1990

Albumin Transcriptional Regulation

Roy Steven Herbst

Albumin Transcriptional Regulation

A thesis presented to the faculty of the Rockefeller University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

by

Roy Steven Herbst

1990
The Rockefeller University
New York

Acknowledgments

The successful completion of this thesis would not have been possible without the guidance and support of numerous individuals. Special thanks are due to the following for their significant contributions to my graduate education:

- I had the good fortune of deriving the benefits from co-advisors during my time at Rockefeller, and I thank both for their contributions to my studies. I wish to thank Dr. J.E. Darnell Jr. for the opportunity to have worked in his laboratory, and for guidance throughout my graduate career. I will always look back fondly at my years spent in the Darnell laboratory and at the numerous collaborations which were made possible as a result of the fertile scientific environment which Jim nurtured.

- I am also deeply indebted to Dr. L. E. Babiss, for the daily insights, enthusiasm and patience with which he provided me throughout my graduate years. It was truly a pleasure to have been Lee's student, and I am grateful to have learned so much from him concerning both the design and implementation of experimental research. Additionally, I will always value his friendship.

- I would also like to thank the numerous members of the Darnell laboratory both past and present who were helpful to me through their advice, technical help and friendship.

- Finally I would like to thank my family and friends for their unfaltering support throughout the course of this work.

Table of Contents

List of Figures

List of Tables

Abstract

An analysis is presented of the molecular mechanisms involved in maintaining the liver specific transcriptional profile of the rat and mouse serum albumin genes. Using DNA transfection and recombinant viral infection, the minimal rat promoter and mouse enhancer elements were identified, functionally dissected and shown independently to be hepatocyte specific elements. Based on these studies, multiple cis-acting DNA regulatory regions were elucidated and the cognate proteins which interact with them characterized and where possible their identity determined. It was found that the promoter proximal -156 nucleotides of the rat albumin gene were necessary and sufficient to insure albumin transcriptional specificity. Furthermore, the transcription factors involved in insuring this effect included several liver enriched proteins and others of a more ubiquitous cellular distribution. It is therefore, the interaction of several factors (only some of which are liver specific or enriched) which dictates the liver specificity of this promoter.

The albumin enhancer element was localized much farther upstream (–10.5 to –8.43kbp) and shown to consist of both positive and negative regulatory domains. Two of the proteins which interact with the albumin enhancer are newly described, NLS1 and ANF, a positive and negative acting protein respectively, which interact in a modular fashion to stimulate the rate

of albumin transcription. NLS1 (non-liver specific 1) DNA binding activity is present in all cell types, yet it stimulates albumin transcription only in the liver. ANF (albumin negative factor) negates the action of NLS1 and can act upon heterologous promoter and enhancer elements in this same fashion. This negative factor is present ubiquitously, but its activity is overridden in the liver by a third protein factor. This protein was determined to be C/EBP, a liver enriched transcription factor, which for the enhancer element performs this novel modulatory role. The biological action of these different factors was subsequently evaluated by correlating their relative levels with both the specificity and rate of albumin gene transcription in a series of rat hepatoma cell lines, which each transcribe the albumin gene at a different rate. An analysis of the expression of the rat transthyretin gene was also included in this series of experiments, and the question of coordinate hepatocyte gene control was addressed. Lastly, the control of some of these transcription factors themselves were analyzed by measuring both their transcription rate and protein binding activity in the rat liver and several hepatoma cell types. These factors are regulated in a complex fashion, involving both transcriptional and post transcriptional mechanisms.

Chapter 1

Introduction

The process of cellular differentiation, which occurs during the development of a multicellular organism, involves a multitude of regulatory events. These include the turning on and off of specific cellular genes that encode proteins which together contribute to the attainment of each developmental stage (Darnell, 1985; Gehring, 1987; Maniatis *et al.*, 1987). The differentiation of the mammalian hepatocyte from a single pluripotent precursor cell early in embryonic development is a well-studied process, and has been shown to involve the sequential and coordinated expression of a large number of genes many of which are activated in a tissue specific fashion (Greengard, 1969; 1970; Greengard *et al.*, 1972; Powell *et al.*, 1984; Panduro *et al.*, 1987a). The initial events which set this complex process into motion, referred to as developmental commitment, manifest themselves at the earliest stages of embryonic life, when a limited number of cells exist, making it difficult to study this process experimentally (Gardner and Rossant, 1976). The end product of these determinative events, however, is differentiation; the patterned expression of a specific set of cellular gene products that serves to embody a cell with its own unique phenotypical characteristics.

By examining the normal expression of tissue specific mRNAs in the adult, during development, and as the result of certain pathophysiological states (ie. regeneration and neoplasia), it is possible to explore the mechanisms involved in their regulation. Numerous genes expressed mainly in hepatocytes have been identified and shown to be differentially

and/or coordinately expressed at various times during the developmental process, many as a result of transcriptional control (Derman *et. al.*, 1981; Tilghman and Belayew, 1982, Cistanzom *et. al.*, 1983; Powell et al., 1984; Ciliberto *et al.*, 1985; Darnell 1985; Clayton and Darnell, 1983; Panduro *et al.*, 1987a). The transcriptional control of an individual liver specific mRNA and the means by which multiple genes are expressed in a concerted fashion to produce the differentiated hepatocyte therefore represent fundamental developmental decisions.

My thesis has been directed towards obtaining a better understanding of the cis-acting DNA sequence elements and the trans-acting cellular regulatory factors that contribute to cellular differentiation in the rodent liver. To accomplish this task, I have focused upon a single gene, that for serum albumin, which is transcribed in a liver specific manner. I examined the transcription of albumin in numerous cell types and tissues (e.g. the mouse and rat liver, human and rat hepatoma cells, CWSV1 rat cells and primary hepatocytes) which each express albumin to a different extent. For each of these experimental systems, I have addressed a set of questions, which includes the following:

I. What is the rate of albumin transcription?

II. What are the cis-acting DNA sequence elements involved in the control of albumin transcription?

III. How many trans-acting factors interact with these regulatory elements? Are these factors liver specific or ubiquitous? Do they resemble any other known transcription factors? How is cell-type specific transcription maintained?

IV. Are these factors shared among different genes and involved in coordinate gene control? How are these factors themselves regulated?

I. Tissue specific gene expression: A general overview

Differentiated cells, such as hepatocytes, pancreatic acinar and islet cells, lens cells, blood cells and lymphocytes, all produce characteristic proteins because the genes encoding them are specifically expressed in that cell type and not elsewhere, despite the fact that all cells maintain identical DNA complements (Derman *et al.*, 1981; Edlund *et al.*, 1985; Boulet *et al.*, 1986; Davis et al., 1987; Lenardo *et al.*, 1987; Scott and Carol, 1987; Maniatis *et al.*, 1987). The regulation of the production of cell-type specific products which is the distinctive characteristic of a differentiated cell, must therefore result from modulations at some point along the complex pathway which generates a functional protein from a DNA encoded gene sequence. These alterations can occur at any one of several different levels of gene control.

Levels of gene control

The production of a specific cellular protein begins with the process of transcription, in which an RNA molecule is copied from a DNA template by the enzyme RNA Polymerase II. This nuclear precursor is then modified via

several processing steps which are nearly always required for the production of a functional mRNA. These include the addition of methylated GTP residues to the 5' end of the RNA (capping), addition of 100-200 adenosine residues (the Poly A tail) to the 3' end of the molecule and splicing in which parts of the RNA (the intervening sequences) are removed (for reviews see Darnell, 1982; Green, 1989). This messenger RNA is then transported from the nucleus to the cytoplasm where it is translated into protein. The regulation of this process can occur at any or all of the above steps and in addition by changing the stability of the various cellular intermediates (i.e. RNA half life; Darnell, 1982). The study of many tissue specific genes in a variety of systems, however, has shown that the primary level of regulation responsible for maintaining cell type specific gene control is transcription (Derman *et al.*, 1981; Tilghman and Belayew, 1982; Gillies *et al.*, 1983; Walker et al., 1983; Darnell, 1985; McKnight and Tjian, 1986; Maniatis *et al.*, 1987; Liu *et al.*, 1988; Mitchell and Tjian, 1989), and an investigation of this process is the central focus of this thesis. It should be kept in mind, however that posttranscriptional mechanisms do occur either as a primary mechanism or in combination with a variable degree of transcriptional modulation. Notable examples of this are the regulated splicing of the genes which dictate the sexual phenotype of Drosophila (Boggs *et al.*, 1987; Nagoshi *et al.*, 1988), or the gene which is involved in the production of calcitonin and its related protein CGRP in a tissue specific fashion (Amara *et al.*, 1982).

Cis-acting DNA sequence elements regulate transcription

Why are specific DNA sequences chosen for transcription in one cell-type and not another? The advent of cloning procedures and the techniques for introducing genes back into cells has enabled a functional characterization of the regulatory elements involved in dictating both the specificity and the rate of transcriptional initiation (Cassio *et al.*, 1981; Ott *et al.*, 1984; Ciliberto *et al.*, 1985; Boulet *et al.*, 1986; Friedman *et al.*, 1986,1987; Babiss *et al.*, 1987; Hay *et al.*, 1987; Grayson *et al.*, 1988a; Godbout *et al.*, 1986, 1988a). Cis-acting DNA regulatory elements have been defined which are necessary and sufficient to permit transcription of tissue specific genes in cell types in which the endogenous genes are normally expressed, and conversely to prevent their expression in inappropriate cell types. These sequences often represent DNase I hypersensitive or methylation free DNA domains, suggesting that these regions are maintained in a transcriptionally active state, with their DNA suitably decondensed and accessible to the transcriptional machinery (Stadler *et al.*, 1980; Weisbrod, 1982; Serfling *et al.*, 1985; Cedar, 1988; Morse and Simpson, 1988).

Protein transcription factors interact with DNA regulatory regions

Trans-acting proteins which bind to these cis-elements have been identified by using *in vitro* DNA binding assays (e.g. DNase I footprinting, gel shift and exonuclease III analysis), and numerous DNA-binding proteins,

many widely distributed and a few that are cell-specific (cell-enriched), are currently being studied to delineate their roles in transcriptional regulation (Briggs *et al.*, 1986; Babiss *et al.*, 1987; Bodner and Karin, 1987; Cereghini *et al.*, 1987; Courtois *et al.*, 1987; Lenardo *et al.*, 1987; Fletcher *et al.*,1987; Ko *et al.*, 1988; Herr *et al.*, 1988; Costa *et al.*, 1988b, 1989; Hai *et al.*, 1988; Ptashne, 1988). Most of the binding sites for these factors are clustered in DNA regions (enhancers or promoter-proximal sequences) that when deleted or mutated lead to a failure of gene expression: these binding proteins or groups of binding proteins are assigned a positive role in transcription (Jones *et al.*, 1987; Lee *et al.*, 1987; Mitchell *et al.*, 1987; Costa *et al.*, 1988b, 1989). In a smaller number of cases, however, DNA regions have been identified that when deleted appear to release a gene into activity or at least to allow higher expression based on the activity of the remaining positive acting regions. These elements presumably act in a negative capacity (Brand *et al.*, 1985; Goodbourn and Maniatis, 1986; Muglia and Rothman Denes, 1986; Colantuoni *et al.*, 1987; Imler *et al.*, 1987; Hay *et al.*, 1989; Kakkis *et al.*, 1989; Levine and Manley, 1989; Scheuermann *et al.*, 1989)

Evidence from the analysis of many elements has suggested that the promoter (usually found within 100-200 nucleotides of the RNA start site; Maniatis *et al.*, 1987; Dynan, 1989) and enhancer elements (which may be located far upstream, downstream or within the coding sequences of a gene (Banerji *et al.*, 1981; Fromm and Berg, 1982; Gilles *et al.*,1983; Queen and

Baltimore, 1983; McKnight and Tjian, 1986) are multifaceted genetic elements containing DNA sequence motifs (7-20bp of DNA), each of which represents a potential site of interaction with a protein transcription factor (Herr and Clark, 1986; Zenke *et al.*, 1986; Wirth *et al.*, 1987; Dynan, 1989; Mitchell and Tjian, 1989; Johnson and McKnight, 1989). A schematized picture of a typical gene regulatory region is shown in Figure 1.1.

The promoter elements of most protein coding genes, though not all, contain an A/T nucleotide rich region found 20-30 nucleotides from the start site of RNA transcription known as the TATA box, which is important for directing the first steps in the formation of an active transcription complex (Horikoshi *et al.*, 1988; Dynan, 1989; Hahn *et al.* 1989). A protein factor, TFIID, interacts with this TATA region, the importance of which is attested to by the fact that it has been functionally conserved from yeast to man. It has recently been demonstrated that the yeast TFIID can functionally complement a TFIID deficient Hela cell transcription system (Buratowski *et al.*, 1988; Cavallini *et al.*, 1988), and this protein has now been cloned (Buratowski *et al.*, 1988). This evolutionary conservation has also been demonstrated for at least one other component of the basic transcriptional machinery (TFIIA; Hahn *et al.*, 1989) as well as for the activating domains of a wide number of DNA binding proteins (Struhl, 1988; Webster *et al.*, 1988; Ptashne, 1988; Guarente, 1988). This suggests a strong conservation exists in the protein-protein interactions dictating the cellular transcriptional mechanism.

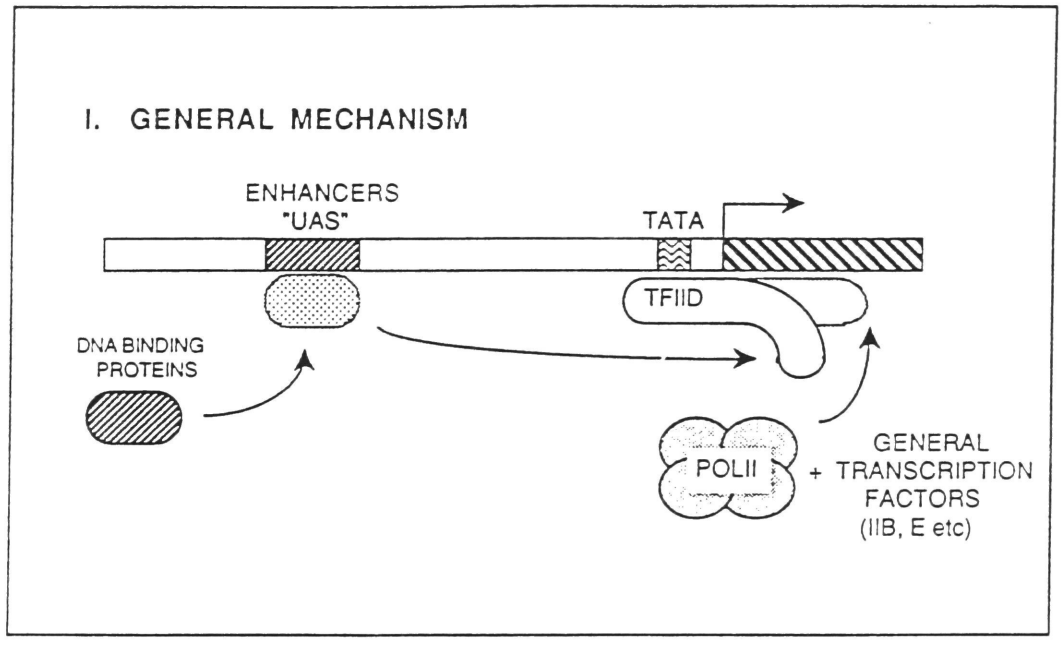


Figure 1.1. A schematized representation of a eukaryotic regulatory element depicting DNA binding proteins interacting with an upstream enhancer element. One possible activation mechanism in which upstream factors interact with the TFIID protein is suggested by this diagram and is described in the text. (from Darnell, 1989).

Mechanism of transcriptional initiation

Studies on the mechanism by which these factors influence transcription are still in their infancy but several models have been developed. One current theory suggests these upstream sequence elements and their cognate proteins increase the functional TFIID binding to the promoter proximal sequences, thereby facilitating transcription (Figure 1.1; Ptashne, 1988; Horikoshi *et al.*, 1988; Nakajima *et al.*, 1988). This model is based on the *in vitro* observation that numerous core promoter elements (containing the TATA box and the cap site alone) are capable of activating transcription in HeLa cell nuclear extracts (Horikoshi *et al.*, 1988). In the cell, however, this is not the case, most probably because TFIID is present in a limited concentration relative to the number of available DNA templates. It therefore has been suggested that upstream sequence elements and their associate factors function by assisting TFIID in the formation of the initiation complex (Horikoshi *et al.*, 1988). These factors could either increase the affinity of TFIID binding to the TATA box or alternatively modify the activity of the protein in such a way as to affect its interaction with the DNA or any of the other components of the initiation complex (Simon *et al.*, 1988; Mitchell and Tjian, 1989). A recent study by Horikoshi *et al.* (1988), demonstrated the latter to be the case for the adenovirus E4 gene. They showed that the upstream transcription factor ATF, qualitatively influences the binding of the TFIID factor upon the E4 promoter proximal sequences (increasing the region

protected from DNase I digestion; Figure 1.1; Horikoski *et al.*, 1988, Van Dyke *et al.*, 1988). Furthermore, this change correlated with increased *in vitro* transcriptional activity, strongly implicating this interaction as a mechanism involved in transcriptional activation (Horikoshi *et al.*, 1988). This, however, represents only one example of an upstream factor interaction, and mounting evidence suggests that these upstream factors might also interact with the large RNA polymerase molecule itself, or any of the at least five associated protein components of the basic transcription complex (Matsui *et al.*, 1980; Samuels *et al.*, 1982; Davison *et al.*, 1983; Reinberg and Roeder, 1987; Mitchell and Tjian, 1989).

For the more upstream DNA regulatory regions, models which include DNA bending and looping have been evoked to explain their ability to interact near the RNA start site (Ptashne, 1986; Schleif, 1986), and studies of the protein-protein interactions surely involved in regulating these effects are being performed for a number of systems (Ptashne, 1988). In any case, it appears that the modular organization of these regulatory elements allows transcriptional regulation to occur in response to a diverse group of cellular signals. Although the interplay between the various proteins is not yet fully understood, it is believed that combinations of cis acting elements arranged in unique configurations confer upon each gene an individualized transcriptional program (Mitchell and Tjian, 1989).

Sequence-specific DNA binding proteins

These transcriptional proteins can be purified based upon their strong affinity for specific DNA binding sites. (Briggs *et al.*, 1986; Mitchell *et al.*, 1987; Scheidereit *et al.*, 1987; Kawakami *et al.*, 1988). Protein sequence is then obtained and these genes can be cloned by preparing oligonucleotide probes or antibodies for use in screening cDNA libraries. Additionally, in recent years several of these proteins have been cloned using the technique of Singh *et al.* (1988) which employs the DNA binding sequence to directly screen λgt11 cDNA expression libraries. Examples of proteins cloned in this way include the H2TF1 and ATF transcription factors (Singh *et al.*, 1988; Hai *et al.*, 1989).

Once cloned, the DNA binding and activating domains of the protein can be functionally dissected. From recent studies is has become evident that several distinctive protein structures are often associated with the DNA binding proteins such as the homeodomain, POU domain, zinc finger, and leucine zipper (reviewed in Mitchell and Tjian, 1989; Johnson and McKnight, 1989). Furthermore, common activating regions are emerging as well, (reviewed by Ptashne 1988; Mitchell and Tjian, 1989), usually consisting of specific classes of amino acids (ie. acidic, glutamine or proline rich groups).

Regulation of transcription factors

In light of their importance in activating transcription, the cellular control of these proteins certainly represents an important developmental

decision. The study of a large number of these factors has demonstrated that they can either be constitutively present (Briggs *et al.*, 1986; Jones *et al.*, 1987; Mitchell *et al.*, 1987; Chodosh *et al.*, 1986; Mermod *et al.*, 1988), tissue specific (Bodner and Karin, 1987; Scheidereit *et al.*, 1987; Kawakami *et al.*, 1988; Frain *et al.*, 1989) or inducible in their distribution (Fletcher *et al.*, 1987; Wu *et al.*, 1987; Lenardo and Baltimore, 1989). Also, it has become apparent in recent years that many of these transcription factors require some specific modifications to become functionally active. For example certain factors have been shown to be glycosylated (Jackson and Tjian, 1988), and included in this group is the ubiquitous factor SP1. This glycosylation is not required for DNA binding, and therefore, the ability to activate transcription could possibly be regulated by the extent of this modification in a specific cell type (Mitchell and Tjian, 1989). Another example of an activated factor is that responsible for transducing the signal for heat shock in yeast. This factor will bind DNA in all cells but must be phosphorylated to be functionally active (Sorger *et al.*, 1987). Conversely, while the above two factors will bind DNA in the absence of modification, the adenovirus E4f factor requires modification of its precursor (by E1A) to efficiently bind DNA and thereby exert its transcriptional effect (Raychaudhuri *et al.*, 1988).

Additionally, a group of factors has been described which are present in the cytoplasm of all cells but in an quiescent form (ie NFκB; Lenardo and Baltimore, 1989). In response to an extracellular signal (LPS) the NFκB

disassociates from its inactivator Iκb in the cytoplasm, assumes its active configuration and is transported to the nucleus. The steroid hormone receptor family functions in a like fashion (Beato, 1989).

Lastly, some activating proteins such as the adenoviral E1A protein or the VP16 protein of herpes virus do not bind to DNA directly but rather activate transcription by interacting with or modifying other proteins (which bind the DNA; Kovesdi *et al.*, 1986; Raychaudhuri *et al.*, 1988; Simon *et al.*, 1988; Triezenberg *et al.*, 1988). Hence, the regulation of the transcription factors can occur in several ways, and it is not merely the presence or absence of a factor, but rather its correct state of modification and the extent of its protein-protein interactions which determines its ultimate transcriptional activity.

How do these factors generate the tissue specific transcriptional profile of a differentiated tissue? Are specific factors involved in each process, or do these genes make use of unique combinations of common cellular factors. Furthermore, how are these factors themselves regulated during development. The approach to these questions calls for the study of a group of genes all active in a specific cell type, from which an analysis of regulatory regions and an identification of the transcriptional proteins might reveal important elements involved in their control. These questions can be approached by studying coordinate expression of a set of genes in hepatocytes, the cell-type that carries out the differentiated functions of the liver.

II. Liver gene expression

Advantages of using the mammalian liver to study development

The liver is the second largest organ (the skin is the largest), and due to its large size, the developing liver can be studied at an early stage of growth, for it always represents a good source of cellular material, whether it be DNA, RNA or protein. The liver constitutes approximately 1-2 percent of the weight of adult animals and as much as 10 percent of the weight of the fetus. (Doljanski, 1960; Jones and Spring-Mills, 1983) Furthermore, the liver is a homogenous tissue composed almost entirely of one parenchymal cell type, the hepatocyte. About 50% of the nuclei in liver are hepatocytes and due to their large size they constitute as much as 80-90% of the volume of the liver (Figure 1.2; Greengard *et. al.*, 1972). In addition, since most hepatocytes are tetraploid, perhaps as much as 90% of the transcription in the liver occurs in hepatocytes(Greengard *et al.*, 1972).

The liver is the center for a multitude of endocrine and exocrine functions and hepatocytes are intricately involved in the maintenance of normal metabolism, detoxification, and in the production of many other tissue specific enzymes and serum proteins of major physiological importance (Greengard, 1969; Jungerman and Katz, 1982). Perhaps more cell specific products are synthesized by hepatocytes than any other cell type (Darnell, 1985). Since many genes which encode these cellular proteins are expressed at high rates, a large number have been cloned and analyzed with

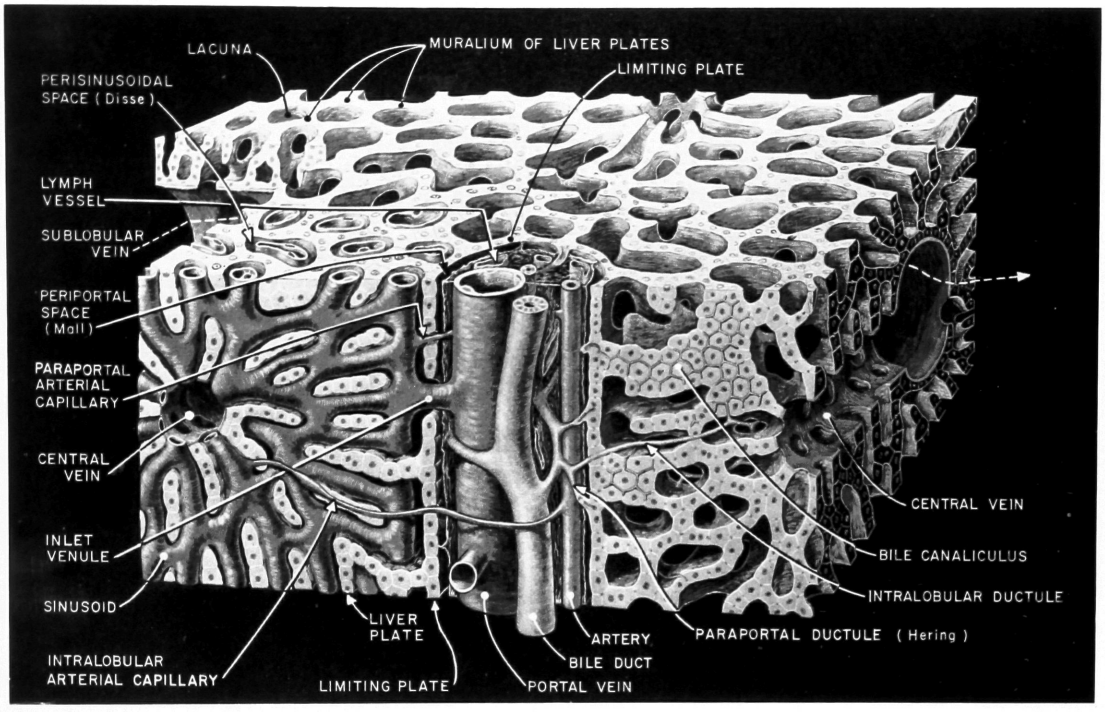
respect to their expression during hepatocyte development. This allows an analysis of the process by which a set of cell-type specific genes are expressed coordinately in a single cell type.

A schematized representation of the adult liver is shown in Figure 1.2, depicting cords of hepatocytes juxtaposed to their blood supply. The liver lobule has been defined to consist of small functional units termed acini (Rappaport, 1976; Jones and Spring-Mills, 1983). These acini are divided into three zones based upon the hepatic microcirculation, and it is now known that several genes in the liver are transcriptionally regulated in a position dependent fashion (e.g. glutamine synthetase; glutathione-s-transferase; Kuo *et al.*, 1988). The serum albumin gene, however, which represents the focus of these studies, is transcribed by all hepatocytes (Bennett *et al.*, 1987).

General approaches to the study of hepatocyte gene expression

To approach the question of hepatocyte-specific gene expression during development and hepatocyte differentiation, several groups isolated cDNA clones representing abundant liver gene mRNAs (Derman *et. al.*, 1981; Tilghman and Belayew, 1982; Cistanzom *et. al.*, 1983). These clones were then used as molecular hybridization probes for the quantitation of mRNA accumulation and an analysis of gene transcription rates in both liver and non-liver tissues using the *in vitro* nuclear run-on assay. This assay involves isolating nuclei from various tissues, and elongating previously initiated

Figure 1.2. A schematized representation of the morphology of the mammalian liver. Depicted are hepatocytes, the major cell type in the liver, juxtaposed to their blood supply. Structures are as marked. (From Elias and Sherrick, 1968).


LACUNA
MURALIUM OF LIVER PLATES
LIMITING PLATE
PERISINUSOIDAL SPACE (Disse)
LYMPH VESSEL
SUBLOBULAR VEIN
PERIPORTAL SPACE (Mall)
PARAPORTAL ARTERIAL CAPILLARY
CENTRAL VEIN
INLET VENULE
SINUSOID
INTRALOBULAR ARTERIAL CAPILLARY
LIVER PLATE
LIMITING PLATE
ARTERY
BILE DUCT
PORTAL VEIN
CENTRAL VEIN
BILE CANALICULUS
INTRALOBULAR DUCTULE
PARAPORTAL DUCTULE (Hering)

RNA polymerase II transcripts in the presence of labeled nucleotides. Since no new initiation reactions occur, this method allows a determination of the relative rates of transcription for any gene in a given cell-type (Weber *et al.*, 1977; Hofer and Darnell, 1981). The transcriptional signals for a group of the cDNA clones (a dozen or so) studied were higher in liver nuclei as compared to brain or kidney nuclei while the relative rates of transcription for clones representing genes transcribed in most all cell types (e.g. actin and tubulin) were similar (Derman *et al.*, 1981; Powell *et al.*, 1984).

From these results, several basic conclusions have emerged. First, it has become evident that two general groups of liver specific transcripts exist: those expressed exclusively in the liver and others preferentially expressed in the liver (though at 100-1000x greater levels; Derman *et. al.*, 1981; Powell *et al.*, 1984; Darnell, 1985). Furthermore, the major level of regulation for many of these genes appears to be transcriptional initiation, for the transcriptional decrease in non-liver tissues completely accounted for the lower levels of accumulated RNA (Derman *et. al.*, 1981; Tilghman *et. al.*, 1982; Cistanzom *et. al.*, 1983; Powell *et. al.*, 1984; Panduro *et. al.*, 1987a). Interestingly, for the common cellular genes analyzed (actin and tubulin) a major component of their regulation appeared to occur posttranscriptionally, since their levels of cytoplasmic RNA were found to be higher than could be accounted for by the transcription rate alone (Powell *et al.*, 1984) The transcription rate of albumin has been evaluated in this fashion in both liver and non-liver cell types

(Table 1.1; Tilghman and Belayew, 1982; Derman *et al.*, 1981; Powell *et al.*, 1984; Liu *et al.*, 1988). Albumin is among the most actively transcribed liver specific genes.

Developmental Regulation of liver gene expression

Many of these liver-specific clones were subsequently used to measure the relative rates of transcription and steady state mRNA accumulation at different developmental stages (Tilghman and Belayew, 1982; Powell *et al.*, 1984; Panduro *et al.*, 1987a). Interestingly, two of the most abundant liver specific mRNAs, albumin and α-fetoprotein, are developmentally regulated in a reciprocal fashion. Figure 1.3 shows a time course of albumin and α-fetoprotein messenger RNA accumulation during mouse development (Tilghman and Belayew, 1982). Both genes are transcriptionally activated at prenatal day 16 with α-fetoprotein expression quickly reaching its maximal level while albumin expression increases more slowly. At birth, α-fetoprotein levels rapidly decrease while albumin steadily increases to its maximum level which is obtained at two weeks of age. At that point, many other "adult" liver genes are expressed at their maximal rates including contrapsin, tyrosine aminotransferase, and cytochrome P450 (Panduro *et al.*, 1987a), along with the genes encoding proteins involved in the pathway of gluconeogenesis. These observations (for albumin and α-fetoprotein) have

Liver	100%
HepG2	10%
Fao	10%
CWSV1	50%
C2	-
C2rev7	0.5%
Hela	-
Brain	-
Kidney	-

Table 1.1. Relative albumin transcription rates in different cell types. Nuclear run on transcription analysis was used to measure the relative albumin transcription rates in rat and human hepatoma cell lines and tissues (as described in the text).

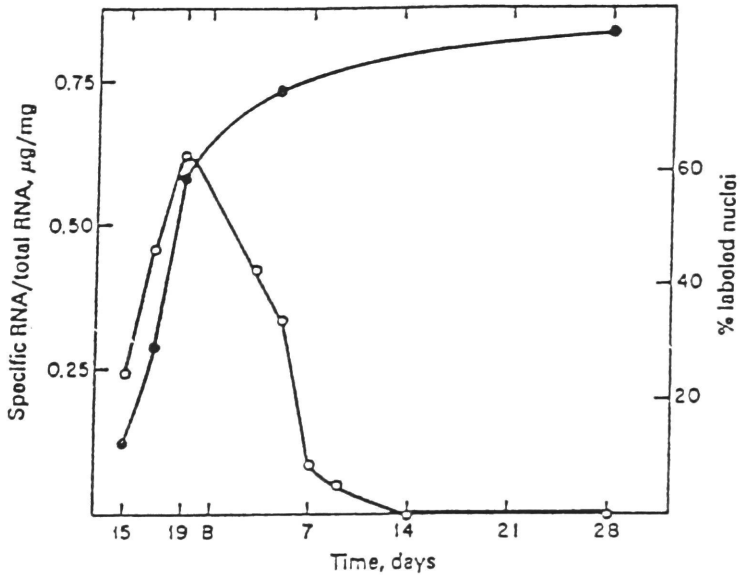


Figure 1.3. Developmental regulation of albumin and α-fetoprotein mRNA levels. At various stages of mouse development, livers were excised, RNA was purified and hybridized to filters containing the albumin (filled-in circles) and AFP (open circles) cDNA clones. These values were then plotted (vertical axis) as a function of the time point in liver development (horizontal axis). (From Tilghman and Belayew, 1982).

been confirmed at the level of transcription (Powell *et al.*, 1984.)

Interestingly, the decrease in α-fetoprotein mRNA following birth coincides with the cessation of cellular DNA replication (which is a characteristic of terminally differentiated hepatocytes), as measured by thymidine incorporation (Tilghman and Belayew, 1982). In fact, the relative levels of albumin and α-fetoprotein is often used as a standard for defining the state of differentiation of hepatocytes. Alphafetoprotein has been shown to increase by as much as 1000 fold in hepatomas, carcinomas, embryonal germ cell tumors, and following liver injury, suggesting that the previously differentiated hepatocytes have dedifferentiated to a more fetal stage (Petropoulos *et al.*, 1983; Panduro *et al.*, 1987b). In many of these situations, the transcription rate of albumin either decreases or stays the same, depending on the exact nature of the inducing stimulus (Petropoulous *et al.*, 1983; Panduro *et al.*, 1986, 1987a). Hence, a high rate of transcription of albumin, in the absence of α-fetoprotein, is characteristic of the end stage, differentiated hepatocyte.

III. *In vitro* systems for the study of hepatocyte development and differentiation

Primary hepatocyte cell culture

Liver differentiation and hepatic gene expression can be studied using a variety of *in vitro* systems. The most obvious choice would be to maintain primary hepatocytes in culture and to use these cells to analyze the regulation of expression of the endogenous and newly introduced liver preferential genes (by DNA transfection or recombinant virus infection). It is possible to perfuse, disaggregate and culture hepatocytes from rodent liver or human surgical biopsies (Clayton and Darnell, 1983; Reid and Jefferson, 1984), using collagenase and the chelating agent EDTA to disrupt the junctions between neighboring hepatocytes. However, under standard serum supplemented culture conditions these primary cells rapidly loose their hepatocyte morphology and with increasing time in culture, do not transcribe their liver-specific genes (Clayton and Darnell, 1983; Reid and Jefferson, 1984). Freshly plated hepatocytes (0-2h in culture in plastic dishes) maintain high rates of liver gene expression, comparable to liver hepatocytes. In contrast, at 24h the rate of transcription of the liver specific, but not common genes (actin and tubulin), decreases dramatically (1-10%) compared to the levels observed in freshly excised liver tissue (Clayton and Darnell, 1983). The liver-specific mRNA levels persist significantly longer (2-5 days), eventually declining as the tissue specific character is lost (Clayton and Darnell, 1983; Jefferson *et al.*,

1984). This suggests a preferential stability of these liver mRNAs which persist in culture. Since the transcriptional rates of this entire series of genes is dramatically lowered in a coordinate fashion, the involvement of a general signal is implied. It has been suggested that certain substances such as hormones, nutrients and metabolic products, present *in situ* are lacking in the culture system (Clayton *et al.*, 1985a, 1985b). Alternatively hepatocytes might need their normal cell-cell contacts as a signal which enables them to coordinately activate a group of characteristic functions (Reid and Jefferson, 1984; Clayton *et al.*, 1985b). This is suggested strongly by work on tissue sections which shows that liver specific gene transcription is maintained for much longer periods of time (Clayton *et al.*, 1985b). Certain experimental systems have tried to extend the period of liver specific gene expression by better mimicking the *in vivo* situation (ie, culture on collagen, media with DMSO) with limited success (Reid and Jefferson, 1984). What these studies do suggest; however, is that the expression of a wide variety of unrelated liver-specific genes is probably controlled by a limited number of liver-specific transcription factors. Therefore, after disaggregation of hepatocytes, the lack of production of any of these regulatory factors would affect the expression of many related genes.

Hepatoma cell lines

Numerous hepatoma cell lines have been isolated as tumor explants from both rodent and human liver (Reuber, 1961; Knowles *et al.*, 1980). Many

clonal cell lines established from these tumors either do not express specific liver genes or more often express specific liver genes at diminished rates (See Table 1.1). This most probably is a reflection of the state of cellular differentiation of the individual hepatoma cell lines. This variability can be both advantageous and disadvantageous from an experimental point of view. For example, the albumin gene is transcribed in human HepG2 cells at only 10% the rate observed in adult liver hepatocytes (Table 1.1; Clayton *et al.*, 1985a; Friedman *et al.*, 1986). While HepG2 cells are easy to maintain and are amenable to transient expression assays, the lack of a fully differentiated hepatocyte phenotype might lead to an inability to detect all levels of liver-gene regulation (particulary the regulatory events responsible for maintaining the rate of transcription of any liver-specific gene). Despite this drawback, however, these cells have been used to characterize the expression of the tissue specific albumin promoter element which I describe in this thesis (Friedman *et al.*, 1986; Babiss *et al.*, 1987) and those of other liver specific genes expressed in hepatocytes including the transthyretin (TTR; Costa *et al.*, 1986; Costa *et al.*, 1988a), α1-antitrypsin (Ciliberto *et al.*, 1985) and α-fetoprotein genes (Godbout *et al.*, 1986, 1988a), each maximally expressed at distinct developmental stages for which the hepatoma cell represents a suitable but non-optimal cell choice.

However, this transcriptional variability can become a benefit since each tumor cell line represents a fixed genetically stable stage of hepatocyte

development. For example, the FaO cell line (Deschatrette *et al.*, 1974; Deschatrette *et al.*, 1980) has been shown by several groups to transcribe the albumin gene at only 10- 20% the rate observed for intact hepatocytes (Cassio *et al.*, 1981; Clayton *et al.*, 1985a; Friedman *et al.*, 1987), suggesting it is only partly maintaining the differentiated liver phenotype. A subclone of FaO cell, however, the C2 cell line (Deschatrette *et al.*, 1974; Deschatrette *et al.*, 1980), was subsequently selected based upon its inability to divide without glucose (this being indicative of an absence of the hepatocyte specific gluconeogenic enzymes; Bertolotti *et al.*, 1987). These cells express few if any liver specific functions (determined at the level of transcription), and are considered to be dedifferentiated (Deschatrette *et al.*, 1987; Clayton *et al.*, 1985a).

Additionally, a C2 revertant (C2rev7), was selected based on its ability to again grow in glucose free medium (indicating reexpression of liver specific gluconeogenic enzymes), and it accumulates cytoplasmic albumin mRNA to the same levels observed in the FaO cell line, but unlike FaO cells, transcribes the albumin gene at barely detectable levels (Friedman *et al.*, 1987). Therefore, although this cell line is synthesizing albumin, it has not fully reverted back to the same state of cellular differentiation as the FaO cell line, for its recovery of albumin gene expression is based mainly on post-transcriptional events (Friedman *et al.*, 1987). Hence, as will be shown in chapter 8, these stable cell lines can be used to correlate the levels of the different transcription

regulatory factors with the transcriptional rate of albumin (and for that matter other liver specific genes), thereby evaluating their biological roles.

Furthermore, additional experiments using somatic cell hybridization techniques have shown that when hepatoma cells are fused with a variety of other cell types, liver-specific functions are most always "extinguished" (Weiss and Chaplain, 1971; Fougere and Weiss, 1978; Mevel-Ninio and Weiss, 1981; Killary and Fournier, 1984). This suggests that diffusible factors are involved in the coordinate control of liver-specific gene expression, and implys that negative regulation is involved. This will be discussed in chapter 7, when a negative element present in the albumin enhancer is described.

CWSV1 Cells

Immortalization of adult hepatocytes with DNA tumor viruses (Adenovirus and SV40), has recently lead to the isolation of several cell lines which express liver specific genes, including albumin, at levels approaching that of normal adult liver cells (Table 1.1; Woodworth *et al.*, 1986; Woodworth and Isom, 1987; Zaret *et al.*, 1988). Like the hepatoma cell lines; however, it appears as though the rate of albumin transcription cannot fully account for the levels of cytoplasmic albumin mRNA observed, again suggesting a change in post-transcriptional regulation exists (Woodworth *et al.*, 1986). Furthermore, since some of these cell lines (ie. CWSV1) are maintained in a chemically defined medium, a more precise analysis of

paracrine substances (ie hormones and growth factors) which may have significant effects on hepatocyte differentiation can be measured experimentally.

CWSV1 cells are SV40 immortalized rat hepatocytes which have been selected for on the basis of albumin production, and they transcribe albumin at 50% the rate of the intact rat liver (Woodworth *et al.*, 1986). Furthermore, like hepatoma cells, these cell lines can be used in transient expression assays, and they have been employed in this thesis to investigate the albumin gene enhancer. The fact that they have the highest level of endogenous albumin transcription of any continuous cell line makes them a good choice for study in cases where rate enhancement is under investigation.

Transgenic mice

Transgenic mice have been used to study gene expression during development in different tissues (Palmiter and Brinster, 1986). By introducing a transgene containing specific amounts of upstream regulatory sequences into the developing mouse embryo, it is possible to assay the expression of this gene in tissues throughout development. This method has been used to define several cis-acting DNA elements that contribute to both the specificity and rate of liver-gene expression, including the α1-antitrypsin, albumin and α-fetoprotein genes (Hammer *et al.*, 1987; Pinkert *et al.*, 1987; Camper and Tilghman, 1989). While time-consuming and technically difficult to prepare,

the major advantage of this approach is that the gene of interest can be studied in the context of chromatin structure, which as already discussed represents an important level of gene control. While corrections must be made for copy number (multiple copies of the endogenous gene might be present) and chromosomal location considered, this approach obviates many problems associated with the hepatoma cell studies, including the potential identification of cis-elements that contribute to transcriptional rates.

In Vitro Transcription

Liver specific gene transcription can also be achieved using cell free nuclear extracts (Gorski *et al.*, 1986). Transcriptionally active liver extracts can be made which transcribe liver gene templates more efficiently than non-liver extracts. The studies of Gorski *et al.* demonstrated that the mouse albumin gene is transcribed at a higher rate in extracts prepared from liver, than in brain or spleen extracts. Similar *in vitro* liver specific transcription has since been observed by investigators using other liver specific genes, and the regulatory regions determined (via deletional analysis) correspond well with those uncovered by transient analyses (Cereghini *et al.*, 1988; Lichtsteiner *et al.*, 1987; Monaci *et al.*, 1988). Now that multiple liver specific factors are both purified and cloned, this *in vitro* method will allow the analysis of a reconstituted liver gene transcription system, in which the mechanistic interactions of the separate components can be precisely analyzed.

IV. The serum albumin Gene

Serum albumin is involved in the maintenance of osmotic pressure and in the transport of many substances throughout the bloodstream (Rothschild *et al.*, 1988.) The substances transported by albumin vary widely from hormones to fatty acids, toxic waste products, bilirubin, fatty acids and exogenous drugs. Serum albumin is known to have a single, unmodified peptide chain of 580 to 585 amino acids, depending on the species, and its secondary structure has a unique pattern of double loops held together by disulfide bridges. Albumin protein levels in the serum are maximal following birth (Rothschild *et al.*, 1988).

Albumin is very similar to the α–fetoprotein gene with respect to genomic structure and protein function. The proteins exhibit nearly identical ligand binding properties and have significant amino acid sequence homologies, suggesting that they might have arisen because of a gene duplication event early in evolution (Gorin *et al.*, 1981; Jagodzinski *et al.*, 1981; Kioussis *et al.*, 1981; Law and Dugaczyk, 1981; Ingram *et al.*, 1982). Furthermore, an even greater homology is observed between their DNA sequences especially in the promoter regions which interact with crucial regulatory proteins. Both genes are found on the same chromosome (chromosome 5, 14, and 4 in mouse, rat and human), with the albumin gene located 5' to the α–fetoprotein gene, separated by 13-14kbp (Figure 1.4 Tilighman and Belawew, 1982; Gorin *et al.*, 1981; Ingram *et al.*, 1982; Urano *et*

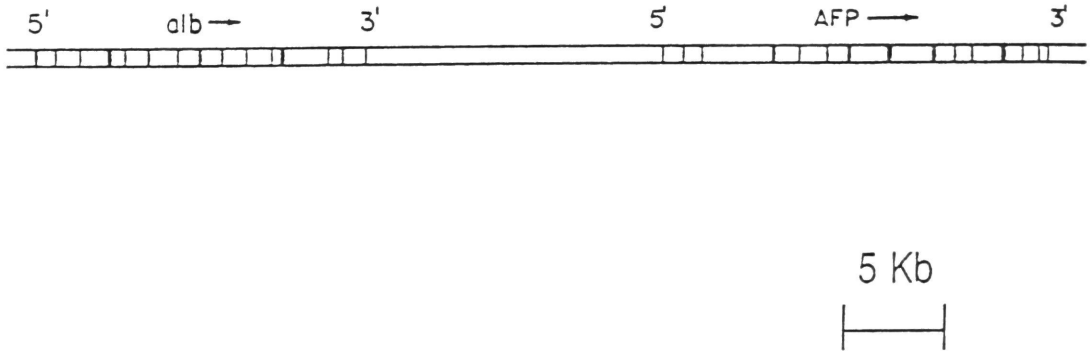


Figure 1.4. Genomic structure of the albumin and α-fetoprotein genes. These genes are very similar with respect to their genomic structure and protein functions. Both genes are found on the same chromosome (chromosome 5, 14, and 4 in mouse, rat and human, respectively), with the albumin gene located 5' to the α-fetoprotein gene and separated by 13-14kbp.

al., 1984; Gal *et al.*, 1985; Tilghman, 1985, Nahon, 1987). Each is transcribed with the same polarity, so it is possible that albumin transcription could extend into the α–fetoprotein gene domain. Further, the mRNAs encoded by both genes are very similar with respect to the complexity of their exon/intron organization. However, as stated earlier, the expression of these genes is regulated in a reciprocal manner; this happening at the level of transcriptional initiation (Tilghman and Belayew, 1982; Powell *et al.*, 1984; Nahon *et al.*, 1987).

Transcriptional regulation of albumin

The serum albumin gene has been cloned and analyzed from many species. The rat gene is composed of 35kbp of contiguous DNA sequence containing 14 exons and 13 introns (Sargent *et al.*, 1979), and the mouse clone is quite similar (Tilghman and Belayew, 1982; Tilghman, 1985). A strong sequence conservation exists 5' to the capsite of the rat, mouse, chicken and human albumin genes, extending to 250bp upstream, suggesting functional importance for these sequences, and the factors which interact with them (Figure 1.5; Cereghini *et al.*, 1987). As will become evident in chapter 3 of this thesis, this was proven to be the case.

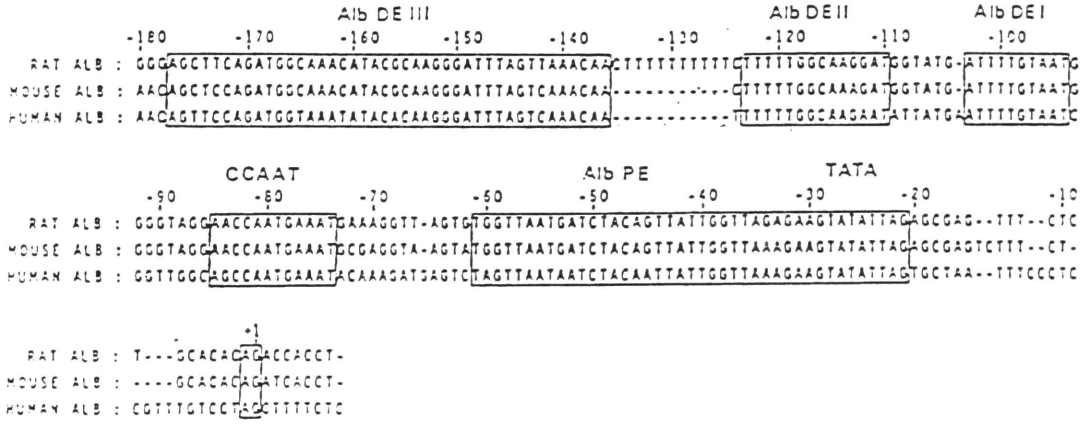


Figure 1.5. Sequence homology between the 5' upstream promoter regions of the rat, mouse and human albumin genes. The regions of greatest homology are boxed, and these conserved elements will be shown in chapter 3 to represent sites for binding of regulatory factors. (from Cereghini *et al.*, 1987).

