There are four open reading frames [ORFs], designated **S**, **P**, **C**, and **X**, strictly conserved in all HBV subtypes, and in the closely related ground squirrel [GSHV] and woodchuck hepatitis virus genomes [WHV] [Figure 1]. These four overlapping ORFs would

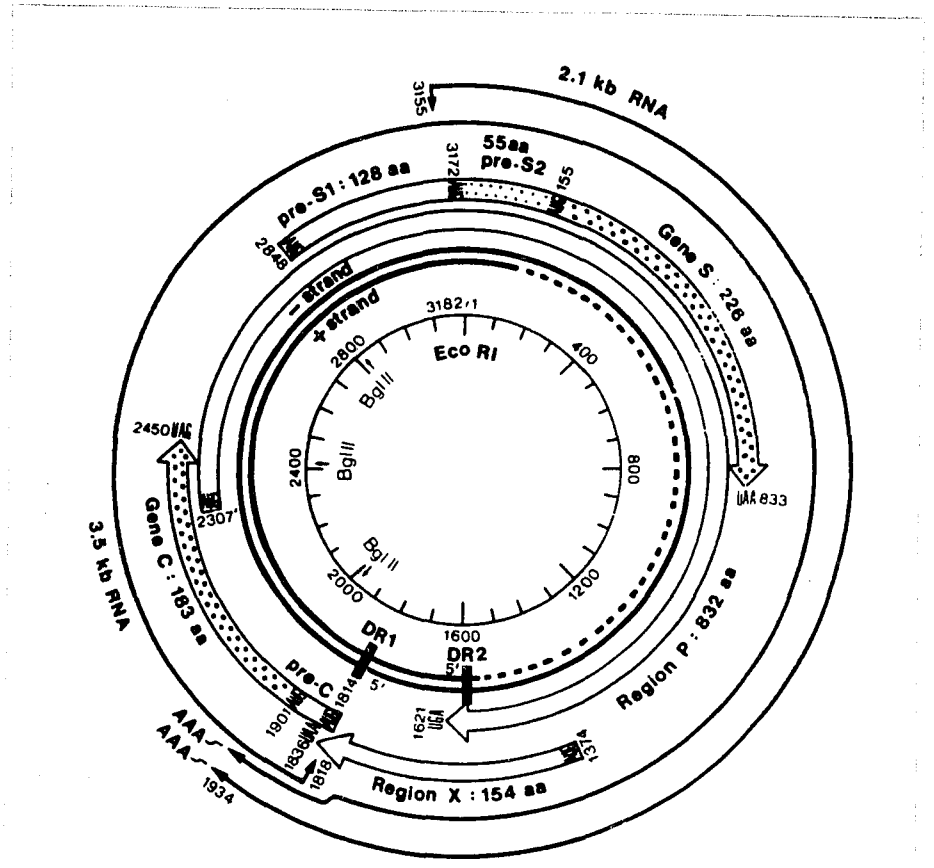

Figure 1 Schematic structure and genetic organization of the HBV genome [subtype ayw]. Starting from the center: restriction enzyme sites and standard reference numbering system beginning at the unique EcoRI site. The two strands labelled - and + are the viral **long** and **short** strands, respectively. The heterogenous 3' end of the + strand is depicted by the dashed completion of this strand. The four conserved open reading frames are shown as broad arrows with in-frame initiation and termination codons indicated. The outermost dark arrows, representing the two major viral transcripts, are marked at their predominant initiation and 3' co-terminal polyadenylation sites. From Tiollais et al, [1985].

be present only in transcripts read from the **long** strand, while there are short, intermittent, and non-conserved ORFs in any potential transcripts encoded by the **short** strand. Thus, at the present time, it appears that all the protein-coding of this small genome resides within these four ORFs of the **long** strand. The compact coding nature of this virus is reflected in the fact that every nucleotide can be grouped in at least one ORF, with half of them participating in two ORF's. All three reading frame registers are used by the virus.

The **S** ORF, is subdivided into pre-S1, pre-S2 and S regions by three in-frame ATG initiation codons. Three variably glycosylated, carboxy co-terminal envelope proteins [termed the large, middle and major surface proteins] are encoded by this ORF. The abundance of each of these proteins in the virion outer membrane has recently been determined, while their functions [presumed to be at least nominally involved in the preferential attachment to, and entry into, human hepatocytes] as a group or individually are unknown [Neurath et al, 1986]. The major protein [HBsAg] is the primary component of highly abundant [up to 10^{13}/ml], 22nm, circulating, spherical and filamentous, non-infectious particles. Since cloning of the genome, HBsAg has been expressed in many systems including yeast cells that produce the recently released commercial vaccine.

The largest ORF, the **P** region, presumably encodes a protein with DNA polymerase/reverse transcriptase activity necessary for viral replication. [These activities have been shown to be involved in replication of the Duck Hepatitis B Virus (DHBV), and are believed to be applicable for all hepadnaviruses; Summers & Mason, 1982.] Homologies have been discovered between a region in this ORF to functionally important regions of cloned retroviral and cauliflower mosaic virus **pol** genes [Toh et al, 1983]. This protein has neither been isolated

from virions, nor expressed in the laboratory.

The **C** ORF, encoding the protein found in nucleocapsid cores, has two in-frame ATG's, thus subdividing it into pre-core and core regions in a similar manner to the **S** ORF. It has not been absolutely determined how the antigenically related HBcAg and HBeAg are encoded by this ORF. Recently published studies suggest that the pre-core:core protein is secreted from cells as HBeAg, while a protein beginning at the second ATG [which does not contain pre-core sequences] is HBcAg [Ou et al, 1986; Roossnick et al, 1986]. The pre-core sequence has homologies to a consensus signal peptide involved in secretion of exported proteins, thus explaining the respective extracellular [primarily serum] and intracellular locations of HBeAg and HBcAg. High quantities of both proteins are produced during infection, with serum HBeAg clinically mirroring the extent of ongoing HBV infection. The nucleocapsid core is an assembly of HBcAg molecules, that are presumed to bind HBV DNA through an arginine-rich region. The function of HBeAg is unknown, but has been proposed as a linker of nucleocapsid to the virion's outer lipid layer, or, in combination with the high quantities of serum HBsAg, may interfere with the immunological response to HBV infection.

The **X** ORF has no known function. Although well-conserved, a protein encoded by this ORF has not conclusively been shown to exist during the course of infection.

Viral transcripts

Information about viral transcripts comes from RNAs isolated from infected livers, virally-associated hepatomas, transfected cell lines, and **in vitro** transcription studies, [in addition to analogous experiments performed with WHV

and GSHV infected animals]. All the protein coding information resides in the **long** strand, which is faithfully reflected in the polarity and span of viral transcripts. Two major unspliced poly-adenylated RNA species, 2.1 and 3.5 kb in length, have been found [Figure 1: Cattaneo et al, 1984; Möröy et al, 1985; Enders et al, 1985]. These 3' co-terminal transcripts are present in approximately equivalent amounts in acutely infected livers. 5' ends of the 2.1 kb mRNA [also called the HBsAg transcript] are heterogenous, dispersed throughout a thirty nucleotide span about the unique EcoRI restriction site in the viral genome [Cattaneo et al, 1983, 1984; Standring et al, 1984; Siddiqui et al, 1986]. All 2.1 kb transcripts can encode the major HBsAg and putative XAg, while the largest and least abundant 2.1 kb transcript includes the pre-S2 AUG, and thus can encode the middle surface protein. The sole polyadenylation signal in use by the virus is at nt 1934 [Simonsen & Levinson, 1983].

The 5' end of the 3.5 kb mRNA is also heterogenous, with the most abundant species mapped within the pre-core region to nt 1816 ± 5 [Will et al, 1987; Enders et al, 1985 for GSHV]. This mRNA is larger than genome length and can encode HBcAg, the putative DNA polymerase/reverse transcriptase, the large, middle, and major envelope proteins, and the XAg. Evidence for the multi-functional coding capability of this transcript is sparse, with most researchers assigning it principally as the HBcAg transcript. One report has described core-pol fusion proteins that would be encoded by this transcript [Will et al, 1986]. Equal in importance to its protein coding capabilities, this mRNA is a replicative intermediate, since it serves as a template for reverse transcription [hence its referral as the pregenome RNA; Summers & Mason, 1982; Seeger et al, 1986]. Only infected liver and recently developed **in vitro** replication systems produce large quantities of 3.5 kb mRNA.

Minor HBV transcripts that originate upstream of the pre-S1 region, the **X** ORF, and the pre-core region have been found [Enders et al, 1985; Will et al, 1987]. These have been proposed to encode proteins containing these regions; replacing undetected splicing, or multiple translational initiations, within the 3.5 kb mRNA as explanations for their expression. **Short** strand transcripts have not been detected in HBV-infected liver, although a transfected mouse fibroblast cell line, under study in this laboratory, has been shown to contain such transcripts which initiate and terminate within transfected viral sequences [Zelent et al, 1987]. One other **short** strand transcript, dependent upon RNA Polymerase III, has been found, using a cell-free **in vitro** transcription system [Standring et al, 1983].

Cis- and trans-acting factors controlling transcription

Components involved in transcriptional regulation can conveniently be divided into **cis**-acting elements and **trans**-acting factors. A cis-acting element is difficult to define, but is typically a regulatory DNA sequence flanking, or within a gene, that influences the expression of that gene. Examples include promoters, enhancers, and inducible elements. The effectiveness of some cis-acting elements [primarily those in promoter regions] is limited to within a specified distance from the start site of transcription; such as the TATA and CCAAT boxes, or Sp1 binding sites [Yaniv, 1984; Dynan & Tjian, 1985; McKnight & Tjian, 1986; Briggs et al, 1987]. On the other hand, enhancers possess the curious characteristic of stimulating transcription in a distance and orientation-independent manner. Enhancers have been discovered in many viral and cellular genes, sometimes placed thousands of nucleotides from the start of transcription [Banerji et al, 1981; Dynan & Tjian, 1982; Khoury & Gruss, 1983;]. Furthermore,

enhancers have also been suggested as playing an integral role in the tissue-specific expression of many genes [Dunn & Gough, 1984; Voss et al, 1986; Dynan, 1987]. Inducible cis-acting elements include the serum responsive element of c-fos, the heat shock responsive element of Drosophila hsp genes, and the metal responsive elements of metallothionein genes [Prywes & Roeder, 1986; Wu, 1984; Karin et al, 1984]. Cis-acting elements are usually mapped by DNA transfection assays, combined with detailed deletional and mutational studies.

Trans-acting factors involved in transcriptional regulation can be general or specific, regarding both gene-specific action and tissue distribution. RNA Polymerase II and auxiliary factors are typical examples of general trans-acting transcriptional factors. Trans-acting factors that selectively affect the expression of certain genes, are usually soluble nuclear proteins possessing a characteristic sequence-specific DNA binding domain [Dynan & Tjian, 1985]. This domain mediates its regulatory action by binding to a specific recognition sequence in the DNA within the locus of the gene it regulates.

The interplay of cis-acting elements and sequence-specific trans-acting factors determines the rate of transcript initiation of many genes, and has been suggested as the means behind tissue-specific, and stimulus-dependent, gene expression [Dynan & Tjian, 1985; Maniatis et al, 1987]. For example: a protein found exclusively in B cells [NF-KB] specifically recognizes a sequence in the kappa gene enhancer, and by binding to that site, activates transcription of B cell-restricted immunoglobulin kappa genes [Sen & Baltimore, 1986; Atchison & Perry,1987]. This model of the control of gene expression is analogous to the prokaryotic lac operon or lambda cro/repressor systems known for many years [Pabo & Sauer, 1984].

It has been suggested that due to the number of eukaryotic genes and the complexity of their expression, regulation consists of multiple cis-acting elements interacting with several trans-acting factors [Dynan & Tjian, 1985; Lewin, 1986; Ptashne, 1986]. Thus, the cells of different tissues would contain distinct mixtures of nuclear trans-acting factors, that establish their characteristic transcription patterns. This would limit the number of different factors to a reasonably finite figure, mathematically estimated to be as little as ten [Voss et al, 1986]. This considerably simplifies the next level of analysis: that is, the determination of the control of expression of the genes encoding these regulatory factors. The methodology of describing, and then purifying, trans-acting factors has recently taken large steps forward, indicating that much more information on these proteins will soon be available. As of now, a few eukaryotic trans-acting factors have been purified to homogeneity, including: the glucocorticoid, estrogen, and progesterone receptors [also cloned and sequenced; reviewed by Green & Chambon, 1986], Sp1 [Briggs et al, 1987], CTF/NF-1 [Jones et al, 1987], and MLTF/USF [Chodosh et al, 1987].

Regulation of HBV mRNA transcription

Some work has been published on cis-acting elements regulating HBV transcription. Promoter regions have been outlined for the major 2.1 kb [Siddiqui et al, 1986] and 3.5 kb [Roossnick et al, 1986] mRNAs, and also for some of the minor transcripts [Siddiqui et al, 1987; Treinin & Laub, 1987]. Rigorous deletional and mutational analyses must still be performed in order to fully understand which sequences within these large domains are important. Sequence analysis of the 2.1 kb mRNA promoter [maximally between nt 2834-34] reveals homology to the SV40 late promoter region, explaining its relatively

ubiquitous expression in HBV DNA transfected cells [Cattaneo et al, 1983]. Unlike many viral and cellular promoters, TATA sequences are not present near the 2.1 kb mRNA initiation site, explaining the heterogeneity of its 5' ends.

A consensus on the region defining the 3.5 kb mRNA promoter [also known as the core promoter] has not been reached. The confusion has arisen from the use of different plasmid constructs, cell lines, and **in vitro** cell extracts, used to study the transcription of this mRNA. Presumably, by analogy to other eukaryotic promoters, the core promoter encompasses a 100-300 nt region containing the start site of 3.5 kb mRNA transcription. A detailed mapping of the cis-acting elements within this promoter region is most important since the expression of this mRNA is restricted to infected livers and human liver-derived cell lines.

A tissue-specific enhancer element has been discovered in the HBV genome, by two groups using different methodologies [Shaul et al, 1985; Tognoni et al, 1985]. Both groups map the enhancer between the **S** and **X** ORFs to a 250bp region centered about nt 1150. This element acts in a distance and orientation-independent manner characteristic of other enhancers. Interestingly, it is highly active in cell lines of human liver origin and virtually inactive in other cell lines, strongly suggesting that it is the guiding force behind the liver-specific expression of the pregenomic, 3.5 kb mRNA.

Other cis-acting elements discovered in the HBV genome include a nuclear factor I [NF-I] binding site upstream of the 2.1 kb promoter [Shaul et al, 1986], and a glucocorticoid responsive element near nt 600 [Tur-Kaspa et al, 1986].

As mentioned previously, full transcriptional competence relies on the interaction of cis-acting DNA sequences with appropriate trans-acting factors. Information on the trans-acting factors controlling HBV transcription is sparse,

including the aforementioned NF-I and glucocorticoid receptor, and recently described nuclear proteins extracted from various tissue culture cells, [some of human liver origin], that bind to four different regions of the HBV enhancer [Shaul et al, 1987]. An apparent liver-specific protein protects one of the regions. An earlier report, by Jameel & Siddiqui [1986], suggested the involvement of liver-specific factors controlling HBV enhancer function. To fully understand the regulation of HBV transcription, [i.e., the cis- and trans-acting factors involved], much more work must be done.

Exploitation of the sequence-specific binding of trans-acting factors has served as the basis for some recently developed methods of study of these proteins, and the interactions with their associated recognition sites in the DNA [Wu, 1985; Singh et al, 1985]. The work presented in this thesis employs one of these techniques, the gel electrophoresis mobility shift assay [Fried & Crothers, 1981; Garner & Revzin, 1981], to demonstrate the presence of distinct proteins in crude nuclear extracts that bind within two HBV DNA restriction fragments. These two fragments, encompassing the enhancer and core promoter domains of the HBV genome, are shown to contain cis-acting DNA elements controlling transcription of the liver-specific 3.5 kb mRNA. Evidence is presented confirming the functionality of these sequences using **in vivo** competition co-transfection assays. Furthermore, binding sites within the two fragments have been mapped at the nucleotide level, and the binding proteins are restricted in appearance to extracts from a human liver-derived cell line. Although more work needs to be done to demonstrate absolutely the involvement of these liver-specific proteins in governing 3.5kb mRNA expression, it is strongly suggested by these studies that these proteins are integral for the expression of this multifunctional, tissue-specific transcript.

Methods

Cells and culture conditions All cells were maintained in sterile Corning 75 cm^2 or 150 cm^2 tissue culture flasks in a humid environment containing 5% CO_2 at 37°C in standard incubators. HepG2 cells [obtained from Barbara Knowles; Knowles et al, 1980] a human hepatoblastoma-derived cell line, were maintained in Minimum Essential Medium (Eagle's) supplemented with 10% fetal calf serum, 100 Units/ml penicillin and 100 µg/ml streptomycin [all tissue culture solutions were purchased from GIBCO Laboratories]. NIH 3T3 cells [mouse fibroblasts] were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 100 Units/ml penicillin and 100 µg/ml streptomycin.

Isolation of nuclei Typically, between five and ten confluent 150cm^2 flasks were used as starting material for isolation of nuclei. Monolayers were first washed twice with ice-cold phosphate buffered saline [PBS: 140 mM NaCl; 2.7 mM KCl; 8.1 mM Na_2HPO_4; 1.5 mM KH_2PO_4, pH 7.4]. Next, 5 ml of Lysis buffer [10 mM Tris-HCl, pH 7.4; 10 mM NaCl; 3 mM $MgCl_2$] was added to each flask and let remain on cells at room temperature for 5 minutes. Lysis of plasma membranes was achieved by subsequent addition of 250 µl of a 10% Nonidet P40 [Fluka Chemical] solution directly to the Lysis Buffer for 10 minutes at room temperature with gentle, intermittent, swirling. Plasma membrane lysis was observed by microscopic observation. The lysate and intact nuclei were removed from the monolayer by pipetting until all was in suspension, and then transferred to a sterile Corning disposable 50 ml polypropylene tube kept on ice. Nuclei were pelleted by centrifugation in the

DuPont Sorvall model RT6000 at 2,000rpm for 10 minutes at 4°C, then washed of residual cytoplasmic debris with 25 ml of ice cold Pellet Washing Buffer [10 mM Tris-HCl, pH 7.4; 10 mM NaCl; 25 mM EDTA] and pelleted as above. The washing procedure was repeated twice more, and the nuclear pellet quick frozen on powdered dry ice and stored at -70°C until ready to be extracted.

Nuclear Extract Preparation Nuclear extract was prepared essentially as described by Dignam et al, [1983] with some modifications. Frozen nuclei were thawed on ice and transferred to a chilled 10 ml glass homogenizer [Wheaton]. 3 ml of ice-cold buffer C [20 mM HEPES, pH 7.9 at 4°C; 25% (vol/vol) glycerol; 0.42 M NaCl; 1.5 mM $MgCl_2$; 0.2 mM EDTA; 0.5 mM phenylmethylsulfonylfluoride (PMSF); 0.5 mM DTT] was added and the nuclei were lysed by homogenization with a close-fitting Teflon pestle [twenty to thirty strokes]. The solution was transferred to a chilled 30 ml Corex tube and stirred gently on ice for 30 minutes. The lysate was then centrifuged in a Sorvall SS-34 rotor at 14,500 rpm for 30 minutes at 4°C. The supernatant was transferred to dialysis tubing [cutoff at 6-8 kDa] and dialyzed against 500 ml of buffer D [20 mM HEPES, pH 7.9; 20% vol/vol glycerol; 100mM KCl; 0.2 mM EDTA; 0.5 mM PMSF; 0.5 mM DTT] for four to seven hours in a 4°C cold room. The dialysate was then centrifuged in the SS-34 rotor at 14,500 rpm for 30 minutes at 4°C to spin down insoluble debris, and the supernatant dispensed in 200 µl aliquots, quick frozen on powdered dry ice, and stored at -70°C until used. When ready for experimentation, one 200 µl aliquot was thawed on ice in the 4°C cold room, and 25 µl aliquots were dispensed, quick frozen, and stored at -70°C as above. Once thawed, these 25 µl aliquots were never re-frozen.

