Title
Characterization and purification of proteins that bind to the mouse beta globin promoter

Author
Hartzog, Grant A

Publication Date
1992

Peer reviewed|Thesis/dissertation

CHARACTERIZATION AND PURIFICATION OF PROTEINS
THAT BIND TO THE MOUSE BETA GLOBIN PROMOTER

by

Grant A. Hartzog

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

[redacted]

. . . . . .

Date **University Librarian**

Degree Conferred: 9/6/92

Contribution by Grant Hartzog of the work described in this thesis

Almost all of the experimental work described in this thesis was performed by Grant Hartzog during the time that he was a graduate student in my laboratory. He designed and performed all the experiments concerning the *cis*-acting CACCC element and the proteins that bind to it, which comprise most of the thesis. The work he describes concerning the proteins that bind to the β-globin CAT box element was performed by both Grant and Alison Cowie, a research technician. Alison began the CAT element experiments and Grant took them over when she left; I estimate that they each contributed equally to the experiments for this part of the work. However, Grant contributed even more signficantly to the interpretation of the results concerning this element and wrote the entire section for this thesis.

Richard M. Myers

Richard M. Myers

CHARACTERIZATION AND PURIFICATION OF PROTEINS THAT BIND TO THE MOUSE BETA-GLOBIN PROMOTER

ABSTRACT

The β-like globin gene locus provides a model system for the study of developmental, inducible and tissue specific regulation of transcription. Additionally, the β-globin gene locus is a model system for the study of chromatin structure in transcriptional regulation. The cis elements that regulate β-like globin gene expression have been studied intensively, however the identities of the proteins that bind to these elements are largely unknown. This work examines the proteins that bind to two elements, CCAAT and CACCC, in the proximal promoter of the adult β-globin gene that are required for transcription.

The CACCC element is bound by at least four nuclear proteins in a gel mobility shift assay. To determine which of these proteins is a potential regulator of b-globin transcription, the ability of promoters bearing point mutations within the CACCC element was correlated with the ability of those promoters to activate transcription. This analysis identified two candidate regulators, Sp1 and an activity we call CACD. A binding site selection procedure was performed on partially pure CACD protein revealing that its high affinity binding site is the same as the β-globin CACCC element. CACD protein was further characterized and purified to reveal that it is a 44 kD protein

that binds DNA in a zinc dependent manner. Finally, CACD DNA-binding activity is inhibited by phosphatase, suggesting that it may be regulated by phosphorylation.

The CCAAT element appears to bind one protein complex in the gel mobility shift assay. This activity, βCP, was purified from nuclear extracts by three different schemes. Purified preparations of βCP contained multiple proteins. βCP activity could be recovered from preparative SDS-polyacrylamide gels only when all of these peptides were mixed together, suggesting that bCP is a multipartite DNA binding protein. A subset of βCP was found to be glycosylated and could be purified on wheat germ agglutinin agarose columns. Finally, βCP stimulated transcription in vitro in a CCAAT site dependent manner.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1.

Chapter 2.

Chapter 3.

Chapter 4.

CHAPTER 1

INTRODUCTION:
ISSUES IN THE STUDY OF TRANSCRIPTIONAL REGULATION AND THE REGULATION OF β-GLOBIN GENE TRANSCRIPTION

INTRODUCTION

The focus of my thesis work has been to characterize and purify proteins that are likely regulators of the mouse β–major globin gene (β^{maj}). The β-globin gene family is unique in the opportunities it presents for studies of transcriptional regulation. It is a model system for developmental, tissue specific and inducible transcription as well as for the study of long range interactions and chromatin structure in the regulation of transcription. Additionally, study of the β–globin family is attractive because of the wealth of phylogenetic data and both naturally occurring and laboratory produced mutations that are available.

My work on the β-globin genes has involved the biochemical characterization of proteins that bind the proximal promoter region of the adult mouse β-globin gene. Although the proximal promoters of the β-globin genes do not have the intrinsic capability of directing either tissue specific or developmentally specific transcription, they are necessary for the transcriptional regulation of their linked genes. Additionally, these regions have been functionally implicated in the transmission of tissue and developmental regulatory signals that originate in other regulatory sequences of the β–globin locus. Therefore, understanding what proteins bind the proximal promoter region is a necessary first step in understanding how an apparently constitutive promoter is converted into one that is regulated in a tissue and developmentally specific manner.

EARLY STUDIES OF TRANSCRIPTIONAL REGULATION

In the 1950's, studies of the regulation of lysogeny in lambda phage and induction of the lactose operon in E. coli led Jacob and Monod to argue that the expression of structural genes was regulated by two other genetic elements, an operator and a regulator gene (Jacob and Monod, 1961). Genetic data showed that the operator occurred very near the genes it regulated, was required for the actions of the regulator gene and could only operate in cis to the genes it regulated. The regulator gene was hypothesized to code for or regulate the actions of a cytoplasmic repressor. The repressor would regulate specific target genes in trans via the intermediacy of the operator. By interacting with metabolites in the cell that activated or inactivated it, the repressor could be used to induce or repress gene expression in response to environmental stimuli.

Subsequently, the operator was shown to be a DNA sequence that served as a binding site for a cytoplasmic protein-the repressor. These studies provided an important paradigm used to understand transcriptional regulation today; the sequence of a particular promoter serves to direct those proteins that regulate it to its vicinity. It is the interactions of the regulator(s) with RNA polymerase (RNAP), facilitated by sequence specific binding of the regulator and RNAP to the promoter, that determines the rate of transcription.

TRANSCRIPTIONAL REGULATION IN PROKARYOTES

In bacteria, RNAP is an enzyme composed of 4 subunits that transiently associates with a fifth subunit, sigma, to create a complex known as the RNAP holoenzyme (Chamberlin, 1982; Helmann and Chamberlin, 1988). This complex is able to locate specific DNA sequences that contain information concerning the transcriptional regulation of downstream genes. Comparison of these DNA sequences, known as promoters, shows that many contain similar sequences at positions -10 and -35 with respect to the transcriptional initiation site at position +1 (McClure, 1985). These conserved sequences are required for promoter recognition by the holoenzyme complex. Genetic and biochemical studies have shown that the sigma subunit of the holoenzyme is required for specific promoter recognition (Helmann and Chamberlin, 1988). The discovery of alternative sigma factors that facilitate recognition of promoters with different upstream sequences led to the concept that sigma factors can be utilized to direct a general transcriptional response; for example, to heat shock and nitrogen limitation (Ishihama, 1988). Once transcription initiates, RNAP no longer binds a single DNA sequence and the sigma subunit is released (Helmann and Chamberlin, 1988; McClure, 1985) (see below).

In addition to the -10 and -35 sequences present in most promoters, individual promoters often contain other sequences that are required for appropriate regulation of their linked gene. These

are the operator sequences mentioned above.[1] They serve as DNA targets for site specific DNA binding proteins that can influence transcription of a nearby gene. Examples of such specific DNA-protein interactions include those of the lambda repressor with its operator sequences in the phage lambda genome and that of the Lac repressor with its operator in the Lac operon. In vitro transcription experiments using highly purified RNAP and regulatory proteins have demonstrated that no other intermediary proteins are required for these regulatory interactions (Hawley, et al., 1985). In addition to repressing transcriptional initiation, regulatory proteins may stimulate transcription, and in fact, the same regulatory proteins may carry out both positive and negative regulation of transcriptional initiation. This is the case for the lambda repressor (Figure 1A). When bound to operator sequences lying between two divergently transcribed promoters in the lambda genome, the repressor negatively regulates one promoter, P_R and positively regulates the other, P_{RM} (Ptashne, 1986).

The DNA-protein structure created by the DNA sequence of a particular promoter is dynamic. That is, promoters do not merely set a monotonous rate of transcription for a particular gene but rather, they allow the rate of transcription of the gene to be changed in response to changing environmental conditions or growth requirements. This may be accomplished by manipulating the abundance of regulatory proteins in the cell, their ability to bind DNA, or their ability to interact productively with the transcription

---

[1] The term operator is most commonly used to denote the DNA binding site of a negative regulator of transcription. I will use the term operator to denote the binding site for both negative and positive regulators of transcription.

apparatus once bound to the promoter. For example, the Lac repressor binds DNA and represses transcription. However, when lactose (or metabolites of lactose) are present, Lac repressor binds lactose and is allosterically converted to a form that no longer binds DNA. This allows transcription of the Lac operon to occur. Thus, lactose, by virtue of its ability to inactivate DNA binding of the Lac repressor, is an inducer of the Lac operon(Glass, 1982) . Other repressors behave in the opposite manner; they bind DNA only when bound to a small molecule. For example, when bound to a molecule of cAMP, the Catabolite Activator Protein (CAP) binds to its operator and activates transcription of the Lac operon (de Crombrugghe, et al., 1984)(Figure 1B). Similarly, DNA damaging agents activate a protease, RecA, that cleaves the lambda repressor. This functionally depletes the cell of the repressor and allows lytic growth of the lysogenic phage (Ptashne, 1986). Finally, NTRC is a protein that activates transcription of promoters recognized by the σ^{54}-RNAP holoenzyme complex (Kustu, et al., 1991)(Figure 1C). This class of promoters directs synthesis of genes that regulate the nitrogen balance of the cell. Although NTRC binds its operator constitutively, it must be phosphorylated by a second protein, NTRB, before it can activate transcription.

More complex regulatory systems in prokaryotes occur when several different operators occur in one promoter. By bringing several different regulators near a promoter via their operators, multiple environmental signals can be integrated at that promoter. As mentioned above, the Lac promoter, in addition to its operator for the Lac repressor, contains an operator for a positive activator of

transcription, CAP. Thus, by combining binding sites for two regulatory proteins, the Lac promoter is active only when both lactose (or its metabolites) and cAMP are present at the appropriate concentrations in the cell (Glass, 1982)(Glass, 1982; Figure 1B).

In addition to interacting with RNAP, regulatory factors can interact among themselves. These interactions take several forms. Regulators can bind DNA cooperatively or anti-cooperatively. For example, the lambda repressor binds to adjacent operator sequences $O_{R}1$ and $O_{R}2$ in a cooperative manner (Ptashne, 1986)(Figure 1A). Because these operators differ slightly in sequence and hence in their affinity for the lambda repressor, interaction of the repressor with the low affinity site, $O_{R}2$, depends on cooperative binding with a repressor molecule bound to an adjacent high affinity site, $O_{R}1$. Thus the ability of repressor to bind to the low affinity site is critically dependent on the presence of repressor bound to the adjacent high affinity site. Because this interaction is between two molecules of repressor, filling of site $O_{R}2$ is dependent upon the square of repressor concentration and is therefore more sensitive to changes in repressor concentration than if cooperative binding with site $O_{R}1$ did not occur. Conversely, binding to operator $O_{R}3$ is low affinity and non-cooperative. This site is only filled at very high repressor concentrations and is relatively insensitive to changes in repressor concentration.

These differing interactions are used to set up complex regulatory circuits (Ptashne, 1986). Site $O_{R}1$ interacts with repressor with high affinity to repress promoter P_{R}. At low and intermediate concentrations, repressor is also bound to site $O_{R}2$ but not to site $O_{R}3$

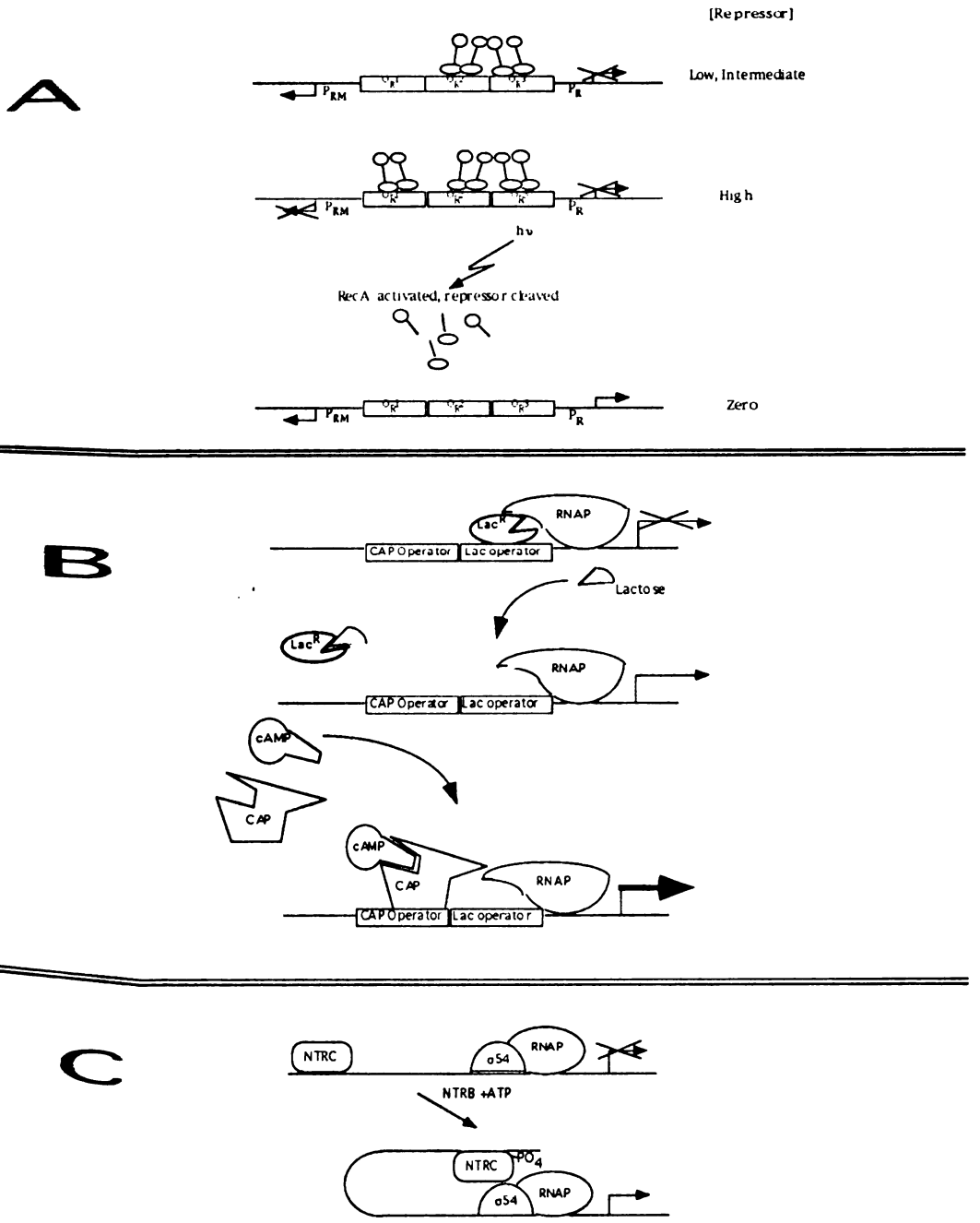
Figure 1. Transcriptional Regulation in Prokaryotes

A. Transcriptional regulation of the lambda phage promoters P_R and P_{RM}. At low and intermediate concentrations of repressor, sites O_R1 and O_R2 are occupied by cooperatively bound dimers of repressor. Site O_R3, which binds repressor non-cooperatively and with low affinity, is empty. The regulatory outcome of this state is that promoter P_R is sterically blocked by repressor and promoter P_{RM} is activated by repressor at operator O_R2. At high concentrations of repressor, site O_R3 becomes filled and repressor inhibits promoter P_{RM}. When inactivated by RecA, repressor vacates its operators and allows transcription of promoter P_R.

B. Transcriptional regulation of the Lac promoter. The Lac promoter is controlled by a positive factor, Catabolite Activator Protein (CAP), and a negative factor, the Lac repressor (Lac^R). Operators for these factors lie near the -10, -35 sequences for RNAP. In the absence of lactose (or it metabolites), the Lac repressor binds the Lac promoter and prevents RNAP from isomerizing to an open complex. In the presence of lactose, the Lac repressor is allosterically converted to a low affinity DNA-binding protein and vacates its operator, allowing transcription to procede. CAP can activate transcription when the Lac repressor is inactivated. For this

to occur, cAMP must be present to convert CAP into a high affinity DNA-binding protein.

C. Regulation of σ54 regulated promoters by NTRC. NTRC can bind DNA constitutively as can σ54-RNAP holoenzyme complexes. However, isomerization to the open complex, and therfore transcription, does not occur unless NTRC is phosphorylated by NTRB. This in turn, allows NTRC to interact with the σ54-RNAP complex to stimulate isomerization to the open complex.


A
[Repressor]
Low, Intermediate
High
hν
RecA activated, repressor cleaved
Zero
B
RNAP
CAP Operator
Lac operator
Lactose
cAMP
CAP
C
NTRC
σ54
NTRB +ATP
PO4

and transcription of P_{RM} is stimulated by the O_R2 bound repressor molecule. At higher concentrations, when it is bound to site O_R3, repressor inhibits transcription from P_{RM}. Thus, by allowing cooperative interactions at promoters, and by using operators with differing affinities for a regulator, the same molecule can differentially regulate different transcription units.

Regulators can also compete with each other to form alternate interactions that will create protein surfaces or structures that may stimulate or repress transcription. The arabinose operon is regulated both positively and negatively by the araC protein. In the absence of cAMP-CAP, araC proteins bound at several sites in the arabinose operon interact with one another to repress transcription. When cAMP-CAP is present however, these repressing interactions are disrupted and araC activates transcription. Thus, the activity of a DNA binding protein can be regulated by specifying alternative protein-protein interactions depending on what other proteins are bound at the promoter.

These types of interactions can occur between proteins bound to sites lying next to one another or to sites separated by tens or hundreds of nucleotides. When bound near but not directly adjacent to one another, DNA bound proteins may interact via bending of the intervening DNA (Ptashne, 1986; Schlief, 1992). However, this type of interaction may be constrained by the intrinsic flexibility of DNA (since a relatively large angle of bending may be required over a short stretch of DNA) and by the rotational phasing of the proteins on the DNA helix (Ptashne, 1986; Schlief, 1992). When separated by large distances, the distortion required of any particular segment of

DNA is slight and DNA bound proteins may interact by looping out the intervening DNA sequences (Schlief, 1992). Thus, not only the presence, but also the relative spacing and orientation of operator sequences in a promoter may determine its transcriptional characteristics.

Regulation Of Transcriptional Initiation

When considering regulation of transcription, it is probably true that any step that can be regulated will be regulated in some system. The functions that RNAP must carry out include site specific recognition of promoters, melting of the DNA helix to gain access to the nucleotides to be transcribed, transcriptional initiation, transcriptional elongation, renaturation of the DNA helix and termination of transcription (McClure, 1985). As discussed above, promoter recognition in prokaryotes is directed by transient association of RNAP with an accessory factor, σ. Promoter recognition can also be regulated by other factors. The lambda repressor, when bound to site O_R1, blocks access of polymerase to promoter P_R (Hawley, Johnson and McClure, 1985). Other regulatory proteins bind cooperatively with RNAP to a promoter, thereby increasing the rate of transcriptional initiation (McClure, 1985).

Once RNAP has bound a promoter, it must then melt the double stranded DNA helix to gain access to the nucleotide template (McClure, 1985). This transition, from the RNAP-double stranded DNA, "closed complex", to the RNAP-melted DNA, "open complex", is followed by initiation of transcription and finally by a transition of

RNAP from a site specific DNA binding mode to a non-specific DNA binding mode (McClure, 1985). This transition is accompanied by release of the sigma factor (Helmann and Chamberlin, 1988). These final steps of initiation are known to be influenced by a variety of regulatory molecules. For example both NTRC and lambda repressor function by stimulating the closed to open complex transition (Hawley, Johnson and McClure, 1985; Popham, et al., 1989)(Hawley, Johnson and McClure, 1985; Popham, et al., 1989; Figures 1A and 1C). The Lac repressor, on the other hand, appears to function by preventing isomerization to the open complex (Straney and Crothers, 1987)(Straney and Crothers, 1987; Figure 1B).

Conclusions

In summary, promoters are the DNA sequences that direct RNAP to the 5' end of a gene and determine what regulatory proteins may bind to DNA in the vicinity of the gene. The regulation of transcription may be dynamic in that the relative abundance, ability to bind DNA and activity of transcriptional regulatory proteins varies in response to the environment. More complex schemes of regulation can be produced in several ways. First, multiple operators can be put in one promoter. This allows for multiple environmental signals to be fed into the promoter. Second, when their operators are spaced appropriately, DNA binding proteins can form cooperative interactions. These sorts of interactions can result in promoter specific modifications of the DNA binding behavior of a regulatory protein. Third, promoter bound proteins interact with one another.

This allows for alternative regulatory functions of one protein depending on the presence of another. Finally, transcriptional initiation involves multiple steps, all of which are potentially regulated.

TRANSCRIPTION REGULATION IN EUKARYOTES

Because the basic functions that eukaryotic RNAP has to perform are the same as those performed by its prokaryotic cousin, our understanding of prokaryotic transcriptional control has provided a schema from which to consider eukaryotic transcription. The major differences between prokaryotic and eukaryotic transcription are fourfold. First and foremost, eukaryotic cells, by virtue of the complexity of structures, cell types and developmental changes that they must produce, integrate much more information and make many more types of transcriptional decisions than prokaryotes. Second, presumably because of the demands of making these many decisions, the eukaryotic transcriptional apparatus is immensely more complicated than that of prokaryotes. Third, eukaryotes segregate their DNA and hence, their transcription apparatus in the nucleus, a compartment distinct from the cytoplasm. Fourth, a true mechanistic difference between the control of prokaryotic and eukaryotic transcription may be found in the role which chromatin structure seems to play in eukaryotic transcriptional regulation.

Eukaryotic Gene Structure

Unlike prokaryotes, eukaryotes utilize three types of RNAP (Lewis and Burgess, 1982). RNAP I and RNAP III transcribe genes coding for structural RNAs and will not be considered here. Genes coding for proteins are transcribed RNAP II. Like prokaryotes, comparison of protein coding genes in eukaryotes showed that many contain an element whose sequence, orientation and position relative to the transcription start site is highly conserved (Corden, et al., 1980). This sequence, the TATA box, is required for high levels of transcriptional initiation and for accurate selection of the transcript start site (Benoist and Chambon, 1981; Concino, et al., 1984; Grosschedl and Birnstiel, 1980; Mathis and Chambon, 1981; McKnight and Kingsbury, 1982; Sawadogo and Sentenac, 1990; Waslyk, et al., 1980). Biochemical studies demonstrated that the TATA box is bound by TFIID, the factor that initiates assembly of RNAPII onto the promoter (Nakajima, et al., 1988; Pugh and Tjian, 1992; Sawadago and Roeder, 1985; Van Dyke, et al., 1989). Promoters that lack a TATA box often have a sequence at the transcription start site, known as the initiator element, that functionally substitutes for the TATA box (Smale and Baltimore, 1989). Recent studies have shown that the protein complex that recognizes the TATA box is also used for transcription of TATA-less promoters, suggesting that the mechanisms of transcriptional regulation of TATA containing and TATA-less promoters will be similar (Pugh and Tjian, 1991; Smale, et al., 1990; Wiley, et al., 1992). Like prokaryotic promoters, eukaryotic promoters contain other elements, varying from promoter to

promoter, that influence the rate of transcriptional initiation (Dynan, 1989). Overall, in their broad outlines, eukaryotic promoters seem to be organized like those of prokaryotes. A conserved sequence, the TATA box, serves to direct RNAPII to the promoter. Other, nearby sequences, serve to confer specific regulatory properties on the promoter by serving as regulatory protein binding sites (See Below).

In addition to sequences nearby the transcription start site, some eukaryotic promoters are regulated by sequences hundreds or thousands of nucleotides distant. These sequences, enhancers, were originally described as elements that increase the rate of transcriptional initiation independent of their orientation and position with respect to the transcription start site of a promoter (Atkinson, 1988; Banerji, et al., 1981). Enhancers have no intrinsic activity of their own, rather they require a linked promoter to influence transcription (Atkinson, 1988). When the nucleotide sequence of enhancers is examined, they are generally found to contain sequence motifs that also occur in proximal promoters (Atkinson, 1988). Thus, enhancers are thought to function by binding particularly potent or large numbers of transcriptional regulators that stimulate the basal transcription apparatus (Atkinson, 1988). Enhancers are not unique to eukaryotes. Although promoters using the most common prokaryotic sigma factor, σ^{70}, tend to interact with nearby regulatory sequences, σ^{54} utilizing promoters can interact with NTRC bound to sites thousands of base pairs distant, upstream or downstream from the gene (Reitzer and Magnasanik, 1986). In both prokaryotic and eukaryotic systems, in vitro experiments support the idea that enhancers work by looping out

intervening DNA to allow protein-protein interactions (Dunaway and Droge, 1989; Muller, et al., 1989; Wedel, et al., 1990). Finally, another class of cis regulatory sequences has been recently described, these sequences appear to regulate the chromatin structure surrounding a gene and will be discussed below.

Protein Components Of The Eukaryotic Transcription Apparatus

The eukaryotic transcription machinery can be divided into three different components. The first, RNAPII, is composed of 11 or more proteins ranging in molecular weight from 10 to 200 kDa (Young, 1991). Several of these proteins bear sequence similarity to bacterial RNAP (Young, 1991). The largest subunit of RNAPII contains a C-terminal domain containing 17-52 repeats of the sequence YSPTSPS (Young, 1991). Genetic and biochemical studies implicate this domain in functional interactions with upstream regulators of transcription, basal transcription factors and the chromatin template (Allison and Ingles, 1989; Koleske, et al., 1992; Peterson, et al., 1991; Usheva, et al., 1992).

The second component of the transcription machinery, variously termed the basal or general factors, includes a number of proteins and or chromatographic fractions named TFIIA, TFIIB, TFIID, TFIIE/F, as well as other, less well characterized factors (Flores, et al., 1992; Roeder, 1991; Sawadogo and Sentenac, 1990). With the possible exception of TFIIA, these factors are required for transcription in vitro of all RNAPII promoters (although see (Green, 1992)Green, 1992 for a different opinion). Identified by biochemical

fractionation and reconstitution of in vitro transcription reactions, these factors can be placed into an ordered assembly pathway (Buratowski, et al., 1989; Flores, et al., 1991; Flores, Lu and Reinberg, 1992; Roeder, 1991). TFIID first identifies and binds in a sequence specific manner to the TATA box. TFIIA and other, less well characterized factors may play roles in regulating or facilitating this process. Next, TFIIB binds to TFIID, this in turn is bound by a complex of TFIIF and RNAPII and finally, TFIIE binds to form the pre-initiation complex. Thus, the basal factors appear to carry out functions analogous to that of the σ factors, they help RNAP identify specific promoters. The preinitiation complex, formed only by the basal apparatus and RNAPII, is competent to specifically initiate transcription. A major role of upstream factors is to modify the rate at which this occurs.

An open question is if, like prokaryotic sigma factors, eukaryotes have alternative forms of the basal factors that direct RNAPII to particular classes of promoters. There are several pieces of evidence that suggest that this may be the case. First, several subtly different sequence variants of the TATA box exist, these vary in their ability to respond to specific upstream activators (Simon, et al., 1988; Taylor and Kingston, 1990). Second, TFIIA and several other factors interact with TFIID and may influence its DNA-binding properties or its ability to interact with other basal factors and thereby regulate promoter recognition by RNAPII (Meisterernst and Roeder, 1991; Meisterernst, et al., 1991; Roeder, 1991). Finally, in addition to the factors described above, there appear to be other factors, variously termed co-activators or adaptors, that are required

for transcriptional regulators to exert their effects (Kelleher, et al., 1990; Martin, et al., 1990; Pugh and Tjian, 1990; Tanese, et al., 1991). There appear to be co-activators specific for each type of regulatory domain (Martin, Lillie and Green, 1990; Tanese, Pugh and Tjian, 1991)(See below). Some of these ill defined factors are associated with the DNA binding component of TFIID, others may associate with the transcription apparatus in a more transient manner. By building complexes with different coactivators, promoters could be made sensitive or insensitive to different classes of upstream regulatory proteins.

The third component of the eukaryotic transcription apparatus consists of the site specific DNA binding proteins, and proteins interacting with them, that regulate transcription from promoter specific DNA binding sites. In contrast to prokaryotes, where repressor proteins are common, the majority of eukaryotic regulatory proteins that have so far been described are activators. Eukaryotic transcriptional regulatory proteins are remarkable in two aspects, their functional conservation across species and their modular construction. Numerous experiments have shown that transcriptional activators from one species can often function in a variety of evolutionarily divergent species. Extreme examples of this include the demonstration that the rat glucocorticoid receptor will activate transcription in yeast and that the yeast Gal4 protein will activate transcription in mammalian cells (Kakidani and Ptashne, 1988; Schena and Yamamoto, 1988). These data argue that the mechanisms of transcriptional activation are largely conserved within the eukaryotes.

A second interesting aspect of eukaryotic transcriptional regulators is their modular construction. The various functions of a transcriptional regulator are often separated into different domains in the protein. These functions include DNA binding, transcriptional activation, ligand binding, nuclear localization, and dimerization (Lamb and McKnight, 1991; Ma and Ptashne, 1987; Picard, et al., 1988; Picard and Yamamoto, 1987). Although some functions are required to observe others, for example, many proteins must dimerize to bind DNA, this modularity allows creation of chimeric proteins (Lamb and McKnight, 1991). For example the activation domain of one protein is often fused to the DNA binding domain of another to create an artificial activator protein of the desired sequence specificity (Brent and Ptashne, 1985). Not only is this modularity experimentally useful, but it is also suggestive of potential mechanisms by which complex regulators and regulatory systems may have evolved.

Like prokaryotes, eukaryotes build complex regulatory circuits in their promoters by combining binding sites for multiple regulatory proteins (Dynan, 1989). Eukaryotic transcriptional regulation is also dynamic, the abundance, DNA binding activity and ability to influence RNAPII are commonly regulated aspects of these proteins. Because eukaryotic cells segregate their DNA in a nucleus, the import of regulators into the nucleus can modulated to regulate their ability to reach their target sequences. This is the case for the transcription factor NF-kB as well as for a protein kinase, c-Raf, that may regulate transcription factor activity (Blank, et al., 1992; Morrison, et al., 1988). Eukaryotic regulators often include subunits that do not

touch DNA. These include proteins that have transcription regulatory functions similar to those found in other DNA binding proteins, proteins that anchor the DNA binding protein in the cytoplasm until an appropriate signal is received and potentially, proteins with enzymatic activities (Blank, Kourilsky and Israel, 1992; Nevins, 1992). Other common mechanisms of regulating factors include using phosphorylation and ligand binding to allosterically regulate function (Larson, et al., 1988; Pabo and Sauer, 1992; Yamamoto, et al., 1988).

Protein-protein interactions are of paramount importance in eukaryotic transcription. Like prokaryotes, these interactions include cooperative binding and they may critically depend on the relative spacing and orientation of protein binding sites. Additionally, eukaryotes make extensive use of protein dimerization to create a diversity of regulatory structures (Lamb and McKnight, 1991). Often, these factors belong to protein families that are highly related in their DNA binding and dimerization domains. Thus, these factors may form a variety of hetero or homo dimers. These dimers may vary in their affinity and specificity for DNA binding sites, ability to interact with other factors and the functions of RNAPII that they affect.

Finally, in vitro studies have shown that some DNA sites are recognized by multiple DNA binding proteins. These proteins may be different members of a single family of DNA binding protein or may come from different families. For example the homeodomain proteins have overlapping sequence specificities (Hayashi and Scott, 1990). Also, C/EBP, a member of the bZip family, as well as CTF and

CP1, members of two other ill-defined families of DNA binding proteins, can all recognize variants of the sequence CCAAT (Johnson and McKnight, 1989). The mechanisms that determine which of several different factors bind a DNA site are not clear, but probably include subtle differences in their affinity for the site and differential abilities to recognize and interact with other proteins bound nearby. This problem is discussed in greater detail in the second chapter of this thesis.

Classes Of Transcriptional Regulatory Protein Structures

In the past few years, many transcriptional regulators have been cloned and sequenced. Comparison of their predicted primary structure shows that many of these proteins belong to large families of related factors. Often, the similarities between family members are limited to their DNA binding and dimerization domains. Comparison of these regions of homology has been instructive in identifying important structural elements (Pabo and Sauer, 1992). This section will discuss some of the better understood families of DNA binding domains.

Zinc Fingers The zinc finger proteins fall into at least three different structural families. These proteins share a common motif that lends them their name; all types of zinc finger proteins chelate one or more zinc ions, with histidines and or cysteines, to create compact DNA binding domains. The presence of the zinc ion is required for DNA binding activity and probably for correct folding of

the DNA binding domain. Three different types of zinc finger protein have been well characterized (Harrison, 1991).

The zinc finger proteins typified by Gal4 comprise a small family of proteins that so far, have only been observed in fungi (Harrison, 1991). Gal 4 binds DNA as a dimer to a dyad symmetric, 17 base pair sequence (Harrison, 1991). The amino terminal 65 amino acids of Gal4 are sufficient for specific DNA binding although additional dimerization functions are found in amino acids 65-94 (Marmorstein, et al., 1992). A monomer of Gal4 binds two molecules of Zn^{2+} using an array of 6 cysteines (Marmorstein, et al., 1992). Each zinc ion is chelated by four cysteines; two of the cysteines make contacts with both ions. The crystal structure of Gal 4 shows that a 30 amino acid protein module, formed around two closely spaced Zn^{2+} ions, constitutes the DNA binding portion of the protein (Marmorstein, et al., 1992). An alpha helix in this module is positioned so that it sticks into the major groove of the DNA and makes several base specific contacts as well as several phosphate contacts (non-specific). The metal binding modules lie at and bind to the 3 nucleotides at the ends of the recognition sequence. In between these modules lies an extended protein linker that makes extensive phosphate contacts with the DNA and ends in a short alpha helical domain over the center of the recognition sequence. This domain is responsible for dimerization with the symmetrically placed monomer of Gal4. The extended structure of the Gal 4 DNA binding domain has led to the suggestion that some other protein may have access to the DNA between the two metal binding modules (Marmorstein, et al., 1992).

The steroid hormone receptor family of zinc finger proteins are ligand regulated DNA binding proteins that bind to dyad symmetric sites as dimers (Lucas and Granner, 1992). Their DNA binding domain is about 70 amino acids in length and binds two Zn^{2+} ions, each with four cysteines (Harrison, 1991). The crystal structure of the glucocorticoid receptor shows that each zinc ion is chelated by two cysteines lying within a protein loop and by two cysteines lying within an alpha helix (Luisi, et al., 1991). The loop and helix pack against one another to form a compact loop-helix domain. The DNA binding domain is composed of two loop-helix structures that form a globular DNA binding domain. The helix of the N-terminal loop-helix fits into the major groove of the recognition site and makes base-specific contacts. The second loop-helix is thought to help position the first loop-helix into its binding site. Residues in the second loop-helix domain are responsible for dimerization contacts with the symmetrically placed monomer of the glucocorticoid receptor.

The Cys_2-His_2 family of zinc finger proteins, typefied by Sp1 and TFIIIA, generally contain three or more directly repeated sequences elements conforming to the sequence; O X C X_{2-4} C X_3 O X_5 O X_2 H X_{3-4} H X_{2-6} where X is any amino acid and O is a hydrophobic amino acid (Harrison, 1991). These domains are often linked by a short conserved sequence, HTGEKP (Pabo and Sauer, 1992). The crystal structure of the Cys_2-His_2 protein Zif 268 shows that each repeated finger domain chelates one zinc ion via the invariant cysteines and histidines (Meisterernst and Roeder, 1991). Each finger folds into a compact domain consisting of an anti-parallel β sheet containing the two cysteines followed by an alpha helix

containing the two histidines. The alpha helix of each finger lies in the major groove and makes specific contacts with three nucleotides. The subsequent finger contacts the next three bases without a gap in the recognition sequence. The recognition sequences of the Cys_2-His_2 finger proteins tend to be G rich (or C rich on the other strand). The crystal structure shows that the predominant protein contacts are to the G rich strand and that the amino to carboxy order of the fingers on this DNA strand is 3' to 5'. Genetic studies suggest that this orientation is a general feature of Cys_2-His_2 proteins (Nardelli, et al., 1991).

bZip proteins The basic region-leucine zipper or bZip proteins make up a diverse family of proteins remarkable for their abilities to heterodimerize with one another (Lamb and McKnight, 1991; Pabo and Sauer, 1992). The bZip proteins are functionally divided into a basic DNA binding domain of ill-defined structure and a dimerization domain-the leucine zipper. The leucine zipper is an approximately 30 amino acid element composed of heptad repeats of non-polar amino acids at every first and fourth amino acid. The fourth amino acid of each repeat is almost always leucine, hence the name. Two zipper domains pair in a parallel, coiled-coil to form a protein dimer. The bZip proteins vary greatly in their dimerization properties. Some do not form homodimers and many are selective in the proteins with which they will heterodimerize (Lamb and McKnight, 1991). These properties allow for many potential differential regulatory complexes to be formed. Both the zipper and basic region of the bZip proteins become strongly alpha helical in character upon dimerization. The recognition sites of the bZIP proteins are dyad

symmetric sequences 9-10 basepairs in length. These data have been used to argue that the bZip proteins form a "scissors grip" over its DNA recognition site with the alpha-helical basic domains extending down from the coiled-coil dimerization domain into the major groove, each making base specific contacts with one half-site of the recognition sequence (Vinson, et al., 1989).