Albumin chromatin structure

Perturbations in the chromatin structure are usually associated with regions of active transcription. It is currently believed that DNase I hypersensitivity indicates DNA sequences upon which the interaction with histones is affected in such a way as to facilitate the accessibility to the transcriptional machinery (Morse and Simpson, 1988). This could be the result of either the complete absence of nucleosomes or alternatively might reflect changes in the composition, conformation or the extent of interactions between different histone proteins (Stadler, 1980; Wu, 1980; Weisbrod, 1982; Roche *et al.*, 1985; Morse and Simson, 1988; Ceder, 1988). Furthermore, these regions of hypersensitivity often contain identified cis-acting regulatory sequences, with corresponding cellular transcription factors bound to them (Jongstra *et al.*, 1984; Zaret *et al.*, 1984), and recent *in vitro* experiments have shown that the formation and stabilization of transcription initiation complexes with sequence specific factors can competitively exclude histones and prevent the inhibitory effect of nucleosomes (Workman and Roeder, 1987). This latter result strongly suggests it is the presence of specific transcription factors at the time of DNA replication that might designate chromosomal regions for potential transcriptional activity. In this way, a gene is committed to the active state and the information contained in the chromosome stably transmitted (Brown, 1984).

Additionally, in many cases methylation has been observed to correlate

with tissue specific gene expression, and treatment of cells with azacytidine, a potent demethylating agent, activates a host of cellular genes (Davis *et al.*, 1987; Ceder, 1988). Methylation therefore represents another mechanism by which genetic information can be passed along during cellular replication, for that is the time when the DNA is methylated. For albumin, it has been shown that decreased methylation correlates with its increased expression as shown in hepatoma cell lines, their variants and extinguished hybrids (Orlofsky and Chasin, 1985). All these studies demonstrate that chromatin organization is a fundamental level of gene control, which in concert with the transcription factors, works to create the potential for transcriptional activation.

Therefore, as an early approach to identifying DNA regions which might contribute to the specificity and rate of albumin gene expression, chromatin structural studies were performed on the rat albumin gene upstream sequences. For this study, DNA was analyzed from cells which did not express albumin [Hela and C2 cells, spleen, kidney, and brain tissues and late cultured (for 24-96h) hepatocytes], express intermediate levels of albumin [FaO and C2rev7 cells, and early cultured (2-24h) hepatocytes] or express the albumin gene at a high rates [liver tissue]. The nuclei from these different cell types were prepared and DNase I hypersensitive sites were mapped (Babiss *et al.*, 1986).

The DNase I hypersensitivity pattern of the rat albumin gene sequences

between -7.3kbp to +2.26kbp (relative to the start site of albumin transcription) was first examined in liver and kidney cells and in cells (ie. early and late primary hepatocytes, hepatoma cell lines) which express albumin to varying extents (Babiss *et al.*, 1986). The major conclusion of this work was that three hypersensitive sites were present in liver nuclei at roughly –0.05, –0.2 and –2.8kbp relative to the start site of albumin transcription, which were liver specific. In cultured hepatocytes (24-96h in culture) or hepatoma cells (FaO), where albumin transcription is 10 - 20% of that in the intact liver, the two more promoter proximal hypersensitive sites were present while the –2.8kbp site was lacking. Therefore, it was reasoned that the hypersensitive sites between -0.05 and -0.2 kbp were involved with the tissue-specific expression of the albumin promoter while the hypersensitive site at -2.8kbp was necessary to maintain high rates of albumin expression. The C2 hepatoma cell line, which does not express the albumin gene, had a unique hypersensitive sites at -1.5kbp, and this site was absent in the C2rev7 cell line, which regained albumin promoter activity. It is still unclear whether this site is responsible for the suppression of albumin gene expression in the C2 cell line.

Liu *et al.* (1988) conducted a similar type of study for the mouse albumin gene, extending their analysis as far upstream as 15kbp and as far downstream as +16.5kbp. They found many hypersensitive sites present only in the liver, including sites at -0.1kbp, -3.5kbp -10.8kbp and -13.7kbp. Pinkert *et al.* also described hypersensitive sites in this region (one at -10.4kbp; Pinkert

et al., 1987). The presence of these multiple hypersensitive sites implicates both the promoter proximal and far upstream sequences for involvement in the regulation of albumin expression.

The albumin promoter

The albumin promoter proximal sequences have been studied by numerous groups and shown to be necessary for tissue-specific expression of the albumin gene. This was initially analyzed both by plasmid transfection studies using hepatoma cells (Ott *et al.*, 1984; Heard *et al.*, 1987), or by using non-replicating adenovirus vectors containing the rat albumin gene sequences extending from -441 to +957 (exons Z and A) and infecting human (HepG2) and rat (FaO) hepatoma cells (Friedman *et al.*, 1986). The studies of Friedman *et al.*, using the viral vector system, are the basis for the albumin promoter analysis which is described in chapter 3. It was shown that the albumin containing viruses direct expression in HepG2 cells, but not in mouse myeloma cells. No such activation in HepG2 cells was observed by the mouse immunoglobulin and mouse β-globin promoters also present on adenoviral genomes, the former of which (with its cell-specific enhancer) was active in myeloma cells (Friedman *et al.*, 1986).

When the rate of albumin-viral (exogenous) transcription and endogenous HepG2 cell albumin transcription were compared, however, that from the viral albumin promoter was lower (Friedman *et al.*, 1986). Therefore

while the promoter proximal element was sufficient for maintaining albumin tissue-specificity, it did not fully reconstitute the rate of albumin expression. This along with the chromatin structural analyses presented above suggested that additional sequences might be important in this regard.

V. Transcription of additional liver specific genes - coordinate gene control

As already stated, a basic understanding of the mechanism(s) by which a wide variety of unrelated genes are coordinately expressed is the liver is a long-range goal. From studies on the regulatory regions of several genes transcribed preferentially in the liver (ie. Transthyretin, α1-antitrypsin and of course albumin), four liver specific (or enriched) transcription factors have been extensively described. These factors (HNF1, C/EBP, HNF3 and HNF4) will be briefly introduced here. However, since all but HNF4 interact with the albumin regulatory sequences, they will be encountered again in later chapters of this thesis.

HNF1

The HNF1 factor was first described by Courtois *et al.*, (1987) and later shown to bind to the promoter proximal sequences of a large number of liver-specific genes including, albumin, α1-antitrypsin, β-fibrinogen, TTR and α-fetoprotein (Courtois *et al.*, 1988; Hardon *et al.*, 1988; Monaci *et al.*, 1988; Feuerman *et al.*, 1989), all of which represent promoters highly active in

hepatocytes. In each case the deletion of the HNF1 binding site produces a large decrease in transcription. Since this protein is present almost exclusively in the liver (though trace amounts are also present in the kidney), it was named hepatocyte nuclear factor 1 (HNF1) by Courtois *et al.* (1987).

The transcriptional role of the HNF1 factor has also been demonstrated *in vitro* where depletion of rat liver nuclear extract of HNF1 by oligonucleotide competition, lead to a specific decrease in albumin promoter function (Cereghini *et al.*, 1988; Lichtsteiner and Schibler, 1989). Additionally, a nuclear extract derived from spleen nuclei can become competent to support albumin gene transcription solely upon the addition of purified HNF1 protein, attesting to the vital role this protein must play in conferring the liver specific phenotype (Lichtsteiner and Schibler, 1989).

HNF1 binds a palindromic consensus sequence containing two 6bp inverted repeats separated by a single nucleotide. It has been purified and shown to be a glycosylated protein with a molecular weight of 80-90kd (Courtois *et al.*, 1988, Frain *et al.*, 1989; Lichtsteiner and Schibler, 1989). Interestingly, a faster migrating variant (as determined by gel-shift analysis) can be detected in dedifferentiated hepatoma cell lines that do not express any liver specific functions, though the origin (same or different gene) and exact role of this product is unclear (Baumheuter *et al.*, 1988; Cereghini *et al.*, 1988.)

The HNF1 factor has now been cloned (Frain *et al.*, 1989) and sequence analysis reveals it to belong to the homeodomain set of proteins (Ghering,

1987), and it contains homology to the recently described POU subfamily of binding proteins (Herr *et al.*, 1988; Frain *et al.*, 1989). HNF1 is a highly liver specific protein and its mRNA seems to be restricted to the liver (Frain *et al.*, 1989). Preliminary transcriptional analysis, however, using a cloned cDNA has detected transcription in several different mouse tissues (Xanthopoulos *et al.*, unpublished observations). This would suggest that a posttranscriptional component is involved in HNF1 gene regulation.

C/EBP

The C/EBP protein was first identified by virtue of its binding to the LTR of Murine Sarcoma virus and the CAAT box of the Herpes Virus TK gene (Graves *et al.*, 1986). It was purified from rat liver (Johnson *et al.*, 1987) and demonstrated to interact with several enhancer and promoter elements (SV40 enhancer, α1-antitrypsin and TTR enhancer and promoter; Costa *et al.*, 1988b), as well as to CAAT box motifs, and hence, its name C/EBP (CAAT/Enhancer Binding Protein). The C/EBP gene has been cloned from the rat and mouse (Landshulz *et al.*, 1988a; Xanthopoulos *et al.*, 1989), and has been found to be an intronless gene possessing the leucine zipper motif, which is indicative of the ability of this factor to interact with DNA sequences as a protein dimer (Landshulz *et al.*, 1988b).

C/EBP is a liver enriched factor and it is present to lesser amounts in fat and kidney tissue (Birkenmeier *et al.*, 1989; Christy *et al.*,1989; Kuo,

Xanthopoulos and Darnell, In Press). Additionally, C/EBP has been shown to be regulated at the level of transcriptional initiation (Xanthopoulos *et al.*, 1989). Furthermore, the transcription rate of C/EBP decreases in late cultures of mouse hepatocytes in a fashion similar to that for many genes regulated by the C/EBP protein (Xanthopoulos *et al.*, 1989). This suggests that the coordinate expression of liver genes during liver development, and suppression of liver-gene expression observed by cultured hepatocytes is controlled to some extent by the level of this factor.

When oligomerized C/EBP directs liver transcription *in vitro* (Marie *et al.*, 1989), implicating it as a positive acting transcriptional protein. Friedman *et al.* (1989) have recently also shown that cotransfection and overexpression of C/EBP in hepatoma cells can boost the transcriptional levels of an albumin reporter gene (which is known to have C/EBP binding sites; Babiss *et al.*, 1987; Cereghini *et al.*, 1987; Lichtsteiner *et al.*, 1987; Chapter 3) by 50 fold. Interestingly, in a similar experiment conducted in fibroblasts, the same C/EBP clone produced no activation of the albumin construct. This cell type dependance suggests that C/EBP might require an additional liver specific protein or a liver specific modification in order to be functionally active (Friedman *et al.*, 1989).

A related protein, DBP, has recently been described which binds to the same sequence as C/EBP (Marie *et al.*, 1989). This factor was cloned by Schibler and colleagues by screening a cDNA expression library with an

C/EBP binding site. Like C/EBP, DBP stimulates the *in vitro* transcription of albumin and functions in cotransfection experiments as above, but in contrast to C/EBP it is not a leucine zipper protein and it appears at later times in embryonic development (Schibler, 1989). Furthermore, DBP, while a liver specific protein, is transcribed in all tissues and hence unlike C/EBP, posttranstscriptional mechanisms are involved in regulating its tissue specificity. The ability of multiple proteins to bind the same DNA sequence is not uncommon for transcription factors and similar situations are observed for example in the cases of the octomer and CAAT binding factors (Dorn *et al.*, 1987; Herr *et al.*, 1988).

HNF3 and HNF4

Two additional liver specific factors HNF3 and HNF4 have been identified from studies on the TTR and α1-antitrypsin genes (Costa *et al.*, 1989). HNF3 was initially found by deletional analysis of the mouse Transthyretin (TTR) promoter and specific point mutagenesis of this HNF3 site blocked transcription in HepG2 cells (Costa *et al.*, 1989). Additionally, oligomerization of the HNF3 binding site stimulated β-globin gene transcription only in liver derived cell lines. This factor has been purified and shown to be a 50kd protein. It has recently been cloned and preliminary analysis suggests that it is transcribed in a liver specific fashion (Lai, Xanthopoulos and Darnell, unpublished observations).

HNF4 sites were identified in both the TTR proximal and TTR enhancer elements, again by transfection analysis, and the binding site of factor very closely resembles the AF-1 factor described by Leff *et al.*, 1989. HNF4 is a 54kd protein which sequence analysis reveals belongs to the zinc finger class of binding proteins. It has also been cloned and its tissue specific regulation appears to be transcriptional (Sladek, Xanthopoulos and Darnell, unpublished observations).

VI. Plan of the thesis

This thesis is divided into three major sections. The first involves a study of the rat albumin promoter (Chapter 3). Recombinant adenoviral vectors were used to introduce the albumin gene into human hepatoma cells. This simplified the identification of the DNA sequences which were both necessary and sufficient for dictating albumin cell-type specificity. An analysis of the multiple DNA-protein interactions which occur upon this promoter element was performed and the significant findings are presented in relation to published reports by other groups who have also analyzed this much studied regulatory region.

In the second section of the thesis, a functional characterization of the far upstream mouse albumin enhancer element is presented (Chapters 4-7). This sequence element was functionally demonstrated in CWSV1 cells to play a critical role in the maintenance of the high rate of albumin transcription,

and I have elucidated both positive and negative regulatory domains involved in producing its tissue specific effect. Two novel regulatory proteins have been identified, NLS1 (Non-Liver Specific protein 1, Chapter 6) and ANF (Albumin Negative Factor, Chapter 7), which along with the liver enriched transcription factor C/EBP (Chapter 5), will be shown to function combinatorally to dictate the liver specificity of the enhancer.

Lastly, in the third section of the thesis (Chapter 8), I explore the question of how these transcription factors combine to modulate the transcription of specific genes in different cell types, and also investigate how the transcription factors themselves are regulated. Taking advantage of the fact that the four liver enriched transcription factors just described have now been cloned, the expression of these factors (at both the transcriptional and post-transcriptional level) was assayed in several hepatoma cell lines and their dedifferentiated variants. The levels of the factors were then correlated with the transcription rates of two of their dependent genes, albumin and transthyretin. These experimental findings help to further elucidate the molecular mechanisms important for maintaining coordinate gene transcription in hepatocyte derived cell-types.

Chapter 2

Materials and Methods

Cells and Viruses

Monolayer cultures of human HeLa and 293 cells (Grahm *et al.*, 1977) were maintained in Dulbeccos modified Eagle's medium supplemented with 10% fetal calf serum. Human and rat hepatoma cells (HepG2, H4II, FaO, C2, C2rev7 and H5) were all grown in monolayer culture in Ham's F12 media supplemented with 5% fetal bovine serum, 1x essential amino acids (GIBCO laboratories) and 25µg of garamycin per ml (Schering corp; Ham, 1960; Knowles *et al.*, 1980; Clayton *et al.*, 1985a). Primary cultures of rat hepatocytes were prepared by perfusion of rat livers *in situ* with EDTA and collagenase using the procedure previously described for mouse hepatocytes (Clayton and Darnell, 1983). Primary cells were cultured in Eagle's medium containing 10% fetal calf serum, 10µg/ml of hydrocortisone and 10µg/µl of insulin.

CWSV1 cells were grown in serum free RPMI media which was supplemented as described by Woodworth and Isom, 1986. This media (1 liter) contained 330ml RPMI (GIBCO), 15ml 7.5% $NaHCO_3$, 4ml Hepes (1.5M), 40ml BSA solution, 4ml Pen-Strep (1g Penicillin + 1 g Streptomycin/100ml) and 4ml 100x trace metals. The BSA solution was prepared by mixing 320mg of fatty acid free BSA to 8 ml of PBS (phosphate buffered saline), and subsequently adding 160µl linolenic acid (5.4µl/ml absolute ethanol), and 160µl of ethanolamine (10µl/ml) dropwise. After stirring for 12hr at 4°C, 400µl transferrin (100mg/ml PBS), 24µl insulin (1mg/ml), 40µl

dexamethasone and 14µl glucagon (1mg/ml) were added. The mixture of trace metals was first prepared as separate solutions (A and B). Solution A (1x) consists of 0.5nM $MnCl_2$ $4H_20$, 0.5nM (NH_4) MO_7O_{24} $4H_20$, 0.25nM $NiSO_4$ $6H_20$, 15.0nM Na_2SeO_3, 0.25nM $SnCl_2.2H_20$, 2.5nM Na_3VO_4, and 50nM $CdSO_4$ while solution B consists of 250nM $Na_2S_10_3$ $.9H_20$. These two solutions were mixed together immediately before use to prepare the 100x stock. While routinely grown in the above serum free medium (SFM), CWSV1 cells were split into SFM containing 5% fetal bovine serum (Hazelton) for 4-6 hours until the cells adhered to the plastic dish. At that time, SFM without serum was again used. Viral stocks were obtained by infection of monolayer cultures of 293 cells and the preparation of crude cellular lysates by repeated cycles of freezing and thawing (Babiss *et al.*, 1984). Titers of viral stocks were determined on 293 cells by fluorescent focus assay (Thiel *et al.*, 1967) with antiserum raised in rabbits against cesium chloride-banded purified virions.

Preparation of plasmid DNA and clone Analysis

Plasmid bearing bacteria (usually the strain HB101) were grown overnight in 500ml N broth (10g bactotryptone, 1g yeast extract, 8g NaCl, 1g Glucose and 0.3g $CaCl_2$ per liter). The turbid cell suspension was spun at 3500rpm for 25min, resuspended in 30ml of Solution I (50mM Glucose, 25mM TrisCl, 10mM EDTA and 2mg/ml lysozyme), and allowed to stand at room temperature for 5min. At that time, 40ml of solution II was added

(0.2N NaOH, 1% SDS), and the mixture was gently swirled. Upon clearing, 40ml of Solution III was added (5M KAc pH 4.8, 57.5ml glacial acetic acid, 145.5 ml H_20), with vigorous shaking. The flocculent white precipitate was removed by centrifugation in a JA-14 rotor for 20min at 4°C, and filtration through two layers of cheesecloth. The filtrate was then combined with 70ml of isopropanol for 15 minutes at room temp. This was centrifuged in JA-14 rotor at 8000rpm for 30min, and the resulting pellet was dissolved in 9ml of TE (10mM Tris pH 8, 1mM EDTA) to which was added 10g of CsCl, and 0.5ml ETBr (10mg/ml). This was added to an ultracentrifuge tube and centrifuged in a VTI65 rotor at 45K for 16hr. Plasmid containing DNA bands were harvested and purified by n-butanol and phenol/chloroform extraction followed by ethanol precipitation.

A mini-prep procedure was used to prepare small quantities of plasmid DNA for clone analysis. Bacteria were grown overnight on N broth plates (N broth + 15mg/liter BBL granulated agar) containing the appropriate antibiotic. A large number of cells was picked with a toothpick and dissolved in 80µl of STET buffer (50mM Tris pH 8.0, 8% sucrose, 5% triton X-100, 50mM EDTA), to which 5µl of a fresh 10mg/ml lysozyme solution was added. This suspension in a 1.5ml microcentrifuge tube was boiled for 40 seconds and then centrifuged (in a microcentrifuge) for 15 min. The pellet was removed and the supernatant extracted with 60µl phenol/chloroform (24:1 v/v isoamyl

alcohol). Following this separation, 20µl of the aqueous phase was added to a fresh microcentrifuge tube containing 40µl of ethanol. The DNA was precipitated, subjected to a single 70% ethanol wash, desiccated and analyzed by restriction endonuclease digestion.

Plasmid and recombinant adenovirus construction

Recombinant adenoviruses were constructed which contained variable amounts of albumin promoter proximal sequence. To accomplish this, rat albumin DNA sequences from -441 to +957 were cloned into pGem1 as a HindIII - EcoR1 fragment. This included 441bp of albumin promoter sequence, the RNA start site and the A and Z exons (Sargent *et al.*, 1979; Sargent *et al.*, 1981). Sequences were serially removed by linearizing this plasmid (using either a HindIII site at -441 or a BstEII site at +47) and performing sequential deletion using the exonuclease Bal31 (Legerski *et al.*, 1978). These digestions were routinely performed by incubating the DNA (1.5pmoles ends) with 0.1unit of Bal 31 (Boheringer Mannheim) in a buffer containing: 20mM Tris pH 7.4, 600mM NaCl, 12.5mM $CaCl_2$, 12.5mM $MgCl_2$, and 1mM EDTA. Reactions were conducted at 30°C for a period of 1-5 min at which point a 10µl aliquot was removed and added to 90ul of precipitation/stop buffer (200mM NaCl, 10mM Tris pH 7.4 and 20mM EDTA). Following phenol chloroform extraction and ethanol precipitation these

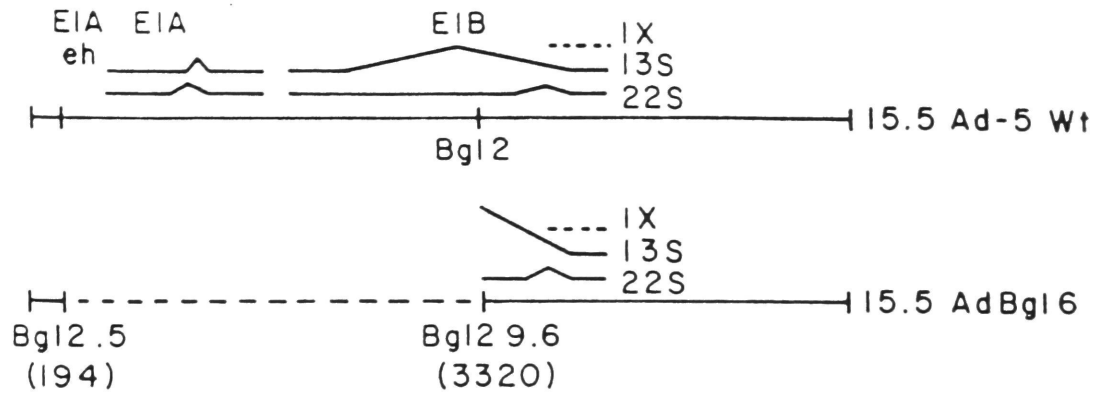
albumin promoter containing fragments were BamH1 linkerized and cloned using standard procedures (Maniatis *et al.*, 1982) into the Bgl II site of the plasmid construct AdBgl6 (Friedman *et al.*, 1986) which contains adenovirus sequences from 0 to 15.5 map units (1 map unit = 365bp) with the nucleotides from 194 to 3320 deleted (Figure 2.1) .

Plasmid DNAs were prepared, linearized at adenovirus nucleotide 3320 and mixed with the subgenomic adenovirus DNA (a fragment containing from 3.85 to 100 map units). This adenovirus segment was prepared by digesting the IN340 viral genome (Hearing and Shenk, 1983) with Cla1 which cuts uniquely at 3.85 map units. This generated linear viral DNA that was almost completely free from infectious molecules. Intracellular recombination between the 3.85 to 100 map unit segment and each of the plasmids resulted in the various adenoviruses described in the text. This method of overlap recombination is described in greater detail elsewhere (Babiss *et al.*, 1984; Friedman *et al.*, 1986).

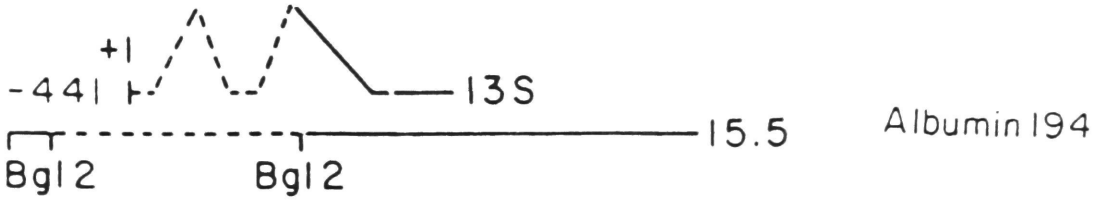
The alb 194 virus series contained the left-end terminal 194 nucleotides of the type 5 adenovirus genome followed by the rat albumin promoter region (-441 to +957) and the first two exons of the albumin gene fused within an intron to viral E1B sequences. This is performed in such a fashion that the 3' E1B exon and E1B poly A+ site allow the formation of a stable mRNA from the recombinant transcription unit (see Figure 2.1). Also included within the E1B region is the adenovirus protein IX transcription unit.

Figure 2.1. Genomic maps of the recombinant albumin adenovirus. In A (top), the left end of the adenovirus genome is depicted with a diagram of the RNA molecules produced from the E1A and E1B regions. The angled lines represent sequences which are spliced out of the mature mRNAs. Shown are sequences from 0 to 15.5 map units of type 5 adenovirus (1 map unit = 365bp). In A, bottom, is shown the cloning vector AdBgl6 (see text) in which sequences from nucleotides 194 to 3320 were removed (represented by the dotted line) and a Bgl II site created. This plasmid served as the recipient for the deleted albumin promoter clones (containing upstream promoter region (up to 441 base pairs) and 3' sequences to +957. In B, is diagrammed the genomic map of the parent 441 base pair containing recombinant albumin adenovirus. The dotted lines represent the albumin sequences. Note that either the 13S or 22S E1B messages can be produced by the fusion albumin/E1B gene (prepared by an intron fusion; Friedman *et al.* 1986).

A


EIA
EIA
eh
EIB
IX
13S
22S
15.5 Ad-5 Wt
Bgl2
IX
13S
22S
15.5 AdBgl6
Bgl2.5
(194)
Bgl2 9.6
(3320)

B


+1
-441
13S
15.5
Bgl2
Bgl2
Albumin 194

Construction of Mouse Albumin Enhancer Containing Expression Plasmids

Plasmid constructs with shortened versions of the mouse albumin enhancer sequences placed at the -300bp position of its promoter proximal region were prepared. These constructions were derived from a BamH1 DNA fragment containing albumin upstream sequences from -12.0Kbp to -8.43kbp fused to the -300 mouse albumin promoter element and extending to +22 of the albumin first exon (alb.EB.hgh, a gift of R. Palmiter, Pinkert *et al.*, 1987) that were cloned into the Bgl II site of the expression vector AdBgl6 which was described in the previous section (Friedman *et al.*, 1986). This plasmid (pRL1) was then digested with HindIII (cuts at -10.5Kbp) and SalI (cuts at nucleotide 5780 of the adenovirus sequence), and ligated into the polylinker of the pGEM1 vector cleaved with the same enzymes. This plasmid (pRL2) was the parent for the CWSV1 transfection constructs.

5' deletions were produced from the HindIII site using the exonuclease Bal31 and verified by sequence analysis (Maxim and Gilbert, 1980). The enhancerless control was constructed by placing the -300(PstI) to +22(BamHI) fragment from the mouse albumin DNA into the BglII site of AdBgl6, and subsequently subcloning the -300 (now a BglII site) to SalI (nucleotide 5780 of adenovirus) fragment into the polylinker of pGEM1 (this plasmid was named pRL3).

Sequences were removed from the 3' end by taking advantage of known restriction endonuclease sites (Acc I -9.9kbp Bgl II -8.9kbp; Pst -8.7kbp)

while the 5' end was resectable by digestion with the endonuclease HindIII. The DNA polymerase repaired termini were subsequently EcoR1 linkerized and the gel isolated fragments were placed in position 5' to the albumin promoter and the orientation determined by restriction endonuclease analysis.

Heterologous Mouse Albumin Enhancer Constructs

Heterologous constructs were designed using the mouse Transthyretin (TTR) promoter. The TTR parent contains 202 base pairs of TTR upstream sequence (of the RNA start site) along with the first two TTR exons, fused to the SV40 large T antigen 3' end and poly A+ addition sites (Costa *et al.*, 1986; 1988a). A unique XbaI site at -202 was used for insertion of the fragments, which were then appropriately oriented by restriction endonuclease digestion. The TTR minimal enhancer sequences (a fragment taken from –1.96 to –1.86Kbp upstream of the cap site) were fused immediately upstream of this promoter element at this same location (Costa *et al.*, 1986). Fragments from the albumin negative region (coordinates as described in chapter 7) were generated by restriction endonuclease digestion, end repaired using klenow polymerase, XbaI linkerized and placed between the promoter/enhancer construct in either orientation as marked (+ or -). These fragments were subcloned into pGem1 and subsequently sequenced (Sanger *et al.*, 1977).

Additional heterologous constructs were also designed using the

mouse β-globin (Konkel *et al.*, 1978) promoter. Albumin enhancer fragments were cut with the appropriate enzyme and inserted in either orientation 341 nucleotides upstream from the β-globin mRNA start site. The SV40 enhancer (Zenke *et al.*, 1986) was inserted in this same location using HindIII linkers. This plasmid served as a positive control in the transfection assays.

Recombinant adenoviruses were isolated as previously described from plasmids which contained either the mouse albumin promoter (-300 nucleotide upstream of the start site) alone or the albumin promoter with the mouse albumin enhancer sequences from -10.5 to -8.43kbp placed immediately upstream. The method of overlap recombination was again used.

Plasmid Transfection, Adenovirus Infection and RNase T_2 analysis of Expression Constructs

Mouse albumin promoter and enhancer-containing plasmid DNA (50μg) along with the SV40 enhancer containing β-globin internal control (20μg) was cotransfected into CWSV1, HeLa, or HepG2 cells at 40% confluency using the calcium phosphate-DNA co-precipitation procedure. Transfections employing Hela and HepG2 cells were performed in DME containing 10% calf serum, changed 3-4 hours prior to the procedure. CWSV1 cells were transfected in Lebovitz L-15 media (GIBCO) containing 8ml of BSA solution

(10gBSA/100ml H_20), 4µl 100x glucagon (150mg/ml), 5ml Pen-Strep, 8ml NaBicarbonate and 20ml of fetal calf serum (Hazelton).

DNA to be transfected was resuspended in 100ul TE (10mM Tris pH8.0, 1mM EDTA). Precipitates were prepared by adding the DNA to 800µl of 1/2 TE and 100µl 2.5M $CaCl_2$ in a 15ml conical tube. Subsequently, 1ml of 2x HBS (50mM Hepes 7.2, 280mM NaCl, 1.5mM Na2HPO$_4$) was slowly added while vortexing constantly. This was allowed to stand at room temperature for 1h prior and then added to each plate. Four hours following transfection, the cells were incubated for 1min with PBS containing 10% glycerol. The PBS/glycerol solution was removed and the cells were washed two times with Hank's buffered salt solution and fresh media added. The next day the media was changed to Hams F12 for the HepG2 cells and serum free media for the CWSV1 cells. Cytoplasmic RNA was harvested from transfected cells at 48 hours after the transfection.

The cytoplasmic RNA was isolated using the method of Nonidet-P-40 (NP-40) lysis (Friedman *et al.*, 1986), The media was removed from the cells and they were washed with 5ml of ice cold PBS, The cells were then scraped into 8ml of PBS and spun at 2000g for 5min at 4°C. Cells were resuspended in 320µl of lysis buffer (0.14M NaCl, 1.5mM $MgCl_2$, 10mM Tris pH 8.6) and 25µl of 0.5% NP-40 was added. Nuclei were removed by centrifugation for 2 min, and the supernatant was added to 300µl of 2x PK buffer (0.2M Tris pH 7.5,

25mM EDTA, 0.3M NACL, 2% w/v SDS) and 200μg/ml proteinase K at 65°C for 30min. This was extracted 2x with phenol/chloroform and the RNA precipitated in ethanol. This cytoplasmic RNA was scored for the accumulation of the albumin-adenovirus hybrid mRNAs by protection against nuclease digestion of labeled antisense RNA probes.

RNA probes were generated using either SP6 or T7 RNA polymerase (Melton *et al.*, 1984; Hart *et al.*, 1985), as indicated in the specific figures and corresponding text. Total cytoplasmic RNA (20μg) was hybridized to a specific probe 7.5×10^5 total cpm and after RNase T_2 digestion (55U/ml) at 30° C for 2h, the products were sized on denaturing polyacrylamide gel. Multiple 5' or 3' assays were used through this thesis and they will be identified appropriately as they appear.

HepG2 cells were infected with the appropriate recombinant adenovirus at 50pfu/cell for 24h in the presence of AraC (40μg/ml) to prevent viral DNA replication (Friedman *et al.*, 1986, Babiss *et al.*, 1987). Following the infection, cytoplasmic RNA was harvested, The poly (A+) containing RNAs were isolated by oligo (dT)-cellulose chromatography (Nakazato and Edmonds, 1974). Oligo (dT) was equilibrated in sterile loading buffer (20mM Tris pH 7.6, 0.5M NaCl, 1mM EDTA and 0.1% SDS, and a 1 ml column was poured and washed with 0.1M NaOH, 5mM EDTA followed by 2-3 volumes of sterile H_20. The pH was monitored and maintained less than 8. RNA was

dissolved in sterile water and heated to 65°C, allowed to cool and loaded twice over the column. Poly A+ RNA was eluted with 2-3 volumes of sterile elution buffer (10mM Tris pH 7.5, 1mM EDTA and 0.5% SDS), and precipitated by adding sodium acetate. The production of stable cytoplasmic RNAs from the fusion E1B gene in virus-infected cells was again scored using an RNaseT_2 protection assay. The Viral E2a gene activity was measured as an internal control for viral infectivity.

Cell and Tissue extracts

Adult male sprague Dawley Rats (Charles river; between 10-14 weeks of age) were used to prepare nuclear extracts. Rat tissue (10g; liver, spleen or kidney) was dissected at 4°C and immediately placed into 50ml of buffer on ice (10mM Tris pH 8; 0.35M NaCl). The TN buffer was decanted off (in a cold room) and the tissue was finely minced and transferred to a 50ml motor driven dounce homogenizer. containing 10ml homogenization buffer (0.32M sucrose, 5mM$CaCl_2$, 2mM MgOAc, O. 1M EDTA, 1mM DTT, 10mM Tris, pH 8.0 and 0.5mM PMSF), and subjected to 10 strokes The homogenate was then filtered through 2 layers of cheesecloth, dounced ten times on ice (hand dounce B pestle) and a sample analyzed for cellular breakage using a phase contrast microscope. This homogenate was next combined with 2 volumes of a solution containing 2M sucrose, 5mM MgOAc, 0.1mM EDTA, 10mM Tris-HCl,pH 8.0, 1mM DTT, and layered over a 10ml cushion of the latter

solution in a Sw27 tube. This was centrifuged at 13,000 rpm for 45 min. in a Beckman L2-65B preparative ultracentrifuge.

The resulting nuclear pellet was resuspended and partially lysed in 20mM Hepes, pH 7.9, 25% glycerol, 0.42M NaCl, 1.3mM $MgCl_2$, 0.2mM EDTA, 0.5mM PMSF, 0.5mMDTT (at a 2 to 1 ratio buffer to nuclei). This mixture was then manually dounced ten times (Type A pestle, loose fitting), and nuclear proteins were extracted over a 45 min. period at 4°C with mild shaking. The viscous mixture was centrifuged at 25,000 rpm for 30min, using a Sw40 rotor. Ammonium sulfate precipitation (0.3gr/ml) and dialysis of the nuclear extracts was performed in BC100 buffer (100mM KCl, 25mM HEPES (N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid), pH 7.9), 20% glycerol and 0.1 M EDTA).

Cultured cell extracts (CWSV1, HepG2, and HeLa) were made according to the procedure of Dignam *et al.* (1983), and these extracts were concentrated by ammonium sulfate precipitation. Extract was stored frozen at -70° C in BC100 buffer. Extracts from the rat hepatoma cells FaO, C2, C2Rev7 and H5 were prepared according to the procedure of Shapiro *et al.* (1987) which differed from the above chiefly in that the nuclei were quickly centrifuged through sucrose prior to extraction. Protease inhibitors used included PMSF (0.5mM final), pepstatin (1µg/ml), Benzamidine HCl (1mM final) and leupeptin (0.5µg/ml final). Heparin agarose columns (5mg of protein per ml of resin) were performed according to the procedure of Briggs *et al.* , (1986).

All extracts were examined to assess their equivalence (for both protein amount and integrity) by performing gel shift assays using an oligonucleotide containing a known binding site for the transcriptional protein AP-1, which is found in most all cell-types (Lee *et al.*, 1987). The specific binding site used consisted of oligonucleotide B site from the α1-antitrypsin enhancer (Oligo B, Grayson *et al.*, 1988a). An example of this analysis is shown in Figure 2.2. This demonstrates that the extracts (prepared in this case from HeLa cells or rat liver, brain or spleen tissue), contained roughly equal amounts of nuclear proteins and that the integrity was comparable.

In addition, these liver extracts have been shown to be transcriptionally active (Figure 2.3). Using a G-less cassette system (Sawadogo and Roeder, 1985) templates containing either the mouse albumin promoter (650 or 50 nucleotides of 5' upstream sequence), the adenovirus major late promoter (400 base pairs) or the β-globin promoter (341 base pairs) were assayed in a reaction containing 45% liver nuclear extract (as described in Gorski *et al.*, 1986). The final transcription buffer contained 25mM Hepes pH 7.6, 50mM KCl, 0.6mM ATP, CTP, 35um UTP, 0.1mM 3'-0-methyl-GTP, 12% glycerol and 1 μl of RNasin. As shown in Figure 2.3, each template generated a specific transcriptional band, which could be inhibited by the inclusion of α-amanitin (data not shown). Furthermore, the addition of albumin sequences (from -50

Figure 2.2. AP1 gel shift of nuclear extracts. As described in the text, the integrity of different nuclear extracts was compared by subjecting them to a gel shift assay using a labeled probe containing an AP1 binding site (Oligo B, α1-antitrypsin enhancer). For each lane, 1ng of labeled probe was incubated with 8μg of nuclear extract in the presence (+) or absence (-) of a 30 fold molar excess of unlabeled AP1 oligonucleotide. The arrow denotes the specific retarded complex.

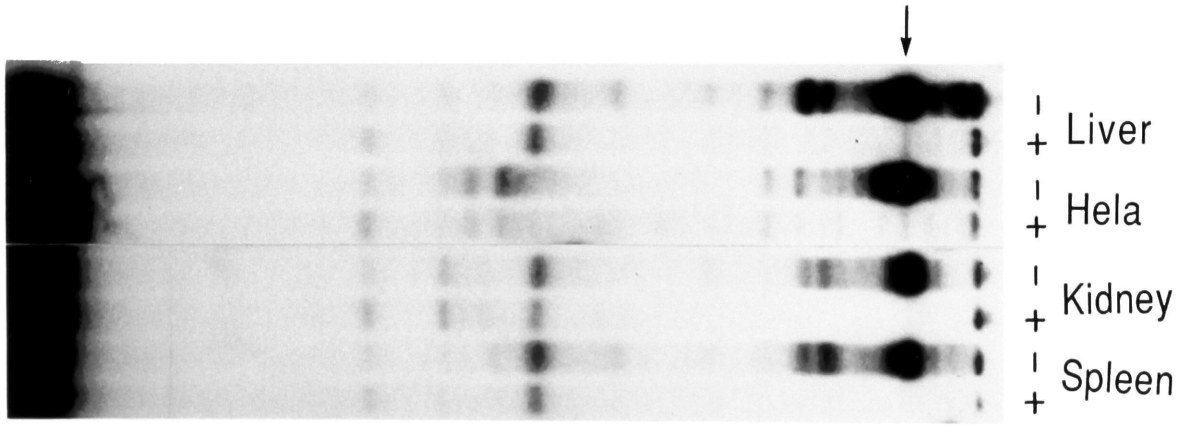
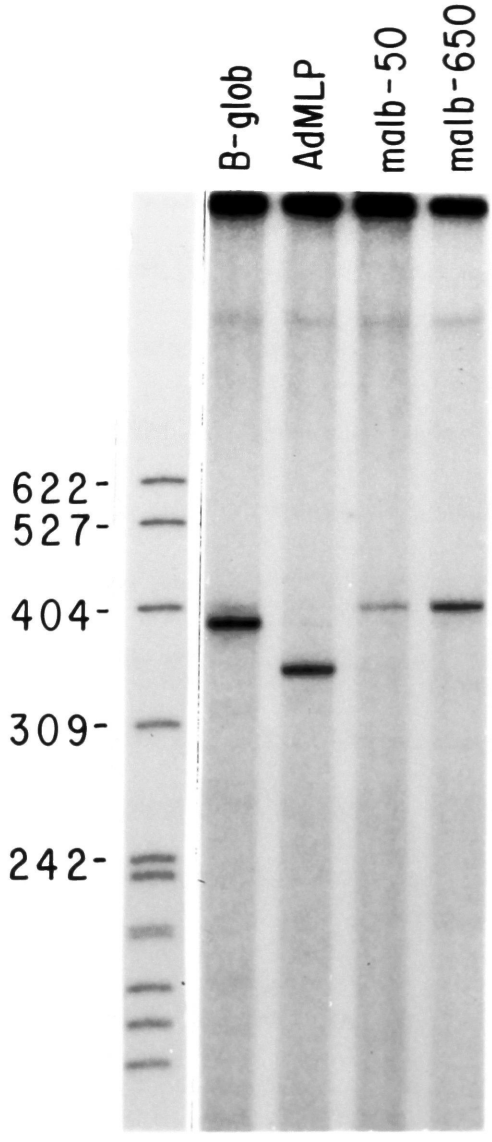

-
+ Liver
-
+ Hela
-
+ Kidney
-
+ Spleen

Figure 2.3. *In vitro* transcription using liver nuclear extract prepared as described in the text. Nuclear extracts from adult rat livers were used to transcribe the mouse albumin gene (the sequences including -650 to +10 relative to the ALB capsite or a deletion to -50bp) which had been cloned into a vector that allows direct measurements of the rate of transcription initiation, the G-less cassette. Also included is the β-globin promoter and an adenovirus major late promoter G-less construct as described in the text. The reactions are performed in the absence of exogenous GTP, which prevents RNA polymerase II elongation from occurring anywhere but at the appropriate cap site.


B-glob
AdMLP
malb-50
malb-650
622-
527-
404-
309-
242-

to -650 nucleotides upstream) provided a stimulation of transcription, demonstrating that upstream trans-acting factors we present and able to activate transcription in these extracts.

In Vitro DNA-protein binding assays

Gel mobility shift assays were used to identify DNA-protein interactions in crude nuclear extracts (Fried and Crothers, 1981; Garner and Revzin, 1981). These assays were performed by incubating 1ng of end labeled restriction fragment or synthetic oligonucleotide with 8-10μg of nuclear extract in a 20μl binding reaction (containing 4μg of poly (dI-dC): poly (dI-dC) (Pharmacia), in 20mM HEPES pH 7.9, 40mM KCl, 2mM $MgCl_2$, 1mM DTT, 0.5mM EGTA and 4% ficoll) for 30min at room temperature. Following incubation, 4μl of this mixture was electrophoresed on a native acrylamide gel (the percentage varying with the size of the probe) in 50mM TBE. When used, an excess of cold competitor DNA or duplex oligonucleotide DNA was employed (the precise fold excess is described with each individual experiment).

Exonuclease III stop reactions were also used to explore DNA-protein interaction (Wu 1985), essentially according to the protocol of Kovesdi *et al.* (1986). For all reactions, 1 μg of 5' end labeled DNA probe (uniquely labeled using T4 polynucleotide Kinase) was combined with 30μg of nuclear extract

and 4μg of poly d (I-C) which was included as a nonspecific competitor of DNA-binding. Following ExoIII digestion (100U per reaction, Boehringer-Mannheim Biochemicals), the products were analyzed on 8% denaturing polyacrylamide gels.

DNase I footprinting studies were also employed (Galas and Schmitz, 1978). These were performed using liver nuclear extract prepared as above which had been fractionated on a heparin agarose column using a KCl salt gradient and Hela extract which had been fractionated on a P11 column using KCl steps (0.25M KCl for liver and 0.4M step for Hela, respectively). The fractions were dialyzed against BC100 buffer and a gel shift reaction was used to determine the fraction possessing maximal binding activity. Subsequently, a 1ng sample of ^{32}P labeled probe was incubated with 5-30μg of extract, 5μg of poly (dI-dC) and 4% polyethylene glycol in a buffer containing 10mM Hepes (pH 7.9), 7 mM $MgCl_2$ and 0.75mM Dithiothreitol at 0°C for 60min. DNase I was added for 3-4min and the reaction was terminated by the addition of proteinase K (200μg/ml) and sodium dodecyl sulfate (0.5%) The DNA was extracted with phenol and ethanol precipitated, and the labeled fragments were analyzed on an 8% denaturing acrylamide gel.

Methylation protection experiments

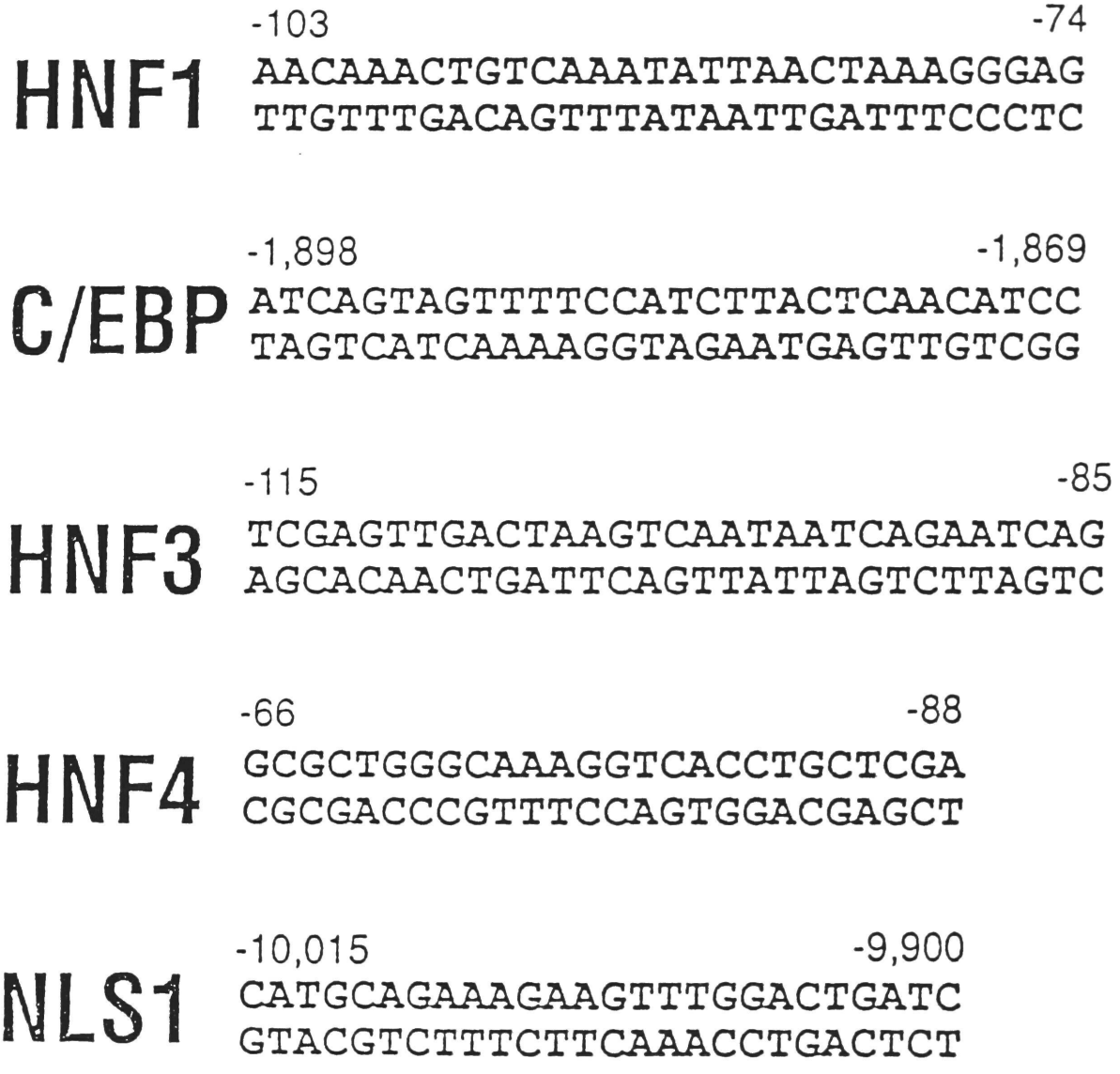
High specific activity 3' end-labeled probes were prepared by filling in a

DNA restriction site with 100μCi of all four ^{32}P deoxynucleotides (New England Nuclear Corp; 3,000 Ci/mmol) with the klenow fragment of DNA polymerase 1 (Maniatis *et al.*, 1982) followed by phenol-chloroform extraction and ethanol precipitation. The DNA was recut with a second restriction enzyme and the fragments were gel purified on a low melting-point agarose gel (Bethesda Research laboratories). The methylation protection (footprint) experiment involved incubating 20ng of this probe with 100μg of HepG2 extract in a 400μl reaction mixture under the same conditions as the gel shift reaction. This mixture was subsequently treated with 2μl of dimethyl sulfate (DMS) for 2 min and immediately loaded on a nondenaturing preparative acrylamide gel (6%). The specific bands were localized by autoradiography and eluted from the acarylamide and deproteinized by phenol-chloroform extraction. The eluted DNA bands were then cleaved with piperidine and analyzed on a denaturing 8% acrylamide gel to reveal G residues which had been protected from DMS methylation by specific protein interactions.