Protein concentrations were determined with the Bio-Rad Protein Assay Kit

and Bovine Serum Albumin as the standard. From 2×10^8 Hep G2 cells, approximately 15 mg of nuclear protein was isolated.

Recombinant Cloning Procedures

Plasmids:

pSVoCAT: obtained from B. Howard (NIH); [Gorman et al, 1982] a promoterless pBR322 plasmid containing the 0.8 kb coding region for the chloramphenicol acetyltransferase gene followed by eukaryotic transcriptional polyadenylation sequences. [See Figure 2]

pTHBV-1: tandem HBV genomes [subtype ayw; isolated and cloned in this laboratory] cloned via the unique EcoRI restriction sites into the EcoRI site of pBR322 [Price et al, 1980]

pIBI31: [International Biotechnologies Inc.; see Figure 3] High yield multi-purpose cloning vector.

Recombinant Plasmid Constructions

RECOMBINANT	VECTOR DNA	INSERT DNA
pEcoAluCAT	pSVoCAT/HindIII	HBV nt 1-1878, pTHBV-1/EcoRI-AluI restriction fragment, isolated, and HindIII linkers added
pCP-1	pIBI31/SmaI	HBV nt 1645-1888; from pTHBV-1/StyI, isolated, blunt-ended with the Klenow fragment of E. Coli DNA Polymerase
pENH-2	pIBI31/SmaI	HBV nt 1082-1378; from pTHBV1/StyI + DdeI, isolated, Klenow fragment blunt-ended
pH-2	pIBI31/EcoRI	complete HBV genome, nt 1-3182; isolated from pTHBV-1/EcoRI
pMLP	pIBI31/SmaI	Adenovirus SmaI "F" restriction fragment; contains the Major Late Promoter region

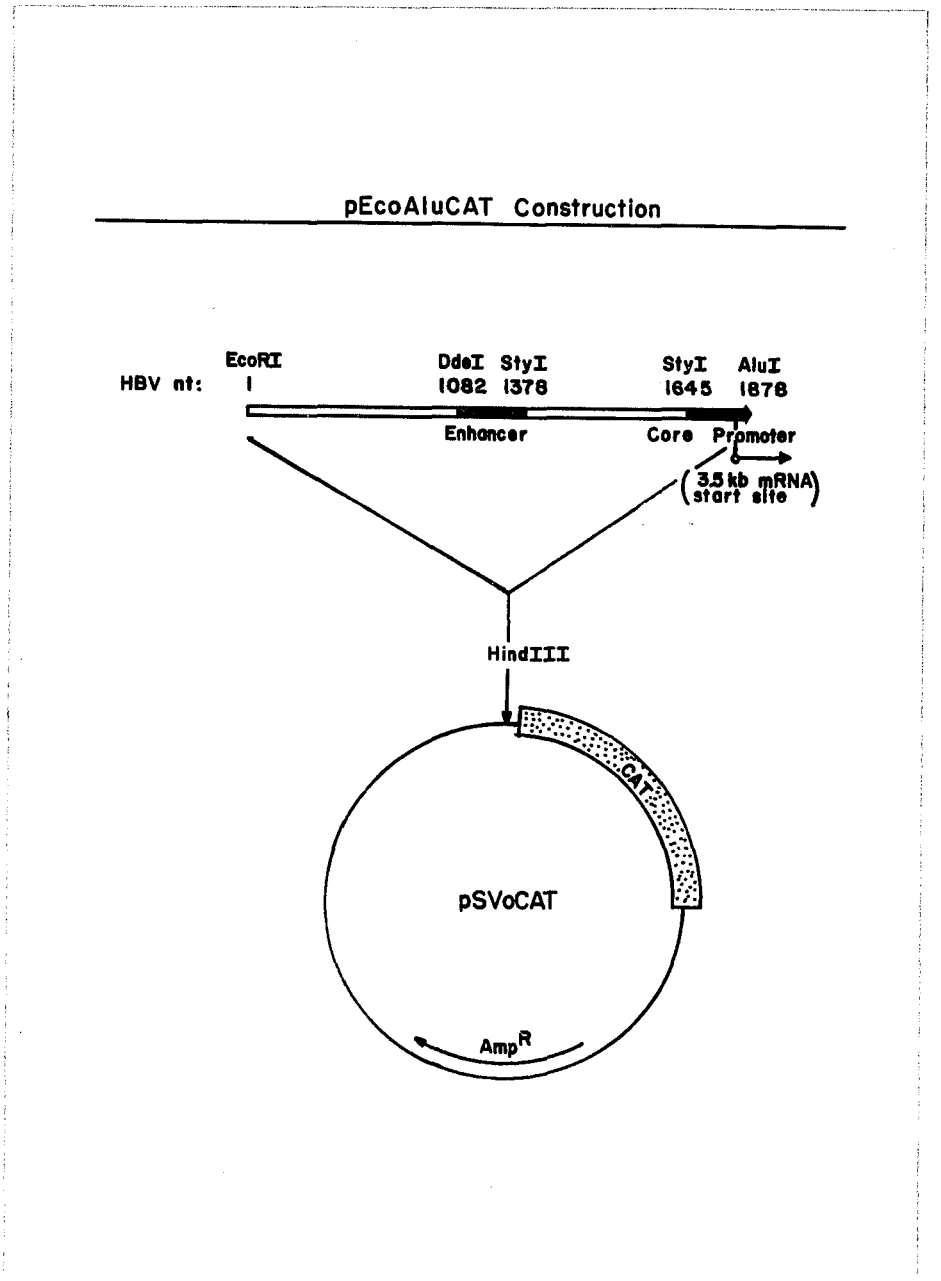

Figure 2 Construction of pEcoAluCAT. The EcoRI-AluI fragment [nt 1-1878] was inserted, via ligated HindIII linkers, into the unique HindIII site of pSVoCAT [Gorman et al, 1982]. The two regions within the EcoRI-AluI insert used for further experiments are indicated by black boxes. Amp^R denotes the ß-lactamase gene used for selection of plasmid-bearing bacteria.

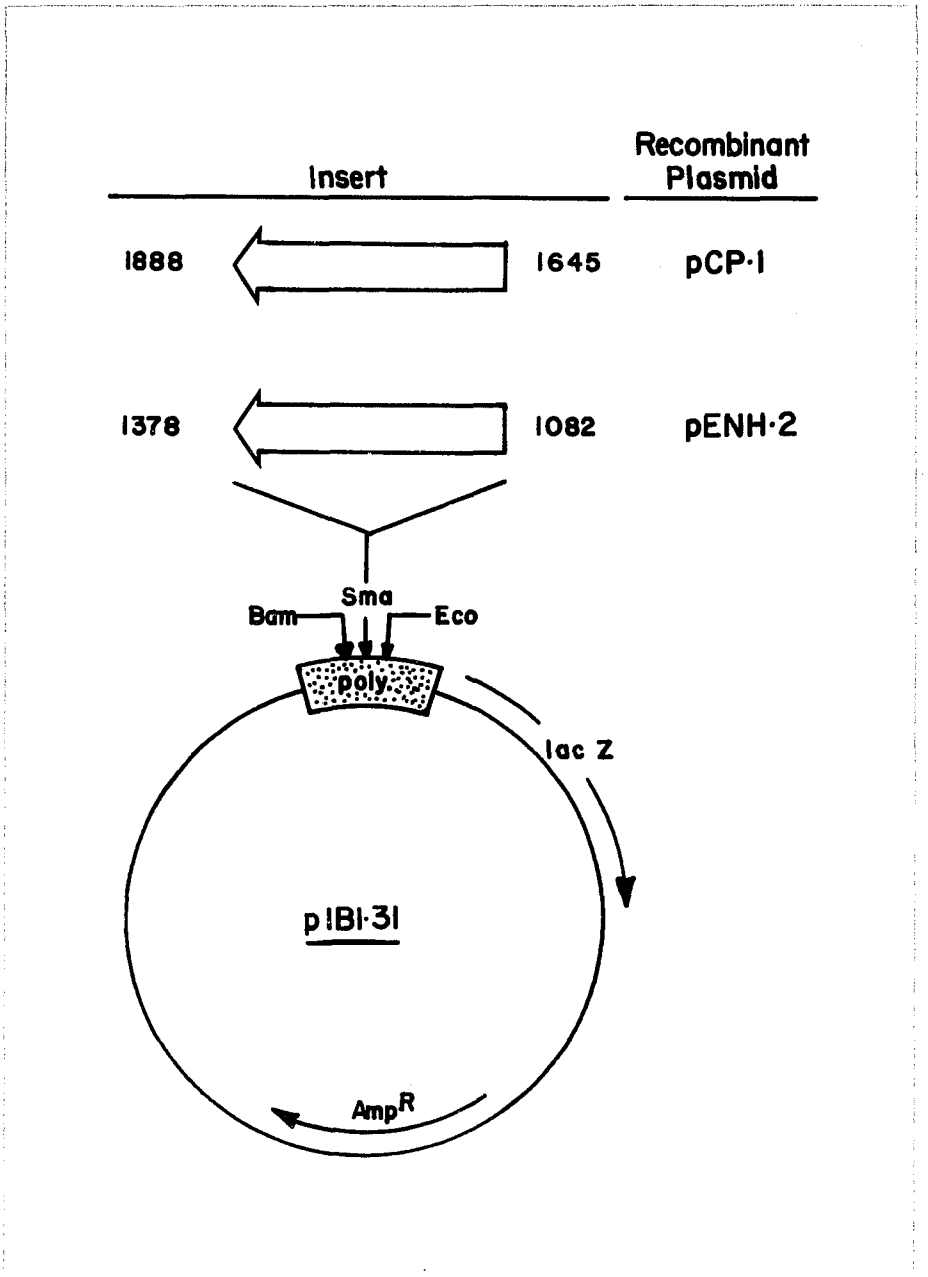

Figure 3 Schematic representation of the construction of pCP-1 and pENH-2. Both gel purified insert DNAs were blunt-ended prior to ligation with SmaI cut pIBI31. Poly. is the polylinker region, and lac Z indicates the ß galactosidase gene, whose product converts the chromogenic substrate X-Gal to a blue, insoluble precipitate. Selected unique restriction sites, BamHI [Bam], SmaI [Sma], EcoRI [Eco] are indicated. 5' kination of BamHI cut pCP-1 or pENH-2 labelled the **long** strand, while 5' kination of EcoRI cut plasmids labelled the **short** strand of HBV.

Restriction enzyme digestions and agarose gel electrophoresis Restriction enzymes were purchased from New England BioLabs [NEB] or Boehringer Mannheim Biochemicals [BMB] and used according to manufacturer's instructions. DNA fragments were separated using submarine agarose gel electrophoresis. Routine DNA fragment preparations and analytical digestions were resolved in 0.8-1.6% SeaKem ME agarose [FMC Corporation, Marine Colloids Division] gels in 1 x TAE buffer [40 mM Tris-HCl; 5 mM EDTA; 1 mM EDTA, pH 8.0]. Samples were loaded into wells with 6 x Loading Buffer [0.25% bromphenol blue; 0.25% xylene cyanoll FF; 30% glycerol]. Simple patterns were separated in an 8 x 10 cm minigel apparatus [Ellard Instrumentation], while more complex ones were run in a 11 x 14 cm gel [Bethesda Research Labs]. Gels were stained in 1 μg/ml ethidium bromide solution. DNA bands were visualized with short wave ultraviolet light, and Polaroid photos were taken when appropriate.

DNA ligation protocol Typically, 0.5 μg of dephosphorylated vector DNA and 1.0 μg of insert DNA were ligated in a 10 μl reaction containing 2 μl of 5x Ligation buffer [250 mM Tris-HCl, pH 7.5; 50 mM MgCl2; 250 mg/ml polyethylene glycol 8000 (Baker Chemical); 5 mM ATP; 0.5 mM DTT; in single-use aliquots stored at -70°C] and 0.5 μl of T4 DNA Ligase [New England BioLabs, at approximately 400 Units/μl]. Incubation was in a 14°C circulating water bath overnight. For blunt-end ligations, vector DNA was not dephosphorylated.

Bacterial transformation Bacterial strains HB101 or DH1 were used to harbor CAT plasmids, while strain JM83 was used with pIBI31 and its derivatives. Transformations were performed as described by D. Hanahan [1983].

Competent bacteria were transformed with one-half of a ligation reaction's contents [typically 0.5-1.0 µg]. Transformed bacteria were selected on plates consisting of L-Broth [per liter: 10 gm Bacto-Tryptone; 5 gm Yeast Extract; 10 gm NaCl; 1 gm glucose] containing 50 µg/ml ampicillin and 1.5% Bacto-Agar [Bacto-Tryptone, Yeast Extract, Bacto-Agar purchased from DIFCO]. JM83 recombinants were selected on these same plates coated with 100µl of X-Gal:IPTG Solution [from a freshly prepared solution of 1000µl 2% w/v X-Gal (dimethyl formamide); 100 µl 200 mM IPTG,; 900 µl L-Broth containing 100 µg/ml ampicillin] a few hours prior to spreading of the transformation mixtures. X-Gal [5-bromo-4-chloro-3-indoyl-ß-D-galactoside] and IPTG [isopropylthio-ß-D-galactoside] were purchased from Boehringer Mannheim Biochemicals [BMB]. Plates were incubated overnight at 37°C prior to screening.

Screening of recombinant bacterial clones Two different screening procedures were used--one for pEcoAluCAT, and the other for pIBI31 derivatives. For pEcoAluCAT, screening was performed as directed by the procedure provided by Bethesda Research Labs, Inc. included with their NACS [Nucleic Acid Chromatography System] columns.

Screening of recombinant pIBI31 plasmids [pCP-1, pENH-2, pH-2, pMLP] was performed in a quicker and simpler fashion. When grown on the plates described above, bacterial colonies harboring vector pIBI31 plasmid were blue in color, demonstrating that the gene for B-galactosidase [poly-lac Z in Figure 3] was uninterrupted and hence faithfully expressed. Bacterial colonies containing recombinant pIBI31 molecules [interrupted in the coding region for B-galactosidase] were unable to convert the substrate [X-Gal] to an insoluble blue precipitate, and were therefore uncolored, or "white." Two white colonies were

picked, grown up in 150ml unamplified overnight cultures of L-Broth containing 50 µg ampicillin, and plasmid was subsequently isolated as detailed below.

Plasmid isolation All plasmids used in these studies were isolated from bacteria using a modification of the original alkaline lysis technique of Birnboim & Doly [1979], followed by CsCl gradient ultracentrifugation. Seed cultures from frozen bacterial stocks were inoculated in 1 ml of growth medium [L-Broth containing the 50 µg/ml ampicillin selection] in sterile Falcon 2059 tubes and grown overnight at 37°C in a Queue Shaker rotating at 225rpm. The next day, 100-200 µl of this seed culture was inoculated into either 150 ml [in Corning 25600 250 ml disposable flasks] or 1.5 liters [in 2800 ml Fernbach flasks] of growth media [L-Broth with 50 µg/ml ampicillin] to grow overnight as described for the seed culture. The next day, bacteria were harvested by centrifugation in 250 ml bottles in Sorvall fixed-angle rotor at 5000rpm x 10 minutes at 4°C. For up to 500 ml of cultures, the bacterial pellet was resuspended in 4 ml of GET buffer [25 mM Tris-HCl; 10 mM EDTA; 1% Glucose, ph 8.0] and transferred to 50 ml Nalge 3139 polyallomer tubes. 0.5 ml of a freshly prepared 25 mg/ml lysozyme [Calbiochem] solution [in GET] was added and the tubes kept on ice for 15 minutes. 9 ml of freshly prepared alkaline lysis solution [0.2 M NaOH; 1% SDS] was then added, mixed and shaken thoroughly, and placed on ice for 15 minutes. 6.75 ml of cold 3 M KOAc, pH 4.8 [prepared according to Maniatis et al, 1982] was added, and the tube vigorously shaken and placed on ice for 15 minutes. Tubes were then centrifuged in an SS-34 rotor at 13,000 rpm x 20 minutes at 4°C. The supernatant was passed through a single layer of cheesecloth into a 50 ml Nalge 3139 tube, and the material precipitated by addition of 14 ml isopropanol at room temperature for 20 minutes. After

centrifugation [HB-4 rotor at 10,000rpm for 15 minutes at 4°C], the pellet was rinsed with 75% EtOH ,re-spun, inverted, and air dried for 10 minutes. The pellet was resuspended in 5 ml of TEN_{10} [10 mM Tris-HCl; 1 mM EDTA; 10 mM NaCl, pH 8.0], 50 µl of RNAase A [BMB; rendered DNAase-free according to Maniatis et al, 1982] was added and incubated at 37°C for 60 minutes. 550 µl of 3 M NaOAc, pH 5.2 [also prepared according to Maniatis et al, 1982] was added, and the solution transferred to a Falcon 2059 tube. The solution was then extracted with 6 ml of Phenol:Chloroform:Isoamyl Alcohol [25:24:1 volume ratio--phenol prepared according to Maniatis et al, 1982], and centrifuged at 7,500 rpm for 10 minutes at 4°C in the HB-4 rotor. The supernatant was re-extracted with Chloroform:Isoamyl alcohol [24:1 volume ratio], centrifuged, transferred to a Falcon 2059 tube and precipitated at room temperature for 20 minutes with 6 ml isopropanol. The precipitate was pelleted in an HB-4 rotor at 8000 rpm for 30 minutes at 4°C, washed with 75% EtOH and re-spun, and then vacuum dried. The pellet was resuspended in 10 ml of 51% weight/volume Cesium Chloride:TE solution [10 mM Tris-HCl; 1 mM EDTA, pH 8.0], 300 µl of 10 mg/ml Ethidium Bromide was then added in relative darkness, and this solution was transferred to one Beckman polyallomer Quick Seal Ti75 tube via a syringe capped with 16 gauge needle. The tube was topped with 51% weight/volume Cesium Chloride:TE solution, heat-sealed and centrifuged in a Beckman Ti75 rotor at ≥55,000 rpm for a minimum of 14 hours, at 20°C.

The plasmid DNA was removed from the tube under long-wave ultraviolet light, by puncturing the tube below the plasmid DNA band and slowly withdrawing the DNA through a 23 gauge needle attached to a 3 ml syringe. Usually very little, if any, chromosomal DNA was present. This solution was transferred to a sterile Corning 50 ml disposable polystyrene tube and extracted

three times with 40 ml of water-saturated isobutanol. The aqueous solution was transferred to a Nalge 3139 50 ml tube, diluted with 2 volumes of sterile ddH20 and precipitated at -20°C for at least two hours with the addition of 1/9th volume of 3 M NaOAc, pH 5.2 and 2 volumes of 95% EtOH. The DNA was pelleted by centrifugation in the HB-4 rotor at 8,000 rpm for 30 minutes at 4°C, washed with 75% EtOH, re-spun, vacuum dried for 30 minutes, and resuspended in TE [10mM Tris-HCl; 1 mM EDTA, pH 8.0]. Yield and purity was determined by spectrophotometry [at 230, 260 & 280 nm wavelengths; 230:260:280 ratio of 1:1.8-2.0:1 considered acceptable] and restriction digestion analysis.