Homeodomain proteins The homeodomain motif is found prominently among factors implicated in developmental control of organisms from yeast to man. Homeodomains are approximately 60 amino acids long and usually bind DNA as monomers (Pabo and Sauer, 1992). Crystal structures of the yeast α2 protein and the Drosophila engrailed protein show that in addition to their primary sequence, the tertiary structure of homeodomain proteins is highly conserved (Kissinger, et al., 1990; Wolberger, et al., 1991). These proteins use a variation of the helix-turn-helix structure, common among bacterial repressors, to form their DNA binding domain (Pabo and Sauer, 1992). One helix is arranged to lie in the major groove of the DNA and the other two helices lie on top of the first. Interestingly, the DNA recognition helix in this complex also makes base specific contacts in the minor groove of the DNA. Homeodomain proteins tend to bind DNA with low specificity (Inostroza, et al., 1991). This raises the problem of how homeodomain proteins discriminate their binding sites in the genome from related sites bound by other homeodomain proteins. Some homeodomain proteins are known to increase the specificity of their interaction with DNA by making cooperative interactions with other proteins. For example the α 2 protein binds cooperatively with either of two other proteins,

MCM1 or a1 (Goutte and Johnson, 1988; Keleher, et al., 1988)(Goutte and Johnson, 1988; Keleher, et al., 1988; Figure 2).

The POU-domain proteins comprise a subclass of homeodomain proteins (Rosenfeld, 1991). Comparison of the cDNAs for the factors Pit-1, Oct-1, Oct-2 and unc-86 led to the observation of a 150 amino acid domain of similarity between these proteins (Herr, et al., 1988). Subsequently, many other POU domain proteins have been described. The POU domain consists of the ~60 amino acid homeodomain separated by 14-25 non-conserved amino acids from a POU specific domain of 70-80 amino acids (Herr, et al., 1988). The POU specific domain of these proteins contributes to the affinity and specificity of DNA binding. The POU proteins contact asymmetric recognition sequences of 12-15 nucleotides and can bind DNA as monomers or cooperatively as dimers. An inhibitory POU protein I-POU, has been described that dominantly inhibits DNA binding of other POU proteins (Rosenfeld, 1991).

bHLH proteins The basic-helix-loop-helix proteins contain a basic stretch of amino acids connected to a domain predicted to form two amphipathic helices separated by a loop. These proteins bind DNA as dimers to dyad symmetric sites. The helix-loop-helix domain appears to be responsible for dimerization and by analogy with the bZip proteins, the basic region is thought to contact DNA (Pabo and Sauer, 1992). The bHLH proteins can form heterodimers and inhibitory bHLH proteins, Id and emc, which lack a basic region, have been described (Benezra, et al., 1990; Garrell and Modolell, 1990). Finally, a subclass of bHLH proteins also contain a leucine zipper. These proteins, which include the proto-oncogene cMyc, may have

additional dimerization specificity due to the presence of this extra dimerization domain (Lamb and McKnight, 1991).

HMG Proteins HMG1 is a protein of unknown function that was described biochemically as an acid soluble, non-histone, nuclear protein. Its cloning revealed that it is composed of an amino-terminal acidic domain followed by two repeats of an 80 amino acid DNA binding domain, the HMG domain. More recent studies have identified several transcriptional regulators that contain the HMG domain (Giese, et al., 1992; Gubbay, et al., 1990; Jantzen, et al., 1990). One member of this family, LEF-1, has been studied extensively and shown to differ from other transcriptional regulators in several important ways. First, LEF-1, unlike most site specific DNA binding proteins, binds DNA in the minor, rather than major, groove. Second, LEF-1 can bend DNA through a large angle. This, combined with data suggesting that LEF-1 has no intrinsic capacity to stimulate transcription, led to the suggestion that LEF-1 functions by configuring DNA in a way that encourages interactions between other promoter bound proteins. This hypothesis is supported by experiments in which a bacterial DNA bending protein, IHF, was functionally substituted for LEF-1 (Giese, Cox and Grosschedl, 1992).

Other DNA binding proteins In addition to the families mentioned above, a number of other DNA binding proteins, some without an obvious structural motif, are known. These include NF-κB, heat shock factor, paired box, ets domain and CCAAT box binding proteins (Pabo and Sauer, 1992). The CCAAT box binding proteins are of particular interest because they appear to be composed of at least three structurally different types of protein that share the

ability to bind to sequences containing or related to CCAAT (Johnson and McKnight, 1989). These are discussed in chapter 4.

Transcriptional Regulatory Domains

The similarities of the proteins in the families described above are generally restricted to their dimerization and or DNA binding domains. With the exception of the HMG proteins, members of the other families described above are thought to regulate transcription by their ability to interact directly or indirectly with the proteins of the basal transcription machinery. Experiments in which portions of these proteins are fused to the DNA binding domain of another protein that lacks intrinsic regulatory functions demonstrate that specific protein domains can increase the rate of transcription from a promoter that binds the chimeric fusion protein (Brent and Ptashne, 1985). Although their function is not understood, a number of poorly defined motifs are repeatedly observed in transcription factors. These motifs include regions rich in acidic amino acids, rich in glutamine, or rich in proline (Courey and Tjian, 1988; Mermod, et al., 1989; Ptashne, 1988). Except for a suggestion that the acidic domains may form amphipathic alpha helices, nothing is known of the higher order structure of these domains (Giniger and Ptashne, 1987).

Evidence That Chromatin Structure Regulates Transcription

Substantial biochemical and genetic evidence that chromatin structure can modulate transcription has accumulated in recent years (Grunstein, 1990). Studies of the in vivo structure of chromatin, as assayed by sensitivity of specific DNA sequences to nucleases, have demonstrated a correlation between active transcription and a nuclease sensitive or open chromatin conformation (Gross and Garrard, 1988). In particular, this point has been clearly demonstrated in the β-globin locus (See below). It appears that domains of active chromatin can be propagated stably through many cell divisions, thus allowing negative regulation by chromatin to be specifically overcome at a particular genetic locus in a particular cell type or tissue (See next section).

A number of cis acting sequence elements that appear to regulate or affect chromatin structure have recently been described. The idea that such elements might exist comes from the observation of position effects. When genes are transferred from their natural context to a new location in the genome, their expression levels often change in a way not explained by the loss of regulatory elements. Rather, these results, "position effects", have been interpreted as being due to dominant effects of surrounding chromatin on expression of the heterologous gene. Three recent studies have identified elements that appear to regulate the effects of chromatin on nearby genes (Eissenberg and Elgin, 1991). Two of these elements, the β–globin LCR (See below) and the chicken lysozyme scaffold attachment regions appear to create a nuclease accessible,

active chromatin conformation (Stief, et al., 1989; Townes and Behringer, 1990). The third element, a "specialized chromatin structure" or scs, was defined as a Drosophila sequence element that provided a boundary between two domains of chromatin but did not change chromatin structure (or gene expression) directly (Kellum and Schedl, 1991). In the case of the scs and the chicken lysozyme scaffold attachment regions, effects were only observed when the test gene fell between two sequence elements. Therefore, these elements may define functional domains of chromatin across which, transcriptional information may not be passed. The β–globin LCR, on the other hand may be active as a single element.

Mechanisms by which chromatin and the transcription machinery interact have only recently been elucidated. Biochemical studies argue that nucleosomes or histone H1 may inhibit promiscuous, non-specific transcription by denying the general factor TFIID access to the TATA box (Croston, et al., 1991; Workman and Roeder, 1987). Some activators appear to act by specifically overcoming this inhibition (Taylor, et al., 1991; Workman, et al., 1988; Workman, et al., 1990; Workman, et al., 1991). The above in vitro studies suggest that the function of many activator proteins may be to counteract or prevent the negative effects of chromatin rather than to positively act on RNAPII itself. This statement is supported by two findings. First, in reconstituted in vitro systems which lack histones, many activators will stimulate transcription only several fold rather than the tens or hundred fold effects observed in vivo. However, when transcription is carried out on reconstituted chromatin templates, the same activators activate transcription at

levels comparable to those observed in vivo. Second, in the above experiments the stimulated levels of transcription observed on the reconstituted chromatin templates is generally slightly *less* than that observed on the naked DNA templates. The dramatic difference in stimulation of transcription in these experiments comes from the great suppression of transcription, in the absence of the activator, by histones.

Genetic studies of histone genes also imply links between transcriptional regulation and chromatin structure. In S. cerevisiae, depletion of histone H4 activates transcription of many genes (Han and Grunstein, 1988). Also, small deletions or single amino acid changes in the amino terminus of histone H4 results in the constitutive repression of a subset of transcriptionally regulated loci and constitutive activation of the silent mating type loci (Durrin, et al., 1991; Kayne, et al., 1988). Mutant histone genes have also been identified in a number of screens for suppressor mutations of transcriptional mutations in yeast (Clark-Adams, et al., 1988). Finally, as mentioned above, a yeast gene with some similarity to the chromatin associated protein HMG1 and a demonstrated non-specific DNA binding activity interacts genetically with the C-terminal repeats of RNAPII (Peterson, Kruger and Herskowitz, 1991).

Given that cells spend a significant amount of their resources in producing the protein components of chromatin and in packaging chromatin, it is worth considering what regulatory advantages chromatin renders to eukaryotic cells. One potential advantage can be found in the usefulness of a general system of transcriptional repression. Eukaryotes often express a variety of differentiated cell

types, each with their own cell specific genes. The expression levels of these genes can vary many orders of magnitude between cell types. By suppressing nonspecific, innappropriate transcription of genes the dynamic range of regulation between cell types can be expanded by many fold. Support for this role of chromatin comes from the biochemical experiments described above. Furthermore, cells appear to create domains of active or open chromatin in a cell type specific manner and appear to be able to stably propagate chromatin structures through cell divisions. Thus the chromatin structure of particular genetic loci within a cell may constitute a stable, epigenetic memory of the cell's history and influence its future regulatory decisions. These issues will be discussed below in the context of tissue specific transcription.

Tissue And Developmental Specific Transcription

The problem of tissue and developmental specific transcription is one of directing the transcription of a particular set of genes that are restricted in their tissue and/or temporal expression. Eukaryotes use several strategies to do this. First, many transcription factors are restricted in their tissue distribution. Second, chromatin appears to used to repress inappropriate gene expression.

Many tissue specific or tissue restricted transcription factors have been described in recent years. Because many of these factors are not truly restricted to a single tissue or developmental stage, and because they often regulate genes with a more restricted temporal or tissue distribution, it is quite likely that combinations of tissue

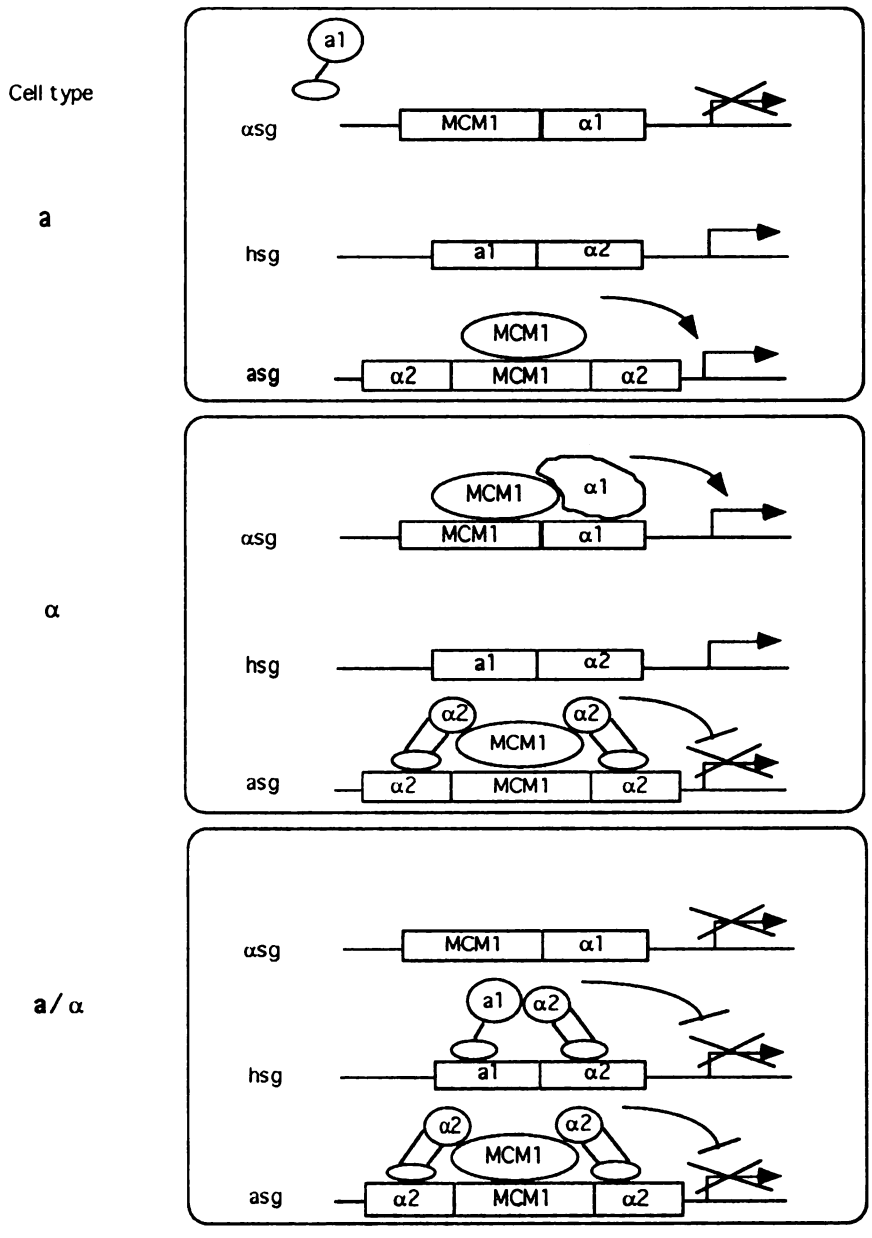
restricted factors combine to form higher order protein structures or surfaces that are specific to one cell type or developmental stage. Thus, the relative spacing of binding sites for tissue restricted factors as well as for more ubiquitous factors may be of critical importance for programming the tissue or developmental specificity of a promoter.

A simple example of these ideas comes from the control of cell type in S. cerevisiae (Herskowitz, 1987)(Herskowitz, 1987; Figure 2). This yeast exists as three distinct cell types: a or α haploid cells and a/α diploids. These cell types are specified by three cell type restricted factors, a1, α 1 and α 2, and one ubiquitous factor, MCM1. α1 and α2 protein can bind cooperatively with MCM1 to the appropriate composite operators. Additionally, a1 and α2 can bind DNA cooperatively with one another. The functional outcomes of these interactions are dependent upon the presence and correct juxtaposition of the binding sites for these various regulators.

a1 protein is produced in a and a/α cells, α 2 protein is produced in α and a/α cells. α 1 protein is found only in α cells. Genes transcribed only in α cells, α specific genes, are positively activated by a complex of α1 and MCM1. This complex binds cooperatively to a composite α1-MCM1 operator in the α specific gene promoters. Similarly, a specific genes are repressed in α cells by a complex of α2 and MCM1 bound to a composite α2-MCM1 operator in a specific genes. In a cells, a specific genes are constitutively active and α specific genes, in the absence of the α1 protein, are transcriptionally quiescent. In diploids, a specific genes are repressed by α 2/MCM1, haploid specific genes are repressed by

Figure 2.

Regulation of cell type in *Saccharomyces cerevisiae.* Three cell types occur in *Saccharomyces cerevisiae.*: haploid a and α cells, and diploid a/α cells. These cell types are determined by the regulated the actions of three classes of cell type specific genes: a specific genes (asg), α specific genes (asg) and haploid specific genes (hsg). The regulation of these genes is shown here. In a cells, asg and hsg are constitutively active and αsg are quiescent because they lack an activator. Note that the a1 protein apparently has no function in a cells. In α cells, hsg are constitutively active, asg genes are activated by a complex of a1 and MCM1 bound cooperatively to a composite α1-MCM1 operator and asg genes are repressed by a compex of α2 and MCM1 bound to a composite α2-MCM1 operator. In diploid a/α cells, hsg are repressed by a complex of a1 and α2 bound cooperatively to a composite a1/α2 operator. asg are repressed by α2-MCM1 complexes as they are in α cells. αsg are repressed indirectly in diploids, their activator protein, α1 is transcriptionally repressed by virtue of the presence of an a1/α2 operator in their promoter. Although diploid specific genes also exist, these appear to be regulated indirectly via a1/α2 repression of a negative regulator of diploid specific genes.


Cell type
a
α
a/α
a1
αsg
MCM1
α1
hsg
a1
α2
MCM1
asg
α2
MCM1
α2
MCM1
α1
αsg
MCM1
α1
hsg
a1
α2
α2
α2
MCM1
asg
α2
MCM1
α2
αsg
MCM1
α1
a1
α2
hsg
a1
α2
α2
α2
MCM1
asg
α2
MCM1
α2

a1/α 2 and α specific genes are repressed indirectly via a1/α2 repression of their positive activator, α 1.

Thus, the combination of one ubiquitous and three cell type restricted regulatory proteins are able to form alternative complexes that, in toto, specify cell type. The downstream targets of these regulators are specified in turn by the presence of particular binding sites for these alternative complexes. Note the role of the ubiquitous activator, MCM1, in the above scheme. It is required for transcription of both a and α specific genes. To activate α specific genes, its operator
must be juxtaposed with an α1 operator to facilitate activation. In a specific genes, MCM1 alone can apparently bind its operator and activate transcription. MCM1 is also required for repression of a specific genes. To accomplish this task, an MCM1 operator must be flanked by two α 2 operators. Finally, a high affinity MCM1 site alone in a promoter can drive constitutive gene expression. Thus the pattern of cell types that a gene is expressed in is dependent on the ability of its promoter to bind the appropriate protein complex. In the case of yeast, the ability of a promoter to facilitate interactions of MCM1 with other proteins plays a pivotal role in determination of cell type specific gene expression.

A de facto method of creating a cell type specific transcriptional regulator, other than by regulating it synthesis, is to modify the protein in a cell type specific manner resulting in a cell type specific activity. Similarly, cell specific regulation of nuclear import or of associated subunits can create a cell type specific regulator.

A second mechanism for cell type specific regulation is to repress inappropriate expression of tissue specific genes via chromatin structure. This scheme imagines that all of the factors required to activate transcription of a gene are present in a cell but that they cannot access the gene due to the chromatin structure surrounding it (Paro, 1990; Weintraub, 1985). The best example of this is the phenomenon of X chromosome inactivation. Also, when transfected into fibroblasts, the erythroid-specific β–globin genes are transcribed at a low but measurable level, yet the endogenous copies of the same gene are not expressed at all (Ptashne, 1988). This result is interpreted as reflecting the alternative chromatin structures of the endogenous and transfected copies of the gene. Several experiments have demonstrated stable, long term changes in an otherwise closed chromatin structure of a gene can be generated by transient signals (Groudine and Weintraub, 1982; Zaret and Yamamoto, 1984). Thus, there appear to be mechanisms whereby chromatin repression of transcription can be stably reversed.

Finally, to return to the problem of yeast cell type control, expression of the transcriptional regulators of yeast cell type, a1, α 1 and α2, are regulated by specialized chromatin domains (Herskowitz, 1987; Figure 3). All yeast cell types contain both the α 1 and α 2 genes as well as an a1 gene. Normally, in a cells, the α 1 and α2 genes are transcriptionally repressed. Likewise, in α cells, a1 gene transcription is repressed. In both cases, this repression is mediated by chromatin. Mutations that affect chromatin structure can result in derepression of these loci (Grosveld, et al., 1982). The resulting

Figure 3.

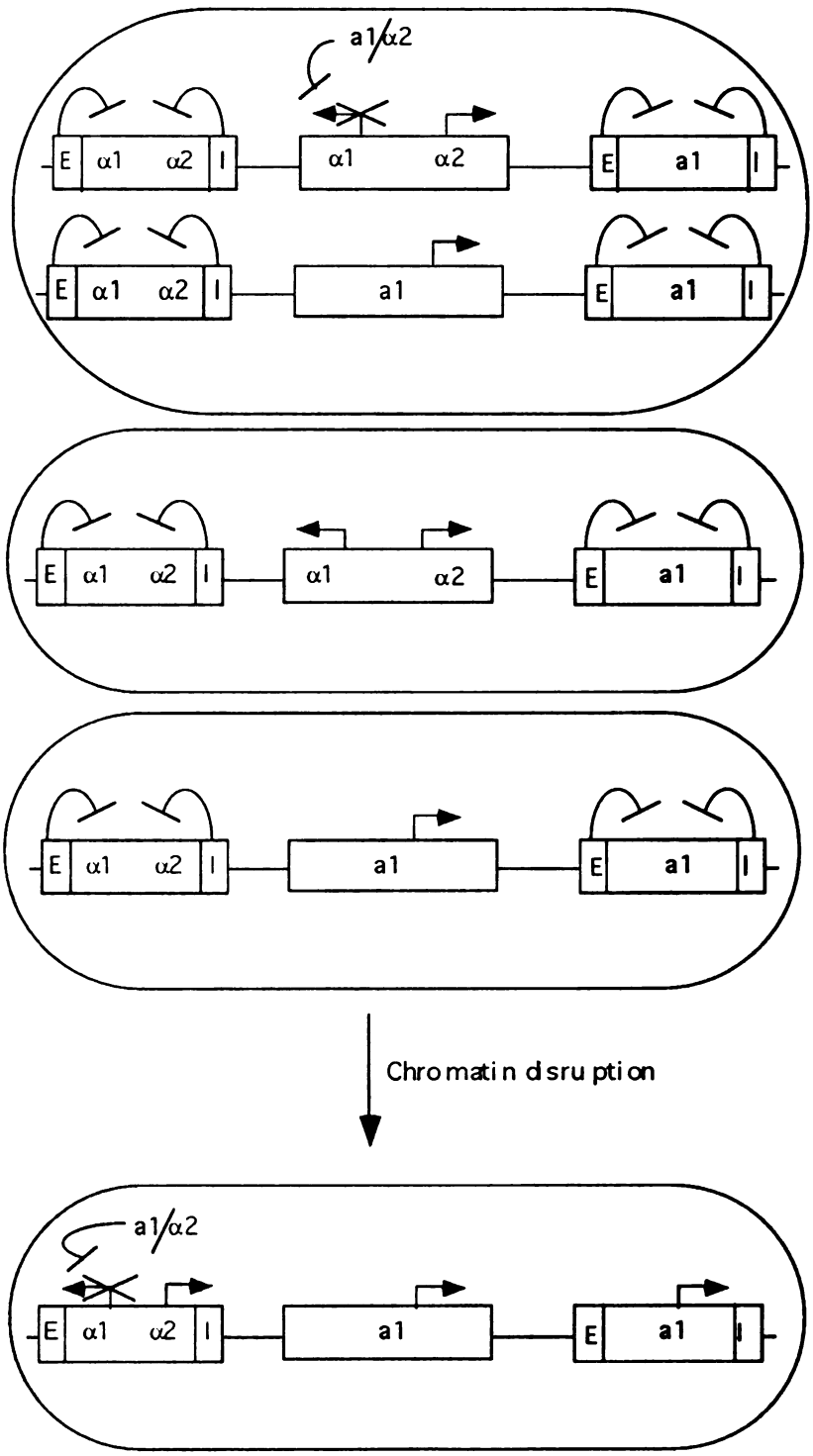
Chromatin control of cell type specific transcription of the yeast regulators a1, α1 and α2. The loci that transcribe the yeast transcriptional regulators that determine yeast cell type are known as MATa and MATα. Transcriptionally silent versions of MATa and MATa flank this locus. Yeast cell type can be changed from a to α and *vice versa* by gene conversion of a silent copy of the opposite type into the MAT locus. The transcriptionally silent state of the silent loci depends upon flanking cis elements, labeled E and I, and a number of protein components. In the diploid, because MAT loci of both types are present, both a and α regulatory proteins are produced. In particular, mutations that perturb chromatin structure can allow transcription of the silent loci giving production of a combination of regulators normally only found in dipoids.

Cell Type:

a/α

α

a



a

(Behaves like a/α)

inappropriate expression of all three regulators causes haploid cells to behave as if they were diploids (Herskowitz, 1987).

Mechanisms Of Transcriptional Initiation And Its Regulation

Because of the complexity of the RNAPII transcription machinery, biochemical studies of initiation of RNAPII are much more difficult to perform than in prokaryotes. In particular, until recently, most of the basal factors have only been available as partially purified column fractions, some of which are quite labile. Furthermore, TFIIA and other factors that interact with TFIID are poorly understood and may form alternative complexes with TFIID that result in promoter dependent pathways of building a transcription initiation complex (Roeder, 1991). However, several in vitro assays of transcription are available (Described in Sawadogo and Sentenac, 1990). Using crude extracts of nuclei as well as more purified systems, many activators have been shown to stimulate transcription in vitro. Biochemical experiments suggest that activators may function by facilitating TFIID-DNA interactions or by increasing the loading of TFIIB into transcription complexes (Abmayr, et al., 1988; Horikoshi, et al., 1988; Horikoshi, et al., 1988; Lin and Green, 1991; Sawadago and Roeder, 1985). As mentioned above, specific factors called co-activators may mediate the effect of transcription factor regulatory domains.

Eukaryotic transcription has a general requirement for ATP during or before the first polymerization reaction (Sawadago and Roeder, 1984). Furthermore, the carboxy-terminal repeated domain

(CTD) of the large subunit of RNAPII is converted from an unphosphorylated to phosphorylated form during this time (Corden, 1990; Lu, et al., 1991). This data has been used to suggest that the unphosphorylated form of the CTD may anchor RNAPII at a promoter prior to initiation and that phosphorylation of the CTD is a required and regulatable step of the initiation process. However, transcriptional initiation by RNAPII lacking the CTD also requires ATP (Laybourn and Dahmus, 1990). Thus phosphorylation of the CTD alone cannot explain the ATP requirement. Clearly, these data are far from complete. The cloning and characterization of the basal apparatus should allow more detailed biochemical experiments and add to our knowledge of mechanisms of regulation in the future.

Conclusions

Although more complicated than the prokaryotic system, the eukaryotic transcription system is thematically similar. Major differences between the two include the presence of a nucleus in eukaryotic cells, the role of chromatin structure in eukaryotes and added complexity of regulating many more genes and a variety of cell types within eukaryotic organisms. The variety of eukaryotic transcription factors and their extensive ability to interact with one another greatly increases the repertoire of potential regulatory structures that eukaryotes can form. A practical problem in investigation of eukaryotic transcriptional regulation is that of determining the nature of any protein(s) bound to a particular cis site and the nature of its interactions with other proteins. A

combined biochemical and genetic approach to this problem is discussed in chapter two. A further problem is presented by the need to be aware and know the consequences of potential post-translational modifications or non-DNA binding protein subunits that may be associated with a bound factor under a particular physiologic state or in a particular tissue.

REGULATION OF THE β–LIKE GLOBIN GENES

The β–like globin genes code for the β–chain of hemoglobin, an erythroid specific gene product. The members of this family produce hemoglobin isoforms that vary in their affinity for oxygen and are expressed sequentially through the course of development. The organization, sequence and regulatory mechanisms of the β–like globin genes are well conserved among the vertebrates (Bunn and Forget, 1986). This section describe studies carried out in human, mouse and chicken systems.

Developmental Regulation of the β-Globin Family

During human development, three classes of β–globin gene are expressed. Early in development, the embryonic β–gene, ε, is expressed in the yolk sac. At about 5 weeks of gestation, the two fetal γ-globin genes, $^{G}\gamma$ and $^{A}\gamma$ begin to be expressed and the site of

erythropoeisis switches to the fetal liver. Beginning around 20 weeks of gestation, the adult genes, β, and a minor variant, δ, begin to be expressed and increase in expression until, at birth, they are the only β–genes transcribed. Also at 20 weeks, erythropoeisis begins in the spleen and later, in the bone marrow. The major location of adult erythropoeisis is in the bone marrow (Bunn and Forget, 1986).

In the mouse, the embryonic β–like genes, y, β h0 and β h1 are expressed in the yolk sac. At 9-10 days of gestation, the site of erythropoeisis switches to the fetal liver and expression of the adult β–like globins, β^{maj} and β^{min} , begins and increases as a proportion of the total β–like gene expression until they are the only β–like genes expressed. Later in mouse development, the site of erythropoeisis switches to the spleen and then to the bone marrow. The mouse, like most species, does not have a fetal specific β–like gene (Reviewed in Kollias, et al., 1986).

The β globin locus

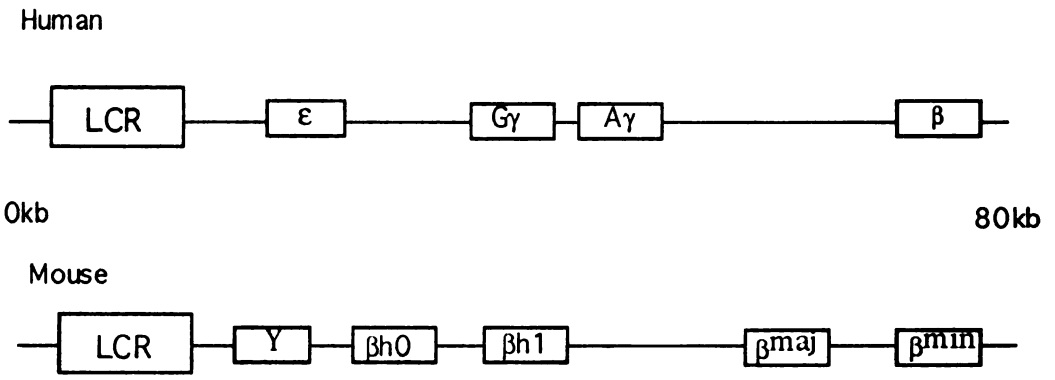
The β–like globin genes are arranged in a single locus, 60-80 kilobases long and, in most species are arranged 5' to 3' in the same order in which they are expressed during development (Figure 4A). The entire nucleotide sequences of both the mouse and human loci have been determined. Additionally, the sequences of the genes and proximal promoter regions of β–like globin genes from many other species have also been determined (Bunn and Forget, 1986; Efstratiadis, et al., 1980). Comparison of these sequences reveals that

Figure 4.

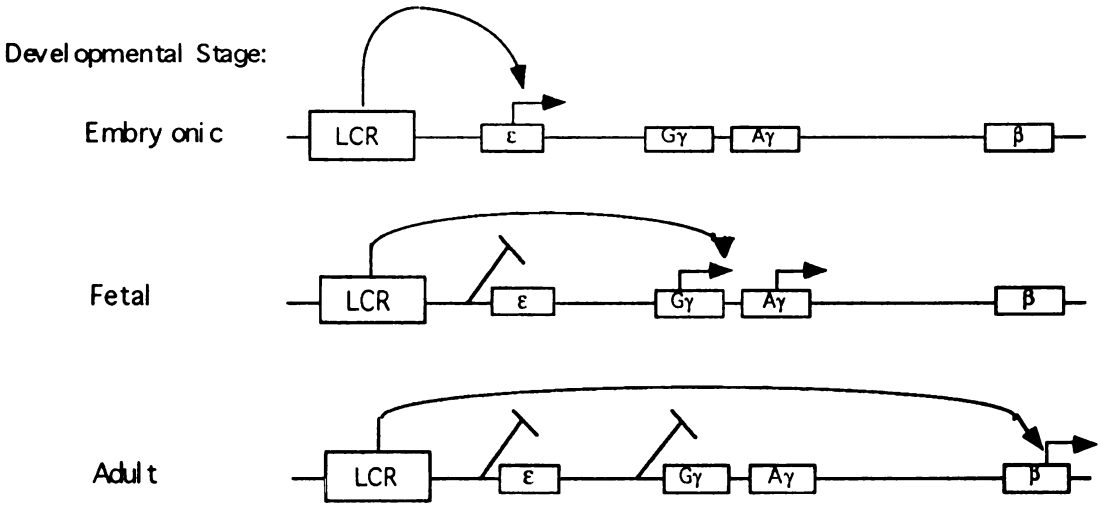
A. The mouse and human β-globin loci.

B. Developmental regulation of human β-globin expression. In the model presented here, the developmental regulation of the β-locus is achieved by a positive competition for the attention of the LCR coupled with a negative regulation of β-gene expression in stages subsequent to that in which it is expressed. For example, in the fetal stage, the g-globin genes out compete the β-globin gene for the effects of the LCR while the embryonic ε gene is negatively regulated. Although this figure implies that the negative regulation occurs at the promoter-LCR interface, its true nature is unknown and could concievably operate directly on the promoter.

A


Human
LCR
ε
Gγ
Aγ
β
0kb
80kb
Mouse
LCR
Y
βh0
βh1
βmaj
βmin

B


Developmental Stage:
Embryonic
LCR
ε
Gγ
Aγ
β
Fetal
LCR
ε
Gγ
Aγ
β
Adult
LCR
ε
Gγ
Aγ
β

the genes and their control regions were highly conserved throughout evolution.

Structure of the β–globin Promoters

In addition to the strong conservation across species of the coding sequence of the β–like globin genes, the non-coding sequences in these genes display a moderate level of conservation. Additionally conservation is seen between developmental variants of the β-like genes both in their coding and non-coding sequences. Overall, the sequence conservation across the promoters of the β–like genes from several different species is about 50%. Notable blocks of sequence conservation include a TATA box at about -30 nucleotides from the transcription start site and a sequence, CCAAT, at position -80(Bunn and Forget, 1986).

The cis acting sequences of β–globin and the other β–like genes were mapped by transfection of mutant β–gene constructs, first into fibroblast cells and later into erythroid like cell lines (Cowie and Myers, 1988; Dierks, et al., 1983; Myers, et al., 1986; Stuve and Myers, 1990). These constructs were linked to heterologous enhancers to drive their expression (See below). These transfection experiments revealed four basic cis elements in the adult β-gene promoters. These are the TATA box at position -30, a directly repeated sequence at -50 called the βDRE, the CCAAT sequence at -80, and a sequence, CCT/ACACCC (CACCC element), at -90. The CCAAT and CACCC elements are duplicated in some promoters (Bunn and Forget, 1986; Efstratiadis, et al., 1980). Supporting the role of

these elements in transcription of β-genes in vivo, certain patients with deficiencies in β–gene transcription have been found to have point mutations in the CACCC, CCAAT and TATA elements (Bunn and Forget, 1986).

Interestingly, mutations in these elements seem to have similar effects on β-globin transcription in both erythroid and non-erythroid cells (Cowie and Myers, 1988; Myers, Tilly and Maniatis, 1986). Furthermore, in erythroid-like cell lines that can be induced to differentiate from a pre-erythroblast stage to a terminally differentiated stage, mutations in the cis elements of the mouse β^{maj} promoter have similar effects before and after differentiation (Cowie and Myers, 1988; Stuve and Myers, 1990). Deletion mutagenesis of the adult β-promoters showed no requirement for sequences upstream of nt -109 (Cowie and Myers, 1988; Dierks, et al., 1983).

Later, more extensive mutagenesis studies, combined with DNA binding studies, identified binding sites, outside of the proximal promoter, for 2 erythroid/megakaryocyte specific factors. The first of these sites conforms to the sequence, T/A GATA A/G, and is bound by an erythrocyte/megakaryocyte specific factor, GATA-1 (Evans, et al., 1988; Romeo, et al., 1990; Wall, et al., 1988). The other factor, NF-E2, binds to sites similar to those bound by the bZIP protein AP1 (Ney, et al., 1990; Romeo, et al., 1990). Therefore, NF-E2 is thought to be a potential erythroid/megakaryocyte specific bZip protein(Mignotte, et al., 1989). The gene for GATA-1 has been cloned and demonstrated to code for a zinc-finger type transcriptional activator (Martin and Orkin, 1990). When the GATA-1 gene was

inactivated in vivo, erythroid development was prevented (Pevny, et al., 1991).