Oligonucleotides

In many of the experiments described in this thesis synthetic oligonucleotides were used as probes for gel shift reactions or as competitors in any of the different i*n vitro* protein assays described above. Five of these double stranded oligonucleotides are shown in Figure 2.4. They are named

Figure 2.4. Sequences of the oligonucleotides used as probes and competitors in gel shift assays. Each is identified on the left by the name of the protein factor known to interact with it. The origin of each binding site is described in the text.


-103
-74
HNF1
AACAAACTGTCAAATATTAACTAAAGGGAG
TTGTTTGACAGTTTATAATTGATTTCCCTC
-1,898
-1,869
C/EBP
ATCAGTAGTTTTCCATCTTACTCAACATCC
TAGTCATCAAAAGGTAGAATGAGTTGTCGG
-115
-85
HNF3
TCGAGTTGACTAAGTCAATAATCAGAATCAG
AGCACAACTGATTCAGTTATTAGTCTTAGTC
-66
-88
HNF4
GCGCTGGGCAAAGGTCACCTGCTCGA
CGCGACCCGTTTCCAGTGGACGAGCT
-10,015
-9,900
NLS1
CATGCAGAAAGAAGTTTGGACTGATC
GTACGTCTTTCTTCAAACCTGACTCT

according to the cognate protein which is known to interact with their sequence. Each oligonucleotides contains a four base pair overhang on each ends which was used for 3' end labeling. The sources of the oligonucleotides are as follows: HNF1, β-fibrinogen promoter -103 to -74bp (Courtois *et al.*, 1987); C/EBP, mouse TTR promoter, -1,898 to -1,869 bp (Costa *et al.*, 1988); HNF3, mouse TTR promoter, -115 to -85 (Costa *et al.*, 1989), HNF4, human ApoCIII gene, -66 to --88bp (Reue *et al.*, 1988; Leff *et al.*, 1989) and NLS1, mouse albumin enhancer, -10.015 to -9,900 base pairs (Herbst *et al.*, 1989).

Nuclear-Run On Transcription

Nuclear run-on analysis was performed to measure relative transcription rates (Weber *et al.*, 1977; Hoffer and Darnell, 1981). Nuclei from different tissues or cell lines were isolated according to the procedures outlined earlier for the preparation of nuclei for extraction. After washing in transcription buffer lacking nucleotides (20 mM Tris, pH 7.9. 20% glycerol, 140 mM KCl, 10mMMgCl_2 and 1mM DDT at 4°C) on ice, the nuclei were resuspended in 100 µl of complete reaction mix (1mM each ATP, CTP and GTP, 0.3 to 0.5 mCi α-^{32}P UTP (3000Ci/mmole; New England Nuclear). The nuclei were then incubated for 15min at 30°C with gentle swirling to chain-elongate previously initiated RNA. At the end of the reaction, the synthesis of nascently labeled RNA was terminated by the addition of 2ml of HSB (500 mM NaCl, 10 mM Tris, pH 7.4, 50mM $MgCl_2$, 2mM $CsCl_2$ and 0.1ml DNAse

(Worthington 1mg/ml stock) at 25°C for 5' with occasional vortexing. Following the addition of SDS (200µl of a 20% solution), EDTA (100µl of 0.5M) and 3ml ETS (10mM tris7.4, 10mM EDTA, 0.2% SDS), nuclear RNA was isolated by hot phenol extraction (heated to 65°C) and ethanol precipitated. Unincorporated triphosphates were removed by precipitation of RNA with trichloroacetic acid (Groudine *et al.*, 1981). The total labeled RNA averaged 1 ct/min per cell nucleus.

Following RNA extraction, labeled RNA was hybridized to an excess of cloned denatured DNA affixed as slots to 0.45µm nitrocellulose (Kafatos *et al.*, 1979), and after RNase digestion, autoradiography was used to detect hybridized RNA. pGem 1 plasmid was always included as a measure of non-specific hybridization. These cloned DNAs were slotted onto nitrocellulose (5µg/dot) and prehybridized for 6h at 65°C in 2x TESS, 5x Denharts solution (0.02% (v/v) Ficoll, 0.02% (w/v) bovine serum albumin, 0.05% (v/v) polyvinylpyrollidone and 500µg yeast tRNA/ml. Nuclear hybridization (15 x 10^7 cts/ml) was conducted for 40h in 2x TESS (4x TESS stock: 0.1M TES pH 4, 0.6 M NaCl, 0.02 M EDTA, 0.4% SDS), 5x Denharts, 0.1mg/ml yeast tRNA, 0.1% sodium pyrophosphate in a final volume of 1 to 3ml. After extensive washing (2x SSC at 65°C), filters were treated with ribonuclease at 37°C (0.45 units/ml pancreatic RNase) in 2xSSC, blotted dry and exposed to Kodak XAR5 X-ray film at -70°C with an intensifying screen for 16-72 hrs.

DNA and Genomic Clones Employed

Spotted DNA included a clone for albumin, a 2400bp genomic fragment (rat B genomic clone, Sargent *et al.*, 1981), TTR, a 440 bp fragment of the rat TTR Gene (Dickson *et al.*, 1985); HNF1 a 3Kb fragment λ16 (Frain *et al.* 1989) (gift of R. Cortese); C/EBP (500bp pPST1 clone, Xanthopoulos *et al.*, 1989); and HNF 3 and 4 were cDNAs of 655 and 1000bp respectively (Lai, Sladek and Darnell, unpublished). A Chicken β-actin clone was used as a control (Cleveland *et al.*, 1980)

DNA sequencing

Most DNA sequencing was performed using the chemical sequencing method (Maxam and Gilbert, 1980). 5μg of each DNA fragment was 5' dephosphorylated using 1 unit of bacterial alkaline phosphatase in 10mM Tris pH 8 for 1h at 65°C, phenol extracted and ethanol precipitated. This DNA was then end labeled using 1 unit of T4 kinase for 1h at 37°C in a buffer containing (0.5M Tris pH 7.6, 0.1M $MgCl_2$, 50 mM DTT, 1mM spermidine, 1mM EDTA and ^{32}P γATP). The DNA was cut at a second site with a restriction endonuclease and this now uniquely end labeled DNA was isolated using low melting point agarose. In some cases, pGem cloned DNA was sequenced using the enzymatic method employing the SP6 or T7 sequencing primers (Sanger *et al.*, 1977).

Chapter 3

The Rat Albumin Promoter

It had previously been shown that the 441 nucleotides upstream of the start site of the rat albumin gene can, when inserted into an adenovirus vector, stimulate cell-specific expression (Babiss *et al.*, 1986; Friedman *et al.*, 1986; Friedman *et al.*, 1987). In the absence of viral DNA replication, cell-specific expression of the albumin promoter is maintained. Rat hepatomas (H4II and FaO), human hepatoma cells (HepG2), and primary mouse hepatocytes all support exogenous albumin transcription from these recombinant viruses (Friedman *et al.*, 1986). In contrast, a number of non-liver derived cell types did not generate exogenous albumin expression. These included myeloma cells (MPC11), HeLa cells, 293 Cells (human) and CREF cells, a rat fibroblast cell line, indicating that this 441 nucleotide region was both necessary and sufficient for maintaining liver specific gene expression (Friedman *et al.*, 1986). Similar conclusions had been reached by Ott *et al.* (1984), by transient transfection analysis of this same promoter element into rat hepatoma cells.

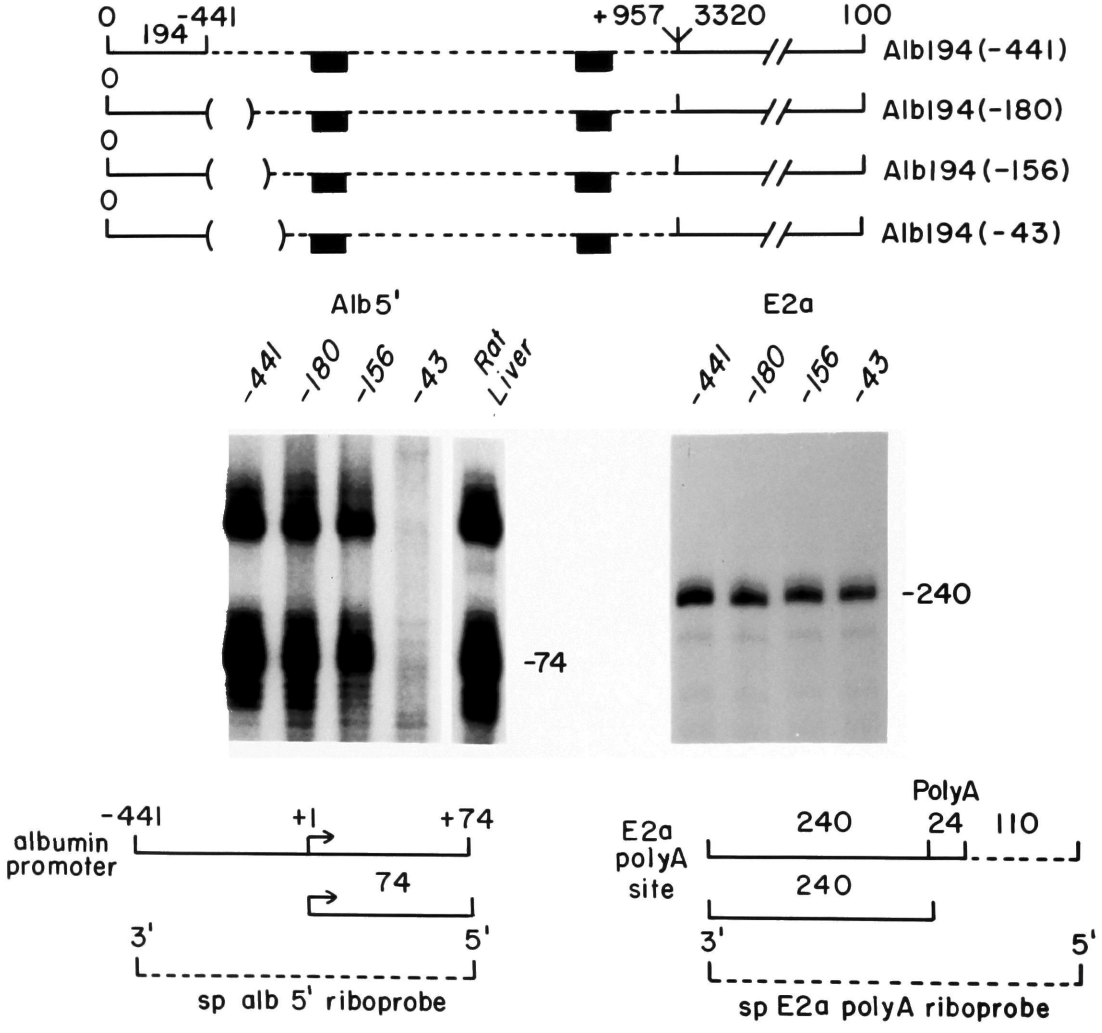
Deletional Analysis of the Rat Albumin Promoter Element

To identify the precise upstream sequence elements that confer tissue-specific expression, Bal31 exonuclease digestion of a segment of the rat albumin promoter was carried out, followed by cloning of the deleted DNA into plasmids, to yield a series of promoter deletions (described in chapter 2). Initial attempts at using DNA transfection to introduce these plasmids in

HepG2 cells gave extremely low signals (below the level of quantitation), even from the largest (441 nucleotide) albumin promoter construct. Therefore, these promoter deleted constructs were introduced into the adenovirus vector system described above (Babiss *et al.*, 1986; Friedman *et al.*, 1986; Chapter 2). These viruses, containing portions of the promoter, yielded stable RNAs generated by transcription initiation at the albumin start site followed by splicing of the albumin Z and A exons, the adenovirus E1b exon and poly A addition sequences. The virus constructs which have deletions 5' to their RNA start sites are named according to the nucleotide at the 5' end of the albumin promoter sequence (Figure 3.1).

Human HepG2 cells were infected at a multiplicity of 50pfu/cell for 24h in the presence of cytosine arabinoside (40μg/ml) to prevent viral DNA replication. Cytoplasmic RNA was isolated and the Poly(A+)containing mRNAs were selected. The accumulation of albumin E1B fusion mRNAs was measured using ^{32}P-labeled sp6 antisense riboprobes, which recognize the 5' ends of the mRNA. The results of the assay are shown in Figure 3.1 (left panel) and demonstrate that albumin promoter sequences extending to position -156 are sufficient for maintaining albumin expression in HepG2 cells. Further deletion of 5' sequences to nucleotide -43 caused a complete loss of albumin expression (the autoradiogram in Figure 1 was purposely overexposed to show the lack of promoter activity from the alb194 (-43) virus). To control for variations in the virus infection and RNA quantitation

Figure 3.1. Functional analysis of the rat albumin promoter by viral infection in HepG2 cells. Adenoviral recombinants containing the rat albumin promoter sequences extending in the 5' direction relative to the known start site for transcription to positions -441, -180, -156 and -43 and in the 3' direction to +957 were isolated by overlap-recombination in 293 cells .(as described in methods). Filled-in boxes show exon regions, open parentheses depict promoter deletions extending towards the cap site, and slanted parallel lines indicate DNA sequences not shown between viral sequences at 9.6 map units and 100 map units (1 map unit = 365bp). HepG2 cells were infected for 24hr with 50 pfu/cell of the appropriate virus in the presence of AraC (40μg/ml) to prevent viral DNA replication. Cytoplasmic poly (A+) mRNA was isolated and 9μg was hybridized with the sp alb 5' riboprobe and 3μg was hybridized with the sp E2a riboprobe. The protected RNAs were separated on a 5% denaturing polyacrylamide gels. The autoradiographic exposure was 12hr for the alb 5' assay and 20min for the E2A assay. As shown in the bottom of the figure, the full-length labeled sp6 alb 5' probe of 515 nucleotides is protected for 74 nucleotides while the E2A riboprobe of 374 nucleotides is protected for 240 nucleotides. For comparison, rat liver RNA (5μg) was also assayed using the sp alb5' riboprobe.


0
194
-441
+957
3320
100
Alb194(-441)
Alb194(-180)
Alb194(-156)
Alb194(-43)
Alb 5'
E2a
-441
-180
-156
-43
Rat Liver
-74
-240
-441
+1
+74
albumin promoter
74
3'
5'
sp alb 5' riboprobe
PolyA
240
24
110
E2a polyA site
240
3'
5'
sp E2a polyA riboprobe

prior to hybridization, a riboprobe assay for the viral E2a mRNA was performed on each of these samples. Since all the viruses led to formation of equivalent amounts of E2a mRNA, the lack of viral-albumin expression from the alb 194 (-43) virus was surely not due to an inability of this virus to infect HepG2 cells (Figure 3.1, right panel). In experiments not shown, the 194 (-156) virus was also active in rat hepatomas (H411 cells), but not in Hela cells (R.S. Herbst and L. E. Babiss, unpublished observation). Thus the sequences between -156 and -43 are able to dictate cell specific expression.

Two additional points should be made that relate to the use of 5' and 3' assays for scoring albumin mRNA expression and for expression analysis in general. First, the larger of the two protected hybrid bands shown in Figure 3.1 (using the 5' albumin assay) is most likely due to incomplete RNase T_2 digestion. This does not to represent a second transcriptional start site or a second splice acceptor site because primer extension analysis of the recombinant viral-albumin mRNA (performed by extending an adenovirus viral-albumin mRNA using an adenovirus E1B specific oligonucleotide) revealed only one start site at the location indicated by the smaller (~74 nucleotide) sp6 band.

A second point is that the 3' assay on poly(A+) selected RNA from infected HepG2 and Hela cells showed that the alb (-180) and alb (-156) viruses accumulated RNAs containing the E1B sequence in both Hela and HepG2 cells (data not shown). However, analysis of these accumulated RNA

molecules (using the 5' albumin sp6 riboprobe) showed that the transcripts did not begin at the albumin start site but rather within the 194 nucleotides at the left end of the adenovirus DNA. This activation of transcription might be based on bringing the albumin initiation region close to the end of the viral genome, which contains a potential enhancer element. Therefore, the only valid measure of albumin promoter activity in the viral constructs was achieved using the 5' sp6 assay scoring the first albumin exon. This is an important point since very often the results of transient assays are reported using assays which do not measure transcription from the actual RNA start site (ie. CAT assays). Based on the above, such results must be interpreted with caution, for they might not reflect transcripts generated exclusively by the regulatory elements under study.

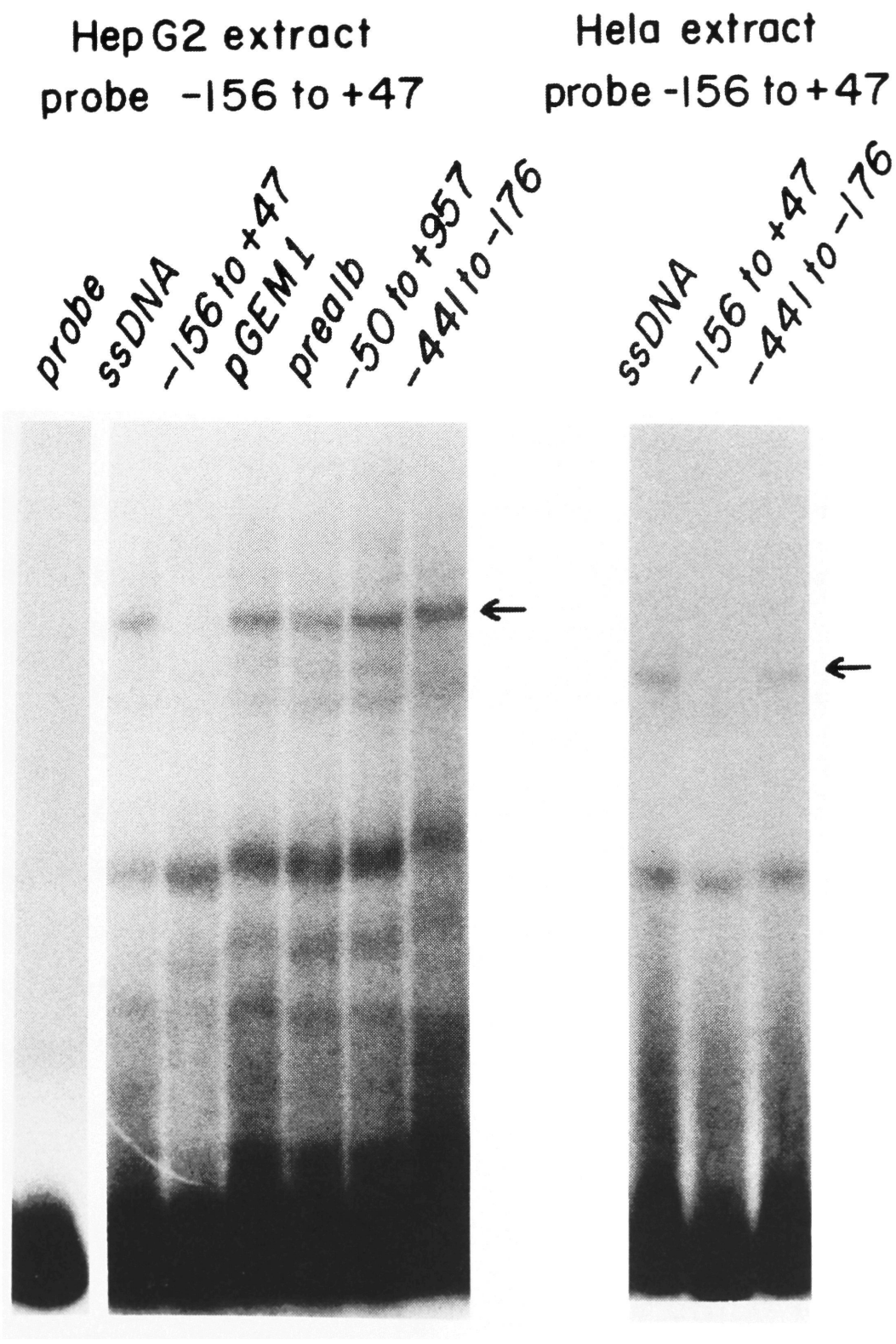
Detection of proteins that bind to albumin sequences between -156 to +47

Experiments were next designed to investigate the DNA-protein interactions occurring on the albumin promoter. To determine if nuclear proteins could physically interact with the albumin promoter sequences between nucleotides -156 and +47, an assay was performed based on the changed electrophoretic mobility of DNA when complexed with protein (the "gel-shift" assay; Fried and Crothers, 1981). End labeled DNA fragments were incubated with nuclear extracts in the presence of a nonspecific competitor of DNA-binding proteins, such as salmon-sperm DNA, and the mixture was

subjected to electrophoresis. The slower mobility of any protein-DNA complexes during electrophoresis was detected upon autoradiography of the polyacrylamide gel When a 203-bp albumin DNA fragment (-156 to +47bp) was reacted with proteins from HepG2 (7μg), HeLa (10μg) or primary hepatocyte (7μg) nuclei, a slower moving DNA-protein complex could be detected (Figure 3.2, data not shown). This shifted band was shown to be specific since the labeled DNA-protein complex could be abolished if excess unlabeled DNA of the same type (-156 to +47bp) was present in the incubation, yet formation was not prevented by other regions of the albumin promoter (sequences from -176 to -441 or -50 to +957), DNA from the mouse prealbumin (TTR) promoter or the α1-antitrypsin promoter (both of which are expressed in hepatocytes; Costa *et al.*, 1988a; Grayson *et al.*, 1988a). Additionally, fragments from the mouse β-globin promoter or from a series of adenovirus promoters (E1A, E1B, E2A, E4) could not compete this shift.

The proteins capable of inducing the specific gel shift were also present in HeLa cells (Figure 3.2). Since neither the endogenous nor recombinant albumin promoter are able to activate transcription in HeLa cells, this result was at first surprising. While this binding assay does not indicate that the proteins from different cell types that produce a gel shift are necessarily identical, this finding was the first indication that all or none differences in the pattern of DNA-protein interaction between liver and non-liver cell types

Figure 3.2. Factor binding to the rat albumin promoter sequences. A gel mobility shift assay was performed as described in Materials and Methods, using 1ng of end-labeled probe (sequences -156 to +957), salmon sperm DNA (1μg), and either HeLa cell nuclear extract (10μg) or HepG2 cell nuclear extract (7μg) followed by electrophoresis through a low ionic strength polyacrylamide gel. As indicated above each lane binding reactions were done in the presence of cold competitor DNA sequences at a 20 fold molar excess relative to the labeled probe. The -50 to + 957 and -441 to -176 DNA fragments were rat albumin sequences. The prealbumin (also referred to as TTR) sequence spans from nucleotide +40 to -202 and represents the region of that gene capable of conferring cell-specific expression. The band resulting from the specific interaction with the probe is indicated by an arrow.


Hep G2 extract
probe -156 to +47
Hela extract
probe -156 to +47
probe
ssDNA
-156 to +47
pGEM1
preAlb
-50 to +957
-441 to -176
ssDNA
-156 to +47
-441 to -176

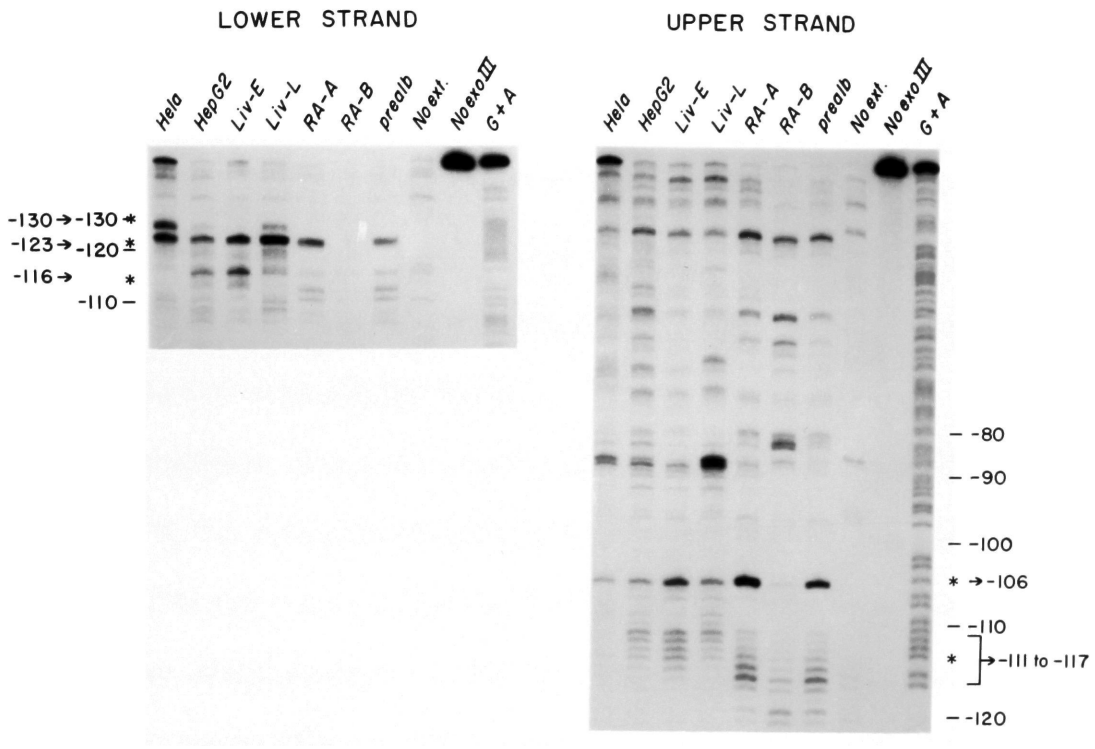
might not be responsible for the maintenance of liver specific gene expression.

Localization of binding specificity

To further define the location of the DNA-protein interactions and the distribution of albumin gene binding proteins in different cell-types, an exonuclease III (exoIII) protection test was designed based on the experiments of Wu (1985). This assay involves incubating a uniquely end labeled DNA fragment with a nonspecific competitor (polydI-dC) and nuclear proteins (30μg for each extract). The complexes that are formed are then exposed to exoIII which digests duplex DNA in a 3' to 5' direction, stalling at regions where DNA-protein interactions have occurred (exoIIIstop; Wu 1985; Kovesdi *et al.*, 1986). The partially digested DNA strands can then be analyzed by gel electrophoresis.

When this type of digestion was carried out on DNA that was cut and labeled at nucleotide +47, and then recut cut at nucleotide -156, a protein binding region was detected with an upstream limit at position -123 (Figure 3.3, lower strand). Nuclear proteins from a wide variety of cell types: hepatocytes, HeLa cells and kidney cells (data not shown) all produced an exonuclease stop at this site. The protein(s) were preferentially bound to the albumin promoter sequences as indicated by the suppression of the exoIII stop band by a homologous competitor DNA (RA-B; -176 to +47) and the

Figure 3.3. Exonuclease III protection assays for the binding of albumin-gene specific factors to albumin upstream sequences. Probes used for binding were 5' end labeled on the upper DNA strand (position -156 of strand equivalent to the RNA) or lower DNA strand (position + 47 of the strand complementary to the mRNA) of the probe. DNA-protein complexes were then incubated with exonuclease III, and partially digested labeled DNA was analyzed on 5% acrylamide denaturing gels. The lanes labeled "No ext" received no protein and the lanes labeled "No exoIII" received no protein or exoIII. Thirty micrograms of extract were used for each reaction. The origin of the extract is shown above each lane. RA-A (-441 to -176), RA-B (-156 to +47) and prealbumin (-202 to +40) represent unlabeled DNA fragments which were used as competitors in a 20 fold molar excess relative to the labeled probe in reactions containing 30μg of liver nuclear extract. The exoIII stop fragments are indicated by asterisks * and by their position relative to the albumin transcription start site. A G + A sequence pattern is shown.


LOWER STRAND
UPPER STRAND
Hela
HepG2
Liv-E
Liv-L
RA-A
RA-B
prealb
No ext.
No exo III
G+A
-130→-130
-123→-120
-116→
-110
-80
-90
-100
*→-106
-110
*→-111 to -117
-120

failure of suppression by a non-homologous DNA competitor (RA-A.; -441 to -176). Furthermore the -202 to +40 region of the prealbumin promoter did not compete this stop. In addition to the exoIII stop at -123, a prominent additional stop occurred at position -130 when extracts from cultured cells were used. This included the Hela cell extract shown as well as the cell lines H4II, FaO, C2 and C2rev7 (data not shown). This extra band was also observed in extracts from the human hepatoma cell line HepG2 and hepatocytes cultured for periods of either 4 or 48 hours. Remember from chapter 1 that under the latter culture condition, albumin transcription declines at least 10-fold. Thus some of the factors capable of binding to the albumin promoter and producing the exoIII stop were still present in cultured cells that have lost the liver specific transcription profile.

Exonuclease digestion experiments designed to determine the 3' boundary of the protein binding region were also carried out with a probe labeled at -156 and cut at +47 so that exonuclease digestion now proceeds from downstream of the putative binding sites (in the opposite direction). ExoIII stops were observed (Figure 3.3, upper strand) indicative of protein binding, and again extracts from different cell-types gave similar protection results. A strong exoIII stop site was observed at -106 and weaker bands were found between -111 and -117. The site at -106 did appear to be stronger when extracts came from early cultured hepatocytes (Figure 3.3: Upper strand Liv-E vs. Liv-L), in this case correlating with albumin transcriptional regulation. Another

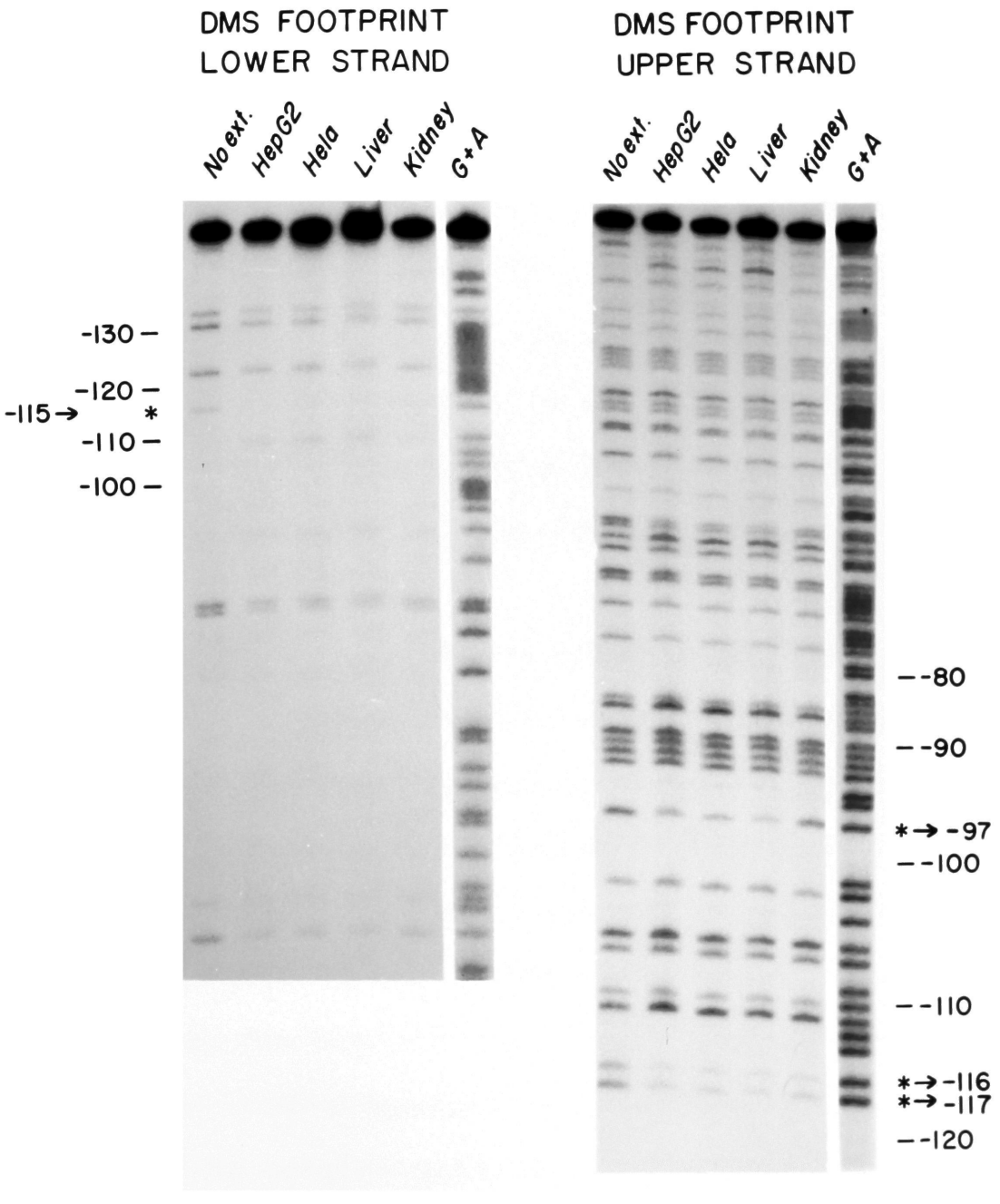
strong exoIII stop site was present at position -90. However, this site, like the –130 exoIII stop on the lower strand, was only present in cultured cell nuclei, and does not correlate with albumin gene expression. Additional exoIII stops are observed but they were present in all the samples, including those to which extract was not added, presumably resulting from difficulty in exoIII digestion of particular DNA sequences (ie. DNA secondary structures).

Protein footprinting of the albumin promoter

To provide evidence of specific sites of DNA-protein contact, suppression of DNA digestion by DNase 1 and methylation by DMS were assayed on DNA sequencing gels (Galas and Schmitz, 1978; Wildeman *et al.*, 1986). HepG2, Hela, liver and kidney nuclear extracts were each tested and incubated with the same uniquely end labeled probe (-156 to +47, labeled on either end), prior to DNase I or DMS treatment.

The DMS footprint on both strands of DNA is shown in Figure 3.4. On the lower strand the only protection observed was at the G residue at position -115. In addition a band at -110 (corresponding to an A residue) was rendered more sensitive to cleavage by all the extracts (Wildeman *et al.*, 1986). An approximately equal effect was afforded by proteins in all extracts, although note that different quantities of each extract was used; HeLa (25μg), HepG2 (2μg), kidney (20μg), and liver (10μg). Hence less liver or HepG2 protein is required to produce this protection than protein from non liver cell-types.

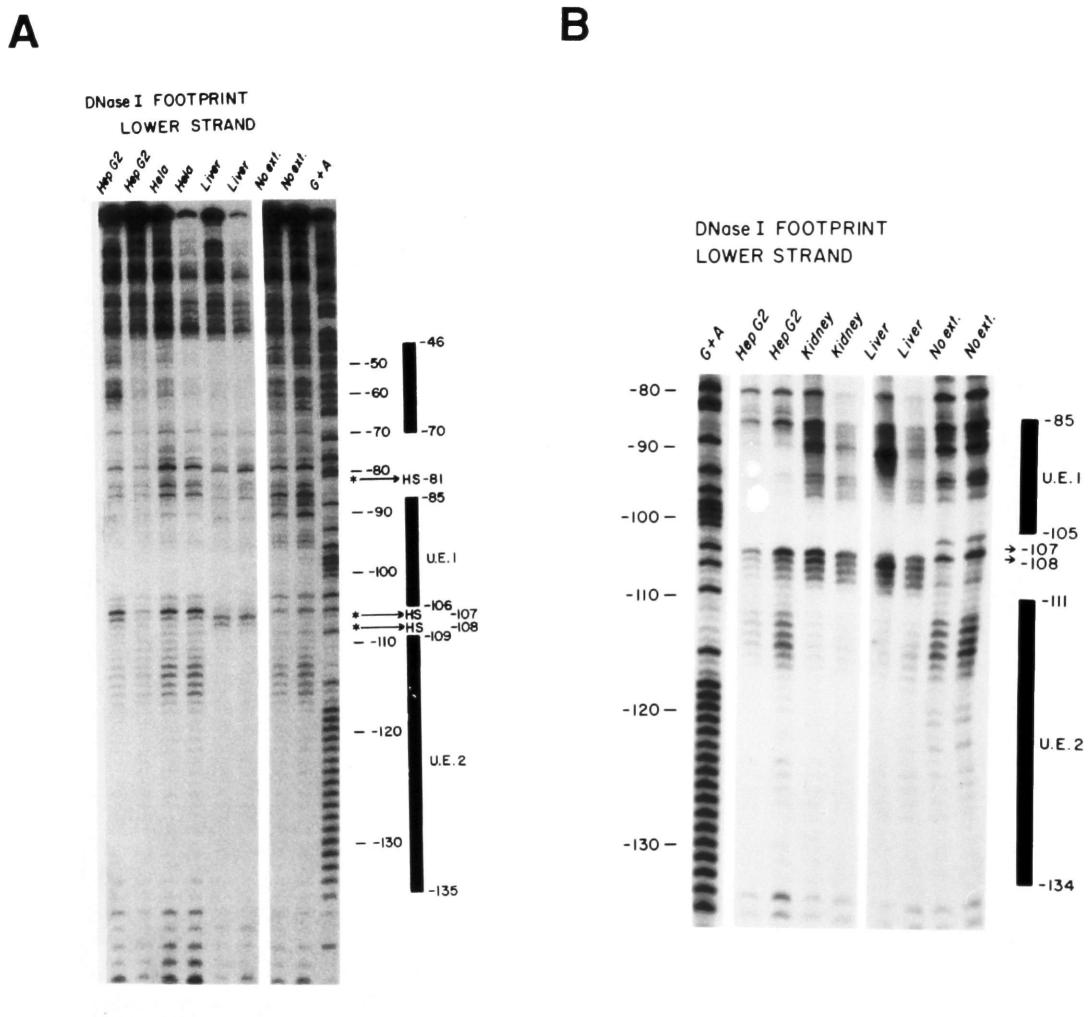
Figure 3.4. Methylation protection experiment to define the albumin DNA sequences that interact with nuclear factors. The -156 to +47bp DNA fragment was labeled on the upper or lower DNA strand and incubated with various concentrations of the following nuclear extracts; HepG2 (20μg), HeLa (25μg), kidney (20μg) and liver (10μg). The mixture was treated with DMS and the reaction products separated on an 8% polyacrylamide-urea gel; (see methods). Complete and/or partially protected G residues are indicated by an asterisk* and labeled according to their position relative to the albumin cap site. A (G+A) sequence reaction of the DNA fragment was electrophoresed along with each DMS protection experiment.


DMS FOOTPRINT
LOWER STRAND
No ext.
HepG2
Hela
Liver
Kidney
G+A
-130
-120
-115→ *
-110
-100
DMS FOOTPRINT
UPPER STRAND
No ext.
HepG2
Hela
Liver
Kidney
G+A
--80
--90
*→ -97
--100
--110
*→ -116
*→ -117
--120

On the upper strand no G residues were completely protected but several G residues were partially protected against methylation by all four extracts (e.g., those at -116 and -117). However, none of the G residues between nucleotides -116 and -97 were protected even though the exoIII stop studies indicated that proteins were bound in this region. Of course proteins might bind to only one side of the DNA, leaving G residues on the opposite side free to react. Also it should be pointed out that the DMS footprint assay could fail to detect interactions if only a limited number of the DNA molecules (e.g. 1/2) react with protein, since unreacted molecules would be free to interact with DMS and thus produce a band in the assay. This contrasts with the exo III assay where protein bindings to only some molecules will still lead to an observable result, for this assay scores for the presence as opposed to the absence of a specific signal.

DNase I footprinting studies were carried out using crude extracts derived from liver and kidney tissues as well as HepG2 and HeLa cells (Figure 3.5A) In all liver extracts three regions of protection were observed: a weak region -46 to -70 (just downstream from the CAAT; see Figure 3.6) ; -85 to -106 (upstream element 1, UE1; just upstream from the CCAAT) and -109 to -135 (upstream element 2, UE2). Additionally, on this same lower strand hypersensitive sites were observed at -108, -107 and -81. Figure 3.5B represents a separate DNase I footprint experiment on the same lower DNA strand. The coordinates of the protected sequences in these DNase I protection

Figure 3.5. Panel A: DNase I footprints induced by nuclear extracts on the rat albumin promoter. The -156 to +47 albumin promoter fragment labeled on the lower strand was incubated with 20μg of the extract shown above each lane. DNA-protein complexes were digested with DNase I at various concentrations depending on the extract used [HepG2 (5 and 15μg/ml), HeLa (15 and 20μg/ml), and liver (2.5 and 5μg/ml)] and the products were analyzed on 8% acrylamide gels. For the no-extract reactions, DNase I was used at a final concentration of 0.5 and 1μg/ml. A (G + A) sequence reaction of the DNA fragment is shown. For each extract, the DNase I concentration is increasing from left to right. Hypersensitive sites, DNA sequences that are cleaved by DNase I to a greater extent than free probe (lanes labeled "No ext."), are indicated by arrows and their position relative to the cap site for albumin transcription is given. Solid vertical bars show regions of DNA that are protected from cleavage by DNase I due to protein binding. Panel B: same as above except the DNase I concentrations used were as follows: HepG2 (15 and 35μg/ml, kidney (10 and 20μg/ml, and liver (10 and 20μg/ml, from left to right).


A
DNase I FOOTPRINT
LOWER STRAND
Hep G2
Hep G2
Hela
Hela
Liver
Liver
No ext.
No ext.
G+A
-46
-50
-60
-70
-70
-80
HS -81
-85
-90
U.E.1
-100
-106
HS
-107
HS
-108
-109
-110
-120
U.E.2
-130
-135
B
DNase I FOOTPRINT
LOWER STRAND
G+A
Hep G2
Hep G2
Kidney
Kidney
Liver
Liver
No ext.
No ext.
-80
-90
-100
-110
-120
-130
-85
U.E.1
-105
-107
-108
-111
U.E.2
-134

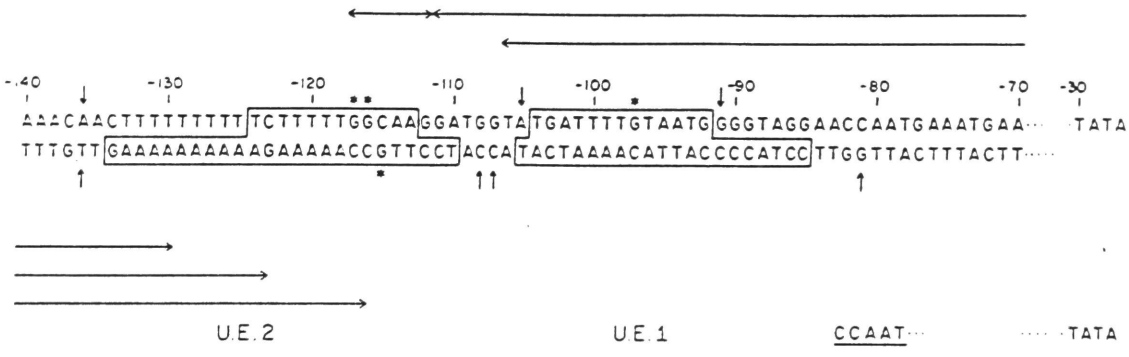


Figure 3.6. Summary of the results of exonuclease III, DNase I footprinting and DMS-methylation protection experiments on the rat albumin upstream promoter sequences. The DNA sequence of the relevant protein binding regions of the albumin promoter is shown. Solid horizontal lines with arrows above the upper and lower DNA sequence show the direction of exoIII cleavage and stop sites. Vertical arrows depict DNase I hypersensitive cleavage sites. Asterisks above G residues show the position of protection from DMS methylation. The sequences in boxes shows regions that are protected from DNase I cleavage on both the upper and lower DNA strands (the upper strand data was not shown). The upper and lower DNA sequences indicate the mRNA non-coding and coding strands respectively.

experiments and all the *in vitro* DNA binding reactions are summarized in Figure 3.6. In all the experiments, the liver cell extract produced the clearest DNase I footprint. The clear result from the two DNase I experiments shown was that all three regions could be footprinted from both extracts as long as a sufficient quantity of extract was used. In the experiments shown in Figure 3.5 and several others that are not shown, the amount of liver nuclear extracts required to give protection was about one-fourth that from Hela or kidney nuclear extracts.

Discussion

The 5' deletion analysis using recombinant adenoviruses defined a region between -156 and -43bp that is required for function of the rat albumin promoter in HepG2 cells. Since the intact promoter only functions in hepatocytes or hepatoma cells, these sequences must include those responsible for cell-specific transcription. Using four different *in vitro* approaches, including gel mobility shifts, exoIII stops and "footprinting" with DNase I and DMS, three regions of DNA-protein interactions were identified, a summary of which is shown in Figure 3.6. The regions protected by liver cell extracts are shown boxed in, and represent the average boundaries obtained from experiments shown in Figures 3.3 - 3.5 and additional experiments not presented.

While extracts from hepatocytes afforded the best protection, it was also

clear that extracts from cells that either do not express the albumin gene, or do so at a reduced rate also contained factors that would at least partially protect the same DNA sequences by both DNase 1 and DMS footprinting and would interact to produce the same gel shift and exoIII stops. There are at least three possible ways to interpret the apparent wide distribution of factors that can interact with the albumin promoter. First, the proteins recognizing these sequences in expressing cells versus non-expressing cells might be encoded by similar but non-identical members of a multigene family with only the member present in liver being capable of activating transcription. Alternatively, it could be that these DNA-binding proteins are in fact the same in the different cell-types but cell specificity is brought about by cell-specific modifications to or interactions with other regulatory proteins. These modifications could also include the packaging the DNA in an accessible chromatin form. Thirdly, it might be the case that the increased amounts of several binding proteins, perhaps interacting cooperatively, are able to combinatorily confer liver specific albumin transcription.

Studies by other laboratories of the albumin promoter

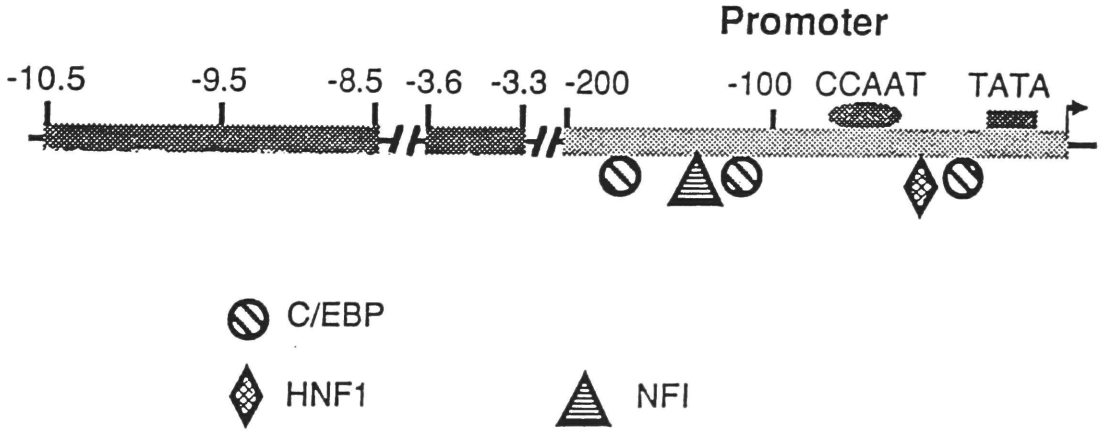
The mouse and rat albumin gene promoter sequences (which are virtually identical in their DNA sequence and protein binding patterns) have been well studied, and similar experiments to those described in this chapter have been published by other groups (Cereghini *et al.*, 1987; Lichtsteiner *et al.*,

1987; Cereghini *et al.*, 1988; Lichtsteiner and Schibler, 1989; Marie *et al.*, 1989; Tronche *et al.*, 1989). Using heparin agarose fractionated liver nuclear extracts a more detailed account of the multiple factors that interact with the albumin regulatory sequences has been obtained. Based on these results at least 6 regions of protection have been detected in addition to the TATA and CAAT binding elements. The current picture of the multiple DNA protein interactions occurring on the albumin promoter element is diagrammed in Figure 3.7.