Plasmid DNA transfection Cultured cells were transfected by the calcium phosphate method of Graham & van der Eb [1973], incorporating modifications published by Wigler et al [1980]. 24 hours prior to transfection, growing cells were seeded in 10 cm plastic culture dishes at a density of approximately 550,000 per dish. Four hours prior to transfection, cells were refed with appropriate growth media. The calcium phosphate:DNA co-precipitate was prepared in the following manner: a 250 mM $CaCl_2$ solution containing 60 µg/ml DNA was slowly added to an equal volume of gently aerated Hepes buffered saline [50 mM Hepes; 280 mM NaCl; 1.5 mM Na_2HPO_4, pH 7.12 ±0.02]. Thirty minutes later, 1 ml of this solution was added to each culture dish, and incubated for four to six hours. Media was then changed, and the cells remained undisturbed--until harvested for CAT activity.

Chloramphenicol acetyltransferase [CAT] assays Gorman et al [1982] first described the CAT transient expression assay. About 48 hours after

transfection, cells were washed three times with PBS. Five minutes after adding 1 ml of Tris Collection Buffer [40 mM Tris-HCl, pH 7.4; 1 mM EDTA; 150 mM NaCl] to each plate, cells were scraped off the plate into a 1.5 ml microcentrifuge tube with a rubber policeman. Tubes were quick-spun to pellet cells. Cell pellets were resuspended [with vigorous vortexing] in 100 µl of 250 mM Tris-HCl, pH 7.8. Three freeze-thaw cycles [each cycle consisted of approximately fifteen minutes alternately in a dry ice/EtOH bath and 37°C water bath] were performed to lyse cells. After a quick microfuge spin to pellet insoluble particles, CAT activity in each lysate was then determined.

The reaction mixture contained 55 µl of cell extract, 70 µl of 1 M Tris-HCl, pH 7.8, 4.5 µl sterile dH_2O, 0.5 µl ^{14}C-chloramphenicol [about 0.2 µCi; New England Nuclear, 50 µCi/µmol] and 20 µl of 4 mM acetyl coenzyme A [Pharmacia Biochemicals; freshly prepared], incubated together at 37 °C in a 1.5 ml microcentrifuge tube for 90 minutes. The radioactive chloramphenicol species were extracted with 1 ml ethyl acetate, and dried down in a Savant Speed-Vac for 45 minutes. Dried samples were resuspended in 30 µl of ethyl acetate, and spotted onto 20 cm x 20 cm silica gel thin layer chromatography plates [EM sciences art. 5735-7]. Plates were developed in glass tanks with chloroform:methanol [95:5], ascending for about 90 minutes. Plates were autoradiographed with Kodak XAR-5 film for a few days.

Radioactive forms were cut from the TLC plates and counted. CAT activity [per cent acetylated per µg protein] was calculated as:

$$\frac{100 \times (\text{1-Acetate cpm} + \text{3-Acetate cpm})}{(\text{1-Acetate cpm} + \text{3-Acetate cpm} + \text{unacetylated chloramphenicol cpm}) \times (\mu\text{g protein per reaction})}$$

Relative CAT activity was normalized to CAT activity in lysates from cells transfected with 2.5 µg of pEcoAluCAT per plate.

DNA radiolabelling protocol

Used for labelling pCP-1 and pENH-2 HBV inserts [Figure 3]. Plasmids were linearized with BamHI or EcoRI, and the 5' ends dephosphorylated. The linear fragment was gel purified, precipitated and kinated with gamma-labelled ATP. The radiolabelled DNA was cleaved with the appropriate enzyme to "free" the HBV insert [i.e., BamHI labelled fragments were cut with EcoRI and vice versa], and the labelled inserts were gel purified. From the inserts' orientations, the **long** strand of HBV was labelled at the BamHI site, while the **short** strand was labelled at the EcoRI site, for both pCP-1 and pENH-2.

Dephosphorylation of 5' ends To increase the efficiency of radioactively labelling DNA, the protruding 5' ends were dephosphorylated with calf intestine alkaline phosphatase. After restriction enzyme digestion, not less than 1 Unit of calf intestine alkaline phosphatase [BMB, molecular biology grade] per 50 pmoles of 5' ends was added and the reaction was continued [in the restriction digestion solution] for 60 minutes at 37°C. EDTA was added to 10 mM final concentration and the reaction was then heated to 68°C for 20 minutes. 6X Loading buffer was added in the appropriate volume, and the solution was loaded into a preparative submarine agarose gel.

Kination of 5' protruding ends Kination of dephosphorylated, 5' protruding ends was carried out at 37°C for 30 to 60 minutes. A typical 50 µl reaction contained 3 pmoles of a linear DNA fragment, 5 µl of fresh gamma-labelled-ATP [Amersham Corporation; 10 µCi/µl, 5000 Ci/mmole specific activity], 5 µl of 10x Reaction Buffer [to give a final solution of 50 mM Tris-HCl, pH 7.5; 10 mM $MgCl_2$; 5 mM DTT; 1 mM spermidine; 1 mM EDTA], and 1 µl of T4

Polynucleotide Kinase [BMB; 10 Units/µl]. The reaction was stopped by heating to 68°C for 20 minutes, and then precipitated with 2 M ammonium acetate and 2 volumes of 95% EtOH.

Kination of blunt ends This reaction was used for labelling of ϕX174/HaeIII DNA markers. 1 µg of dephosphorylated ϕX174/HaeIII DNA was resuspended in 35 µl of "A" buffer [20 mM Tris-HCl, pH 9.5; 1 mM spermidine; 0.1 mM EDTA], heated at 90°C for 2 minutes and chilled on ice for 2 minutes. 5 µl of "B" buffer [500 mM Tris-HCl, pH 9.5; 100 mM $MgCl_2$; 50 mM DTT; 50% glycerol], 5 µl of gamma-labelled-ATP, and 1 µl of T4 Polynucleotide Kinase were added and the reaction proceeded at 37°C for 60 minutes. The reaction was then neutralized with 2 µl of 2 M Tris-HCl, pH 7.5. Unincorporated ATP was removed by passage of the reaction over a 10 ml Sephadex G-50 medium [Pharmacia] column as per Maniatis et al [1982].

Isolation of gel-separated DNA fragments with NA-45 membrane Used primarily for recovery of small quantities of single-end-labelled DNA for mobility shift and methylation interference assays. After preparative gel electrophoresis [cast and run in 1 x TAE buffer with 0.5 µg/ml ethidium bromide] yielded suitable separation of restricted DNA fragments [as visualized under long wave ultraviolet light], a slit slightly wider than the sample well was cut in the gel ahead of the band of interest. A strip of NA-45 membrane [Schleicher & Schuell; pretreated as recommended by manufacturer] was placed in the incision, and electrophoresis continued until all the staining of the band of interest was at the membrane [usually within 30 minutes]. The membrane was removed and rinsed in NET buffer [150 mM NaCl; 0.1 mM EDTA; 20 mM

Tris-HCl, pH 8.0] for 5 minutes. Elution of DNA from the membrane was performed by submerging the membrane in a sufficient volume of High Salt NET Buffer [1.0 M NaCl; 0.1 mM EDTA; 20 mM Tris-HCl, pH 8.0] in a 1.5 ml Eppendorf. This tube was incubated at 65°C, with occasional vortexing, for 30 to 45 minutes. After a quick microcentrifuge spin, the supernatant was transferred to another tube, and the membrane washed with 100 µl of High Salt NET Buffer, which was combined with the first supernatant. The complexed ethidium bromide was removed with three successive water-saturated isobutanol extractions, and the DNA precipitated with two volumes of 95% EtOH. After an overnight precipitation at -20°C, the tubes were microcentrifuged for 30 minutes, and the pellet washed with 75% EtOH, re-spun, vacuum dried and resuspended in 400 µl of 0.3 M NaOAc, pH 5.2. This solution was then sequentially extracted with equal volumes of phenol, phenol:chloroform:isoamylalcohol, and chloroform:isoamyl alcohol, and then precipitated upon addition of 1000 µl of 95% EtOH. After precipitation overnight at -20°C, the precipitate was pelleted, washed and vacuum dried as above, and resuspended in 50 to 100 µl of TE. If radioactively labelled, 1 µl of this solution was counted in 10 ml of Liquiscint [National Diagnostics, Inc.] in a scintillation counter.

Gel electrophoresis mobility shift assays The procedures employed in my gel electrophoresis DNA-binding assays were from protocols initially described by Fried & Crothers [1981], Garner & Revzin [1981], and Strauss & Varshavsky [1984], incorporating subsequent modifications as described by Carthew, Chodosh & Sharp [1986], and Singh, Sen, Baltimore & Sharp [1985].

Analytical binding reactions Each 20 µl reaction was performed in a 1.5 ml Eppendorf tube. Reactants were added to the appropriate volume of sterile ddH_2O in the following order: 2 µl of 10X Binding Buffer [100 mM Tris-HCl, pH 7.5; 500 mM NaCl; 10 mM EDTA; 10mM DTT; 500 µg/ml BSA; 40 mM $MgCl_2$; 50% vol/vol glycerol], 0 to 10 µg of "bulk" DNA [poly(dI-dC)-poly(dI-dC) or poly(dA-dT)-poly(dA-dT); Pharmacia] and 5 to 10µg of nuclear protein were mixed together gently with a pipet tip, and this pre-incubation step proceeded at room temperature for 10 minutes. 5000-10,000 cpm of end-labelled DNA [between .2 and .6 ng] was then added, mixed as above, and the incubation continued for another 20 minutes, also at room temperature. For DNA competition experiments, unlabelled competing DNA was added to the reaction mixture at the same time as labelled probe.

Gel preparation Non-denaturing 4% polyacrylamide gels were prepared at least two hours prior to performing the binding reactions. In a Corning 50 ml sterile disposable centrifuge tube, 34.9 ml of sterile ddH_2O, 400µl of 100X Running Buffer [670 mM Tris-HCl, pH 8.0; 330 mM NaOAc; 100 mM EDTA], 4 ml of a 40% stock acrylamide solution [30:1 acrylamide:bisacrylamide ratio; Schwarz/Mann Biotech, de-ionized with Bio-Rad AG501-X8(D) resin], and 711 µl of 3% ammonium persulfate [Schwarz/Mann Biotech] were mixed together. 40 µl of TEMED [Schwarz/Mann Biotech] was then added to catalyze polymerization, and this solution was poured into a 1.5 mm thick, 14 cm wide, 16 cm long glass plate mold [Hoefer Scientific SE600 Vertical Gel Apparatus], the appropriately sized comb [15 wells for analytical gels, 3 x 3 cm wide wells for preparative gels] was inserted, and the gel was allowed to polymerize for over one hour at room temperature. For 6% gels, 6 ml of the acrylamide stock solution and 32.9

ml of sterile ddH_2O was included in the gel solution recipe. To facilitate the exclusive removal of one glass plate after electrophoresis, the surface of one plate was coated with PAM non-stick cooking spray, wiped with Kimwipes until an even layer was produced, and placed in the gel forming apparatus prior to pouring the gel solution. After electrophoresis, plates were cleaned in a tank filled with Micro Cleaning Solution [American Scientific Products].

Gels were pre-electrophoresed at 160 Volts [approximately 25 mA] at room temperature for over 90 minutes before loading the reaction mixtures. Buffer was recirculated between compartments.

Sample loading Wells were thoroughly rinsed with running buffer prior to sample loading. Total reaction mixtures were loaded with a 50 µl Unimetrics syringe pipettor [Hoefer Scientific; flushed four or more times with running buffer between sample loadings] or disposable flat sequencing gel tips [Stratagene Corp.]. 6X loading dye solution was included only in marker lanes. The DNA size markers used were either ϕX174/HaeIII or pBR322/MspI DNA [New England BioLabs], 5' end-labelled as previously described.

Electrophoresis After all samples were loaded, the gel was run at 300 Volts for five minutes to enter all the material into the gel quickly, and then decreased to 160 Volts [10 Volts/cm] for the appropriate length of time to achieve suitable separation of bands--usually one and one-half to three hours for 4% gels, and four hours for 6% gels. The XCFF dye comigrates with 400 bp DNA in a 4% gel and 230 bp DNA in a 6% gel. After removal of the PAM-coated plate, the gel was transferred to Whatman 3mm paper, dried at 80°C under vacuum for one hour [BioRad Slab Gel Dryer model 483], and

autoradiographed with Kodak X-OMAT AR film [at -70°C, in the presence of DuPont Cronex Intensifying screens] overnight.

Initial titrations Independent titrations of $MgCl_2$ [0 to 10 mM], "bulk" DNA [0 to 10 µg of poly (dI-dC):poly (dI-dC) or poly (dA-dT):poly (dA-dT)], and nuclear extract [2 to 20 µg], were necessary to optimize both the magnitude and appearance of each binding complex.

Methylation interference assays

DNA methylation with dimethyl sulphate Following guidelines given by Maxam and Gilbert [1980], 0.2 to 1.0 x 10^6 cpm of single end-labelled DNA was methylated with 1 µl of dimethyl sulfate [DMS; Fluka Chemical] in 200 µl of DMS Buffer [50 mM Na-cacodylate; 1 mM EDTA, pH 8.0] for three minutes at 20°C. 50 µl of DMS Stop Solution [1.5 M NaOAc, pH 7.0; 1 M 2-mercaptoethanol; 200 µg/ml poly (dI-dC):poly (dI-dC)] was then added, and the DNA was precipitated with 750 µl of 95% EtOH on powdered dry ice for five minutes. The precipitate was pelleted in an Eppendorf fixed-angle microcentrifuge spun for fifteen minutes, resuspended in 250 µl of 0.3 M NaOAc, pH 5.2, and re-precipitated with 750 µl of 95% EtOH. The pellet was precipitated as above, washed twice with 75% EtOH, vacuum dried and resuspended in 11 µl of TE. 1 µl was counted in aqueous solvent.

Preparative binding reactions and gel electrophoresis Preparative binding reactions for methylation interference experiments were carried out with a ten-fold scale up of predetermined optimal analytical reaction conditions [to give a

binding reaction volume of 200 µl], containing 50,000 to 100,000 cpm of single end-labelled methylated DNA. Preparative gel electrophoresis proceeded exactly as described for analytical gel electrophoresis, except that each binding reaction was loaded into a 3.5 cm wide preparative well.

Elution of bands from preparative gels After preparative electrophoresis, the wet gel was wrapped in plastic wrap [while still attached to one glass plate], and autoradiographed at 4°C overnight to Kodak X-OMAT AR film. Bands of interest were localized by superimposing the gel on top of its autoradiograph, and the bands were excised from the gel with a scalpel blade. The gel slice was crushed on a clean glass plate with the side of the scalpel blade, and placed in 3.5 ml of Elution Buffer [10 mM Tris-HCl, pH 8.0; 1 mM EDTA; 300 mM NaCl; 0.2% wt/vol SDS; 1 µg/ml poly dI-dC; 10 µg/ml Proteinase K] in a Falcon 2059 tube, and eluted by shaking at 275 rpm at 37°C overnight. Gel pieces were pelleted by centrifugation in an HB-4 rotor spun at 8000 rpm for 15 minutes at >15°C. The supernatant was passed through a 5 ml disposable plastic pipette [stoppered with its cotton plug] into a Falcon 2059 tube. Gel pieces were washed with 0.5 ml of Elution Buffer, which was then combined with the above supernatant. Sequential extractions with equal volumes of phenol, phenol:chloroform:isoamyl alcohol, chloroform:isoamyl alcohol were performed, 10 µg of carrier DNA [poly (dI-dC):poly (dI-dC)] was added, and the DNA precipitated with 8.5 ml of 95% EtOH at -20°C overnight. The precipitate was pelleted in an HB-4 rotor at 8000 rpm for 45 minutes at 4°C, washed with 75% EtOH and respun, vacuum dried and resuspended in 3 ml of NACS Buffer A [10 mM Tris, pH 7.5; 0.1 mM EDTA; 200 mM NaCl]. This solution was then passed over a NACS column [Bethesda Research Labs; hydrated and equilibrated

as per manufacturer's instructions], washed with 5 ml of NACS Buffer A, and eluted with 2 x 150 µl NACS Buffer B [10 mM Tris-HCl, pH 7.5; 0.1 mM EDTA; 1000 mM NaCl] into a 1.5 ml Eppendorf tube, and precipitated overnight at-20°C with 1 ml of 95% EtOH. Incorporation of NACS column purification greatly improved band resolution in sequencing gels.

Tubes were spun in an Eppendorf microcentrifuge for 45 to 60 minutes, and the pellet was washed with 1 ml of 75% EtOH and respun. The 75% EtOH wash and spin was repeated twice more, and the pellet vacuum dried.

Methylated DNAs were cleaved in 100 µl of freshly prepared 1M piperidine [Fluka Chemical] at 90°C for 40 minutes. Tubes were then placed on wet ice for two minutes, quick-spun, frozen on finely powdered dry ice for five minutes, and lyophilized. DNAs were redissolved in 50 µl of sterile water, frozen and lyophilized. Two more rounds of lyophilization, after resuspension in 10 µl of water, were performed. The samples were resuspended in 3 µl of 2mM Tris-0.05 mM EDTA (pH 7.5). 6 µl of Formamide-EDTA-Dye Solution [100 µl consisted of 88 µl of deionized formamide; 10 µl of 2% (w/v) xylene cyanol FF and bromphenol blue; 2 µl of 0.5 M EDTA, pH 8.0] was mixed in, and 2 µl of this solution was counted in aqueous fluour. Samples were now ready for loading onto sequencing gels.

8% Polyacrylamide:7M urea sequencing gels All gels were run using the IBI STS45 Apparatus [International Biotechnologies Inc.] with 0.2mm wedge spacers. Gels were cast the day prior to electrophoresis.

The gel solution was prepared by dissolving 90 gm urea [Schwarz/Mann Biotech, ultrapure grade] into a solution of 50.7 ml water and 36 ml of 40% acrylamide stock solution [38:2 w/w acrylamide:bisacrylamide, Schwarz/Mann

Biotech, ultrapure grade]. Approximately two grams of Bio-Rad AG501-X8(D) mixed-bed resin was added, the stirring continued for ten minutes, and the solution was filtered through a Nalgene 0.45µm filter. To 130.2 ml of filtrate, 15 ml of 10 x TBE [one liter contained 110 gm Tris base; 55 gm boric acid; 50 ml 0.5 M EDTA, pH 8.0] and 4.8 ml 1.6% ammonium persulfate was added. This mixture was divided into three 50 ml aliquots in Corning disposable 50 ml centrifuge tubes. To one aliquot, 40 µl of TEMED was mixed in, and this solution was drawn into a 60 ml syringe and poured through a 20 gauge needle into the gel mold [plates were silanized according to Maniatis et al, 1982]. TEMED was added to the next two aliquots when more gel solution was needed to eventually fill the mold. A 32 well comb was inserted, and the gel was allowed to polymerize overnight.