The LCR

In addition to the proximal promoter regions of the β–like globin genes, there is a regulatory region that is shared by the entire locus. The discovery of this region, the Locus Control Region or LCR, was driven by the inability to obtain reproducible, high efficiency expression of β–like genes in tissue culture cells or transgenic mice. In either assay, levels of β–globin gene expression varied from clone to clone, were not consistently erythroid specific, and were not expressed at levels comparable to the endogenous globin genes or proportional to the number of copies of the gene transferred (Townes and Behringer, 1990). Inclusion of a heterologous enhancer increased expression somewhat but did not overcome the other problems mentioned. These results were interpreted as being due to two problems. First, the lack of erythroid specificity and low expression levels were thought to be due to the absence of regulatory sequences, present in the β–globin locus but lacking in the small plasmid borne constructs used to generate the transgenic mice and long term transfectants. Second, the inconsistent expression levels, the lack of copy number dependence on expression level and the lack of erythroid specificity were thought to reflect effects of DNA sequences surrounding the inserted gene on its expression. Since the insertion site of the transfected gene into the recipient is relatively random, the effects of the insertion site would vary from

recipient cell to recipient cell. Supporting these interpretations, experiments in which an entire chromosome containing the human β-locus was transferred to a heterologous erythroid cell showed that, in its natural chromosomal context, the β-genes were appropriately regulated (Baron and Maniatis, 1986).

Given these results, several groups looked for other regulatory sequences in the β-globin locus. While several enhancer and silencer type sequences were found near individual globin genes, none of these solved all of the expression problems mentioned above (Behringer, et al., 1987; Bodine and Ley, 1987; Kollias, et al., 1987; Trudel and Costantini, 1987). Two groups took a slightly different approach to finding potential regulatory sequences by first mapping DNaseI hypersensitive sites within the β-globin locus (Forrester, et al., 1986; Tuan, et al., 1985). This strategy was based on the correlation of sensitivity to DNaseI with transcriptional activity, the association of DNaseI hypersensitive sites with transcriptional control regions and the hypothesis that certain chromatin structures may be stably propagated through the cell cycle and used as a developmental regulatory signal. Additionally, several cases of β-thalassemia had been shown to due to deletions upstream of the β-locus that did not affect the β-genes themselves (Townes and Behringer, 1990). Therefore, these regions represented sites of potential control elements and were assayed for DNase I hypersensitive sites.

In erythroid cells, the β-globin locus was found to exhibit three levels of DNaseI sensitivity. First, the entire locus was moderately sensitive to DNaseI. Second, minor DNaseI hypersensitive sites were

observed at the 5' end of actively transcribing β–like genes (Sheffery, et al., 1982). And finally, several erythroid specific, but developmental stage and transcription independent, DNaseI super-hypersensitive sites were found to flank either side of the locus (Forrester, et al., 1986; Tuan, et al., 1985). These sites, when linked to a β–globin gene, acted as an erythroid specific enhancer in transient transfection assays (Tuan, et al., 1989). In transgenic mice or long term transfection assays, this element conferred integration-site independent, copy number dependent, high level, erythroid specific transcription to linked β–like genes as well as to heterologous non-erythroid specific genes (Forrester, et al., 1989; Grosveld, et al., 1987; van Assendelft, et al., 1989). In addition to copy number dependence, β–globin genes linked to the LCR were transcribed at levels per gene that were similar to those of the endogenous globin genes (Grosveld, et al., 1987).

The above results suggested that the LCR plus the proximal promoter regions of the β–like genes constitute the major transcriptional regulatory elements of the β–globin locus. Subsequent studies have shown that the hypersensitive site 3' to the locus has no detectable transcription regulatory activity and that the 5' hypersensitive sites provide redundant functions (Collis, et al., 1990; Forrester, et al., 1989; Townes and Behringer, 1990). Additional support for the LCR as a regulator of β–like gene transcription comes from thalassemia cases in which deletions, that remove all or part of the LCR but leave the downstream β–like genes intact, inactivate transcription of the downstream β–like genes and transform the affected β–locus into a transcriptionally inactive,

nuclease resistant chromatin conformation (Forrester, et al., 1990; Townes and Behringer, 1990). This observation is consistent with models in which the LCR creates an erythroid specific, open chromatin structure that facilitates subsequent expression of the individual β–like genes.

Discovery of the human LCR in turn led to discovery of similar sequences in other organisms and gene families. In the chicken, an enhancer, between the adult β and a downstream embryonic β gene, is associated with a DNaseI hypersensitive site and confers site independent expression on either gene in transfections (Nickol and Felsenfeld, 1988). When the chicken β^{maj} gene and its enhancer are inserted into transgenic mice, the enhancer functions analogously to the human LCR (Reitman, et al., 1990). The human α globin locus also has an LCR like element. This element colocalizes with a DNaseI hypersensitive site and confers high level, copy number dependent, site independent, erythroid specific regulation on linked α globin genes (Higgs, et al., 1990). In contrast to the β-globin locus, the α-globin locus is in a constitutively open conformation (Vyas, et al., 1992). Thus, whetherthe α-LCR to regulates chromatin structure is unknown. Finally, sequences near the CD2 gene and the chicken lysozyme gene (See above) seem to have LCR like elements (Bonifer, et al., 1990; Greaves, et al., 1989).

Fine Structure of the LCR

The fine structure of the LCR has been mapped, first by deletion mutagenesis of LCR sequences that were subsequently assayed for function in gene transfer experiments, and second by a combination of in vitro and in vivo footprinting studies. The deletion studies have narrowed the functional regions at each of the DNaseI hypersensitive sites to two to three hundred nucleotides (Collis, Antoniou and Grosveld, 1990; Curtin, et al., 1989; Forrester, et al., 1989; Philipsen, et al., 1990; Ryan, et al., 1989; Talbot, et al., 1990; Tuan, et al., 1989). Sequence analysis of these regions revealed numerous potential CACCC, GATA-1 and NF-E2 binding sites. Footprinting studies showed that many of these sites are bound by proteins in vitro and that some of the sites are protected from methylation by DMS in vivo (Ikuta and Kan, 1991; Ney, et al., 1990; Philipsen, et al., 1990; Qiliang, et al., 1990; Reddy and Shen, 1991; Strauss and Orkin, 1992; Talbot and Grosveld, 1991; Talbot, et al., 1990). The in vivo protections were erythroid specific and some were partially inducible during induction of erythroid differentiation. The in vivo protected CACCC and GATA-1 elements observed in these studies tended to occur near one another (Ikuta and Kan, 1991; Strauss and Orkin, 1992). This is interesting given several studies that suggest that CACCC binding proteins can functionally cooperate with factors bound nearby. Interestingly, the human α globin LCR and the chicken LCR-like β-globin enhancer are also composed of closely spaced CACCC, GATA-1 and NF-E2 binding sites that have been shown to be protected from DMS or DNaseI in vivo (Emerson, et al., 1987; Jackson, et al., 1989; Strauss, et al., 1992). In transient transfection assays, the second hypersensitive site of the human β-

globin LCR functions as an enhancer (Ney, et al., 1990). This enhancing activity has been mapped to a pair of tandem NF-E2 sites that occur in an eighteen basepair region (Ney, et al., 1990; Talbot and Grosveld, 1991). Additional sequences required to provide position independent expression, include CACCC and GATA-1 elements that are footprinted in vivo (Ney, et al., 1990; Talbot and Grosveld, 1991). Finally, in the chicken globin enhancer, transfection assays have shown that the CACCC, NF-E2 and GATA-1 sites are all required for optimal β-globin transcription (Reitman and Felsenfeld, 1988).

Developmental Regulation of the β–like genes

The developmental regulation of the β–like genes is particularly puzzling because the proximal promoters of these genes do not exhibit strong stage specificity and because they all share the same upstream regulatory element, the LCR. During development, expression of a particular β–like gene is not limited to a single tissue. For example, some adult β–globin expression is observed in the fetal liver. Furthermore, during fetal development, both γ– and β–globin gene expression can be observed in the same cell and the proportion of β–like gene expression made up by the adult β–globin gene rises during the course of development (Stamatoyannopoulos, et al., 1987). This is consistent with a model in which the regulation of the switch from γ–globin to β–globin expression is graded rather than being due to utilization of fundamentally different or developmental stage specific erythroid precursor cells. Finally, even in the mature adult,

a subset of red blood cells contain a small amount of γ-globin chains in addition to β-globin (Stamatoyannopoulos, et al., 1987). This suggest that the switching mechanism is not absolutely rigorous.

A model of switching with strong experimental support was first described in the chicken system. Recall that the chicken β locus contains an enhancer with LCR like functions that lies between the adult and embryonic genes (Reitman, et al., 1990). When the enhancer is transfected into erythroid cells with either the adult or embryonic gene alone, that gene is expressed in an erythroid specific but developmental stage non-specific manner (Choi and Engel, 1988; Nickol and Felsenfeld, 1988). However, when both the adult and embryonic genes are transfected together on a single construct that also contains the enhancer, normal developmental regulation of the β genes is observed (Choi and Engel, 1988). Furthermore, mutations that decrease the transcriptional efficacy of either the adult or embryonic gene in this construct lead to a reciprocal increase in transcription of the other gene and a loss of developmental specificity (Foley and Engel, 1992). Finally, insertion of a second enhancer in constructs with both the adult and embryonic gene allows expression of both genes in a developmental stage non-specific manner (Choi and Engel, 1988; Foley and Engel, 1992). These results led to the idea that developmental regulation of β-like globin genes may occur via competition of the β-gene promoters for LCR sequences (See Figure 4B). In this case differences between promoters of the β-like genes would lead to differing abilities to compete for the LCR at different stages of development. If competition were the sole mechanism for developmental regulation

of the b-genes, discrimination between them would be limited. Several results that suggest a role for negative regulation in this system will be discussed below.

Support for a competitive mechanism of globin switching in mammals came from experiments in which multiple β–like genes were linked to the LCR on a single construct and inserted into transgenic mice. These experiments showed that in contrast to the case for a single gene linked to the LCR, which is often expressed constitutively in erythroid cells, multiple β-like genes would undergo switching (Behringer, et al., 1990; Enver, et al., 1990). Moreover, switching depended not only on the presence of two β-like genes but also on their relative orientation with respect to one another. Genes nearer to the LCR tended to be expressed earlier and longer during development of the transgenic mice than genes farther away from the LCR (Hanscombe, et al., 1991). Additional support for this model comes from an experiment in which the strong erythroid promoter of the Friend virus was inserted into the human β–globin LCR by homologous recombination in an erythroid cell line. In this cell line, the downstream β–like genes are all transcriptionally inactive and DNaseI resistant while transcription of the inserted Friend virus promoter is induced by erythroid differentiation (Kim, et al., 1992). This result was interpreted to reflect the ability of the inserted promoter to out compete the β–genes for the LCR.

One question of importance is how the functional interaction of the LCR with the downstream β-genes occurs. Two studies, one in the context of the human β–locus and the other in the chicken, argue that the regions of the proximal promoter encompassing the CACCC,

CCAAT, βDRE and TATA elements are all required for full cooperation with the LCR (Antoniou and Grosveld, 1990; Foley and Engel, 1992). Additionally when heterologous genes were placed adjacent to the LCR, most were expressed in an erythroid specific manner (Hanscombe, et al., 1991; van Assendelft, et al., 1989). However, when the mouse histone H4 promoter is placed next to the LCR, it is not activated (Antoniou and Grosveld, 1990). Intriguingly, this promoter, in contrast to the others placed next to the LCR, lacks a functional CACCC, CCAAT or βDRE box. Thus interactions of the proximal promoters with the LCR may be mediated by proteins binding the CACCC, CCAAT or βDRE elements. These elements appear to be bound by ubiquitous factors (See chapters 2 and 4 and Stuve and Myers, 1990), suggesting that a ubiquitous protein or proteins is required for the LCR to generate erythroid specific transcription of the β–genes.

If the competition model for β–gene developmental regulation is correct, then some difference between the individual β-genes must explain the ability of the LCR to be functionally redirected during the course of development. Three potential ways in which this might occur will be considered here. First, developmental regulation may come from the addition of silencer elements that abrogate transcription of earlier developmental variants of the β-genes at later stages. There is evidence that such a mechanism occurs with the embryonic gene during the fetal and adult stages of development and with the fetal genes in the adult stage. When an embryonic or fetal gene alone was linked to the LCR, expression extinguished at the correct developmental stage (Dillon and Grosveld, 1991; Lindenbaum

and Grosveld, 1990). Furthermore, the γ-LCR construct was expressed at the embryonic stage (Dillon and Grosveld, 1991). Thus, the addition of negative regulatory elements appears to confer greater developmental specificity to the β-genes.

Second there may be stage specific transcription factors that bind only to the promoter of the β-like gene active at that stage and function to recruit the LCR. If this model were correct, then mutations within the binding site for such a factor would disrupt the ability of the promoter to compete for the LCR and allow expression of normally silent β-genes. Mutations of this type have recently been described in the human γ-globin promoter (Jane, et al., 1992).

A third mechanism for recruiting the LCR in a stage specific manner is to use ubiquitous factors that vary in abundance or activity during development to specify alternate interactions with the LCR. For example, if the affinity of the CACCC factor for its binding sites in the β-genes was highest for the adult β-gene and lowest for the embryonic β-gene, then by decreasing the level of the CACCC binding factor during development, the occupancy of CACCC elements could be developmentally regulated and in turn could regulate interactions with the LCR. Within this model, it is possible to further consider two potential mechanisms of LCR-proximal promoter interactions. The LCR may form specific protein-protein interactions with a factor bound to a β-gene proximal promoter or, alternatively, or it may recognize pre-initiation complexes assembled at a promoter and discriminate between promoters based on their ability to form such complexes.

Conclusions

As discussed in the sections on general transcriptional regulation, gene specific regulation depends on the protein binding sites present in a promoter, whether or not the promoter is accessible to proteins, the particular transcriptional regulatory proteins present in the nucleus and the protein-protein interactions on the promoter. In the β-globin locus, the cis elements of the proximal promoters have been mapped at high resolution. The cis elements of the LCR are beginning to be mapped. Two erythroid specific factors are known to bind regulatory elements in the locus, and one of these, GATA-1, has a proven role in erythroid development. Furthermore, in erythroid cells, the chromatin structure of the locus is known to be permissive to transcription.

Several major issues in globin regulation have yet to be adequately addressed. First, what is the identity of the factors that bind the cis elements in the proximal promoters and the LCR? Second, how does the LCR regulate chromatin structure? Third, how do the LCR and β-gene promoters interact? Fourth, how do the above interactions lead to developmental specificity?

For my thesis work, I have chosen to lay the groundwork for answering the above questions by biochemically characterizing factors that interact with the promoter of the mouse β^{maj} gene. Specifically, I have characterized and purified factors that bind the CACCC and CCAAT elements of this promoter (Chapters three and four). These studies have highlighted some of the complexities of gene regulation that were discussed above. First, the CCAAT factor is

composed of multiple subunits and may exist in alternate conformations. Second, multiple proteins can bind to the CACCC element in vitro. Chapter two of this thesis describes a methodology for biochemically discriminating between multiple factors that bind to a single cis element and determining which of these is most likely responsible for the in vivo function of the element.

References

Abmayr, S.M., Workman, J.L., and Roeder, R.G. (1988). The pseudorabies immediate early protein stimulates in vitro transcription by facilitating TFIID: promoter interactions. Genes Dev *2*(5), 542-53.

Allison, L.A., and Ingles, C.J. (1989). Mutations in RNA polymerase II enhance or suppress mutations in Gal4. Proc. Natl. Acad. Sci. USA *86*, 2794-2798.

Antoniou, M., and Grosveld, F. (1990). β-globin dominant control region interacts differently with distal and proximal promoter elements. Genes and Dev. *4*, 1007-1013.

Atkinson, M.L. (1988). Enhancers: mechanisms of action and cell specificty. Annu. Rev. Cell Biol. *4*, 127-153.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 sequences. Cell *27*, 299-308.

Baron, M., and Maniatis, T. (1986). Rapid reprograming of globin gene expression in transient heterokaryons. Cell *46*, 591-602.

Behringer, R., Hammer, R.E., Brinster, R.L., Palmiter, R.D., and Townes, T.M. (1987). Two 3' sequences direct adult erythroid-specific expression of human β-globin genes in transgenic mice. Proc. Natl. Acad. Sci. USA *84*, 7056-7060.

Behringer, R.R., Ryan, T.M., Palmiter, R.D., Brinster, R.L., and Townes, T.M. (1990). Human γ- to β-globin gene switching in transgenic mice. Genes Dev. *4*, 380-389.

Benezra, R., Davis, R.L., Lockshon, D., Turner, D.L., and Weintraub, H. (1990). The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell *61*, 49-59.

Benoist, C., and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promoter region. Nature *290*, 304-310.

Blank, V., Kourilsky, P., and Israel, A. (1992). NF-κB and related proteins: Rel/dorsal homologies meet ankyrin-like repeats. Trends Biochem. *17*, 135-140.

Bodine, D.M., and Ley, T. (1987). An enhancer element lies 3' to the human $^{A}\gamma$ globin gene. EMBO J. *6*(10), 2997-3004.

Bonifer, C., Vidal, M., Grosveld, F., and Sippel, A.E. (1990). Tissue specific and position independent expression of the complete gene domain for chiken lysozyme in transgenic mice. EMBO J. *9*(2843-2848),

Brent, R., and Ptashne, M. (1985). A eukaryotic transcriptional activator with the specificity of a prokaryotic repressor. Cell *43*, 729-736.

Bunn, H.F., and Forget, B.G. (1986). Hemoglobin: molecular genetic and clinical aspects. Philadelphia, W.B. Saunders Co.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell *56*, 549-561.

Chamberlin, M.L. (1982). Bacterial DNA-dependent RNA polymerases. *In:* The enzymes. New York, Academic Press, pp. 61-86.

Choi, O.-R.B., and Engel, J.D. (1988). Developmental regulation of β-globin gene switching. Cell *55*, 17-26.

Clark-Adams, C.D., Norris, D., Osley, M.A., Fassler, J.S., and Winston, F. (1988). Changes in histone gene dosage alter transcription in yeast. Genes Dev *2*(2), 150-9.

Collis, P., Antoniou, M., and Grosveld, F. (1990). Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. EMBO J. *9*(1), 233-240.

Concino, M.F., Lee, R.F., Merryweather, J.P., and Weinmann, R. (1984). The adenovirus major late promoter TATA box and initiation site are both necessary for transcription in vitro. Nucl. Acids Res *12*(9), 7423.

Corden, J., Wasylk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., and Chambon, P. (1980). Promoter sequences of eukaryotic protein-coding genes. Science *209*, 1406-1414.

Corden, J.L. (1990). Tails of RNA polymerase II. Trends Biochem. *15*, 383-387.

Courey, A.J., and Tjian, R. (1988). Analysis of Sp1 reveals multiple transcription activation domains, including a novel glutamine-rich activation motif. Cell *55*, 887-898.

Cowie, A., and Myers, R.M. (1988). DNA sequences involved in transcriptional regulation of the mouse β-globin promoter in murine erythroleukemia cells. Mol. Cell. Biol. *8*(8), 3122-3128.

Croston, G.E., Kerrigan, L.A., Lira, L.M., Marshak, D.R., and Kadonaga, J.T. (1991). Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription. Science *251*(4994), 643-9.

Curtin, P.T., Liu, D., Liu, W., Chang, J.C., and Kan, Y.W. (1989). Human β-globin gene expression in transgenic mice is enhanced by a distant DNase I hypersensitive site. Proc. Natl. Acad. Sci. USA *86*, 7082-7086.

de Crombrugghe, B., Busby, S., and Buc, H. (1984). Cyclic AMP receptor protein: role in transcriptional activation. Science *224*, 831-837.

Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J., and Weissman, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell *32*, 695-706.

Dillon, N., and Grosveld, F. (1991). Human γ-globin genes silenced independently of other genes in the β-globin locus. Nature *350*, 252-254.

Dunaway, M., and Droge, P. (1989). Transactivation of the Xenopus rRNA gene promoter by its enhancer. Nature *341*, 657-659.

Durrin, L.K., Mann, R.K., Kayne, P.S., and Grunstein, M. (1991). Yeast histone H4 N-terminal sequence is required for promoter activation *in vivo*. Cell *65*, 1023-1031.

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

Efstratiadis, A., et al. (1980). The Structure and evolution of the human β-globin gene family. Cell *21*, 653-668.

Eissenberg, J.C., and Elgin, S.C.R. (1991). Boundary functions in the control of gene expression. Trends Genet. *7*(10), 335-339.

Emerson, B.M., Nickol, J.M., Jackson, D.P., and Fesenfeld, G. (1987). Analysis of the tissue specific- enhancer at the 3' end of the chicken adult β-globin gene. Proc. Natl. Acad. Sci. USA *84*, 4786-4790.

Enver, T., Raich, N., Ebens, A.J., Papayannopoulou, T., Costantini, F., and Stamatoyannopoulos, G. (1990). Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. Nature *344*, 309-313.

Evans, T., Reitman, M., and Felsenfeld, G. (1988). An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. Proc. Natl. Acad. Sci. USA *85*, 5976-5980.

Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z., and Reinberg, D. (1991). the small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. Proc. Natl. Acad. Sci. USA *88*, 9999-10003.

Flores, O., Lu, H., and Reinberg, D. (1992). Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. J. Biol. Chem. *267*, 2786-2793.

Foley, K.P., and Engel, J.D. (1992). Individual stage selector element mutations lead to reciprocal changes in β- vs. ε-globin gene transcription: genetic confirmation of promoter competition during globin gene switching. Genes and Dev. *6*(730-744),

Forrester, W.C., Epner, E., Driscoll, C., Enver, T., Brice, M., Papayannopoulou, T., and Groudine, M. (1990). A deletion of the human β-globin locus activation region causes a major alteration in

chromatin structure and replication across the entire β-globin locus. Genes Dev. *4*, 1637-1649.

Forrester, W.C., Novak, U., Gelinas, R., and Groudine, M. (1986). A developmentally stable chromatin structure in the human β-globin gene cluster. Proc. Natl. Acad. Sci. USA *83*, 1359-1363.

Forrester, W.C., Novak, U., Gelinas, R., and Groudine, M. (1989). Molecular analysis of the human β-globin locus activation region. Proc. Natl. Acad. Sci. USA *86*, 5439-5443.

Garrell, J., and Modolell, J. (1990). The Drosophila extramacrochaetae locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop-helix protein. Cell *61*, 39-48.

Giese, K., Cox, J., and Grosschedl, R. (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein complexes. Cell *69*, 185-195.

Giniger, E., and Ptashne, M. (1987). Transcription in yeast activated by a putative amphipathic α helix linked to a DNA binding unit. Nature *330*, 670-672.

Glass, R.E. (1982). Gene Function. Berkeley, University of California Press.

Goutte, C., and Johnson, A.D. (1988). a1 protein alters the DNA binding specificity of alpha2 repressor. Cell *52*, 875-882.

Greaves, D.R., Wilson, F.D., Lang, G., and Kioussis, D. (1989). Human CD2 3'-flanking sequences confer high-level T cell-specific, position-independent gene expression in transgenic mice. Cell *56*, 979-986.

Green, M.R. (1992). Gene regulation. Transcriptional transgressions. Nature *357*, 364-365.

Gross, D.S., and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. Annu. Rev. Biochem. *57*, 159-197.

Grosschedl, R., and Birnstiel, M.L. (1980). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Natl. Acad. Sci. USA *77*, 1432-1436.

Grosveld, F., van Assendelft, G.B., Greaves, D.R., and Kollias, G. (1987). Position-independent, high level expression of the human β-globin gene in transgenic mice. Cell *51*, 975-985.

Grosveld, G.C., de Boer, E., Shewmaker, C.K., and Flavell, R.A. (1982). DNA sequences necessary for transcritpion of the rabbit β-globin gene *in vivo*. Nature *295*, 120-126.

Groudine, M., and Weintraub, M. (1982). Propagation of globin DNAase i-hypersensitive site in absence of factors required for induction: a possible mechanism for determination. Cell *30*, 131-139.

Grunstein, M. (1990). Histone function in transcription. Annu Rev Cell Biol *6*, 643-78.

Gubbay, J., Collington, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. Nature *346*, 245-250.

Han, M., and Grunstein, M. (1988). Nucleosome loss activates yeast downstream promoters *in vivo*. Cell *55*(6), 1137-45.

Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N., and Grosveld, F. (1991). Importance of globin gene order for correct developmental expression. Genes Dev. *5*, 1387-1394.

Harrison, S.C. (1991). A structural taxconomy of DNA-binding domains. Nature *353*, 715-719.

Hawley, D.K., Johnson, A.D., and McClure, W.R. (1985). Functional and physical characterization of transcription initiation complexes in the bacteriophage λ O_R region. J. Biol. Chem. *260*(14), 8618-8626.

Hayashi, S., and Scott, M. (1990). What determines the specificity of action of Drosophila homeodomain proteins? Cell *63*, 883-894.

Helmann, J.D., and Chamberlin, M.J. (1988). Structure and function of bacterial sigma factors. Annu. Rev. Biochem. *57*, 839-872.

Herr, W., et al. (1988). The POU domain: a large conserved region in the mammalian *pit-1, oct-1, oct-2* and *Caenorhabditis elegans unc-86* gene products. Genes and Dev. *2*, 1513-1516.

Herskowitz, I. (1987). A master regulatory locus that determines cell specialization in yeast. *In:* The Harvey Lectures. Alan R. Liss inc., pp. 67-92.

Higgs, D.R., Wood, W.G., Jarman, A.P., Sharpe, J., Lida, J., Pretorius, I.-M., and Ayyub, H. (1990). A major positive regulatory region located far upstream of the human α-globin gene locus. Genes Dev. *4*, 1588-1601.

Horikoshi, M., Carey, M.F., Kakidani, H., and Roeder, R.G. (1988). Mechanism of action of a yeast activator: direct effect of Gal4 derivatives on mammalian TFIID-promoter interactions. Cell *54*, 665-669.

Horikoshi, M., Hai, T., Lin, Y.S., Green, M.R., and Roeder, R.G. (1988). Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiation complex. Cell *54*(7), 1033-42.

Ikuta, T., and Kan, Y.W. (1991). *In vivo* protein-DNA interactions at the β-globin gene locus. Proc. Natl. Acad. Sci. USA *88*, 10188-10192.

Inostroza, J., Flores, O., and reinberg, D. (1991). Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of general transcription factor IIE. J. Biol. Chem. *266*, 9304-9308.

Ishihama, A. (1988). Promoter selectivity of prokaryotic RNA polymerases. Trends Genet. *4*(10), 282-286.

Jackson, D.P., Evans, T., Nickol, J.M., and Felsenfeld, G. (1989). Developmental modulation of protein binding to β-globin gene regulatory sites within chicken erythrocyte nuclei. Genes Dev. *3*, 1860-1873.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318-356.

Jane, S.M., Ney, P.A., Vanin, E.F., Gumucio, D.L., and Nienhuis, A.W. (1992). Identification of a stage selector element in the human γ-globin gene promoter that fosters preferential interaction of the 5' HS2 enhancer when in competition with the β-promoter. EMBO J. *11*, 2961-2969.

Jantzen, H.-M., Admon, A., Bell, S.B., and Tjian, R. (1990). Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. Nature *344*, 830-836.

Johnson, P.F., and McKnight, S.L. (1989). Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. *58*, 799-839.

Kakidani, H., and Ptashne, M. (1988). Gal4 activates gene expression in mammalian cells. Cell *52*(2), 161-167.

Kayne, P.S., Kim, U.J., Han, M., Mullen, J.R., Yoshizaki, F., and Grunstein, M. (1988). Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. Cell *55*(1), 27-39.

Keleher, K.A., Goutte, C., and Johnson, A.D. (1988). The yeast cell-type-specific repressor alpha2 acts cooperatively with a non-cell-type-specific protein. Cell *53*, 927-936.

Kelleher, R.J., Flanagan, P.M., and Kornberg, R.D. (1990). A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. Cell *61*, 1209-1215.

Kellum, R., and Schedl, P. (1991). A position-effect assay for boundaries of higher order chromosomal domains. Cell *64*, 941-950.

Kim, C.G., Epner, E.M., Forrester, W.C., and Groudine, M. (1992). Inactivation of the β-globin gene by targeted insertion into the β-globin locus control region. Genes and Dev. *6*, 928-938.

Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T., and Pabo, C.O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.1 Å resolution: a framework for understanding homeodomain-DNA interactions. Cell *63*, 579-590.

Koleske, A.J., Buratowski, S., Nonet, M., and Young, R.A. (1992). A novel transcription factor reveals a functional link between the RNA polymerase II CTD and TFIID. Cell *69*, 883-894.

Kollias, G., Hurst, J., deBoer, E., and Grosveld, F. (1987). The human β-globin gene contains downstream developmental specific enhancer. Nucle. Acids Res. *15*(14), 5739-5747.

Kollias, G., Wrighton, N., Hurst, J., and Grosveld, F. (1986). Regulated expression of human Aγ-, β-, and hybrid γβ-globin genes in transgenic

mice: manipulation of developmental expression patterns. Cell *46*, 89-94.

Kustu, S., North, A.K., and Weiss, D.S. (1991). Prokaryotic transcriptional enhancers and enhancer-binding proteins. *16*, 397-401.

Lamb, P., and McKnight, S.L. (1991). Diversity and specificty in transcriptional regulation: the benefits of heterotypic dimerization. *16*, 417-422.

Larson, J.S., Schuetz, T.J., and Kingston, R.E. (1988). Activation in vitro of sequence-specific DNA binding by a human regulatory factor. Nature *335*, 372-375.

Laybourn, P.J., and Dahmus, M.E. (1990). Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before initiation of transcription. J. Biol. Chem. *265*(22), 13165-13173.

Lewis, M.K., and Burgess, R.R. (1982). Eukaryotic RNA polymerases. *In:* The enzymes. Academic Press inc., pp. 109-153.

Lin, Y.S., and Green, M.R. (1991). Mechanism of action of an acidic transcriptional activator in vitro. Cell *64*(5), 971-981.

Lindenbaum, M.H., and Grosveld, F. (1990). An in vitro globin gene switching model based on differentiated embryonic stem cells. Genes Dev. *4*, 2075-2085.

Lu, H., Flores, O., Weinman, R., and Reinberg, D. (1991). The non-phosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. Proc. Natl. Acad. Sci. USA *88*, 10004-10008.

Lucas, P.C., and Granner, D.K. (1992). Hormone response domains in gene transcription. Annu. Rev. Biochem. *61*, 1131-1173.

Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R., and Sigler, P.B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. Nature *352*, 497-505.

Ma, J., and Ptashne, M. (1987). A new class of yeast transcriptional activators. Cell *51*, 113-119.

Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S.C. (1992). DNA recognition by GAL4: structure of a protein-DNA complex. Nature *356*, 408-414.

Martin, D.I.K., and Orkin, S. (1990). Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. Genes Dev. *4*, 1886-1898.

Martin, K.J., Lillie, J.W., and Green, M.R. (1990). Evidence for interaction of different transcriptional activators with distinct cellular targets. Nature *346*, 147-152.

Mathis, D., and Chambon, P. (1981). The SV40 early region TATA box is required for accurate in vitro initiation of transcription. Nature *290*, 310-315.

McClure, W.R. (1985). Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. *54*, 171-204.

McKnight, S.L., and Kingsbury, R. (1982). Transcriptional control signals of a eukaryotic protein-coding gene. Science *217*, 316-324.

Meisterernst, M., and Roeder, R.G. (1991). Family of proteins that interact with TFIID and regulate promoter activity. Cell *67*, 557-567.

Meisterernst, M., Roy, A.L., Lieu, H.M., and Roeder, R.G. (1991). Activation of class II gene transcription by regulatory factors is potentiated by a novel activity. Cell *66*, 981-993.

Mermod, N., O'Neill, E.A., Kelly, T.J., and Tjian, R. (1989). The proline-rich transcriptional activator of CTF/NF1 is distinct from the replication and DNA binding domain. Cell *58*, 741-753.

Mignotte, V., Eleouet, J.F., Raich, N., and Romeo, P.-H. (1989). Cis- and trans-acting elements involved in the regulation of the erythroid promoter of the human porphobilinogen deaminase gene. Proc. Natl. Acad. Sci. USA *86*, 6548-6552.

Morrison, D.K., Kaplan, D.R., Rapp, U., and roberts, T. (1988). Signal transduction from membrane to cytoplasm: growth factors and membrane bound oncogene products increase Raf-1 phosphorylation and associated protein kinase activity. Proc. Natl. Acad. Sci. USA *85*, 8855-8859.

Muller, H.P., Sogo, J.M., and Schaffner, W. (1989). An enhancer stimulates transcription in trans via a protein bridge. Cell *58*, 767-777.

Myers, R.M., Tilly, K., and Maniatis, T. (1986). Fine structure genetic analysis of a β-globin promoter. Science *232*, 613-618.

Nakajima, N., Horikoshi, M., and Roeder, R.G. (1988). Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificty, and TATA box-promoter interactions of TFIID. Molec. Cell. Biol. *8*(10), 4028-4040.

Nardelli, J., Gilson, T., Vesque, C., and Charnay, P. (1991). Base sequence discrimination by zinc-finger DNA-binding domains. Nature *349*, 175-178.

Nevins, J.R. (1992). A closer look at E2F. Nature *358*, 375-376.

Ney, P., Sorrentino, B.P., McDonagh, K.T., and Nienhaus, A.W. (1990). Tandem AP-1-binding sites within the human β-globin dominant control region function as an inducible enhancer in erythroid cells. Genes and Dev. *4*, 993-1006.

Ney, P.A., Sorrentino, B.P., Lowrey, C.H., and Nienhuis, A.W. (1990). Inducibility of the HSII enhancer depends on binding of an erythroid specific nuclear protein. Nucle. Acids Res. *18*(20), 6011-6017.

Nickol, J.M., and Felsenfeld, G. (1988). Bidirectional control of the chicken β- and ε- globin genes by a shared enhancer. Proc. Natl. Acad. Sci. USA *85*(2548-2552),

Pabo, C.O., and Sauer, R.T. (1992). Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. *61*, 1053-1095.

Paro, R. (1990). Imprinting a determined state into the chromatin of Drosophila. Trends Genet. *6*(12), 416-421.

Peterson, C.L., Kruger, W., and Herskowitz, I. (1991). A functional interaction between the C-terminal domain of RNA polymerase II and the negative regulator *SIN1*. Cell *64*, 1135-1143.

Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.-F., D'Agati, V., Orkin, S.H., and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature *349*, 257-260.

Philipsen, S., Talbot, D., Fraser, P., and Grosveld, F. (1990). The β-globin dominant control region: hypersensitive site 2. EMBO J. *9*(7), 2159-2167.

Picard, D., Salser, S., and Yamamoto, K.R. (1988). A movable and regulable inactivation function within the steroid binding domain of the glucocorticoid receptor. Cell *54*(1073-1080),

Picard, D., and Yamamoto, K.R. (1987). Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. EMBO J. *6*(11), 3333-3340.

Popham, D.L., Szeto, D., Keener, J., and Kustu, S. (1989). Function of a bacterial activator protein that binds to transcriptional enhancers. Science *243*(629-635),

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

Ptashne, M. (1986). A genetic switch. Cambridge, Cell Press.

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

Pugh, B.F., and Tjian, R. (1990). Mechanism of transcriptional activation by Sp1: evidence for coactivators. Cell *61*, 1187-1197.

Pugh, B.F., and Tjian, R. (1991). Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes Dev. *5*, 1935-1945.

Pugh, B.F., and Tjian, R. (1992). Diverse transcriptional functions of the multisubunit eukaryotic TFIID complex. J. Biol. Chem. *267*, 679-682.

Qiliang, L., Zhou, B., Powers, P., Enver, T., and Stamatoyannopoulos, G. (1990). β-globin locus actvation regions: conservation of organization, structure and function. Proc. Natl. Acad. Sci. USA *87*, 8207-8211.

Reddy, P.M.S., and Shen, C.-K.J. (1991). Protein-DNA interactions in vivo of an erythroid- specific, human β-globin locus enhancer. Proc. Natl. Acad. Sci. USA *88*, 8676-8680.

Reitman, M., and Felsenfeld, G. (1988). Mutational analysis of the chicken β-globin enhancer reveals two positive-acting domains. Proc. Natl. Acad. Sci. USA *85*, 6267-6271.

Reitman, M., Lee, E., Westphal, H., and Felsenfeld, G. (1990). Site-independent expression of the chicken β^A-globin gene in transgenic mice. Nature *348*, 749-752.

Reitzer, L.J., and Magnasanik, B. (1986). Transcription of glnA in E. Coli is stimulated by activator bound to sites far from the promoter. Cell *45*, 785-792.

Roeder, R.G. (1991). The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *16*, 402-408.