Two of the binding regions which have been described in this thesis (UE1 and UE2) correspond to the C/EBP and NF1 binding sites depicted in Figure 3.7, and the more promoter proximal site which was described corresponds to the HNF1 site (Figure 3.7). The previous inability to observe the other three sites is readily explainable. For one thing, the two additional C/EBP sites have a much weaker affinity for the protein (greater than ten fold less; Lichtsteiner *et al.*, 1987), and hence are difficult to detect using crude nuclear extracts. Additionally, the binding of the CCAAT and HNF1 have been shown to be mutually exclusive (Lichtsteiner *et al.*, 1987), and therefore without using fractionated extracts it is difficult to observe DNase I protection upon these elements.

The results from these other studies support the initial conclusion that several of the factors which interact with the albumin promoter are present in cells which either do or do not transcribe albumin, though at somewhat

Figure 3.7. Summary of the known DNA-protein interactions that occur on the mouse and rat albumin gene promoters. At the bottom of the figure is a legend identifying each transcription factor. For a detailed explanation of this figure see the text.


Promoter
-10.5
-9.5
-8.5
-3.6
-3.3
-200
-100
CCAAT
TATA
C/EBP
HNF1
NFI

reduced levels in the non-liver tissues. Hence the albumin promoter is a multiprotein complex, with nearly every nucleotide bound by protein, and these control regions coincide with the sequence blocks that have been highly conserved throughout mammalian evolution (see Figure 1.5, Chapter 1). Of the factors that bind the albumin promoter only one is highly liver specific (HNF1), two are liver enriched (NF1 and C/EBP)[1] and the CAAT factor is present in all tissues. Therefore hepatocyte restricted albumin transcription seems to be controlled by several liver enriched binding factors interspersed with several ubiquitous elements, which combinatorily produce cell type specificity, for no single one of these factors is exclusively present in the liver. Interestingly two of these proteins, C/EBP and HNF1 interact with the regulatory regions of several other liver genes, and their possible involvement in coordinate liver gene control will be addressed in Chapters 8 and 9 of this thesis.

---

[1]The transcription factor NF1 was originally described to be a ubiquitous cellular protein (Jones *et al.*, 1987). Recent studies, however, suggest that a liver specific form of this factor exists, most probably generated by alternative RNA splicing (Lichtsteiner *et al.*, 1987; Paonessa *et al.*,1988).

Chapter 4

Functional Analysis of the Mouse Albumin Enhancer

Throughout the course of the rat albumin promoter studies described in the previous chapter, it was observed that the rate of transcription of the virally introduced albumin gene (containing a maximum of 441 base pairs of upstream promoter sequence) was much lower than that for the endogenous gene (Chapter 3; Friedman *et al.*, 1986; Babiss *et al.*, 1986b). This was determined by performing nuclear run-on transcription analysis on the viral DNA template and comparing it with the rate of transcription of the endogenous albumin gene. The ability to perform this type of analysis represents a major advantage of the viral vector system, as opposed for example, to the transient transfection assay, where far too few cells take up and express the introduced DNA to permit measurements of this type. (Babiss *et al.*, 1986b; Friedman *et al.*, 1986). While albumin cell-type specificity was observed, high rates of transcription, comparable to the endogenous gene in HepG2 cells, were only scored at high multiplicities of infection (when upwards of 300 copies/cell of template were used) or when enhancer elements (ie. the E1A enhancer) were added to the viral template (Friedman *et al.*, 1986). Hence while 441 nucleotides upstream of the albumin mRNA start site could dictate tissue specific transcription these sequences did not generate quantitatively correct levels of albumin expression (Babiss *et al.*, 1986b; Friedman *et al.*, 1986).

This difference in transcription rate can be explained in several ways. For one thing, it might represent an effect of localizing the albumin promoter

on the viral genome since viral templates might lack the essential chromatin structural configurations which play an active role in the regulation of the endogenous gene. Alternatively, the presence of the albumin promoter on the viral chromosome could itself lead to an inhibition of its expression. Equally likely, however, is the possibility that this 441bp promoter region is missing essential cis-acting DNA sequences involved in maintaining the high rate of albumin transcription (Friedman *et al.*, 1986; Babiss *et al.*, 1987).

As mentioned in the introduction (Chapter 1), previous work from our laboratory had identified a DNase I hypersensitive site at a position –2.8kbp upstream of the rat albumin promoter which correlates with the maximal rate of albumin transcription (Babiss *et al.*, 1986a), and it was subsequently determined that this region contains a binding site for the transcription factor C/EBP. Recombinant adenoviruses were constructed which contained these added DNA sequences and assayed for their expression in both HepG2 cells and primary rat hepatocytes. In neither case, however, was any enhancement observed from this region (even when the region containing the hypersensitive site was moved immediately adjacent to the 441bp promoter element).

Concurrently, Richard Palmiter's laboratory was performing an analysis of the mouse albumin gene using transgenic mice (Pinkert *et al.*, 1987). By analyzing mice which either contained 300bp of the mouse albumin upstream promoter element or as much as 12 kilobase pairs (kbp) of upstream

sequence fused to the human growth hormone gene, they identified a far upstream albumin enhancer element, capable of producing a 5-10 fold stimulation of albumin promoter directed expression. By analyzing several additional transgenic constructs, Palmiter's group further determined the functional enhancer sequences to reside in a region between –10.4 and –8.5kbp upstream of the RNA start site. Notably, removal of the most 3' part of this region (from –8.5 to –9.3kbp) enabled the remaining DNA sequences (–10.4 to –9.3kbp) to produce weak enhancement in certain non-liver tissues (e.g. lung, spleen, kidney; Pinkert *et al.*, 1987). This suggested that separate domains might be involved in determining both the specificity and rate enhancing functions of this enhancer.

I have examined this region of the mouse albumin gene that in transgenic mice is needed for full hepatocyte specific expression. I was interested in identifying the cis-acting elements and the proteins that might bind to the enhancer region to compare them to those that interact with the albumin promoter proximal region (Chapter 3), and to the known factors that activate hepatocyte-specific expression in several other genes (e.g. TTR and α1-antitrypsin). To perform this analysis, however, a transient transfection assay was required, for transgenic mice are not an efficient means of assaying the large number of constructs involved in this type of functional enhancer dissection.

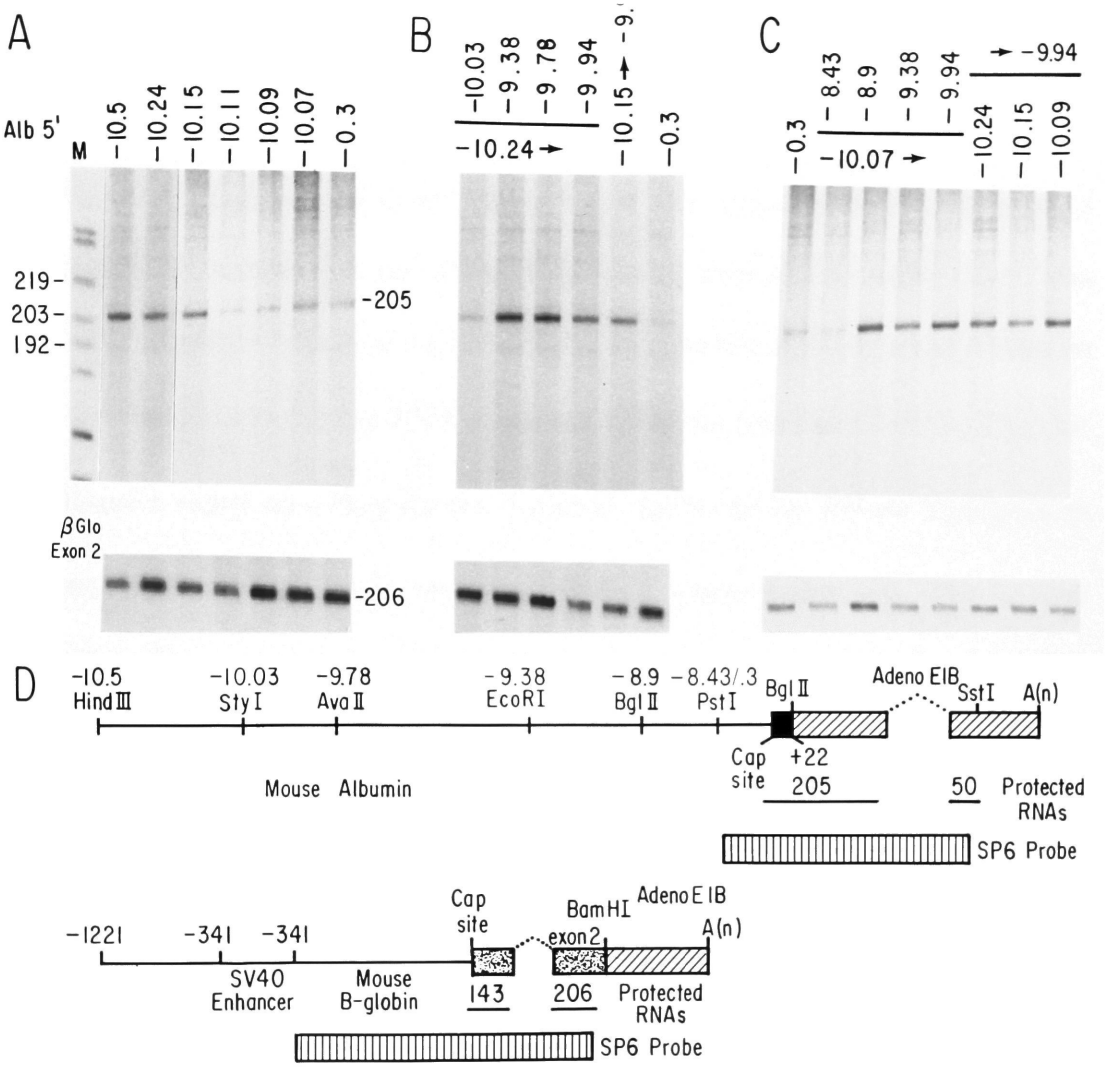
Transient transfection of the mouse albumin enhancer in CWSV1 Cells

To specifically localize the mouse albumin DNA sequences that function to increase the rate of transcriptional initiation from the albumin cap site, transient plasmid transfection assays using rat CWSV1 cells were performed (Figure 4.1). Since these cells transcribe albumin at a rate about 50% of that in the rat liver, they seemed to represent a good choice for use in studying the sequences involved in enhancing albumin transcription (see Chapter 1; Table 1.1). By comparison, the HepG2 cell line which had been used for the promoter analysis only transcribes its endogenous albumin gene at 10% that of the liver gene (Clayton *et al.*, 1985), and it was initially thought that this might be due to an inability of the enhancer to function in these cells.

The parental plasmid used contained albumin promoter proximal sequences extending from -10.5 to -8.43kbp relative to the transcriptional start site fused to albumin sequences from -300 to +22 nucleotides (Figure 4.1D). The mouse albumin sequence (-10.5 to 8.43kbp followed by -300 to +22bp) was then joined to the first two exons of the adenovirus E1b gene including the polyA+ addition sequences needed for the production of a stable mRNA (Figure 4.1D).

A 5' deletion analysis was performed using constructs beginning at –10.5kbp (a nominal number because the entire 2 kilobase pair (kbp) upstream region has not been sequenced) with a constant 3' boundary at –8.43kbp.

Figure 4.1. Functional analysis of the mouse albumin enhancer element in CWSV1 cells. The parent construct (top, D) and deletions (marked by position in kbp) were co-transfected into CWSV1 cells with a control plasmid (an SV40 enhancer containing β-globin promoter (bottom, D) to standardize for variations in transfection efficiency (see Materials and Methods, Chapter 2). Assays for expression were carried out by protection of labeled sp6 probes (shown in D) as described in the text. In all panels the lane marked 0.3 refers to an enhancer-less mouse albumin promoter construct (containing the -300 proximal promoter element.). In Panel A, all constructs have a 3' boundary of -8.43kbp and a 5' boundary as marked. In Panels B and C the introduced enhancer sequences are as marked.


A
Alb 5'
M
-10.5
-10.24
-10.15
-10.11
-10.09
-10.07
-0.3
219-
203-
192-
-205
βGlo
Exon 2
-206
B
-10.03
-9.38
-9.78
-9.94
-10.15
-10.24
-0.3
C
-0.3
-8.43
-8.9
-9.38
-9.94
-9.94
-10.07
-10.24
-10.15
-10.09
D
-10.5
Hind III
-10.03
Sty I
-9.78
Ava II
-9.38
EcoRI
-8.9
Bgl II
-8.43/.3
Pst I
Bgl II
Adeno E1B
Sst I
A(n)
Cap
site
+22
205
50
Protected
RNAs
Mouse Albumin
SP6 Probe
-1221
-341
-341
Cap
site
BamHI
exon2
Adeno E1B
A(n)
SV40
Enhancer
Mouse
B-globin
143
206
Protected
RNAs
SP6 Probe

Next, a 3' deletion set was constructed using -10.24kbp as a 5' boundary and deleting sequence in the 3' direction as marked (Figure 4.1 B, C). These plasmids were then tested for the effects of upstream enhancer sequence on expression by transient transfection assay in the rat CWSV1 cells. All transfection experiments included an independent plasmid containing a SV40 enhancer driven β–globin promoter which was used to monitor transfection efficiency (Figure 4.1D). Since the experimental and control constructs protected similar sized fragments, expression from each was assayed in separate reactions by hybridizing two different ^{32}P-labeled antisense RNA probes to cytoplasmic RNA prepared from the transfected cells (25μg for the albumin assay and 10μg for the β-globin) followed by RNase T_2 digestion and electrophoresis of the T_2 resistant RNA on a sequencing gel. The labeled antisense RNA probe for albumin expression was protected over 205 nucleotides representing the fusion exon made up of the first 22 bases of albumin gene added to 183 base pairs of the E1b first exon (see Figure 4.1D). A smaller 50 nucleotide protected fragment (not shown) corresponds to part of the adenovirus second exon. A 206 nucleotide band representing the second exon of the β-globin gene was protected (see figure 4.1D), along with a 143 nucleotide band corresponding to the first exon (not shown in Figure 4.1).

The autoradiograms of gels representing the RNase T_2 analysis of three separate experiments are presented in Fig. 4.1. The entire enhancer region

stimulated albumin gene expression by at least 5 fold in these cells (Fig. 4.1A, –10.5 vs –0.3), a result similar to the 5 to 10 fold stimulation reported in transgenic mice (Pinkert *et al.*, 1987). This stimulation was lost when sequences between –10.15 and –10.11Kbp were removed (Figure 4.1A), implicating this region as having a major role in enhancer function.

The initial 3' deletion construct had a 5' boundary at –10.24kbp (therefore containing the 5' activating region mentioned above) and an initial 3' boundary at –8.43kbp. Sequences were removed to –9.3kbp (EcoRI site), –9.78kbp (AvaII site) or –9.94kbp (AccI site; Figure 4.1B). Following transfection of CWSV1 cells, no effect was noted upon removal of these regions. However, removal of sequence to –10.03kbp completely abolished transcriptional stimulation, identifying an essential 3' sequence between –10.03 and –9.94kbp (Figure 4.1B). This observation, along with the 5' analysis described above, identified the minimal enhancer element as a nominal 210 nucleotide sequence (–10.15 to –9.94kbp; Figure 4.1B).

A detailed examination of expression from further recombinant constructs within this region yielded a surprising result. A plasmid containing sequences from –10.07 to –8.43kbp, lacking 80 nucleotides from the essential 5' element delineated above, failed to generate any enhancer function in CWSV1 cells (Figure 4.1 A, C). However, removal of 1550, 850 or 470 nucleotides at the 3' side of this segment (between –8.43 and –9.94kbp) reactivated enhancer activity from the previously 5' deleted plasmid (Figure

4.1C). Furthermore, return of the DNA sequence at the 5' side to –10.09, –10.15 or –10.24kbp gave no boost above the already maximal transcriptional level obtained from a plasmid whose 5' boundary was at –10.07kbp (Fig. 1C). Therefore , if at least 470bp of 3' sequence (from –8.9 to –8.43kbp) was deleted, then complete enhancement in CWSV1 cells was obtained by the 130 nucleotides from –10.07 to –9.94kbp. Furthermore, this minimal element could function in either orientation (data not shown) .

Summary of SV1 Transfection Analysis

These analyses established three DNA regions as contributing to the function of the mouse albumin enhancer. Alone, Region II (–10.03 to –9.94kbp) is capable of enhancing the albumin promoter. When Region III –8.9 to –8.43kbp) is added to Region II the enhancement is negated. Region I –10.15 to –10.11kbp) alone is not active, but its addition restores the enhancer activity that is blocked by Region III (see Figure 4.2 for a summary of these three regions). Table 4.1 summarizes the functional consequences of the eight possible combinations of these three elements. Over the course of many experiments the mean enhancer effect was 5 fold. Hence, a value of 100% has been given to constructs demonstrating maximal enhancement while a level of 20% represents the enhancerless promoter construct.

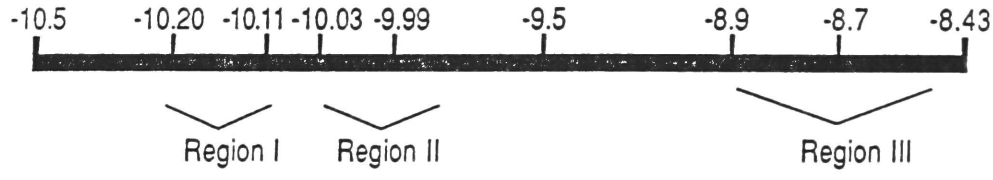


Figure 4.2. Summary of the three functionally important regions of the mouse albumin enhancer as identified by CWSV1 transfection analysis. Three functional regions: I, II, and III (shown by the brackets below the line drawing) were delineated. Region I (-10.15 to -10.11kbp) alone does not contribute to the rate of albumin gene expression, but serves to override the negative effect of region III (-8.9 to -8.43kbp). Region II (-10.03 to -9.94kbp) contains the DNA sequences that contribute positively to the increased rate of transcription initiation from the albumin promoter.

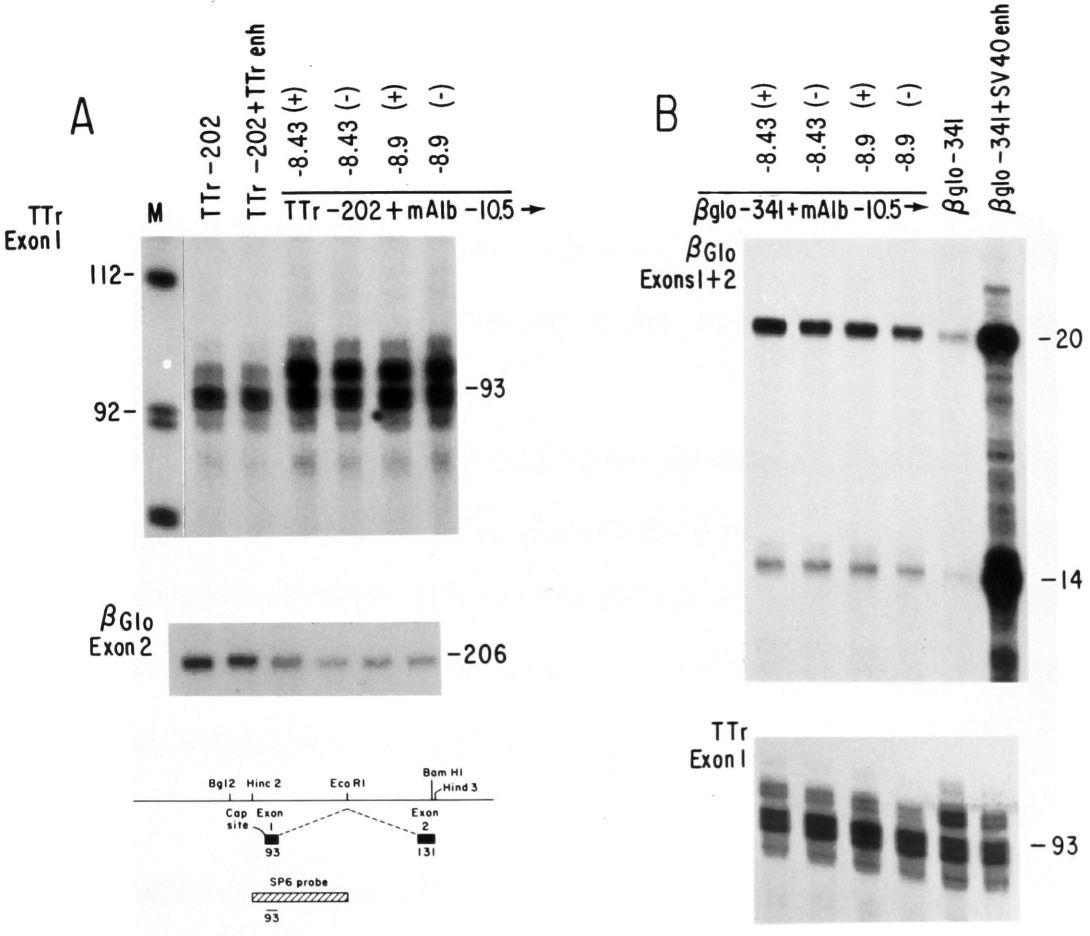
I	II	III	Relative Expression
-	-	-	20%
+	+	+	100%
-	+	+	20%
-	+	-	100%
+	+	-	100%
+	-	-	20%
-	-	+	20%
+	-	+	20%

Table 4.1. Combinatorial effects of the three functional enhancer regions in CWSV1 cells. Over the course of many experiments, the average enhancer effect was five fold which has been arbitrarily ascribed a value of 100%. The basal level of transcription from the enhancerless mouse albumin promoter (–300) is 20%. The transfection results upon which the results in this table are based are all presented in Chapter 4 except the final permutation (+, -, +) which represents data not shown.

The mouse albumin upstream enhancer can activate a heterologous promoter in an orientation independent manner

Three properties usually ascribed to enhancers are the ability to function at a distance (in a position independent fashion), an orientation independence and the ability to function to activate a heterologous promoter (Serfling, 1985; Maniatis *et al.*, 1987; Guarente, 1988). The sanctity of this first criterion seems assured, for the albumin enhancer is found greater than 10kbp upstream of the albumin start site, where it was shown by Pinkert *et al.*, (1987), to function in transgenic mice. To test the remaining facets of enhancer function in cultured cells, the –10.5 to –8.43kbp enhancer fragment or a –10.5 to –8.9Kbp enhancer fragment (lacking the negative element described above) was placed in both orientations in front of either a promoter active in hepatocytes and hepatoma cells, the mouse transthyretin (TTR) promoter (Costa *et al.*, 1986), or the mouse β-globin promoter (Konkel *et al.*, 1987), a promoter that is by itself minimally active in most all cells. These constructs along with control plasmids were used to carry out transfections of several cell types. The assay of the transthyretin (TTR) promoter depends on a nominal 93 base pair RNase T_2 protected band that is complementary to the TTR first exon (Fig. 4.3A, bottom, Costa *et al.*, 1986). The TTR -202 promoter element functions without the albumin enhancer to give a messenger RNA that protects a 93 and 92 nucleotide doublet following RNase T_2 digestion plus a small amount of a doublet several bases longer (nominally 95 and 96 bases).

Figure 4.3. The mouse albumin enhancer can activate heterologous promoters of both liver and non-liver specificity following transfection into rat CWSV1 cells. In Panel A, the mouse albumin enhancer sequences from either -10.43 to -8.43kbp or -10.5 to -8.9kbp (the region between -8.43 to -8.9kbp is the negative region) were placed in either orientation (designated +/-), 5' to the -202 position of the previously described mouse Transthyretin (TTR) promoter element (Costa *et al.*, 1986, 1988a). Also in this assay are a -202 TTR promoter construct (an enhancerless control), and this same construct containing the 100 nucleotide TTR upstream enhancer element (Costa *et al.*, 1986, 1988a). The sp6 RNA probe specific to the 5' end of the TTR mRNAs, protected a 93 nucleotide RNA as depicted below. A SV40 enhancer containing mouse β-globin promoter construct is included as an internal control (see figure 1), and the 206nt second exon is shown. In B, these same sequences are placed 5' to the β-globin promoter, in either orientation (+/-). An enhancerless construct β-glo -341 along with the SV40 enhancer containing construct are also assayed. For this series, the TTR -202 promoter construct is used as an internal control.


A
M
TTr -202
TTr -202+TTr enh
-8.43 (+)
-8.43 (-)
-8.9 (+)
-8.9 (-)
TTr -202+mAlb -10.5
TTr Exon I
112-
92-
-93
βGlo Exon 2
-206
Bgl2
Hinc 2
EcoRI
Bam HI
Hind 3
Cap site
Exon 1
Exon 2
93
131
SP6 probe
93
B
-8.43 (+)
-8.43 (-)
-8.9 (+)
-8.9 (-)
βglo-341+mAlb -10.5
βglo-341
βglo-341+SV40enh
βGlo Exons1+2
-20
-14
TTr Exon I
-93

The albumin enhancer, in either orientation (+ or -), with and without the DNA sequences of region III between -8.9 to -8.43kbp, did not affect the promoter function that gives the 92 or 93 nucleotide bands but did increase by five-fold the message from starts that occur just upstream. This partial activation of the TTR promoter by the albumin enhancer contrasts with the complete lack of effect in the CWSVl cells of the TTR enhancer element on its own promoter (Figure 4.3A, lane 3 compared to lane 2). Thus, the TTR enhancer element from the mouse gene that functions in human hepatoma cells (Costa *et. al.*, 1986, 1988a) does not function in CWSV1 cells, indicating that these CWSV1 cells may not engage in completely normal hepatocyte transcription.

Figure 4.3B proves that the albumin enhancer is also capable of stimulating transcription of a foreign (non-liver) promoter in CWSV1 cells. Either the complete enhancer (–10.5 to –8.43kbp) or the Region I plus Region II fragments alone in either orientation, inserted at the –341 promoter position increased β -globin expression by five-fold (Fig. 4.3B).

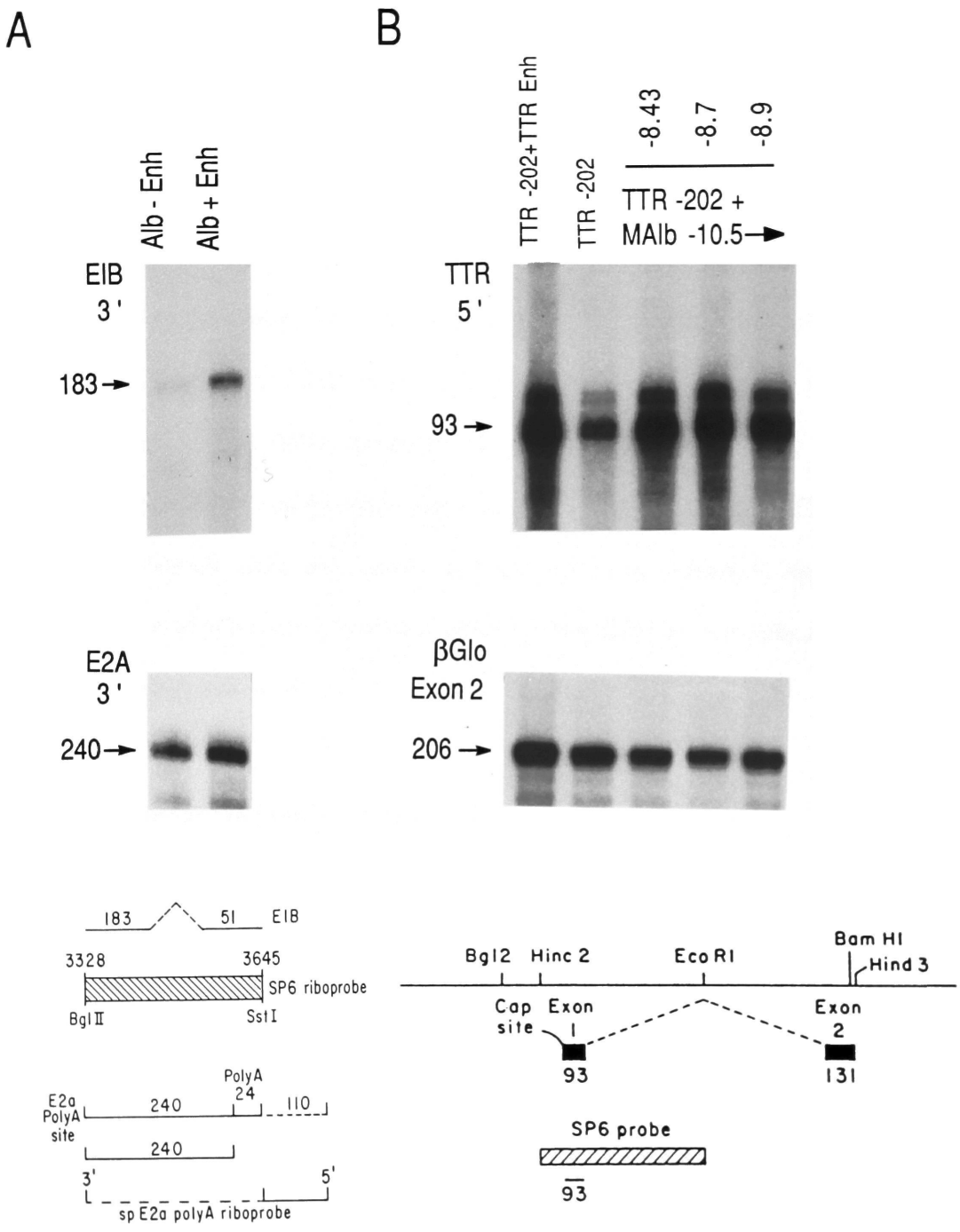
The mouse albumin enhancer will function in HepG2 cells

While CWSV1 cells were used to analyze the regulation of the albumin enhancer, most studies of this kind have used rodent and human hepatoma cell lines (including FaO and HepG2). One problem with hepatoma cells, however, is that they only transcribe their endogenous albumin genes at

roughly 10% the level found in the rat liver (Clayton *et al.*, 1985; Babiss *et al.*, 1987; Friedman *et al.*, 1987, Table 1.1), suggesting the possibility that the albumin enhancer may not be contributing to the rate of albumin transcription in these cells. I was interested in addressing this question but as perviously described, the albumin promoter region gives very weak signals upon transient transfection into HepG2 cells (Chapter 3, Babiss *et al.*, 1987). The enhancer was therefore tested in HepG2 cells using a recombinant adenovirus construct containing the albumin upstream enhancer (–10.5 to –8.43kbp) fused to the albumin promoter and an adenovirus reporter gene (the E1B promoter; Figure 4.4A). In addition, plasmids were prepared for transfection that had the albumin enhancer driving the promoter of the mouse transthyretin gene which is known to be active and susceptible to stimulation by its own enhancer in HepG2 cells (Figure 4.4B).

The promoter alone (-300 to +22) or the promoter plus the –l0.5 to –8.43kbp enhancer element were separately inserted into an adenovirus vector as previously described (Chapter 2; Babiss *et al.*, 1986; Friedman *et al.*, 1986). Upon infection with recombinant viruses, the mouse albumin regulatory sequences resulted in production of a stable RNA species generated by transcription initiation at the albumin start site followed by splicing of the fusion albumin-E1b exon (Figure 4.4, bottom). Expression from these plasmids was scored using an antisense RNA probe specific for the first two exons of the adenovirus E1B gene (probe sp641, Friedman *et al.*, 1986). The

Figure 4.4. The mouse albumin enhancer functions in human hepatoma cells (HepG2 cells). Panel A represents infections (50pfu/cell for 24h in the presence of AraC using recombinant adenoviruses containing the albumin -300 promoter (Alb - enh) or the same construct containing the -10.5 to -8.43Kbp enhancer sequence (alb + enh) into HepG2 cells. Following isolation of cytoplasmic RNA, the accumulation of exogenously expressed albumin-viral mRNA is scored using the probe depicted in the diagram. Expression of the viral E2A gene was also quantitated to monitor variations in virus uptake (Babiss *et al.*, 1987). In panel B, the same mouse albumin enhancer/TTR heterologous constructs (50μg) depicted in Figure 4.3 are transfected into human HepG2 cells, with the SV40 enhancer containing β-globin plasmid (10μg) as an internal control. Sequences are again placed 5' to the TTR -202 nucleotide and the 3' boundary of mouse albumin enhancer sequence is labeled above each lane.


A
B
Alb - Enh
Alb + Enh
EIB
3 '
183
E2A
3 '
240
TTR -202+TTR Enh
TTR -202
-8.43
-8.7
-8.9
TTR -202 +
MAlb -10.5
TTR
5 '
93
βGlo
Exon 2
206
183
51
EIB
3328
3645
SP6 riboprobe
BglII
SstI
PolyA
24
E2a
PolyA
site
240
110
240
3'
5'
sp E2a polyA riboprobe
Bgl2
Hinc 2
Eco RI
Bam HI
Hind 3
Cap
site
Exon
1
Exon
2
93
131
SP6 probe
93

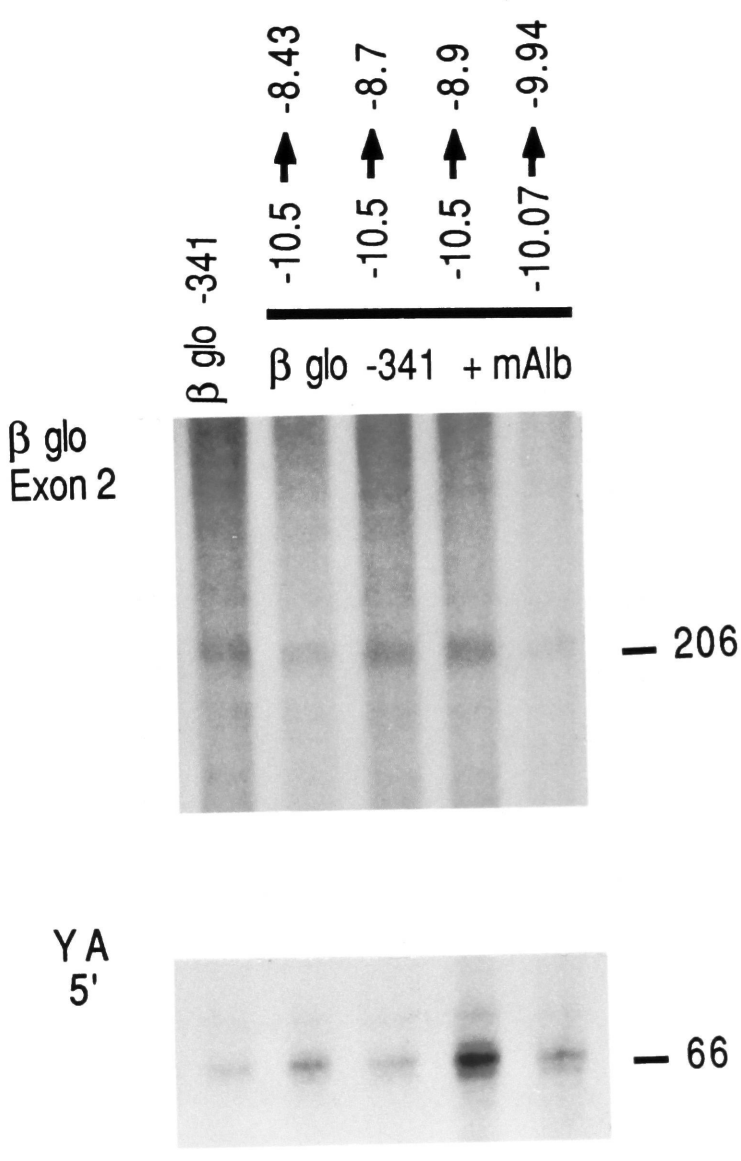
enhancer containing construct produced a five to ten fold stimulation compared to the construct beginning at -300 of the albumin gene (Figure 4.4A). As an internal control, expression of the viral E2a gene was measured in each infection.

Additionally, HepG2 cells were transfected with plasmid constructs containing the TTR promoter plus either the albumin or TTR enhancers (the chimeric albumin enhancer/TTR promoter constructs are identical to those described in the previous chapter). The clear result was that the albumin enhancer did stimulate TTR transcription in these cells and the stimulation was from both RNA start sites, enhancing equally the 93-94 and 95-96 doublets, in a fashion nearly equivalent to that achieved by the TTR enhancer itself. Thus, HepG2 cells can surely use the albumin enhancer sequences to increase TTR and albumin promoter activity five fold as was the case in the CWSV1 cells (Fig. 4.4 A, B).

The mouse albumin enhancer is cell-specific

While the albumin enhancer can function in both CWSV1 and HepG2 cells, it did not function in HeLa cells. The plasmid constructs containing the albumin enhancer plus the β -globin promoter or the β-globin promoter alone were introduced into HeLa cells (Figure 4.5) Neither the complete enhancer sequences (–10.5 to –8.43kbp) nor any part of the enhancer including Region I and II together or II alone (–10.5 to –8.9kbp and –10.07 to –9.94kbp)

Figure 4.5. The mouse albumin enhancer does not function in Hela cells. This figure depicts transfection of different portions of albumin enhancer sequence (boundaries shown) placed 5' to the mouse β -globin promoter (at base pair -341) into Hela cells. The specific β-globin protected RNA is 206 nucleotides as shown in figure 4.1 (the second exon). The glutathione-S- transferase YA gene which is active in Hela cells is included as an internal control, and produces a 66 nucleotide protected RNA as shown.


β glo -341
-10.5 → -8.43
-10.5 → -8.7
-10.5 → -8.9
-10.07 → -9.94
β glo -341 + mAlb
β glo
Exon 2
206
YA
5'
66

resulted in a decrease in β-globin expression. These constructs were also introduced into MC cells, a neuroblastoma line, with like negative results (data not shown). These transfections were internally controlled by the inclusion of a construct which contained the glutathione-S-transferase YA promoter which has been shown to be active in HeLa cells (Paulson *et al.*, In Press). It was assayed by and RNase T_2 protection assay which protected a RNA of 66 nucleotides (Figure 4.5, bottom). Hence each part of the albumin enhancer functions in a cell-type specific fashion, including the region II sequences, which alone can produce the entire positive enhancer effect.

Discussion

The functional analysis of the albumin enhancer has shown it to be a complicated genetic locus containing both positive and negative regulatory elements. The modularity of this element is similar in many respects to other complicated regulatory regions such as those found in the yeast haploid specific genes (Herskowitz, 1989), the c-myc gene (Hay *et al.*, 1987, 1989) and the SV40 enhancer region (Zenke *et al.*, 1986; Herr and Clark, 1986). Furthermore, the SV40 enhancer has recently been demonstrated to contain individual modules termed "enhancons" each of which interact with a single protein, and these units function in pairs (which in some cases require specific partners) to facilitate the diverse transcriptional effects of that regulatory element (Herr and Clark, 1986; Ondek *et al.*, 1988). The albumin

enhancer regions could function in the same way.

The ability of the enhancer to function in hepatoma cells is deserving of special note. Albumin is an adult liver specific gene, only reaching its maximal level of expression several weeks following birth in the rodent (Camper and Tilghman, 1989) In fact, Tilghman and coworkers have suggested that the albumin enhancer does not function in the fetus and that instead the albumin gene is activated during fetal development by the neighboring α-fetoprotein enhancers. While the α-fetoprotein enhancer could still play role, the fact that the albumin enhancer functions in these partially differentiated HepG2 cells demonstrates that the albumin enhancer can work to some extent in more fetal cell types.

Thus as was the case for the promoter (chapter 3), multiple regulatory elements are involved in regulating the enhancer, yet the two differ significantly since the albumin enhancer contains a negative region and a unique type of modulatory element. Having identified these three important regions, I proceeded with an *in-vitro* DNA protein binding analysis to identify the transcription factors which interact with them, with the goal of better understanding the mechanism by which this enhancer exerts its liver specific effect. Chapters 5-7 sequentially present the results of protein binding studies on each of these regions.

Chapter 5

Albumin Enhancer Region I

As described in the preceding chapter, three regions of functional importance were identified based on a deletion analysis of the albumin enhancer in CWSV1 cells. From these studies, the enhancer was shown to consist of both positive and negative regulatory sequences: a negative region (III) that suppresses an otherwise positive, liver-specific enhancer element (II) and a third region that by itself is inactive but in conjunction with the remaining elements overrides the effect of the negative region (I). To gain a better understanding of the mechanism by which these cis-acting elements function, each of these regions was analyzed for DNA-protein interaction using a combination of *in vitro* binding analyses. Region I, the overriding element is described here while the other elements are presented in successive chapters.

Two nuclear protein-binding sites within the mouse albumin enhancer are recognized by C/EBP, a previously identified liver enriched transcription factor

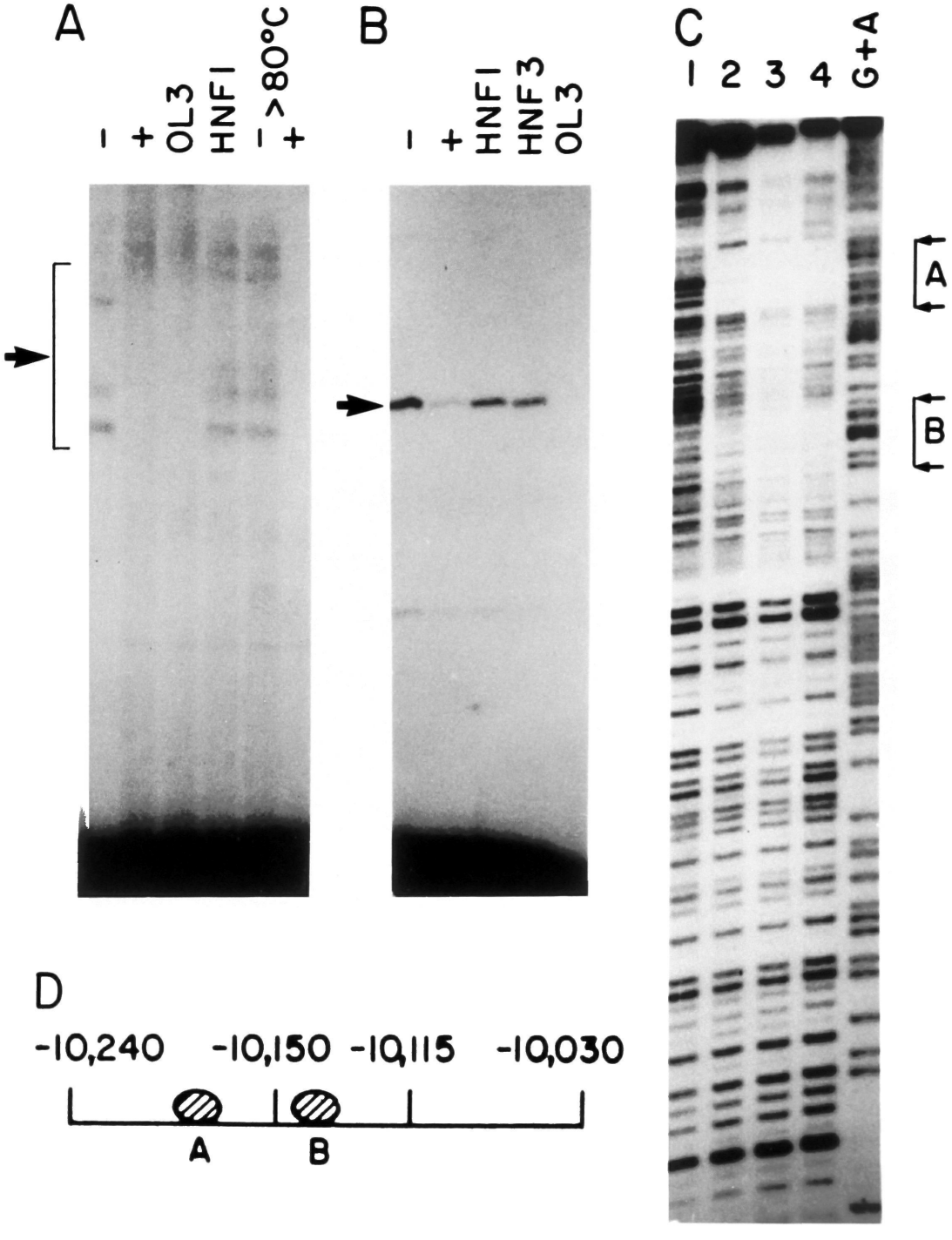
The search for protein binding sites within the enhancer began with a 35 nucleotide probe (-10.15 to -10.115) containing the functionally acting Region I sequences (deletion of DNA sequence from -10.15 to -10.11kbp completely ablated enhancer function; see chapter 4). Protein binding was first explored using the gel retardation shift assay which takes advantage of the fact that end-labeled DNA complexed with protein migrates more slowly than free-DNA (Fried and Crothers, 1981; Garner and Revzin, 1981). When mixed

with hepatocyte extract the labeled -10.15 to -10.115Kbp probe produced three gel shift bands (Fig. 5.1A). Interestingly, these bands were competed during DNA-protein interaction both by an unlabeled, homologous DNA fragment and by an oligonucleotide representing a binding site in the TTR enhancer known to bind the liver enriched transcription factor C/EBP (TTR oligonucleotide 3, Costa *et al.*, 1988a). No competition, however, occurred when the labeled -10.15 to -10.115kbp Region I probe was competed with an oligonucleotide binding site for hepatocyte-specific nuclear HNF1. Furthermore, the binding activity in Region I was heat stable (80°C), a known characteristic of the C/EBP protein (Johnson *et al.*, 1987; Costa *et al.*, 1988b). Lastly, purified C/EBP/β–galactosidase fusion protein (gift of S. McKnight) produced a gel shift with this probe that was competed by TTR oligonucleotide 3, but not by oligonucleotides binding sites for HNF1 or HNF3 (Figure 5.1B). Hence, all these lines of evidence strongly implicated C/EBP as the protein which is interacting with the Region I sequences, although point mutagenesis of these binding sites in conjunction with a functional analysis of the mutants is certainly warranted to confirm this model.

DNase I footprinting of the C/EBP binding site (s)

To more precisely locate the C/EBP binding site within Region I, a 210 nucleotide fragment (-10.240 to -10.030kbp) was end labeled and a DNase I

Figure 5.1. Definition of a C/EBP site(s) within the mouse albumin enhancer Region I. In panel A the mouse albumin HindIII/Rsa1 fragment (–10.15 to –10.115kbp, 1ng) was 3' end labeled and incubated with 10μg of liver nuclear extract, and 4μg of poly (dIdC): poly (dIdC) in the absence (-) or presence (+) of homologous unlabeled DNA (50 fold molar excess). The binding reaction (30min at 20°C) was then analyzed for gel retardation on an 8% acrylamide gel. The unbound and hence faster migrating probe is shown at bottom while a doublet of specific retarded bands are depicted in the region flanked by the arrow. Oligonucleotides (in 50 fold molar excess) representing a known C/EBP binding site (0L3) and a binding site for HNF1, were assayed for competition as well. The lanes labeled 80°C represent reactions where the liver extract was heated prior to incubation with the probe. In panel B, the same probe is assayed but now using a C/EBP/β-galactosidase fusion protein (gift of S. McKnight; see text), in the presence (+) or absence (-) of specific competitor. Again, oligonucleotides representing binding sites for C/EBP (OL3) and HNF1 were used in addition to an oligonucleotide with a site for HNF3. In C, DNase I footprints of C/EBP binding sites in the mouse albumin enhancer, using either oligonucleotide affinity purified C/EBP protein (lanes 2 and 3) or C/EBP β-galactosidase fusion protein (lane 4). For this probes labeled uniquely at the 5' end spanning from the HindIII site at -10.24kbp to the Sty1 site at -10.03kbp were incubated with buffer alone (lane 1), affinity purified protein (lane 2- 5ng, lane 3, 20ng), or fusion protein (lane 4, enriched bacterial extract), treated with DNaseI, and the products analyzed by polyacrylamide gel electrophoresis. Two regions of protection from DNaseI digestion are denoted as A and B.