After mounting the gel in the STS45 Apparatus, the chambers were filled with 1 x TBE [89 mM Tris base, 89 mM boric acid, 2.5 mM EDTA; pH 8.0] as running buffer. Wells were rinsed clean of unpolymerized acrylamide, and the gel was pre-electrophoresed at a constant 100 Watts [approximately 2500 Volts] until the surface temperature reached 50-55°C, usually within 60 to 90 minutes. Tubes containing DNA samples [with equivalent concentrations of radioactivity] were denatured at 90°C for three minutes, chilled on wet ice for two minutes, and quick-spun in the microcentrifuge. Wells were rinsed again. Two µl of each sample was loaded into each well via a flattened sequencing gel disposable pipet tip [Stratagene Corp.] attached to a 20 µl Pipetman. Electrophoresis was begun at 100 Watts, and continued until a suitable separation was achieved, noting that the bromphenol blue and the xylene cyanol FF dyes comigrate with ≈15nt and ≈60 nt single-stranded DNA, respectively. The surface temperature was kept between 50-55°C for the duration of electrophoresis.

After electrophoresis, one plate was removed, and the gel was fixed in three liters of 10% methanol-10% glacial acetic acid for twenty minutes. The fixative solution was removed, the gel was transferred to Whatman 3mm chromatography paper, vacuum dried at 80°C in the Bio-Rad Slab Gel Dryer for one hour, and autoradiographed by exposure to Kodak X-OMAT AR film in the presence of DuPont Cronex intensifying screens at -70°C. Typically, for 1000 cpm of radioactivity loaded per lane, exposure times of two to five days was sufficient to visualize all bands.

Results

Published reports have repeatedly demonstrated the dissimilar nature of the regulation of transcription of the two major HBV mRNAs. HBV DNA-transfected cell lines regularly produced high quantities of 2.1kb mRNA [regardless of species or tissue of origin], while substantial quantities of 3.5kb mRNA was evident only in cells capable of supporting viral replication--human liver cells [Standring & Rutter, 1986 and references within]. Transcription of 3.5kb mRNA is liver cell-specific, and I have performed experiments examining the regulatory cis- and trans-acting factors governing the tissue-specific expression of this transcript.

Cis- and trans-acting factors mediating the tissue-specific and general expression of many cellular and viral genes have been studied with the recently developed CAT transient expression assay [Gorman et al, 1982]. Recombinant plasmids containing the eukaryotic promoter element under investigation inserted directly upstream of the bacterial chloramphenicol acetyltransferase [CAT] gene, are transfected into cell lines of interest. 48 hours after transfection, crude freeze-thaw lysates are made from the transfected plates, and CAT enzyme activity in these lysates is measured. Measured CAT enzyme activity is proportional to the amount of steady-state CAT mRNA, which is primarily dependent on the rate of initiation of transcription. Furthermore, the rate of initiation of transcription from these recombinant plasmids is governed by the eukaryotic DNA sequences inserted upstream of the CAT gene, and their interactions with trans-acting factors. Similarly, appropriate constructs can be made to study the role of enhancers and inducible cis-acting sequences. In summary, the CAT transient expression assay can provide a sensitive,

reproducible, and reliable account of the cis- and trans-acting factors involved in the regulation of eukaryotic gene expression.

Tissue specificity of pEcoAluCAT expression

The plasmid pEcoAluCAT was constructed to study the cis and trans-acting factors regulating 3.5kb mRNA expression. A DNA restriction fragment containing HBV sequences 1 to 1878 was inserted directly upstream of the CAT gene of pSVoCAT [see Figure 2]. This EcoRI-AluI fragment of HBV included the enhancer and core promoter regions [containing the start site of 3.5kb mRNA transcription], but excluded the HBcAg ATG at nt 1901, the polyadenylation signal at nt 1934, and the sequences controlling 2.1kb mRNA expression [i.e., the ≈200 nt region upstream of the EcoRI site].

As an initial test of the tissue- and species-specificity of pEcoAluCAT expression, the plasmid was transfected into various cell lines, including: NIH 3T3 [mouse fibroblast], Rat 6 [rat fibroblast], CV-1 [monkey kidney], HepG2 [human hepatoma], and 2.2.15 [a HepG2 derived cell line harboring replicating HBV; Sells et al, 1987]. If pEcoAluCAT contained all the regulatory elements to faithfully mimic the observed liver-specific pattern of 3.5kb mRNA expression, then CAT activity would be significantly higher in HepG2 and 2.2.15 cells than in the other cell lines. Lysates from pEcoAluCAT-transfected HepG2 and 2.2.15 cells contained high CAT enzyme activities, while lysates from the other transfected cells produced very little CAT [data not shown]. In addition, ongoing HBV replication had little effect on pEcoAluCAT expression, since HepG2 and 2.2.15 cell lysates contained equally abundant quantities of CAT activity. Thus, transfection of pEcoAluCAT into HepG2 cells provided an

acceptable system to study the factors controlling the tissue-specificity of 3.5kb mRNA expression. Further studies were undertaken to determine the role of trans-acting factors in the expression of pEcoAluCAT in HepG2 cells.

pEcoAluCAT expression in HepG2 cells is dependent on limited concentrations of positively-acting trans-acting factors

The action of many cis-acting transcriptional regulatory elements appears to be mediated primarily through interactions with trans-acting nuclear factors, and many of these factors are sequence-specific DNA-binding proteins [Dynan & Tjian, 1985]. Trans-acting, liver-specific factors that interact with sequences within the HBV genome have been proposed to account for the liver-specificity of 3.5kb mRNA transcription [Jameel & Siddiqui, 1986]. In order to determine whether pEcoAluCAT expression in HepG2 cells was dependent on such interactions, **in vivo** competition co-transfection experiments were performed [Schöler & Gruss, 1984]. Prior to these experiments, a titration curve of pEcoAluCAT expression in HepG2 cells determined the linear range of CAT activity to be between 0.5 and 10 µg of pEcoAluCAT transfected per dish [Figure 4 and Table 1].

Table 1

Titration of pEcoAluCAT activity in HepG2 cells

pEcoAluCAT (µg per dish)	Relative CAT activity
0.25	20%
0.5	37%
1.0	22%
2.5	100%
10.0	110%

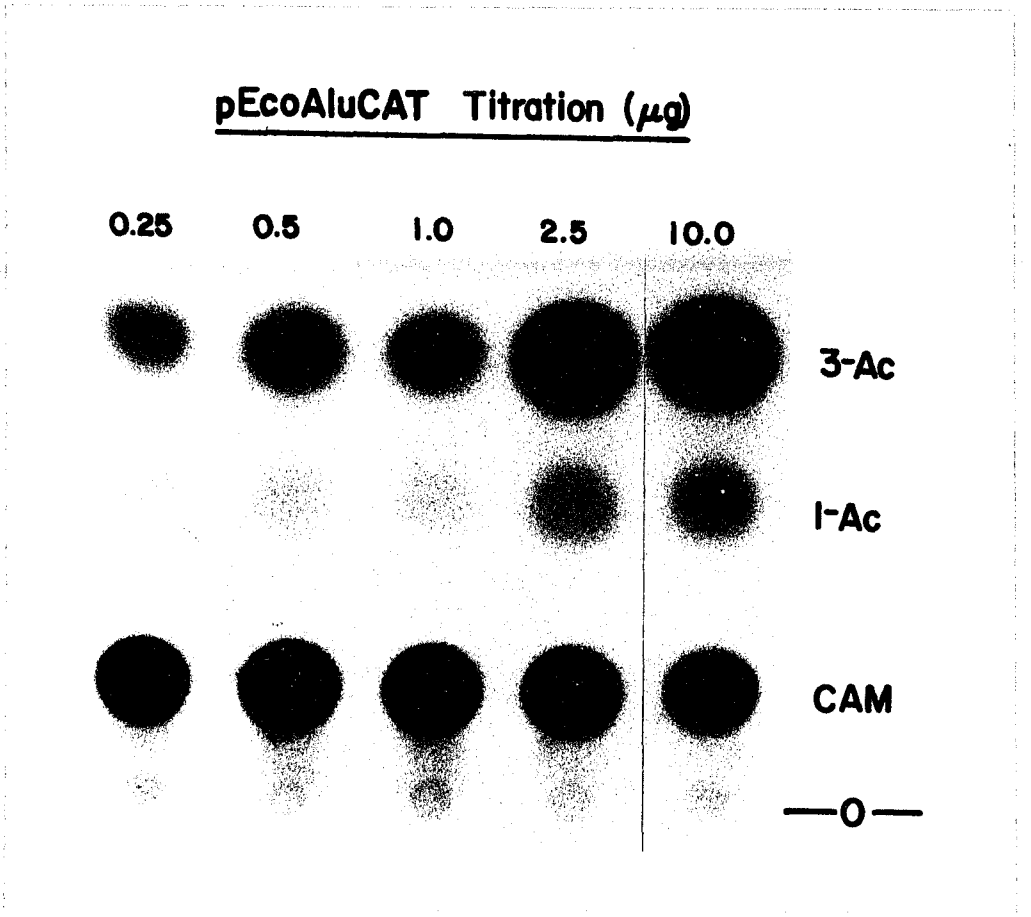

Figure 4 Titration of pEcoAluCAT expression in HepG2 cells. A total of 30 μg of DNA was transfected per plate, consisting of the indicated amounts of pEcoAluCAT plus the appropriate amount of pIBI31 as carrier. -O-, CAM, 1-Ac, and 3-Ac represent the origin, unacetylated ^{14}C-chloramphenicol, 1-acetate chloramphenicol and 3-acetate chloramphenicol, respectively. Normalized CAT activities are given in Table 2.

If sequence-specific DNA binding proteins were required for pEcoAluCAT expression in HepG2 cells, and present in limited concentrations, then inclusion of increasing amounts of specific competitor HBV DNA in the transfection cocktail would be coupled with a decrease in pEcoAluCAT activity. A non-saturating amount of pEcoAluCAT [2.5μg] was co-transfected with plasmids containing the entire HBV genome [pH-2], the enhancer region [pENH-2], or the core promoter region [pCP-1] [Figure 5 and Table 2]. Optimal DNA uptake in HepG2 cells had previously been determined to be between 20-30 μg per 10^6 cells/10 cm dish [data not shown], using pIBI31 as carrier.

Table 2

Competition co-transfection experiments with pEcoAluCAT

Competitor plasmid		Relative CAT activity
None		100%
pH-2:	5 μg (2:1)	50%
	15 μg (6:1)	23%
	27.5 μg (11:1)	12%
pENH-2:	5 μg (4:1)	65%
	15 μg (12:1)	19%
	27.5 μg (22:1)	6%
pCP-1:	27.5 μg (22:1)	64%

[Parentheses contain molar ratios of competitor:pEcoAluCAT plasmids]

These data strongly suggested that HepG2 cells contained titratable, trans-acting, positive regulators of pEcoAluCAT expression, and that at least some of these factors were DNA-binding proteins, recognizing sequences within the enhancer and core promoter regions. There was not a significant difference between competitions with the enhancer fragment and the total HBV genome,

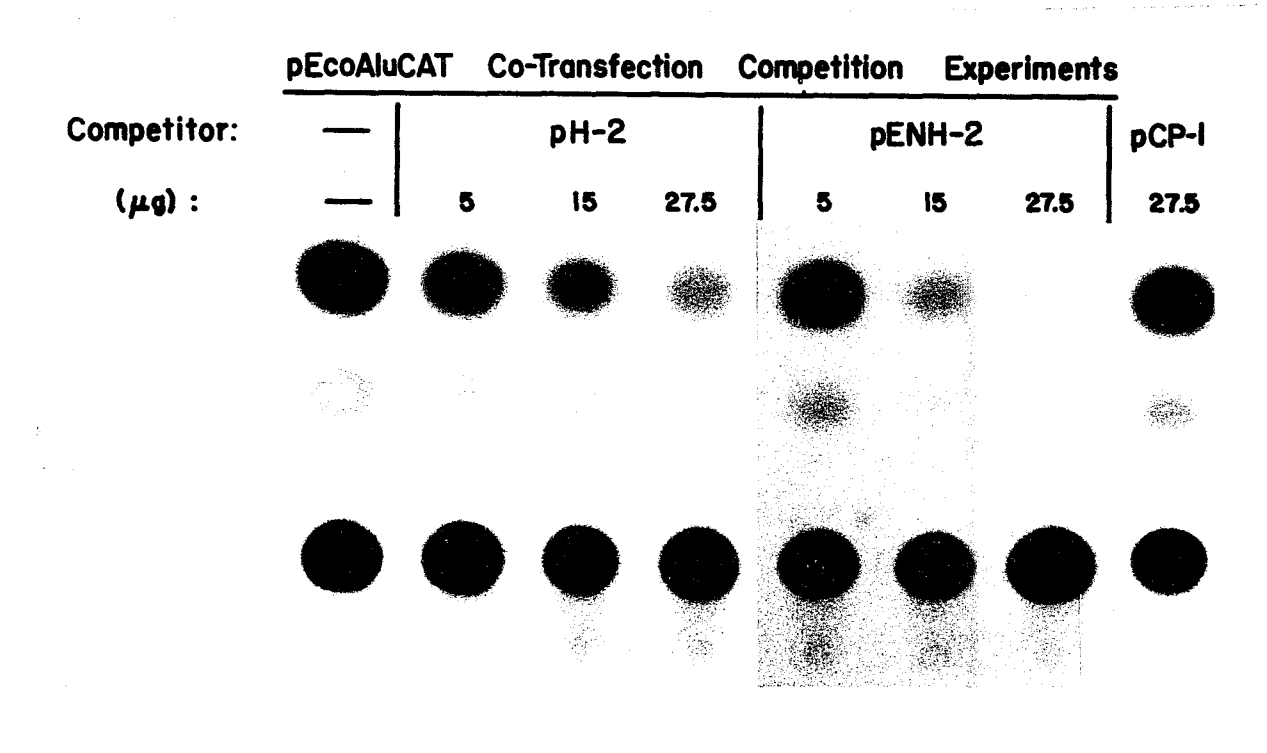

Figure 5 pEcoAluCAT **in vivo** co-transfection competition experiments. Various concentrations of competing plasmids [described in the Methods section] were co-transfected with 2.5 μg of pEcoAluCAT, and the appropriate amount of carrier pIBI31, for a total of 30 μg of DNA per dish. Normalized CAT activities are presented in Table 3.

indicating that this 295 bp contained most of the cis-acting elements driving pEcoAluCAT expression, consistent with the results of Jameel & Siddiqui [1986], who used a slightly different plasmid construct.

Plasmids containing strongly active enhancers make the detection of promoter-binding proteins, using **in vivo** competition CAT assays, more difficult; especially if the promoter is weakly functional on its own [Schöler & Gruss, 1984]. This seemed to be the case with the HBV core promoter, shown in a different construction to be virtually inactive without its accompanying enhancer, by Roosnick et al [1986]. A 22:1 molar ratio of core promoter DNA [in the plasmid pCP-1] to pEcoAluCAT decreased CAT activity ≈40%, suggesting the presence of HepG2 nuclear protein(s) that bind to this region. This amount of competition, although relatively little, is reproducible and significant, and can be explained if HepG2 cells contained large amounts of this binding protein(s).

Identification of HepG2 nuclear factor binding within the core promoter using gel mobility shift assays

Gel mobility shift assays can demonstrate whether or not an extract contains such proteins. The assay rests on the property that a specific protein-DNA complex ["bound" DNA] can be resolved from uncomplexed ["free"] DNA by migration through a polyacrylamide matrix. The advantages of this approach over other **in vitro** assays that detect DNA-binding proteins include:

1.) crude, unfractionated extracts can be used;

2.) low-abundance proteins can readily be detected;

3.) low-affinity binding complexes are eliminated since "bulk", non-specific competitor DNA [e.g. poly (dI-dC):poly (dI-dC) or poly (dA-dT):poly (dA-dT)] is included in the binding reaction;

4.) only proteins with an affinity for sequences on the radiolabelled DNA of greater than 10^4 will enter the gel as a DNA-protein complex; and,

5.) results can usually be seen with an overnight exposure.

Using this assay, one slower migrating band appeared when radiolabelled core promoter restriction fragment [CP;nt 1645-1888] was incubated with crude HepG2 nuclear extract [Figure 6, Lane 2]. No slower migrating band appeared when assays were performed with Proteinase K-treated extract, or when 0.1% SDS was included in the binding reaction, implicating a DNA-protein complex as the slower migrating band [data not shown]. These data strongly suggested the presence of a CP DNA-specific binding protein in HepG2 nuclei. Additional experiments were undertaken to verify this finding and explore some characteristics of the protein and its interaction with CP DNA.

Thermal stability of the CP binding protein

Results of a thermal stability assay are presented in Table 3. The CP binding protein was relatively heat-labile, with all detectable binding activity destroyed by heating the extract to 37°C for ten minutes. This was probably not due to non-specific activation of proteases, since this same 37°C-treated extract still contained HBV enhancer binding activity [Figure 11].

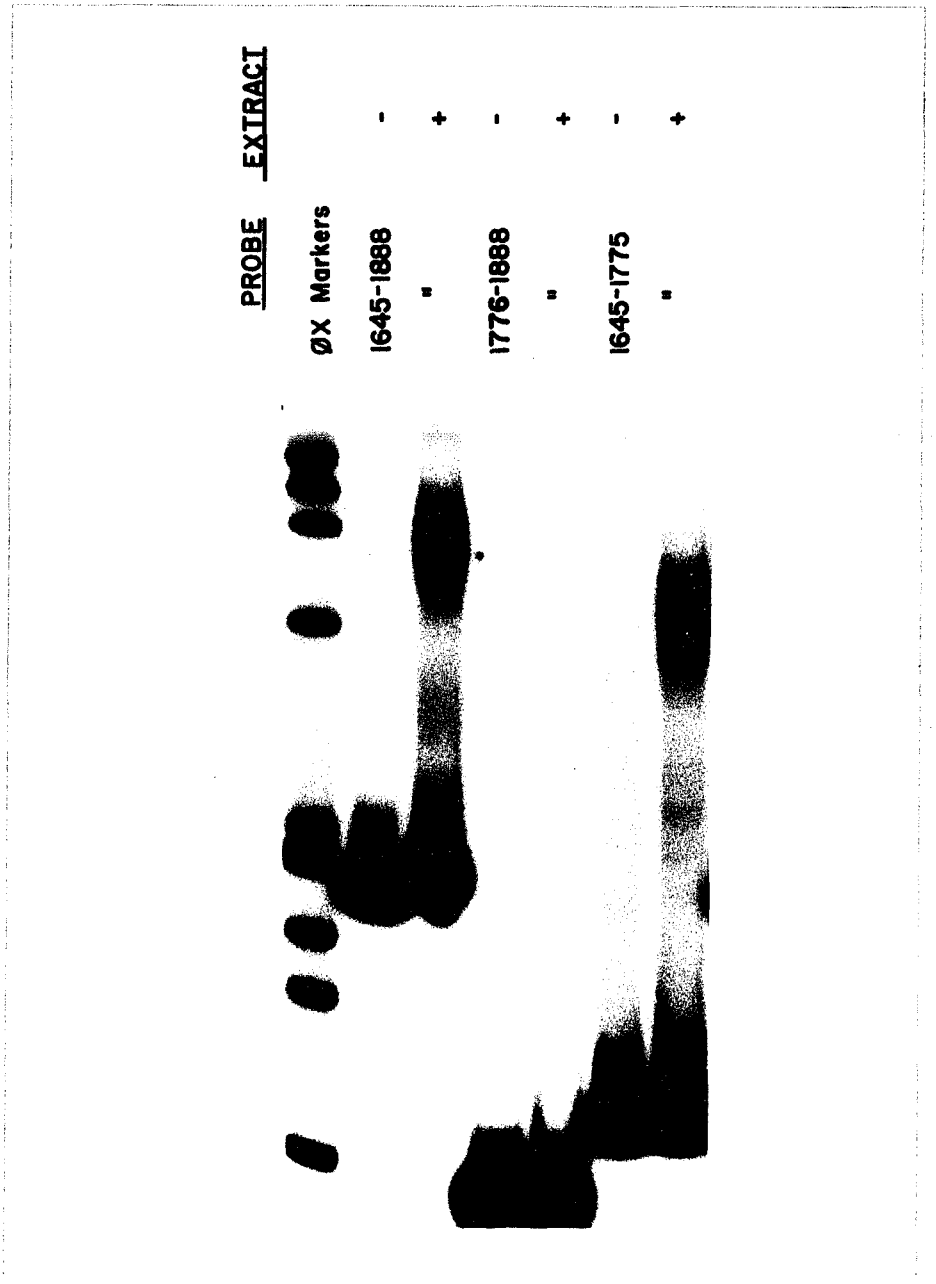

Figure 6 Gel electrophoresis mobility shift assays with the core promoter. Indicated single-end labelled DNAs were incubated in the absence [-] or presence [+] of 10μg of HepG2 nuclear extract. Each 20 μl reaction mixture contained 2 μg of poly (dA-dT):poly (dA-dT) as bulk competitor. Samples, including kinated ϕX174/HaeIII size markers, were loaded onto a 6% polyacrylamide gel and run for two hours at 10 Volts/cm.