Romeo, P.-H., Prandinin, M.-H., Joulin, V., Mignotte, V., Prenant, M., Vaichenker, W., Marguerie, G., and Uzan, G. (1990). Megakaryocytic and erythrocytic lineages share specific transcription factors. Nature *344*, 447-449.

Rosenfeld, M.G. (1991). POU-domain transcription factors: pou-er-ful developmental regulators. Genes and Dev. *5*, 897-907.

Ryan, T.M., Behringer, R.R., Martin, N.C., Townes, T.M., Palmiter, R.D., and Brinster, R.L. (1989). A single erythroid-specific DNase I super-hypersensitive site activates high levels of human β-globin gene expression in transgenic mice. Genes Dev. *3*, 314-323.

Sawadago, M., and Roeder, R.G. (1984). Energy requirement for specific transcription initiation by the human RNA polymerase II system. J. Biol. Chem. *259*, 5321-5326.

Sawadago, M., and Roeder, R.G. (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., and Sentenac, A. (1990). RNA polymerase B(II) and general transcription factors. Annu. Rev. Biochem. *59*, 711-754.

Schena, M., and Yamamoto, K.R. (1988). Mammalian glucocorticoid receptor derivatives enhance transcription in yeast. Science *241*, 965-967.

Schlief, R. (1992). DNA looping. Annu. Rev. Biochem. *61*, 199-223.

Sheffery, M., Rifkind, R.A., and Marks, P.A. (1982). Murine erythroleukemia cell differentiation: DNase1 hypersensitivity and DNA methylation near the globin genes. Proc. Natl. Acad. Sci. USA *79*, 1180-1184.

Simon, M.C., Fisch, T.M., Benecke, B.J., Nevins, J.R., and Heintz, N. (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),

Smale, S.T., and Baltimore, D. (1989). The initiator as a transcription control element. Cell *57*, 103-113.

Smale, S.T., Schmidt, M.C., Berk, A.J., and Baltimore, D. (1990). Transcriptional activation by Sp1 as directed through TATA or intitiator: specific requirement for mammalian transcription factor IID. Proc. Natl. Acad. Sci. USA *87*(4509-4513),

Stamatoyannopoulos, G., Nienhuis, A.W., Leder, P., and Majerus, P.W. (1987). The molecular basis of blood diseases. Philadelphia, W.B.Saunders Co.

Stief, A., Winter, D.M., Stratling, W.H., and Sippel, A.E. (1989). A nuclear DNA attachment element mediates elevated and position independent gene activity. Nature *341*, 343-345.

Straney, S.B., and Crothers, D.M. (1987). Lac repressor is a transient gene-activating protein. Cell *51*, 699-707.

Strauss, E.C., Andrews, N.C., Higgs, D.R., and Orkin, S.H. (1992). In vivo footprinting of the human α-globin locus upstream regulatory element by guanine and adenine ligation-mediated polymerase chain reaction. Mol. and Cell. Biol. *12*(5), 2135-2142.

Strauss, E.C., and Orkin, S.H. (1992). *In vivo* protein-DNA interactions at hypersensitive site 3 of the human β-globin locus control region. Proc. Natl. Acad. Sci. USA *89*(5809-5813),

Stuve, L.L., and Myers, R.M. (1990). A directly repeated sequence in the β-globin promoter regulates transcription in murine erythroleukemia cells. Mol. Cell. Biol. *10*(3), 972-981.

Talbot, D., and Grosveld, F. (1991). The 5' HS2 of the globin locus control region enhances transcription through the interaction af a multimeric complex binding at two functionally distinct NF-E2 binding sites. EMBO J. *10*, 1391-1398.

Talbot, D., Philipsen, S., Fraser, P., and Grosveld, F. (1990). Detailed analysis of the site 3 region of the human β-globin dominant control region. EMBO J. *9*, 2169-2178.

Tanese, N., Pugh, B.F., and Tjian, R. (1991). Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. Genes Dev. *5*, 2212-2224.

Taylor, I.C.A., and Kingston, R.E. (1990). Factor substitution in a human HSP70 gene promoter: TATA-dependent and TATA-independent interactions. Mol. Cell. Biol. *10*, 165-175.

Taylor, I.C.A., Workman, J.L., Schuetz, T.J., and Kingston, R.E. (1991). Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains. Genes Dev. *5*, 1285-1298.

Townes, T., and Behringer, R.R. (1990). Human globin locus activation region (LAR): role in temporal control. Trends Genet. *6*, 219-223.

Trudel, M., and Costantini, F. (1987). A 3' enhancer contributes to the stage-specific expression of the human β-globin genes in transgenic mice. Genes Dev. *1*, 954-961.

Tuan, D., Soloman, W., Qiliang, L., and London, I.M. (1985). The "beta-like-globin" gene domain in human erythroid cells. Proc. Natl. Acad. Sci. USA *82*, 6384-6388.

Tuan, D.Y.H., Solomon, W.B., London, I.M., and Lee, D.P. (1989). An erythroid specific, developmental-stage-independent enhancer far upstream of the human "β-like globin" genes. Proc. Natl. Acad. Sci. USA *86*, 2554-2558.

Usheva, A., Maldanado, E., Goldring, A., Lu, H., Houbavi, C., Reinberg, D., and Aloni, Y. (1992). Specific interaction between the nonphosphorylated form of RNA polymerase II and the TATA-binding protein. Cell *69*, 871-881.

van Assendelft, G.B., Hanscombe, O., Grosveld, F., and Greaves, D.R. (1989). The β-globin dominant control region activates homologous

and heterologous promoters in a tissue specific manner. Cell *56*, 969-977.

Van Dyke, M.W., Sawadago, M., and Roeder, R.G. (1989). Stasbility of transcription complexes on class II genes. Mol. Cell. Biol. *9*(1), 342-344.

Vinson, C.R., Sigler, P.B., and McKnight, S.L. (1989). A scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science *246*, 911-916.

Vyas, P., Vickers, M.A., Simmons, D.L., Ayyub, H., Craddock, C.F., and Higgs, D.R. (1992). Cis-acting sequences regulating expression of the human α-globin cluster lie within constitutively open chromatin. Cell *69*, 781-793.

Wall, L., deBoer, E., and Grosveld, F. (1988). The human β-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. Genes Dev. *2*, 1089-1100.

Waslyk, B., Derbyshire, R., Guy, A., Molko, D., Roget, A., Teoule, R., and Chambon, P. (1980). Specific in vitro transcription of conalbumin gene is drastically decreased by single-point mutation in T-A-T-A box homology sequence. Proc. Natl. Acad. Sci. USA *77*, 7024-7028.

Wedel, A., Weiss, D.S., Popham, D., Droge, P., and Kustu, S. (1990). A bacterial enhancer functions to tether a transcriptional activator near a promoter. Science *248*, 486-490.

Weintraub, H. (1985). Assembly and propagation of repressed and derepressed chromosomal states. Cell *42*, 705-711.

Wiley, S.R., Kraus, R.J., and Mertz, J.E. (1992). Functional binding of the "TATA" box binding component of transcription factor TFIID to

the -30 region of TATA-less promoters. Proc. Natl. Acad. Sci. USA *89*, 5814-5818.

Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D., and Pabo, C.O. (1991). Crystal structure of a Mat a2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell *67*, 517-528.

Workman, J.L., Abmayr, S.M., Cromlish, W.A., and Roeder, R.G. (1988). Transcriptional regulation by the immediate early protein of pseudorabies virus during *in vitro* nucleosome assembly. Cell *55*(2), 211-9.

Workman, J.L., and Roeder, R.G. (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*(4), 613-22.

Workman, J.L., Roeder, R.G., and Kingston, R.E. (1990). An upstream transcription factor, USF (MLTF), facilitates the formation of preinitiation complexes during *in vitro* chromatin assembly. Embo J *9*(4), 1299-308.

Workman, J.L., Taylor, I.C., and Kingston, R.E. (1991). Activation domains of stably bound GAL4 derivatives alleviate repression of promoters by nucleosomes. Cell *64*(3), 533-44.

Yamamoto, K.K., Gonzalez, G.A., Biggs, W.H., and Montimony, M.R. (1988). Phosphorylation-induced binding and transcriptional efficacy of nuclear factor CREB. Nature *334*, 494-498.

Young, R.A. (1991). RNA polymerase II. Annu. Rev. Biochem. *60*, 689-715.

Zaret, K.S., and Yamamoto, K.R. (1984). Reversible and persistent changes in chromatin structure accompany activatioon of a glucocorticoid dependent enhancer element. Cell *38*, 29-38.

CHAPTER 2

DISCRIMINATION AMONG POTENTIAL ACTIVATORS OF THE β-GLOBIN CACCC ELEMENT BY CORRELATION OF BINDING AND TRANSCRIPTIONAL PROPERTIES

INTRODUCTION

The adult β-like globin promoters contain several elements whose sequence and location are highly conserved (Efstratiadis, et al., 1980). One of these is the CACCC element, CCA/TCACCC (Table 1). A large body of work has demonstrated that this element is required for wild type levels of transcription of transfected β-like globin genes in HeLa, L and 3T6 cells (Dierks, et al., 1983; Myers, et al., 1986; Wright, et al., 1984). Similarly, this element is also required for transcription of transfected β-globin genes in erythroid cells (Cowie and Myers, 1988; Stuve and Myers, 1990). More striking, several different naturally occurring single base substitution (point) mutations in the CACCC element of the human β-globin promoter that are known to decrease transcription cause β-thalassemia (Gonzalez-Redondo, et al., 1989; Kulozik, et al., 1991; Orkin, et al., 1984; Ristaldi, et al., 1990). Also, there are indications that the β-globin promoter CACCC element may be involved in mediating the effects of the far upstream Locus Control Region (LCR; Antoniou and Grosveld, 1990).

Transcription of several other genes depends upon sequences that resemble or are identical to the β-globin CACCC site. These include the SV40 enhancer, the erythroid specific GATA-1 transcriptional activator gene, the tryptophan oxidase gene and the porphobilinogen deaminase gene (Frampton, et al., 1990; Schule, et al., 1988; Tsai, et al., 1991; Xiao, et al., 1987). These genes, and other CACCC element containing genes, are active in many different cell types. In some cases, it is known that the CACCC element is required for the activity of an additional nearby cis element. For example, in

Table 1

ALIGNMENT OF CACCC ELEMENTS FROM DIFFERENT GENES [a]

Adult β-Globin Genes	
	101
Human β [b]	C A G A C C T C A C C C T G
	A G A G C C A C A C C C T A
	88 87
Mouse β major	A G A G C C A C A C C C T G
Mouse β minor	G A A G C C T C A C C C T G
Rabbit β1	C A G A C C T C A C C C T G
	A G A G C C A C A C C C T G
Goat βA and βC	C A A G C C T C A C C C T G
Chicken β	T G A G C C C C A C C C T G
Other Genes with CACCC Elements	
Tryptophan Oxidase	A G A G C C A C A C C C A G
Porphobilinogen Deaminase	C A G G C C C C A C C C T C
SV40 GT-1C Motif	C T T T C C A C A C C C T A
GATA-1	G G C G G C A C A C C C C C
	A C T G C C C C A C C C A C

a. The sequences shown here were derived from the following references: Tryptophan Oxidase (45), SV40 GT-1C (53), GATA-1 (50), all others in (12). The boxed sequences are those that are most highly conserved and required for transcription.

b. The human CACCC element is duplicated. Naturally occuring, individual C to T point mutations at positions -101,-88,-87 or a C to G mutation at position -87 result in β-thalassemia.

the tryptophan oxidase promoter, a CACCC element and a glucocorticoid response element (GRE) occur within 10 nucleotides of one another, and transcriptional induction of this gene by glucocorticoids occurs only if both the GRE and the CACCC element are intact (Schule, et al., 1988). Additionally, the tryptophan oxidase GRE and CACCC element must be spaced an integral number of 10 nucleotides apart for maximal activity, suggesting that proteins binding these elements interact in a specific spatial orientation to activate transcription cooperatively. A similar interaction appears to occur in the porphobilinogen deaminase promoter between a CACCC element and a GATA-1 binding site (Frampton, et al., 1990).

The above data suggest that the CACCC element serves as a binding site for a factor or factors, widely distributed across tissues, that activates transcription and that likely co-operates with other transcriptional regulatory proteins. For these reasons, many laboratories have been interested in characterizing potential transcriptional regulatory proteins that bind to this sequence element and a number of such putative factors have been identified in many different cell types (Catala, et al., 1989; Davidson, et al., 1988; de Boer, et al., 1988; Giglioni, et al., 1989; Gumicio, et al., 1988; Lewis, et al., 1988; Mantovani, et al., 1988; Mitchell, et al., 1987; Spanopoulou, et al., 1991; Xiao, et al., 1987). Because we are interested in the regulation of β-globin gene transcription, we have studied CACCC binding proteins from MEL cells, a mouse erythroid precursor immortalized at the pre-erythroblast stage (Marks and Rifkind, 1978).

When treated with any of a variety of chemical agents, MEL cells differentiate into an erythrocyte-like cell. During this process, adult globin genes and other erythroid-specific genes are transcriptionally activated. We have found that several CACCC binding activities occur in MEL nuclear extracts, in agreement with previous reports (de Boer, et al., 1988; Mantovani, et al., 1988). Although these DNA-binding activities can be distinguished by their mobility's in native gels, they share many other properties. We sought to differentiate between these binding activities and deduce which activity(s) is most likely to activate β-globin transcription from the CACCC element in vivo.

Given the rapid expansion in the number of known DNA-binding transcriptional regulatory proteins, many of which bind to similar sites, our problem of differentiating between several different activities binding a single site is not unusual (Hayashi and Scott, 1990; Johnson and McKnight, 1989). To address this problem, we first looked for a correlation between the ability of CACCC binding proteins to bind to β-globin promoters with mutations in or around the CACCC element and the ability of these mutant promoters to activate transcription. Three CACCC element binding species bound less well to β-globin promoters bearing mutations in the CACCC element that decreased transcription; this is behavior that might be expected of a transcriptional regulatory protein acting through this site. Two of these potential activators were examined in greater detail. We demonstrated that one CACCC element binding protein is the transcriptional activator SP1. A second candidate activator was subjected to a site selection procedure to find its preferred high-

affinity binding site. The results of this experiment lead us to believe that it is this activity, rather than SP1, that is most likely to be the transcriptional regulator that binds the CACCC element and activates transcription of the adult β-globin genes. Finally, we suggest that this combination of binding, transfection and site selection experiments may provide a general approach for differentiating between multiple proteins binding to a particular cis element.

MATERIALS AND METHODS

DNAs

The β-globin promoter was assayed in the context of a fusion of a minimal β-globin promoter (nucleotides -106 to +26) to the mouse metallothionein-1 gene (nucleotides +65 to +2200). Most of the mutant β-globin promoters, as well as the histone H4 reference gene, were derived from previous work of this lab, however several other new mutations were generated for this work (Myers, Tilly and Maniatis, 1986; Stuve and Myers, 1990). These mutations, -95A, -94T, -92C, -92T, -90T and -92C/-93T, were all constructed by site directed mutagenesis (Kunkel, et al., 1987).

Transfections

MEL cells ($aprt^-$; (Deisseroth and Hendrick, 1978)) were grown in Dulbecco's modified Eagle medium (DME; Irvine Scientific) supplemented with 10% fetal calf serum (HyClone Laboratories, Inc.) and penicillin-streptomycin (100 U/ml penicillin G, 100 ug/ml Streptomycin SO_4). Cells were transfected by the DEAE dextran

method, and where indicated, induced to differentiate exactly as described in (Cowie and Myers, 1988).

RNA analysis

Cytoplasmic RNA was isolated from the transfected cells by the method of Favalaro et al. (Favalaro, et al., 1980). The β-metallothionein and H4-metallothionein fusion RNAs were quantitated by an S1 nuclease protection assay as described in (Cowie and Myers, 1988). Single stranded, end labeled probe DNAs complementary to the expected products of the transfected genes were made as described previously (Cowie and Myers, 1988; Stuve and Myers, 1990); See Figure 1A). Signals were visualized byautoradiography and were quantitated on a Phosphorimager (Molecular Dynamics). An example of the data collected is shown in Figure 1B.

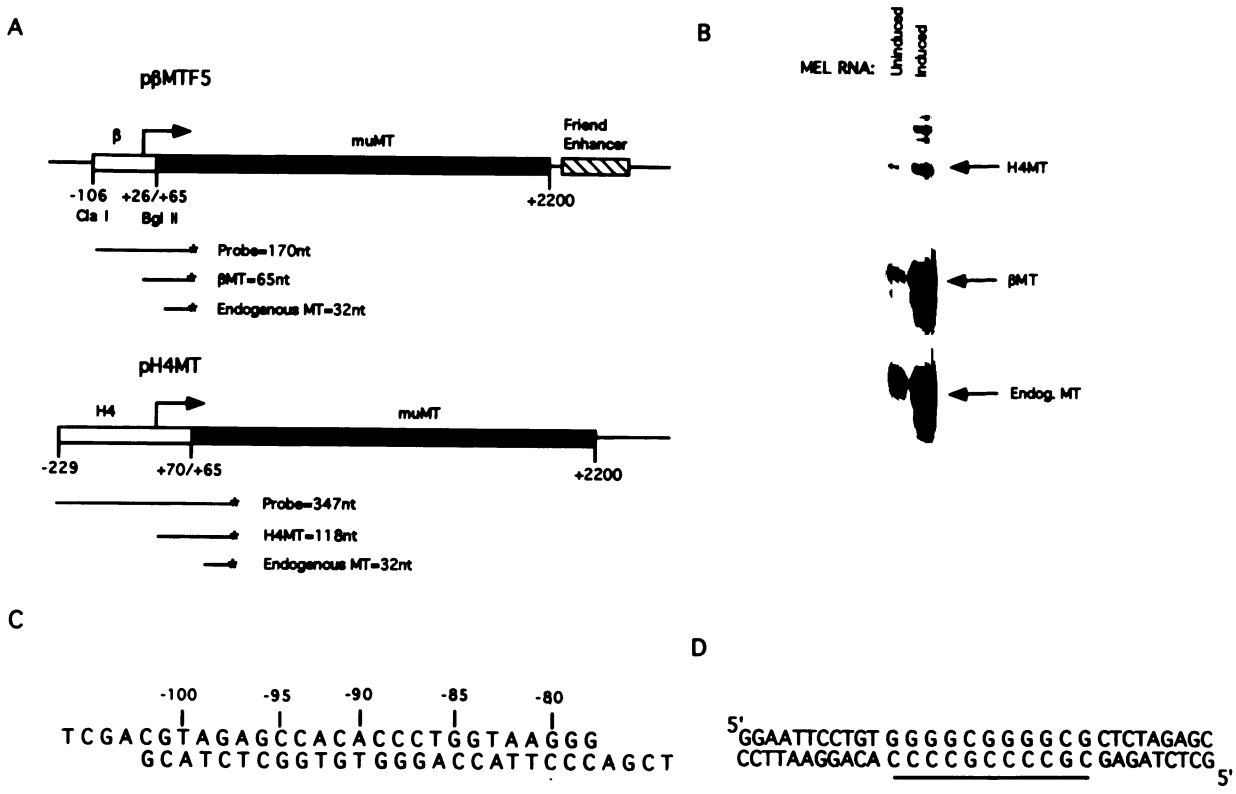
DNA Binding Assays

1. Probes

The CAC1 probe was generated by cloning an oligonucleotide (Figure 1C) into the XhoI site of plasmid pSP72 (Krieg and Melton, 1987). This cloned product was excised from the plasmid by digesting with Eco RI and Xba I and was ^{32}P labeled by filling in the overhanging ends created by the restriction enzymes with Klenow enzyme in the presence of α-^{32}P-dATP (3000 Ci/mmol, Amersham). The labeled DNA was then gel-purified in 6% native polyacrylamide gels, eluted in crush soak buffer overnight at 37°C and ethanol precipitated (Maxam and Gilbert, 1980). Probes were stored in 10

Figure1 Plasmid DNAs used in transfection and mobility shift experiments.

(A) Chimeric β-globin/mouse metallothionein-1 and histone H4/mouse metallothionein-1 constructs (Cowie and Myers, 1988; Stuve and Myers, 1990) that were transfected into MEL cells. The single stranded DNA probes used in the S1 nuclease protection assay are shown below each construct. In addition to detecting the transcript to which they are directed, each probe detects a small portion of the transcript from the other transfected construct in the region that is homologous to the mouse metallothionein-1 transcript as well as the endogenous metallothionein-1 transcript itself. In both cases, this extra signal is detected as a 32 nucleotide DNA fragment. (B) An example of the S1 nuclease assay. Cells were transfected with pβMTF5 and pH4MT, and a portion of the cells were subsequently induced to differentiate. RNA was isolated and subjected to the S1 nuclease protection assay in the presence of both the βMT and H4MT probes. The products of this assay were fractionated on a 50% urea-polyacrylamide gel and are shown here. The bands above the H4MT signal are due to upstream starts in plasmid pH4MT. (C) Sequence of the CACCC oligonucleotide used to generate probe CAC1 and also used in competition assays. This oligonucleotide contains the mouse β-major-globin promoter sequence between nucleotides -78 and -102 and is numbered accordingly. (D) Sequence of the oligonucleotide with a high affinity SP1 binding site used as a competitor and, when end labeled with ^{32}P, as a probe in mobility shift assays. The SP1 binding site is underlined.


A
pβMTF5
β
muMT
Friend Enhancer
-106
Cla I
+26/+65
Bgl II
+2200
Probe=170nt
βMT=65nt
Endogenous MT=32nt
pH4MT
H4
muMT
-229
+70/+65
+2200
Probe=347nt
H4MT=118nt
Endogenous MT=32nt
B
MEL RNA:
Uninduced
Induced
H4MT
βMT
Endog. MT
C
-100
-95
-90
-85
-80
TCGACGTAGAGCCACACCCTGGTAAGGG
GCATCTCGGTGTGGGACCATTCCCAGCT
D
5'GGAATTCCTGT GGGGCGGGGCG CTCTAGAGC
CCTTAAGGACA CCCCGCCCCGC GAGATCTCG 5'

mM tris pH 8, 1 mM EDTA and 100 mM NaCl. β-globin promoter probe DNAs were prepared as above using Bgl II and Cla I to remove the 132 nucleotide β-globin promoter (-106 to +26) fragment from the pβMTF5 plasmid and its mutant derivatives.

For DNase I footprinting, CAC1 probes labeled at a specific end were generated by cutting with either Eco RI or Xba I alone, treating with calf intestinal alkaline phosphatase (Boehringer), and end-labeling with γ-^{32}P-ATP (7000 Ci/mmol, ICN) and polynucleotide kinase (NEB). The DNA was subsequently digested with the other restriction enzyme to release the end labeled probe and then gel purified.

The SP1 probe was generated using γ-^{32}P-ATP and polynucleotide kinase to end-label a double stranded oligonucleotide bearing a high affinity SP1 binding site (Kadonaga, et al., 1986; Figure 1D).

The sequence of the two strands of the NF-1 binding site oligonucleotide used as a competitor in figure 2A is 5'GATCCTTTTGGATTGAAGCCAATATGAT and 5'CTAGATCATATTGGCT TCAATCCAAAAG (Jones, et al., 1987). Annealing these oligonucleotides produces a 5'-GATC overhang at each end.

2. Mobility Shift Assay

For mobility shift assays, protein samples were mixed with 1-12 fmol probe DNA in a buffer consisting of 2 mM MgCl, 10% Ficoll 400 (Pharmacia), 20 mM Tris-HCl pH 8, 0.4 mM EDTA in a final reaction volume of 20 µl. 2 µg non-specific competitor DNA (poly dI·dC; Sigma) was included when crude protein preparations were

used, whereas no competitor was used in reactions with partially purified CACCC species D or with pure SP1. Assays of purified SP1 or partially pure CACCC species D included 0.5 mg/ml BSA (Cal Biochem, nuclease and protease free) as a carrier protein; in the case of partially purified CACCC species D, no mobility shift was observed without carrier protein. Binding reactions were incubated at room temperature for 5-20 minutes and loaded onto a 6%, 39:1 (acrylamide: bis-acrylamide) crosslinked polyacrylamide gel. The gel was run at room temperature at approximately 6 volt/cm in a buffer consisting of 50 mM Tris-HCl pH 8.5, 190 mM glycine (BioRad) and 1 mM EDTA (Ausubel, et al., 1991). For assays with specific competitor DNAs, the protein sample was added after the probe and competitor DNAs were mixed; the reaction was then incubated for at least 20 minutes at room temperature before being loaded on the gel.

The proteolytic clipping band shift assay (Schreiber, et al., 1988) was performed by combining 50 µl of purified SP1 or partially pure CACD with 10 µg BSA carrier protein in 1 mM $CaCl_2$. 50 ng proteinase K (Boehringer Mannheim) was added and the proteins were allowed to digest at room temperature. At intervals of 2, 5 and 10 minutes, 10 µl aliquots were withdrawn and treated with PMSF to inhibit the proteinase K. Four µl each of these proteolyzed samples was then subjected to mobility shift analysis.

3. DNase 1 Footprinting

DNase 1 footprinting assays were carried out in conjunction with the mobility shift assay. The mobility shift assay allowed the separation of the different CACCC binding species in crude extracts as

well as easy separation of bound and unbound probe. For the DNaseI reactions, a typical mobility shift reaction was scaled up in size 3-5 times. After binding for 10 minutes at room temperature, the reaction was brought to 1.25 mM in $CaCl_2$ and immediately treated with 5 to 100 ng DNase I (Worthington) for 1 minute at room temperature. EGTA was added to 1.25 mM and the samples were immediately loaded onto a mobility shift gel. After electrophoresis, the free and shifted probe bands were excised and purified as described in (Ausubel, et al., 1991). The eluted DNAs were fractionated on 10% polyacrylamide-50% urea gels and visualized by autoradiography.

Proteins

MEL cells were grown in suspension at 0.5 to 1.5 x 10^6 cells per ml in Jokliks MEM (Irvine Scientific) supplemented with 5% newborn calf serum (Gemini Biosciences) and penicillin-streptomycin. HeLa cells were grown in the same media at a density of 5 to 8 x 10^5 cells per ml. L cells were grown in Jokliks MEM, penicillin-streptomycin and 10% calf serum (Hyclone) at approximately 5 x 10^5 cells per ml. A204 cells were grown in RPMI (Irvine Scientific), penicillin-streptomycin and 10% fetal calf serum at 5 x 10^5 cells per ml.

Nuclear extracts were prepared as in Dignam et al. (Dignam, et al., 1983). All buffers used in the preparation of nuclear extracts and in the purification of proteins included 0.1% NP-40 (Sigma), 1 mM DTT, 0.5 mM phenylmethylsulfonyl fluoride (PMSF; Sigma), 1 mM Benzamidine (Sigma), 1 μg/ml aprotinin (Boehringer), 1 μg/ml Leupeptin (Boehringer) and 1 μg/ml Pepstatin A (Sigma). All

manipulations of protein were carried out at 0-4°C. Proteins were stored at -70°C.

CACCC species D (CACD) was purified from MEL cell nuclear extracts by a combination of conventional and site specific DNA affinity chromatography. This data will be presented in detail elsewhere (GAH and RMM in preparation).

SP1 was purified from MEL cells by a combination of wheat germ agglutinin agarose and site-specific DNA affinity chromatography as described by Jackson and Tjian (Jackson and Tjian, 1989). The protein produced by this procedure was 95% pure SP1 protein as judged by silver staining and was authentic SP1 as judged by immunoblot analysis.

Site Selection Assay

The site selection assay follows the protocol described by Blackwell and Weintraub (Blackwell and Weintraub, 1990). Briefly, an oligonucleotide, 5'-TCCGAATTCCTACAGN_{21}TGCAATGGATCCGTCT, with 21 random internal nucleotides and flanking primer sites was synthesized. The oligonucleotide was made double stranded by annealing primer A (5'-AGACGGATCCATTGCA), and primer extending with Klenow enzyme in the presence of 60 μCi α-^{32}P-dATP and 0.5 mM each of the 3 other unlabeled dNTPs. The probe was gel purified as above and, because it was internally labeled, it was used as soon as possible after gel purification. The amount of probe was quantitated by counting the amount of incorporated ^{32}P and assuming a random distribution of adenines in the internal, random sequences. Mobility shift reactions were performed with 4 fmol of

the labeled probe and the most highly purified fractions of CACD (see below). Marker lanes consisting of CACD binding reactions with the CAC1 probe were also run. After electrophoresis, gels were dried and exposed to film. The region of the gel containing probe DNA migrating at the same mobility as the CACD shift with the CAC1 probe was excised and the DNA was eluted by the crush soak method and collected by ethanol precipitation.

The recovered probe was immediately amplified by PCR by using primer A and primer B (5'-TCCGAATTCCTAG). PCR conditions were: 2.5 nM each primer, 0.1 mM dNTPs, 1.5 mM Mg^{2+}, 50 mM KCl, 10 mM Tris-HCl pH 8.4 and 100 μg/ml gelatin and 0.5 μl Taq DNA polymerase (Cetus-Perkin Elmer) in 100 μl total volume. Samples were put through 30 cycles of: 94°C for 20 sec, 55°C for 20 sec and 72°C for 30 sec. After PCR, the products were separated from the unincorporated nucleotides and primers on a Sephadex G-25 (Pharmacia) spin column and concentrated by ethanol precipitation in the presence of 10 μg carrier tRNA. These products were labeled with ^{32}P and made double stranded as above except that both primers were now included since the starting template was double stranded. Subsequent mobility shift reactions were performed with lower amounts of probe and protein to select for higher affinity protein-DNA interactions. After five rounds of selection, the PCR products were digested with Eco RI and Bam HI and purified on a G-25 spin column. These products were ligated into Eco R1/Bam H1 digested pSP72. Sequences of individual clones were determined by the dideoxy chain termination method.

Antibodies

The anti-SP1 antibodies used in this work were kindly provided by the Tjian laboratory. One, 2892, is a polyclonal rabbit antiserum produced against SP1 protein made in *E. coli*. The other, 2990, is a rabbit polyclonal antiserum produced against a peptide comprising most of SP1 zinc finger 2.

Immunoblot Analysis

Immunoblots were performed using standard protocols (Harlow and Lane, 1988). Proteins were separated on 10% polyacrylamide-SDS gels and electrophoretically transferred to 0.2 μm nitrocellulose (Schleicher and Schule). The filter was blocked using 2% gelatin (BioRad, EIA grade) and was probed with antibody 2892 at a 1:1000 dilution. The filter was subsequently probed with an anti-rabbit alkaline-phosphatase conjugated antibody (Promega) and was developed using the Promega Protoblot reagents according to their instructions.

RESULTS

DNA Binding Studies

To identify potential transcriptional trans-activating proteins that bind the CACCC element, we used the gel mobility shift assay. When a ^{32}P-labeled, double-stranded DNA probe (CAC1) containing nucleotides -101 to -78 of the mouse β-globin promoter was mixed with MEL nuclear extracts and electrophoresed on a native gel, four shifted species, labeled A-D, were observed (Figure 2A). Occasionally, we observed species that migrated with a greater

mobility than species D. These occurred more commonly in extracts made without protease inhibitors and in extracts that were several months old. Thus, we assume that these higher mobility bands represent proteolytic artifacts (See also Figure 7A).

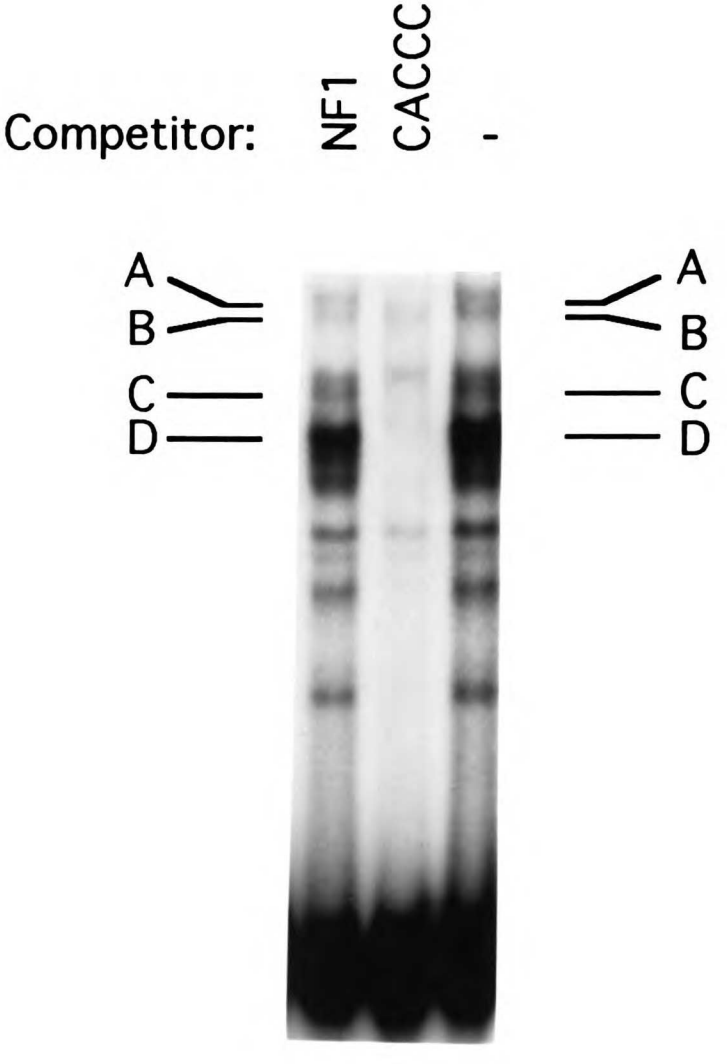
Competition with an excess of the unlabeled, double stranded CACCC oligonucleotide abolished each of the 4 CACCC binding species. Species D was much more sensitive to competitor than species A-C (Figure 2A, Figure 2B and see below). None of the shifted species was affected by competition with a non-specific DNA fragment (Figure 2A, NF1 competitor). When the entire mouse β-globin promoter proximal region, from nucleotide -106 to +26, was used as a probe, essentially the same pattern of shifts occurred (See below). Each shifted species observed with the promoter probe behaved in competition assays like the species of similar mobility observed with the CAC1 probe (data not shown). Therefore, under our mobility shift conditions with both the promoter and CAC1 probes, four CACCC binding activities were detected.

To ask if CACCC binding activities change during the course of chemically-induced differentiation of MEL cells, mobility shift assays were performed with the CAC1 probe and extracts from MEL cells that had been induced for various periods of time (Figure 2B). Although the total amount of CACCC binding activity increased slightly during induction, we did not observe any large changes in the relative abundance of any one of the binding species nor did we observe any new bands indicative of an inducible CACCC binding activity. Also, nuclear extracts from L cells, HeLa cells and the rhabdomyosarcoma cell line A204 gave the same spectrum of site

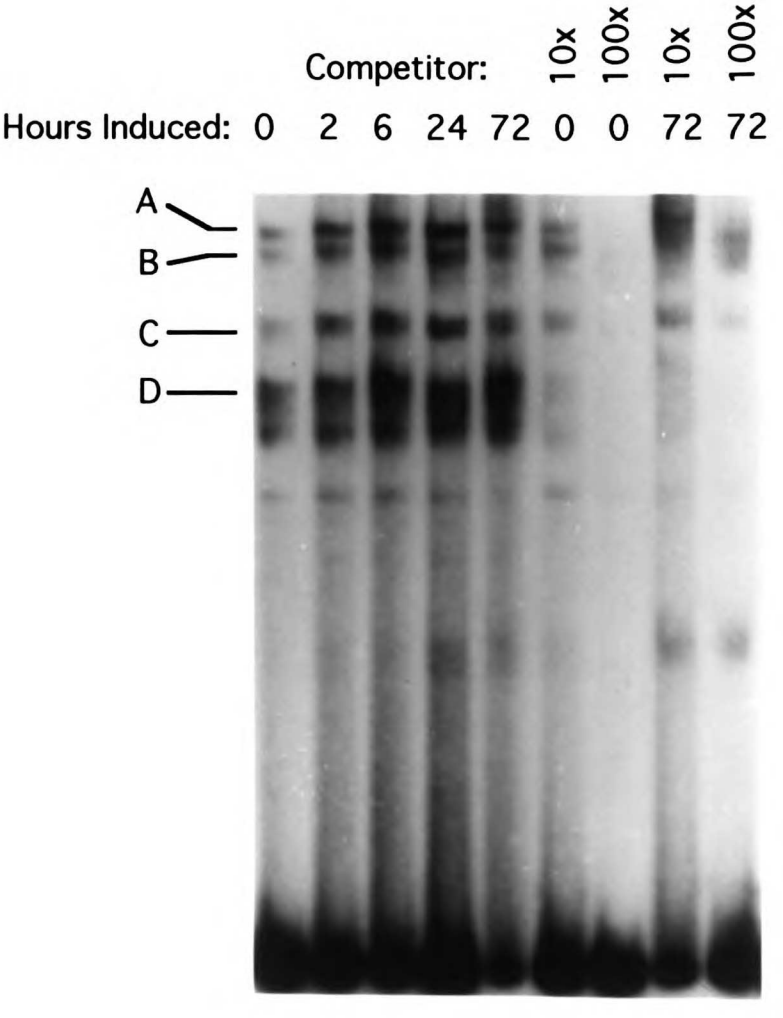
Figure 2. CACCC mobility shift assay.