A
- + OL3 HNF1 - >80°C +
B
- + HNF1 HNF3 OL3
C
- 2 3 4 G+A
A
B
D
-10,240 -10,150 -10,115 -10,030
A
B

footprint analysis was performed using protein purified by oligonucleotide affinity chromatography [using oligonucleotide 2 of TTR (Costa *et al.*, 1988a)] or an the *E. coli* produced fusion protein containing a 20 KD segment of C/EBP (as above). Two clear footprints designated sites A and B were observed on both DNA strands, though only the results from the lower strand are shown in Figure 5.2. Thus, not only did C/EBP bind within the 35 nucleotide functionally important region used in the first gel shift experiment (the B site, Figure 5.1D) but to a second site as well (the A site, Figure 5.1D). The specificity of this second site for C/EBP was confirmed by a gel shift assay which demonstrated that the complex was specifically competed with TTR oligomer 2, another known C/EBP binding site (Figure 5.2, Costa *et al.*, 1988a). In addition, the binding activities interacting with both these sites were liver enriched, producing retarded shift complexes with these probes only in liver and not kidney, brain or spleen nuclear extracts (data not shown), as has been previously described for the C/EBP protein (Grayson *et al.*, 1988a; Costa *et al.*, 1988a, 1988b). In light of this redundancy of C/EBP sites, Region I for the discussion has been expanded to include both sites A and B.

Discussion

Sequence comparison of C/EBP binding sites

Sequence comparison of the footprint region showed a good match

Figure 5.2. Identification of a second C/EBP site by gel shift analysis using a DNA probe extending from -10.24 to -10.14kbp. This probe was uniquely end labeled and subjected to a gel retardation assay as in figure 5.1. Several shifted complexes are observed in the absence (-) but not in the presence (+) of a 50 fold molar excess of self competitor fragment. Competition assays using specific DNA fragments were as previously described. TTR Ol2 represents a known C/EBP binding site (Costa *et al.*, 1988b).

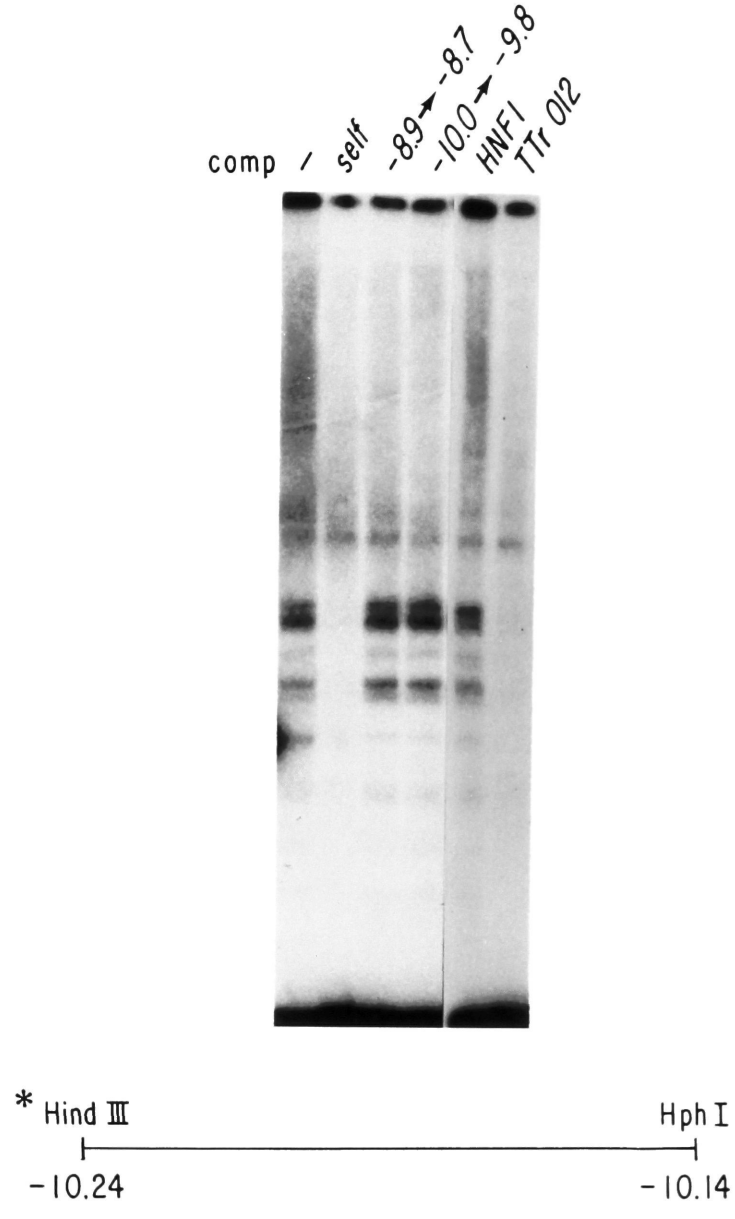
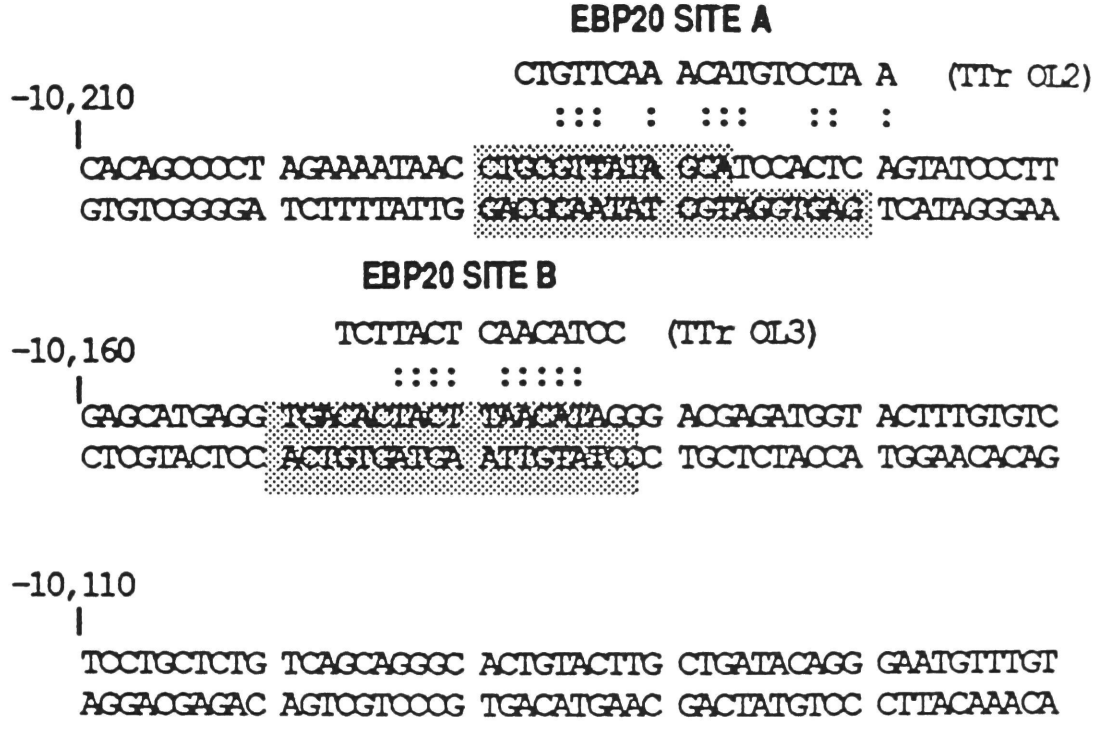

comp
-
self
-8.9→-8.7
-10.0→-9.8
HNF1
TTr 012
* Hind III
Hph I
-10.24
-10.14

Figure 5.3 Summary of the DNA-protein interactions occurring on the mouse albumin enhancer Region I. This region has two sites for the protein factor C/EBP. The sequence of this region is shown and the footprintable regions (labeled A and B) respectively are highlighted. Shown above each region is the best homology match with other known C/EBP sites (Costa *et al.*, 1988b).


EBP20 SITE A
CTGTTCAA ACATGTCCTA A (TTr OL2)
-10,210
CACAGCCCCT AGAAAATAAC CTGCGTATA GCATCCACTC AGTATCCCTT
GTGTCGGGGA TCTTTTATTG GACGCATAT CGTAGGTGAG TCATAGGGAA
EBP20 SITE B
TCTTACT CAACATCC (TTr OL3)
-10,160
GAGCATGAGG TGTCACTACT TAACATAGGG ACGAGATGGT ACTTTGTGTC
CTCGTACTCC ACTGTGATGA ATTGTATCCC TGCTCTACCA TGGAACACAG
-10,110
TCCTGCTCTG TCAGCAGGGC ACTGTACTTG CTGATACAGG GAATGTTTGT
AGGACGAGAC AGTCGTCCCG TGACATGAAC GACTATGTCC CTTACAAACA

between Region I, site B, and the TTR oligonucleotide site 3 (Figure 5.3; 9 out of 11 bases in the relevant region). Site A in Region I did not match well with any of the C/EBP footprinted regions, the best match being a 7 out of 10 match with the TTR oligo 2 site (Figure 5.3; Costa *et al.*, 1988). This lack of a strong consensus among C/EBP binding sites (Costa *et. al.*, 1988b) is characteristic of this protein, with the good match between TTR oligo 3 and site B in the albumin enhancer being an exception (Costa *et al.*, 1988b). This well documented ability of C/EBP to bind to a wide array of divergent sequences is like the situation observed for the hap1 protein of yeast which binds to the UAS elements of the cyc1 and cyc7 genes and is responsible for their Heme inducibility (Pfeiffer *et al.*, 1987). In that case it has been shown that the same binding domain has somehow evolved to recognize two very different DNA sequence regions.

C/EBP functions in a novel capacity in the mouse albumin enhancer

In all previously characterized functional situations C/EBP protein has been demonstrated to have a positive effect on transcription (see chapter 1). Past experience in our laboratory working with the mouse transthyretin (TTR) and α1-antitrypsin genes localized many C/EBP sites on both these elements (Costa *et al.*, 1988b; Grayson *et al.*, 1988b), and removal of a site produced a decrease in transcription, although some sites had only a modest (20-50%) effect on expression. This latter fact, however, could have been due

to the fact that HepG2 cells (in which most of these studies were done) have reduced levels of C/EBP protein as compared to the liver (Freidman *et al.*, 1989).

In the mouse albumin enhancer the C/EBP binding sites alone have no demonstrable positive effect on transcription, and their addition to Region II, in the absence of Region III, produced no transcriptional boost above that previously observed from Region II alone. The C/EBP sites did (site B alone), however, play an important modulatory role, counteracting the negative effect of Region III. This is the first demonstration of C/EBP functioning in this capacity.

The function that the C/EBP protein displays in the mouse albumin enhancer resembles the role played by the α1 protein in yeast, which specifically activates the α specific genes in haploid cells (Herskowitz, 1989). It that system, it has been proven that α1 is not involved in transcriptional stimulation directly, but acts by modulating the activity of a second factor termed MCMI (also referred to in the literature as PTRF or GRM; Keleher *et al.*, 1988, Bender *et al.*, 1987; Herskowitz, 1989) which binds to an adjacent DNA sequence. It has been shown that the α1 protein increases the affinity of MCMI for its binding site (Jarvis *et al.*, 1988; Herskowitz, 1989) and additional studies have suggested that it also changes the MCMI protein conformation to increase its ability to effect transcription (Tan *et al.*, 1988). This precedent

suggests a mechanism by which C/EBP might be functioning in the present situation.

Exactly how the albumin C/EBP proteins(s) are influencing the other albumin enhancer factors is an issue which still needs to be explored. A thorough understanding of this will require a purification of the binding activities which interact with all three regulatory regions and a careful analysis of the exact nature of their protein-protein interactions. The fact, however that the mouse albumin promoter contains three C/EBP sites provides a clue concerning the mechanism by which the C/EBP site (s) are functioning. For example, the transfection constructs which lacked Region I were inactive for enhancement, although three promoter C/EBP sites (in the 300bp albumin promoter element, see chapter 3) were immediately adjacent to negative Region III. The fact that these sites could not override the negative effect might indicate that the C/EBP exerts its effect by interacting with the neighboring Region II (the NLS1 protein, chapter 6).

C/EBP interacts with both the albumin promoter and enhancer elements

The presence of C/EBP sites in both the promoter and the enhancer of the albumin gene, as well as at a site between the two (at -3.5kbp; Liu *et al.*, 1988) is quite interesting, and suggests that this protein might in some way act to facilitate communication between these disparate regulatory elements. Additionally, both the mouse TTR and α1-antitrypsin genes also have C/EBP

sites in both their promoter and enhancer elements (Figure 9.1; Costa *et al.*, 1988a,1988b; Grayson *et al.*, 1988b) further implicating C/EBP in this type of role. This theory becomes even more tenable with the discovery that C/EBP possess the leucine zipper motif and therefore is able to dimerize. Interactions between like proteins (ie. Ara C) are known to be important in several prokaryotic systems to facilitate the interaction of DNA sequences located large distances apart, since their interaction facilitates DNA looping (Huo *et al.*, 1988; Schleif *et al.*, 1988). Hence while C/EBP has no positive effect, perhaps it plays a dual modulatory role: counteracting the negative region while also assisting these far upstream sequences to function. A model for how C/EBP might work in insuring albumin tissue specificity will be presented in the discussion of chapter 7, after all three regulatory regions have been described.

Chapter 6

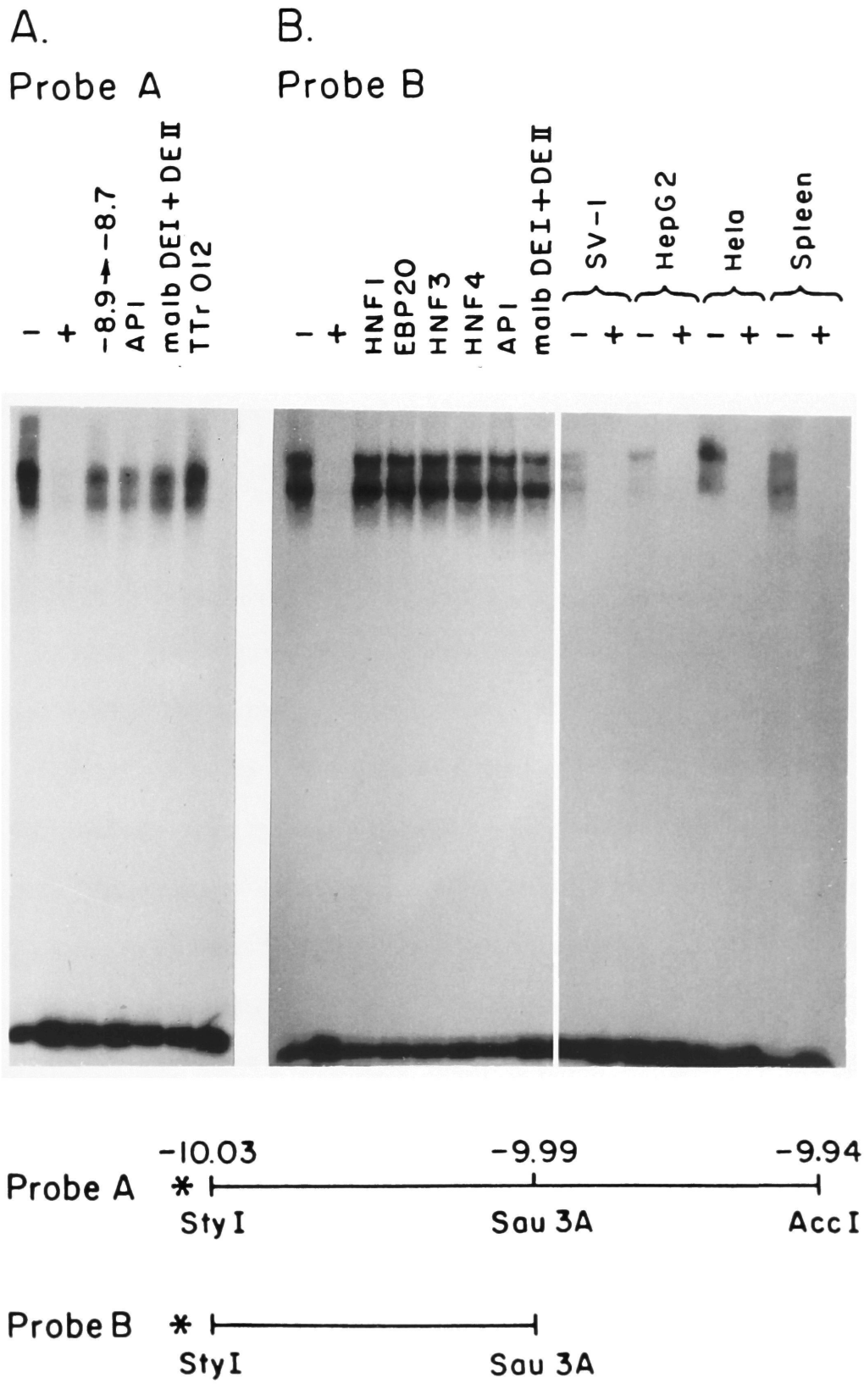
Albumin Enhancer Region II

The focus of this chapter is Region II, the positive acting, liver specific enhancer element. It was this region which alone could provide complete enhancement in CWSV1 cells. This same region, however, could not stimulate transcription of the heterologous β-globin promoter in HeLa cells.

Region II binds proteins in extracts from liver and non-liver cell types

Region II (-10.03 and 9.94kbp) was fully active for enhancer function in the CWSV1 transfection assays. Therefore, this 90 nucleotide DNA fragment was analyzed for protein interactions by performing a gel shift using crude liver and spleen extracts as well as extracts from HepG2, CWSV1 and HeLa cells. Two retarded bands that could be specifically competed by unlabeled homologous oligonucleotide were observed (Figure 6.1A). These bands were not competed, however, by a collection of binding sites representing previously identified proteins. These include sites for two widely distributed proteins, AP1 and NF1 (Jones *et al.*, 1987; Lee *et al.*, 1987), C/EBP and the -8.9 to -8.7kbp portion of Region III, as well as sites for the recently described hepatocyte enriched factors HNF 3 and HNF 4 that occur in the TTR and α1-antitrypsin regulatory regions (Costa *et al.*, 1989). The protein binding region of the 90 nucleotide region II probe was subsequently narrowed to the 40 nucleotides between –10.03 and –9.9kbp because this fragment produced an identical gel-shift result as the entire 90 nucleotide fragment (Figure 6.1B) while the corresponding 50 nucleotide fragment from -9.99 to -9.94kbp did not

Figure 6.1. The factor which interacts with Region II is present in extracts from liver and spleen tissue and CWSVl, HepG2 and HeLa cells. In panel A, a 5' labeled DNA probe (probe A) spanning from -l0.03 to -9.94Kbp (Region II) was subjected to a gel retardation assay. A shifted doublet was observed in the absence (-) but not the presence (+) of specific competitor. The additional competitors have been described in the text. A second labeled probe was prepared by truncating the fragment described in panel A at the unique Sau3A site at -9.90 kbp (probe B), and its shift pattern is shown in Panel B. This probe produces a similar gel shift pattern after incubation with extracts derived from the cell lines and tissues marked above each lane.

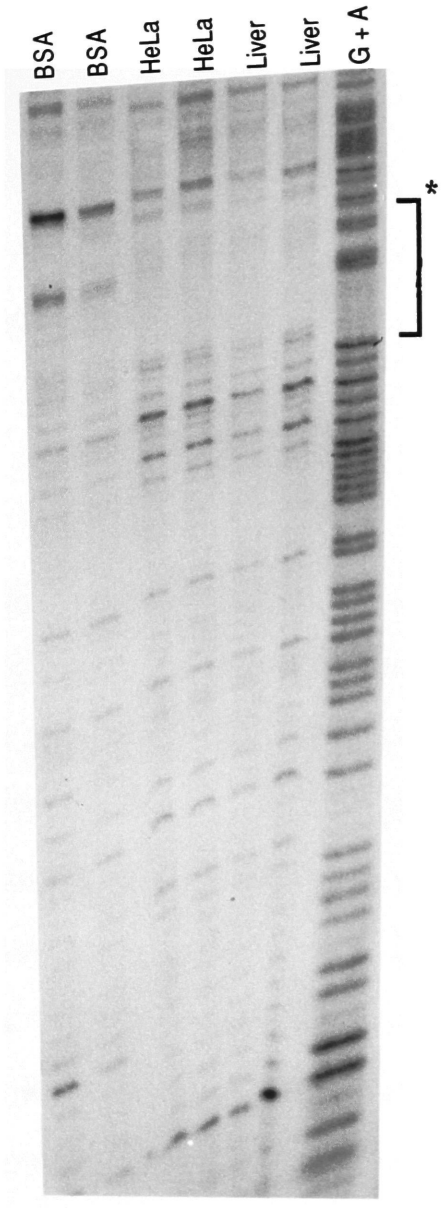

A.
Probe A
- +
-8.9→-8.7
API
malb DEI+DEII
TTr 012
B.
Probe B
- +
HNF1
EBP20
HNF3
HNF4
API
malb DEI+DEII
SV-1
HepG2
Hela
Spleen
- + - + - + - +
-10.03
-9.99
-9.94
Probe A
Sty I
Sau 3A
Acc I
Probe B
Sty I
Sau 3A

bind any protein (data not shown). The protein(s) responsible for the gel shift of the Region II probe were also found in lesser abundance in the CWSV1 and HepG2 cells, where the enhancer was shown to function. However, nuclear extracts obtained from HeLa cells and spleen tissue also produced gel shift results similar to those obtained with liver extracts (Figure 6.1B). Thus, DNA Region II that clearly seems involved in enhancement in hepatoma (or liver) cells but not in HeLa cells, can be bound by proteins from cells in which the enhancement does not occur. For this reason, this newly described protein was named NLS1 (non-liver specific factor 1).

DNase I footprint analysis of NLS1 binding

To more precisely define the sequences which interact with NLS1, a DNase I footprint experiment was performed. A DNA fragment extending from -10.03 to -9.94kbp was uniquely 5' end labeled and subjected to a footprint analysis using nuclear extract from either liver or HeLa cells. The liver and HeLa nuclear extracts were fractionated over a heparin agarose column and the maximal NLS1 binding activity was determined to be present in the 0.4M KCl peak for both. In the result shown, however, a 0.4M P11 HeLa column step fraction was used which also has a high level of NLS1 binding activity (gift of X. Fu). The DNase I footprint(Figure 6.2) employed the extracts just described or BSA as a control. Note that a nearly identical footprinted region was observed for both extracts. The footprinted region is

Figure 6.2 DNase I footprint of the Region II sequences. The functionally active DNA fragment from -10.03 to -9.94 was 5' end labeled and subjected to a DNAse I footprint assay as described in the methods section of the text. Heparin agarose purified liver fraction or P11 purified HeLa extracts were used plus a BSA control. Either 3 or 6μg of extract was used (the left or right lane respectively). The footprinted region is depicted at the bottom by the brackets



-10,030 NLS1 -9.980

CAGTTTCCATGTACATGCAGAAAGAAGTTTGGACTGATCAATACAGTCCT

GTCAAAGGTACATGTACGTCTTTCTTCAAACCTGACTAGTTATGTCAGGA

denoted by the black bar, and the protected DNA sequences are shown at the bottom of the figure depicted by a bracket (Figure 6.2, bottom). In an additional DNase I experiment (data not shown), this same footprinted region could be competed by a –10.03 to –9.9kbp albumin enhancer fragment, thereby proving the specificity of this interaction. Therefore, no differences are observed between the protected regions obtained with extract from rat liver where the NLS1 could direct transcription, or HeLa cells where NLS1 was inactive.

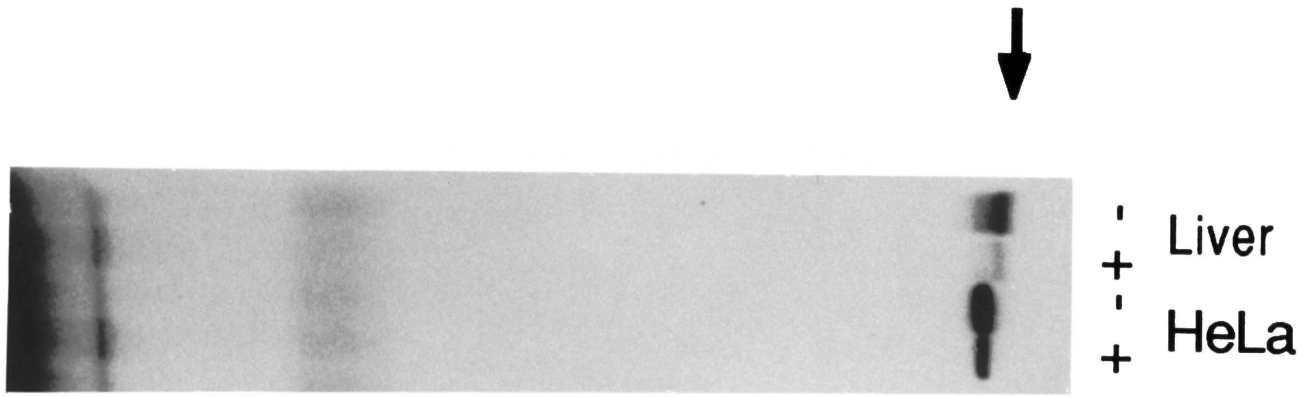
Discussion

In contrast, to region I, Region II on its own did have enhancing ability, but only in hepatocytes and not in HeLa cells. It was therefore somewhat surprising to find the proteins that bound to Region II were widely distributed in several cell types, including HeLa and spleen cells. It is possible that the protein(s) binding to Region II might either be modified correctly only in liver cells or alternatively that a second protein, present only in liver cells, interacts with the NLS1 to produce gene enhancement. Additionally, NLS1 binding activity could represent a multigene family, with only the liver version capable of activating transcription. Several previously identified transcription factors have been shown to constitute multigene families (ie. ATF, AP1, CAAT; Dorn *et al.*, 1987; Hai *et al.*, 1988) each member of which might activate transcription in its own unique manner.

Is NLS1 modified specifically in the liver? A final answer to this question will require the purification and cloning of this factor from liver and at least on other cell source. The data above, however, suggests that no differences exist in the DNase I protection pattern generated from two different cell sources, and the shift mobility of complexes from several cell types was similar.

Recently, however, a difference in the gel-shift has been detected using an oligonucleotide constructed according to the footprint data (Figure 6.2) and the shifted complexes were analyzed by electrophoresis on higher percentage acrylamide gels (10% as opposed to 8%). Under these newer conditions (shown in Figure 6.3), the HeLa protein produced a shifted complex with a slightly increased electrophoretic mobility. This difference could be the result of a post transcriptional modification event (ie. cleavage or phosphorylation) that might change the size or charge of the protein, or perhaps indicates that the two proteins are products of different genes. Of note, the Hela cells are a human cell line while the liver extract was prepared from the rat, and hence a species difference might produce this effect. In any event, this shift mobility difference is similar to the situation observed with the ApoCIII gene where the positive acting factor AF-1 is present in two different gel shift forms in expressing vs. non-expressing tissues (Leff *et al.*, 1989), or for the factor HNF1, which exists as a faster migrating variant (vHNF1) in dedifferentiated hepatoma cells (Cereghini *et al.*, 1988). The

Figure 6.3. Liver and Hela cell nuclear extracts produce NLS1 gel-shifts of different mobility. An oligonucleotide representing the footprinted sequences shown in Figure 6.2 (bottom) was end labeled an subjected to a gel shift using crude liver or HeLa cell nuclear extract, in the presence (+) or absence (-) or a 100 fold excess of unlabeled homologous oligonucleotide competitor, and electrophoresed on a 10% acrylamide gel. Under these condition, a difference in gel shift mobility is observed using extracts from these two different cell-types


-
+
Liver
-
+
HeLa

analysis of the albumin enhancer NLS1 site described above is still preliminary and as the protein(s) are purified these different alternatives can be explored in greater detail.

Chapter 7

Albumin Enhancer Region III

From the functional studies described in chapter 4, Region III was demonstrated to prevent the enhancing action of Region II, an effect that was overcome by the presence of Region I. To characterize the nuclear factors that might interact with the negative acting DNA sequences of Region III (shown from the functional studies presented in Chapter 4 to lie between –8.9 and –8.43kbp), and to determine the tissue distribution of any such factors, gel retardation assays and DNase I footprint analyses of this region were performed.

Multiple Factors Interact with Region III

From the CWSV1 transfection analysis, the region between –8.9 to –8.43kbp was proven to have a negative effect. The sequences extending from –8.9 to –8.43kbp were therefore divided into three segments (probes 1-3, Figure 7.1), and subjected to separate gel shift analyses (it had previously been determined by exonuclease III studies that the region between –8.9 and –8.84kbp did not contain any sites of DNA-protein interaction, and hence, they were not included for further study). By competition with unlabeled oligonucleotides, two binding sites for known proteins were localized: NFl which binds to probe 1 (Figure 7.2) and HNF3 which binds to probe 2 (Figure 7.4 and 7.5). The third probe bound several proteins for which no known oligonucleotides could compete (probe 3, Figure 7.6).

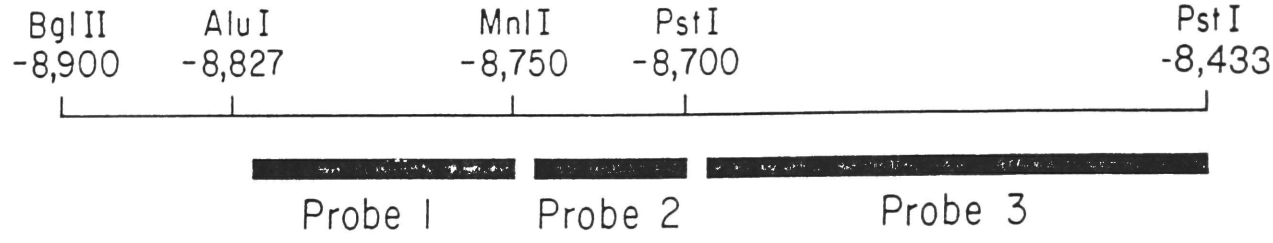


Figure 7.1. Restriction map of albumin enhancer Region III. Using the restriction endonuclease sites shown, this DNA fragment was divided into three different end-labeled DNA probes (denoted 1, 2 and 3) which were used to assay for DNA-protein interaction. No probe was prepared for the –8,900 to –8,827bp segment for it had previously been shown not to interact with any protein.

The mouse albumin enhancer contains a NF1 binding site

A 77 base pair fragment (–8.827 to –8.750kbp) was 3' end labeled and subjected to a gel retardation shift assay. A protein binding site was detected, which appears to be an NF1 site based on competition by an oligonucleotide containing mouse albumin upstream elements DEI and DEII, but not by an oligonucleotide containing DEI alone (Figure 7.2A). Since DEI represents an C/EBP binding site and DEII a site for NF1 binding, this competition must therefore be resulting from the NF1 site (Babiss *et al.*, 1986; Cereghini *et al.*, 1987; Costa *et al.*, 1988.). Furthermore, a gel-shift of this region using nuclear extracts prepared from various tissues and cell types showed this activity to be generally distributed among the different extract sources (Figure 7.2B), as has been previously shown for NF1, and a reasonable similarity is evident with the known NF1 consensus sequence (Figure 7.3; Jones *et al.* , 1987). The mouse albumin promoter also contains a binding site for NF1 (see chapter 3), which has been shown to bind both the ubiquitous NF1 protein and a liver specific form of this factor which results from alternative RNA splicing (Paonessa *et al.*, 1988). The albumin enhancer NF1 binding site presumably binds both of these different forms of the factor, which are not readily distinguishable by gel shift analysis (Lichtsteiner *et al.*, 1987).

Region III contains and HNF3 binding site

The DNA fragment extending from -8.75 to -8.7kbp (probe 2) produced

Figure 7.2. Identification of an NF1 binding site in the mouse albumin enhancer by gel- shift analysis. Probe 1 (–8,827 to –8,750bp; see Figure 7.1) was end-labeled and 1ng was incubated with 8μg of liver nuclear extract (Panel A) or 8μg of nuclear extract from HeLa cells, brain or spleen tissue (Panel B), in the presence (+; self) or absence (-) of self competitor. Specific competitors used include the –8.75 to –8.7kbp fragment of the albumin enhancer, mouse albumin DEI + DEII or oligonucleotide 2 from the TTR enhancer (TTR OL2, a C/EBP binding site; Costa *et al.*, 1988a). The mouse albumin DEI is a C/EBP binding site while DEII is a site of NF1 binding (Cereghini *et al.*, 1987).

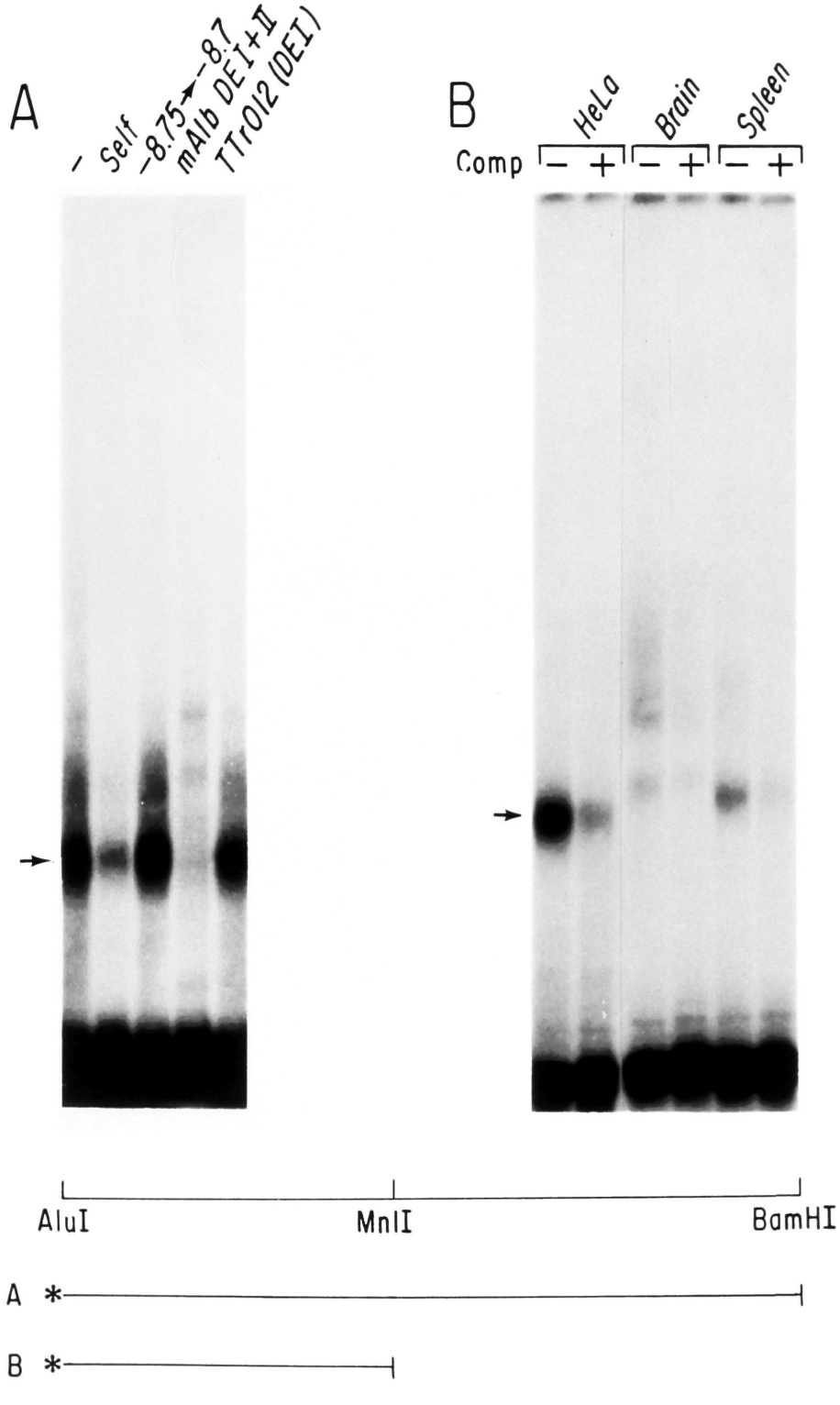
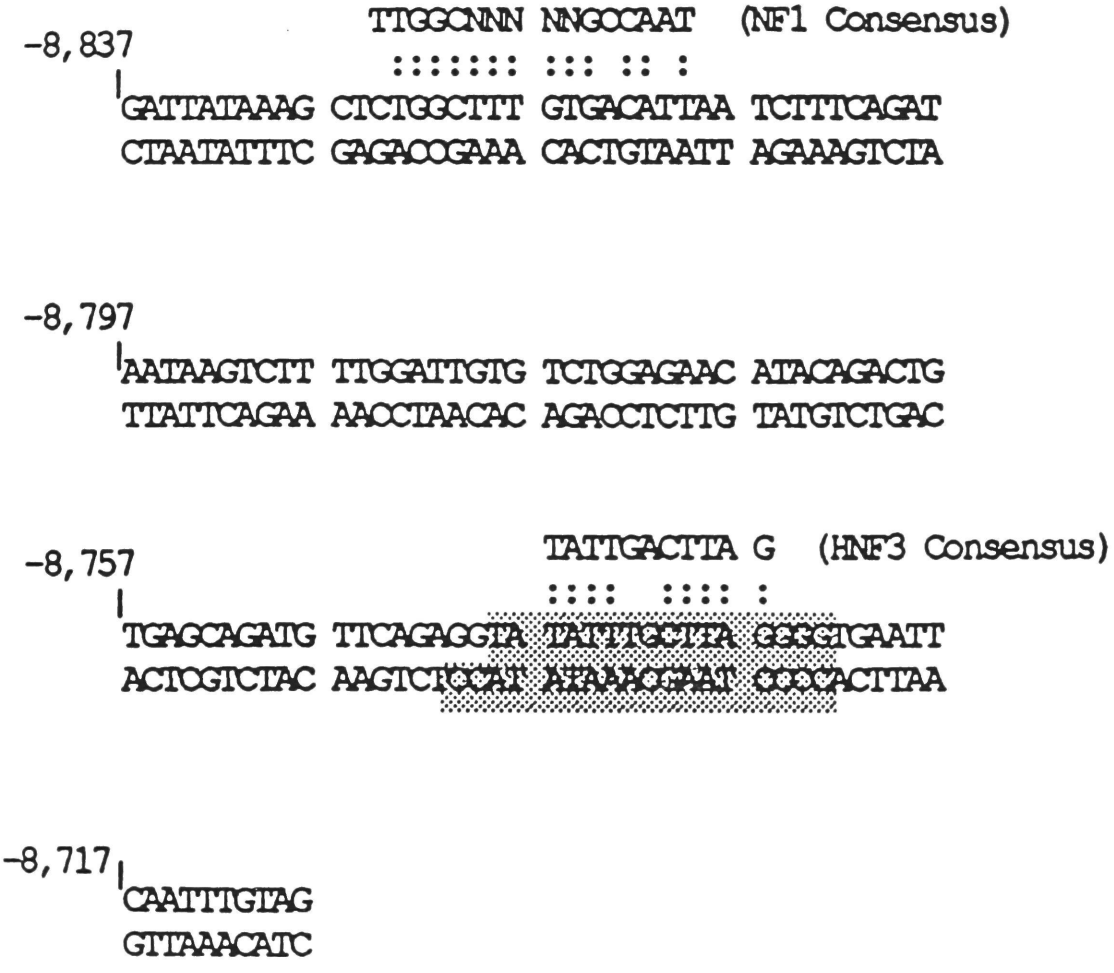

A
- Self
-8.75→-8.7
mAlb DE1+II
TTrO12 (DE1)
B
HeLa
Brain
Spleen
Comp
- +
- +
- +
AluI
MnlI
BamHI
A *
B *

Figure 7.3. DNA sequence of the -8.837 to -8,700bp portion of Region III which contains the NF1 and HNF3 binding sites. The DNA sequences footprinted by purified HNF3 are shaded and above each site is shown the homology with known consensus sequences (Jones *et al.*, 1987, Costa *et al.*, 1989).


TTGGCNNN NNGCCAAT (NF1 Consensus)
-8,837
GATTATAAAG CTCTGGCTTT GTGACATTAA TCTTTCAGAT
CTAATATTTC GAGACCGAAA CACTGTAATT AGAAAGTCTA
-8,797
AATAAGTCTT TTGGATTGTG TCTGGAGAAC ATACAGACTG
TTATTCAGAA AACCTAACAC AGACCTCTTG TATGTCTGAC
TATTGACTTA G (HNF3 Consensus)
-8,757
TGAGCAGATG TTCAGAGGTA TATTGCTTTA GGGGTGAATT
ACTCGTCTAC AAGTCTCCAT ATAATCGAAT CCCCACTTAA
-8,717
CAATTTGTAG
GTTAAACATC

a gel-shift band with protein from hepatocytes that was competed by itself (+) and by an oligonucleotide containing the binding site for the liver specific protein HNF3 which interacts with the TTR promoter proximal sequences between -85 to -111bp (Figure 74A; Costa *et al.*, 1989). A series of additional oligonucleotides, however, did not compete with the labeled probe. This strongly suggested the existence of a HNF3 binding site in the mouse albumin enhancer. Additionally, this binding activity is liver specific as shown in Figure 7.4B, for the shifted complex is not present in nuclear extracts prepared from kidney, spleen or brain tissue and HeLa cells have a single band of lesser mobility which is always present in cultured cells but does not represent HNF3 (Costa *et al.*, 1989).

This HNF3 site was confirmed by using oligonucleotide affinity purified HNF3 protein (gift of R. Costa; Costa *et al.*, 1989) to footprint this region. (Figure 7.5). Two different amounts of purified HNF3 protein were used (lanes 1 and 2: 5 or 10μg respectively) plus a BSA control. A G+A DNA sequence ladder is included for orientation. The protected nucleotides on each DNA strand are depicted by the black bar in Figure 7.5 and interestingly there exists a DNase I hypersensitive site in the middle of the footprinted region (marked by a "*"). The footprinted DNA sequences are highlighted in Figure 7.3 where it is shown that an 8/10 base homology exists with the known HNF3 element. The mismatched bases are in a region which Costa *et al.* (1989) has shown dictates the affinity of the HNF3 protein for its DNA

Figure 7.4. Identification of a HNF3 site in the mouse albumin enhancer by gel shift analysis. One ng of Probe 2 (-8,750 to -8,700bp; see Figure 7.1) was end-labeled and subjected to a gel shift analysis as previously described using either 8ug of rat liver nuclear extract (Panel A) or 8ug of nuclear extracts prepared from kidney, spleen, brain tissue and HeLa cells (Panel B), in the presence (+, Alu) or absence (-) of self competitor. Specific competitors used include oligonucleotides containing binding sites for C/EBP (TTr Ol2), HNF3 (TTr –85 to –111), and DEI and DEII as described in Figure 7.2. A 100bp pGem fragment is also included as a competitor. M1 and M2 identify the two shifted complexes which are characteristically observed for HNF3

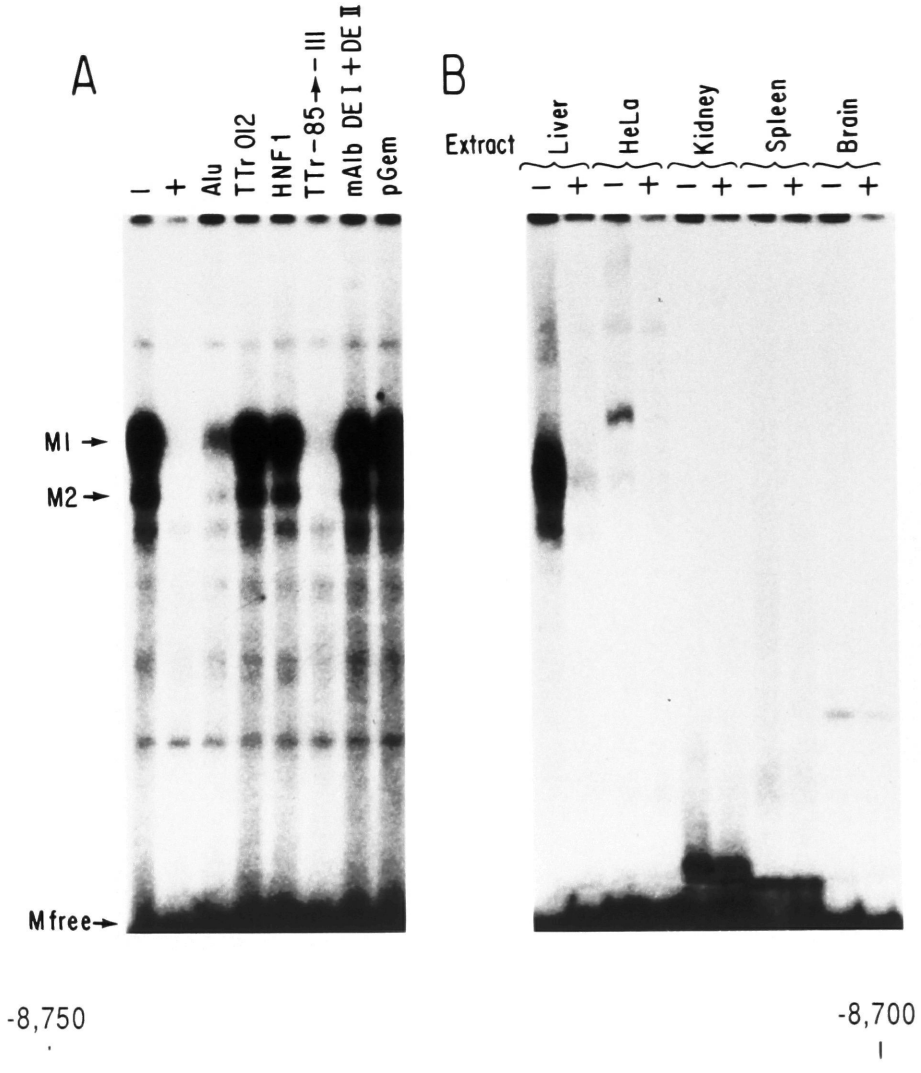
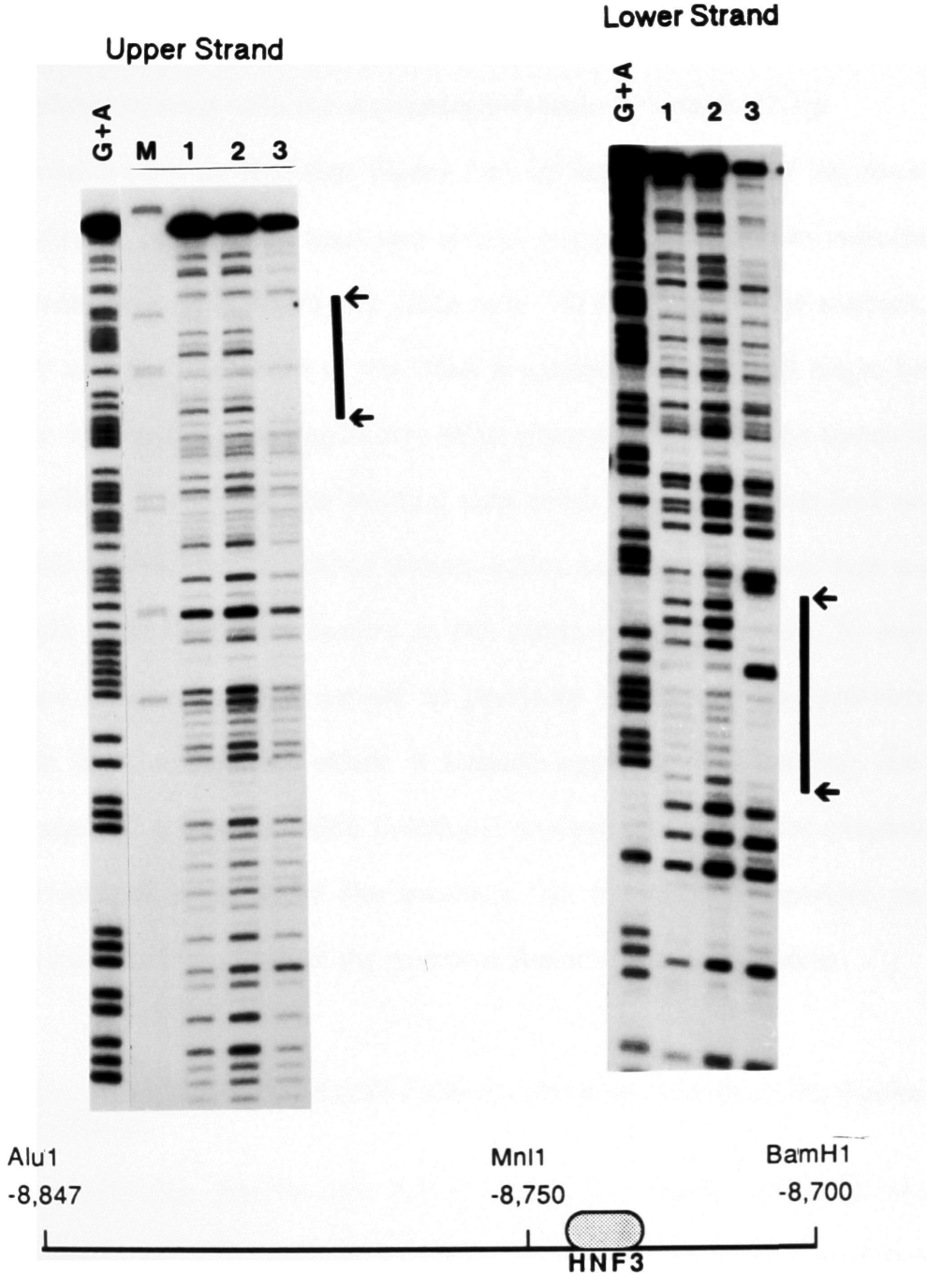

A
B
-
+
Alu
TTr OI2
HNF1
TTr-85→-III
mAlb DE I + DE II
pGem
Extract
Liver
HeLa
Kidney
Spleen
Brain
M1
M2
Mfree
-8,750
-8,700

Figure 7.5. DNase I footprint analysis of the HNF3 binding site. A 5' end labeled probe from -8,827 to -8,700bp was incubated with affinity purified HNF3 protein (approx. 1000 fold; Gift of R. Costa) prior to digestion with the enzyme DNase I. The products were then analyzed on this 8% denaturing acrylamide gel. Lane 1 was incubated with BSA as a control and lanes 2 and 3 with 5 or 10µg of the HNF3 protein respectively. A (G+A) sequencing ladder and markers (M) are included for orientation purposes. The footprinted regions are denoted by the black bars. Note the appearance of several hypersensitive sites including the most prominent in the center of the footprinted region marked with a "*".