Table 3

Thermal stability of the CP binding protein

Extract pre-treatment temperature	Binding activity
0 °C	+
23 °C	+
37 °C	-
68 °C	-
95 °C	-

The CP binding protein contains a CP sequence-specific binding domain

To determine whether the CP binding protein specifically recognized the CP region of HBV, or was a general DNA binding protein, appropriate competition experiments were performed. If the binding protein specifically recognizes a sequence in CP DNA, then the appearance of the shifted band would be titrated away with excess cold CP DNA, but not with various other non-homologous DNAs. Inclusion of only 6 ng of cold CP DNA in the binding mixture dramatically decreased the magnitude of the bound complex, while 100ng of cold viral promoter/enhancer DNAs [SV40 promoter-enhancer, HBV enhancer, or HBsAg promoter] barely competed for binding to the protein [Figure 7; also, the adenovirus major late promoter did not compete for binding (data not shown)]. 100ng of specific CP competitor DNA in the binding reaction eliminated the appearance of the bound complex. These competition experiments established that at least one HepG2 nuclear factor [referred to as the CP binding protein] bound to specific sequences within the HBV core promoter with a much higher affinity than to other selected DNA sequences.

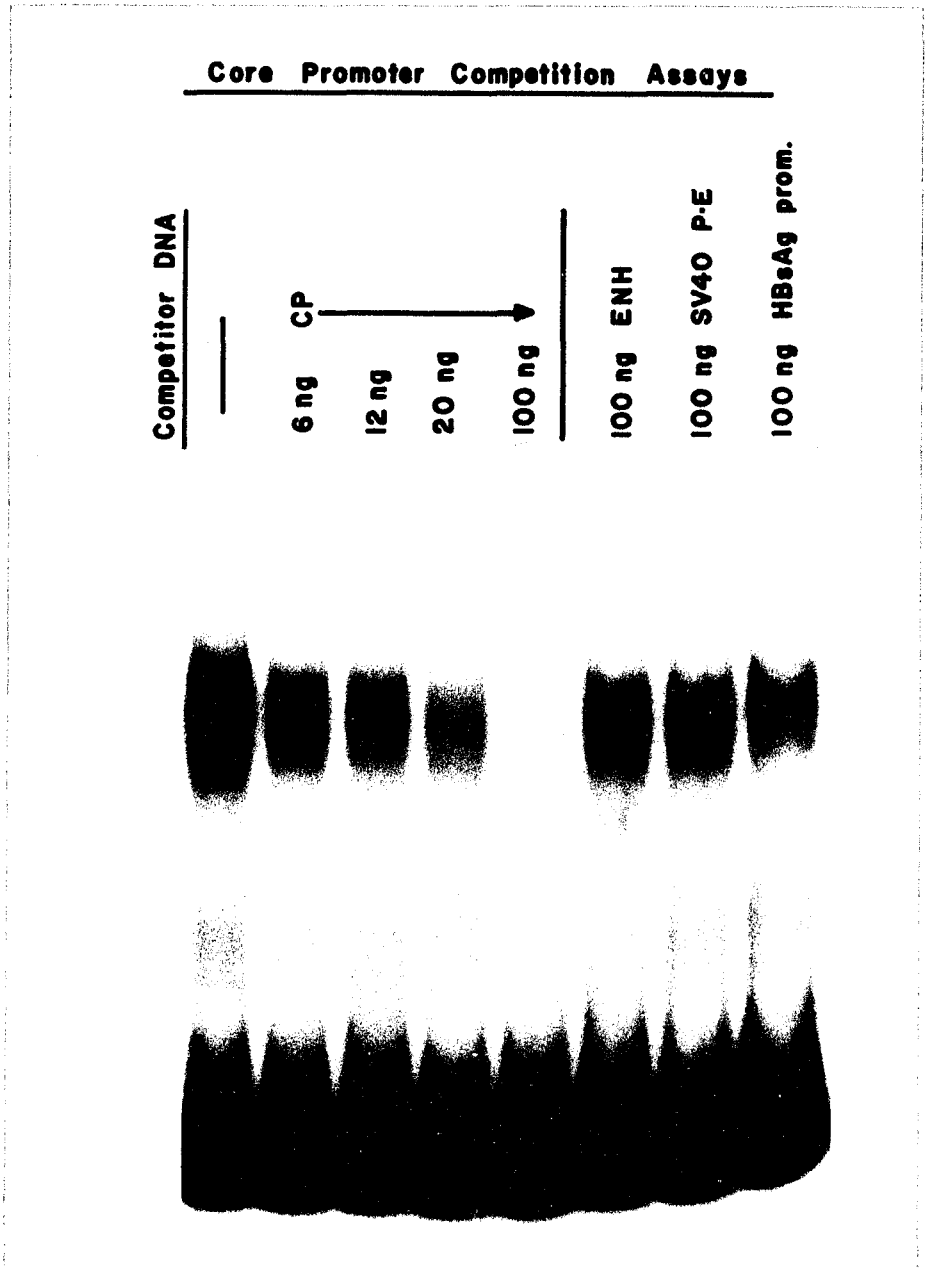

Figure 7 DNA competition experiments with labelled core promoter. Competing DNAs: CP [core promoter, HBV nt 1645-1888], ENH [HBV enhancer, HBV nt 1082-1378], SV40 P-E [HindIII-PvuII SV40 early promoter-enhancer fragment, SV40 nt 5171-(5243/1)-270], HBsAg prom. [HBsAg promoter, BamHI-EcoRI fragment (HBV nt 2906-3182)]. Conditions as for Figure 6, but gel was run for four hours.

Tissue distribution of the CP binding protein

If the CP binding protein is ubiquitously distributed, [e.g., the transcription factor Sp1, found in virtually all cell extracts analyzed; Dynan & Tjian, 1985; Briggs et al, 1987], then mobility shift assays performed with extracts from various cell types would demonstrate shift patterns similar to that seen with the HepG2 nuclear extract. On the other hand, if the CP binding protein is restricted in its expression [e.g., to human or liver-derived cells], then assays performed with extracts from other cells would reveal either absent or non-specific shift patterns. Gel mobility shift assays were then performed with HepG2, HeLa [human cervical carcinoma], C127 [mouse fibroblast], and NIH 3T3 [mouse fibroblast] nuclear extracts [Figure 8]. Interestingly, each extract contained proteins that bound CP DNA, however these DNA-protein complexes did not migrate with the same mobility as the one seen with HepG2 nuclear extract. Furthermore, competition experiments demonstrated that the bound complexes in HeLa, C127, and 3T3 diminished in intensity regardless of cold competitor DNA, whereas only specific CP DNA competed for binding of the HepG2 binding protein. All together, these data indicated that the CP binding protein in HepG2 nuclei:

1.) was not a general DNA binding protein;

2.) possessed a sequence-specific DNA binding domain;

3.) was limited in its distribution among various tissues, and appeared to be liver cell-specific.

Whether the HepG2-derived CP binding protein was the same titratable factor suggested by the previously described pEcoAluCAT **In vivo** CAT competition experiments remains to be seen, but a first step involved precise localization of the binding site.

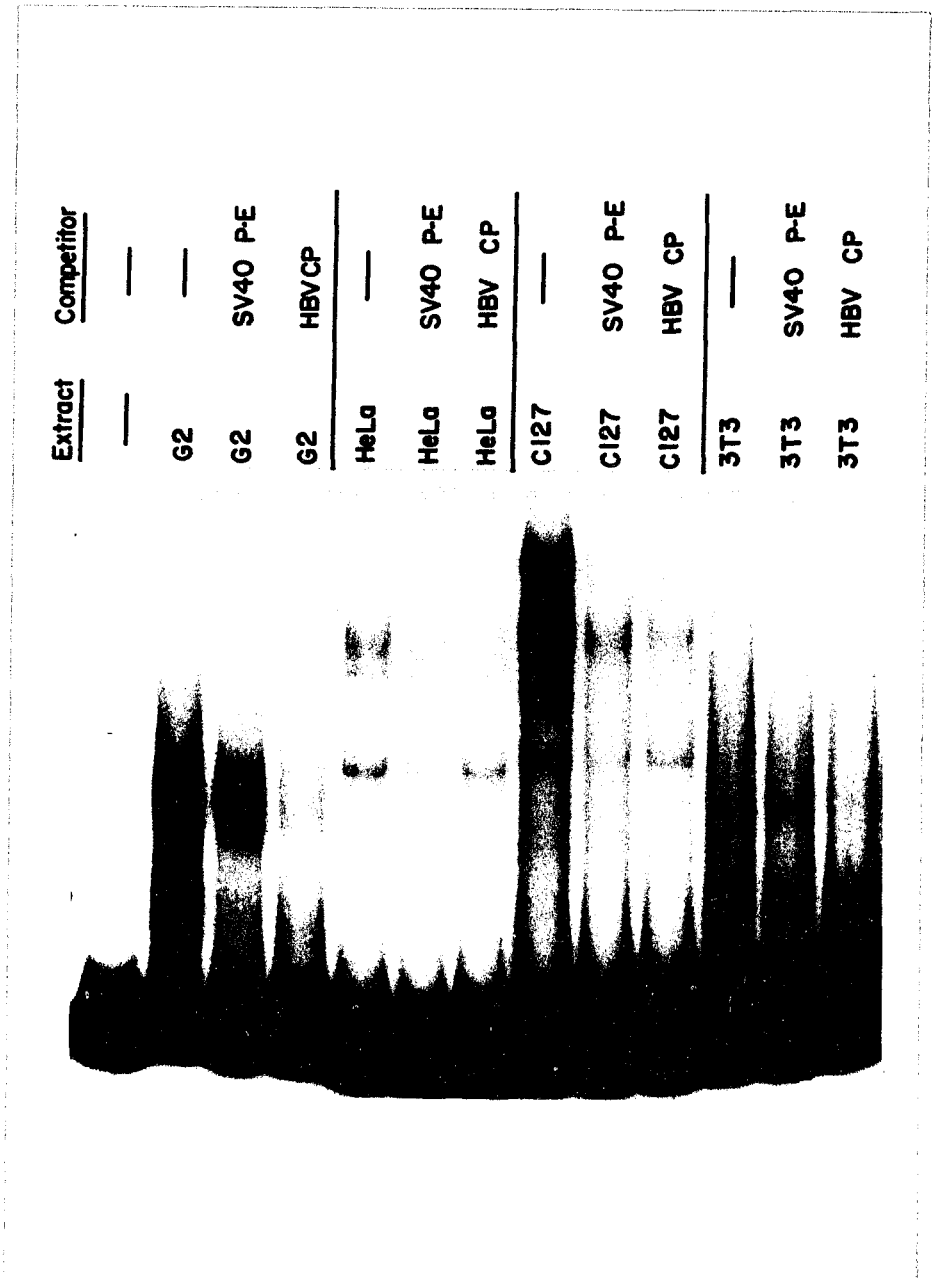

Figure 8 HepG2 cells contain a protein that specifically recognizes sequences within the core promoter. Labelled core promoter DNA was incubated with 4 μg of poly (dI-dC):poly (dI-dC) and nuclear extract from HepG2 [10 μg], HeLa [7 μg], C127 [15 μg], or NIH3T3 [15 μg] cells. 100 ng of SV40 promoter-enhancer, or 50 ng of HBV core promoter, were included as indicated. Using poly (dI-dC):poly (dI-dC) instead of poly (dA-dT):poly (dA-dT) resulted in the appearance of an artifactual bound complex with HepG2 nuclear extract [completely absent with the inclusion of SV40 DNA in the binding mixture]. This 4% polyacrylamide gel was run for two hours.

Localization of the CP binding protein recognition site

Single-end labelled CP DNA was cut at the unique MaeI site at nt 1775. Assuming that digestion at the MaeI site did not destroy binding, incubation of HepG2 nuclear extract with these smaller substrates would demonstrate which part of the CP DNA contained the binding site; either the 1645-1775 or 1776-1888 fragments. Results, presented in Figure 6, demonstrate no binding within the 1776-1888 fragment, while full binding resided within the rest of the CP DNA--i.e., between nucleotides 1645 and 1775. Predictably, use of a smaller fragment as substrate increased the mobility of the bound complex as compared to using the entire core promoter.

High resolution mapping of the CP binding site

Gel mobility shift assays demonstrated the presence of a non-ubiquitous protein present in HepG2 nuclei that specifically bound somewhere between nt 1645 and 1775 of the HBV genome. In order to map this binding site with greater precision, a procedure that resolved the binding at the nucleotide level was undertaken. The methylation interference assay [Siebenlist & Gilbert, 1980; Hendrickson & Schlief, 1985], using full length single-end labelled CP DNA, was performed.

Dimethyl sulfate methylates the N-7 of guanine residues [Maxam & Gilbert,1980]. This methyl group extends into the major groove of double stranded DNA, which can interfere with DNA-protein interactions. Methylated guanine residues that prohibit protein binding can be identified with the methylation interference assay. These guanine residues are then presumed as important for factor recognition of the binding site. Partially methylated, single-end labelled CP DNA [a randomly modified, non-homogenous substrate]

was mixed with HepG2 nuclear extract in a ten-fold scale-up of optimal analytical binding reaction conditions. Bound and free complexes were separated by 6% polyacrylamide gel electrophoresis, and the DNA in each complex was localized, eluted, purified, and cleaved at methylated G residues with piperidine, then run on a sequencing gel alongside appropriate markers, including piperidine-cleaved substrate DNA. Bands corresponding to methylated G residues that interfered with binding of the CP binding protein would be significantly under-represented in the lane loaded with DNA extracted from the Bound complex, in comparison to the lane loaded with Free DNA.

The results of such an experiment [using **short** strand end-labelled methylated CP DNA as binding substrate] are shown in Figure 9. All binding sites within the 1645-1888 core promoter restriction fragment would be revealed, [since the entire region is represented in this Figure]. Five bands, corresponding to methylated guanine residues that prohibited binding of the factor, were absent in the Bound lane. Methylation of any guanine between nt 1660-1671 of the **short** strand, eliminated factor binding, thus implicating this region within the core promoter as the recognition site for CP binding protein. A repeat of this assay with an appropriately labelled probe demonstrated which methylated G residues of the **long** strand interfered with factor binding; those between nt 1665-1680 [data not shown]. All together, the binding site is minimally between nt 1660-1680, and by extension to the nearest non-interfering guanine residue, spans nt 1655-1684. Schematically, these sequences are:

CORE PROMOTER BINDING SITE: [Nucleotides 1651 to 1690 displayed]:

```
...5′-cttacataaG aGGactcttG Gactctcagc aatgtcaacg-3′... Short
...3′-gaatgtattc tcctGaGaac ctGaGaGtcG ttacagttgc-5′... Long
```

[Capital Gs represent methylated guanine residues that interfered with factor binding.]

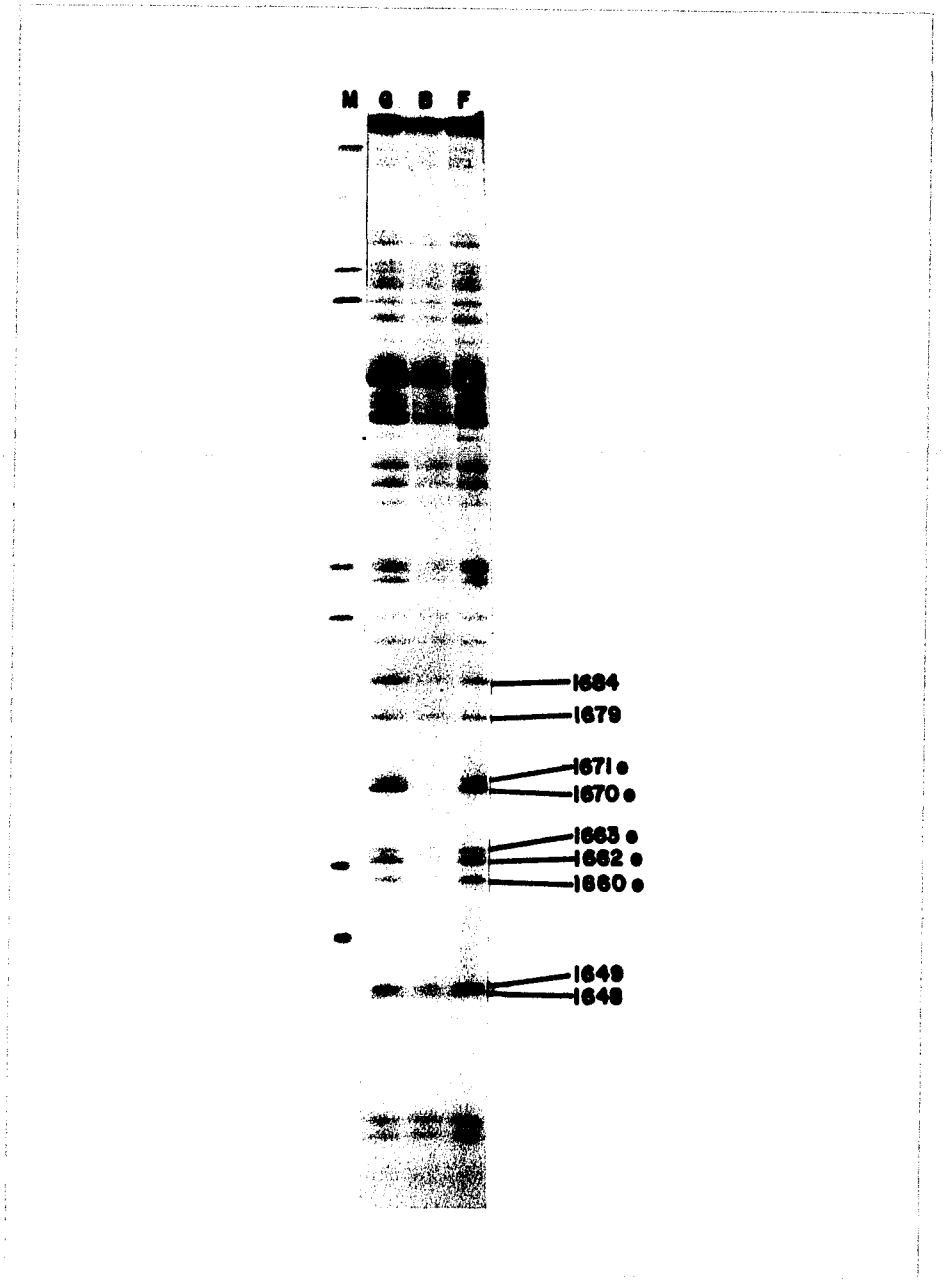

Figure 9 Methylation interference mapping of the core promoter binding site. An 8% polyacrylamide:7 M urea, 0.2 mm wedged, sequencing gel was loaded with pBR322/MspI markers [**M**], piperidine-cleaved probe to provide a guanine ladder [G], and DNA eluted and piperidine-cleaved from the bound [**B**] and free [F] complexes. Relevant **short** strand guanine nucleotides are indicated, with fully protected residues marked with filled ovals. Probe: BamHI-EcoRI fragment of pCP-1, labelled at the EcoRI site.