(A) When crude MEL nuclear extracts are mixed with the CAC1 probe and separated on a native gel, several bands are observed. These bands are competed by a 100 fold molar excess of a specific (CACCC) but not by a non-specific (NF1) competitor. The four upper bands that are routinely observed with crude MEL nuclear extracts are labeled A-D. The species migrating faster than species D are variable. Band C sometimes appears to be dimeric and other times monomeric. (B) Mobility shift analysis of nuclear extracts of induced MEL cells. MEL cells were treated with DMSO for 0, 2, 6, 24 or 72 hours and then used to prepare nuclear extracts. Extracts from cells that were uninduced or cells that were induced for 72 hours were used for competition assays. Addition of a 10 fold or 100 fold molar excess of the CACCC oligonucleotide over the CAC1 probe demonstrates that the CACCC binding activities in the induced extract behave similarly to those in the uninduced extract. (C) Nuclear extracts from other cell types contain CACCC binding activities similar to those present in MEL cells. Nuclear extracts from the fibroblast cell lines HeLa and L cells, the human rhabdomyosarcoma line A204 and MEL cells were subjected to mobility shift analysis with the CAC1 probe.

A

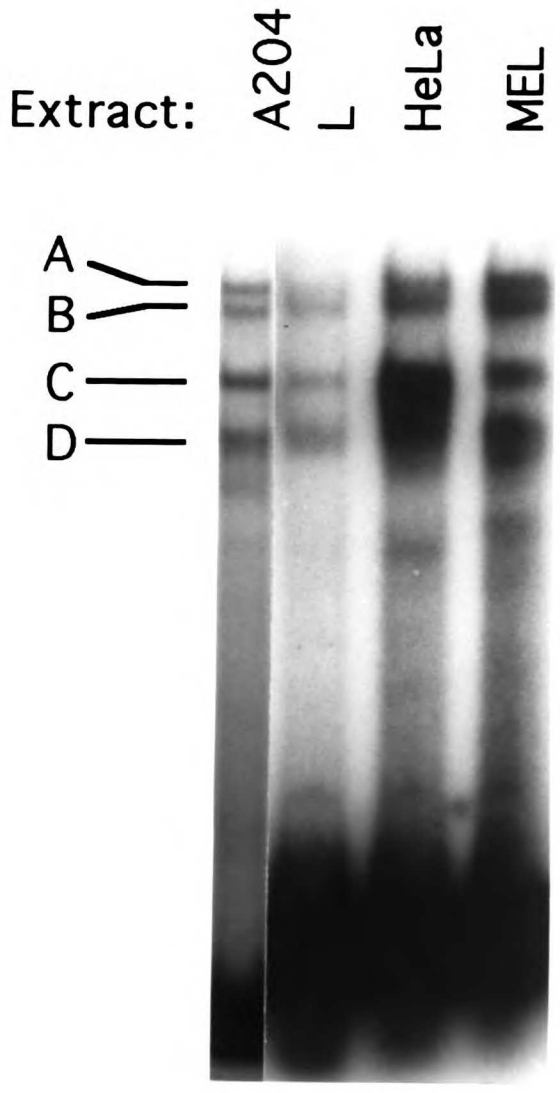

Competitor:
NF1
CACCC
-
A
B
C
D
A
B
C
D

B


Competitor:
10x
100x
10x
100x
Hours Induced: 0 2 6 24 72 0 0 72 72
A
B
C
D

100

C



specific CACCC binding activities as MEL cells, although the relative abundance of these binding activities varied between extracts (Figure 2C). Thus, we observed no inducible or erythroid specific CACCC element binding activity.

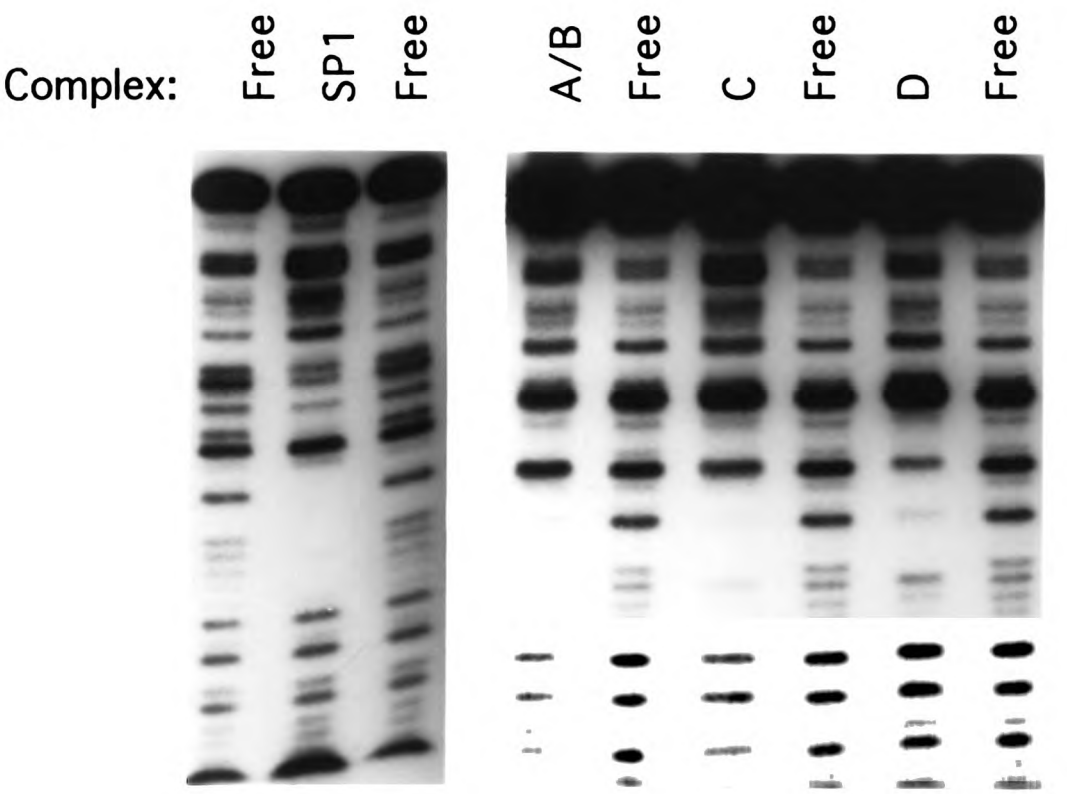
The four CACCC binding activities were further analyzed by DNase1 footprinting. Binding reactions with crude MEL nuclear extracts and end labeled CAC1 probe were assembled and treated briefly with DNase1. These reactions were electrophoresed on mobility shift gels to separate species A-D and the shifted probe DNAs were eluted and then separated on sequencing gels (Figure 3). Because species A and B run near one another on the mobility shift gels, these species could not be separated and were eluted as a single entity. Therefore, the footprint for species A and B represents the composite of two individual footprints. The data from these experiments reveal that species A to D have very similar but not identical interactions with the CACCC binding site.

Point mutant binding studies

Because it appears that multiple factors with similar properties bind the CACCC element, we needed some way to differentiate these potential transactivators. We reasoned that if a particular CACCC binding species activates β-globin transcription, its ability to bind a mutant β-globin promoter fragment might parallel the ability of that mutant promoter to activate transcription. While mutations in the CACCC element that decrease transcription need not necessarily decrease binding of a CACCC transactivator protein, we expected that mutations that do not decrease transcription should not decrease

Figure 3. DNAse 1 footprinting.

MEL nuclear extracts and CAC1 probe were used to set up mobility shift reactions. Before loading onto the gel, the sample was treated briefly with DNAse 1. After electrophoresis, the bands corresponding to bound and free probe were cut out and the DNA was eluted and fractionated on a denaturing gel. Reactions using CAC1 probe labeled at the 5' end of either the sense or antisense strand are shown. Purified SP1 protein was also footprinted by this method.

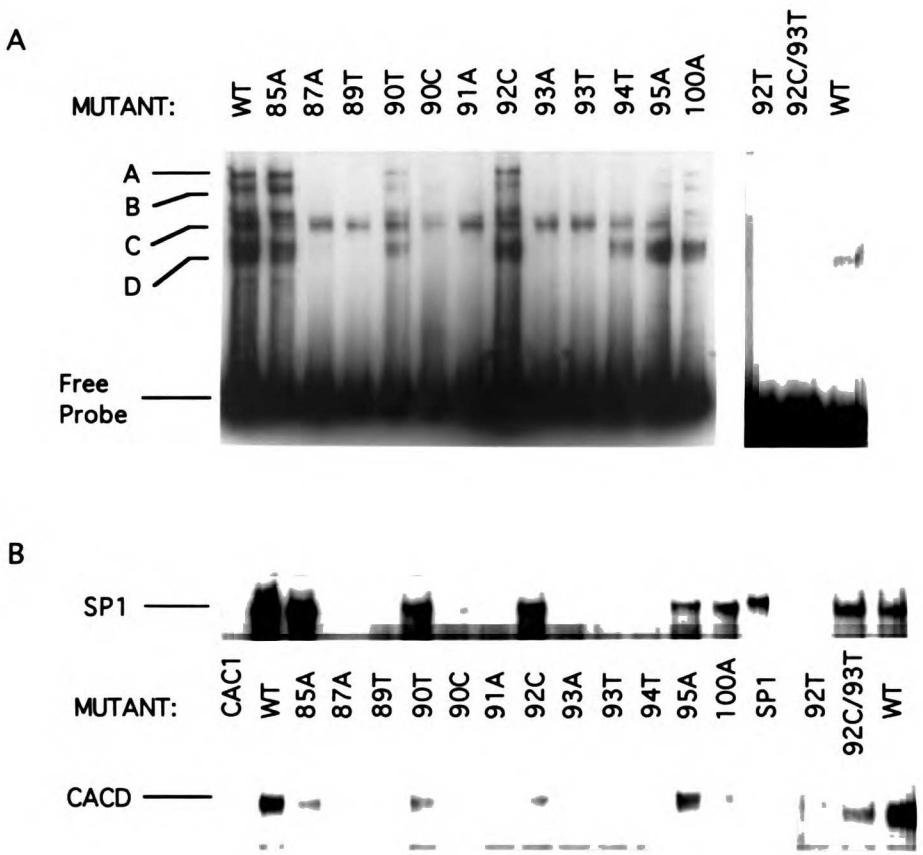

Complex:
Free
SP1
Free
A/B
Free
C
Free
D
Free

binding of a CACCC binding transactivator protein. Using a fragment of the β-globin promoter including nucleotides -106 to +26 as a probe in the mobility shift assay, we asked whether promoters with point mutations in or around the CACCC sequence could bind proteins with the same relative affinity as the wild type sequence.

With crude nuclear extracts, the mobility shifts observed with the mutant promoter probes could be placed into several different classes (Figure 4A). One class of mutants gave rise to shifted species that were identical to those observed with the wild type probe (Figure 4A, WT) or to a pattern that had only slightly decreased levels of species A, B and D. This class of mutations includes -100A, -95A and -85A, which flank the CACCC site, as well as -92C and -90T, which fall within the conserved sequence, CCACACCC. The second class of mutant promoters bound negligible amounts of species A, B and D. This class includes mutations -87A, -88T, -89T, -91A, -92T, -93A and -93T. Three mutants that do not fit into the two classes above are -90C, -94T and -92C/-93T. Mutation -90C bound species B and C at reduced levels and did not bind species A or D. In contrast, mutant -94T bound species C and D but not species A or B. Mutation -92C/-93T, increases binding of species A and C, does not affect binding of species B and decreases binding of species D. Finally, only mutations -92C/-93T and -92T seemed to affect the affinity of the promoter probes for species C. Overall, these results are reminiscent of those for experiments that observed binding of HeLa and F9 cell proteins to the SV40 GT1 motif, which is similar to the CACCC element (Xiao, et al., 1987); Table 1). However, in contrast to this

Figure 4. Mobility shift analysis of point mutant promoter probes.

The 132 nucleotide Cla I/Bgl II fragment of the wild type and mutant β-globin promoters were labeled with ^{32}P to similar specific activities. Three fmole each of these probes were used in the mobility shift assay with (A) crude nuclear extracts of MEL cells, (B) pure SP1 protein and (C) partially purified CACD protein. In (B) and (C) only a single shifted species of invariant mobility was observed for either protein on any of the probes and only this species is shown here. Figure 6 shows an entire mobility shift with the protein preparations used in this figure. Mutants -92T and -92C/93T were assayed in a separate experiment and are shown alongside a reaction with the wild type probe that was performed at the same time for comparison. In (B), SP1 binding to the CAC1 and SP1 probes, in addition to the point mutant promoter probes, was also assayed. Mutant -88T gave mobility shifts identical to those observed for mutants -87A and -89T (data not shown).


A
MUTANT:
WT
85A
87A
89T
90T
90C
91A
92C
93A
93T
94T
95A
100A
92T
92C/93T
WT
A
B
C
D
Free
Probe
B
SP1
CAC1
MUTANT:
WT
85A
87A
89T
90T
90C
91A
92C
93A
93T
94T
95A
100A
SP1
92T
92C/93T
WT
CACD

previous study, several mutations that we examined (but that they did not) helped to differentiate between the binding specificities of the 4 CACCC binding activities. These include, -90T,-94T, -95A and -92C/-93T.

Transfections

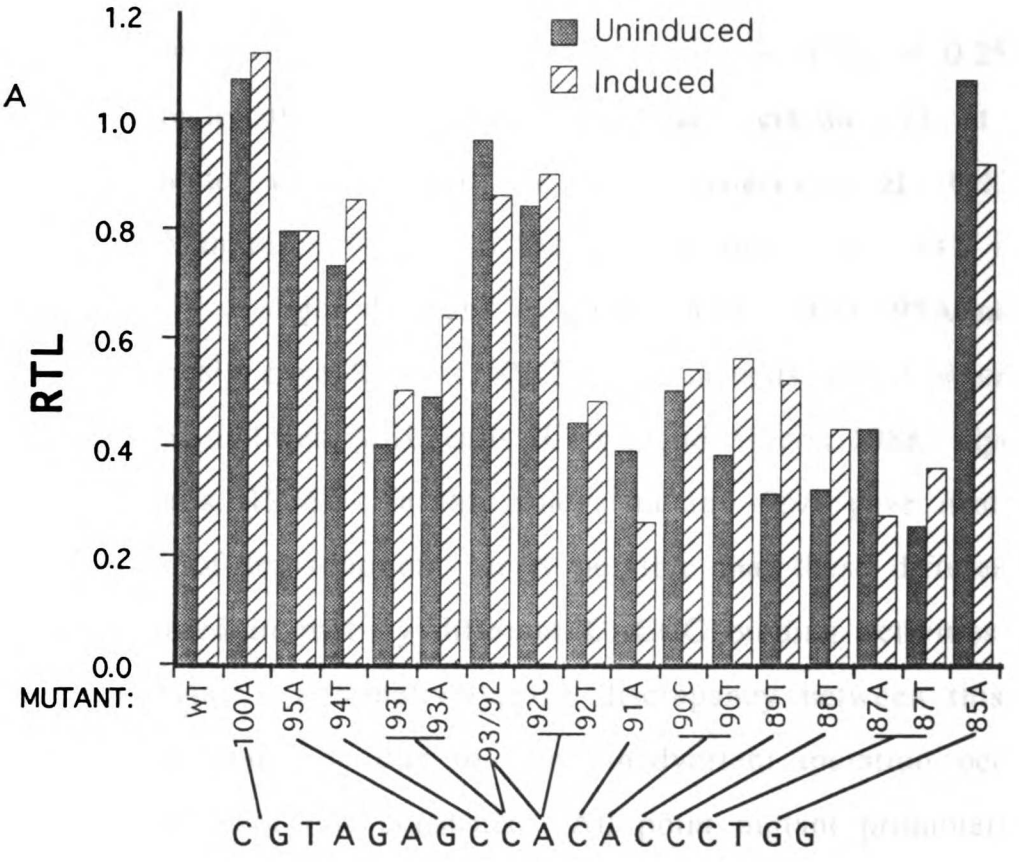
Although previous studies of mutations in the β-globin CACCC element have examined the transcriptional effects of point mutations at nearly every nucleotide of this conserved sequence, these studies were carried out primarily in non-erythroid cells (Charnay, et al., 1985; Cowie and Myers, 1988; Dierks, et al., 1983; Grosveld, et al., 1982; Myers, Tilly and Maniatis, 1986). To confirm the extent of the CACCC element and its nature as a cis sequence required for β-globin transcription in erythroid cells, point mutant β-globin promoters linked to the mouse metallothionein-1 gene were transiently co-transfected into MEL cells along with an internal reference gene (Figure 1A). A portion of the transfected cells was subsequently induced to differentiate with DMSO. The amount of β-globin-metallothionein mRNA produced by these transfected cells was measured with an S1 nuclease assay, quantitated and expressed as the level of transcription relative to wild type (RTL). The levels of transcripts measured in this assay are known to reflect accurately the transcriptional activity of the template that produced them (Chao, et al., 1983).

The results of these experiments, summarized in Figure 5, are in good agreement with previous studies in HeLa, MEL and 3T6 cells with the exception of the results from the -94T and -95A mutations

Figure 5. Quantitated results of transfection assays.

All data represent the average of a least 3 separate transfections except for the induced value of mutant -89T which is the average of 2 experiments. Relative transcription levels (RTLs) were calculated as the ratio of βMT signal to the H4MT signal for the test plasmid relative to the same ratio for the wild type βMT construct {(βMT/H4MT) $_{\text{test plasmid}}$/(βMT/H4MT) $_{\text{wild type βMT plasmid}}$}. RTLs of constructs in induced cells were calculated relative to the wild type βMT signal in induced cells and, in uninduced cells, the RTLs were calculated relative to the wild type βMT signal in uninduced cells. In the experiments reported here, DMSO induction resulted in an average increase in wild type β-globin-metallothionein fusion gene transcript levels of 6.5 fold. (A) The data are presented graphically with values from uninduced cells shown in the dark bars and from induced cells in the light bars. (B) The average RTL for each construct is given for both induced and uninduced cells. The standard deviation for each value is given in parenthesis.

A



B

MUTANT	RTL UNINDUCED	RTL INDUCED
WT	1.00 (0.11)	1.00 (0.02)
100A	1.07 (0.35)	1.12 (0.10)
95A	0.79 (0.17)	0.79 (0.21)
94T	0.73 (0.34)	0.85 (0.07)
93T	0.40 (0.13)	0.50 (0.10)
93A	0.49 (0.15)	0.64 (0.11)
92C/93T	0.96 (0.21)	0.86 (0.18)
92C	0.84 (0.29)	0.90 (0.21)
92T	0.44 (0.10)	0.48 (0.07)

MUTANT	RTL UNINDUCED	RTL INDUCED
91A	0.39 (0.27)	0.26 (0.06)
90C	0.50 (0.21)	0.54 (0.23)
90T	0.38 (0.11)	0.56 (0.38)
89T	0.31 (0.10)	0.52 (0.25)
88T	0.32 (0.14)	0.43 (0.20)
87A	0.43 (0.20)	0.27 (0.04)
87T	0.25	0.36
85A	1.07 (0.20)	0.92 (0.09)

(Cowie and Myers, 1988; Dierks, et al., 1983; Myers, Tilly and Maniatis, 1986). The -94T mutant gave an RTL of 0.25 when transfected into HeLa cells in one study and an RTL of 1.0 when transfected into 3T6 cells in another (Dierks, et al., 1983; Myers, Tilly and Maniatis, 1986). In the present study, the -94T mutant resulted in only a slight decrease in transcription. The -95A mutant previously gave RTLs of 0.17 in HeLa cells and 0.34 in uninduced and 0.64 in induced MEL cells (Cowie and Myers, 1988; Myers, Tilly and Maniatis, 1986). In this study, mutant -95A gave nearly wild type levels of transcription. It is possible that these differences can be explained in part by different CACCC binding activities in different cell types. Alternatively, the discrepancy between this and previous studies could be due to some inadvertent mutation occurring elsewhere in this construct. All point mutant promoters used in this study were resequenced and shown to have the expected sequence in the -106 to +1 region.

When comparing the transfection data to the point mutant binding data, it is clear that the binding behavior of CACCC species A, B and D on the point mutant promoters correlates well with the transcriptional activity of those promoters. CACCC species C DNA binding does not seem to correlate well with the transcriptional activity of the point mutant templates, indicating that this activity is less likely to activate the CACCC element in vivo.

CACCC Species B is Sp1

Other groups have suggested that the transcriptional activator SP1 can bind to the CACCC element (Giglioni, et al., 1989; Gumucio, et

al., 1991; Lewis, et al., 1988; Spanopoulou, Giguere and Grosveld, 1991; Xiao, et al., 1987; Yu, et al., 1991). If true, it is possible that SP1 is responsible for activating transcription through the β-globin CACCC element. The results of five experiments described below showed that SP1 can bind to the CACCC element.

First, we compared the ability of unlabeled oligonucleotides bearing either a high affinity SP1 binding site or a CACCC binding site to compete for binding to the CAC1 probe (Figure 6A). The CACCC oligonucleotide competed for species D binding (Figure 6A, lanes 2-6) at concentrations 4-5 fold below those required for equal levels of competition by the SP1 oligonucleotide (Figure 6A, lanes 7-11). Species A, B and C, on the other hand, were competed by concentrations of the SP1 oligonucleotide 4-5 fold below those required for equal levels of competition by the CACCC oligonucleotide. Thus, species D binds with higher affinity to the CACCC element than to a high affinity SP1 binding site, whereas species A, B and C bind with a higher affinity to an SP1 high affinity binding site than to a CACCC element.

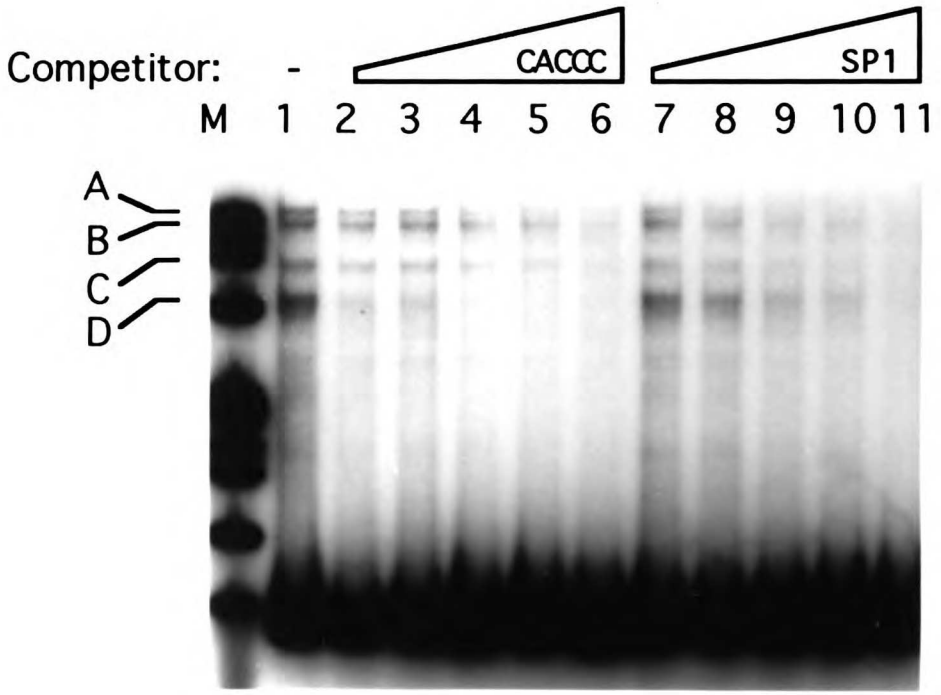
Second, we purified SP1 from MEL cells and directly examined its ability to bind the CACCC probe in the mobility shift assay. These experiments demonstrated that SP1 shifts the probe to a position identical to that of CACCC species B (Figure 6B, Compare lanes 7 and 8).

Third, when a polyclonal anti-SP1 antibody was included in a mobility shift binding reaction with pure SP1, the SP1 binding activity was diminished and a new, lower mobility binding species appeared (Figure 6C). Pre-immune serum did not affect SP1 binding

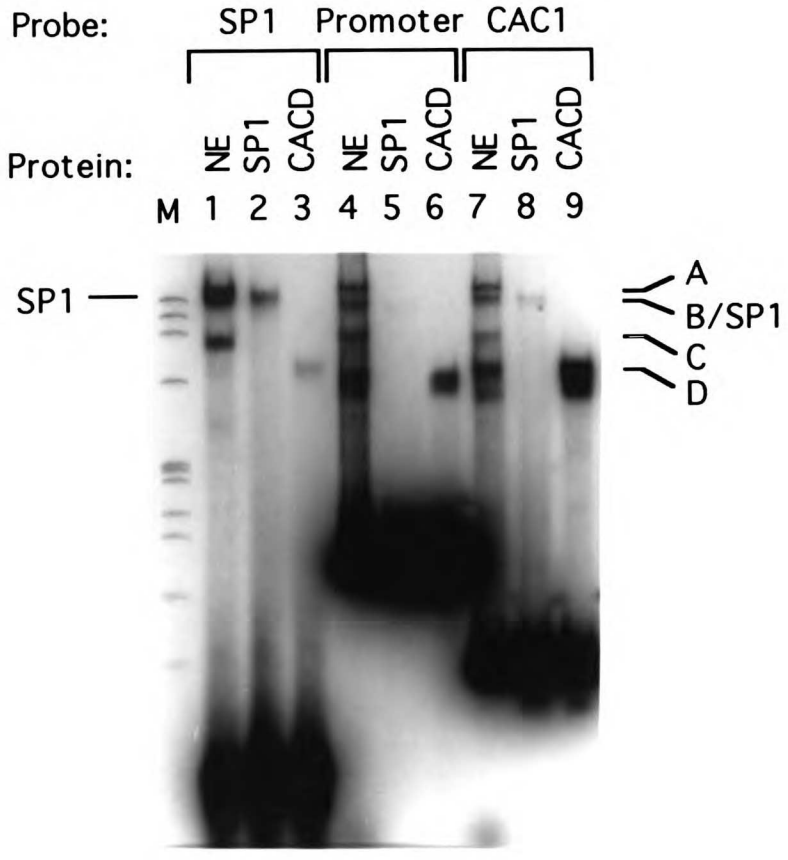
Figure 6. Comparison of CACD and SP1.

(A) Competition assays with CACCC and SP1 oligonucleotides. Mobility shift reactions with crude nuclear extracts, 12 fmole of the CAC1 probe and competitor DNA were assembled. These reactions included: no competitor (lane 1) or 4 (lanes 2 and 7), 8 (lanes 3 and 8), 20 (lanes 4 and 9), 40 (lanes 5 and 10) or 80 (lanes 6 and 11) fold molar excess of competitor oligonucleotide. The competitors were the CACCC oligonucleotide (lanes 2-6) or the SP1 oligonucleotide (lanes 7-11). (B) Mobility shift assays using the SP1 (lanes 1-3), β-globin promoter (lanes 4-6) and CAC1 (lanes 7-9) probes. Mobility shift reactions were assembled with each probe and either: crude nuclear extracts of MEL cells (lanes 1, 4 and 7), pure SP1 protein (lanes 2, 5 and 8) or partially pure CACD protein (lanes 3, 6 and 9). Each reaction contained 3 fmole of probe and the amount of each protein used was kept constant across the 3 probes. Note that in lane 5 a very faint shift comigrating with CACCC species B was seen on the original autoradiograph. (C) The indicated protein was mixed with the CAC1 probe in the presence of no (-), or an immune (I) or preimmune (PI) antiserum directed against SP1. These reactions were subjected to mobility shift analysis. (MEL NE= MEL cell nuclear extract). The arrow indicates the mobility of the new, low mobility complex observed when anti-SP1 antisera was included with either pure SP1 or crude MEL NE's.

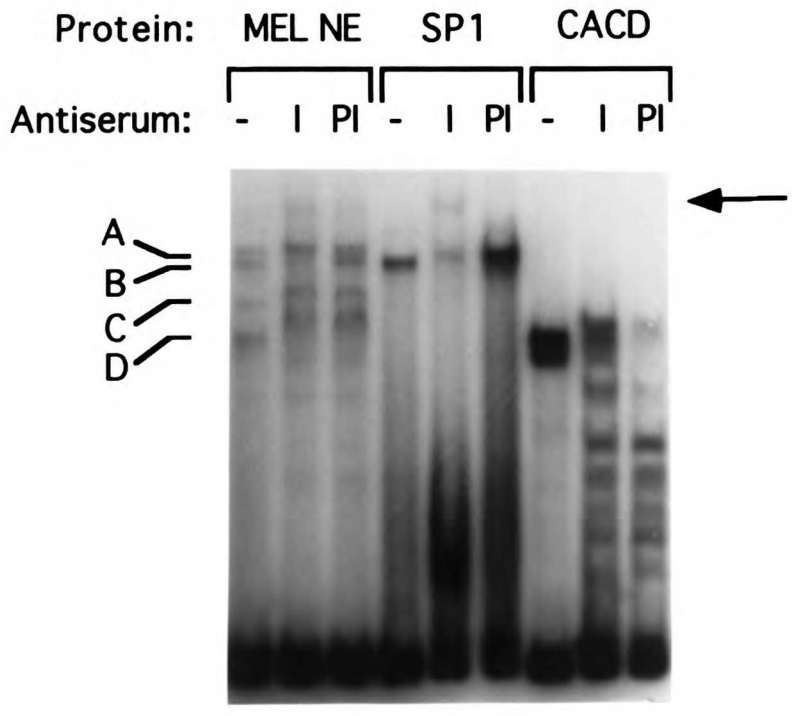
A



B


Probe:
SP1
Promoter
CAC1
Protein:
NE
SP1
CACD
NE
SP1
CACD
NE
SP1
CACD
M 1 2 3 4 5 6 7 8 9
SP1
A
B/SP1
C
D

C


Protein:
MEL NE
SP1
CACD
Antiserum:
- I PI - I PI - I PI
A
B
C
D

in the mobility shift assay. When the same experiment was performed with a crude MEL nuclear extract, species B was clearly diminished and a new lower mobility species appeared in the gel, presumably representing a supershifted DNA/antibody/species B complex (Figure 6C).

Fourth, when SP1 protein was used in the gel mobility shift assay with the point mutant probes, it gave results similar to those observed with CACCC species B in crude MEL nuclear extracts (Figure 4B). Finally, when purified SP1 was DNase1 footprinted on the CAC1 probe, it gave a footprint almost identical to that of the CACCC species A/B complex previously footprinted (Figure 3). In aggregate, the results of these experiments argue that SP1 in the crude MEL nuclear extract binds and retards the mobility shift probe to give rise to CACCC species B. While species A and C bind with higher affinity to an SP1 site than a CACCC site, they appear to be unrelated to SP1 by the criteria of the supershift experiment.

Characterization of Species D

Since species D is by far the most prominent activity observed in a mobility shift of the CAC1 probe with crude MEL nuclear extracts, we decided to characterize this activity in more detail. By a combination of conventional and site-specific DNA affinity chromatography, we purified CACCC binding species D approximately 3000 fold to ~1% purity (In preparation). We named this activity CACD for CACCC element binding species D.

We observed no other CACCC binding activities in this preparation of CACD. Non-specific competitor DNA was not required

in mobility shifts of partially pure CACD, presumably because the bulk of non-specific DNA-binding activities were removed during the purification procedure. The mobility of CACD in the gel shift assay, its relative binding affinity for the point mutant probes and its DNaseI footprint were the same as that observed for species D in the crude extract (See Figures 3, 4B and 6B). The partially pure CACD generated in this preparation was used for the remainder of this work.

Comparison of CACD to Species B/SP1

The similar behaviors of CACD and species B/SP1 in the footprinting and point mutant binding assays raised the possibility that CACD is related to SP1, perhaps as an alternately spliced or proteolytic product of SP1. Alternatively, CACD and SP1 may have distinct but related DNA binding domains. We compared CACD and SP1 in several different assays, and in each case, CACD appeared to differ from SP1.

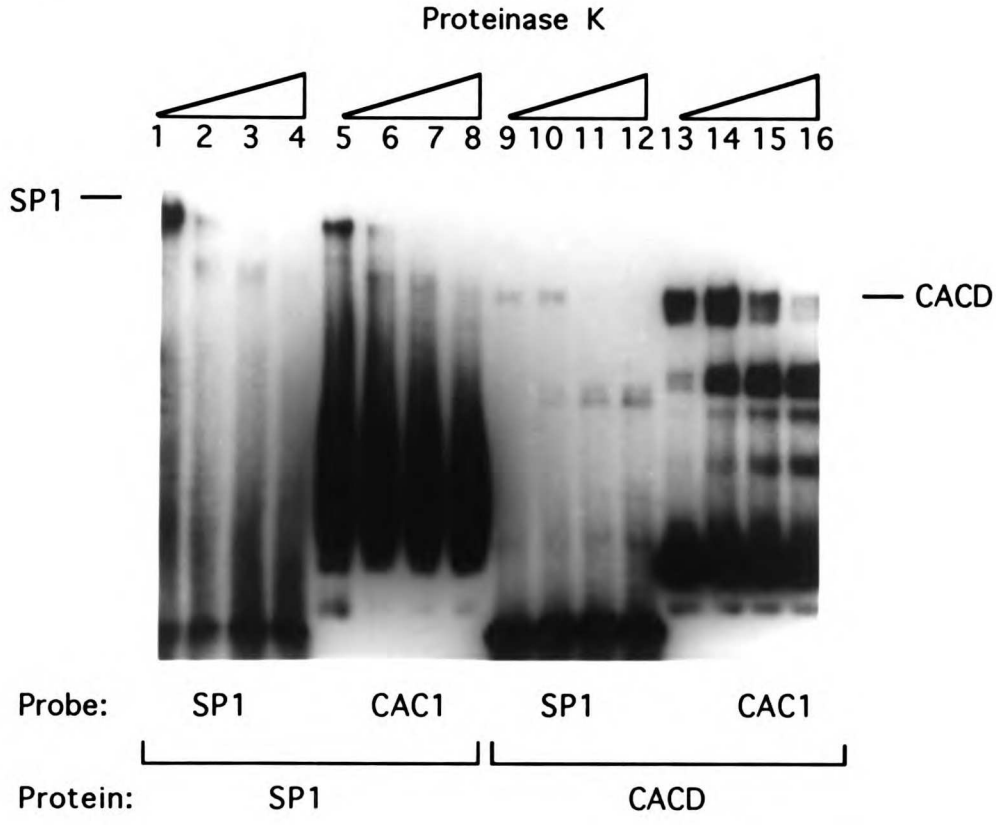
First, we examined the ability of CACD to bind to an SP1 probe. Consistent with the competition experiments of Figure 6A, we found that CACD could bind to an SP1 probe, although less efficiently than to the CAC1 probe (Figure 7A, lanes 9 and 13; also see Figure 6B lanes 3 and 9). Conversely, SP1 protein bound to the CAC1 probe, but less efficiently than it bound to the SP1 probe (Figure 6B lanes 2 and 8; Figure 7A, lanes 1 and 5).