Upper Strand
G+A
M
1
2
3
Lower Strand
G+A
1
2
3
Alu1
-8,847
MnI1
-8,750
BamH1
-8,700
HNF3

sequence. This albumin sequence was subsequently demonstrated to be a weak HNF3 binding site by gel shift analysis (data not shown).

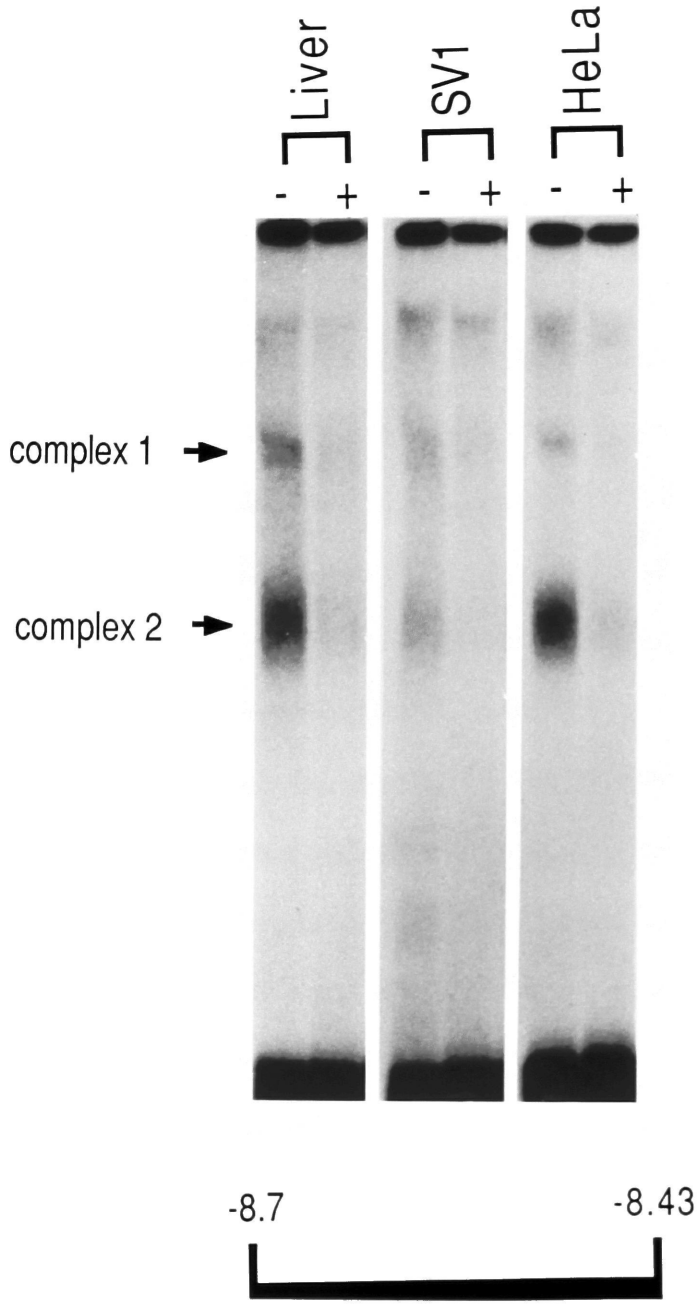
Several Proteins interact with the sequences between -8.7 and -8.43Kbp

Probe 3 (–8.7 to –8.43kbp; Figure 7.6), by far the largest of the three probes analyzed, produced at least two shifted complexes present in extracts prepared from rat liver, CWSV1 and HeLa cells. At this point in the analysis, however, it was unclear which of the DNA sequences in Region III might be responsible for the negative, regulatory effect observed. Clearly the removal of the described HNF3 and NFl binding sites (with Region I intact) had no affect on the overall level of mRNA accumulation, suggesting at least that no positive role exists for these factors in the transient assay system. In any event, none of these results served to precisely single out the proteins responsible for the negative effect. It became apparent that settling this question required a more detailed deletional analysis of this region coupled with a functional analysis of the mutants. At this point, therefore an additional functional analysis of the negative element was undertaken.

The negative acting sequences reside between -8.7 and-8.43kbp of the mouse albumin enhancer

To specifically localize the DNA sequences which conferred the negative effect on the mouse albumin enhancer, transient plasmid transfection assays were again performed using rat CWSV1 cells (see chapter

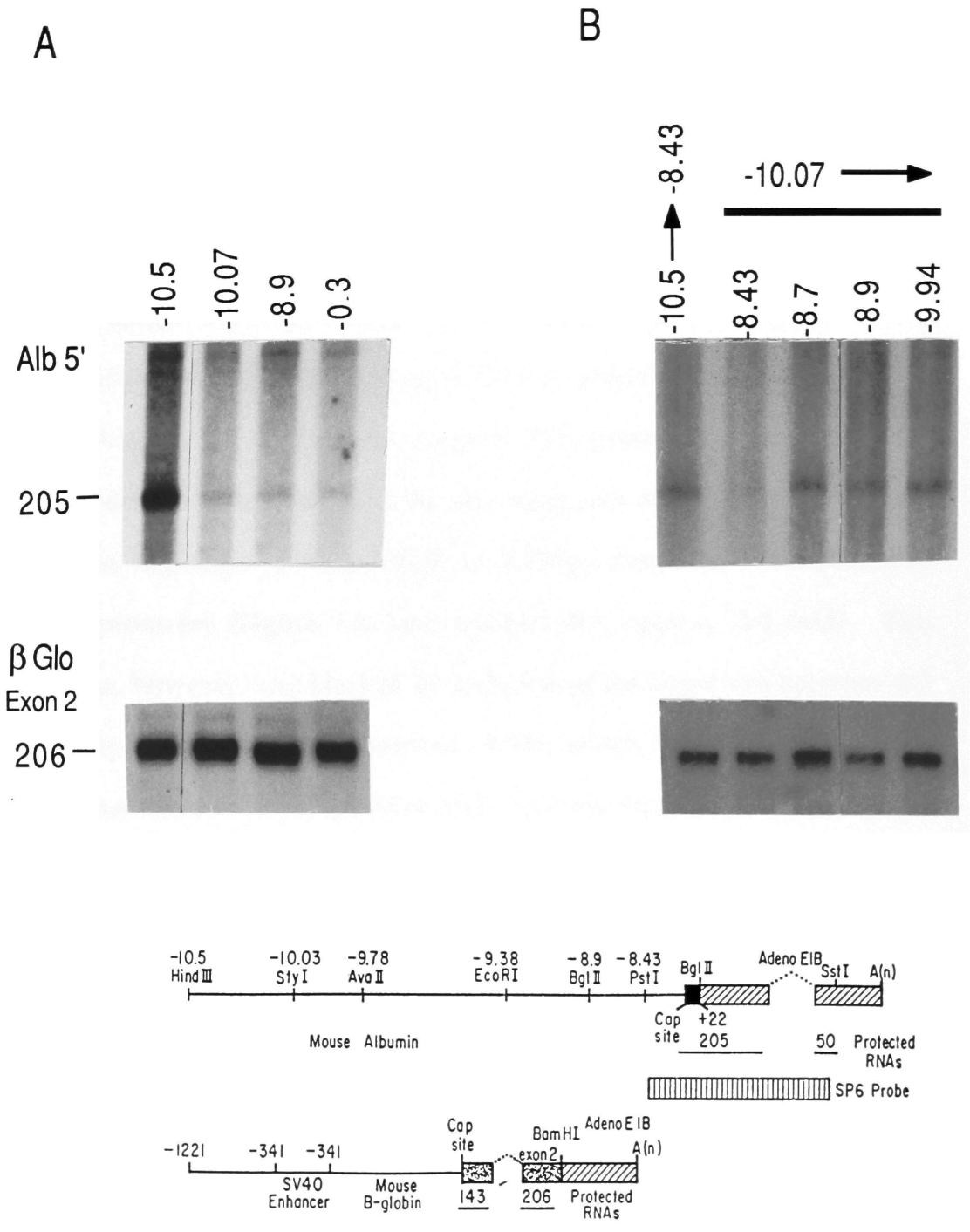
Figure 7.6 Gel-shift analysis of protein binding to Probe 3 of albumin enhancer Region III. Probe 3 (-8700 to -8,433bp) the largest of the probes was end-labeled and subjected to gel-shift analysis with 8ug of nuclear extract from liver tissue, CWSV1 or HeLa cells in the presence (+) or absence (-) of self competitor. Specific shifted DNA-protein complexes are shown by arrows.


Liver
SV1
HeLa
-
+
-
+
-
+
complex 1
complex 2
-8.7
-8.43

4, Figure 4.1). Agreeing with the results reported in chapter 4, a five-fold stimulation in albumin promoter function was found when the entire albumin enhancer was included (Fig. 7.7A; -10.5 vs. -0.3). Furthermore, when sequences between –10.5 and –10.07kbp (which includes Region I) were removed, a complete loss of enhancer function was observed (Figure 7.7A, –10.07), owing to the absence of the enhancer modulating sequences. Also shown in figure 7.7A, however, is a DNA fragment which contains the negative element exclusively (sequences from –8.9 to –8.43kbp) placed 5' to the -300 albumin promoter element. Interestingly, the Region III sequences did not suppress the basal level of albumin promoter function (compare lanes marked –8.9 and –0.3, figure 7.7A). Therefore, the negative element would seem to have no direct effect on the albumin promoter but rather appears to exert its effect solely by negating the effects of the enhancer.

To define the functional boundaries of the negative element, a plasmid containing mouse albumin regulatory domain II and III (–10.07 to –8.43kbp) was used to make deletions from the 3' boundary using three different restriction sites: Pst1; (–8.7kbp), Bgl II (–8.9kbp); or Acc1 (–9.94kbp). Plasmids with these deletions (maintaining a set 5' end at –10.07kbp) were assayed in CWSV1 cells (Figure 7.7B). The plasmid construct with a 3' coordinate of –8.43kbp (containing the entire –8.9 to –8.43 region), did not show any enhancer function, due to the presence of the negative element. In contrast, when 270bp were deleted to the site at –8.7kbp, full enhancer

Figure 7.7 Additional functional analysis of the mouse albumin enhancer element in CWSV1 cells. In panel A the parent construct (bottom) and deletions (marked by their 5' boundary in kbp) were co-transfected into CWSV1 cells with a control plasmid (bottom, a SV40 enhancer containing β-globin promoter) to standardize for variations in transfection efficiency (see Materials and Methods, chapter 2). Assays for expression were carried out by protection of labeled SP6 probes. The lane marked –0.3 indicates the enhancerless albumin promoter plasmid containing 300bp of promoter sequence. In panel B, the parent construct is assayed (-10.5 to -8.43kbp), along with an additional construct as shown which maintains a 5' enhancer boundary of -10.07kbp while having a variable 3' end position as indicated above each lane.


A
B
-10.5
-10.07
-8.9
-0.3
Alb 5'
205
β Glo
Exon 2
206
-8.43
-10.07
-10.5
-8.43
-8.7
-8.9
-9.94
-10.5
Hind III
-10.03
Sty I
-9.78
Ava II
-9.38
EcoRI
-8.9
Bgl II
-8.43
Pst I
Bgl II
Adeno E1B
Sst I
A(n)
Cap site
+22
205
50
Protected RNAs
Mouse Albumin
SP6 Probe
-1221
-341
-341
Cap site
BamHI
exon 2
Adeno E1B
A(n)
SV40 Enhancer
Mouse B-globin
143
206
Protected RNAs

function was restored, locating the functional, negative sequences of Region III between -8.7 and -8.43kbp. As observed previously, further 3' deletion of DNA sequences to -9.94kbp resulted in identical enhancement, indicating that no additional negative regulatory elements exist.

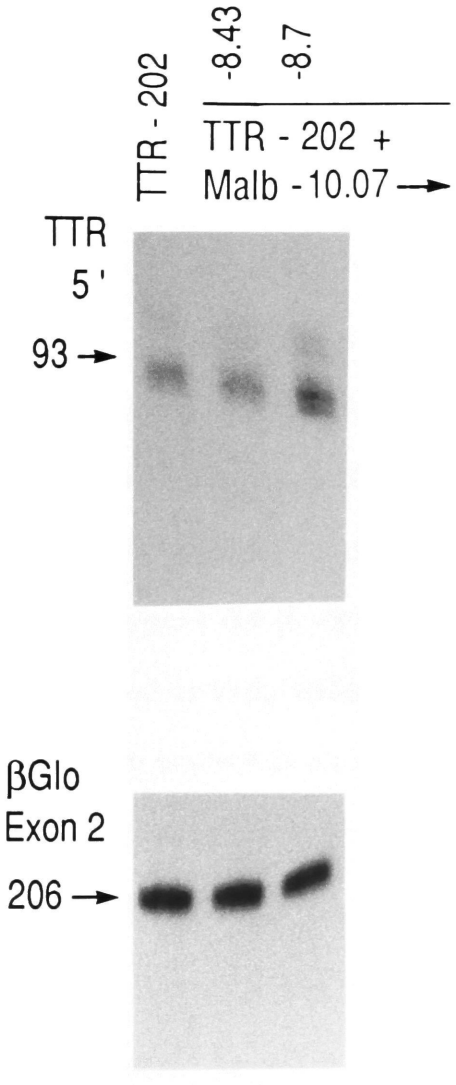
The negative element functions in HepG2 cells

An added experiment (Figure 7.8) showed that the negative acting region of the albumin enhancer (Region III) also produced an effect in HepG2 cells, in this case upon the heterologous TTR promoter. When the TTR promoter alone was coupled with the activating part of the albumin enhancer without the negating sequences (-10.07 to -8.7kbp), there was a stimulation of the TTR promoter (Figure 7.8; lane marked -8.7; approx. 3-5 fold). This stimulation, however, was blocked by inclusion of the sequences between -8.7 to -8.43kbp (Figure 7.8, lane marked -8.43), which includes the negative region. Therefore, both the positive and negative, regulatory regions of the albumin enhancer function in HepG2 cells (see Chapter 4). Subsequent experiments aimed at delineating the exact location and function of the albumin enhancer negative element have therefore used the HepG2 cell line.

The albumin negative element can act to down-regulate the TTR enhancer

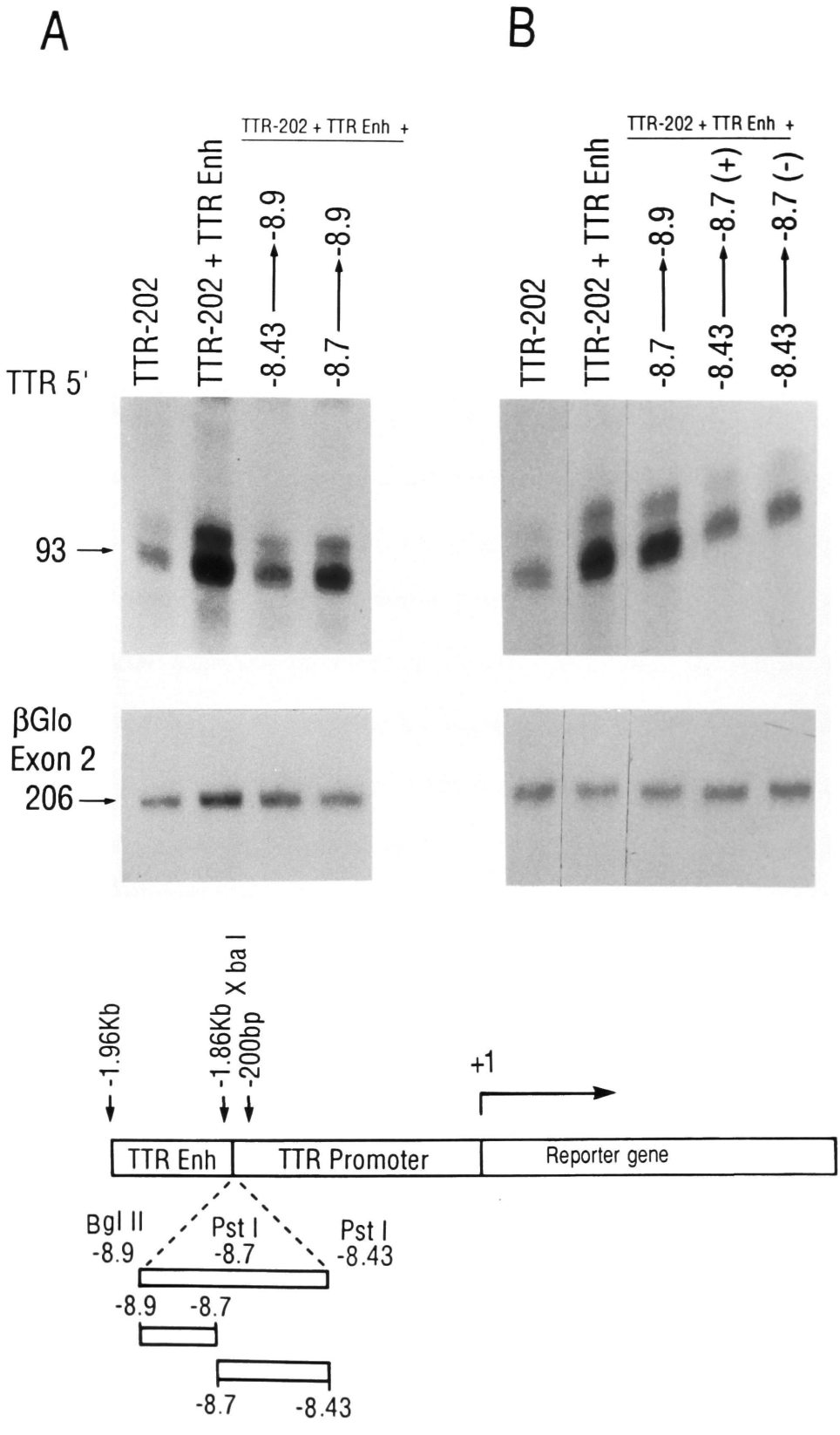
I next wanted to determine if the negative element could act upon a heterologous enhancer; in this case the TTR enhancer. The effect of the

Figure 7.8 The mouse albumin enhancer negative region can function in human hepatoma cells (HepG2 cells) upon the heterologous TTR promoter. Albumin enhancer sequences as indicated were placed 5' to the TTR promoter element at position -202 relative to the start site of RNA transcription. These constructs were then transfected into human HepG2 cells, along with the SV40 enhancer containing β-globin plasmid as an internal control. The RNase T_2 assay used to score the TTR expression protects a 93 nucleotide band as shown and is described in the text.


TTR - 202
-8.43
-8.7
TTR - 202 +
Malb -10.07
TTR
5'
93
βGlo
Exon 2
206

negative region in abolishing the action of the TTR enhancer on its own promoter was demonstrated by inserting various parts of the albumin enhancer region between the TTR enhancer and promoter as shown (Figure 7.9, bottom). The assay of the transthyretin (TTR) promoter depends on a nominal 93 base pair RNase T_2 protected band that is complementary to the TTR first exon (Fig. 7.9, bottom), although a band slightly larger is also observed (nominally 95 and 96 bases) which is thought to represent either an upstream start site or an RNase T_2 artifact (see chapter 4). In any event, these bands are usually proportional. The approximately 5 fold stimulation of the TTR enhancer on its own promoter (Fig. 7.9, left most two lanes) was blocked by the presence of the albumin negative region (Figure 7.9, –8.43 to –8.9). This negative effect was mapped more precisely by dividing the negative element into two roughly equal sized pieces (–8.9 to –8.7kbp) and (–8.7 to –8.43kbp). The negative effect was contained entirely within to –8.7 to –8.43kbp region which could exert its action when present in either orientation (indicated by + or – in Figure 7.9). Similar experiments were performed by placing the same negative region fragments between the SV40 enhancer-β-globin promoter construct described previously, and assaying for expression in both HepG2 and HeLa cells (data not shown). In each case, no negative effect was observed; either the SV40 enhancer is too potent to achieve suppression or there are a limited group of enhancers that can be suppressed by this negative element.

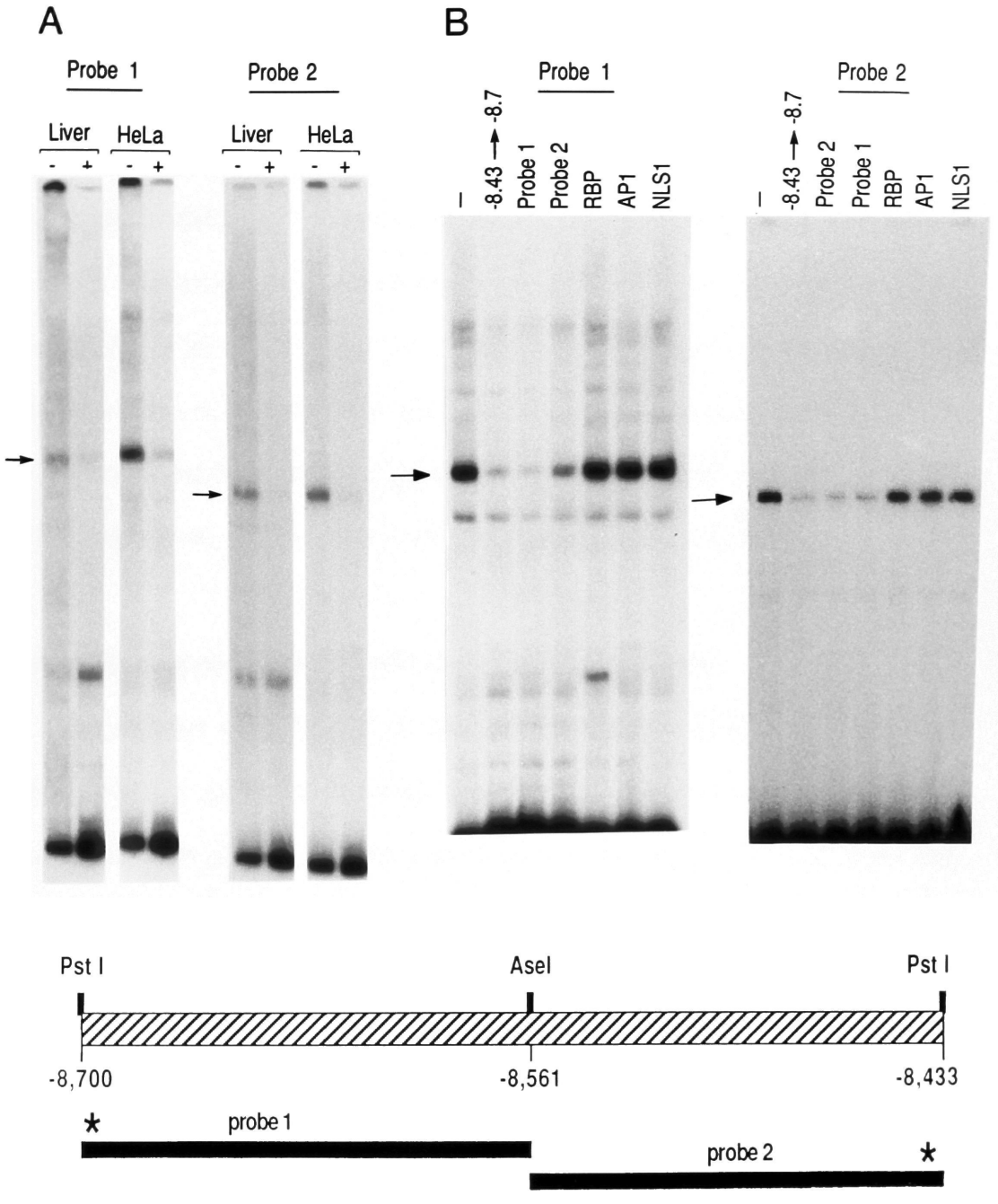
Figure 7.9. Mouse albumin enhancer Region III can negate the function of the TTR enhancer. Constructs were prepared as described in the methods section and diagrammed at bottom, which placed portions of the mouse albumin enhancer negative region (as marked above each lane) between the TTR enhancer and promoter at a unique XbaI site as shown. These constructs were transfected into human hepatoma cells (50μg) along with the SV40/β-globin plasmid (20μg) as an internal control. The TTR -202 promoter containing plasmid and the TTR promoter/TTR minimal enhancer construct are included as controls in both panels A and B, which represent separate experiments.


A
B
TTR-202 + TTR Enh +
TTR-202
TTR-202 + TTR Enh
-8.43 ⟶ -8.9
-8.7 ⟶ -8.9
-8.7 ⟶ -8.9
-8.43 ⟶ -8.7 (+)
-8.43 ⟶ -8.7 (-)
TTR 5'
93
βGlo
Exon 2
206
-1.96Kb
-1.86Kb
X ba I
-200bp
+1
TTR Enh
TTR Promoter
Reporter gene
Bgl II
-8.9
Pst I
-8.7
Pst I
-8.43
-8.9 -8.7
-8.7 -8.43

Two binding sites for the same protein reside between -8.7 and -8.43kbp

Experiments were designed that would identify protein binding sites within the 273 base pair portion of the albumin enhancer containing the negative element. As shown earlier in this chapter, this region gives a complex gel shift pattern suggesting that several proteins might be interacting with these sequences (Figure 7.5). Therefore, to simplify the analysis, this DNA fragment was bisected at a unique AseI restriction site (-8,651bp). End-labeled DNA fragments from -8,700 to -8,652bp and -8,651to -8,430bp were prepared and mixed with nuclear extracts from either rat liver or HeLa cells. Assays for DNA binding were carried out by gel retardation shifts using these labeled DNA fragments. Both fragments bound protein from both extracts and one particular complex (marked in Figure 7.10A with an arrow) was specifically prevented from forming by including an excess of unlabeled homologous oligonucleotide. Furthermore, the two parts of the negative region (–8,700 to –8,562 and –8,561 to –8,430, probes 1 and 2, respectively) showed cross competition with each other indicating a binding site in each half for the same (or a similar) DNA binding protein (Figure 7.10B). Early efforts at purification of this binding activity are consistent with this conclusion, for upon fractionation with heparin agarose the peak binding activity for both probes is found in the same fraction (data not shown). Since the specific activity of the two probes was the same and probe 1 gave a stronger gel shift band and was better able to compete for itself than an equal

Figure 7.10. Definition of the ANF (albumin negative factor) sites within the mouse albumin negative region -8.7 to -8.43kbp by gel-shift assay. Two roughly equal sized probes were prepared by restriction endonuclease digestion at the unique AseI site as shown. These two end labeled probes were then used in separate gel shift assays (Probe 1 and Probe 2, respectively). In panel A the probes were 3' end labeled and incubated with liver or Hela cell nuclear extract in the absence (-) or presence (+) of homologous unlabeled DNA (50 fold molar excess) and the products analyzed by gel retardation on an 5.5% native acrylamide gel. In Panel B the same two probes are used (as marked) except now other competing fragments and oligonucleotides (also in 50 fold excess) are used. These include the entire -8.7 to -8.43kbp fragment (represented by (+), unlabeled probes 1 and 2 as well a DNA fragment from the retinol binding protein (RBP) gene promoter containing the functional negative acting region (gift of R. Cortese). Oligonucleotides representing binding sites for the factors NLS1 and AP1 are also included as competitors.

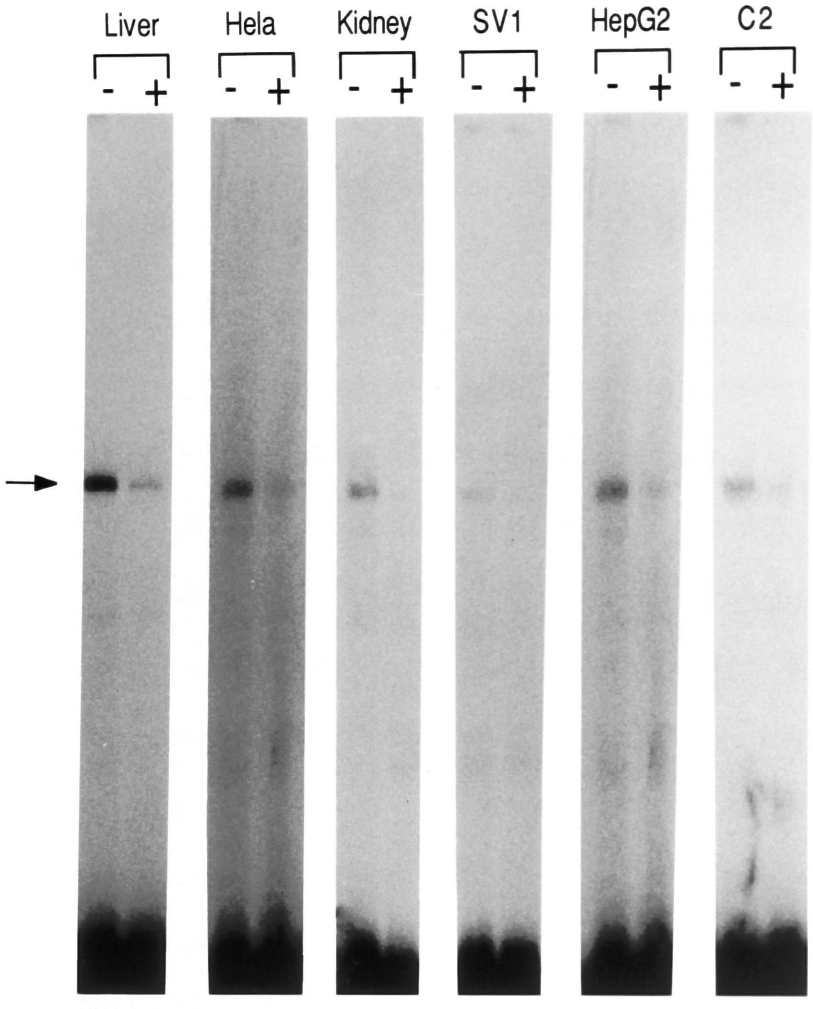

A
Probe 1
Probe 2
Liver
HeLa
Liver
HeLa
- + - + - + - +
B
Probe 1
Probe 2
– -8.43 → -8.7 Probe 1 Probe 2 RBP AP1 NLS1
– -8.43 → -8.7 Probe 2 Probe 1 RBP AP1 NLS1
Pst I
AseI
Pst I
-8,700
-8,561
-8,433
probe 1
probe 2

molar amount of probe 2, it appears that probe 1 binds this protein with a higher affinity than probe 2. The specificity of this complex was further tested by competition during complex formation with excess molar amounts of several other oligonucleotides. No other oligonucleotide tested reduced complex formation (Figure 7.10B). These unsuccessful competitors included the previously described negative elements in the retinol binding protein gene (RBP; Colantuoni *et al.*, 1987) and a group of other positive acting factors (shown in Fig. 7.10 are sites for AP1 and NLS1, the protein that binds to the activating region of the albumin enhancer; not shown but also with no effect were oligonucleotides representing SP1, HNF1, 3, 4, C/EBP and Oct1 binding sites (Courtois *et al.*, 1988; Costa *et al.*, 1988; 1989; Jones *et al.*, 1988).

The albumin negative factor (ANF) is present ubiquitously

The cellular distribution of this newly identified protein which was termed ANF for albumin enhancer negative factor was determined next. The same amount of total protein from nuclear extracts from six different cell sources were used in a gel shift reaction with negative region probe 2 (-8,651 to -8,430), shown in Figure 7.11. Each extract used was first subjected to a gel shift reaction using a known AP1 binding site to insure its integrity (data not shown, chapter 2), and each was normalized tin this way. All the cell lines or tissues contained similar amounts of the albumin negative factor except for CWSV1 cells where the level was reduced by 50%. It was in these CWSV1

Figure 7.11. The ANF binding activity is present ubiquitously. Probe 2 (see Figure7.10) was end labeled and subjected to a gel shift assay with equal protein amounts (8μg) of nuclear extracts prepared from the cell-types identified above each lane. The arrow depicts the specific ANF complex.


Liver
Hela
Kidney
SV1
HepG2
C2
- +
- +
- +
- +
- +
- +

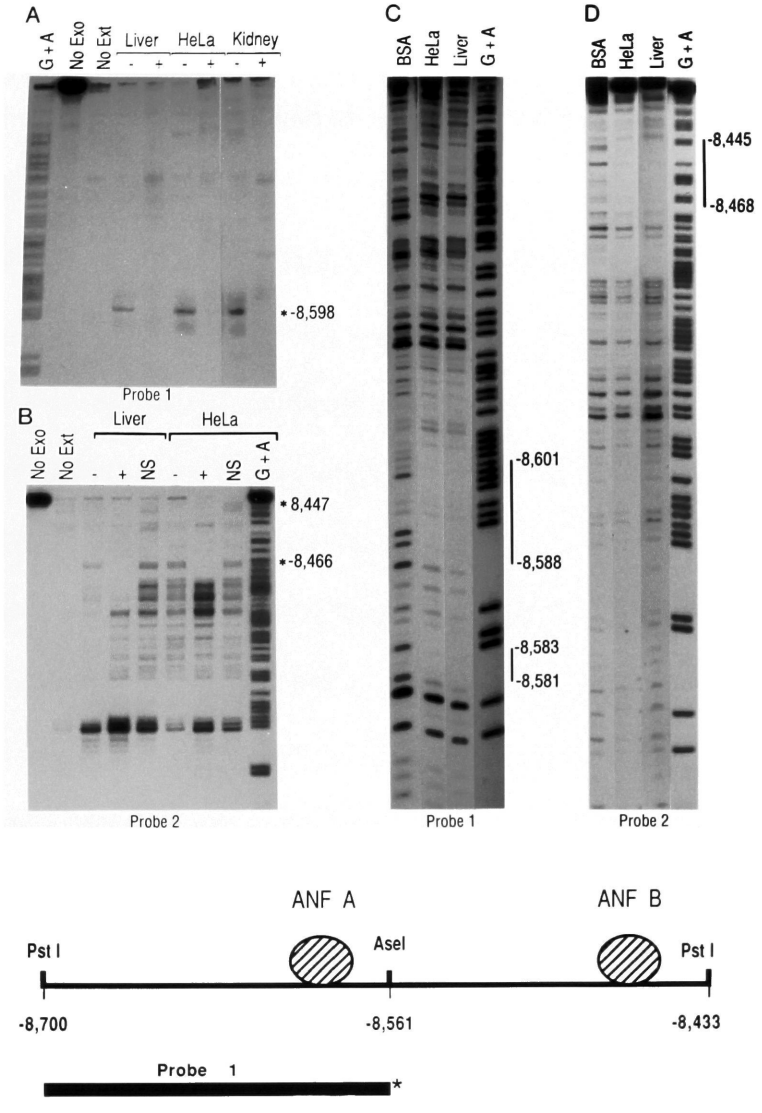
cells, however that the function of the negative element was first identified, suggesting that this decrease might not be functionally significant. These results illustrate that the factor recognizing the negative region of the albumin enhancer is widely distributed, and is present in cell and tissue types which do (liver, SV1 and HepG2) or do not (HeLa, Kidney and C2) transcribe albumin.

The ANF binding site

To define the binding sites for the ANF proteins in each region of the negative element two types of experiments were carried out, both of which used end-labeled DNA probes. Each of the fragments (Probe 1 or Probe 2, Figure 7.12) were uniquely 5' end labeled (as shown in Figure 7.12, bottom) and subjected to their separate analyses. Note that only the results for one DNA strand (the bottom strand for probe 1 and top strand for probe 2), are shown in Figure 7.12, yet the data representing the results from both strands are summarized in figure 7.13.

First, exonuclease III digestion of the labeled probe in the presence of nuclear proteins showed blocks to digestion that were specifically competed by the presence of excess unlabeled probe (Figure 7.12A and B). For each exoIII analysis, 1ng of labeled probe was incubated with 30μg of nuclear protein (either liver, HeLa or kidney) for 30min and then subjected to exonuclease III digestion. The starred bands indicate specific, self-competable, exonuclease

Figure 7.12. Exonuclease III stop and DNase I protection experiments to delineate the boundaries of ANF binding. The two probes were 5' end labeled at the position shown and subjected to the assays as described in the methods section. The ExoIII assays are shown in parts A and B for probe 1, and 2 respectively while C and D represent the DNase I footprints for the same two probes. Note that in panel A the lane labeled (+) contains homologous self competitor while the (-) lane contains an equal molar amount of pGem1. In B, however, the (-) lane contains no unlabeled non-specific competitor. This control is instead now included in the lane labeled NS containing an equal amount of pGem1. The lanes labeled "No ext" received no protein extract and the lanes labeled "No exoIII" received neither protein extract nor exoIII. The exoIII stop fragments are indicated by asterisks and by their position relative to the albumin transcription start site. A "G + A" sequence ladder is included. The DNase I protection assay (C and D) used the same probes incubated with either BSA (as a control) or either P11 purified Hela extract, or a heparin agarose purified liver protein fraction. These complexes were treated with DNaseI, and the products analyzed by 8% denaturing polyacrylamide gel electrophoresis as shown. Regions of protection from DNaseI digestion are denoted by the bar, with the coordinates as shown. See Figure 7.13 for a summary of the ANF binding sites and a sequence comparison.

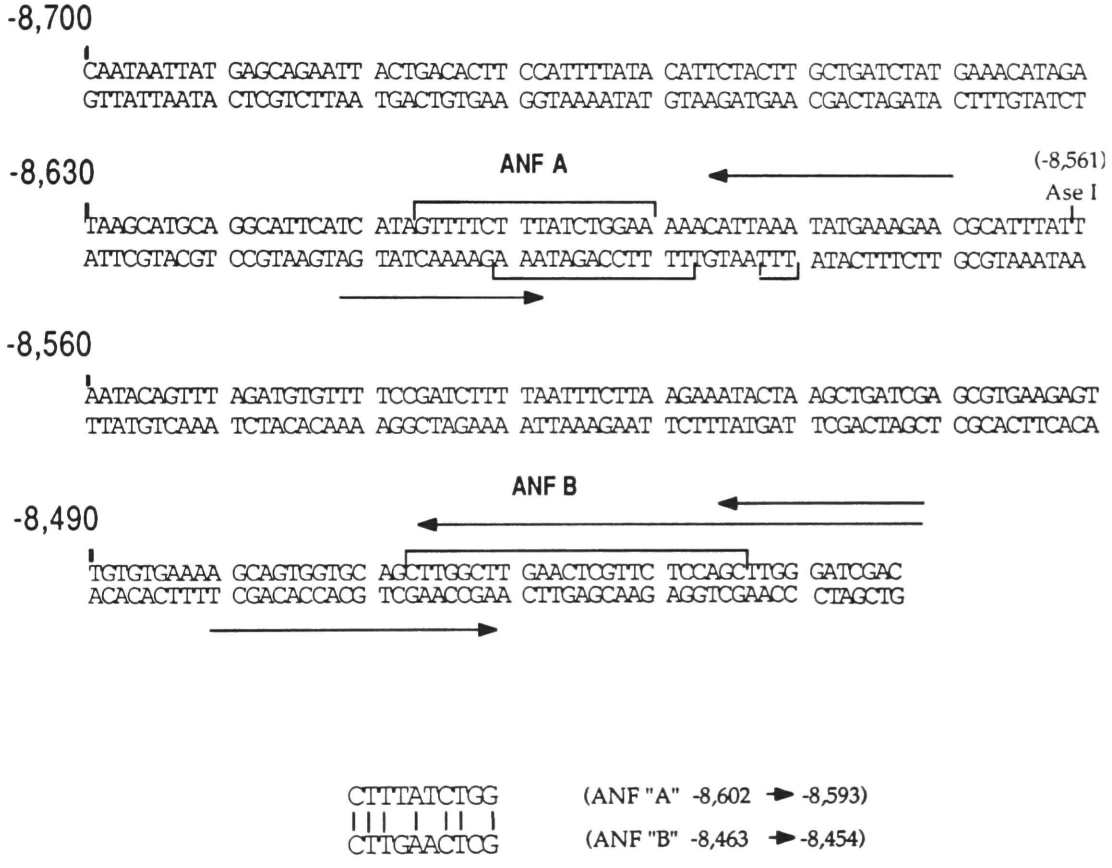

A
G + A
No Exo
No Ext
Liver
HeLa
Kidney
- + - + - +
* -8,598
Probe 1
B
Liver
HeLa
No Exo
No Ext
- + NS - + NS G + A
* 8,447
* -8,466
Probe 2
C
BSA
HeLa
Liver
G + A
-8,601
-8,588
-8,583
-8,581
Probe 1
D
BSA
HeLa
Liver
G + A
-8,445
-8,468
Probe 2
ANF A
ANF B
Pst I
AseI
Pst I
-8,700
-8,561
-8,433
Probe 1
*

stop sites presumably representing points of DNA-protein interaction. Using probe 1 (Figure 7.12A), which contains ANF site A (the stronger binding site), a predominant exonuclease stop was seen at nucleotide -8,598 with liver, HeLa and kidney nuclear extract (Figure 7.12A). Lanes marked (+) contain a 30 fold molar excess of cold unlabeled self competitor while the lanes marked (-) contain an equal fold molar excess of a pGem fragment to keep the total amount of DNA constant.

For ANF site B, an exonuclease stop were observed with both liver and HeLa extracts (Figure 7.12B). In these experiments the (+) again refers to a 30 fold excess of self competitor while the lane marked (NS) contains an equal molar excess of pGem. The (-) lane contains no unlabeled DNA and hence, a reduced amount of total DNA. A strong exoIII stop is observed at -8466 while a much weaker stop is seen at nucleotide -8,447 (Figure 7.12B).

A DNase I protection assay (Figure 7.12 C and D) using the same probes further delineated the boundaries of DNA-protein interaction. These footprint reactions (as described in the methods) were done using either BSA as a control or 5μg of a heparin agarose purified liver fraction or HeLa extract from a 0.4M salt P11 column step (a gift of X. Fu). For ANF site A the footprinted region is depicted (somewhat split) in Figure 7.12C, and includes the DNA sequence defined by the exoIII stop site (see Figure 7.13). The ANF site B is protected as shown in Figure 7.12D by the region encompassed by the black bar. Again, the exoIII stop site is included within this sequence. These

Figure 7.13. Sequence of the functionally important portion of the albumin enhancer negative region (-8.7 to -8.43kbp). The exonuclease stop sites from both strands represented by the arrows, and the DNase I protected regions are represented by the brackets. No footprint has been obtained for the ANF site B lower strand which represents the weaker binding site. A comparison of the ANF A and B binding sites are shown at bottom, indicating a 7/10 sequence match between the two protected regions.



CTTTATCTGG (ANF "A" -8,602 → -8,593)
||| | || |
CTTGAACTCG (ANF "B" -8,463 → -8,454)

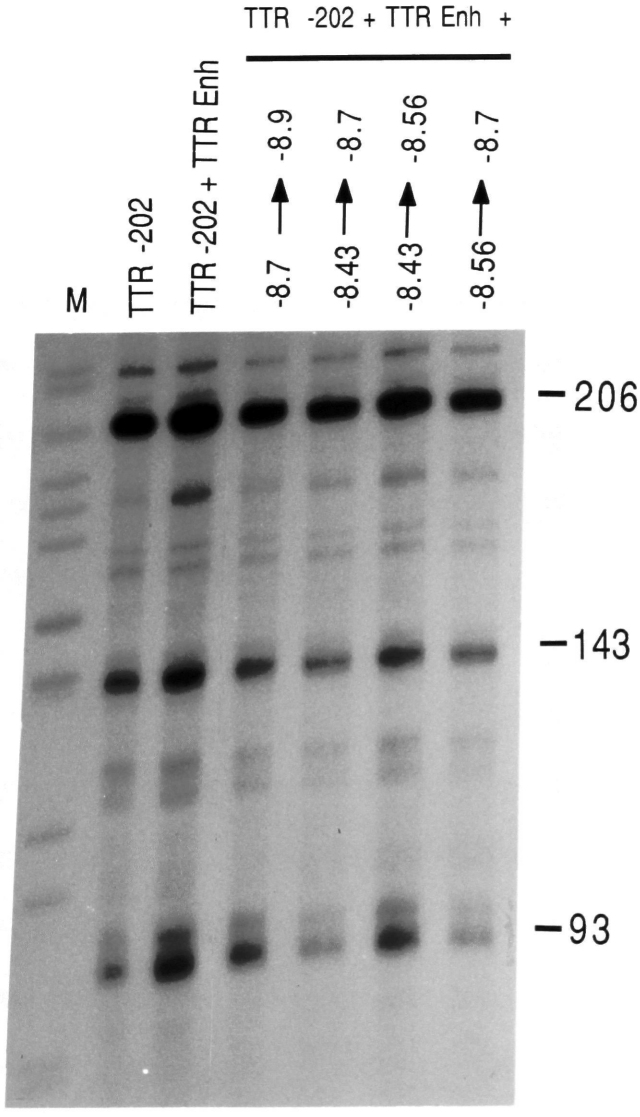
results are summarized in figure 7.13, and a comparison of the sequence within the two protected regions revealed a 7 base pair identity within a 10 base pair sequence. The relative importance of these nucleotides for protein binding will require further mutagenesis and functional studies.

The ability of the ANF sites to interfere with the TTR enhancer is proportional to their affinity

Having identified these two ANF binding sites, a final experiment compared the effectiveness of the two ANF sites in suppressing TTR gene enhancer activity. Transient transfections were done using similar plasmid constructs as described in Figure 7.9 with fragments containing the ANF A or B binding sites now independently inserted between the TTR elements (all constructs shown are in the + orientation). Following co-transfection into HepG2 cells (50μg of TTR DNA) using a SV40 enhancer driven β-globin plasmid (20μg) as a control, both TTR and β-globin gene expression were scored in the same reaction. The 93 nucleotide specific TTR signal (labeled 93 in Figure 7.14) can be evaluated after correcting for variations in transfection efficiency and RNA extraction using the β- globin signal which here consists of a summation of the 206 and 143 nucleotide bands, representing the second and first exons respectively (labeled -206 and -143).