Multiple nuclear proteins bind within the HBV enhancer region

The competition CAT co-transfection experiments shown in Figure 5 provided evidence for the functional importance of the enhancer [ENH] region for pEcoAluCAT expression in HepG2 cells. Moreover, these results strongly implicated titratable trans-acting factor(s), that bind in the ENH region, as principal mediators of enhancer function. Mobility shift assays were undertaken to provide **in vitro** corroboration of the presence of such factors.

Mobility shift assays with end-labelled ENH DNA [nt 1082-1378, DdeI-StyI restriction fragment] and HepG2 nuclear extract were performed. Two easily distinguishable bound complexes were present, with the upper one consisting of two closely migrating bands [Figure 10]. Assays performed with Proteinase K-treated extract, or in the presence of 0.1% SDS, destroyed the appearance of all shifted bands indicating, as with the CP experiments, that the slower migrating bands were due to protein-DNA complexes [data not shown].

Thermal stability of the enhancer binding proteins

Heat-treated extracts [performed in parallel with the experiments shown in Figure 5 for the CP binding protein], were incubated with labelled ENH DNA [Figure 11]. Interestingly, the proteins responsible for both bound complexes were relatively heat-stable; incubation with 68°C-treated extract destroyed the the appearance of the upper band, while the lower band was present with 95°C-treated extract. Thus, on the basis of heat-stability and rate of migration through polyacrylamide, the proteins responsible for the ENH shift patterns were different not only from each other, but also from the protein(s) involved in forming the CP binding complex.

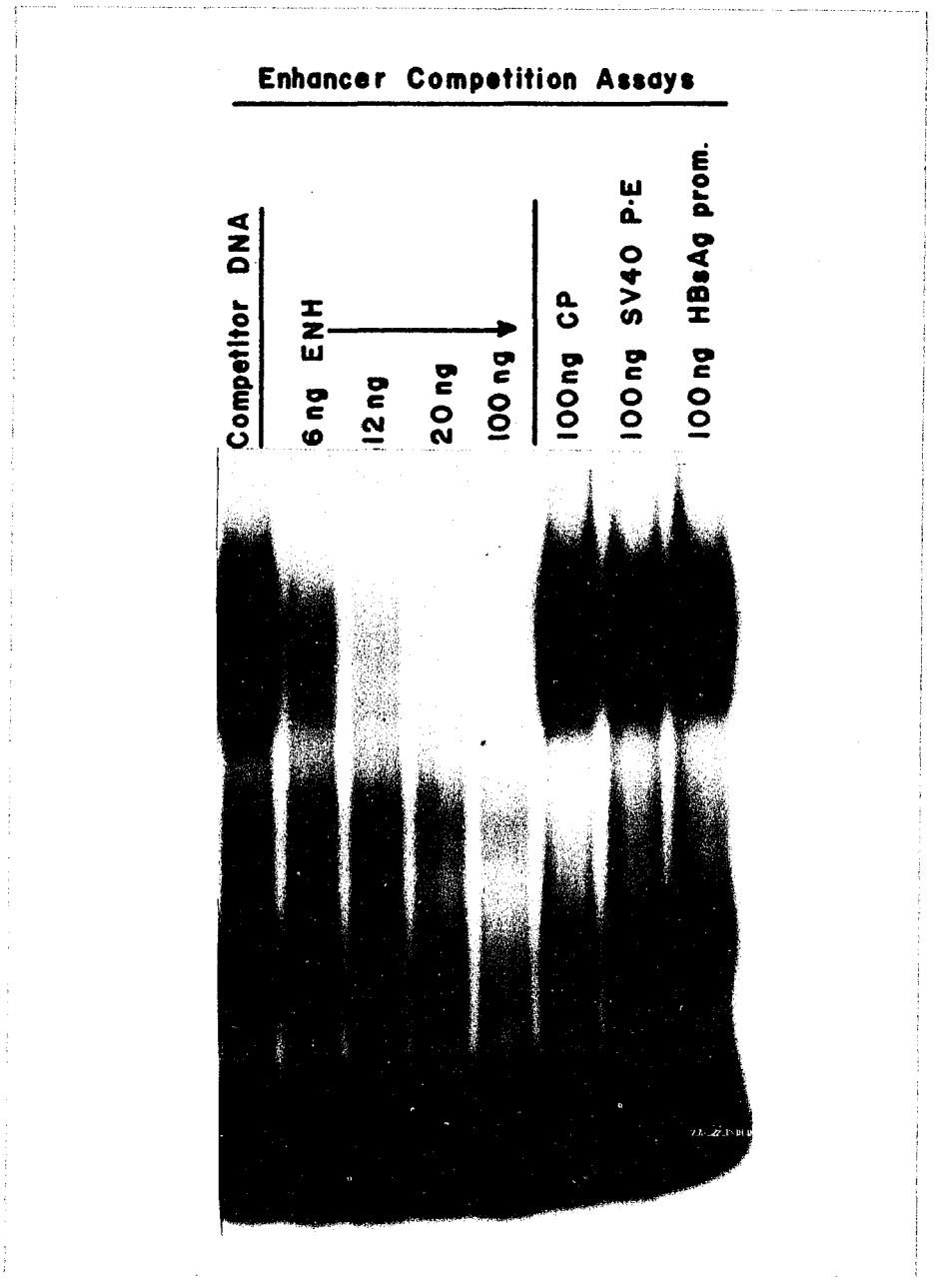

Figure 10 Competition for HBV enhancer DNA binding. An end-labelled fragment [approximately 0.5 ng], spanning HBV nt 1082-1378 in the HBV enhancer region, was incubated with 10 µg of HepG2 nuclear extract and 8 µg of poly (dI-dC):poly (dI-dC) with various competitor DNAs [as outlined in Figure 7]. Resolved on a 6% polyacrylamide gel.

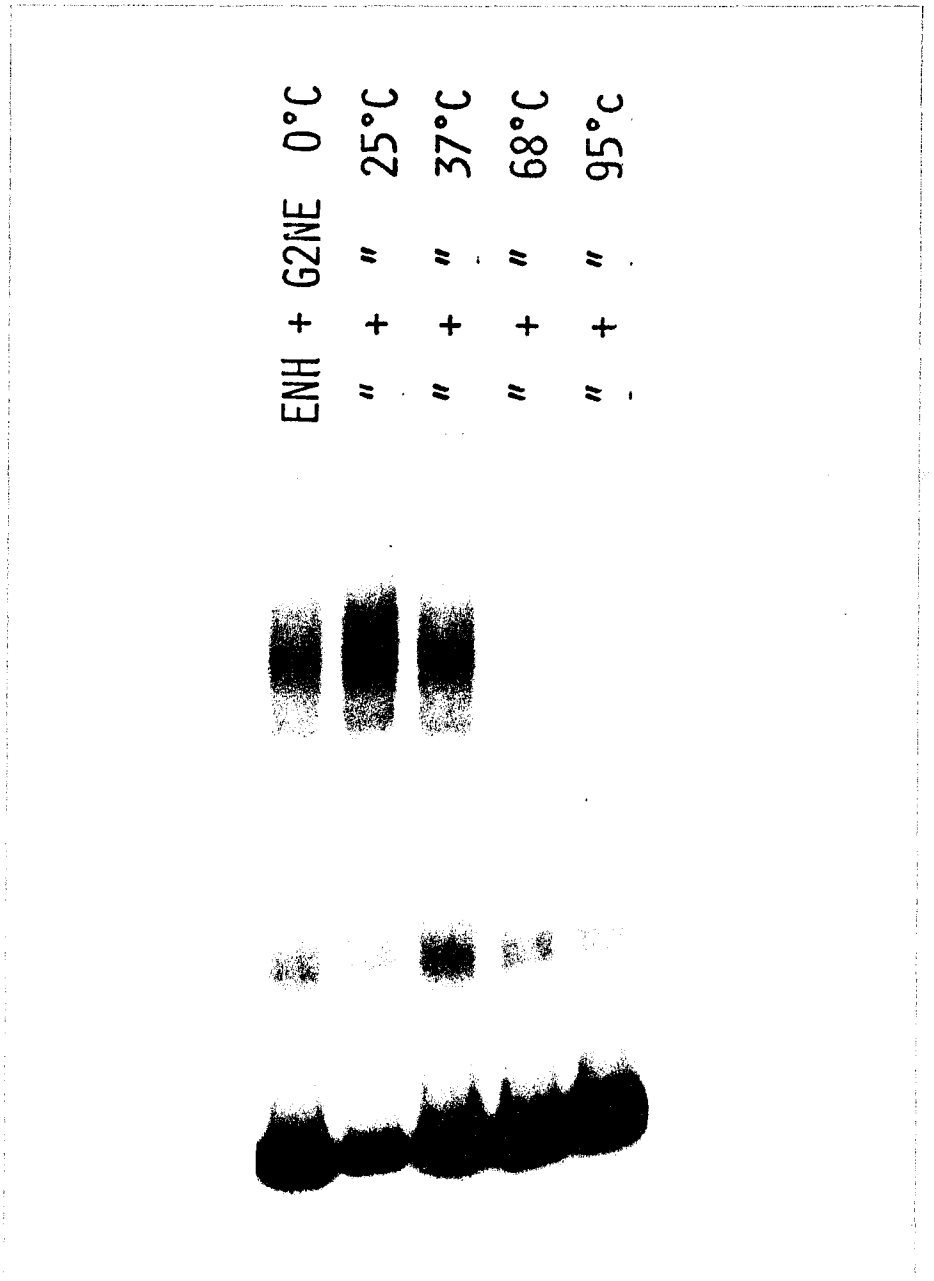

Figure 11 Temperature sensitivity of enhancer binding proteins. Incubations of radiolabelled enhancer DNA with heat-treated HepG2 nuclear extract [pre-heated to the indicated temperatures for 10 minutes, then chilled on ice for 10 minutes prior to room temperature binding]. Binding assays performed as in Figure 10, resolved on a 4% polyacrylamide gel.

DNA competition experiments with the enhancer region

Results from binding experiments with various amounts of "cold" DNAs, both specific and non-specific, are presented in Figures 10 and 12 [lanes 1-3]. The two shifted bands, differing in heat-stability, also differed with respect to competition with homologous [ENH] and non-homologous [SV40 promoter-enhancer] DNAs. SV40 DNA partially competed for binding to the enhancer fragment, appearing to affect the faster migrating complex moreso than the upper one [more obvious in the first three lanes of Figure 13]. If one assumes that the upper complex is a composite of multiple proteins bound to the enhancer, [including the heat-stable protein responsible for the faster migrating complex], then any decrease in the magnitude of the faster migrating complex would be followed with a proportional decrease in the amount of slower-migrating bound complex. Many enhancers, both viral and cellular, are bound by multiple sequence-specific binding proteins, including sites with homologies to the SV40-derived, consensus enhancer sequence [e.g., the immunoglobulin kappa enhancer; Sen & Baltimore, 1986], so it appears that the HBV enhancer proves not to be an exception. Specific competition with the enhancer region was observed with only 6ng of cold ENH DNA included in the binding mixture, whereas no competition was observed with 100 ng of core promoter, or HBsAg promoter DNAs. The faster migrating complex required more cold HBV enhancer DNA to completely compete for binding, than the slower migrating one.

Tissue distribution of the enhancer binding proteins

Given the tissue-specific activity of the HBV enhancer, it would be important to determine whether the factors in HepG2 nuclear extract that bound

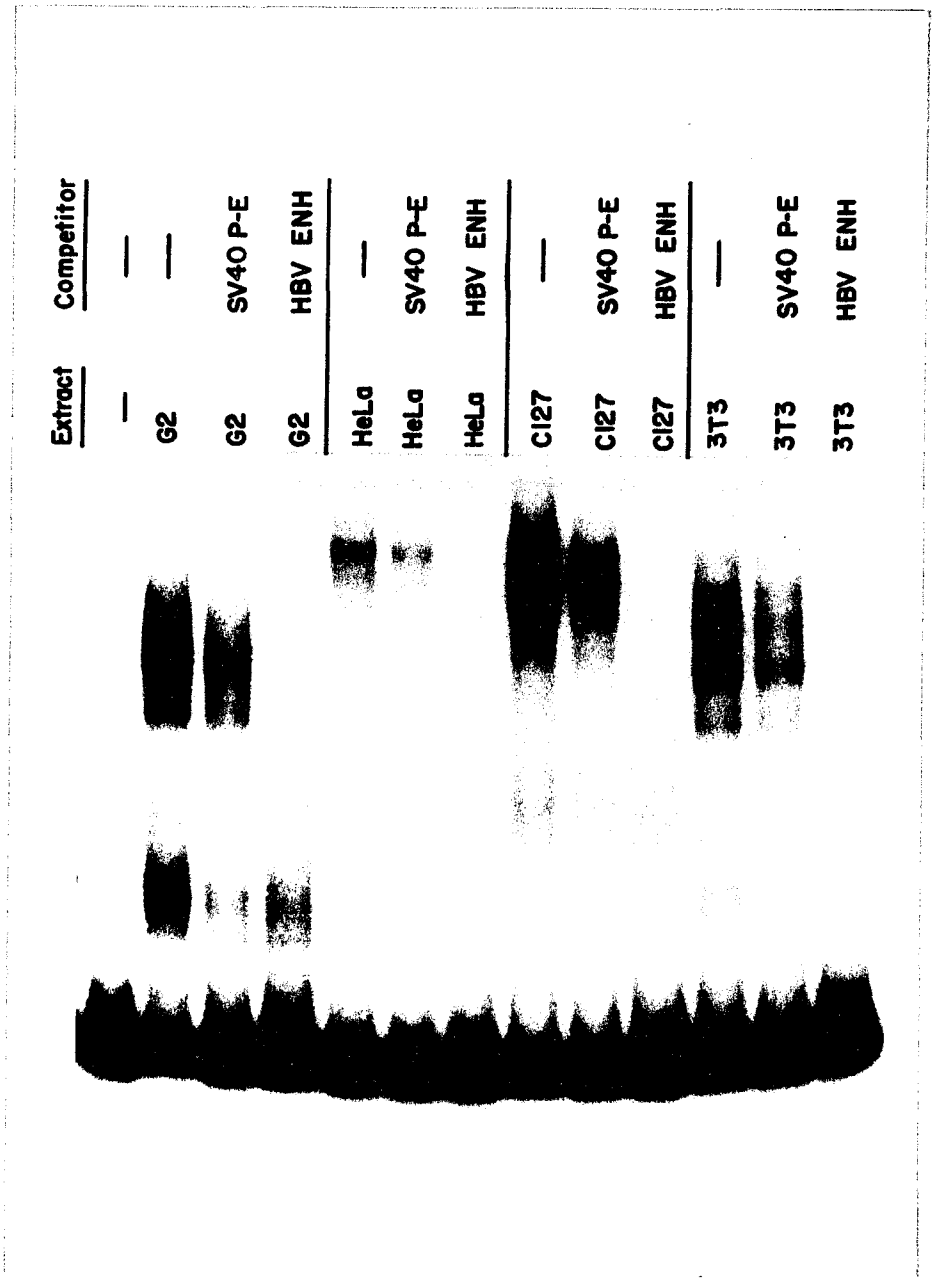

Figure 12 Tissue distribution of HBV enhancer binding proteins. Labelled enhancer fragment incubated with various nuclear extracts in the presence of 8 µg of poly (dI-dC):poly (dI-dC) [source and amounts in Figure 8]. Competing DNAs [Figure 8] were included where indicated.

to the HBV enhancer region were present in extracts from other cell types. If bands appeared with these other cells' extracts, and these bands comigrated with the bound complexes formed with HepG2 extract, then the HepG2-derived factors would not be liver-specific. Incubation of labelled ENH DNA with each extract resulted in the appearance of bound complexes, although of widely differing mobilities and magnitudes [Figure 12].

Competition experiments, with cold SV40 promoter-enhancer and specific HBV enhancer DNAs, exhibited results of some complexity. For HeLa, C127 and 3T3 cells, proteins were present in each extract that specifically recognized sequences in the HBV enhancer, since the binding was only marginally competed by SV40 DNA and completely competed by HBV ENH DNA. These cells therefore contained enhancer-binding proteins with similar characteristics to those present in HepG2 nuclei, yet from the mobility and number of the DNA-protein complexes, it would be assumed that each cell contained different proteins, even if the binding sites were identical.

Localization of the enhancer binding site

Single end-labelled ENH DNA was digested with XmnI [at nt 1247], and both fragments were individually incubated with HepG2 nuclear extract. Incubation with the XmnI-StyI subfragment [representing HBV nt 1247-1378] did not result in the appearance of a shifted band. However, when the remainder of the enhancer region was used as the substrate [DdeI-XmnI, HBV nt 1082-1246] a shift pattern identical to that obtained with the entire ENH DNA was observed. As was the case with using digested CP DNA as substrate for binding, the mobility of the bound complexes was noticeably increased with respect to the use of the entire ENH DNA as substrate [data not shown].

High resolution mapping of protein-binding sites within the HBV enhancer

Mobility shift assays have demonstrated factors present in HepG2 nuclei that bound within the entire enhancer domain [nt 1082-1378], and XmnI digestion of single-end labelled probes placed the binding between nt 1082-1246. Methylation interference assays, as described above, were performed to precisely map the binding sites at the nucleotide level. Results from the methylation interference experiment [performed with full-length single-end labelled methylated ENH DNA] are presented in Figure 13. A large binding region, maximally between nucleotides 1089-1170 of the **short** strand, is apparent. Methylated guanine residues varied in their abililty to inhibit factor binding throughout the region. This span, too large to be the binding site for one factor, most likely involves interactions with multiple DNA-binding proteins. Methylated guanine residues of the **long** strand that interfered with factor binding was much more limited in span in comparison to that which was seen in the **short** strand [data not shown].