Second, immunoblot analysis of the most purified fractions of CACD did not show cross-reactivity to either of two poly-clonal anti-SP1 antibodies (one of which is an anti-peptide antibody directed

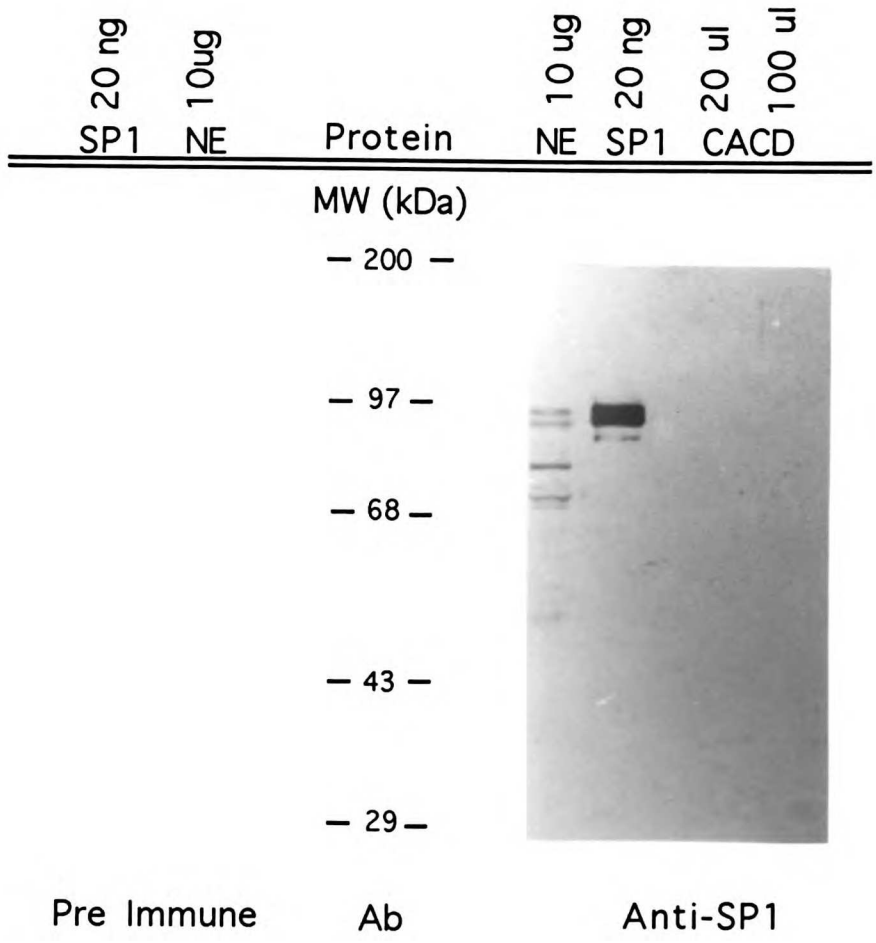
Figure 7. Comparison of CACD and Sp1

(A) Proteolytic clipping bandshift assay of SP1 and CACD protein. Mobility shift reactions using the SP1 (lanes 1-4 and 9-12) or CAC1 (lanes 5-8 and 13-16) probes were assembled with either pure SP1 (lanes 1-8) or partially pure CACD (lanes 9-16). The protein in these reactions had been treated with proteinase K for 0 (lanes 1, 5, 9 and 13), 2 (lanes 2, 6 ,10 and 14), 5 (lanes 3, 7 ,11 and 15) or 10 minutes (lanes 4, 8,12 and 16) prior to being assayed by mobility shift. Similar amounts of protein were assayed at each time point except for the no treatment samples which contained approximately 60% as much protein as the remainder of the reactions. (B) Immunoblot analysis of SP1 and CACD. The indicated amounts of MEL nuclear extracts, SP1 protein and CACD protein were separated on a 10% SDS gel, transferred to nitrocellulose and probed with an antiserum (#2892) directed against SP1 or a preimmune serum. Using the amount of CACD binding activity present in the fractions blotted and our estimate that CACD is a 50 kDa protein (In preparation), a minimum of 15 ng of CACD was used in this experiment. Immunoblots using antiserum # 2990 gave similar results (data not shown).

A


Proteinase K
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
SP1
CACD
Probe: SP1 CAC1 SP1 CAC1
Protein: SP1 CACD

B


20 ng SP1
10ug NE
Protein
10 ug NE
20 ng SP1
20 ul CACD
100 ul
MW (kDa)
200
97
68
43
29
Pre Immune
Ab
Anti-SP1

against finger 2 of the SP1 DNA binding domain; Figure 7B). Third, these anti-SP1 antisera did not supershift CACD when included in the mobility shift assay (Figure 6C).

The fourth approach for looking for relatedness between CACD and SP1 was to use the proteolytic clipping band shift assay (Schreiber, et al., 1988). In this assay, proteins are subjected to gentle protease treatment before being used in the mobility shift assay. If portions of the protein not essential for DNA binding are removed, then the mobility of the resulting DNA-protein complex will be increased compared to that of the intact protein. By subjecting both SP1 and CACD to proteinase K for varying times and then observing their mobility shifts, we asked if these proteins showed similar sensitivities to protease and if the pattern of proteolytic products still active for DNA binding was similar between the two proteins. SP1 and CACD differed substantially in this assay (Figure 7A). SP1 was relatively sensitive to proteinase K treatment and only produced one fragment of higher mobility. CACD gave rise to several fragments of higher mobility and the relative abundance of these fragments suggests that they were produced stoichiometrically from the intact protein.

Selection of Binding Sites for CACD

Because we had tested CACD binding only on the SP1 and CACCC probes, and did not know if sites with a higher affinity for the CACD protein existed, CACD was subjected to a site selection procedure (Blackwell and Weintraub, 1990). An oligonucleotide containing 21 random internal nucleotides and defined flanking sequences that

serve as primer binding sites was used to make a pool of random, double-stranded ^{32}P-labeled probe molecules. This random probe pool was used in mobility shift assays with the most purified fractions of CACD. DNA-protein complexes migrating with the same mobility as the CACD-CAC1 probe complex were eluted from the gel, PCR amplified, ^{32}P labeled and subjected to subsequent rounds of selection. By using low concentrations of CACD and by decreasing the amount of PCR amplified DNA probe used in later rounds of selection, we hoped to select for high affinity CACD binding sites. After 5 rounds of selection, the PCR products were cloned and the sequences of 22 individual clones were determined. Of the 22 clones examined, 16 contained sequences that matched the sequence CCACACCC at 6 or more of these 8 nucleotides (Table 2). These sequences were aligned by their CACCC element homology and the frequency of individual nucleotides occurring at each position of this alignment was determined. The nucleotides occurring most frequently in the selected CACD binding sites are the cytosines within the CACCC element. These nucleotides were shown by our mutational analysis to be most important for CACCC species D binding activity and for transcription of the β-globin gene. The adenines, which are not as important for either DNA binding of CACD or transcription of the β-globin gene, occur less frequently in the selected sequences. Overall, a consensus sequence, ANCCACACCCAT (N=any nucleotide), emerges from these data.

The results of the selection experiment suggest that the primary determinant of CACD binding affinity is the conserved sequence CCACACCC. In addition, an AT dinucleotide was commonly

TABLE 2

SITE SELECTION DATA

OLIGO	GAATTCCTACAG NNNNNNNNNNNNNNNNNNNN TGCAATGGATCCGTCT		
	AGAG	CCACACCC	TGG = WT BETA GLOBIN SEQUENCE
CLONE[a]			
2	GAATTCCTACAGCGAA	CCACCGCC	ATTGCAATGTGCAATGGATCCGTCT
3	AGACGGATCCATTGCATCGGGT	CCGCGCCC	ATCGTCCCTGTAGGAATTC
5	GAATTCCTACAGG	CCACACCC	ATTTAAAGGTAGTGCAATGGATCCGTCT
6	GAATTCCTACAGGC	CAACACCC	ATGATCCATCATGCAATGGATCCGTCT
7	GAATTCCTACAGC	CCACACCT	ATTTTAGGGAGGTGCAATGGATCCGTCT
9	AGACGGATCCATTGCACAGAGACCCG	CCACACCC	GTGCTGTAGGAATTC
11	AGACGGATCCATTGCATGGGAGA	ACACACCC	ATATGGCTGTAGGAATTC
13	AGACGGATCCATTGCACAACGG	CCACGCCC	TTCATCGCTGTAGGAATTC
16	AGACGGATCCATTGCAGATTGTGT	ACCACGCC	TTATGCTGTAGGAATTC
20	GAATTCCTACAGGC	CAACACCC	ATGATCCATCATGCAATGGATCCGTCT
22	GAATTCCTACAGC	CAACGCCC	TGCGCTACATGTGCAATGGATCCGTCT
23	AGACGGATCCATTGCA	CCACGCCC	CAACCTTATTGGCCTGTAGGAATTC
25	GAATTCCTACAGCGAA	CCACGCCC	ATTGCAATGTGCAATGGATCCG
27	AGACGGATCCATTGCATAACCATGATATGT	CCACGCCC	TGTAGGAATTC
28	GAATTCCTACAG	CCCCACCC	AGGAGACGCGCTATGCAATGGATCCGTCT
29	GAATTCCTACAGTCCGGGTGGGG	CCGCATCC	AGTGCAATGGATCCGTCT
29(rev.)	AGACGGATCCATTGCACTGGATGCGG	CCCCACCC	GGACTGTAGGAATTC

FREQUENCY OF INDIVIDUAL NUCLEOTIDES OCCURRING IN ALIGNED SEQUENCES

	A	G	A	G	C	C	A	C	A	C	C	C	T	G	G	T	A	A
G	5	7	12	6			2		6	2			2	5	4	5	4	4
A	5	4	3	4	2	3	12	1	9				10	1	4	5	1	6
T	1	2		3						1		1	4	11	6	4	7	2
C	6	4	2	4	15	14	3	16	2	14	17	16	1		3	3	5	5
CONSENSUS	N	N	a	N	C	C	a	C	a	C	C	C	a	t	N	N	t	N

a. The sequences shown here are aligned to their homology with the β-globin CACCC sequence. Six sequences that had a less than a 6 of 8 match to the β-globin CACCC sequence are not shown. Also, note that some of the sequences vary in length from the starting oligonucleotide. Presumably, this is due to misincorporation of nucleotides by Taq DNA polymerase.

recovered at the 3' end of the consensus sequence (corresponding to positions -86 and -85) and might also contribute to binding affinity. The -85A point mutant probe was bound by CACCC species A-D as efficiently as the wild type promoter (Figure 4), but mutations -85T and -86A, which change the wild type promoter to the consensus sequence, had not been tested. We therefore made a -86A/-85T double mutant oligonucleotide that was otherwise identical to the CACCC oligonucleotide used in the competition assays. We found that it competed for CACD binding at half the concentration required by the wild type oligonucleotide, indicating that the -86A/-85T mutation results in higher affinity CACD DNA-binding (data not shown). Finally, site selection experiments, similar to those above, in which the starting oligonucleotide was fixed as NNNNNCNNCACCCNNN, returned the sequence cNNNaCCACACCCa/ttt (lower case letters indicate a lower but still non-random frequency of occurrence, N indicates a random nucleotide; data not shown), again consistent with the idea that the primary determinant of CACD binding affinity is the sequence of the core conserved sequence found in the β-globin promoters.

DISCUSSION

Several common activities can bind the CACCC element.

To identify potential transcription factors acting through the CACCC element, we employed a gel shift assay using MEL nuclear extracts and a short probe containing the conserved CACCC sequence. This assay identified at least three activities that bind the CACCC element in a specific manner, species A, B and D, with species D

binding most tightly. A fourth activity, CACCC species C, bound less specifically than these other species. Each activity is present in several different cell types and each increases slightly in abundance during MEL cell differentiation. Although this last observation could indicate coordinate regulation of these binding activities in MEL cells, we feel it is more likely due to a change in overall nuclear structure or protein content during terminal differentiation, which leads to a slight increase in these factors on a per protein basis.

CACCC Species B is SP1

Using purified SP1 protein in gel shifts with the CAC1 probe, the SP1-CAC1 complex migrated with the same mobility as CACCC species B in native gels. Species B and SP1 also behaved similarly in binding competition experiments. When either of two anti-SP1 antibodies was included in mobility shift binding reactions with crude MEL extracts, species B was specifically abolished and a new, lower mobility, band appeared. Therefore, we conclude that CACCC species B contains SP1.

Although species A and C behaved similarly to species B in competition assays, and exhibited DNAse 1 footprints nearly identical to SP1, they were not reproducibly perturbed by anti-SP1 antisera in the super shift assay. Occasionally, CACCC binding species A, C and D were perturbed when SP1 antisera were included in mobility shift reactions. However, these other interactions also occurred with preimmune serum, suggesting that they were not the result of specific antigen-antibody interactions.

What activity is responsible for transcriptional activation of β-globin through the CACCC element?

One phenotype expected of mutations in the binding site of a transcriptional activator is that those mutations that prevent the activator from binding DNA should decrease the activity of the linked promoter. Although it is possible that mutations in the CACCC element can decrease β-globin transcription by some other mechanism, we felt that at least some mutations were likely to exert their effect in this manner. Starting with the assumption that the mobility shift assay, as performed here, allows examination of any CACCC binding activity relevant to the activity of the β-globin promoter, we develop the following model of CACCC element function.

Because none of the point mutations in the CACCC element that decreased transcription also decreased binding of CACCC species C, this activity is unlikely to be responsible for activation of β-globin transcription. SP1, CACCC species A and D are all candidate activators since their relative binding affinities for the point mutant promoters correlate well with the transcriptional activity of those promoters in the transfection assay. However, the few exceptions to this last statement are potentially instructive.

Mutants -94T and -95A, which bound wild type levels of species D and low levels of SP1 and species A, were transcribed at near wild type levels. One model that would account for this observation is that species D normally binds the CACCC element in vivo and activates transcription of the β-globin gene. Therefore, CACCC element mutations that do not affect species D binding, such as -94T and -95A, would not affect β-globin transcription. Mutant

-92C/-93T, which was transcribed at near wild type levels, bound low levels of species D, wild type levels of SP1 and increased amounts of species A and C. Thus, this CACCC element would be less likely to use CACD as a transactivator in vivo. These data can be incorporated into our model by assuming that some other factor, perhaps SP1 or species A, compensates for the loss of affinity for species D by binding the CACCC element and activating transcription. Indeed, this type of CACCC element has been shown to confer SP1 responsiveness on a linked promoter in vivo (Gumucio, et al., 1991). Since the -92C/-93T mutation creates a CACCC element of the type found in fetal and embryonic globin genes as well as the α- and β-globin LCRs, this result suggests that more than one CACCC element binding protein may regulate globin gene transcription in vivo (Philipsen, et al., 1990; Strauss, et al., 1992).

Because CACCC species A appears only when both species D and SP1 can also bind, we cannot say that CACCC species A is sufficient for transcription. However, because several mutations appear to abolish CACCC species A binding without affecting transcription, species A must not be necessary for transcription. Because it is not abundant compared to species D and because we have not characterized it extensively, we shall not consider it further in this discussion. One mutation, -90T, results in decreased transcription, yet it binds significant amounts of SP1, CACD and species A. It is possible that this mutation affects transcription by altering interactions of an activator with the rest of the transcriptional machinery rather than by preventing the activator from binding. Finally, if our initial assumption is incorrect - if other activities bind

the CACCC element in vivo that are not observed in our in vitro assay - then the above model may be incorrect. To test this possibility, any candidate activator of the CACCC element must be specifically depleted from a cell or cell extract and be shown to be required for β-globin transcription.

The CACCC Element is a High Affinity Target of CACD

From the above analysis we conclude that SP1, a known activator of transcription, can probably stimulate transcription from the β-globin promoter but that it is not the only activity that can provide this function. To further assess the role of CACCC species D in β-globin transcription, a site selection experiment to find high affinity CACD binding sites was performed. These experiments showed that the high affinity site for CACD is the site that naturally occurs in the mouse β-globin promoter. The site selection experiment also showed that the -86A/85T mutation slightly increases affinity for CACD. Whether or not this mutation has any functional consequence remains to be seen.

The CACCC elements of the adult β-globin like genes are most likely conserved across species as a result of an evolutionary selection for a high affinity binding site for a protein that prefers to bind to the sequence CCACACCC. This argument predicts that positions that are less well conserved among CACCC elements should be less important both for binding of the activator protein and for transcription of the β globin promoter; this is what we observed for binding of CACD to promoters with the -92C mutation. Although the -92T mutation reduces DNA-binding and transcription, this is not

inconsistent with our argument since CACCC elements with a thymidine at this position make only a modest contribution to β-globin transcription (Dierks, et al., 1983; Gonzalez-Redondo, et al., 1989; Ristaldi, et al., 1990). Furthermore, β-globin promoters with this type of CACCC element often carry 2 CACCC elements, one with a -92T type sequence and one with a -92A type sequence (Table 1).

It is formally possible that SP1 provides all transactivation functions through the CACCC element. However, although SP1 appears to be able to function through the CACCC element, its binding to the mouse β-major CACCC element is weak compared to its affinity for a consensus SP1 site or compared to the affinity of CACD for the β-globin CACCC site (compare the binding of these proteins to their cognate and heterologous sites in Figures 6B and 7A). Site selection experiments similar to the ones described here have been performed with SP1 and defined a consensus high affinity binding site, GGGGCGGGGT, that is different from the CACCC element (Thiesen and Bach, 1990). If SP1 normally functions through the CACCC element, it is hard to understand why higher affinity consensus SP1 sites are not found in the β-like-globin promoters. If low affinity SP1 binding at this position were important, then, by the arguments given above, lower levels of sequence conservation in the CACCC elements of the β-like globin promoters would be expected. Another piece of evidence that argues that CACCC and SP1 binding sites are functionally distinct, and therefore likely to be the targets of different proteins, comes from the work of Wang and Gralla (Wang and Gralla, 1991). When testing the ability of proximal promoter elements to interact with remote enhancer type elements, they found

that CACCC and SP1 sites behaved quite differently even when compared in nearly identical contexts. This suggests that CACCC and SP1 sites are the targets of different transcriptional activators with distinct functions. Based on the above arguments, we believe that the CACCC element represents a high affinity protein binding site and, on the strength of the site selection result and the correlation of CACD binding to and the transcriptional activity of the point mutant promoters, that this element is a target for the CACD protein.

The problem of multiple proteins binding to a single DNA sequence element is one of general significance (Johnson and McKnight, 1989). Many DNA binding proteins are members of families that share similar DNA binding domains and sequence specificities. For example, different homeodomain proteins of *Drosophila* can bind to very similar sites in vitro yet they act at different sites in vivo (Hayashi and Scott, 1990). Furthermore, the relative affinity differences between homeodomain proteins for a particular binding site can be a good predictor of which protein(s) will act on the binding site in vivo (Dessain, et al., 1992). In some cases, the difference between the site preferences of these related proteins has been shown to be due to small differences in structure between the proteins (Hayashi and Scott, 1990). The site selection procedure as used here provides an avenue of inferring what protein or proteins are most likely to operate through a particular site in vivo.

An additional level of target specificity is often provided to a DNA-binding protein by interactions with another site specific DNA-binding protein bound nearby. This is true for the yeast

homeodomain protein α2 (Goutte and Johnson, 1988; Keleher, et al., 1988). The α2 protein's DNA-binding specificity is determined in part by its ability to interact cooperatively with either of 2 proteins, MCM1 or a1. In this situation, the site selection procedure would be of use only if the protein-protein interaction that occurred in vivo was recognized and reproduced in vitro. Mutations near a protein's DNA-binding site that decrease transcription but do not affect its binding in vitro may be indicative of an in vivo protein-protein interaction that is directed by the correct apposition of two proteins on their DNA-binding sites. Thus, a prerequisite for using the site selection assay to discriminate between potential activators is a thorough understanding of the binding site of interest and the functional consequences of mutations in or around the site. In the case of the β-globin CACCC element, mutations at positions flanking the CACCC element have no effect on transcription (Cowie and Myers, 1988; Myers, Tilly and Maniatis, 1986). Also, combinations of mutations in the CACCC element and other cis sequences in the β-globin promoter result in lower transcription levels than those for either single mutation alone. Finally, the affinity of CACD is similar for both the CAC1 and β-globin promoter probes (Figure 6B, compare lanes 4 and 7). All of these results are consistent with the model that the protein that binds the CACCC element does so independently of proteins binding other cis sequences in the β-globin promoter.

On the Role of the CACCC Element in Transcription

In contrast to a previous study, we found no evidence for an inducible CACCC element binding activity in extracts of induced MEL

cells (Mantovani, et al., 1988). This finding is consistent with the similar effects of CACCC element mutations on β-globin transcription in uninduced and induced cells, arguing that the CACCC element mediates the effects of a factor that provides activation functions not specific to terminally differentiating MEL cells. Likewise, the presence of CACCC binding factors and the activity of CACCC elements in non-erythroid cells suggests that CACCC binding transcription factors are widely distributed across tissues.

CACD

Several groups have described a DNA-binding activity that has a sequence specificity and mobility in the mobility shift assay that is similar to that of CACD. These include TEF-2 and AP3 (Davidson, et al., 1988; Mitchell, Wang and Tjian, 1987; Xiao, et al., 1987). However, these activities have not been described in detail. The proteolytic clipping bandshift assay (Figure 7A) shows that CACD has a protein component that can be partially degraded and still retain its DNA binding activity. We have found that the DNA-binding activity of CACD is inhibited by chelators of divalent cations and that low concentrations of zinc can restore binding activity (GAH and RMM, in preparation). The zinc dependent DNA-binding activity of CACD is reminiscent of that demonstrated for several zinc finger proteins, including SP1 and an activity, likely to be SP1, that binds the chicken CACCC element (Kadonaga, et al., 1987; Lewis, et al., 1988).

Given the similarity of SP1 and CACD biochemical and DNA-binding properties, it seemed possible that CACD is related to SP1.

This might occur via an alternate splice of the SP1 mRNA or by proteolysis of the 92.5 kDa SP1 protein to produce the smaller CACD protein. However, alternate forms of the SP1 cDNA have not been reported, its mRNA appears to be a single species on Northern blots, and low stringency Southern blots fail to reveal SP1 homologues (Kadonaga, et al., 1987; Saffer, et al., 1990). Furthermore, by four experimental criteria presented here, relative affinities for their DNA binding sites, immunoblot analysis, supershift analysis and protease clipping, SP1 and CACD appear to be unrelated. Further evidence suggesting that SP1 and CACD are different entities includes their divergent chromatographic behavior. On both wheat germ agglutinin and site specific DNA affinity columns, these activities behave quite differently (Our unpublished data). While the above experiments do not rule out that CACD is related to SP1 by an alternate splice or proteolytic event, they make it unlikely. We believe that CACD, like SP1, is likely to be a member of the Cys_2-His_2 zinc finger protein family but only distantly related to SP1.

In conclusion, we characterized 4 CACCC element DNA binding activities that occur in MEL cell nuclear extracts. By correlating the ability of these activities to bind to mutant promoters with the ability of the mutant promoters to activate transcription, we found that two activities, CACD and SP1, are likely activators of the β-globin CACCC element. We subsequently used a site selection experiment to determine the high affinity site for CACD. This experiment gave the surprising result the CACD's high affinity site is nearly identical to that of the β-globin CACCC element.

We suggest that these experiments illustrate a general methodology for discriminating between potential activators of a single cis element. Observation of the effects of mutations in and around a cis element allows determination of its extent in a functional sense. Binding studies using the same mutant sites allows one to search for proteins with binding properties consistent with those suggested by the functional studies. If more than one candidate activator remains at this point, the site selection assay can be used to determine the preferred high affinity sites of the candidate activators. Additionally, this assay will reveal the allowed variations in the high affinity site that still allow DNA binding. These data can be used in conjunction with the functional studies to ascertain what activity(s) has the properties expected of the in vivo activator.

REFERENCES

Antoniou, M., and Grosveld, F. (1990). β-globin dominant control region interacts differently with distal and proximal promoter elements. Genes and Dev. *4*, 1007-1013.

Ausubel, F.K., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1991). Current Protocols in Molecular Biology. Greene Publishing Associates/Wiley-Interscience.

Blackwell, T.K., and Weintraub, H. (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science *250*, 1104-1110.

Catala, F., de Boer, E., Habets, G., and Grosveld, F. (1989). Nuclear protein factors and erythroid transcription of the human A gamma-globin gene. Nucleic Acids Res. *17*(10), 3811-38-27.

Chao, M.V., Mellon, P., Charnay, P., Maniatis, T., and Axel, R. (1983). The regulated expression of beta-globin genes introduced into mouse erythroleukemia cells. Cell *32*, 483-493.

Charnay, P., Mellon, P., and Maniatis, T. (1985). Linker scanning mutagenisis of the 5'-flanking region of the mouse β-major-globin gene: sequence requirements for transcription in erythroid and non-erythroid cells. Mol. Cell. Biol. *5*(6), 1498-1511.

Cowie, A., and Myers, R.M. (1988). DNA sequences involved in transcriptional regulation of the mouse β-globin promoter in murine erythroleukemia cells. Mol. Cell. Biol. *8*(8), 3122-3128.

Davidson, I., Xiao, J.H., Rosales, R., Staub, A., and P., C. (1988). The HeLa cell protein TEF-1 binds specifically and cooperatively to two SV40 enhancer motifs of unrelated sequence. Cell *54*, 931-942.

de Boer, E., Antoniou, M., Mignotte, V., Wall, L., and Grosveld, F. (1988). The human β-globin promoter: nuclear protein factors and erythroid specific induction of transcription. EMBO J. *7*(13), 4203-4212.

Deisseroth, A., and Hendrick, D. (1978). Human alpha-globin gene expression following chromosomal dependent gene transfer into mouse erythroleukemia cells. Cell *15*, 55-63.

Dessain, S., Gross, C.T., Kuziora, M.A., and McGinnis, W. (1992). Antp-type homeodomains have distinct DNA binding specificities that correlate with their different regulatory functions in embryos. EMBO J. *11*(3), 991-1002.

Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J., and Weissman, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell *32*, 695-706.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. *11*, 1475-1489.

Efstratiadis, A., et al. (1980). The Structure and evolution of the human β-globin gene family. Cell *21*, 653-668.

Favalaro, J., Treisman, R., and Kamen, R. (1980). Transcription maps of polyoma virus-specific RNA: analysis by two dimensional nuclease S1 gel mapping. Methods Enzymol. *65*, 718-749.

Frampton, J., Walker, M., Plumb, M., and Harrison, P.R. (1990). Synergy between the NF-E1 erythroid-specific transcription factor and the CACCC factor in the erythroid-specific promoter of the human porphobilinogen deaminase gene. Mol. Cell. Biol. *10*(7), 3838-3842.

Giglioni, B., Comi, P., Ronchi, A., Mantovani, R., and Ottolenghi, S. (1989). The same nuclear proteins bind the proximal CACCC box of the human beta-globin promoter and a similar sequence in the enhancer. *164*(1), 149-155.

Gonzalez-Redondo, J.M., et al. (1989). A C----T substitution at nt--101 in a conserved DNA sequence of the promotor region of the beta-globin gene is associated with "silent" beta-thalassemia. Blood *73*(6), 1705-1711.

Goutte, C., and Johnson, A.D. (1988). a1 protein alters the DNA binding specificity of alpha 2 repressor. Cell *52*, 875-882.

Grosveld, G.C., de Boer, E., Shewmaker, C.K., and Flavell, R.A. (1982). DNA sequences necessary for transcritpion of the rabbit β-globin gene *in vivo*. Nature *295*, 120-126.

Gumicio, D.L., Rood, K.L., Gray, T.A., Riordan, M.F., Sartor, C.I., and Collins, F.S. (1988). Nuclear proteins that bind the human γ-globin gene promoter: alteration in binding produced by point mutations associated with hereditary persistance of fetal hemoglobin. Mol. Cell. Biol. *8*(12), 5310-5322.

Gumucio, D.L., Rood, K.L., Blanchard-McQuate, K.L., Gray, T.A., Saulino, A., and Collins, F.S. (1991). Interaction of Sp1 with the human gamma globin promoter: binding and transactivation of normal and mutant promoters. Blood *78*(7), 1853-1863.

Harlow, E., and Lane, D. (1988). Antibodies: A Laboratory Manual. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory .

Hayashi, S., and Scott, M. (1990). What determines the specificity of action of Drosophila homeodomain proteins? Cell *63*, 883-894.

Jackson, S.P., and Tjian, R. (1989). Purification and analysis of RNA polymerase II transcription factors by using wheat germ agglutinin affinity chromatography. Proc. Natl. Acad. Sci. *86*(6), 1781-5.

Johnson, P.F., and McKnight, S.L. (1989). Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. *58*, 799-839.

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

Kadonaga, J.T., Carner, K.R., Masiarz, F.R., and Tjian, R. (1987). Isolation of cDNA encoding transcription factor SP1 and functional analysis of the DNA binding domain. Cell *51*, 1079-1090.

Kadonaga, J.T., Jones, K.A., and Tjian, R. (1986). Promoter specific activation of RNA polymerase II transcription by SP1. Trends Biochem. Sci. *11*, 20-23.

Keleher, K.A., Goutte, C., and Johnson, A.D. (1988). The yeast cell-type-specific repressor alpha 2 acts cooperatively with a non-cell-type-specific protein. Cell *53*, 927-936.

Krieg, P.A., and Melton, D.A. (1987). In Vitro RNA synthesis with SP6 RNA polymerase. *155*, 397.

Kulozik, A.E., Bellan-Koch, A., Bail, S., Kohne, E., and Kleihauer, E. (1991). Thalassemia intermedia: moderate reduction of beta globin gene transcriptional activity by a novel mutation of the proximal CACCC promoter element. Blood *77*(9), 2054-2058.

Kunkel, T.A., Roberts, J.D., and Zakour, R.A. (1987). Rapid and efficient site-specific mutagenisis without phenotypic selection. Methods Enzymol. *154*, 367-382.

Lewis, C.D., Clark, S., Felsenfeld, G., and Gould, H. (1988). An erythrocyte-specific protein that binds to the poly(dG) region of the chicken β-globin promoter. Genes and Dev. *2*, 863-873.

Mantovani, R., Malgaretti, N., Nicolis, S., Giglioni, B., Comi, P., Cappellini, N., Bertero, M.T., Caligaris-Cappio, F., and Ottolenghi, S. (1988). An erythroid specific nuclear factor binding to the proximal CACCC box of the beta-globin gene promoter. Nucleic Acids Res. *16*(10), 4299-4313.

Marks, P.A., and Rifkind, R.A. (1978). Erythroleukemic differentiation. Annu. Rev .Biochem. *47*, 419-448.

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

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

Myers, R.M., Tilly, K., and Maniatis, T. (1986). Fine structure genetic analysis of a β-globin promoter. Science *232*, 613-618.

Orkin, S.H., Antonarakis, S.E., and Kazazian, H.H.J. (1984). Base substitution at position -88 in a β-thalassemic globin gene. J. Biol. Chem. *259*, 8679-8681.

Philipsen, S., Talbot, D., Fraser, P., and Grosveld, F. (1990). The b-globin dominant control region: hypersensitive site 2. EMBO J. *9*(7), 2159-2167.

Ristaldi, M.S., et al. (1990). The C-T substitution in the distal CACCC box of the beta-globin gene promoter is a common cause of silent beta thalassaemia in the Italian population. British Journal of Haematology *74*(4), 480-486.

Saffer, J.D., Thurston, S.J., Annarella, M.B., and Compton, J.G. (1990). Localization of the gene for the trans-acting transcription factor Sp1 to the distal end of mouse chromosome 15. Genomics *8*, 571-574.

Schreiber, E., matthias, P., Muller, M.M., and Schaffner, W. (1988). Identification of a novel lymphoid specific octamer binding protein (OTF-2B) by proteolytic clipping bandshift assay (PCBA). EMBO J. *7*(13), 4221-4229.

Schule, R., Muller, M., Otsuka-Murakami, H., and Renkawitz, R. (1988). Cooperativity of the glucocorticoid receptor and the CACCC-box binding factor. Nature *332*, 87-90.

Spanopoulou, E., Giguere, V., and Grosveld, F. (1991). The functional domains of the murine *Thy-1* gene promoter. Mol. Cell. Biol. *11*(4), 2216-2228.

Strauss, E.C., Andrews, N.C., Higgs, D.R., and Orkin, S.H. (1992). In vivo footprinting of the human a-globin locus upstream regulatory element by guanine and adenine ligation-mediated polymerase chain reaction. Mol. and Cell. Biol. *12*(5), 2135-2142.

Stuve, L.L., and Myers, R.M. (1990). A directly repeated sequence in the β-globin promoter regulates transcription in murine erythroleukemia cells. Mol. Cell. Biol. *10*(3), 972-981.

Thiesen, H., and Bach, C. (1990). Target detection assay (TDA): aversatile procedure to determine DNA binding sites as demonstrated on SP1 protein. Nucleic Acids Res. *18*(11), 3203-3209.

Tsai, S.F., Strauss, E., and Orkin, S.H. (1991). Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter. Genes and Dev. *5*(6), 919-931.

Wang, W.D., and Gralla, J.D. (1991). Differential ability of proximal and remote element pairs to cooperate in activating RNA polymerase II transcription. Mol. Cell. Biol. *11*(9), 4561-4571.

Wright, S., Rosenthal, A., Flavell, R., and Grosveld, F. (1984). DNA sequences required for regulated expression of β-globin genes in murine erythroleukemia cells. Cell *38*, 265-273.

Xiao, J.H., Davidson, I., Macchi, M., Rosales, R., Vigneron, M., Staub, A., and Chambon, P. (1987). In vitro binding of several cell-specific and ubiquitous nuclear proteins to the GT-I motif of the SV40 enhancer. Genes and Dev. *1*(8), 794-807.

Yu, C.Y., Motamed, K., Chen, J., Bailey, A.D., and Shen, C.K. (1991). The CACC box upstream of human embryonic epsilon globin gene binds Sp1 and is a functional promoter element *in vitro* and *in vivo*. J. Biol. Chem. *266*(14), 8907-8915.

CHAPTER 3

CHARACTERIZATION AND PURIFICATION OF THE CACCC ELEMENT BINDING PROTEIN CACD

INTRODUCTION

The CACCC element of the adult β-globin promoters is a sequence, CCACACCC, that occurs at position -90 with respect to the transcription start site, is conserved across species and required for wild type promoter activity (Charnay, et al., 1985; Cowie and Myers, 1988; Dierks, et al., 1983; Efstratiadis, et al., 1980; Myers, et al., 1986). In MEL cells, a pre-erythroblast like cell line that can be induced to terminally differentiate, the CACCC element appears to be equally active before and after induction (Cowie and Myers, 1988; Marks and Rifkind, 1978). CACCC elements are also found in other genes, many of which are active in non-erythroid cell types (See chapter 2). These data suggest that the CACCC element is the target of a transcriptional activator protein(s) that provides a general, non-cell-type specific transcriptional activation function.

In order to understand β-globin transcriptional regulation, we are studying the CACCC element and the factors that bind it. Previously, we showed that MEL cell nuclear extracts contain 4 activities that can bind the CACCC element in a gel mobility shift assay (See chapter 2). No change in the pattern of these activities or induction of a new CACCC binding activity occurred during MEL cell differentiation. Also, activities with the same mobiltiy in the gel shift assay and binding pattern in DNAse1 footprinting assays were observed in non-erythroid cell types. We and other labs have demonstrated that one of the CACCC element binding activities is the transcriptional activator protein Sp1. Any of these four activities might represent a potential transcriptional activator(s) of β-globin transcription.

To diffferentiate between these activities and their potential to activate β-globin transcription, we looked for a correlation between their ability to bind to β-globin promoters with mutant CACCC elements and the transcriptional activity of those promoters. This strategy led to the hypothesis that at least two different activities can activate transcription from the β-globin promoter CACCC element: SP1 and an activity that we named CACD. When CACD was partially purified and subjected to a site selection assay, the consensus high affinity binding site of this protein was discovered to be the same as the conserved CACCC sequence that occurs in the adult β-globin promoters. On the strength of these results and the abundance of CACD activity in MEL nuclear extracts, we suggested that CACD represents the activity that binds the β-globin CACCC element in vivo and stimulates transcription.

Testing this hypothesis requires demonstrating that CACD has the ability to activate transcription when bound to the β-globin CACCC element and then showing that CACD is the protein that carries out this function in vivo. As a step towards these goals, we describe here the biochemical characterization and partial purification of the CACD activity. We show that CACD contains a 44 kD protein that specifically binds the β-globin CACCC element in a zinc dependent manner. Also, the CACD DNA binding activity is inhibited by phosphatase treatment, suggesting that it is regulated by phosporylation. We further characterized CACD by DNAse1 footprinting and DMS interference assays. Finally, we have purified CACD several hundred fold from nuclear extracts of MEL cells.

Materials and Methods

Enzymes

Except where noted, all enzymes were purchased from New England Biolabs.