As shown in Figure 7.14, the ANF A containing fragment (–8.56 to –8.7kbp) had about the same negative activity as the larger fragment

Figure 7.14 The mouse albumin enhancer negative region sequences were placed between the TTR enhancer and promoter as in Figure 7.9. For this assay, however, both probes (for the TTR or the β-globin control) were assayed in the same reaction. Again, the TTR promoter (TTR -202) and TTR promoter-enhancer controls are shown (TTR -202 + TTR Enh). The coordinates of the inserted albumin sequences are as marked and are all in the (+) orientation. The 93 nucleotide band represents the specific TTR signal while the 143 and 206bp bands represent exons 1 and 2 of the β–globin gene.


TTR -202 + TTR Enh +
M
TTR -202
TTR -202 + TTR Enh
-8.7 → -8.9
-8.43 → -8.7
-8.43 → -8.56
-8.56 → -8.7
206
143
93

previously described (–8.43 to –8.7) while ANF B containing fragment (–8.43 to –8.56) had at best a slight negative effect. Recall that figure 7.9 showed that the affinity of the ANF A site (-8.56 to -8.7, probe 1) was stronger than that for site B (-8.56 to -8.43, probe 2). Thus, the apparent stronger binding of site A correlates with the greater negative effect it has in transfection assays. Therefore, this assay has served to locate the entire, functional negative effect of region III to the ANF A site located on the –8.56 to –8.7Kb fragment. Due to the close proximity of these sites, however, and the possibility that they each might function *in vivo*, both will be included in the following discussion of Region III. In any event, mutational analysis of these binding sites (ie. using the technique of point mutagenesis) will be required to confirm the role of ANF in albumin enhancer function.

Discussion

The experiments described in this chapter prove that the albumin enhancer contains a functionally acting negative sequence between –8.43 and –8.7Kb upstream of the albumin gene RNA start site. This sequence, presumably acting through the newly described protein ANF, negates the effect of the positive acting region of the albumin enhancer. A negative action is also evident when the element is placed between the enhancer and promoter of the mouse TTR gene. The negative element had no direct effect on the mouse albumin promoter alone, nor was this element able to suppress

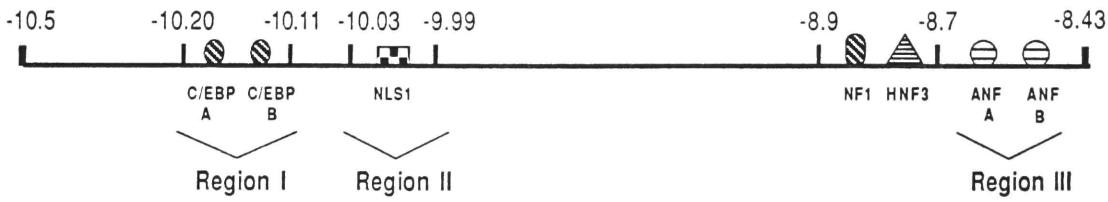
the very strong β-globin activation by the SV40 enhancer of the β-globin promoter in either HepG2 or HeLa cells.

Both the TTR enhancer-promoter and the albumin enhancer-promoter sequences have binding sites for many different proteins, including four factors present more prominently in hepatocytes than in other cell types (HNF1,3,4 and C/EBP) and the NLS1 protein described in chapter 6 (Cereghini *et al.*, 1987; Lichtsteiner *et al.*; Courtois *et al.*, 1988). In the albumin enhancer, the ANF sites between –8.43 and –8.7kbp presumably negate the action of the NLS1 factor that binds at ~10kbp. In the experimental constructs where Region III divides the TTR elements, the negative element is presumably able to prevent the action of one of the other site-specific DNA binding proteins mentioned above.

A model for the tissue specificity of the mouse albumin enhancer

At this juncture, all three of the albumin enhancer regions have been described, and it seems reasonable to discuss a working hypothesis to explain how they might interact to generate tissue specific gene enhancement. The three enhancer regions and the proteins which have been identified as binding to them are shown in Figure 7.15. It has been demonstrated in the preceding chapters that the NLS1 protein is responsible for the complete positive action of this enhancer which can be negated by the action of the ANF binding sites. The C/EBP protein, however, modulates the negative

Figure 7.15 Summary of the DNA-protein interactions occurring on the mouse albumin enhancer. The three regions of functional importance (identified as I, II and III) and the proteins that interact with them have been identified and are shown schematically. Region I has two sites for the protein C/EBP (see chapter 5). Region II was shown to interact with NLS1 (chapter 6), and Region III contains two binding sites for ANF which have been described in this chapter.


-10.5
-10.20
-10.11
-10.03
-9.99
-8.9
-8.7
-8.43
C/EBP A
C/EBP B
NLS1
NF1
HNF3
ANF A
ANF B
Region I
Region II
Region III

effect and enables the enhancer to function. How then, does this overriding mechanism insure that albumin transcription occurs in a liver specific fashion?

The ANF binding activity is present ubiquitously, and while to date it has only been shown to negate transcription in liver derived cell types (HepG2 and SV1), assume for the purpose of this model that ANF is present and able to negate albumin transcription in all cell types. Next consider NLS1, a factor for which binding activity is present in all tissues but specific modification (or an additional protein) is required for stimulatory function. The results presented in this thesis have demonstrated that in the liver cell-types NLS1 is active, whereas in HeLa cells, although binding activity is present, function is lacking. Furthermore, the results from the Palmiter laboratory mentioned earlier strongly suggest (but of course don't prove) that the NLS1 is functionally modified in other differentiated tissues (Pinkert *et al.*, 1987; e.g. spleen, lung, kidney). This is based on the fact that the albumin enhancer sequences in the absence of the negative region (identified in this thesis) could enhance albumin transcription in several non-liver tissues of transgenic mice (a fragment from –10.4 to –9.3kbp produced this effect, and NLS1 is the only identified positive acting element in this region). With these facts as a backdrop several specific situations may now be examined.

In the hepatocyte, NLS1 is functionally active although the ANF is present as well. However, the presence of C/EBP overrides the negative

mechanism and enables the enhancer to function. On the other hand, in a non-liver tissue such as the spleen, NLS1 is also present and functionally active (extrapolating from Pinkert *et al.*, 1987) again in the presence of the negative factor. In this case, however, C/EBP is lacking and therefore the negative effect persists. The end result is no enhancement in this tissue. Finally, in a HeLa cell a dual mechanism appears to keep enhancer directed albumin transcription silent, for NLS1 is not functionally modified and C/EBP is lacking. Hence, it would seem to be both the presence/absence of C/EBP and the state of modification of the NLS1 which controls albumin enhancer function.

In regard to the final point above, perhaps NLS1 is present and functionally active exclusively in fully differentiated tissues (hence the absence of function in the HeLa cells). It that event, C/EBP becomes the key to enhancer function. C/EBP is a liver enriched factor and this model would explain the lack of albumin transcription in most tissues. It could not, however, account for regulation of the albumin enhancer in adipocytes, which are known to produce significant levels of C/EBP (Birkenmeier *et al.*, 1989). In this situation, perhaps the relative levels of all three different factors are such that enhancement does not occur (ie. for example NLS1 is reduced or only partially functional while C/EBP is at a less than optimal concentration). Alternatively, it is possible that a low level of enhancer function does occur in this tissue but its effect is not manifested due to the tissue specific nature of

the albumin promoter. This latter point is significant for as demonstrated in this thesis, the albumin gene has a dual mechanism for insuring its specificity: both a highly tissue specific promoter and enhancer element. It should be keep in mind, however, that *in vivo* the above mechanism could be secondary, for the factors might have already functionally altered the chromatin structure during terminal differentiation of the cell. Hence, the decision whether to express albumin may already be stored on the chromosome.

The testing of the above model provides the basis for future functional studies involving this enhancer element. A number of points must still be examined: 1) the ability of the ANF to function in non-liver cell types; 2) The ability of NLS1 to function in other differentiated tissues; 3) an analysis of enhancer function in adipocytes. Probably, the best way to address these questions in multiple cell-types would be to prepare transgenic mice containing different combinations of these three regulatory regions. Now that the binding sites have been delineated, this type of experiment is quite feasible. Additionally, it has recently been found that the NLS1 oligonucleotide (described in chapter 6) can be oligmerized and used to enhance β–globin expression in HepG2 cells. This same construct can now be used to analyze NLS1 function in a host of differentiated cell types (including adipocytes) by transfection. Finally, the negative element can be tested in non-liver cell types by making it part of promoters known to be active in

HeLa cells.

Molecular extinction suggested the existence of negative regulation

Somatic cell genetic studies have shown that when certain hepatoma cells are fused with each other or with a variety cell types, suppression of liver gene transcription occurs (Killary and Fournier, 1984, Petit *et al.*, 1986; Gourdeau *et al.*, 1989). This extinction is subject to gene dosage effects and is reversible, often following the loss of a particular chromosome in the extinguishing parent. This has led to the conclusion that these "extinguisher" genes act in a trans-dominant manner. Tse 1 (tissue-specific extinguisher) was the first locus identified as a discrete genetic element residing on mouse chromosome 11 or human chromosome 17. (Killary and Fournier, 1984), and the product (s) encoded by this locus specifically extinguishes tyrosine aminotransferase transcription in hepatoma x fibroblast cell hybrids (Gourdeau *et al.*, 1989). Extinguishers such as these have also been found to be involved in the regulation of many other liver specific genes including albumin. Petit *et al.* (1986) have described cell fusion studies of FaO rat hepatomas with microcells (which contain only a small number of chromosomes) that extinguish albumin expression, and they have narrowed this locus to a single unidentified chromosome (Petit *et al.*, 1988).

Is ANF the factor involved in albumin extinction; the product of this extinguishing locus? The answer to this question is no for several reasons.

First of all, the ANF is present ubiquitously, including in liver cells and hence is not acting in the same dominant fashion as the extinguishers described above. Furthermore, Petit *et al.* demonstrated that extinction occurred upon a transfected albumin promoter containing only 400bp of upstream sequence, and therefore this enhancer is not even necessary for that process. It seems more likely that the extinguishing gene product is acting indirectly to effect the activity of one of the factors involved in albumin transcriptional regulation. For example, if it interfered with the regulation of C/EBP production it could negate the enhancer (for C/EBP is required to override the negative element) and severely restrict the activity of the promoter. Likewise a decrease in the activity of any of the other factors involved in albumin transcription would have a similar effect. It has recently been suggested that the loss of HNF1 activity and the appearance of a non-functional variant (vHNF1; see introduction) might be involved in this process (discussed in chapter 8).

Mechanisms of action of the negative element

The exact protein-protein interactions involved in manifesting the negative effect of ANF is currently under study. The fact that the negative element does not extinguish the basal level transcription from the albumin promoter (-300 sequences) nor does it require its own promoter, suggests that the ANF doesn't specifically inhibit initiation complex formation but rather

blocks the enhancing activity of upstream factors (either by interfering with their binding or by masking their activating functions). Perhaps the binding of the ANF interferes with the binding of NLS1 in the absence C/EBP. There is precedent for these types of protein interaction in several systems including most prominently the yeast mating type genes where for example, the MATα1 gene has been shown to change the α2 protein in such a fashion that it no longer recognizes a specific binding site (Goutte *et al.*, 1988). Alternatively, the ANF sites either independently or perhaps by interacting with each other, might form a large complex or secondary structure that prevents the enhancer from functioning .

The negative element resembles a yeast silencer

In many ways this albumin negative element resembles the silencers of yeast in their ability to function in an orientation and distance free fashion to block transcription (Brand *et al.*, 1985, 1987). Like the yeast silencer, this element is not promoter specific and works through a cis-acting sequence known to bind cellular factors. As the SIR proteins and binding sites from yeast and the ANF protein are better characterized a comparison will certainly be warranted.

Conclusion

Future work will attempt to address the mechanism by which this

negative region functions and by which the C/EBP overrides the effect. As each of the three albumin enhancer proteins are purified and cloned it will become possible to explore the protein-protein interactions responsible for the this intricate regulatory scheme. From this discussion, however, further support is drawn for the conclusion that combinations of transcription factors interact to produce regulated gene activity in differentiated cells. While at the the present time not enough is known to suggest specific mechanisms of action for these proteins, the first step in solving such interactions is the clear delineation of DNA binding sites and the characterization of their cognate proteins.

Chapter 8

Transcriptional Regulation in Rat Hepatoma Cells

Hepatoma cells while useful recipients to test recombinant genes transcribed specifically in hepatocytes, do not exhibit normal rates of transcription (as compared to the rat liver) for an array of hepatocyte expressed genes (Clayton *et al.*, 1985; Friedman *et al.*, 1987; chapter 1). Additionally, numerous hepatoma cells lines have been isolated which have dedifferentiated and lost the ability to manifest liver specific functions completely (Descharette and Weiss, 1974). Since many genes, including albumin, are differentially transcribed in these cell-types, they represent a useful system for exploring the molecular mechanisms underlying gene control. Using gel-shift analyses, the relative levels of known transcriptional proteins can be probed in nuclear extracts from the different cell lines and correlated with the transcription rates of their dependent genes. Furthermore, with the availability of clones for several liver enriched transcription factors, it is possible to measure their rates of transcription (the factors) in each cell-type. This approach enables a beginning understanding of the process by which the activity of the transcriptional regulators themselves are controlled in a specific cell type.

Work in several laboratories has identified, characterized and now cloned at least four nuclear protein factors, which have been shown to be involved in hepatocyte specific gene expression (HNF1, C/EBP, HNF3, and HNF4; chapter 1). These four liver enriched proteins, along with the factors NLS1 and ANF (chapters 6 and 7), are known to be involved in the

transcriptional regulation of both the mouse albumin and transthyretin (TTR) genes. A summary of the DNA-protein interactions that occur upon the albumin and TTR regulatory regions is diagrammed in Figure 8.1.

With the availability of DNA clones for each of these genes and a set of DNA oligonucleotides that each contains a binding site for one of these six factors, a series of rat cell lines was examined (Figure 8.2; Deschatrette and Weiss, 1974; Deschatrette *et al.*, 1980) whose profile for hepatocyte-like transcription has been described (Clayton *et al.*, 1985; Friedman *et al.*, 1987). FaO cells represent the parent hepatoma cell line and are considered most similar to hepatocytes; C2 is a variant of FaO that has simultaneously lost gluconeogenic capacity and albumin production, and several other liver specific functions; C2 rev7 is a revertant of C2 that has regained gluconeogenic capacity and some ability to produce albumin. H5 is a variant of the cell line H4II, which like C2 is dedifferentiated cell line that has lost liver-like functions.

Nuclear run-on transcription rate analysis

To determine the transcription rate of various liver specific genes, including the four transcription factors, nuclear run-on assays were performed with nuclei from rat liver and the cell lines described above (Figure 8.3; Table 8.1). Two genes encoding secreted serum proteins transcribed at high rates: albumin and transthyretin show very different

Figure 8.1. Summary of the DNA-protein interactions occurring in the promoter/enhancer regions of the mouse albumin and TTR genes. This figure is based on the results of functional analysis and *in vitro* binding studies on the 5' flanking regions of these genes (albumin; this thesis and references therein; TTR, Costa *et al.*, 1986,1988a, 1989). The legend (bottom), denotes the identity of the different factors.

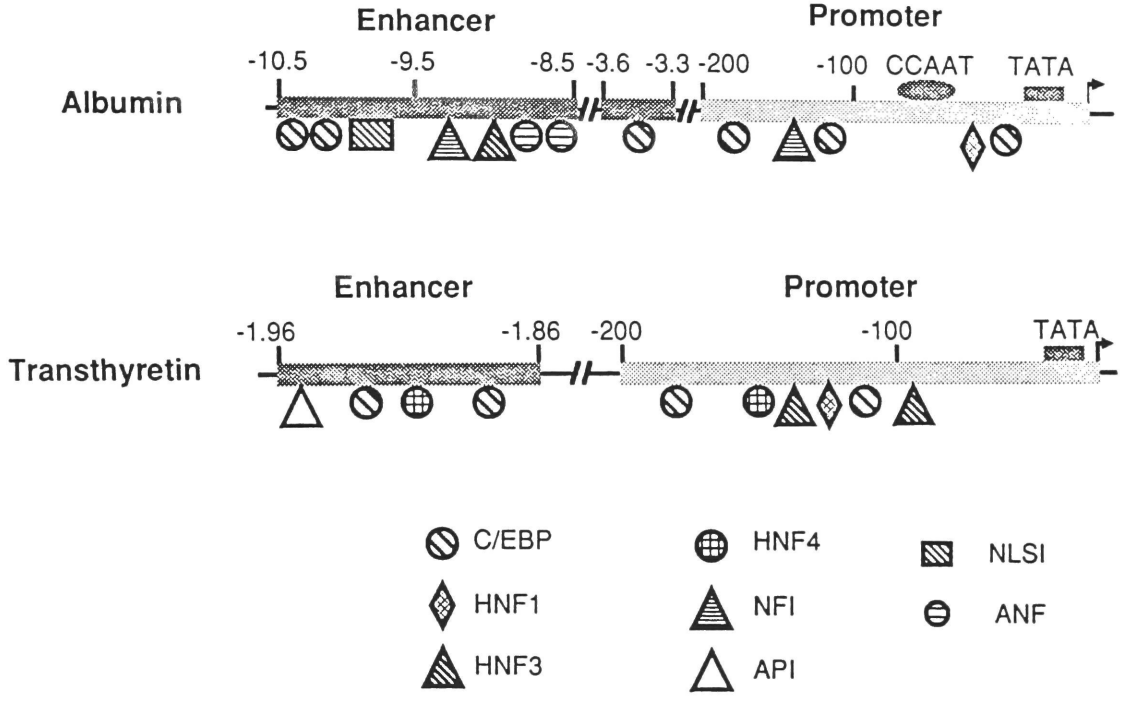
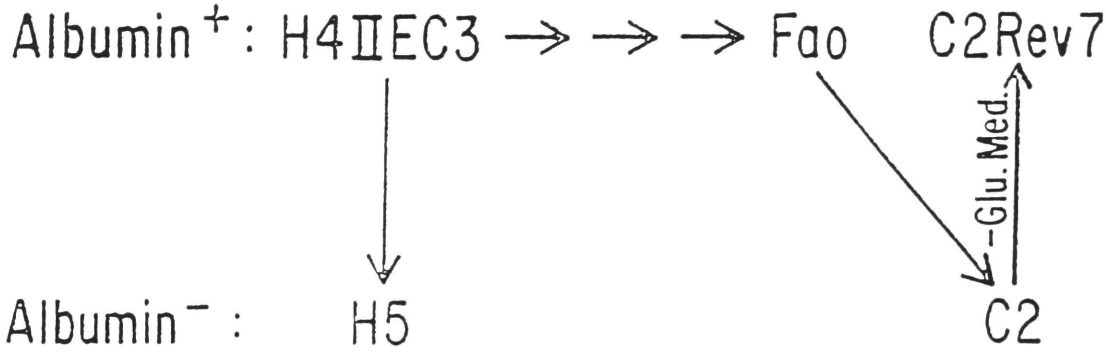

Enhancer
Promoter
Albumin
-10.5
-9.5
-8.5
-3.6
-3.3
-200
-100
CCAAT
TATA
Enhancer
Promoter
Transthyretin
-1.96
-1.86
-200
-100
TATA
C/EBP
HNF1
HNF3
HNF4
NFI
API
NLSI
ANF

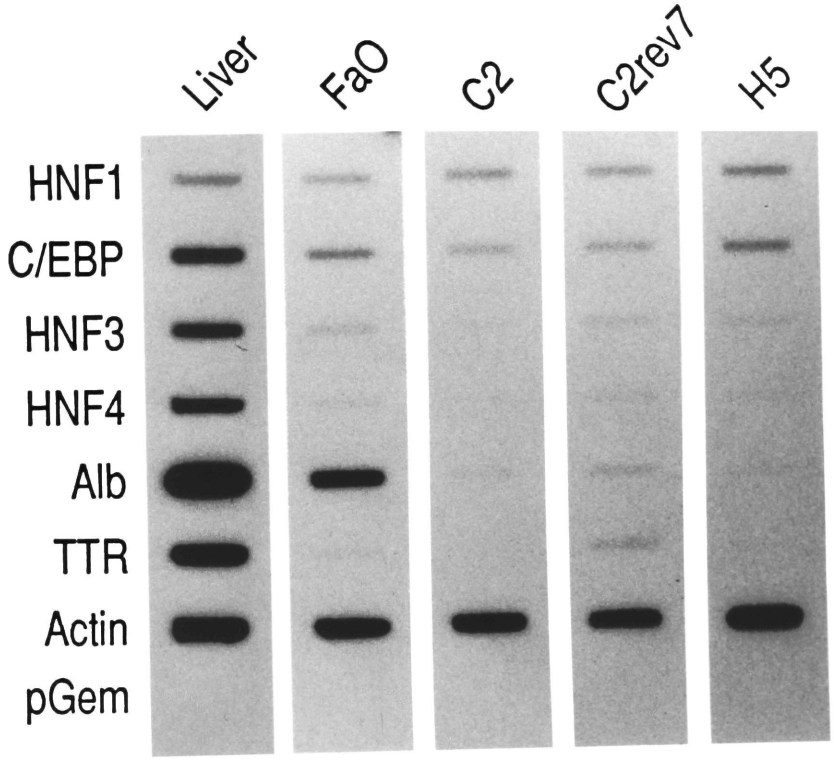
Figure 8.2. Origin of rat hepatoma cell clones (Deschatrette and Weiss, 1974; Deschatrette *et al.*, 1980). These clones are divided into two groups; the albumin producing, differentiated cell lines (top) or the non-albumin producing, dedifferentiated cell types (bottom). The arrows extends from the parent of each cell line. These cell lines were selected based on the ability to grow in the presence or absence of glucose. For a more thorough discussion of these cell lines see text.


Albumin+ : H4IIEC3 → → → Fao C2Rev7
-Glu. Med.
Albumin− : H5 C2

transcription patterns in FaO cells. Albumin is transcribed in FaO cells at about 18% the rate it is in liver but TTR transcription is transcribed at only 1% that in rat liver. The transcription rate in C2 cells of the albumin gene was barely above the baseline level and that for the TTR was undetectable. In C2rev7, the hepatocyte-like revertant, a small but definite resumption of transcription rate was observed both for the albumin and TTR genes, whereas the H5 did not transcribe either of these genes. These results are quite similar to those described earlier for these cell lines (Clayton *et al.*, 1985; Friedman *et al.*, 1987). Therefore, while both albumin and TTR maintain identical transcriptional specificity in the different cell lines (activated in liver, FaO and C2rev7 and inactive in C2 and H5) their relative transcription rates vary significantly.

The results of the nuclear run-on assays for the transcription factors are more complex. The most striking, finding is the five-fold decrease in the transcription of the C/EBP gene in FaO cells compared to liver and the 40-fold decrease in HNF3 and HNF4 transcription. In contrast, to the HNF3, HNF4 and C/EBP genes, however, the transcription rate of the HNF1 gene is almost the same in FaO cells as in liver and does not change significantly in any of the other cell lines. C/EBP transcription is also similar (within a factor of 2) in all the hepatoma cell lines (Figure 2; Table 1) while transcription of HNF3 and HNF4 is very low in FaO and C2rev7 cells and almost undetectable in C2 and H5 cells.

Figure 8.3. Nuclear run-on transcription analysis of albumin (ALB), Transthyretin (TTR) and the genes coding for HNF1, HNF3, HNF4 and C/EBP. Nuclei from rat liver tissue or the FaO, C2, C2rev7 and H5 cells lines were isolated and allowed to chain-elongate nascent RNA in the presence of ^{32}P-dUTP as described previously (Weber *et al.*, 1987; Hofer and Darnell 1981) The RNA was extracted and hybridized to nitrocellulose filters containing excess DNA affixed as dots. Unhybridized DNA was digested and washed away leaving the hybridized RNA to be detected by autoradiography. Spotted DNAs include: rat albumin (a 2400bp genomic fragment; rat B clone Sargent *et al.*, 1981), rat TTR (a 440 bp fragment of the rat TTR cDNA, Dickson *et al.*, 1985); HNF1 (the 3052bp λ16 cDNA clone, Frain *et al.*, 1989) C/EBP (500bp pPst1 fragment, Xanthopoulos *et al.*, 1989). HNF 3 (655 nucleotides) and HNF4 (972 nucleotides) were cDNAs in plasmids (Lai, Sladek and Darnell, unpublished). A Chicken β-actin clone was used as a control (Cleveland *et al.*, 1980).


Liver
FaO
C2
C2rev7
H5
HNF1
C/EBP
HNF3
HNF4
Alb
TTR
Actin
pGem

Normalized Nuclear Run-On Transcription Data

	Liver	FaO	C2	C2Rev7	H5
Albumin	100	18	0.1	1.0	-
TTR	189	0.6	-	1.2	-
HNF1	2.0	1.8	1.9	2.0	1.7
C/EBP	92	26	18.5	19.5	16
HNF3	46	1.0	0.5	1.0	0.5
HNF4	39	0.6	-	0.6	-

Table 8.1. The results for the nuclear run on analysis were normalized based on the size of the DNA probes and the actin transcription value. The albumin transcription signal in the liver was arbitrarily given a values of 100 and serves as the standard of reference. This table represents the averaged results of two different experiments, one of which is shown in Figure 8.3.

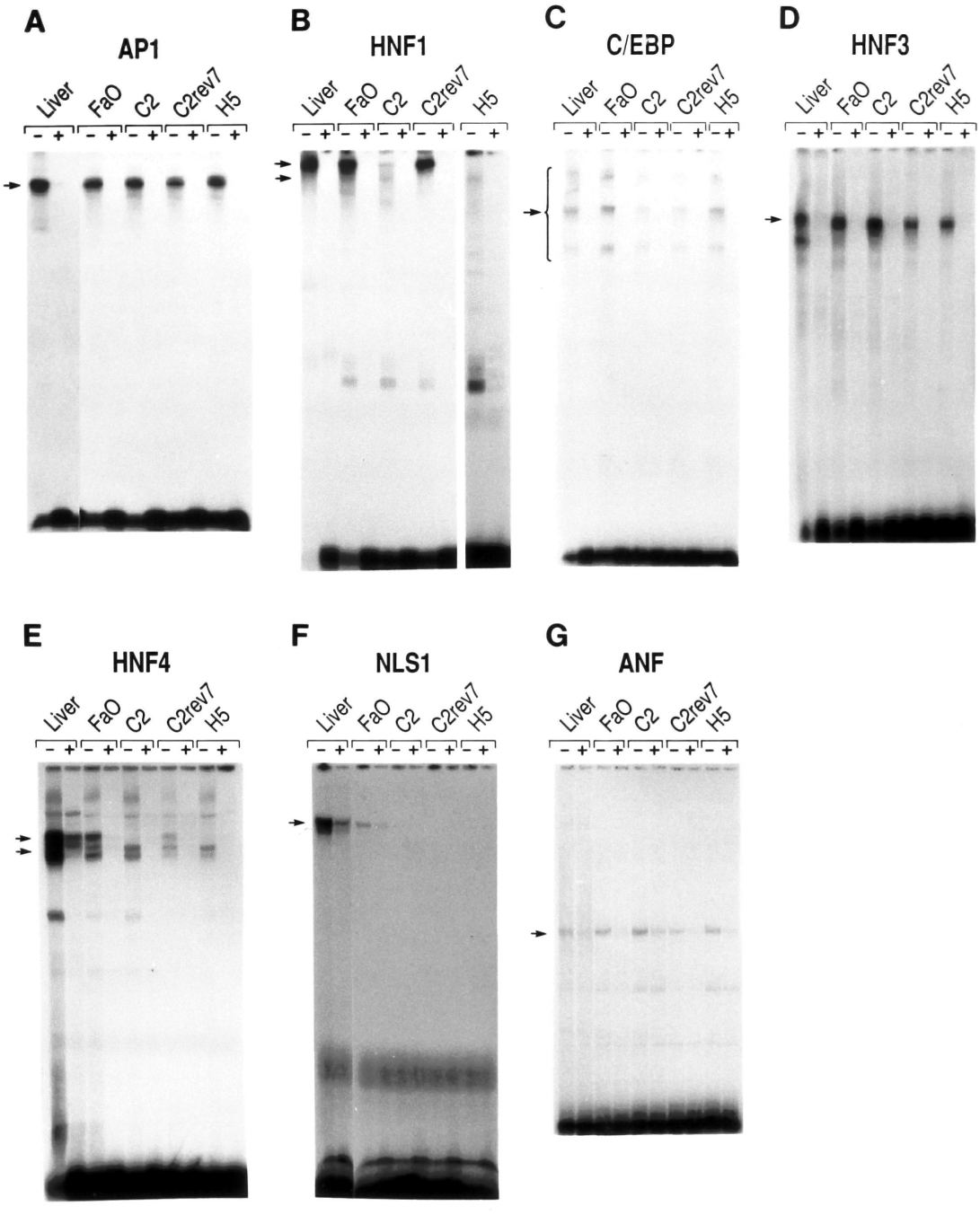
Gel-shift analysis of Nuclear Factors in Hepatoma Cell Lines

Having established the transcriptional patterns for these four factors, the nuclear protein levels of these factors was established in the nuclei of the different rat cell types to determine if transcriptional regulation accounted for their control. Furthermore, this allowed an analysis of the NLS1 and ANF factors for which clones are not yet available (Figure 8.4). Nuclear extracts of rat liver cells and of the four cell lines were prepared. Each extract contained about the same amount of protein capable of binding the AP-1 site of α 1-antitrypsin (Figure 8.4A; oligonucleotide B, Grayson *et al.*, 1988) indicating that the method for obtaining factors was similarly effective in all cell types.

Surprisingly, the pattern for HNF1 binding activity in the cell lines was very different from that of its transcription rate, which varied little between the cell types. While FaO, C2rev7 and liver nuclear extracts contained similar amounts of HNF1 DNA binding activity, C2 cells had at least ten to twenty-fold less HNF1 binding activity, and in H5 cells it was undetectable (Figure 8.4B). Since the transcription rate of albumin is very low in C2rev7 cells compared to FaO and liver cells, this result shows that HNF1 alone cannot insure high albumin transcription rates.

Another interesting result of the HNF1 shift assays is the appearance of a faster moving DNA-protein species in the C2 and H5 nuclear extracts (Figure 4B, bottom arrow; note that the H5 lanes represent a five fold longer autoradiographic exposure indicating the low level of this complex). This

Figure 8.4. Binding activities for AP1, HNF1, C/EBP, HNF3, HNF4 and NLS1 in nuclear extracts prepared from the rat liver or from the hepatoma cell lines FaO, C2, C2rev7 and H5. Oligonucleotide probes were prepared to represent these binding sites and in Table 4.2, except for ANF which is the same probe used in Figure 7.11. DNA probes were 3' end labeled and incubated (1ng) with 10µg of nuclear extract in a 20µl binding reaction containing 4µg of poly (dI-dC):poly (dI-dC) in 20mM HEPES pH 7.9, 40mM KCl, 2mM $MgCl_2$, 1mM DTT, 0.5mM EGTA and 4% ficoll, for 30min at room temperature. Following incubation, 4µl of this mixture was electrophoresed on 5 - 9% acrylamide gels run in 25mM Tris-borate and 0.25mM EDTA, in the absence (-) or presence (+) of homologous unlabeled DNA (50 fold molar excess).


A
AP1
B
HNF1
C
C/EBP
D
HNF3
E
HNF4
F
NLS1
G
ANF
Liver
FaO
C2
C2rev7
H5
- +

activity has been observed by others, also in dedifferentiated, non-liver expressing cells, and has been named variant HNF1 (vHNF1; Baumheuter et al, 1988; Cereghini *et al.*, 1988). The makeup of this vHNF1 protein is far from clear for it could be a modified form (either pre-or post-translational) of the HNF1 gene or alternatively the product of a different gene. The function of vHNF1 (e.g. is it a positive or negative acting factor) is equally obscure (Tronche *et al.*, 1989), but it appears that it is the lack of HNF1 and not necessarily the presence of vHNF1 in the C2 and H5 cell types that is most probably responsible for the lack of albumin and TTR gene transcription.

Reference to C/EBP binding activity is helpful in trying to interpret the albumin and TTR transcriptional results (Figure 8.3C): FaO cells have abundant C/EBP binding activity and both C2 and C2rev7 cells have less. Therefore, unlike the case with HNF1, C/EBP binding activity better parallels its transcription rate. Also, since both the albumin promoter and enhancer elements contain functional C/EBP binding sites (Figure 8.1) , this lower amount of C/EBP could partly explain the much reduced albumin transcription rate in C2rev7 cells.

The results of the HNF3 binding activity were remarkably different from both the HNF1 and C/EBP results (Figure 8.4D). HNF3 binding activity is present at approximately equal amounts in all 4 hepatoma cell types tested (Figure 8.4E), although it is not transcribed well in any of them, as compared to the high rate observed in liver. These results suggest that post-

transcriptional events are involved in maintaining the level of this protein in hepatoma cells (In animal tissues, however, HNF3 is regulated transcriptionally not post-transcription, Xanthopoulos et al., unpublished.)

HNF4 binding activity is present in FaO cells at 10% the level of liver cells, and this drop off corresponds to a transcriptional decrease. Furthermore, a low amount of HNF4 binding activity is present in C2 and H5 cells where transcription is barely detectable as compared to liver cells. However, the gel-shift pattern of this binding activity was not completely normal. C2 extracts produce two HNF4 complexes (lower arrow figure 3E) but lack the third slowest moving complex (upper arrow, Figure 3E). C2rev7 which regained some TTR transcriptional activity have at least some proteins that form all three HNF4 complexes, but at a much reduced level (even though its transcription rate is nearly identical to that in FaO cells). This is similar to the situation observed for the vHNF1 factor and suggests that posttranslational modification might influence HNF4 function.

NLS1 and ANF protein levels

Lastly, the relative protein abundance was determined for the factors NLS1 and ANF which confer the positive and negative actions of the albumin enhancer. In the series of hepatoma derivatives tested here, only FaO cells contain detectable amounts of the NLS1 binding protein (Figure 8.3F), although at a much lower level than that found in liver. This result

correlates with the higher albumin transcription rate in FaO cells compared to C2, H5 or C2rev7 and the reduced albumin transcription observed in FaO cells compared to liver. Additionally, the levels of ANF (albumin negative factor) were measured in the cell lines and shown to be fairly constant. Hence, increased production of this negative factor does not seem to be involved in turning albumin off in these cell types.

Discussion

These results contribute to two lines of inquiry about transcriptional control in specialized cells: 1) the regulation of transcription factors themselves and 2) how various factors combine to regulate transcription of specific genes and specific cell types. First, certain hepatoma cells even though they are markedly dedifferentiated, still transcribe some liver-enriched transcription factors at rates comparable to their differentiated parents. HNF1 and C/EBP are in this category. Two other factors HNF3 and HNF4, originally described as proteins necessary for high rates of transcription of TTR, are not transcribed well in hepatoma cells, at least under the cell culture conditions used in this study. Despite this minimal level of HNF3 transcription, however, HNF3 DNA binding activity was abundant in all the cells tested. This result is consistent with earlier studies which demonstrated that in C2rev7 cells considerable albumin mRNA is present with very low transcription rates (Jefferson *et al.*, 1984; Clayton *et al.*, 1985; Friedman *et al.*,

1987), presumably resulting from posttranscriptional stabilization of the mRNA.

Similarly, the HNF1 gene was transcribed well in C2 and H5 cells, although little or no protein capable of DNA binding was found in these cell extracts. In both cases a smaller amounts of the faster migrating bands were evident. Again, whether these bands are derived from post-transcriptional modifications of HNF1 or are the product of a distinct gene is not known. In any case, the major event is the loss of the larger 80KD HNF1 protein that produces the slower migrating band, and the role, if any, of this variant HNF1 remains to be established. Since the gel retardation assay scores DNA binding activity and not amounts of protein, antibodies to the various transcription factors will be needed to determine how much protein is actually present. Then, it might be possible to decipher how much of the apparent post-transcriptional regulation is in fact, post-translational modification.

Correlation with Albumin and TTR gene expression

Can the transcriptional behavior of the genes for albumin and TTR, two prototypical liver products, be explained by the transcriptional activity of their transcription factor genes? Some correlations seem evident but numerous questions remain. First, albumin transcription in FaO cells seems to be the most easily explained. HNF1 and C/EBP are transcribed at a relatively high rate and their DNA binding activities are equally abundant in

FaO cells and in liver. Both factors are known to be important in transcription from the albumin promoter (Figure 8.1). However, the protein level of NLS1, the single known positive activator of the albumin enhancer (Chapter 6), is decreased significantly in FaO cells as compared to liver. In the C2rev7 cells, the lower level of NLS1 DNA binding activity could account for the further decrease in the albumin gene transcription observed. Additionally, it appears that C2 and H5 do not transcribe any albumin because they do not contain a functional HNF1 protein. This implicates HNF1 as having the essential role in determining the specificity of albumin transcription. It would seem that a minimum level of HNF1 is required to allow a basal level of albumin transcription to occur, and that the other factors work primarily to modulate transcriptional rate.

Regarding the TTR ene, four of the factors assayed here are known to have binding sites either in its upstream enhancer or in the promoter proximal sequences (C/EBP, HNF1, HNF3 and HNF4; see Figure 8.1). While the transcription rates for HNF3 and HNF4 in the cultured cells correlates with that of the TTR gene (all decreased with respect to the liver) the levels of DNA binding activities of these factors found in the nuclear extracts did not. For example, HNF3 binding activity was the same in all extracts. These results suggest that either these proteins can bind DNA *in vitro* but can't activate transcription inside the cells or the decreased level of HNF4 alone dictates the loss of TTR transcription. There have been other reported cases

where DNA binding activity *in vitro* has been disassociated from transcriptional activation *in vivo* (e.g. serum response factor levels in cells not stimulated by serum or PDGF. (Treisman, 1986 Treisman *et al.*, 1989). In those cases, it is assumed that either some change in the factor or some other protein is required to render the factor active *in vivo*. The possibility that HNF4 concentration is critical, however, is made plausible by the results in C2rev7 cells in which there is a small amount of HNF4 binding activity and a low but detectable level of TTR transcription (as opposed to the rat liver). Another obvious possibility is that several of the proteins apparently needed for TTR expression are all present in lowered concentrations (ie. in the FaO or C2rev7 cells) and therefore, cannot, under these circumstances, interact properly in FaO cells to stimulate transcription.

Conclusions

Almost all genes studied thus far which are regulated in a cell-specific fashion have multiple binding sites for a variety of proteins. Therefore, as these experiments suggest, it is very likely the interaction between those proteins that define a given level of transcription for a gene and not just the presence or absence of individual factors. Hence, the transcriptional variability inherent in these hepatoma cell systems is extremely useful for sorting through this complicated task. Experiments can be planned in the not too distant future where recombinant genes for these different factors can be

introduced into a dedifferentiated hepatoma cell (ie. C2) to observe if the liver-like phenotype can be reestablished.

Lastly, these experiments demonstrate that the cell-type control of these transcription factors is a complicated matter, with regulatory events occurring at both the transcriptional and post transcriptional level. The ultimate understanding of how these factors are differentially controlled in the cells (including the elucidation of potential autoregulatory mechanisms) is important for it will elevate the study of hepatocyte differentiation one step up the regulatory cascade towards an understanding of the molecular events which underlie cellular determination.

Chapter 9

Conclusion

This thesis has analyzed the regulatory elements involved in the control of the serum albumin gene, and through this process identified two novel DNA binding activities (NLS1 and ANF). Having determined the different proteins which interact with the albumin gene, their tissue specificity and possible means of regulation, the mechanisms responsible for albumin transcriptional control are now better understood. Another major reason for studying albumin (an adult hepatocyte-expressed gene), however, was to address the question of coordinate gene control in the maintenance of the differentiated hepatocyte. In this regard, the concluding chapter of this thesis will include both a recapitulation of the results obtained for the albumin promoter and enhancer regions and a comparison of these elements with the regulatory regions of Transthyretin (TTR; Costa *et al.*, 1986, 1988a, 1989) and α1-antitrypsin (Grayson *et al.*, 1988a, 1988b), two additional serum protein genes transcribed primarily in the liver (Figure 9.1). From the study of these genes and the identification of their upstream transcription factors, several general rules for the design and function of their regulatory regions have emerged.

I. Multiple sites of DNA-protein interaction are present on the regulatory elements of hepatocyte-expressed genes.

The first and most obvious conclusion from observing Figure 9.1 is that multiple proteins interact with the upstream regulatory sequences of all

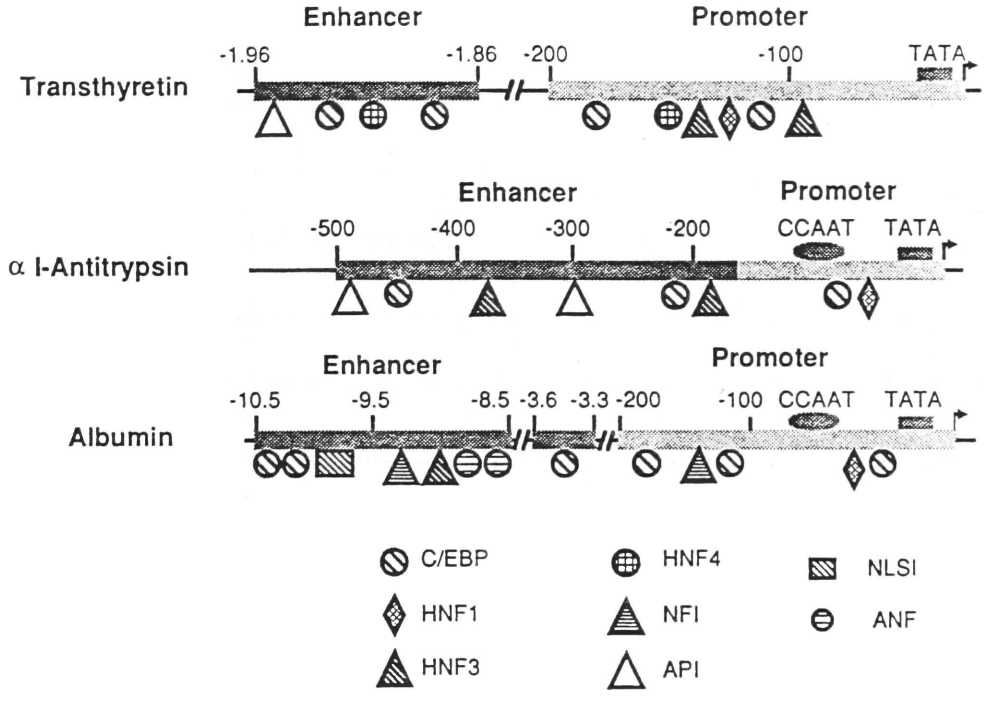


Figure 9.1. Summary of the known sites of DNA-protein interaction that occur on the mouse transthyretin (TTR), mouse α1-antitrypsin and rat and mouse albumin genes (the rat promoter and the mouse enhancer; as described in this thesis the regulatory regions of the two genes are virtually identical). At the bottom of the figure is a diagram representing each cellular factor. This figure is based on results from the following: TTR gene (Costa *et al.*, 1986,1988a, 1989) a1-antitrypsin (Grayson *et al.*, 1988a; Costa *et al.*, 1989) and albumin (this thesis; Babiss *et al.*, 1987; Cereghini *et al.*, 1987; Lichtsteiner *et al.*, 1987; Liu *et al.*, 1988). For a more detailed explanation of this figure see text.

three of these genes. This thesis has demonstrated the numerous protein interactions that occur on the albumin enhancer (as many as 6) and promoter (at least 5; note that the Figure 9.1 includes data generated by other research groups studying the albumin promoter; Cereghini *et al.*, 1987; Lichtsteiner *et al.*, 1987). In addition, Zaret and colleagues have recently reported yet another C/EBP site to exist at -3.5kbp which is shown in the diagram (Liu *et al.*, 1988). Interestingly, this same degree of protein interaction is present on the other two genes described here. For example, there are as many as 10 sites of DNA-protein contact on the mouse transthyretin gene upstream sequences and eight on the mouse α1 antitrypsin regulatory regions. Therefore, these liver specific genes have complicated genetic elements, containing many binding sites for regulatory proteins.

II. Both Promoters and Enhancer elements exist

These sites of protein interaction can be found either within the first 100 or 200 nucleotides or as far as 10 kilobases upstream from the start site of RNA synthesis. In all three of the cases shown in Figure 9.1, both a hepatocyte specific promoter and enhancer element exist, and each is required to achieve maximal rates of transcription. Furthermore, for some hepatocyte-expressed genes (ie. α-fetoprotein) several distinct enhancer elements have been elucidated (Godbout *et al.*, 1988a).

One intriguing aspect of these promoter and enhancer elements,

however is the great similarity between them, for the same group of proteins appears to interact with both. For example, multiple C/EBP binding sites are found on both the albumin promoter and enhancer DNA sequences. This same multiplicity of sites is observed for the two other genes under discussion (see Figure 9.1). Additionally, HNF3 and HNF4 sites occur on both elements of the TTR gene. This common link between these liver specific enhancer and promoter elements is similar to the situation observed for the immunoglobulin genes where for example, the octomer motif is found in several places in the upstream regions (Sen and Baltimore, 1986; Lenardo *et al.*, 1987).

It is now known that several of these liver specific promoter elements (including albumin and TTR; Heard *et al.*, 1987; Costa *et al.*, 1988a, 1989) can be placed upstream of a heterologous promoter (in either orientation) where they act to boost the rate of transcription in a fashion similar to classical enhancer elements. Therefore, the above facts probably indicate that the mechanisms by which promoter and enhancer elements activate transcription are not very different, and perhaps enhancer elements work simply by increasing the concentration of given trans-activating protein at the RNA start site.

III. Both cell specific and ubiquitous factors are involved in liver specific gene transcription

The factors involved in stimulating transcription of a particular liver specific gene are a mixture of proteins that can be found in extracts of many cells (ie. AP1, NF1 and CAAT binding proteins) and a limited number found predominantly in hepatocytes or hepatoma cells and at low levels (or not at all) in other tissues. The four cloned factors already described, HNF1, HNF3, HNF4, C/EBP are examples of the latter group (Courtois et al., 1987; Frain et al., 1989; Graves *et al.*, 1986; Landshulz *et al.*, 1988; Costa *et al.*, 1989). Furthermore, it is known that certain proteins can exist which might be present in all cell-types but specifically "modified" only in certain situations. NLS1 which was described in this thesis is a candidate for such a factor as are AF-1 (the ApoCIII gene Leff *et al.*, 1989) or ϕNF1 (Lichtsteiner et al., 1987, Paonessa et al., 1988). Alternatively, these proteins might represent members of multigene families for which only the liver representative is active.

In any event, the above suggests a model in which the ubiquitous factors act only to increase the magnitude of the transcriptional effect without determining specificity. Instead, it is the presence of specific combinations of liver enriched factors that generates tissue specific gene expression. This type of regulation allows more variable control for no single activating switch exists. That is, by possessing multiple sites for liver enriched proteins, the genes can presumably be modulated by a larger number of different stimuli.

IV. Positive and negative regulatory factors exist

While most of the elements described so far appear to be positive acting and additive, a smaller group of functionally negative acting DNA-regions and proteins also exists Negative acting elements have been described for several hepatocyte-expressed genes including apolipoprotein CIII, retinol binding protein, human α1-antitrypsin gene and α-fetoprotein (Muglia and Rothman Denes, 1986; Reue *et al.*, 1988; Colantuoni *et al.*, 1987; D'onofrio et al., 1989; Camper and Tilghman, 1989). In this thesis the albumin gene was shown to have a negative element which contains two binding sites for a negative acting protein, ANF (albumin negative factor). Whether ANF plays a general role in liver regulation by interacting with other liver specific genes remains to be explored. Naturally, the precise transcriptional arrangements which exists upon the regulatory regions of liver specific genes (positive/negative/both) are as varied as the different DNA-binding protein combinations which can occur.

V. The liver enriched factors interact with multiple genes

Most all the liver enriched or modified factors identified so far interact with more than one gene expressed chiefly in hepatocytes (Figure 9.1) This suggests their involvement in coordinate control, simultaneously activating several genes involved in maintaining the hepatocyte-like phenotype. For example, the C/EBP and HNF1 and HNF3 proteins interact with all three of

the genes under discussion. In a like fashion, it has recently been found that NLS1 also binds to the constitutive activating region of the mouse glutathione-s-transferase gene. While this gene is expressed at high levels pericentrally in the liver, it is believed that the NLS1 interacts with the promoter element of the gene which directs the basal level expression in all hepatocytes (Paulson et al., In press). Hence, numerous genes transcribed predominantly in the liver interact with the same group of liver enriched regulatory proteins.