Schematically, the enhancer binding region involves the following span:

ENHANCER BINDING REGION:[Nucleotides 1088 to 1170 represented]

```
...5'-ggc tttcactttc tctccaactt acaaGGcctt tctGtGtaaa
...3'-ccg aaagtgaaag agcggttgaa tgttccggaa agacacattt

caatacctGa acctttaccc cGttGcccGG caacGGccag-3'... Short strand
gttatggact tggaaatggg gcaacGGGcc GttGccggtc-5'... Long strand
```

[Capital Gs represent methylated guanine residues that interfered with factor binding]

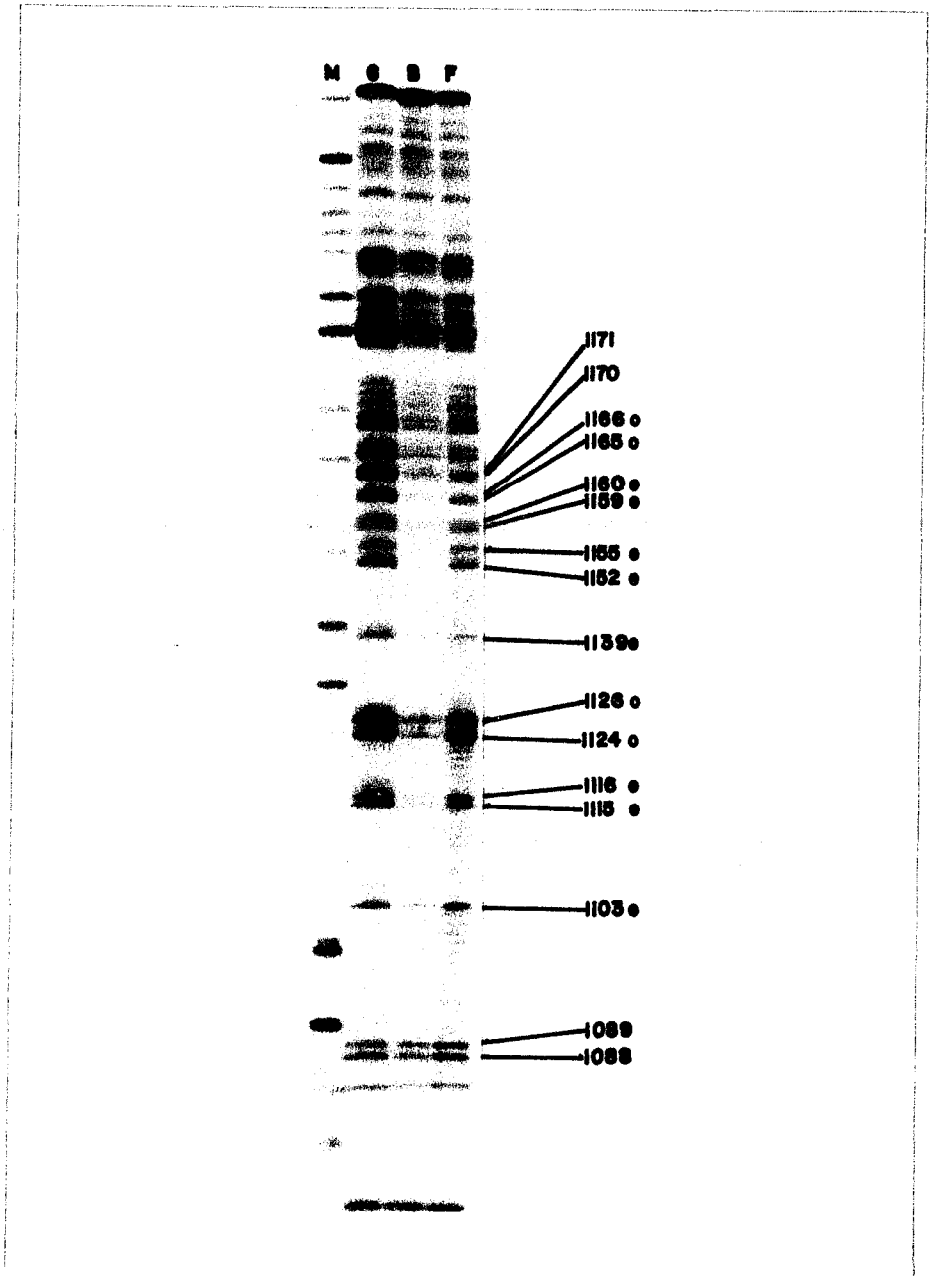

Figure 13 Determination of the protein binding region in the HBV enhancer by methylation interference. **Short** strand labelled methylated enhancer DNA was incubated with HepG2 nuclear extract as described in the Methods. After elution from the preparative gel, the slowest migrating bound complex [**B**] and free, uncomplexed [**F**] DNAs were cleaved with piperidine [as was labelled fragment to provide a guanine ladder (**G**)], and loaded onto a sequencing gel alongside radiolabelled markers [**M**; pBR322/MspI]. Pertinent guanine residues are indicated, filled ovals representing complete interference, and unfilled ovals representing partial intereference of factor binding when these residues were methylated.

Discussion

One of the primary goals of molecular biology has been to understand the underlying mechanisms regulating eukaryotic gene expression. Studies have been performed on many of the steps involved in gene expression--from the status of the chromatin to the translation of encoded protein--including considerable work that has provided much new information on the factors controlling transcription. These studies were a direct result of the application of several newly-developed techniques. Focusing on the expression of a few selected genes, these methods have significantly expanded our knowledge of cis-acting DNA sequences, trans-acting DNA-binding proteins, and their regulatory interactions. But perhaps even more importantly, these techniques can be easily extrapolated to study the factors controlling the expression of any cloned gene. Applying some of these techniques, the present study has described two transcriptional regulatory regions in the HBV genome, and demonstrated the binding of nuclear proteins to specific sequences within these regions.

pEcoAluCAT expression proved to be a faithful mirror of the liver-specific pregenomic, 3.5 kb mRNA transcript of HBV. Highly active in a human hepatoblastoma-derived cell line, HepG2, and virtually inactive in most other cell lines tested, it was well-suited to study the factors involved in this mRNA's tissue-specific expression. Use of the Schöler & Gruss [1984] competition co-transfection CAT assay, demonstrated the presence of titratable, trans-acting factors [HBV DNA binding proteins] in HepG2 nuclei that positively regulated pEcoAluCAT expression. More precisely, some of these factors recognized sequences within the enhancer and core promoter domains of the HBV genome. Further characterization of these DNA-protein interactions was the focus of

this thesis.

Use of the gel electrophoresis mobility shift assay demonstrated the presence of HepG2 nuclear proteins that specifically recognized sites within the enhancer and core promoter domains, supporting the **in vivo** CAT competition co-transfection results with **in vitro** DNA binding data. Binding was not efficiently competed with non-homologous DNAs, but specific competition was observed using each region's DNA, indicating the sequence-specific, non-covalent nature of the binding. Expression of these proteins was restricted among various tissue types and appeared to be liver cell-specific. The binding sites within each region have been resolved at the nucleotide level using the technique of methylation interference. Combining number of bound complexes with the span and composition of the binding sites implicated multiple proteins bound to the enhancer region, and probably only one or two proteins bound within the core promoter.

Palindromes are components of each binding site

The binding site within the core promoter region is between nucleotides 1655 and 1684 [setting the limits to the nearest guanine residue appearing to be excluded from the recognition site; Figure 9 and page 48]. Analysis of the binding site reveals a palindrome based on the sequence 5'-**AAGAG**-3', and a virtually identical repeat of this sequence [substituting a G for the first A] in the **long** strand three nucleotides away. Beginning at nucleotide 1658:

```
...AAGAGgactcttggactctc...  Short
...ttctcctGAGAAcctGAGAG...  Long
```

[Methylation of any of these guanine residues interferes with factor binding]

Sequence analysis of the enhancer binding site [maximally between nt 1089-1170; Figure 14 and page 56] revealed some items of interest--including a perfect palindrome, beginning at nucleotide 1150:

...CCGTTGCCcggcaacgg... **Short**
...ggcaacgggCCGTTGCC... **Long**

Palindromes are frequently seen as constituents of recognition sites for prokaryotic transcriptional regulatory DNA-binding proteins [reviewed in Pabo & Sauer, 1984]. These proteins [e.g., lambda cro, lambda repressor, or lac repressor] bind and function as dimers. The few cellular eukaryotic promoter binding factors analyzed thus far bind as monomers; e.g., MLTF/USF that binds to the -60 region of the adenovirus major late promoter [Carthew et al, 1985; Chodosh et al, 1986; Sawadoga & Roeder, 1986], or the general transcription factor Sp1 [Briggs et al, 1987] even when the binding site is a near-perfect palindrome, as is the MLTF/USF binding site.

The bidirectional functioning of enhancers and some enhancer-like sequences [e.g., the serum responsive element of c-fos or the heat-shock responsive element in Drosophila; Gilman et al, 1986; Prywes & Roeder, 1986; Topol et al, 1985; Wu, 1985] can be explained if the factor's recognition sequence is a palindrome, or if multiple binding sites are inversely repeated within the enhancer. Whether or not preservation of each palindrome is important for binding of the CP or ENH binding factors awaits a detailed mutational analysis of each binding site and the effect of these mutations on factor binding and transcription enhancement.

Binding site length suggests a single factor binding to the core promoter, and multiple factors binding within the enhancer

The length of the binding site yields information concerning the

number of factors bound. The CP binding site is between 21-30 nucleotides long, and therefore in the same size range as prokaryotic operator sequences that are bound by protein dimers, and slightly longer than the 17 nt MLTF recognition sequence in the adenovirus major late promoter, which binds as a monomer. The appearance of a single bound complex in gel mobility shift assays does not exclude the possibility of more than one DNA-binding protein, [or a single protein that binds only as a dimer], but the binding of these proteins would not be independent; rather, strongly cooperative binding would be assumed. Stepwise binding of a dimer would be readily observed with gel mobility assays.

In clear contrast to the CP binding site, the binding site in the enhancer region is extremely large, between 60-80 nucleotides long. All guanine residues between nt 1103-1166 are implicated in factor binding, with varying degrees of involvement [Figure 14]. A single protein cannot be responsible for binding to this region, thus indicating the binding of several factors within the HBV enhancer. Shaul et al [1985] when first describing the enhancer element, showed that various deletions in the enhancer diminished, but did not eliminate, its function. Each deletion was suggested to remove at least one of many trans-acting factor binding sites. Deletions did not venture into the binding region determined in this thesis, although they did point out SV40 enhancer-derived "consensus" enhancer sequence homologies [also termed "core" enhancer sequences] that are included in this large binding domain.

The sequence 5'-**GTGG**($^{TTT}/_{AAA}$)**G**-3' found in the SV40 enhancer has been proposed as a consensus enhancer sequence [Khoury & Gruss, 1983]. The sequences **GTGTAAAC**, **GGGTAAAG**, and **AGTGTTTG** are in the HBV enhancer region beginning at nucleotides 1124, 1140, and 1181, respectively. In the

present experiments, the first two sequences are bound by HepG2 nuclear proteins, while the last is not [although Shaul et al, 1987 describe a protein bound to this region]. This consensus enhancer sequence is part of the SV40 enhancer, shown to be bound by nuclear factors governing SV40 enhancer function [Schöler & Gruss, 1984; Wildeman et al, 1986]. The faster migrating band seen in my gel mobility shift assays can be partially competed with excess SV40 DNA, indicating that this complex consists of liver-specific factors that probably recognize at least one of the two protected, SV40 DNA-derived consensus enhancer sequences.

The region bound by protein in the slower migrating enhancer complex was determined, and is conceivably a composite of the consensus enhancer binding protein(s) plus additional, liver-specific factor(s) binding to the palindrome. A definitive statement on the formation of this bound complex will rest on select deletions within various portions of the enhancer, and their resulting effects on complex formation and band shift patterns.

Other explanations for slower migrating bands appearing on mobility shift assays include protein bound to a previously formed nucleoprotein complex, or an altered secondary structure of the DNA-protein complex [see review by Ptashne, 1986]. Binding site analysis of complexes formed in these ways would be identical for each bound species, since only one protein-DNA contact would be required.

Computer generated searches for homologies between HBV DNA binding sites and sequenced genes

Although discovered as proteins that recognized sequences within the HBV genome, it is reasonable to assume that these liver-specific DNA-binding proteins exist primarily not for the control of **viral** gene expression, but as components involved in the transcriptional regulation of **cellular**, liver-specific genes. Within the loci of these genes [regulated by the binding of liver-specific proteins] would be binding sites similar to those found in the HBV genome. If any of these analogous genes have been cloned and sequenced, then a search for sequences [homologous to the binding sites in the HBV genome] within the appropriate database would reveal them. A computer search of all animal and viral sequences in the GenBank database [via the Bionet IFIND program that uses the Wilbur & Lipmann algorithm] was performed looking for homologies to the enhancer and core promoter palindromic binding sites.

Due to the nature of the search algorithm, a search for homologies to the entire HBV enhancer binding region would probably not yield much significant information. If enhancers are composites of various transcriptionally-regulatory sequence "cassettes" [Voss et al, 1986], and the HBV enhancer is not an exception, then a dissection of the region into appropriate fragments for computer-generated homologies would be in order. Unlike the SV40-derived consensus enhancer sequence homologies, no attention had previously been paid to the palindrome within the HBV enhancer, and a search for homologies to a 30 nt region encompassing this sequence was conducted. The 100 top-ranked matches, were analyzed, and those which seem of most interest are presented below in Table 4.

Table 4

Homologies to the enhancer palindrome

1141 1151 1161 1170 [HBV nt]	
ACCTTTACCCCGTTGCCCGGCAACGGCCAG	HBVayw
ACCTTTACCCCGTTGCCgaGCAACGGCCAG	HBVadyw
ACCTTTACCCCGTTGCCCGGCAACGGtCAG	HBVadr
ACCTTTACCCCGTTGCtCGGCAACGGCCtG	HBVadw
AgtTgTggCCCGTTGCCaGaCAACGtggtG	GSH
AgtTgTggCCCGTTGtCaGGCAACGtggcG	WHV
tCtTTTAattaGTTGCtaGGCAACtGCCct	Polyomavirus

As expected, the greatest similarities are with corresponding sequences in the genomes of other HBV subtypes, with more variation in the other sequenced hepadnaviral genomes [from ground squirrel hepatitis virus (GSH) and woodchuck hepatitis virus (WHV)]. The mismatches do not disturb the palindromic nature of the enhancer binding site, appearing either at the center, or outside of, the palindrome. This indicates the importance of preserving this moiety for HBV. The binding site is within the polymerase open reading frame and the conservation of these nucleotides among hepadnaviruses might reflect protein coding capability, in addition to preservation of a trans-acting factor binding site.

There was one highly significant homologous non-HBV sequence uncovered by this search, also centered about the palindrome. From the polyomavirus genome, [beginning at nucleotide 5150] this sequence mismatches at the end, and the midpoint of the palindrome, similar to the mismatches across HBV subtypes. Also, two of the mismatched nucleotides are complementary, revealing a palindrome of the sequence **AGTTGC...GCAACT**. This sequence is within the polyomavirus "B" enhancer region, is important for polyomavirus enhancer

function, and is bound by nuclear proteins from 3T6 cells [Piette et al, 1985]. It would be interesting to determine whether these two viral sequences can compete for binding of the same nuclear factors, or can complement each other's enhancer activities.

Since the HBV enhancer functions primarily in liver cells, it would have been expected to uncover some liver-specific genes with this search. Although many liver-specific genes have been cloned and sequenced, they are mainly cDNA clones. Cellular enhancers reside primarily outside the coding region of cellular genes [i.e., upstream and downstream of coding regions, or within introns], and thus would not be found within their cDNA sequences. The HBV enhancer was the first [and so far, only] enhancer sequence in a protein-coding region found to date [Tognoni et al, 1985]. When the loci of these liver-specific genes are sequenced more extensively, it would not be surprising to then find homologies to the HBV enhancer palindromic binding site.

The search for CP binding site homologies [to HBV nt 1654-1683] produced some items of interest, including liver-specific genes [Table 5].

Table 5
Homologies to the core promoter binding site

SEQUENCE	GENE [Distance from mRNA 5' end]
CTTACATAAGAGGACTCTTGGACTCTCAGC	HBVayw [-150]
CTTACATAAGAGGACTCTTGGACTCTCAGC	HBVadr ["]
CTTACATAAGAGGACTCTTGGACTCcCAGC	HBVadw ["]
aTgcCAagcaAGGACctTGGACTCcttat	WHV ["]
aTgAtgcAAaAGGACTtTTGGACTgcttat	GSHV ["]
gggtggaAAGAGGAC*CTTGG*CTCTCAGa	Mouse apolipo. A [-300]
CgTggtgAAGAGGAgTCTTGG*CTCTCAtg	Rat fatty acid BP [+100]
tgacCATcAGAGGACcCTTGGAtTCTCcag	Rat albumin [+80]
acactccAgGAGGcCTCTTGGACTtTaAtg	Rat apolipo. A [-100]

[apolipo. A ≡ apolipoprotein A; BP ≡ binding protein; * indicates various types of discontinuities no larger than several nucleotides]

In addition, the sequence 5'-($^{A}/_{G}$)**AGAG**-3' appears numerous times in a recently published sequence analysis of the 5' flanking regions of the human albumin and α-fetoprotein genes [Urano et al, 1986]. Within 450 nucleotides of the start of transcription, this sequence appears twice in the albumin, and six times in the α-fetoprotein 5' flanking regions.

Function of the CP binding protein

The mRNA isolated from HBV-infected livers indicates that the 3.5 kb transcript initiates within the pre-core region, at nt 1816 ± 5 [Cattaneo et al, 1984; Will et al, 1987]. Recently, a transfected HepG2 cell line that supports all aspects of HBV replication **in vitro** has been made in our laboratory [Sells et al, 1987]. Analysis of mRNA isolated from this cell line confirms the fact that it contains correctly initated pre-genome mRNA. Presumably then, untransfected HepG2 cells contain all necessary factors to faithfully produce authentic pre-genome mRNA. The core promoter region used in my mobility shift assays covers a region from approximately 170 nucleotides upstream, to 70 nucleotides downstream from the pre-genome mRNA initiation site. From available information on eukaryotic promoter regions, this fragment should be sufficiently large enough to contain all major cis-acting domains involved in the regulation of pre-genome mRNA transcription [reviewed by Yaniv, 1984 and Maniatis et al, 1987]. Within this fragment, I found only one site bound by a factor present in HepG2 extract, and this site is 120 nt upstream from the start of transcription. Only a detailed deletional analysis across the entire core promoter region will tell if this binding site is important for pre-genome transcription, and if any other cis-acting regulatory regions exist, possibly closer to the start of transcription. This binding site is not too distant from the initiation of

transcription to exert an effect, since Baldwin & Sharp [1987] found a functional protein binding site within the mouse H-2K^b class I major histocompatibility gene promoter, located 166 nucleotides upstream of this mRNA's initiation site. This sequence also contains a palindrome.

The binding site at nt 1660 is close to the start site of transcription of some minor HBV transcripts--namely those found in infected chimpanzee liver at nt 1680 by Will et al [1987], and an RNA Polymerase II initiation site as determined in a cell free **in vitro** transcription system also at nt 1680 [Rall et al, 1983]. GSHV transcripts from infected ground squirrel livers include a number of minor polyadenylated RNAs whose 5' ends map between HBV nt 1710 to 1750 [Enders et al, 1985], and minor transcripts that start within this region have also been detected in the previously mentioned transfected HepG2 cell line [Zelent et al; manuscript in preparation]. Unlike the pregenome mRNA [which starts within the pre-core region], these transcripts can encode a polypeptide beginning with the pre-core AUG. Therefore, it is possible that the factor that binds at nt 1660 might be involved in the expression of these minor transcripts, and not the pre-genome mRNA, or possibly both.

Comparison of the enhancer binding region to published data

During the completion of this work, a report describing nuclear factors that bind to sites within the HBV enhancer region was published [Shaul et al, 1987]. DNAse I footprinting revealed four sites between HBV nucleotides 827-1374 bound by factors from crude HepG2 nuclear extract. One of these protected regions, [from nt 1180 to 1206, called the "E" site], is apparently bound by a liver-specific factor, while the other regions are bound by ubiquitous factors [between nt 881 and 1050, and therefore outside the limits of

the enhancer fragment used in the studies presented in this thesis]. The protein that binds to the E site is extremely heat stable, much like the protein responsible for the faster migrating, shifted band seen in gel mobility shift assays [Figure 11]. Furthermore, the E site has strong homology to the SV40 enhancer core sequence [as previously discussed], and in my mobility shift assays, this faster migrating shifted band diminished in intensity when excess cold SV40 DNA was included in the binding mixture. But more than one sequence in the HBV enhancer has SV40 consensus enhancer DNA homologies, including two that were bound by HepG2 nuclear proteins, as described in this work. When Shaul et al [1987] used more HepG2 extract in their binding reactions, virtually the entire enhancer region was protected from DNAase I digestion, including the large binding region found in my experiments the gel mobility shift/methylation interference assays. It is conceivable that with fractionation of their extract, dissection of the factors responsible for these extended DNAase I-resistant regions would corroborate my findings.