Probes

The probe CAC1, described in chapter 2, was used in all mobility shift assays. For UV crosslinking, in a volume of 10 µl, 5 µg of the plasmid pSPCAC1, from which the CAC1 probe is derived, was treated with 1.1 µl 1N NaOH for 3 minutes. 1.1 µl 3M NaOAc pH4.5 was then added along with 100 µl EtOH and the sample was chilled and precipitated. The pellet was washed with EtOH and then allowed to anneal to 5 pmol of a primer (5'-AATACGACTCACTATAG-3') homologous to the T7 phage promoter that flanks the polylinker sequences of pSPCAC1. This primer is oriented so that the antisense strand of the CACCC element is 3' to it; the resulting probe, therefore, is labeled with ^{32}P and 5-bromo-deoxy-uridine (BrdU) on this strand. The annealed template was primer extended in a 20 µl reaction containing1mM dCTP, 1 mM dATP, 60 µCi α-^{32}P-dGTP (3000 Ci/mM, Amersham), 0.2 mM 5-bromo-dUTP (Sigma), 2 mM DTT, 5U Klenow enzyme in a buffer containing 10 mM Tris-HCl pH 7.5 and 10 mM $MgCl_2$. The reaction was incubated for 45 minutes at room temperature. 20U Eco R1 and 20U Bam H1 were the added to release the double stranded probe from the plasmid and the reaction was incubated an additional 2 hours at 37°C. The probe was purified in a native 6% polyacrylamide gel and eluted by the crush soak method

(Maxam and Gilbert, 1980). For DNase I footprinting, CAC1 probes labeled at a specific end were generated by cutting with either Eco RI or Xba I alone, treating with calf intestinal alkaline phosphatase (Boehringer), and end-labeling with γ-^{32}P-ATP (7000 Ci/mmol, ICN) and polynucleotide kinase. The DNA was subsequently digested with the other restriction enzyme to release the end labeled probe and then gel purified as above.

Competitor DNAs

The sequences of the competitor oligonucleotides used in this work are: CCAAT: 5'-GATCCTTTTGGATTGAAGCCAATATGAT-3' and 5'-CTAGATCATATTGGCTTCAATCCAAAAG-3' ,
CACCC: 5'-TCGACCCTTACCAGGGTGTGGCTCTACG-3' and 5'-TCGACGTAGAGCCACACCCTGGTAAGGG-3'. The oligonucleotides were annealed in Tris-HCl pH8.0, 1mM EDTA, 0.1M NaCl by heating to 95°C and allowing them to cool slowly to room temerature.

Gel Mobility Shift

The gel mobility shift assay was carried out exactly as described in chapter two.

UV Crosslinking Assay

For UV crosslinking, normal mobility shift reactions were set up with 4 fmol of the internally labeled, BrdU substituted probe and allowed to incubate for 10 minutes at room temperature. The samples were spotted onto Saran Wrap on the surface of a Fotodyne UV 300 light box (λmax=300 nanometer) and irradiated for 4

minutes. These reactions were then transferred to an eppendorf tube, brought to 6 mM Mg^{2+}, 10 mM Ca^{2+}, and treated with approximately 50 U Micrococcal nuclease (Boehringer) and 1 μg DNaseI (Worthington) for 10 minutes at room temperature. The nuclease treated samples were brought to 20 mM EDTA and precipitated with 5 volumes of cold acetone. The pellets were resuspended in loading buffer and separated on a 10% polyacrylamide-SDS gel. Marker proteins (BioRad, low molecular weight) were visualized by staining with coomassie brilliant blue and the crosslinked protein-DNA complexes were visualized by autoradiography.

Gel Elution, Denaturation-Renaturation Assay

SDS gel elution followed by denaturation and renaturation were carried out according to the method of Hager and Burgess (1980) (Hager and Burgess, 1980). Because we were concerned about oxidation of the protein, the SDS used in the preparative gel, the loading buffer and gel running buffer was prepared as described in (Hunkapiller, et al., 1983). Also, the running buffer included 0.1mM Na thioglycolate (Sigma) and the gel was pre-run 1-2 minutes prior to loading.

DNaseI footprinting and DMS interference.

DNase 1 footprinting assays were carried out in conjunction with the mobility shift assay. For the DNaseI reactions, a typical mobility shift reaction was scaled up in size 3-5 times. After binding

for 10 minutes at room temperature, the reaction was brought to 1.25 mM $CaCl_2$ and immediately treated with 5 to 100 ng DNase I (Worthington) for 1 minute at room temperature. EGTA was added to 1.25 mM and the samples were immediately loaded onto a mobility shift gel. After electrophoresis, the free and shifted probe bands were excised and purified as described in (ref red book). The eluted DNAs were fractionated on 10% polyacrylamide-50% urea gels and visualized by autoradiography. For DMS interference, CAC1 probe was lightly methylated as described in (Maxam and Gilbert, 1980). This probe was used in a preparative scale mobility shift reaction and the bound and free probe were separated and gel purified as above. To cleave at the methylated residues, the gel purified DNAs were treated with 10% piperidine at 90°C for 30 minutes. The samples were then lyophilized, dissolved in water and lyophilized again. After multiple cycles of resuspension and lyophilization, the samples were separated on a 10% polyacrylamide-50% urea gel and visualized by autoradiography.

Phosphatase treatment

For the experiment in Figure 5A, 4U calf-intestinal alkaline-phophatase (Boerhinger, 24U/µl) was combined with 10 µl of CACD affinity eluate. At the indicated time, 2 µl CACD was withdrawn, added to a mobility shift binding reaction and loaded onto a running mobility shift gel. The reactions that included vanadate were assembled separately and the vanadate was combined with the CACD sample prior to the phosphatase.

Proteins

MEL cells (Deisseroth and Hendrick, 1978) were grown in suspension at 0.5 to 1.5 x 10^6 cells per ml in Jokliks MEM (Irvine Scientific) supplemented with 5% newborn calf serum (Gemini Biosciences) and penicillin-streptomycin (100U/ml pennicillin G, 100 ug/ml streptomycin SO_4).

Nuclear extracts were prepared as in Dignam et. al. except that cells were resuspended in only 1.5 ml buffer C per 10^9 cells (Dignam, et al., 1983). All buffers used in the preparation of nuclear extracts and in the purification of proteins included 0.1% NP-40 (Sigma), 1 mM DTT, 0.5 mM phenylmethylsulfonyl fluoride (PMSF; Sigma), 1 mM Benzamidine (Sigma), 1 ug/ml aprotinin (Boehringer), 1 ug/ml Leupeptin (Boehringer) and 1 ug/ml Pepstatin A (Sigma). All manipulations of protein were carried out at 0-4 ^oC. Proteins were stored at -70 ^oC.

CACD was purified from MEL cell nuclear extracts prepared from 1.7 x 10^{11} cells. Extracts were dialyzed for 2.5 hours at 4^oC in approximately 20 volumes of buffer containing 20 mM Hepes pH 7.9, 10% glycerol, 0.4 mM EDTA, 0.1% NP40 and 0.1 M KCl. The dialyzed extracts were divided into five separate aliquots and chromatographed individually on a 5 cm x 61 cm, 1.2 L S-300 column that had been equilibrated in dialysis buffer containing 0.5 M KCl. Fractions were assayed for CACD binding activity by the mobility shift assay. Compared to marker proteins, CACD reproducibly eluted from the column in the position of 180 kDa (Between aldolase and catalase). Peak fractions were pooled and dialyzed in 8 L of buffer D (20% glycerol, 20mM Hepes pH 7.9, 0.4

mM EDTA, 0.1% NP40) containing 0.1 M KCl for 2 hours. After a clearing spin, these fractions (294 ml) were loaded onto a 60 ml DNA-cellulose (Sigma) column equilibrated in buffer D containing 0.1 M KCl. This column was washed with buffer D containing 0.2 M KCl and eluted with buffer D containing 0.4 M KCl. The peak fractions were pooled and diluted 1:1 with buffer D that lacked KCl to bring the KCl concentration to 0.2 M. Poly dI·dC (Sigma)was added to 20 ug/ml. Because we have found that in highly purified preparations of CACD, carrier protein helps to maintain DNA-binding activity, insulin (Boehringer Mannheim) was added to 0.1mg/ml. These fractions were divided into two aliquots. Each aliquot was individually loaded onto a 14 ml CACCC-Sepharose column, washed with 30-40 ml of buffer D containing 0.4 M KCl and eluted in Buffer D containing 0.6 M KCl. The CACCC-Sepharose column was prepared as described in (kadonaga methods) using the CACCC oligonucleotide described above.

Protein Determinations

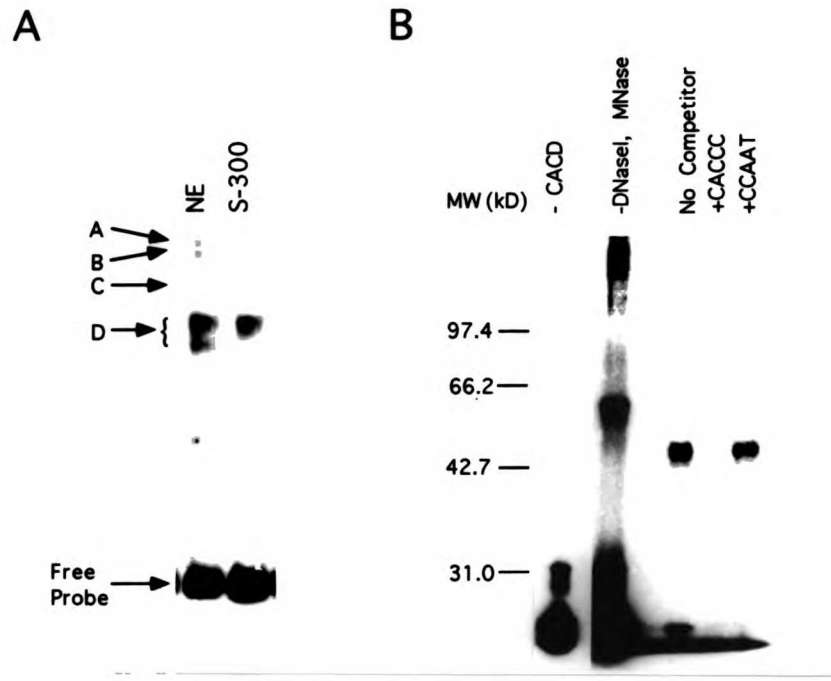
Protein concentrations were measured by using the Bradford reagent (Pierce), with BSA as a standard, except for the purified fractions, which were run on SDS polyacrylamide gels, visualized by silver staining and compared with known amounts of standard proteins.

RESULTS

Initially, we fractionated crude nuclear extracts of MEL cells on various chromatographic matrices. On several of these, it was possible to separate CACD binding activity from the other CACCC binding activities present in the extracts. For example, Figure 1A shows that passage of a nuclear extract over a Sephacryl S-300 column allows separation of CACD from other CACCC element binding activities. This separation allowed us to quickly determine the approximate molecular weight of CACD by a UV-crosslinking procedure. Partially pure fractionsof CACD were mixed with a CACCC element probe that had been internally labeled with ^{32}P and that contained bromo-deoxy-uridine rather than thymidine. This mixture was briefly irradiated with UV light to cross-link the protein-DNA complexes and was then treated with DNAse1 and micrococcal nuclease to trim away most of the probe from the cross-linked complexes. The products of this reaction were separated on a SDS-polyacrylamide gel, revealing a protein-DNA complex of 45-46.5 kD (Figure 1B). By two criteria, the signal observed in this experiment is specific to CACD. First, no other CACCC binding activities are observed in mobility shift assays of the fractions used for UV crosslinking and second, inclusion of specific competitor DNA in the crosslinking reaction abolished the signal while a non-specific competitor had no effect (Figure 1B).

With an estimate of the size of the CACD protein in hand, we proceded to purify this activity by gel filtration on a Sephacryl S-300 column followed by non-specific and then site-specific DNA-affinity

Figure 1. Panel A shows a gel mobility shift assay of a MEL cell nuclear extract (NE) and an S-300 column CACD pool fraction (S-300) using the CAC1 probe. CACCC binding activities A-D are marked on the right. Panel B shows a UV-crosslinking assay performed using affinity purified CACD and the ^{32}P-labeled, BrdU substituted CAC1 probe. In the lane labeled -DNaseI, MNase (MNase=micrococal nuclease), the UV irradiated sample was not treated with DNaseI or micrococal nuclease. Specific (CACCC) or non-specific (CCAAT) competitor DNAs were used at a 33 fold molar excess to the CAC1 probe. UV-crosslinking of less purified fractions yielded similar results. No sample was loaded in the lane between the no nuclease sample and the no competitor sample.


A
NE
S-300
A
B
C
D
Free
Probe
B
MW (kD)
- CACD
-DNaseI, MNase
No Competitor
+CACCC
+CCAAT
97.4
66.2
42.7
31.0

chromatography. These efforts resulted in a 94 fold purification of the CACD binding activity (Table 1). Silver staining SDS-polyacrylamide gels of the most purified fractions of this activity revealed a 44 kDa protein that appeared to copurify with the binding activity (Figure 2A). A number of other bands appeared on silver stained SDS-polyacrylamide gels of fractions from the DNA affinity column. Most of these bands clearly do not co-fractionate exactly with the binding activity. Two bands, one that migrates just above the 44 kD band and another that migrates at approximately 40 kD, nearly co-fractionate with CACD activity. However, both of these bands occur in fractions that lack CACD activity, 37 and 38. Also note that the CACD binding activity migrates as a doublet in the gel mobility shift assay. This is a variable behavior observed for several CACCC binding activities and may represent an artifact of the gel electrophoresis system.

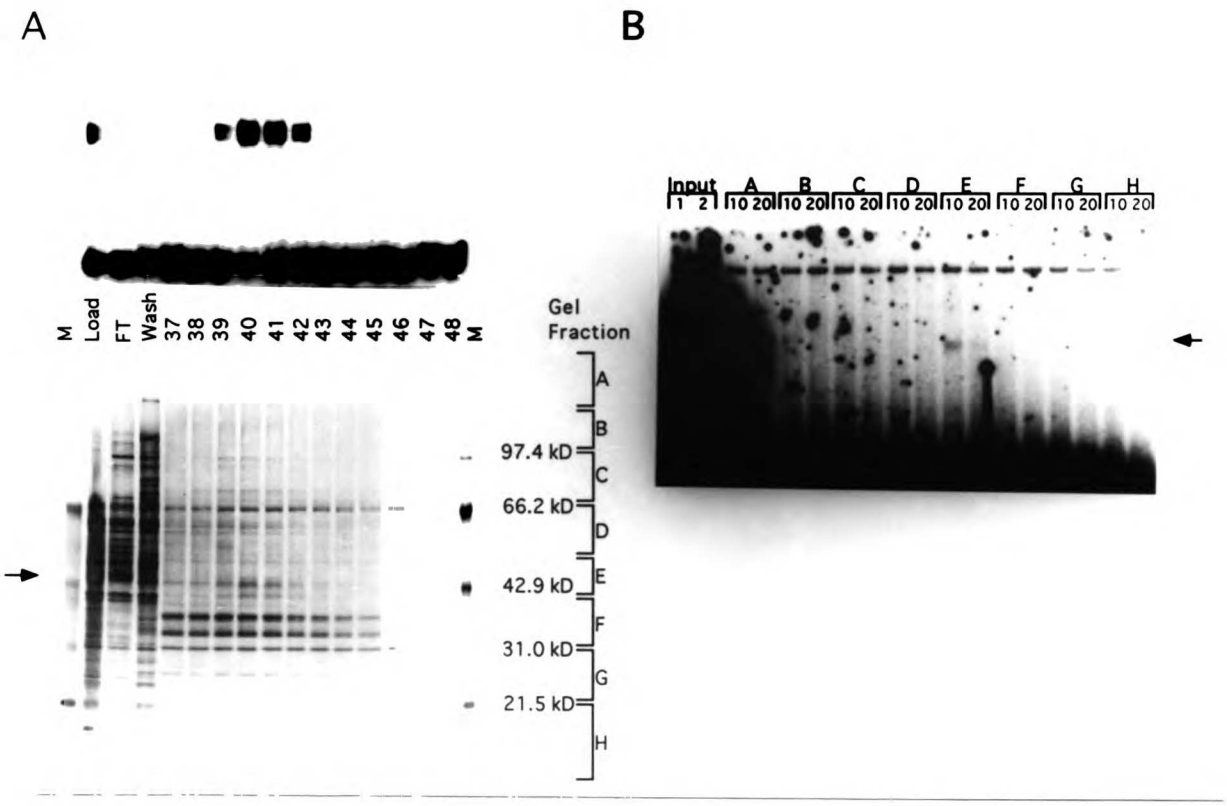
As an additional test of the identity of the 44 kDa band, we attempted to gel purify CACD, denature it and then renature the protein and assay for its activity. Affinity purified CACD was separated on a preparative SDS gel, the gel was divided into fractions (As indicated in figure 2A) and the proteins in each fraction were eluted. The eluted proteins were denatured with 6M guanidine, renatured by progressive dilution in D*0.1 buffer and assayed for CACCC binding activity (Figure 2B). The proteins eluted from the SDS gel in the 52 kD to 42 kD molecular weight range gave rise to a CACCC binding activity with a mobility identical to that of CACD (Figure 2B). No other CACCC binding activites were observed in this experiment. Although it is possible that our CACD fractions contain

Table 1. Purification of CACD

Fraction	Volume (ml)	Protein Conc. (mg/ml)	Total Protein (mg)	Specific Activity (U/mg)	Total Activity (U)*	Yield (%)	Fold Purification
Nuclear Extract	295 ml	6.55	1932	413	798×10^3	100%	
S-300	294 ml	1.23	362	541	196×10^3	25%	1.3x
DNA Cellulose	67 ml	0.28	18.8	14.3×10^3	268×10^3	34%	35x
CACCC Sepharose	24 ml	~.075	1.8	38.9×10^3	70.1×10^3	8.8%	94x

* 1 unit of activity is defined as the amount of CACD required to shift 0.15 fmol probe in a binding reaction containing a total of 3 fmol probe.

Figure 2. Panel A. The top portion of this figure shows a gel mobility shift assay of a CACCC-Sepharose column run. 1μ l each of the load (DNA cellulose pool), flow through, wash and eluted (1ml each) fractions is assayed here in a reaction that contained 3 fmol CAC1 probe. In the bottom panel, 10μl of the DNA cellulose (2.8 μ g), 20 μl of the flow through and wash fractions, and 100 μ l each of the eluted fractions were separated on an SDS-polyacrylamide gel and silver-stained. The arrow indicates the position of the 44 kD band that copurifies with CACD activity. Note that a band that migrates just above the 44 kD band is present in fractions that lack CACD activity. Also, a band at ~40 kD also appears to copurify with CACD activity, but is clearly present in fractions that lack CACD activity (37,38). Panel B shows a gel mobility shift assay of denatured-renatured protein samples that had been eluted from a preparative SDS-polyacrylamide gel of affinity purified CACD. The letters indicate the gel fraction assayed (see panel A) in each lane. The mobility shift reactions in this case were scaled up to 50μl and contained either 10 or 20μl of renatured material.

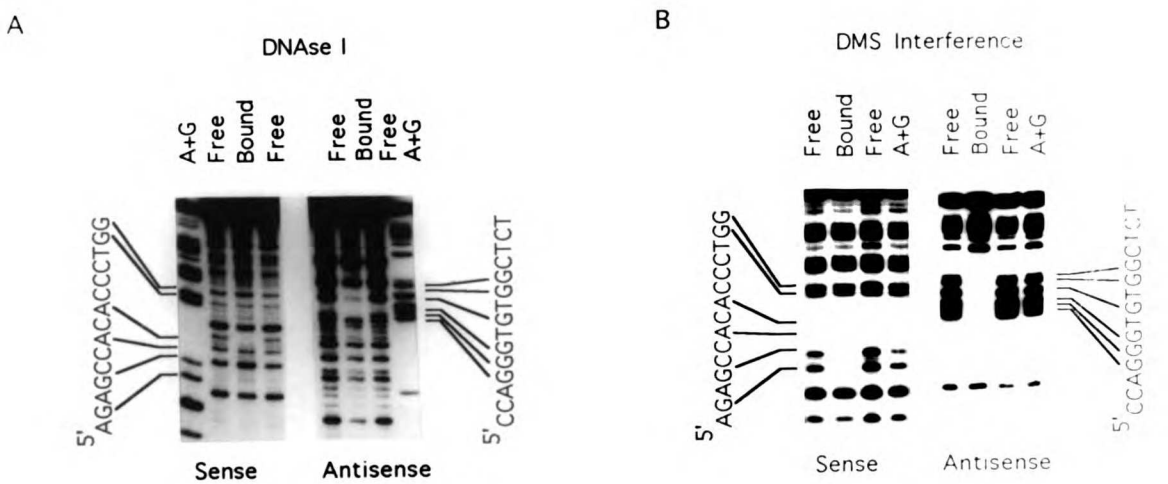

A
B
M
Load
FT
Wash
37
38
39
40
41
42
43
44
45
46
47
48
M
Gel
Fraction
97.4 kD
66.2 kD
42.9 kD
31.0 kD
21.5 kD
A
B
C
D
E
F
G
H
Input
1
2
10
20

other activities that bind the CAC1 probe and give rise to a gel mobility shift with the mobility we ascribe to CACD, at a minimum, a protein that migrates between 42 and 52 kD on an SDS-polyacrylamide gel can give a CACD-like mobility shift.

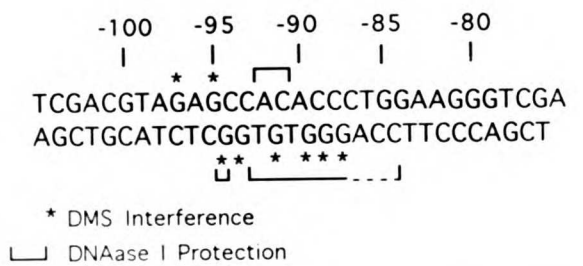
We further characterized CACD using DNase1 footprinting and DMS interference assays (Figure3). The DNase1 footprint of CACD was extensive on the G-rich, antisense strand of the CAC1 probe but was restricted to only a few nucleotides on the C-rich sense strand of this element (Figure 3A, summarized in Figure 3C). Extensive protein-DNA contacts on the antisense strand were also indicated by the results of DMS interference experiments (Figure 3B). Since no G residues occur in the core of the sense strand of the β-globin CACCC element, DMS interference provided little data for this strand. However, methylation of either of the two G residues 5' of the CACCC element on the sense strand prevented CACD binding. Additionally, under the conditions that our DMS interference experiments were performed, a low level of minor groove, N3 methylation of As was observed. Methylation of either of the A residues in the center of the CACCC sequence on the sense strand also inhibited binding of CACD. This result may be indicative of minor groove protein-DNA contacts or of indirect effects of minor groove methylation on CACD-DNA interactions. These data are summarized in figure 3C.

Because CACD shares many binding characteristics with Sp1, a zinc finger DNA binding protein (See chapter 2), we suspected that CACD might also be a zinc finger protein. This suspicion is based on the fact that DNA binding proteins that fall within a structural family tend to bind to similar sequences (Johnson and McKnight, 1989).

Figure 3. DNAse I footprinting and DMS interference assays of CACD. Panel A. DNase 1 footprint of CACD on the CAC1 probe. A+G refers to an A+G sequencing track marker lane. Free refers to unbound, DNase I treated probe that was separated from CACD bound probe (Bound) on a preparative mobility shift gel. Sense refers to the strand that is in the sense orientation with respect to the β-globin gene. This is the top strand shown in panel C. Antisense refers to the bottom strand in panel C. Panel B shows DMS interference analysis of the interaction of CACD with the CAC1 probe. The lane designations are as above. Panel C shows the CAC1 probe with a summary of the DNase I footprinting and DMS interference data. Bracketed regions indicate nucleotides protected from DNase I. The dashed brackets indicate partial protection. The starred nucleotides interferedd with CACD binding when methylated. The numbers above the sequence indicate the position relative to the transcription start site.


A
DNAse I
A+G
Free
Bound
Free
Free
Bound
Free
A+G
5' AGAGCCACACCCTGG
5' CCAGGGTGTGGCTCT
Sense
Antisense
B
DMS Interference
Free
Bound
Free
A+G
Free
Bound
Free
A+G
5' AGAGCCACACCCTGG
Sense
Antisense

C


-100
-95
-90
-85
-80
TCGACGTAGAGCCACACCCTGGAAGGGTCGA
AGCTGCATCTCGGTGTGGGACCTTCCCAGCT
* DMS Interference
DNAase I Protection

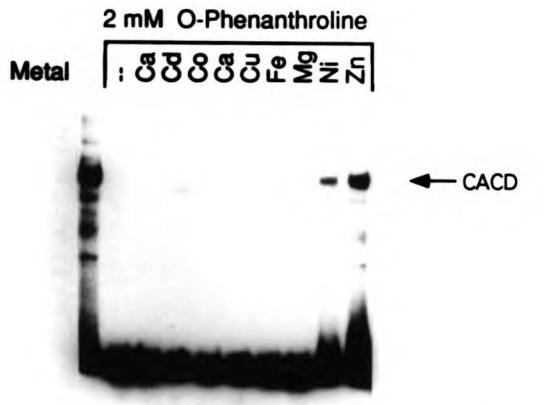
While some DNA sequences are known to be bound by proteins from different families of DNA-binding proteins, these appear to be the exception rather than the rule.

To test our hypothesis that CACD was a zinc finger protein, gel mobility shift binding reactions containing CACD were treated with the chelator 1,10-O-phenanthroline to remove divalent cations. This treatment resulted in loss of CACD binding activity (Figure 4A, lane 2). Substantial CACD binding activity could be restored when the divalent cations zinc or nickel were included in addition to the 1,10-O-phenantroline (Figure 4A, lanes 9 and 10). Other divalent cations including magnesium, did not restore CACD binding activity (Figure4A, lanes 3-8). Thus, CACD does not have a general requirement for divalent cations but rather it has a specific requirement for zinc ornickel. Since zinc ions were by far the most efficacious in restoring binding activity, we interpreted these experiments as being consistent with CACD being a zinc finger protein.

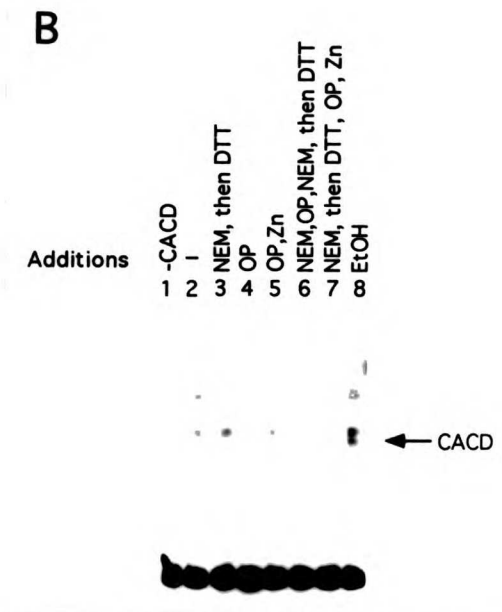
A second predicition of our model that CACD is a zinc finger protein is that the zinc ion should interact with a thiol group of a cysteine and that this would potentially protect it from alkylation by N-ethyl-maleimide (NEM). To test this hypothesis, gel mobility shift binding reactions containing crude MEL nuclear extracts were assembled (Figure 4B). When treated for 5 minutes with 1 mM NEM and then chased with excess DTT, no change in CACD binding was observed (Figure 4B, lane3). As before, treatment with 1-10-O-phenanthroline abolished all CACCC binding and the addition of zinc to these inhibited reactions restored CACD binding (Figure 4B, lanes 4

Figure 4. Inhibition of CACD binding by 1,10-O-phenanthroline and by NEM. Panel A shows a gel mobiltiy shift assay of MEL nuclear extracts using the CAC1 probe. In lane 1, no additions were made. In subsequent lanes, 2 mM 1,10-O-phenanthroline was included in the binding reaction. In lanes 3-10 mM of the indicated metal was included. The position of the CACD-CAC1 protein-DNA complex is indicated by the arrow. Panel B shows a mobility shift assay reaction that contained MEL nuclear extracts and the CAC1 probe. In lane 1, the nuclear extract was omitted. In lane 2, no additions were made. In lane 3, 1 mM NEM was added for 5 minutes and then DTT was added to 2.5 mM. In lane 4, 1 mM 1,10-O-phenanthroline was added. In lane 5, 1 mM 1,10-O-phenanthroline and 1 mM Zinc acetate were present. In lane 6, 1 mM 1,10-O-phenanthroline, 1 mM zinc acetate and 1 mM NEM were incubated with the nuclear extract for 5 minutes. DTT was then added to 2.5 mM. In lane 7, the nuclear extract was first treated with 1 mM NEM for 5 minutes, then with 2.5 mM DTT for 2 minutes followed by addition of 1 mM 1,10-O-phenathroline and 1 mM zinc acetate. In lane 8, 1 ml of 20% EtOH (The same solution that the 1,10-O-phenathroline was made up in) was included.

A


2 mM O-Phenanthroline
Metal
: Ca Cd Co Ca Cu Fe Mg Ni Zn
CACD

B

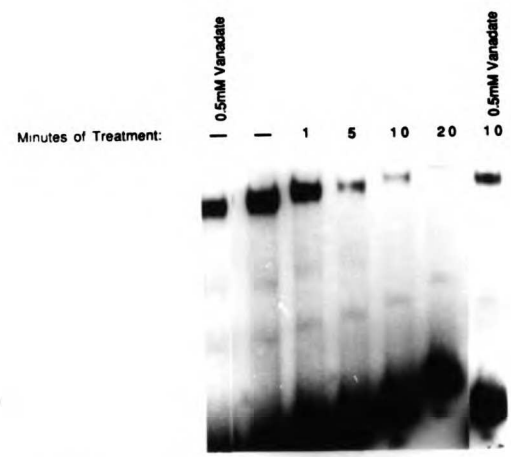

Additions
-CACD
-
NEM, then DTT
OP
OP,Zn
NEM,OP,NEM, then DTT
NEM, then DTT, OP, Zn
EtOH
1 2 3 4 5 6 7 8
CACD

and 5). If 1mM NEM was included during treatment with 1,10-O-phenanthroline, no CACCC binding activity was recovered, even if zinc was present during this time (Figure 4B, lane 6). However, if nuclear extracts were first treated with NEM, then chased with DTT and subsequently treated with 1,10-O-phenanthroline, binding activity was maintained Figure 4B, lane 7). Thus, CACD binding activity (as well as the other CACCC binding activities), was sensitive to NEM treatment only if 1,10-O-phenanthroline was similtaneously present. We interpret these results to mean that chelation of zinc causes a reversible structural change in CACD that exposes a thiol group whose integrity is required for CACD binding activity. This is consistent with the model that CACD is a zinc finger protein.

A number of DNA binding activities depend upon phosphorylation for their DNA binding activity (Larson, et al., 1988; Yamamoto, et al., 1988). To ask if CACD DNA binding activity is dependent on phosphorylation, CACD containing gel mobility shift reactions were briefly incubated with calf-intestinal-alkaline phosphatase (CIP) before being loaded on a mobility shift gel (Figure 5). These experiments show that CACD binding activity decreases in a time dependent manner during treatment with CIP and that this decrease is prevented by the phosphatase inhibitor vanadate or by inactivating CIP by heat treatment (Figure 5 and data not shown). Also, when crude nuclear extracts were used in these experiments, the only CACCC binding activity affected by phosphatase treatment was CACD (Data not shown). This result suggests that the inhibition of CACD binding was due to a specific dephosphorylation event

Figure 5. Gel mobility shift assay of CACD treated with calf intestinal alkaline phosphatase. Affinity purified CACD was combined with calf-intestinal alkaline phosphatase for the indicated and then combined with the CAC1 probe loaded onto a mobility shift gel. To demonstrate the specificity of the effect of phosphatase, vanadate was included in one phosphatase treated reaction. As can be seen from the reaction on the left, vanadate alone slightly inhibits CACD binding.

Phosphatase Inhibits CACC Binding



rather than the presence of a contaminating protease that destroyed CACD.

DISCUSSION

In this work we have partially purified and biochemically characterized a CACCC element binding protein, CACD. We first partially purified CACD and then used that material in a UV-crosslinking procedure to identify a 45-46.5 kD protein that specifically binds the CACCC element (Figure 1B). Crosslinking in the presence of specific and non-specific competitor DNA suggested that this interaction was sequence specific. Although it is possible that CACD is composed of multiple subunits, only one of which is visualized in this assay, we expected that any pure preparation of CACD would minimally include a protein of approximately 45-46.5 kD. CACD was purified to near homogeneity using a combination of gel-filtration and DNA-affinity chromatography. Silverstaining of SDS-polyacrylamide gels of affinity purified fractions revealed a slightly diffuse band at 44 kD that copurified with CACD binding activity (Figure 2A). No other band on the silverstained gels correlated with CACD activity. We believe the small discrepancy between the mobility of the signal observed in the crosslinking experiment versus that of the silverstained band identified in affinity purified CACD fractions is due to the added mass of the covalently bound DNA in the crosslinked DNA-protein complex (Note that without nuclease treatment, the protein-DNA complex is much larger; Figure 1B). To reinforce our impression that CACD binding

activity correlated with this 44 kD band, we separated affinity purified CACD fractions on a preparative SDS-polyacrylamide gel, eluted the proteins, put them through a denaturation-renaturation procedure and assayed for CACCC binding (Figure 2C). This experiment demonstrated that the 42 kD to 52 kD molecular weight range contained a protein(s) that gave rise to a CACCC binding activity with a mobility identical to that of CACD. It is possible that one of the less abundant bands in this region of the SDS gel gives rise to the binding activity observed in the renatured samples. However, given the relative abundance of these bands compared to the 44 kD band and the fact that they do not copurify exactly with CACD activity, we feel this is unlikely. Therefore, we conclude that we have purified a CACCC binding activity 94 fold, and that this activity is most likely due to the 44 kD protein described above. Definitive proof that the 44 kD band is CACD will require either a CACD specific antibody or further purification of CACD.

Characterization of CACD led to two interesting findings. First, treatment with phosphatase inhibits CACD binding activity. This effect is specific in that inhibition of phosphatase with vanadate or heat treatment preserved CACD activity (Figure 5 and data not shown). A further argument for specificty comes from the fact that when crude nuclear extracts were treated with phosphatase, none of the CACCC binding activities except CACD was affected (GH and RMM unpublished information). However, as yet, we have not been able to demonstrate restoration of CACD binding activity by addition of a protein kinase. Since CACD has not been purified to homogeneity, two interpretations of these results are possible. First, CACD may be

a phosphoprotein whose DNA binding activity depends upon its phosphorylation state. Second, the effect of phosphatase on CACD binding activity may be the result of some intermediate protein that can affect CACD in a phosphatase-dependent manner. Either explanation has potentially important implications for the regulation of CACD binding activity. However, so far we have not observed large changes in CACD activity that indicate situations in which CACD levels are manipulated, i.e. specific induction or down regulation of CACD activity during differentiation of MEL cells (See chapter 2). Therefore, at the moment, we are unaware of situations in which CACD activity is likely to be modulated. Additional reagents, CACD antibodies and larger amounts of purified CACD protein, should facilitate more definitive tests of the role of phosphorylation in the regulation of CACD activity.

A second set of findings during our characterization of CACD suggests that CACD is a zinc-finger protein. First, CACD binding activity is dependent upon divalent cations, most likely in a zinc specific manner (Figure 4A). Second, CACD binding activity is sensitive to NEM when it is depleted of divalent cations and resistant to NEM when divalent cations are in excess (Figure 4B). At the very least, these results suggest that CACD undergoes a reversible structural change, when depleted of divalent cations, that exposes a cysteine, necessary for DNA binding, to the solvent. These results are consistent with the model that CACD is a zinc finger protein and that zinc protects the cysteines of the finger(s) from irreversible alkylation by NEM.

If CACD is a zinc finger protein, several of its properties suggest that it belongs to Cys_2-His_2 class of zinc finger protein that is typefied by Sp1 (Harrison, 1991). First, unlike the Gal4 or steroid hormone receptor types of finger proteins, Cys_2-His_2 protein binding sites are not dyad symmetric (Harrison, 1991). The β-globin CACCC site lacks dyad symmetry and site selection experiments demonstrate that the β-globin CACCC element, rather than some symmetrized version of it, is the high affinity CACD binding site (See chapter 2). Second, Cys_2-His_2 protein binding sites tend to be G-rich on one strand with most protein-DNA contacts occuring on this strand (Harrison, 1991). Again, the anti-sense strand of the CACCC site is G rich, GGGTGTGG, and methylation interference shows that methylation of any of these Gs inhibits CACD binding. Also, the DNase1 footprint of CACD is much more extensive on the G rich strand than it is on the C rich strand. Finally, the binding characteristics of CACD in DNase1 footprinting and DMS interference assays are very similar to those of SP1 (See chapter 2). As discussed in chapter 2, CACD and Sp1 are unlikely to be directly related; CACD is not recognized by an anti-Sp1 antibody, a proteolytic clipping assay differentiates between Sp1 and CACD and they show slight but significant differences in their abilities to bind to sequence variants of the CACCC and Sp1 binding sites. In addition to these differences between Sp1 and CACD, Sp1 protein does not appear in the most highly purified CACD fractions. In fact, Sp1 protein that does co-fractionate with CACD through the DNA cellulose column flows through the CACD-Sepharose affinity column (Data not shown). Therefore, we conclude that CACD is most likely a Cys2-His2 zinc

finger protein that is minimally composed of a 44 kD subunit that is related to but distinct from Sp1. Proof of this hypothesis and exploration of CACD's role in β-globin transcription in vivo requires additional reagents: higher purity preparations of CACD, CACD specific antibodies and preferably, CACD's cDNA. The material generated in this work is being devoted to obtaining these reagents.