Perspectives

This thesis therefore suggests that hepatocyte specific gene expression relies on the combinatorial action of several, different regulatory factors. Through the interplay of these factors, only some of which are tissue specific, a regulated differentiated cell phenotype is established. Importantly, this type of model can account for the fact that certain liver genes are expressed in other tissues. For example, TTR is transcribed in certain regions of the brain (the choroid plexus) and α1-antitrypsin in macrophages and the kidney (Costa *et al.*, 1989). It is therefore reasonable to assume that the ubiquitous factors or reduced levels of one of the liver enriched factors combine with additional factors in the alternate tissue to activate transcription (Costa *et al.*, 1989). In this way, regulation is achieved in each tissues through different but overlapping subsets of regulatory molecules

All the factors that have been identified and described in this thesis represent the end stage of what is probably an autoregulatory, cascade of factors which are initiated by the determinative events in the early embryo. The process which dictates this regulatory network of transcription factors and controls factor production in a given cell type, however, remains unknown. The answers to this question can be obtained by working backwards as described in this thesis; asking both how these transcription factors interact to stimulate transcription and exploring the mechanism by which they themselves are modulated. In this way, the molecular mechanisms responsible for hepatocyte differentiation and development can be further explored.

References

Amara, S.G., V. Jones, M.G. Rosenfeld, E.S. Ong and R.M. Evans. 1982. Alternative RNA processingin calcitonin gene expression generates mRNAs encoding different polypeptide products. Nature **298**:240-244.

Babiss, L.E., P.B. Fisher and H.S. Ginsberg. 1984. Transformation of mutations in the early region 1B-encoded 21-and 55 kilodalton proteins of adenovirus type 5. J. Virol. **52**:389-395.

Babiss, L. E., A. Bennett, J. M. Friedman, and J. E. Darnell, Jr. 1986a. DNase I-correlation between chromatin structure and transcriptional activity. Proc. Natl. Acad. Sci. USA **83**:6504 -6508.

Babiss, L. E., J. M. Friedman, and J. E. Darnell, Jr. 1986b. Cellular promoters incorporated into the adenovirus genome: effects of viral regulatory elements on transcription rates and cell specificity of albumin and β-globin promoters. Mol. Cell. Biol. **6**:3798-3806.

Babiss, L. E., R. S. Herbst, A. L. Bennett and J. E. Darnell, Jr. 1987. Factors that interact with the rat albumin promoter are present both in hepatocytes and other cell types. Genes Dev. **1**:256 - 267.

Baldwin, A and P. Sharp. 1988. Two factors NF-κB and H2TF1 interact with a single regulatory sequence in the class I MHC promoter. Proc. Natl. Acad. Sci. USA. **85**:723-727.

Banerji, J., L. Olson, and W. Schaffner. 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell **33**:729-740.

Baumhueter, S., G. Courtois and G. R. Crabtree. 1988. A variant nuclear protein in dedifferentiated hepatoma cells binds to the same functional sequences in the β-fibrinogen gene promoter as HNF-l. EMBO J. **7**:2485 - 2493

Beato, R. 1989. Gene regulation by steroid hormones. Cell 56: 335-344.

Bennett, A.L., K.E. Paulson, R.E. Miller and J.E. Darnell, Jr. 1987. Acquisition of antigens characteristic of adult pericentral hepatocytes by differentiating fetal hepatoblasts in vitro. J. Cell Biol. **105**:1073-1085.

Bertolotti, R. 1977. A selective system for hepatoma cells producing gluconeogenic enzymes. Somatic Cell Genet. **3**:365-380.

Birkenmeier, E.H., B. Gwynn, S. Howard, J.Jerry, J.I. Gordon, W.H. Landschulz and S.L. McKnight. 1989. Tissue-specific expression, developmental regulation, and genetoc mapping of the gene encoding CCAAT/enhancer binding protein. Genes Dev. **3** :1146-1156.

Bodner, M. and M. Karin 1987. A pituitary specific trans-acting factor can stimulate transcription from the growth hormone promoter in extracts of nonexpressing cells. Cell **50**:267-270.

Boggs,R.T., P. Gregor, S. Idrisss, J.M. Belote, MM. McKeown. 1987. Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the transform gene. Cell 50:739-47.

Boulet,A.M., C.R. Erwin, and W.J. Rutter. 1986. Cell-specific enhancers in the rat exocrine pancreas. Proc. Natl. Acad. Sci. USA. **283**:3599-3603.

Brand, A.H., L. Breeden, J. Abraham, J. Sternglanz and K. Nasmyth. 1985. Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. Cell **41**:41-48.

Brand,A.H., G. Micklem and K. Nasmyth. 1987. A yeast silencer contains sequences that can promote autonomous plasmid replication and transcriptional activation. Cell **51**:709-719.

Briggs, M.R., J.T. Kaaadonaga, S.P. Bell and R. Tjian. 1986. Purification and biochemical characterization of the promoter-specific transcription factor Sp1. Science **234**:47-52.

Brown,D.D. 1984. The role of stable complexes that repress and activate eukaryotic genes. Cell **37**:359-365.

Buratowski, S., S. Hahn and P.A. Sharp and L. Guarente. 1988. Function of a yeast TATA element binding protein in a mammalian transcription system. Nature **334**:37-42.

Camper, S.A. and S.M. Tilghman. 1989. Postnatal repression of the a-fetoprotein gene is enhancer independent. Genes Dev. **3**:537-546.

Cassio, D., M. C. Weiss, M.-O. Ott, J. M. Salat-Trepat, and T. Erdos. 198l. Expression of the albumin gene in rat hepatoma cells and their dedifferentiated variants. Cell **27**:35l-358.

Cavallini, B., J. Huet, J. Plassat, A. Sentenac, J. Egly and P. Chambon. 1988. A yeast activity can substitute for the Hela cell TATA box factor. Nature **334**: 77-80.

Ceder, H. 1988. DNA methylation and gene activity. 1988. Cell **53**:3-4.

Cereghini, S., M. Raymondjean, A. G. Carranca, P. Herbomel, and M. Yaniv. 1987. Factors involved in control of tissue-specific expression of albumin gene. Cell **50**:627- 638.

Cereghini, S., M. Blumenfeld, and M. Yaniv. 1988. A liver-specific factor essential for albumin transcription differs between differentiated and dedifferentiated rat hepatoma cells. Genes Dev. **2**:957-974.

Chodosh, L.A., T.W. Carthew and P.A. Sharp. 1986. A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. Mol. Cell. Biol. 6:4723-4733

Christy, R.J., V.W. Yang, J.M. Ntambi, D.E. Geiman, W.H. Landshulz, A.D. Friedman, Y. Nakabeppu and T.J. Kelly. 1989. Differentiation-induced gene expression in 3T3-L1 preadipocytes: CCAAT/enhancer binding protein interacts with and activates the promoters of two adipocyte-specific genes. Genes Dev. **3**:1323-1335.

Ciliberto, G., L. Dente, and R. Cortese. 1985. Cell-specific expression of a transfected human α1-antitrypsin gene. Cell **41**:531-540.

Cistanzom, F., L. Castagnoli, L. Dente, P. Arcari, M. Smith, P. Costanzo, G. Raugei, P. Izzo, T. C. Pietropaolo, L. Bougueleret, F. Cimino, F. Salvatore and R. Cortese. 1983. Cloning of several cDNA segments coding for human liver proteins. EMBO J. **2**: 57-62.

Clayton , D.F. and J.E. Darnell, Jr. 1983. Changes in liver-specific compared to common gene transcription during primary culture of mouse hepatocytes. Mol. Cell. Biol. **3**:1552-1561.

Clayton, D. F., M. Weiss, and J. E. Darnell, Jr. 1985a. Liver-specific RNA metabolism in hepatoma cells: variations in transcription rates and mRNA levels. Mol. Cell. Biol. **5**:2633 -2641.

Clayton, D. F., A. L. Harrelson, and J. E. Darnell, Jr. 1985b. Dependence of liver-specific transcription on tissue organization. Mol. Cell Biol. **5**:2623-2632.

Cleveland, D.W., M.W. Lopata, R.J. MacDonald, N.J. Cowan, W.J. Rutter and M.W. Kirschner. 1980. Number and evolutionary conservation of alpha- and beta-tubulin and cytoplasmic beta- and gamma-actin genes using specific cloned cDNA probes. Cell **20**:95-105.

Colantuoni, V., A. Pirozzi, C. Blance and R. Cortese. 1987. Negative control of liver-specific gene expression: cloned human retinol-binding protein gene is repressed in HeLa cells. EMBO J. **6**:631 - 636.

Costa, R. H., E. Lai, and J. E. Darnell, Jr. 1986. Transcriptional control of the mouse prealbumin (transthyretin) gene: Both promoter sequences and a distinct enhancer are cell specific. Mol. Cell. Biol. **6**:4697-4708.

Costa, R. H., E. Lai, D. R. Grayson, and J. E. Darnell, Jr. 1988a. The cell-specific enhancer of the mouse transthyretin (prealbumin) gene binds a common factor at one site and a liver-specific factor(s) at two other sites. Mol. Cell. Biol. **8**:8l-90.

Costa, R. H., D. R. Grayson, K. G. Xanthopoulos, and J. E. Darnell, l988b. A liver-specific DNA-binding protein recognizes multiple nucleotide sites in regulatory regions of transthyretin, α1-antitrypsin, albumin and simian virus 40 genes.Proc. Natl. Acad. Sci. USA. **85**:3840 - 3844.

Costa, R. H., D. R. Grayson, and J. E. Darnell, Jr. 1989. Multiple hepatocyte-specfic nuclear factors function in the regulation of the transthyretin and α1-antitrypsin genes. Mol. Cell. Biol. **9**:1415-1425.

Courtois, G., J. G. Morgau, L. A. Campbell, G. Fourel, and G. R. Crabtree. 1987. Interaction of a liver-specific nuclear factor with the fibrinogen and α1-antitrypsin promoters. Science **238**:688-692.

Courtois, G., S. Baumhueter and G.R. Crabtree. 1988. Purified hepatocyte nuclear factor 1 interacts with a family of hepatocyte-specific promoters. Proc. Natl. Acad. Sci. USA. **85**:7937-7941.

Darnell, J. E., Jr. l982. Variety in the level of gene control in eukaryotic cells. Nature **297**: 365 - 371.

Darnell,J.E., Jr. 1985. Considerations on the basis for tissue-specificity. Proceedings of The Robert A. Welch Foundation COmferences on Chemical Research XXIX. Genetic chemistry: The Molecular Basis of Heredity. November 4-6 pp. 301-307.

Darnell, J.E., Jr. 1989. Thoughts about gene regulation and mammalian development. Arzneim.-Forsch./Drug Res. **39**:1011-1014.

Davison, B.L., J.-M. Egly, E.R. Mulvihill and P. Chambon. 1983. Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. Nature **301**:680-686.

Davis, R.L., H. Weintraub and A.B. Lassar. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell **51**:987-1000.

Derman, E., K. Krauter, L. Walling, C. Weinberger, M. Ray and J. E. Darnell, Jr. 1981. Transcriptional control in the production of liver- specific mRNAs. Cell **23**:73l-739.

Deschatrette, M., and M. C. Weiss. 1974. Characterization of differentiated and dedifferentiated clones from a rat hepatoma. Biochimie **56**:l603-1611.

Deschatrette, J., E. M. Moore, M. Dubois, and M. C. Weiss. 1980. Dedifferentiated variants of a rat hepatoma: reversion analysis. Cell **19**:1043-1051.

DeSimone, V., G. Ciliberto, E. Hardon, G. Paonessa, F. Palla, L. Lundberg, and R. Cortese. l987. *Cis-* and *trans*-acting elements responsible for cell-specific expression of the human alpha-l-antitrypsin gene. EMBO J. **6**:2759-2766.

Desplan, C., J. Theis and P.H. O'farrel. 1988. The sequence specificity of homeodomain-DNA interactions. Cell **54**:1081-1090.

Dickson, P.W., G.S. Howlett and G. Schreiber. 1985. Rat transthyretin (prealbumin) molecular cloning, nucleotide sequence and gene expression in the liver and brain. J. Biol. Chem. **260**:8214-8219.

Dignam, J. D., R. M. Lebowitz, and R. G. Roeder. l983. Accurate transcription initiation byRNA polymerase II in a soluble extract from isolated mammalian nuclei. Nuc. Acids Res. **11** :l475-l489.

Doljanski, F. 1960. The growth of the liver with special reference to mammals. Int. Rev Cytol. **10**:217-241.

D'Onofrio, C., V. Colantuoni, and R. Cortese. l985. Structure and cell-specific expression of a cloned human retinol binding protein gene: the 5'-flanking region contains hepatoma specific transcriptional signals. EMBO J. **4**: l98l- 1989.

Dorn, A.J., J. Bollekens, J. Staub, C. Benoist and D. Mathis. 1987. A multiplicity of CAAT box binding proteins. Cell **50**: 863 -866.

Dynan, W.S. 1989. Modularity in promoters and enhancers. Cell **58**:1-4.

Edlund, T., M.D. Walker, P. J.Barr, and W. J. Rutter. l985. Cell-specific expression of the rat insulin gene: evidence for role of two distinct 5' flanking elements. Science **230**:912-916.

Edmondsen, D.G. and E.N. Olson. 1989. A gene with homology to the myc sumularity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle dfferentiation program. Genes Dev. **3**:628-640.

Elias, H., and J.C. Sherrick. 1969. Morphology of the Liver. Academic Press, New York. 390 pages.

Feuerman, M.H., R. Godbout, R.S. Ingram and S.M. Tilghman. 1989. Tissue-specific transcription of the mouse a-fetoprotein gene promoter is dependent on HNF1. Mol. Cell. Biol. **9**: 4204-4212.

Fire,A., M. Samuels and P.A. Sharp. 1984. Interactions between RNA polymerase II factors, and templates leading to accurate transcription. J. Biol. Chem. **259**:2509-2516.

Fletcher, C., N.Heintz, and R.G. Roeder. 1987. Purification and characterizationm of OTF-1, a transcription factor regulating cell cycle expression of the human histone H2b gene. Cell **51**: 773-781.

Fougere, C. and M.C. Weiss. 1978. Phenotypic exclusion in mouse melanoma-rat hepatoma hybrid cells: pigment and albumin production are not reexpressed simultaneously. Cell **15**:843-854.

Frain, M. G. Swart, P. Monaci, A. Nicosia, S. Stampfli, R. Frank and R. Cortese 1989 The liver-specific transcription factor LF-B1 contains a highly diverged homeobox DNA binding domain. Cell **59**:145-157.

Fried, M., and D. M. Crothers. 1981. Equilibria and kinetics of Lac repressor operator interactions by polyacrylamide gel electrophoresis. Nuc. Acids Res. **9**:6505-6525.

Friedman, J. M., L. E. Babiss, D. F. Clayton, and J. E. Darnell, Jr. 1986. Cellular promoters incorporated into the adenovirus genome: Cell specificity of albumin and immunoglobulin expression. Mol. Cell. Biol. **6**:3791-3797.

Friedman, J. M., L. E. Babiss, M. Weiss, and J. E. Darnell, Jr. 1987. Hepatoma variants (C2) are defective for transcriptional and post-transcriptional actions form both endogenous and viral genomes. The EMBO J. **6**:1727-1731.

Friedman, A.D., W.H. Landschulz and S.L. McKnight. 1989. CCAAT/enhancer binding protein activates the promoter of the serum albumin gene in cultured hepatoma cells. Genes Dev. **3**:1314-1322.

Fromm, M., and P. Berg. 1982. Deletion mapping of DNA regions required for SV40 and early region promoter function in vivo. J. Mol. App. Genet. **1**:457-481

Gal, A., J.-L. Nahon, M. Gomez-Garcia, I. Tratner and J.M. Sala-Trepat. 1985. Organization of the albumin and alpha-fetoprotein genes in fetal and adult rat tissues and rat hepatomas. Differentiation **29**:238-245.

Galas, D., and A. Schmitz. 1978. DNase footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. **5**:3157-3170.

Gardner, R. L., and J. Rossant. 1976. Determination during embryogenesis. In Embryogenesis in Mammals. K. Elliott, and M. O'Connor, ed.Elsevier, New York. 5-25.

Garner, M.M. and A. Revzin. 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the E. coli. lactose operon regulatory system. Nucleic Acids Res. **9**:3047-3060.

Gehring, W.J. 1987. Homeo boxes in the study of development. Science **236**:1245-1252.

Gillies, S.D., S.L. Morrison, V.T. Oi and S. Tonegawa. 1983. A tissue specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain. Cell **33**:717-728.

Godbout, R., R. Ingram, S. M. Tilghman. 1986. Multiple regulatory elements in the intergenic region between the α-fetoprotein and albumin genes. Mol Cell. Biol. **6**:477-587.

Godbout, R., R. Ingram, and S. M. Tilghman. 1988a. Fine-structure mapping of the three mouse α-fetoprotein gene enhancers. Mol Cell. Biol. **8**:1169-1178.

Goodbourn, S., H. Burstein and T. Maniatis. 1986. The human β-interferon gene enhancer is under negative control. Cell **45**:601-610.

Gorin, M. B., D. L. Cooper, F. Eiferman, P. Van de Rijn, and S. M. Tilghman. 1981..The evolution of a-fetoprotein and albumin. I. A comparison of the primary amino acid sequences of mammalian a fetoprotein and albumin. J. Biol. Chem. **256**: 1954-1962.

Gorski, K., M. Carneiro, and U. Schibler. 1986. Tissue-specific *in vitro* transcription from the mouse albumin promoter. Cell **47**:767-776.

Gourdeau, H., T.C. Peterson, and R.E.K. Fournier. 1989. Differential activity of a tissue-specific extinguisher locus in hepatic and nonhepatic cells. Mol. Cell. Bio. **9**:1813-1822.

Goutte, C. and A.D. Johnson. 1988. a1 protein alters the DNA binding specificity of the α2 repressor. Cell **52**:875-882.

Graham, F.L., J. Smiley, W.C. Russell and R. Nairu. 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J. Gen. Virol. **36**:59-77.

Graves,B. J., P. F. Johnson, and S. L. McKnight. 1986. Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. Cell **44**: 565-576.

Grayson, D.R., R.H. Costa, K.G. Xanthopolos and J.E. Darnell, Jr. 1988a. A cell specific enhancer of the mouse α1-antitrypsin gene has multiple functional regions and corresponding protein binding sites. Mol. Cell. Biol. **8**:1055-1066.

Grayson, D. R., R. Costa, K. G. Xanthopoulos and J. E. Darnell. 1988b. One factor recognizes the liver-specific enhancers in α1-antitrypsin and transthyretin genes. Science **239**:786-788.

Green, M.R. 1989. Pre-mRNA processing and mRNA nuclear export. Current Opinion Cell Bio. **1**:519-525.

Greengard, O. 1969. Enzymatic differentiation in mammalian liver. Science: **163**:891-895.

Greengard, O. 1970. The developmental formation of enzymnes in rat liver. In Biochemical Actions of Hormones. Vol. I. G. Litwack, ed., Acad. Press, N. Y. 53-87.

Greengard, O., M. Federman, and W. E. Knox. 1972. Cytophotometry of developing rat liver and its application to enzymatic differentiation. J. Cell Biol. 52:261-262.

Grosschedl, R. and D. Baltimore. 1985. Cell-type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. Cell **41**:885-897.

Groudine, M., M. Peretz,, and H. Weintraub. 1981. Transcriptional regulation of hemoglobin switching in chicken embryos. Mol. Cell. Biol. **1**:281-288.

Guarente. L. 1988. UASs and Enhancers: Common mechanism of transcriptional activation in yeast and mammals. Cell **52**:303-305.

Hahn, S., S. Buratowski, P.A. Sharp and L. Guarente. 1989. Isolation of the gene encoding the yeast TATA binding protein TFIID: A gene identical to the SPT15 suppressor of Ty element insertions. Cell **58**:1173-1181.

Hai, T., F. Liu, E.A. Allegretto, M. Karin and M.R. Green 1988. A family of immunologically related transcription factors that included multiple forms of ATF and AP-1. Genes Dev. **2**:1216-1226.

Hai, T., F. Liu, E.A. Allegretto, M. Karin and M.R. Green. 1989. Transcription factor ATF cDNA clones: an extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers. Genes Dev. 3:2083-2090.

Ham, R.G. 1965. Clonal growth of somatic cells in a chemically defined medium. Proc. Natl. Acad. Sci. USA **53**:288-293.

Hammer, R. E., R. Krumlauf, S. A. Camper, R. L. Brinster, S. M. Tilghman. 1987. Diversity of α-fetoprotein gene expression in mice is generated by a combination of separate enhancer elements. Science: **235**:53-58.

Hardon, E. M., M. Frain, G. Paonessa and R. Cortesi. 1988. Two distinct factors interact with the promoter regions of several liver-specific genes. EMBO J. **7**: 17ll-1719.

Hart, R.P., M.A. McDevitt, H. Ali and J.R. Nevins. 1985. Definition of essential sequences and functional equivalence of elements downstream of the adenovirus E2A and early simian virus 40 polyadenylation sites. Mol. Cell. Biol. **5**:2975-2983.

Hay, N., J.M. Bishop and D. Levens. 1987. Regulatory elements that modulate expression of human c-myc. Genes Dev. **1**:659-71.

Hay, N., M. Takimoto and J.M. Bishop. 1989. A FOS protein is present in a complex that binds a negative regulator of MYC. Gene Dev. **3**:293-303.

Heard, M., P. Herbomel, M.-O. Ott, A. Mottura-Rollier, M. Weiss, and M. Yaniv. 1987. Determinants of rat albumin promoter tissue specificity analyzed by an improved transient expression system. Mol. Cell. Biol. **7**:2425 -2434.

Hearing, P. and T. Shenk. 1983. The adenovirus type 5 E1A transcriptional control region contains a duplicated enhancer element. Cell **33**:695-703.

Herbomel, P., A. Rollier, F. Tronche, M.O. Ott, M. Yaniv and M.C. Weiss. 1989. The rat albumin promoter is composed of six distinct positive elements within 130 nucleotides. Mol. Cell. Biol. **9**:4750-4758.

Herbst, R.S., N. Friedman, J.E. Darnell and L.E. Babiss. 1989. Positive and negative regulatory elements in the mouse albumin enhancer. Proc. Natl. Acad. Sci. USA **86**:1553-1557.

Herr, W., and J. Clarke. 1986. The SV40 enhancer is composed of multiple functional elements that can compensate for one another. Cell **45**:461 -470.

Herr, W., R.A. Sturm, R.G. Clerc, L.M. Corcoran, D. Baltimore, P.A. Sharp, H.A. Ingrahm, M.G. Rosenfeld, M. Finney, G. Ruvkin and H.R. Horvitz. The POU domain: A large conserved region in the mammalian pit 1, oct-1, oct-2 and Caenorhabditis elegans unc-86 gene products. Genes Dev. **2**:1513-1516.

Herskowitz, I. 1989. A regulatory hierarchy for cell specialization in yeast. Nature **342**:749-757.

Hoey, T., R. Warrior, J. Manak and M. Levine. 1988. DNA-binding activaties of the *Drosophila melanogaster* even-skipped protein are mediated by its homeo domain and influenced by protein context. Mol. Cell. Bio. **8**:4598-4607.

Hofer, E. and J.E. Darnell, Jr. 1981 The primary transcription unit of the mouse β-major globin gene. Cell **23**:585-593.

Horikoshi, M., T Hai, Y.-S. Lin, M. Green and R.G. Roeder. 1988. ATF interacts with the TATA factor to facilitate establishment of the preinitiation complex. Cell **54**:1033-1042.

Huo,L., K.J. Martin and R. Schleif. 1988. Alternative DNA loops regulate the arabinose operon in Escherichia coli. Proc. Natl. Acad. Sci. USA. **85**:5444-5448.

Imler, J.L., C. Lemaire, C. Wasylyk and B. Wasylyk. 1987. Negative regulation contributes to tissue specificity of the immunoglobulin heavy-chain enhancer. Mol. Cell. Bio. **7**: 2558-2567.

Ingraham, H.A., R. Chen, H.J. Mangalam, H.P. Elsholtz, S.E. Flynn, C.R. Lin, D.M. Simmons, L. Swanson and M.G. Rosenfeld. 1988. A tissue specific transcription factor containing a homeo domain specifies a pituitary phenotype. Cell **55**:519-529.

Jackson, S.P. and R. Tjian. 1988. O-glycosylation of eukaryotic transcription factors: Implications for mechanisms of transcriptional regulation. Cell **55**:125-133

Jagodzinski, L. L., T. D. Sargent, M. Yang, C. Glackin and J. Bonner. 1981. Sequence homology between RNAs encoding α-fetoprotein and rat serum albumin. Proc. Natl. Acad. Sci. USA **78**: 3521-3525.

Jefferson, D. M., D. F. Clayton, J. E. Darnell, Jr. , and L. M. Reid. 1984. Posttranscriptional modulation of gene expression in cultured rat hepatocytes. Mol. Cell Biol. **4**:1929-1934.

Johnson,P.F., W.H. Landschulz, B.J. Graves and S. McKnight. 1987. Identification of a rat liver protein that binds to the enhancer core element of three animal viruses. Genes Dev. **1**:133-146.

Johnson, S.A., J.M. Salmeron, Jr. and S.S. Dincher. 1987. Interaction of positive and negative regulatory proteins in the galactose regulon of yeast. Cell **50**: 143-1446.

Johnson, P.F. and S.L. McKnight 1989. Eukaryotic Transcriptional Regulatory Proteins. Annu. Rev Biochem **58**:799-839.

Jones, A. L. and E. Spring-Mills. 1983. The liver and gall bladder. pp. 707-748. In Weiss, L. (ed.), Histology: cell and tissue biology, 5th ed. Elsevier, New York, N.Y.

Jones, K. A., J.T. Kadonaga, P. J. Rosenfeld, T. J. Kelly, and R. Tjian. 1987. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell **48**:79-89.

Jones, N.C., P.W. J. Rigby and E. Ziff. 1988. *Trans*-acting protein factors and the regulation of eukaryotic transcription: lessons from studies on DNA tumor viruses. Genes Dev. **2**:267-281.

Jongstra,J., T.L. Reudelhumer, P. Oudet, C. Benoist, C.B. Chae, J.M. Jeltch, D.J. Mathis and P. Chambon. 1984. Induction of altered chromatin structures by simian virus 40 enhancer amnd promoter elements. Nature **307**:708-714

Jose-Estanyol, M. and J.L. Danan. 1988. A liver-specific factor and nuclear factor I bind to the rat a-fetoprotein promoter. J. Biol. Chem. **263**:10865-10871.

Jungermann, K. and N. Katz. 1982. Functional hepatocellular heterogeneity. Hepatology **2**:385-395.

Kadonaga, J. and R. Tjian. 1986. Affinity purification of sequence-specific DNA binding proteins. Proc. Natl. Acad. Sci. USA **83** :5889-5893.

Kafatos, F.C., C.S. Jones and A. Efstradiadis. 1979. Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. Nucleic Acids Res. **7**:1541-1551.

Kakkis, E., K.J. Riggs, W. Gillespie and K. Calame. 1989. A transcriptional repressor of c-myc. Nature **339**:718-721.

Kawakami,K., C. Schheiddereit and R.G. Roeder. 1988. Identification and purification of a human immunoglobulin enhancer binding protein (NFkB) that activatess transcription from a humaan immunodefficiency virus 1 (HIV-1) promoter in vitro. Proc. Natl. Acad. Sci. USA **85**:47000-4704.

Keleher, C.A., C. Goutte, and A.D. Johnson. 1988. The yeast cell-type specific repressor a2 acts cooperatively with a non cell-type-specific protein. Cell **53**:927-936.

Killary, A.and R. E. K. Fournier. 1984. A genetic analysis of extinction: Trans-dominant loci regulate expression of liver-specfic traits in hepatoma hybrid cells. Cell **38**:523-534.

Kioussis, D., F. Eifferman, P. vander Rijn, M. D. Gorin, R.S.. Ingram, and S. M. Tilghman. 1981. The evolution of a-fetoprotein and albumin II: the structures of the α–fetoprotein and albumin genes in the mouse. J. Biol. Chem. **256**: l960-1967.

Knowles, B., C. C. Howe, and D. P. Arden.1980. Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. Science **290**:497-499.

Konkel, D. A., S. M. Tilghman and P. Leder. 1978. The sequence of the chromosomal mouse β-globin major gene: homologies in capping, splicing and poly(A) sites. Cell **15**:1125-1132.

Ko, H.S., P. Fast, W. McBride and L.M. Staudt. 1988. A human protein specific for the immunoglobulin octamer DNA motif contains a functional homeo-box domain. Cell **55**:135-144.

Kovesdi, I., R. Reichel and J.R. Nevins. 1986. Identification of a cellular transcription factor involved in E1A trans-activation. Cell **45**:219-228.

Kovesdi, I., R. Reichel and J.R. Nevins. 1987. Role of an adenovirus E2 promoter binding factor in E1A-mediated coordinate gene control. Proc. Natl. Acad. Sci. USA. **84**:2180-2184.

Kuo, C. F., K. E. Paulson and J. E. Darnell, Jr. 1988. Positional and developmental regulation of glutamine synthetase expression in mouse liver. Mol Cell. Bio. **8**:4966-4971.

Landschulz, W. H., P. F. Johnson, E. Y. Adashi, B. J. Graves, and S. L. McKnight. 1988a. Isolation of a recombinant copy of the gene encoding C/EBP. Genes Dev. **2**:786-800.

Landschulz, W.H., P.F. Johnsom and S.L. McKnight. 1988b The leucine zipper: A hypothetical structure common to a new class of DNA-binding proteins. Science **240**:1759-1764.

Law, W. and A. Dugaiczyk. 1981. Homolohy between a primary structure of α-fetoprotein deduced from a cDNA sequence and serum albumin. Nature **291**:201-205

Lee, W., A. Hasslinger, M. Karin, and R. Tjian. 1987. Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. Nature **325**: 368-372.

Lefevre, C., M. Imagawa, S. Dana, J. Grindlay, M. Bodner and M. Karin. 1987. Tissue-specific expression of the human growth hormone gene is conferred in part by binding of a specific trans-acting factor. EMBO J. **6**:971-981.

Leff, T., K. Reue, A. Melian, H. Culver and J.L. Breslow. 1989. A regulatory element in the ApoCIII promoter that directs hepatic specific transcription binds to proteins in expressing and non-expressing cell types. J. Biol. Chem. **264**:16132-16137.

Legerski, R.J., L. Hodnett and H.B. Gray Jr. 1978. Extracellular nucleases of pseudomonas BAL 31: Use of the double strand deoxy-riboexonuclease activity as the basis of a convenient method for the mapping of fragments of DNA produced by cleavage with restriction enzymes. Nucleic Acids Res. **5**:1445-1464.

Lenardo, M., J. W. Pierce and D. Baltimore. 1987. Protein-binding sites in Ig gene enhancers determine transcriptional activity and inducibility. Science **236**:1573-1577.

Lenardo, M.J. and D. Baltimore, 1989. NFκB: A pleiotropic mediator of inducible and tissue specific gene control. Cell **58**:227-229.

Levine, M., and Manley, J. 1989. Transcriptional repression of Eukaryotic promoters. Cell **59**:405-408.

Lichtsteiner, S., J. Wuarin, and U. Schibler. 1987. The interplay of DNA-binding proteins on the promoter of the mouse albumin gene. Cell **51**: 963- 973.

Lichtsteiner, S. and U. Schibler. 1989. A glycosylated liver-specific transcription factor stimulates transcription of the albumin gene. Cell **57**:1179-1187.

Liu, J.-K., Y. Bergman, K. S. Zaret. 1988. The mouse albumin promoter and a distal upstream sites are simultaneously DNase I hypersensitive in liver chromatin and bind similar liver-abundant factors *in vitro*. Genes Dev. **2**:528-541.

Maire, P., J. Wuarin and U. Schibler. 1989. The role of cis-acting promoter elements in tissue-specific albumin gene expression. Science **244**:343-346.

Maniatis, T., Fritsch, E.F. and J. A. Fischer 1982. Molecular Cloning: a laboratory manuel. Cold spring Harbor Laboratory, Cold spring Harbor, N.Y.

Maniatis, T., S. Goodbourn and J.A. Fischer. 1987. Regulation of inducible and tissue-specific gene expression. Science **236**:1237-1244.

Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499-560.

Matsui, T., J. Segall, P.A. Weil and R.G. Roeder. 1980. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J. Biol. Chem. **255**:11992-11996.

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

Melton, D.A., P.A. Krug, M.R. Rebaglicti, T. Maniatis, K. Zinn and M.R. Green. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes for plasmids containing bacteriophage sp6 promoters. Nucleic Acids Res. **12**:7035-7056

Mermod, N., T.J. Williams, and T. Tjian. 1988. Enhancer binding factors Ap-4 and Ap-1 act in concert to activate SV40 late transcription in vitro. Nature **332**:557-561.

Mevel-Ninio, M. and M.C. Weiss. 1981. Immunofluorescence analysis of the time course of extinction, reexpression, and activation of albumin production in rat hepatoma-mouse fibroblast heterokaryons and hybrids. J. Cell Biol. **90**:339-350.

Mitchell, P.J., C. Wang and R. Tjian. 1987. Possitive and negative regulation of transcription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. Cell **50**:847-861.

Mitchell, P.J. and R. Tjian. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245**:371-378.

Monaci, P., A. Nicosin, R. Cortese. 1988. Two different liver-specific factors stimulate *in vitro* transcription from the human α1-antitrypsin promoter. EMBO J. **7**:2075 - 2087.

Morse, R.H. and R.T. Simpson. 1988. DNA in the nucleosome. Cell **54**:285-87.

Muglia, L., and L. B. Rothman-Denes. 1986. Cell type-specific negative regulatory element in the control region of the rat α-fetoprotein gene. Proc. Natl. Acad. Sci. USA **83**:7653 - 7657.

Nahon, J.-L. 1987. Developmental regulation of albumin and a-fetoprotein expression in the rat. Nuc. Acids Res. **12**:6751-6762.

Nagoshi, R.N., M. McKeown, K.C. Burtis, J.M., Belote, and B.S. Baker. 1988. The control of alternative splicing at genes regulation sexual differentiation in D. melanogaster. Cell **53**:229-36..

Nakazato, H. and M. Edmonds. 1974. Purification of messenger RNA and heterogeneous nuclear RNA containing poly (A) sequences. Methods Enzymol. **29**:431-443.

Nir, U., M.D. Walker and W.J. Rutter. 1986. Regulation of rat insulin 1 gene expression: Evidence for negative regulation in nonpancreatic cells. Proc. Natl. Acad. Sci. USA. **83**:3180-3184.

Ondek, B. L. Gloss and W. Herr. 1988.The SV40 enhancer contains two distinct levels of organization. Nature **333**:40-45.

Orlofsky, A. and L.A. Chasin. 1985. A domain of methylation change at the albumin locus in rat hepatoma cell variants. Mol. Cell. Biol. **5**:214-225.

Ott, M.O., L. Sperling, D. Cassio, J. Levilliers, J.M. Sala-Trepat and M.C. Weiss. 1982. Undermethylation at the 5' end of the albumin gene is necessary but not sufficient for albumin production by rat hepatoma cells in culture. Cell **30**:825-833.

Ott, M. O., L. Sperling, P. Herbomel, M. Yaniv, and M. Weiss. 1984. Tissue specific expression is conferred by a sequence from the 5' end of the rat albumin gene. EMBO J. **3**:2505 - 2510.

Palmiter, R.D., and R.L. Brinster. 1986. Germ-line transformation of mice. Annu. Rev. Genet. **20**:465-499.

Panduro, A., F. Shalaby, and D. A. Shafritz. 1987a. Changing patterns of transcriptional and post-transcriptional control of liver-specific gene expression during rat development. Genes Dev. l:ll72-1182.

Panduro, A., F. Shalaby and D. A. Shafritz. 1987b. Liver-specific gene expression in various pathophysiologic states. Hepatology **4**:10S-14S.

Petropoulos, C., S. G. Andrews, T. Tamaski, T., and N. Fausto. 1983. Alpha-fetoprotein and albumin mRNA levels in regenerating liver and carcinogenesis. J. Biol. Chem. **258**: 490l-5103.

Petit, C., J. Levilliers, M. O. Ott, M. C. Weiss. l986. Tissue-specific expression of the rat albumin gene: genetic control of its extinction in microcell hybrids. Proc. Natl. Acad. Sci. USA **83**:256l-2565.

Pinkert, C. A., D. M. Ornitz, R. L. Brinster, and R. D. Palmiter. 1987. An albumin enhancer located l0 kb upstream functions along with its promoter to direct efficient, liver-specific expression in transgenic mice. Genes Dev. **1**:268-276.

Powell, D. J., J. M. Friedman, A. J. Oulette, K. S. Krauter and J. E. Darnell, Jr. 1984. Transcriptional and post-transcriptional control of specific messenger RNAs in adult and embryonic live. J. Mol. Biol. **179**:2l-35.

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

Ptashne, M. 1988. How eukaryotic transcriptional activators work. Nature **335**:683-689.

Queen, C. and D. Baltimore. 1983. Immunoglobulin gene transcription is activated by downstream sequence elements. Cell **33**:741-748.

Rappaport, A. M. 1976. The microcirculatory acinar concept of normal and pathological hepatic structure. Beitr. Path. Bd. **157**: 215-243.

Rauscher III, F.J., P.J. Voulalas, D.R. Franza Jr. and T. Curran. 1988. Fos and Jun bind cooperatively to the AP-1 site: Reconstitution *in vitro*. Genes Dev. **2**:1687-1699.

Raychaudhuri, P., S. Bagchi and J.R. NEvins. 1989. DNA-binding activity of the adenovirus-induced E4F transcription factor is regulated by phosphorylation. Genes. Dev. **3**:620-627.

Raymondjean, M., S. Cereghini and M. Yaniv. 1988. Several CCAAT binding proteins coexist in eukaryotic cells. Proc. Natl. Acad. Sci. USA **85**:757-761.

Reid, L. and D.M. Jefferson. 1984. Cell culture studies using extracts of extracellular matrix to study growth and differentiation in mammalian cells, pp. 239-280. In Mather,J. (ed.), Mammalian cell culture. Plenum

Reinberg, D and R.G. Roeder. 1987. Factors involved in specific transcription by mammalian RNA polymerase II: purification and functional analysis of initiation factors IIB aabd IIE. J. Biol. Chem. **262**:3322-3330.

Reuber, M.D. 1961. A transplantable bile-secreting hepatocellular carcinola in the rat. J. Nat. Cancer Inst. **26**:891-899.

Reue, K., T. Leff and J.L. Breslow. 1988. Human apolipoprotein CIII is regulated by positive and negative cis-activating elements and tissue-specific protein factors. J. Biol. Chem. **263**:6857-1408.

Roche, J., C. Gorka, P. Goeltz, and J. J. Lawrence. 1985. Association of histone H1° with a gene repressed during liver development. Nature **314**:197-8.

Rosenfeld, P.J. and T.J. Kelly. 1986. Purification of NF-1 by DNA recognition site affinity chromatography. J. Biol. Chem. **261**:1398-1408.

Rothschild, M. A., M. Oratz, S. S. Schreiber. 1988. Serum albumin. Hepatology **8**:385-401.

Sanger, F., S. Nicklen and A.R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. USA **74**:5463-5467.

Sargent, T. D., J.-R. Wu , J. M. Sala-Trepat, R. B. Wallace, A. A. Reyes, and J. Bonner. 1979. The rat serum albumin gene: analysis of cloned sequences. Proc. Natl., Acad. Sci. USA **76**:3256-3260.

Sargent, T.D., L.L. Jagodzinski, M. Yang and J. Bonner. 1981. Fine structure and evolution of the rat serum albumin gene. Mol. Cell. Biol. **1**:871-883.

Sawadogo, M. and R.G. Roeder. 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell **43**:165-175.

Sawadogo, M., M.W. Van Dyke, P.D. Gregor and R.G. Roeder. 1988. Multiple forms of the human gene-specific transcription factor USF. J. Biol Chem. **263**: 11985-11993.

Scheidereit, C., C. Heguy and R.G. Roeder. 1987. Identification and purification of a human lymphoid-specific octamer-binding protein (OTF-2) that activates transcription of an immunoglobulin promoter *in vitro*. Cell **51**:783-793.

Scheidereit, C., J.A. Cromlish, T. Gerster, K. Kawakami, C.-G. Balmaceda, R.A. Currie, and R.G. Roeder. 1988. A human lymphoid-specific transcription factor that activates immunoglobulin genes is a homeobox protein. Nature **336**:551-557.

Scheuermann, R. and U.Chen. 1989. A developmental-specific factor binds to suppressor sites flanking the immunoglobulin heavy-chain enhancer. Genes Dev. **3**:1255-1266.

Schibler, U. 1989. Relative importance of cis-acting promoter elements in tissue specific albumin gene expression. In Regulation of Liver Gene Expression. Abstracts of papers presented at the May 3 - May 7 1989 meeting. Cold Spring Harbor Laboratory, New York.

Schleif, R. 1988. DNA looping Science **240:** 127-8.

Scott, M.P. and S.B. Carroll. 1987. The segmentation and homeotic gene network in early Drosophila development. Cell **51**:689-698.

Sen, R., and D. Baltimore. 1986. Multiple nuclear factors interact with the immunoglobulin enhancers. Cell **46**: 705-711.

Serfling,E., M. Jasin and W. Schaffner. 1985. Enhancers and eukaryotic gene transcription. Trends Genet. **1**:224-230.

Serfling, E. 1989. Autoregulation-a common property of eukaryotic transcription factors? Trends Genet. **5**:131-133.

Shapiro, D.J., P.A. Sharp, W.W. Wahli and M.J. Keller. 1988. A high-efficiency HeLa cell nuclear transcription extract. DNA **7**: 47-55.

Simon, M.C., T.M. Fisch, B.J. Benecke, J.R. Nevins and N. Heintz. 1988. Definition of multiple, functionally distinct TATA elements, one of which is a target in the hsp70 promoter for E1A regulation. Cell **52**:723-729.

Singh,H., R. Sen, D. Baltimore and P.A. Sharp. 1986. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature **323**:640-643.

Singh, H., J.H. LeBowitz, A.S. Baldwin and P.A. Sharp. 1988. Molecular cloning of an enhancer binding protein: Isolation by screening of an expression library witrh a recognition site DNA. Cell **52**:415-423.

Sorger,P.K., M.J. Lewis, H. Pelham. 1987. Heat shock factor is regulated differentially in yeast and HeLa cells. Nature 329:81-84.

Stalder, J.A., A. Larsen, J.D. Engel, M. Dolan, M. Groudine, and H. Weintraub. 1980. Tissue-specific DNA cleavages in the globin chromatin domain induced by DNase I. Cell **20**:451-460.

Staudt, L.M., H. Singh, R. Sen, T. Wirth, P.A. Sharp and D. Baltimore. 1986. A lymphoid specific protein binding to the octomer motif of immunoglobulin genes. Nature **323**:640-643.

Struhl,K. 1988. The Jun oncoprotein, a vertebrate transcription factor, activates transcription in yeast.. Nature **322**:649-651.

Thiel, J.F. and K.O. Smith. 1967. Fluorescent focus assay of viruses on cell monolayers in plastic petri dishes. Proc. Soc. Exp. Biol. Med. **125**:892-894.

Tilghman, S. M., and A. Belayew. 1982. Transcriptional control of the murine albumin/α -fetoprotein locus during development. Proc. Natl. Acad. Sci. USA **79**:5254-5257.

Tilghman, S. M. 1985. The structure and regulation of a-fetoprotein and albumin genes. In Oxford surveys in eukaryotic genes vol2 pp160-206. R. Dawkins and M. Ridley ed. Oxford university Press. Oxford 1985.

Treisman, R. 1986. Identification of a protein-binding site that mediates transcriptional response of the c-fos gene to serum factors. Cell **46**:567-574.

Triezenberg, S.J., R.C. Kingsbury, and S.L. McKnight. 1988. Functional analysis of VP16, the transactivator of herpes simplex virus immediate early gene expression. Genes Dev. **2**:718-724.

Tronche, F., A. Rollier, I. Bach, M.C. Weiss and M. Yaniv. 1989. The rat albumin promoter: cooperation with the upstream elements is required when binding of APF/HNF1 to the proximal element is partially impaired by mutation or bacterial methylation. Mol. Cell. Biol. **9**:4759-4766.

Urano, Y., K. Watanabe, M. Sakai, and T. Tamaoki. l986. The human albumin gene: Characterization of the 5' and 3' flanking regions and the polymorphic gene transcripts. J. Biol. Chem. **26l**:3244 -3251.

VanDyke, M.W., R. Roeder and M. Sawadogo 1988. Physical analysis of transcriptional preinitiation complex assembly on a class II gene promoter. Science **241**:1335-1338.

Walker, M. D. T. Edlund, A. M. Boulet, and W. J. Rutter. l983. Cell-specific expression controlled by the 5'-flanking region of insulin and chymotrypsin genes. Nature **306**:557- 56l.

Wasylyk, C. and B. Wasylyl. 1986. The immunoglobulin heavy- chain B-lymphocyte enhancer efficiently stimulates transcription in non-lymphoid cells. Embo J. **5**:553-560.

Weber, J., W. Jelinek and J.E. Darnell, Jr. 1977. The definition of a large viral transcription unit ladt in Ad2 infection of HeLa cells: mapping of nascent RNA in isolated nuclei. Cell **10**:611-616.

Webster, N., J.R. Jin, S. Green., M. Hollis and P. Chambon. 1988. The UAS_G is a traanscriptional enhancer in human HeLaa cells in the presence of the GAL4 trans-activator Cell **52**:169-178.
Weisbrod, S. l982.. Active chromatin. Nature **297**: 289-295.

Weiss, M. C., and M. Chaplain. 1971. Expression of differentiated functions in hepatoma cell hybrid: reappeaaarance of tyrosine aminotransferase inducibility after the loss of chromosomes. Proc. Natl. Acad. Sci. USA **68**:3026-3030.

Wirth, T., L. Staudt and D. Baltimore. 1987. An octamer oligonucleotide upstream of a TATA motif is sufficient for lymphoid-specific promoter activity. Nature **329**:174-178.

Woodworth, C., T. Secott, and H. C. Isom. 1986. Transformation of rat hepatocytres by transfection with simian virus 40 DNA to yield proliferating differentiated cells. Cancer Res. **46**:4018 - 4026.

Woodworth, C. D., and H. C. Isom. 1987. Regulation of albumin gene expression in a series of rat hepatocyte cell lines immortalized by simian virus 40 and maintained in chemically defined medium. Mol. Cell. Biol. **7**:3740 - 3748.

Workman, J.L. and R.G. Roeder. 1987. Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. Cell **51**:613-622.

Wu,C. The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase 1. Nature **286**:854-860.

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

Wu, C., S. WIlson, B. Walker, I Dawid, T. Paiisley, V. Zimarino aand H. Ueda. 1987. Purification aand properties of Drosophilllaa heat shock activator protein. Science **238**:1247-1253.

Yamamoto, K.R. 1985. Steroid receptor regulated transcription of specifici genes and gene networks. Annu. Rev. Genet. **19**:209-262.

Xanthopoulos, K.G., J. Mirkovitch, T. Decker, C.F. Kuo and J.E. Darnell Jr. 1989. Cell-specific transcriptional control of the mouse DNA-binding protein mC/EBP. Proc. Natl. Acad. Sci. USA **86**:4117-4121.

Zaret,K.S. and K.R. Yamamoto. 1984. Reversible and persistent changes in chromatin structure accompany activation of glucocorticoidd dependent enhncer element. Cell **38**: 29-39.

Zaret, K. S., C. M. DiPersio, D. A. Jackson, W. J. Montigny, and D. L. Weinstat. 1988. Conditional enhancement of liver-specific gene transcription. Proc. Nat'l. Acad. Sci. USA **85**: 9076 - 9080.

Zenke, M., T. Grundstrom, H. Matthes, M. Wintzerith, C. Schatz, A. Wildeman, and P. Chambon. 1986. Multiple sequence motifs are involved in SV40 enhancer function. EMBO J. **5**:387-397.

End