Even though Shaul et al used the same cell line and HBV DNA subtype in their experiments as I did, they did not detect the palindromic binding site centered about nt 1159, nor the adjoining binding region; and I did not detect a the E binding site. Plausible explanations for these discrepancies can be summarized as follows:

1.) since the nuclear extraction methods were different, the observed binding was with crude extracts of differing protein content;

2.) without proper competition experiments, DNAase I footprinting with crude extracts can result in spurious artifactual binding;

3.) DNAaseI footprinting can miss a binding site if it occurs in a relatively nuclease-resistant region, whereas methylation interference, which relies on

a chemically altered substrate, would be less likely to miss the presence of a binding site; and

4.) Shaul et al [1987] did detect binding to the same enhancer region discussed here, but only when binding reactions included more extract. The specificity of this binding would have to be established.

Conclusions and possibilities for further studies

The goal of these experiments was to understand the mechanisms involved in the tissue-specific expression of the 3.5 kb mRNA of HBV, primarily emphasising the role of sequence-specific DNA-binding proteins regulating transcription. Nuclear proteins bearing HBV enhancer and core promoter DNA binding determinants were found in a human hepatoblastoma-derived cell line. These proteins were restricted in their expression to this cell type. The specificity of binding has been determined, as has the identification of the respective protein-binding sites in each region. Chemical synthesis of oligonucleotides containing these binding sites [and those with specific mutations, focusing on disruption of the respective palindromes] would be the next step. Gel mobility shift assays with these oligonucleotides would be expected to corroborate the findings presented in this thesis. **In vivo** competition co-transfection experiments with pEcoAluCAT would establish whether or not these oligonucleotides contained the specific DNA sequences responsible for HBV enhancer and core promoter functions, whereas mutated oligonucleotides, [without protein-binding activity], should not. Also, pEcoAluCAT derivatives, containing various deletions and mutations in these binding sites, should be constructed. CAT activity would be expected to be profoundly diminished in the appropriate pEcoAluCAT subclones.

Once transcriptional regulatory activity has been conclusively shown to reside within the DNA sequences contained in the synthesized oligonucleotides, purification of the binding proteins is in order. Following protocols by Briggs et al [1987], and Chodosh et al [1986], affinity purification of the factors from HepG2 cells could be accomplished within six months. Purification would proceed by passage of the crude extract over two to three columns [e.g., size, heparin-agarose, ion exchange] prior to passage over an affinity column consisting of oligonucleotide-bound sepharose. Gel mobility shift assays would ascertain the presence of binding in each fraction. Once purified, the functionality of these proteins should be assayed in an **in vitro** transcription system. And finally, peptide sequence analysis of these purified proteins would lead to the possible cloning of these genes, providing the opportunity for the analysis of the control of expression of factors that regulate liver-specific transcription.

Bibliography

Atchison ML, Perry RP [1987]: The role of the kappa enhancer and its binding factor NF-KB in the developmental regulation of kappa gene transcription. Cell **48**,121-128

Baldwin AS Jr., Sharp PA [1987]: Binding of a nuclear factor to a regulatory sequence in the promoter of the mouse H-2K^b class I major histocompatibility gene. Mol. & Cell. Biol. **7**,305-313

Banerji JS, Rusconi S, Schaffner W [1981]: Expression of a ß-globin gene is enhanced by remote SV40 sequences. Cell **27**,299-308

Birnboim HC, Doly J [1979]: A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. **7**,1513-1523

Böhnlein E, Gruss P [1986]: Interaction of distinct nuclear proteins with sequences controlling the expression of polyomavirus early genes. Mol. & Cell. Biol. **6**, 1401-1411

Briggs MR, Kadonaga JT, Bell SP, Tjian R [1986]: Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. Science **234**,47-52

Carthew RW, Chodosh LA, Sharp PA [1985]: An RNA Polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell **43**,439-448

Cattaneo R, Will H, Schaller H [1984]: Hepatitis B virus transcription in the infected liver. EMBO J. **3**,2191-2196

Cattaneo R, Will H, Darai G, Pfaff E, Schaller H [1983]: Detection of an element of the SV40 late promoter in vectors used for expression studies in COS cells. EMBO J. **2**,511-514

Cattaneo R, Will H, Hernandez N, Schaller H [1983]: Signals regulating hepatitis B surface antigen transcription. Nature **305**,336-338

Chakraborty PR, Ruiz-Opasa N, Shafritz DA [1981]: Transcription of human hepatitis B virus core antigen gene sequences in an **in vitro** HeLa cellular extract. Virology **111**,647-652

Chodosh LA, Carthew RW, Sharp PA [1986]: A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. Mol. & Cell. Biol. **6**,4723-4733

Christman JK, Gerber M, Price PM, Flordellis C, Edelman J, Acs G [1982]: Amplification of expression of hepatitis B surface antigen in 3T3 cells cotransfected with a dominant-acting gene and cloned viral DNA. Proc. Natl. Acad. Sci. USA **79**,1815-1819

Dane DS, Cameron CH, Briggs M [1970]: Virus-like particles in serum of patients with Australia-antigen associated hepatitis. Lancet **1**,695-698

Dignam JD, Lebowitz RM, Roeder RG [1983]: Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucl. Acids Res. **11**, 1475-1489

Dunn AR, Gough N [1984]: Tissue-specific enhancers. TIBS **9**,81-82

Dynan W, Tjian R [1982]: Transcriptional enhancer sequences: a novel regulatory element. TIBS **7**,124-125

Dynan WS, Tjian R [1985]: Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature **316**,774-778

Dynan WS, Sazer S, Tjian R, Schimke RT [1986]: Transcription factor Sp1 recognizes a DNA sequence in the mouse dihydrofolate reductase promoter. Nature **319**,246-248

Dynan WS [1987]: What mechanisms underlie tissue-specific gene transcription? Trends in Genetics **3**,121-122

Enders GH, Ganem D, Varmus H [1985]: Mapping the major transcripts of ground squirrel hepatitis virus: the presumptive template for reverse transcriptase is terminally redundant. Cell **42**,297-308

Fried M, Crothers DM [1981]: Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucl. Acids Res. **9**,6505-6525

Galibert F, Mandart E, Fitoussi F, Tiollais, Charnay P [1979]: Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in E. coli. Nature **281**,646-650

Garner MM, Revzin A [1986]: The use of gel electrophoresis to detect and study nucleic acid-protein interactions. TIBS **11**,395-396

Garner MM, Revzin A [1981]: A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucl. Acids Res. **9**,3047-3059

Gilman MZ, Wilson RN, Weinberg RA [1986]: Multiple protein-binding sites in the 5'-flanking region regulates c-fos expression. Mol. & Cell. Biol. **6**,4305-4316

Gorman C, Moffat LF, Howard BH [1982]: Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. & Cell. Biol. **2**,1044-1051

Gough NM [1983]: Core and e antigen synthesis in rodent cells transformed with hepatitis B virus DNA is associated with greater than genome length viral messenger RNAs. J. Mol. Biol. **165**,683-699

Graham FL, van der Eb AJ [1973]: A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**,456-457

Hanahan D [1983]: Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. **166**,557-580

Hendrickson W, Schleif R [1985]: A dimer of AraC protein contacts three adjacent major groove regions of the aral DNA site. Proc. Natl. Acad. Sci. USA **82**,3129-3133

Jones KA, Kadonga JT, Rosenfeld PJ, Kelly TJ, Tjian R [1987]: A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell **48**,79-89

Karin M, Haslinger A, Holtgreve H, Cathala G, Slater E, Baxter JD [1984]: Activation of a heterologous promoter in response to dexamethasone and cadmium by metallothionein gene 5'-flanking DNA. Cell **36**,371-379

Khoury G, Gruss P [1983]: Enhancer elements. Cell **33**,313-314

Knowles B, Howe CC, Aden DP [1980]: Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. Science **209**,497-499

Laub O, Rall LB, Truett M, Shaul Y, Standring DN, Valenzuela P, Rutter WJ [1983]: Synthesis of hepatitis B surface antigen in mammalian cells: expression of the entire gene and the coding region J. Virol. **48**,271-280

Lee W, Mitchell P, Tjian R [1987]: Purified transcription factor AP-1 interacts with TPA-inducible elements. Cell **49**,741-752

Lewin B [1985]: Genes (second edition) John Wiley & Sons, New York, p. 293ff

Loyter A, Scangos GA, Ruddle FH [1982]: Mechanisms of DNA uptake by mammalian cells: fate of exogenously added DNA monitored by the use of fluourescent dyes. Proc. Natl. Acad. Sci. USA **79**,422-426

Malpiece Y, Michel ML, Carloni G, Revel M, Tiollais P, Weissenbach J [1983]: The gene S promoter of hepatitis B virus confers constitutive gene expression. Nucl. Acids Res. **11**,4645-4654

Maniatis T, Fritsch EF, Sambrook J [1982]: Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Maniatis T, Goodbourn S, Fischer JA [1987]: Regulation of inducible and tissue-specific gene expression. Science **236**,1237-1245

Matsui T, Segall J, Weil A, Roeder RG [1980]: Multiple factors required for accurate initiation of transcription by purified RNA Polymerase II. J. Biol. Chem. **255**,11992-11996

Maxam AM, Gilbert W [1980]: Sequencing end-labeled DNA with base specific chemical cleavages. (L. Grossman and K. Moldave, editors) Methods in Enzymology, **65**,499-560

McKnight S, Tjian R [1986]: Transcriptional selectivity of viral genes in mammalian cells. Cell **46**,795-805

Miller RH, Robinson WS [1986]: Common evolutionary origin of hepatitis B virus and retroviruses. Proc. Natl. Acad. Sci. USA **83**,2531-2535

Mizushima-Sugano J, Roeder RG [1986]: Cell-type-specific transcription of an immunoglobulin kappa light chain gene **in vitro**. Proc. Natl. Acad. Sci. USA **83**,8511-8515

Möröy T, Etiemble J, Trépo C, Tiollais P, Buenida M-A [1985]: Transcription of woodchuck hepatitis virus in the chronically infected liver. EMBO J. **4**,1507-1514

Neurath AR, Kent SBH, Strick N, Parker K [1986]: Identification and chemical synthesis of a host cell receptor binding site on hepatitis B virus. Cell **46**,429-436

Ono Y, Onda H, Sasada R, Igarashi K, Sugino Y, Nishioka K [1983]: The complete nucleotide sequence of the cloned hepatitis B virus DNA: subtype adr and adw. Nucl. Acids Res. **11**,1747-1757

Ou J-H, Laub O, Rutter WJ [1986]: Hepatitis B virus gene function: the precore region targets the core antigen to cellular membranes and causes the secretion of the e antigen. Proc. Natl. Acad. Sci. USA **83**,1578-1582

Ou J-H, Rutter WJ [1987]: Regulation of secretion of the hepatitis B virus major surface antigen by the pre-S1 protein. J. Virol. **61**,782-786

Pabo CO, Sauer RT [1984]: Protein-DNA recognition. Ann. Rev. Biochem. **53**,293-321

Piette J, Kryszke M-H, Yaniv M [1985]: Specific interaction of cellular factors with the B enhancer of polyoma virus. EMBO J. **4**,2675-2685

Pourcel C, Louise A, Gervais M, Chenciner N, Dubois M-F, Tiollais P [1982]: Transcription of the hepatitis B surface antigen gene in mouse cells transformed with cloned viral DNA. J. Virol. **42**,100-105

Price P, Ostrove S, Flordellis C, Sells MA, Thung S, Gerber M, Christman JK, Acs G [1983]: Characterization of RNA transcripts and virally coded proteins synthesized in mouse fibroblasts transfected with hepatitis B DNA. Biosci. Rep. **3**,1017-1026

Price PM, Hirschman SZ, Garfinkel E [1980]: DNA cloned from the ayw subtype of hepatitis B virus. J. Med. Virol. **6**,139-145

Prywes R, Roeder RG [1986]: Inducible binding of a factor to the c-fos enhancer. Cell **47**,777-784

Ptashne M [1986]: Gene regulation by proteins acting nearby and at a distance. Nature **322**,697-701

Rall LB, Standring DN, Laub O, Rutter WJ [1983]: Transcription of hepatitis B virus by RNA polymerase II. Mol. & Cell. Biol. **3**,1766-1773

Roossnick MJ, Jameel S, Loukin SH, Siddiqui A [1986]: Expression of hepatitis B viral core region in mammalian cells. Mol. & Cell. Biol. **6**,1393-1400

Sawadoga M, Roeder RG [1985]: Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell **43**,165-175

Schöler HR, Gruss P [1984]: Specific interactio between enhancer-containing molecules and cellular components. Cell **36**,403-411

Seeger C, Ganem D, Varmus HE [1986]: Biochemical and genetic evidence for the hepatitis B virus replication strategy. Science **232**,477-484

Sells MA, Chen M-L, Acs G [1987]: Production of hepatits B virus particles in HepG2 cells transfected with cloned hepatitis B virus DNA. Proc. Natl. Acad. Sci. USA **84**,1005-1009

Sen R, Baltimore D [1986]: Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell **46**,705-716

Shaul Y, Rutter WJ, Laub O [1985]: A human hepatitis B viral enhancer element. EMBO J. **4**,427-430

Shaul Y, Ben-Levy R [1987]: Multiple nuclear proteins in liver cells are bound to hepatitis B virus enhancer element and its upstream sequences. EMBO J. **6**,1913-1920

Shaul Y, Ben-Levy R, De-Medina T [1986]: High affinity binding site for nuclear factor I next to the hepatitis B virus S gene promoter. EMBO J. **5**,1967-1971

Siddiqui A [1983]: Expression of hepatitis B virus surface antigen gene in cultured cells by using recombinant plasmid vectors. Mol. & Cell. Biol. **3**,143-146

Siddiqui A, Jameel S, Mapoles J [1986]: Transcriptional control elements of hepatitis B surface antigen gene. Proc. Natl. Acad. Sci. USA **83**,566-570

Siddiqui A, Jameel S, Mapoles J [1987]: Expression of the hepatitis B virus X gene in mammalian cells. Proc. Natl. Acad. Sci. USA **84**,2513-2517

Siebenlist U, Gilbert W [1980]: Contacts between Escherichia coli RNA Polymerase and an early promoter of phage T7. Proc. Natl. Acad. Sci. USA **77**,122-126

Simonsen CC, Levinson AD [1983]: Analysis of processing and polyadenylation signals of the hepatitis B surface antigen gene by using simian virus 40-hepatitis B virus chimeric plasmids. Mol. & Cell. Biol. **3**,2250-2258

Singh H, Sen R, Baltimore D, Sharp PA [1985]: A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature **319**,154-158

Standring DN, Rutter WJ, Varmus HE, Ganem D [1984]: Transcription of the hepatitis B surface antigen gene in cultured murine cells initiates within the presurface region. J. Virol. **50**,563-571

Standring DN, Rall LB, Laub O, Rutter WJ [1983]: Hepatitis B virus encodes an RNA Polymerase III transcript. Mol. & Cell. Biol. **3**,1774-1782

Standring DN, Rutter WJ [1986]: The molecular analysis of hepatitis B virus. in Progress in Liver Diseases, vol. VIII, 311-333

Strauss F, Varshavsky A [1984]: A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. Cell **37**,889-901

Summers J, Mason WS [1982]: Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. Cell **29**,403-415

Summers J [1981]: Three recently described animal virus models for human hepatitis B virus. Hepatology **1**,179-183

Sureau C, Romet-Lemonne J-L, Mullins JI, Essex M [1986]: Production of hepatitis B virus by a differentiated human hepatoma cell line after transfection with cloned circular HBV DNA. Cell **47**,37-47

Tiollais P, Charnay P, Vyas GN [1981]: Biology of the hepatitis B virus. Science **213**,406-411

Tiollais P, Pourcel C, Dejean A [1985]: The hepatitis B virus. Nature **317**,489-495

Tognoni A, Cattaneo R, Serfling E, Schaffner W [1985]: A novel expression selection approach allows precise mapping of the hepatitis B virus enhancer. Nucl. Acids Res. **13**,7457-7472

Toh H, Hayashida H, Miyata T [1983]: Sequence homology between retroviral reverse transcriptase and the putative polymerase of hepatitis B virus and culiflower mosaic virus. Nature **305**,827-829

Topol J, Ruden DM, Parker CS [1985]: Sequences required for in vitro transcriptional activation of a Drosophila hsp 70 gene. Cell **42**,527-537

Treinin M, Laub O [1987]: Identification of a promoter element located upstream from the hepatitis B virus X gene. Mol. & Cell. Biol. **7**,545-548

Tur-Kaspa R, Burk RD, Shaul Y, Shafritz DA [1986]: Hepatitis B virus contains a glucocorticoid responsive element. Proc. Natl. Acad. Sci. USA **83**,1627-1631

Urano Y, Watanabe K, Sakai M, Tamaoki T [1986]: The human albumin gene: characterization of the 5' and 3' flanking regions and the polymorphic gene transcripts. J. Biol. Chem. **261**,3244-3251

Valenzuela P, Quiroga M, Zaldivar J, Gray P, Rutter WJ [1980]: The nucleotide sequence of the hepatitis B viral genome and the identification of major viral genes. In Animal Virus Genetics, (B. Fields, R. Jaenisch and C.F. Fox, editors) 57-70

Voss SD, Schlokat U, Gruss P [1986]: The role of enhancers in the regulation of cell-type-specific transcriptional control. TIBS **11**,287-289

Wigler M, Perucho M, Kurtz D, Dana S, Pellicer A, Axel R, Silverstein S [1980]: Transformation of mammalian cells with an amplifiable dominant-acting gene. Proc. Natl. Acad. Sci. USA **77**,3567-3570

Wildeman AG, Zenke M, Schatz C, Wintzerith M, Grundström T, Matthes H, Takahashi K, Chambon P [1986]: Specific protein binding to the SV40 enhancer in vitro. Mol. & Cell. Biol. **6**,2098-2105

Will H, Reiser W, Weimer T, Pfaff E, Büscher M, Sprengel R, Cattaneo R, Schaller H [1987]: Replication strategy of human hepatitis B virus. J. Virol. **61**,904-911

Will H, Salfeld J, Pfaff E, Manso C, Theilmann L, Schaller H [1986]: Putative reverse transcriptase intermediates of human hepatitis B virus in primary liver carcinomas. Science **231**,594-596

Wu C [1984]: Two protein-binding sites in chromatin implicated in the activation of heat-shock genes. Nature **309**,229-234

Wu C [1985]: An exonuclease protection assay reveals heat-shock elements and TATA box DNA-binding proteins in crude nuclear extracts. Nature **317**,84-87

Yaniv M [1984]: Regulation of eukaryotic gene expression by transactivating proteins and cis acting DNA elements. Biol. Cell. **50**,203-216

Zelent AZ, Sells MA, Price PM, Mohammed A, Acs G, Christman JK [1987]: Murine cells carrying integrated tandem genomes of hepatitis B virus DNA transcribe RNAs from endogenous promoters on both strands and express middle and major viral envelope proteins. J. Virol. **61**,1108-1115