REFERENCES

Charnay, P., Mellon, P., and Maniatis, T. (1985). Linker scanning mutagenisis of the 5'-flanking region of the mouse β-major-globin gene: sequence requirements for transcription in erythroid and non-erythroid cells. Mol. Cell. Biol. *5*(6), 1498-1511.

Cowie, A., and Myers, R.M. (1988). DNA sequences involved in transcriptional regulation of the mouse β-globin promoter in murine erythroleukemia cells. Mol. Cell. Biol. *8*(8), 3122-3128.

Deisseroth, A., and Hendrick, D. (1978). Human alpha-globin gene expression following chromosomal dependent gene transfer into mouse erythroleukemia cells. Cell *15*, 55-63.

Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J., and Weissman, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell *32*, 695-706.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. *11*, 1475-1489.

Efstratiadis, A., et al. (1980). The Structure and evolution of the human β-globin gene family. Cell *21*, 653-668.

Hager, D.A., and Burgess, R.R. (1980). Elution of proteins from sodium dodecyl sulfate, and renaturation of enzymatic activity: results with sigma subunit of E. coli RNA polymerase, wheat germ topoisomerase, and other enzymes. Anal. Biochem. *109*, 76-86.

Harrison, S.C. (1991). A structural taxconomy of DNA-binding domains. Nature *353*, 715-719.

Hunkapiller, M.W., Lujan, E., Ostrander, F., and Hood, L.E. (1983). Isolation of microgram quantities of proteins from polyacrylamide gels for amino acid sequence analysis. Methods Enzymol. *91*, 227.

Johnson, P.F., and McKnight, S.L. (1989). Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. *58*, 799-839.

Larson, J.S., Schuetz, T.J., and Kingston, R.E. (1988). Activation in vitro of sequence-specific DNA binding by a human regulatory factor. Nature *335*, 372-375.

Marks, P.A., and Rifkind, R.A. (1978). Erythroleukemic differentiation. Annu. Rev .Biochem. *47*, 419-448.

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

Myers, R.M., Tilly, K., and Maniatis, T. (1986). Fine structure genetic analysis of a β-globin promoter. Science *232*, 613-618.

Yamamoto, K.K., Gonzalez, G.A., Biggs, W.H., and Montimony, M.R. (1988). Phosphorylation-induced binding and transcriptional efficacy of nuclear factor CREB. Nature *334*, 494-498.

CHAPTER 4

CHARACTERIZATION AND PURIFICATION OF A CCAAT BINDING FACTOR FROM MURINE ERYTHROLEUKEMIA CELLS THAT STIMULATES β–GLOBIN TRANSCRIPTION

INTRODUCTION

Transcriptional control of the adult β-globin genes is mediated by at least three distinct sets of cis elements. The Locus Control Region, LCR, is required to create the open chromatin structure of the locus in erythroid cells (Forrester, et al., 1990). The LCR also provides cell type specific enhancer functions (Curtin, et al., 1989; Forrester, et al., 1989; Grosveld, et al., 1987; Townes and Behringer, 1990; Tuan, et al., 1989; van Assendelft, et al., 1989). The promoters of the β-globin genes are required for wild type levels of transcription, specify the transcriptional start site and mediate a functional communication with the LCR (Antoniou and Grosveld, 1990; Charnay, et al., 1985; Dierks, et al., 1983; Myers, et al., 1986; Stuve and Myers, 1990; Wright, et al., 1984). Finally, because a single LCR regulates transcription of all of the developmental variants of β-globin within the locus, other elements direct the LCR away from transcriptionally quiescent β-globin genes (Choi and Engel, 1988; Foley and Engel, 1992; Jane, et al., 1992). Understanding the nature and mechanisms of these various functional interactions will require detailed knowledge of the proteins that bind to these various cis sequences.

Analysis of deletion mutants of the mouse β-globin promoter in transient assays in MEL cells has demonstrated that the only 5' sequences necessary for accurate, high level transcription lay between nucleotides -106 and +26 relative to the mRNA cap site (Cowie and Myers, 1988). The same study showed that three sequence elements, the CACACCC, CCAAT and TATA elements, are required for efficient transcription in both uninduced and induced

MEL cells (Cowie and Myers, 1988). A later study found an additional, directly repeated element, the βDRE, required for maximal transcription that lies between the CCAAT and TATA elements (Stuve and Myers, 1990). In addition to specifying the transcription start site and making direct contributions to transcription activity, the proximal promoter mediates interactions with the LCR. Both the CCAAT and CACCC elements (Mutations in the βDRE have not been tested) must be intact for the maximal functional interaction between these two control regions (Antoniou and Grosveld, 1990). Therefore, the CCAAT and CACCC box binding factor(s) may mediate interactions with the LCR.

The CCAAT sequence motif occurs frequently in the control regions of eukaryotic genes transcribed by RNA polymerase II in many different cell types (Maniatis et al., 1987). However, there is not a single, ubiquitous protein or protein family that interacts the CCAAT element, but rather several different proteins. These proteins have overlapping sequence specificities and cellular distributions (Barberis, et al., 1987; Barnhart, et al., 1988; Chodosh, et al., 1988; Cohen, et al., 1986; Dorn, et al., 1987; Jones, et al., 1987; Kim, et al., 1988; Kim and Sheffery, 1990; Landschulz, et al., 1988; Maity, et al., 1988; Raymondjean, et al., 1988; Santoro, et al., 1988). Determination of what CCAAT factor(s) binds the β–globin promoter is a necessary first step to understanding its function.

We describe here the identification and characterization of a CCAAT binding activity from MEL cells, which we call βCP, that interacts specifically with the CCAAT sequence of the mouse β-globin promoter. We have purified βCP from MEL cell nuclear extracts to

near homogeneity by a combination of conventional chromatography techniques and DNA affinity chromatography based on the mouse β-globin CCAAT sequence. βCP activity co-purifies with a complex of polypeptides that range from 31 to 38 kD in mobility on an SDS gel. A similar set of polypeptides copurified in two other purification schemes. When eluted from a preparative polyacrylamide-SDS gel, only fractions that include all of the peptides in the 31 to 42 kD molecular weight range had CCAAT binding activity. Thus, βCP may be multipartite. Based on binding competition characteristics, its appearance in silver stain gels and the renaturation experiment, this protein appears to be identical to an α-globin CCAAT element binding protein, αCP1, and to be a homologue of the HeLa factor CP1 (Barnhart, et al., 1988; Chodosh, et al., 1988; Kim, Barnhart and Sheffery, 1988; Kim and Sheffery, 1990). Finally, partially pure βCP stimulates transcription from the β-globin promoter in vitro in a sequence-specific manner.

Materials and Methods

Cells

MEL cells (aprt-, Deisseroth and Hendrick, 1978) were grown in 6 liter suspension flasks in Jokliks MEM supplemented with 5% newborn calf serum (Gibco). Cell density was maintained at 8 x 10^5 to 1.5 x 10^6 cells/ml.

DNA

The β−globin promoter fragment that was used as a probe in the mobility shift assays was excised from the plasmid pβMT4F by

digestion with *ClaI* (at -106) and *XhoI*. This plasmid is a derivative of pβMTF (Cowie and Myers, 1988) in which nt -53 to -33 is deleted and replaced with an XhoI linker, The fragment was 5'-end-labeled at either the *ClaI* or *XhoI* site by polynucleotide kinase and γ-^{32}P-ATP and isolated on 7% native polyacrylamide gels. The βCCAAT probe was excised from pβCCAAT with *HindIII* and *BamHI*. This plasmid contains the double stranded βCCAAT oligonucleotide cloned into the *SalI* site of pSP73 (Figure 1B; Krieg and Melton, 1987). The β-globin promoter fragment, -106 to +26, was isolated from pβMTF by digestion with *ClaI* and *BglII*. The human alpha-globin competitor fragment was isolated from a *SmaI* /*NcoI* digest of pSVa1Pst (Mellon, et al., 1981). The polyomavirus early promoter fragment was an *ApaI* to *BglI* fragment from pPyCAT (Veldman, et al., 1985). The histone H4 promoter was excised from pBRH4 (Grosschedl and Baltimore, 1985) with *EcoRI* and *PvuII*; the HSV tk promoter was isolated from pSPtkneo (Alison Cowie unpublished) with *PvuII* and *PstI* and the Friend virus promoter from pβMTF with *SacI* and *SmaI*. All fragments were isolated on 7% native polyacrylamide gels and eluted by the crush-soak method (Maxam and Gilbert, 1980). The labeled fragments were quantified by TCA-precipitation and scintillation counting to determine the specific radioactivity, unlabeled fragments were quantified on agarose gels against known standards. The sequence of the CACCC double stranded oligonucleotide used for the competition assays is: 5'-TCGACGTAGAGCCACACCCTGGTAAGGG-3' / 5'-TCGACCCTTACCAGGGTGTGGCTCTACG-3'. The sequence of the CTF/NF1 binding site double stranded oligonucleotide is:

5'-GATCCTTTTGGATTGAAGCCAATATGAT-3' / 5'-
CTAGATCATATTGGCTTC AATCCAAAAG-3'.

Plasmids used in the *in vitro* transcription reactions were: pSPβM T (pβMTF with the Friend enhancer removed), pSPtkneo [the tkneo fragment from pMV6 (Maddon, et al., 1985) subcloned into pSP73 (Krieg and Melton, 1987)] and pBRH4 (Grosschedl and Baltimore, 1985).

Binding assays

Binding reactions for mobility shift assays were performed in a final volume of 20µl and included 3 fmol labeled DNA, 0.2µg/ml poly dI.dC (Sigma), 2% Ficoll (Sigma), 20mM Tris-HCl pH 7.9, 1mM EDTA, 1mM DTT, 80mM NaCl and 1 to 5µl nuclear extract or column fraction. The reactions were preincubated at 0°C for 5 min prior to addition of the labeled DNA, followed by a further 20 min at room temperature. The samples were applied to 5% polyacrylamide gels (40:1 acrylamide: bis-acrylamide) in 6.7mM Tris.HCl pH7.9, 3.3mM NaOAc, 1mM EDTA and electrophoresed in the same buffer which was recirculated. Gels were dried prior to autoradiography.

Preparative binding reactions were performed in a final volume of 50µl in the same buffer conditions as above but with 3 to 5 x 10^5 dpm labeled DNA, 0.1µg/ml polydI.dC and up to 20µl fractionated protein. Following electrophoresis, the wet gels were covered with Saranwrap and exposed to X-Ray film. Bands were electroeluted by the method of (Ausubel, et al., 1991). For interference assays, DMS treatment was as described (Maxam and Gilbert, 1980). For footprinting, DNase I digestion was carried out by addition of 1 ul 10

μg/ml DNase I (Worthington) for 10 min at room temperature followed by addition of EDTA to 20mM to stop digestion. Reactions were separated by preparative mobility shift gels as described above. Following elution, samples were further fractionated on 20% denaturing polyacrylamide gels.

The binding assays on the SDS polyacrylamide gel-eluted protein fractions were done in a final volume of 50 μl, to accommodate a larger volume of protein, with 6 fmol labeled probe.

Binding activity was quantified by excision of the gel slices corresponding to the bound and free DNA. One unit of activity is defined as the amount of protein required to shift 1 fmol of DNA under the assay conditions used.

In vitro Transcription

Reactions were performed in a final volume of 20μl containing 0.3μg supercoiled plasmid DNA, 0.2mM rNTPs, 2.5mM $MgCl_2$, 5mM Spermidine, 3% PEG 8000, 5U RNasin (Promega-Biotec) and up to 10μl nuclear extract or fractionated protein. For the experiments in which βCP was added to the *in vitro* transcription reaction, 5 to 10 ng purified protein was used. Incubation was for 60 min at 30°C. Following phenol extraction and ethanol precipitation, specific RNA was determined by S1 analysis (Favalaro, et al., 1980) with 5'-end labeled single stranded DNA probes (Cowie and Myers, 1988).

RNA polymerase II assays were done according to the method of (Reinberg and Roeder, 1987).

Fractionation and Purification

Scheme 1: Nuclear extracts from 2 x 10^{10} cells were prepared according to the method of (Dignam, et al., 1983) except that the final dialysis was against twice 50 volumes of buffer TM0.1M (TM buffer is 50mM Tris.HCl pH 7.9, 12.5mM $MgCl_2$, 1mM EDTA, 1mM DTT, 20% glycerol, phenylmethylsulfonyl fluoride (PMSF) and Benzamidine to 0.5mM and KCl added as indicated). The nuclear extract was loaded onto a 50ml heparin agarose column and washed with 2 column volumes TM0.1M. Bound protein was eluted in steps of 3 column volumes of TM0.2M and TM0.4M sequentially. The 0.2M KCl fraction from the heparin agarose column was precipitated with 50% saturated ammonium sulfate, redissolved in a small volume buffer T*0.1M (TM buffer minus $MgCl_2$, plus 0.1 % NP40) and loaded onto a 2.4 x 100cm Sephacryl S300 column (Sigma) and run in the same buffer. Fractions were assayed by mobility shift gels and the activity peak was pooled and loaded directly onto a 2ml DNA affinity column. The affinity column was made according to Kadonaga and Tjian (1986) using the βCCAAT oligos shown in Figure 2D. The column was washed with 10 volumes T*0.1M and bound protein was eluted with 2 volumes T*0.5M. The first eluate was diluted to 0.1M KCl and reapplied to the affinity resin. The bound protein eluted the second time was flash frozen at -70°C in small aliquots. This fraction was diluted five-fold in T* buffer minus KCl before use in the binding assays or transcription reactions.

Scheme 2: A nuclear extract from 2.8 x 10^{10} cells were prepared and fractionated as above through the Sephacryl S-300 step. Half of the βCP activity peak of the S-300 column was then applied to a 8 ml calf thymus DNA cellulose column equilibrated in TM*0.1M. After

loading, the column was washed with 2 volumes of TM*0.1M followed by a 40ml TM*0.1M-TM*0.6M gradient. Most βCP activity eluted between 0.4 and 0.6M from this column. These fractions were pooled, diluted 1 to 4 with TM*Ø and applied to a 1 ml βCCAAT-Sepharose column as above.

Scheme 3: Approximately 1 fourth of the S-300 activity pool described in scheme 2 was applied to a 1 ml wheat germ agglutinin agarose column equilibrated in TM*0.1. The flow through was reapplied twice to the column. The column was washed with 10 volumes TM*0.1M and then eluted with 5 volumes of TM*0.1M containing 0.3M N-acetyl-glucosamine. The eluted fractions of βCP from this column were applied to a βCCAAT column as described in scheme 1.

In vitro transcription system: The 0.4M heparin agarose fraction described in scheme 1 was dialyzed against TM0.1M and loaded onto a 20ml DEAE Sepharose column (Sigma). The column flow through was pooled and frozen at -70°C in small aliquots. The column was step-eluted with TM0.25M. NP40 was added to the pooled 0.25M peak to 0.1% and the pool was dialyzed against TM*0.1M (TM buffer plus 0.1% NP40) before freezing in aliquots at -70°C.

Protein Determinations

Protein concentrations were measured by using the Bradford reagent (Pierce) except for the purified fractions, which were run on SDS polyacrylamide gels and visualized by silver staining. Polypeptides were eluted and denaturation and renaturation were carried out according to the method of (Hager and Burgess, 1980).

Results

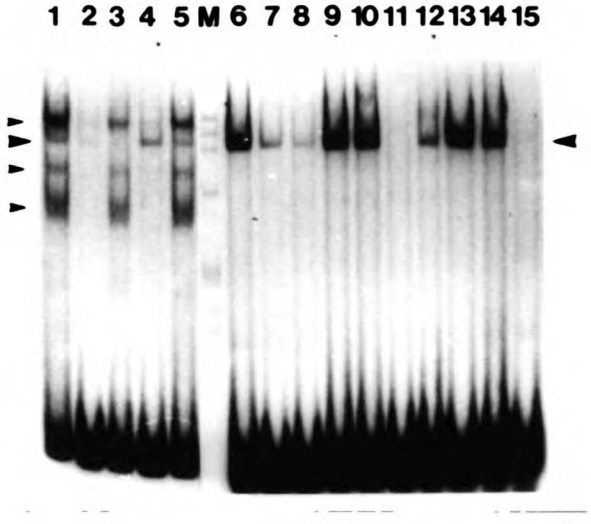
Identification of a β-globin CCAAT binding activity

We have used the gel mobility shift assay to detect a specific DNA binding activity, present in MEL cell nuclear extracts, that interacts with the mouse β–globin promoter CCAAT sequence. When a fragment of the mouse β-globin promoter (nt -106 to -54) was used as a probe in these assays, four bands of reduced mobility were observed (Fig. 1A, lane 1). This DNA fragment contains both the CCAAT sequence and the CCACACCC (CACCC) sequence. Because MEL cells contain nuclear proteins that bind to the CACCC sequence (Chapters 2 and 3), some of the shifted bands may be due to these binding activities. To determine which binding activity gave rise to each band, various unlabeled competitor DNAs were added to the binding reactions. Addition of a 100-fold excess of a DNA fragment containing the β-globin promoter from nt -106 to +26 resulted in loss of all binding species (Figure 1A, lane 2). When a DNA fragment containing only the β-globin CCAAT sequence (see Figure 1B) was added in excess, only a single binding species was abolished (Figure 1A, lane 3). Addition of a double stranded oligonucleotide corresponding to the β-globin CACCC sequence (see Materials and Methods) resulted in the loss of the other bands (Figure 1A, lane 4). When a double-stranded oligonucleotide containing a high affinity binding site for the CCAAT binding protein CTF/NF1 (see Materials and Methods; (Jones, et al., 1987)) was added in excess to the binding reaction, no loss of binding to the β-globin sequences was observed (Figure 1A, lane 5).

Figure 1.

Panel A. Competition for β CP binding assayed by gel mobility shift assay. Binding reactions, containing 3 fmol probe DNA and 17 ug MEL cell nuclear extract were carried out using either the β-globin promoter fragment as probe, lanes 1 to 5; or the pβCCAAT fragment as probe, lanes 6 to 15. Binding reactions with no added competitor DNA are shown in lanes 1 and 6; with the β-globin promoter fragment added as competitor, in lanes 2 and 7; with the pβCCAAT DNA fragment as competitor in lanes 3 and 8; with the CACCC double-stranded oligonucleotide added as competitor in lanes 4 and 9, and with the CTF/NF1 double-stranded oligonucleotide added in lanes 5 and 10. Competition for β CP binding to the pβCCAAT probe, with DNA fragments from other promoters are shown in lanes 11 to 15. Lane 11 is with the alpha-globin promoter fragment added as competitor; lane 12 with the HSV tk promoter fragment added; lane 13 with the histone H4 promoter added; lane 14 with the polyomavirus early promoter fragment added and lane 15 with the Friend virus promoter fragment as competitor. All competitor fragments were added at 100-fold excess. M denotes a labelled Hae III digest of øX174, used as a size marker. Panel B. Sequence of the double stranded oligonucleotide used to make probe βCCAAT, as a competitor DNA and to make the β CCAAT-Sepharose resin.

A



B

ß- CAAT Sequence

Xho I

TCGAGAAGGGCCAATCTGCTCG

CTTCCCGGTTAGACGAGCAGCT

Sal I

To simplify interpretation of the mobility shift gels, a DNA fragment containing the β-globin CCAAT sequence with polylinker sequences at each end of the CCAAT sequence was used as a probe (βCCAAT probe, see materials and methods and Figure 1B). When used in binding reactions with crude nuclear extracts, this probe produced a single band of reduced mobility (Figure 1A, lane 6), that migrated with the same mobility as the band abolished by CCAAT competitor when the β-globin promoter was used as a probe (compare Figure 1A, lanes 1, 4 and 6). Competition experiments carried out with this probe confirmed that specific CCAAT binding was responsible for the mobility shift. The β-globin promoter fragment (nt -106 to +26) competed efficiently for binding (Fig. 1. lane 7), as did the unlabelled βCCAAT fragment (Fig. 1, lane 8). Addition of double-stranded oligonucleotides corresponding to either the β-globin CACCC sequence or the CTF/NF1 sequence did not compete for binding (Fig. 1, lanes 9 and 10, respectively).

To determine if the CCAAT binding activity identified by the mobility shift assay was specific for the β-globin promoter CCAAT element, we used DNA fragments from other promoters in the competition experiments. Addition of a DNA fragment from the human α-globin promoter competed efficiently for binding to the βCCAAT probe (Fig. 1A, lane 11), as did the Friend virus promoter (Fig. 1A, lane 15). The HSV thymidine kinase (TK) promoter competed only weakly (Fig. 1A, lane 12), whereas the histone H4 and polyomavirus early promoters did not appear to compete at all (Fig. 1A, lanes 13 and 14). We conclude that the βCCAAT probe detects

the same CCAAT binding activity as the β-globin promoter fragment, even in the absence of the binding activities interacting at the CACCC sequence. Thus it is apparent that the CACCC sequence is not required for CCAAT binding and that, under our conditions, CCAAT binding and CACCC binding factors do not detectably interact.

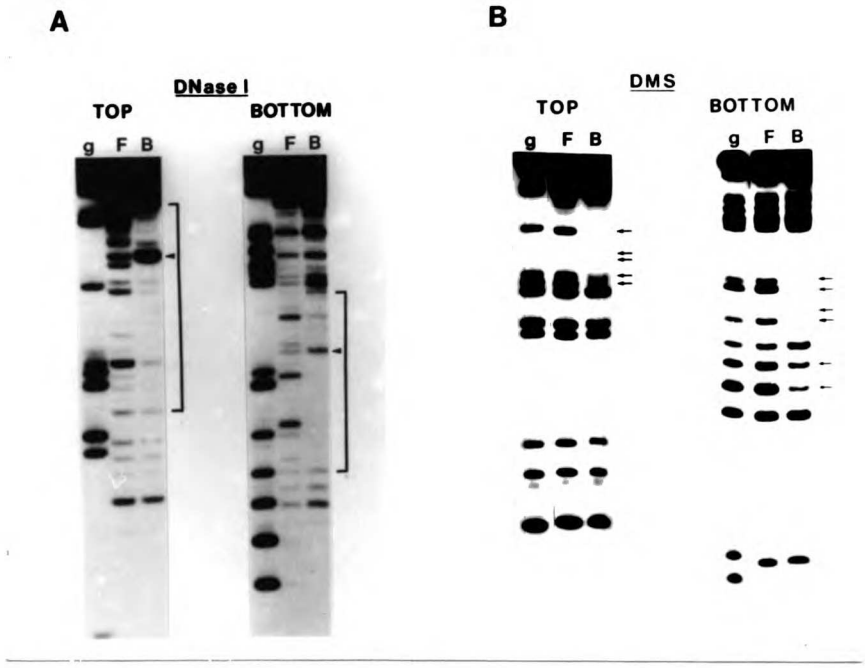
Characterization of βCP binding

The binding of βCP was further characterized by determining the extent of protection from DNase I digestion it provided to the mouse β-globin promoter (Galas and Schmitz, 1978). For these experiments, the labeled β-globin promoter fragment (nt -106 to -54) was combined with 6μg of protein from the βCP activity peak of an S-300 sizing column (See below) and treated briefly with DNase I. The bound and free probe were separated on a preparative mobility shift gel, eluted and further fractionated by denaturing polyacrylamide gel electrophoresis. The results of such an experiment are shown in Figure 2A. Protection was observed on both strands of the DNA centered around the GGCCAATCTG sequence motif. The extent of the DNase I protection observed, with respect to the promoter sequence, is diagrammed in Figure 2C.

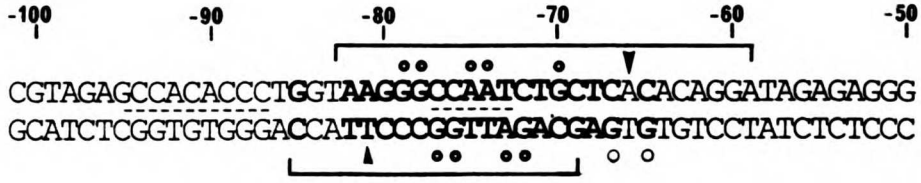
The requirements for close contacts between βCP and G and A residues in the CCAAT sequence were analyzed by DMS interference analysis, again using the -106 to -54 probe and 6μg of the S-300 activity peak. The results of this analysis are shown in Figure 2B. Bands corresponding to methylated residues that interfered with binding are indicated by arrows and their positions in the β-globin sequence are marked in Figure 2C. Methylation of any of the G and

Figure 2. Characterization of binding of βCP to the β-globin promoter.

Panel A shows the results of DNase I footprints following a preparative mobility shift gel. TOP and BOTTOM refer to the upper and lower DNA strands as written in Figure 2C. "g" is the DMS cleavage pattern of the naked DNA fragment. F and B refer to the free and bound DNA fractions from the mobility shift gel. The region of the DNA protected from DNase I digestion is bracketed, with the enhanced or unprotected bands marked by arrowheads. Panel B shows the results of DMS-methylation interference experiments following a preparative mobility gel. Designations are as in panel A. Bands corresponding to interfering G and A residues are indicated by arrows. Panel C shows a partial nucleotide sequence of the mouse β-globin promoter. The bases in bold type indicate those that are also present in the βCAAT oligonucleotide. The regions of DNase I protection are bracketed with enhanced or unprotected bands indicated by arrowheads. Dark circles mark bases that interfere strongly with binding, and open circles mark bases that interfere only partially.


A
DNase I
TOP
BOTTOM
g F B
g F B
B
DMS
TOP
BOTTOM
g F B
g F B

C


-100
-90
-80
-70
-60
-50
CGTAGAGCCACACCCTGGTAAGGGCCAATCTGCTCACACAGGATAGAGAGGG
GCATCTCGGTGTGGGACCATTCCCGGTTAGACGAGTGTGTCCTATCTCTCCC

A bases within the GGCCAATCTG sequence motif interfered strongly with βCP binding. Methylation of some of the bases downstream of this sequence also interfered with binding, although to a lesser degree than those within the CCAAT sequence (Figures 2B and 2C).

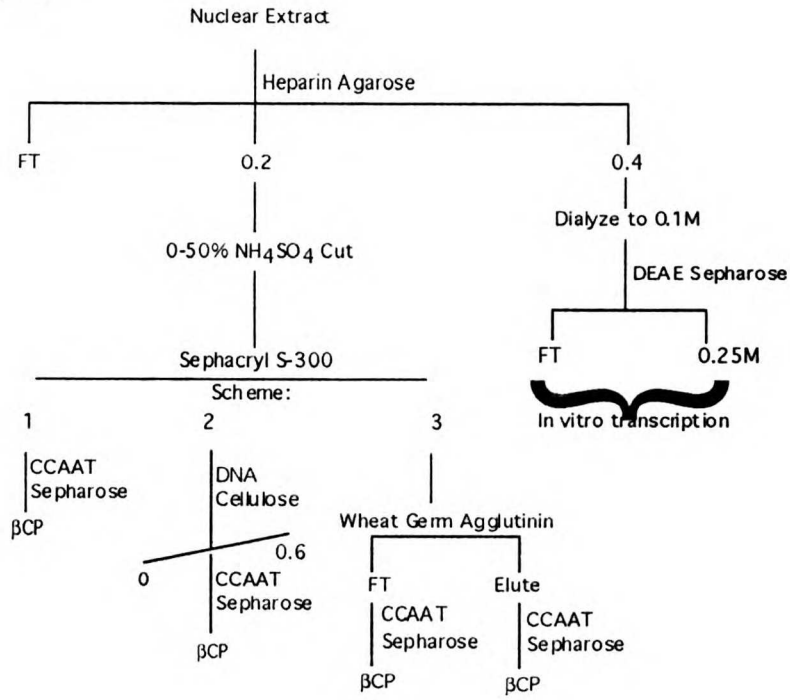
Purification of βCP

To further characterize the β-globin CCAAT binding activity, we undertook its purification from MEL cell nuclear extracts. The protocol used is outlined in Figure 3A, scheme 1. Nuclear extracts from uninduced MEL cells were applied to a heparin agarose column in 0.1M KCl. Fractionation of binding activity was assayed by mobility shift gels using the βCCAAT probe (Figure 3B). βCP binding activity eluted from the column at 0.2M KCl (Figure 3B, lanes 3). This fraction was ammonium sulfate precipitated and applied to a Sephacryl S-300 sizing column. βCP activity eluted from this column at a volume between that which catalase (240 kD) and aldolase (160 kD) marker proteins elute. Thus, βCP behaves as a large entity. The binding activity of the pooled peak is shown in Figure 3B, lanes 5. The pooled S-300 activity peak was applied to a βCCAAT-DNA Sepharose affinity column. The sequence of the double stranded oligonucleotide used to make the affinity column is given in Figure 1B; this is the same double stranded oligonucleotide used to make the probe βCCAAT. After extensive washing with buffer containing 0.1M KCl, specific CCAAT binding proteins were eluted from the column in buffer containing 0.5M KCl. The eluate was diluted to 0.1M KCl, reapplied to the column, and the washing and elution steps were repeated. Binding activity of this final, twice-eluted, fraction is

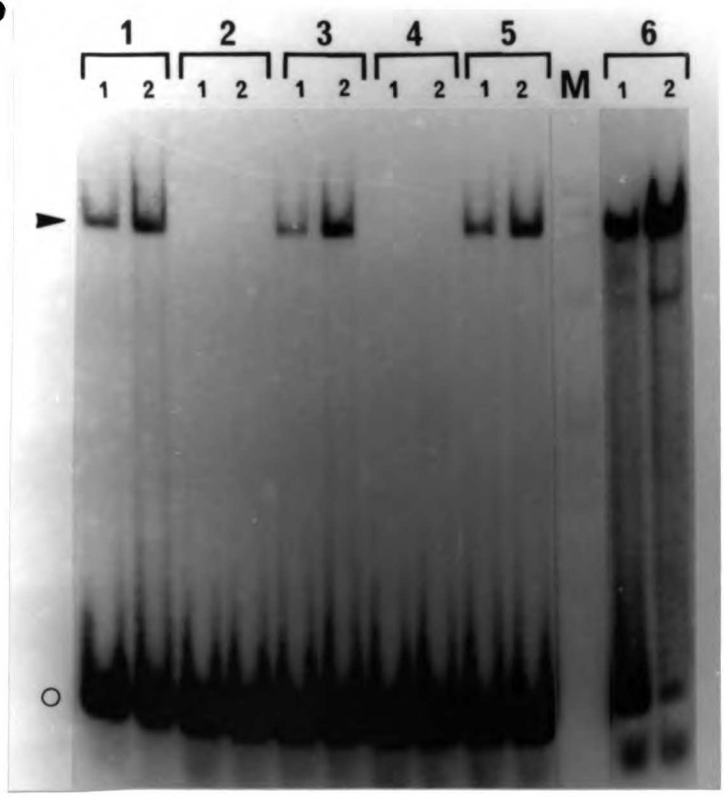
Figure 3. Purification of β CP.

Panel A shows the outline of the purification schemes for β CP and fractionation of the *in vitro* transcription reaction. Panel B shows the result of mobility shift gel electrophoresis on fractions obtained during the purification. In lanes 1:1 and 1:2, 8 and 16 ug nuclear extract are used in the binding reaction with 3 fmol of the pβCCAAT DNA probe. In lanes 2:1 and 2:2, 4.5 and 9 ug of the heparin agarose 0.1M KCl flow through fraction is assayed. In lanes 3:1 and 3:2, the binding assay contains 2.5 and 5 ug, respectively, of the heparin agarose 0.2M fraction. In lanes 4:1 and 4:2, the binding reactions contain 3.2 and 6.4 ug of the heparin agarose 0.4M fraction. In lanes 5:1 and 5:2, 0.6 and 1.2 ug of the Sephacryl S300 activity peak are assayed and in lanes 6:1 and 6:2, 4 and 8 ng from the DNA-Sepharose Elute 2 fraction is assayed. M is the labelled øX174 HaeIII digest size marker. Panel C shows a silver stained SDS-polyacrylamide gel of the final, 2nd pass, affinity purified βCP from scheme 1.

A


Nuclear Extract
Heparin Agarose
FT
0.2
0.4
0-50% NH_4SO_4 Cut
Dialyze to 0.1M
DEAE Sepharose
Sephacryl S-300
FT
0.25M
Scheme:
In vitro transcription
1
2
3
CCAAT Sepharose
βCP
DNA Cellulose
0.6
0
CCAAT Sepharose
βCP
Wheat Germ Agglutinin
FT
Elute
CCAAT Sepharose
βCP
CCAAT Sepharose
βCP

B


1
2
3
4
5
6
1 2 1 2 1 2 1 2 1 2 M 1 2

C

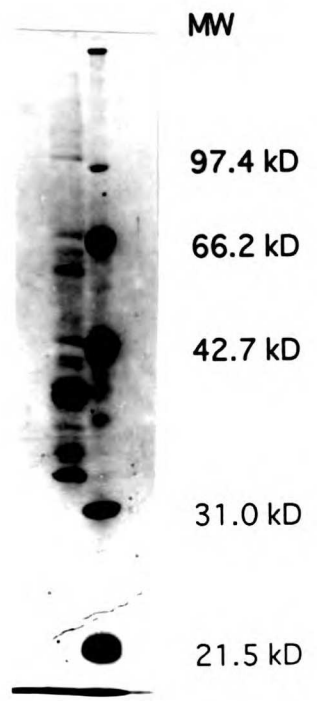

MW
97.4 kD
66.2 kD
42.7 kD
31.0 kD
21.5 kD

Table 1. Purification of βCP.

Fraction	Volume (ml)	Protein Conc (mg/ml)	Protein Total (mg)	%	Activity* Total (units)	Activity* Yield (%)	Sp. Act. (units/mg)	Purification
Nuclear Extract	51	8.8	449	100	16164	100	36	1
Heparin agarose	35	1.26	44	10	4136	26	94	2.6
S300	22	0.3	6.6	1.5	3432	21	520	14.4
DNA affinity	2	0.004	0.008	0.0018	480	3	60,000	1700

* One unit of activity is defined as the amount of protein required to bind 1 fmol of probe under standard mobility shift gel conditions.

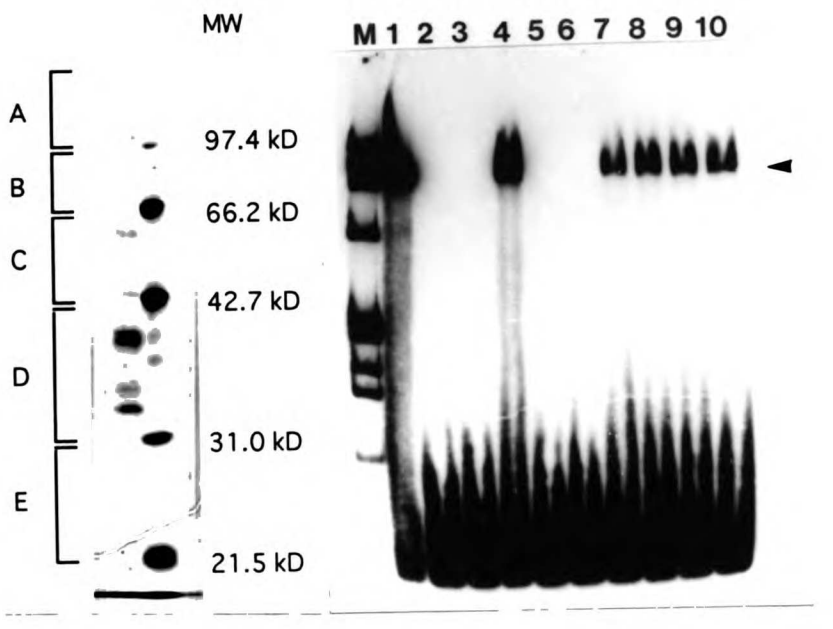
shown in Figure 3B, lanes 6. An aliquot of affinity purified βCP was separated by SDS-PAGE and visualized by silver staining (Figure 3C).

After two passes over the DNA affinity column, the eluted protein fraction still contained multiple proteins. To ascertain which of these polypeptides is responsible for βCP activity, a preparative scale SDS-polyacrylamide gel was run and sections of the gel, A to E as indicated in Figure 4, were excised and the proteins were eluted. Each eluted fraction was denatured in 6M guanidine hydrochloride and renatured by dilution (Hager and Burgess, 1980); CCAAT binding activity of each fraction was assayed by the gel mobility shift assay. A binding activity, with a mobility identical to that of βCP, was recovered from fraction C alone (Fig. 4, lane 4), the fraction containing peptides in the 31-43 kD range. When these polypeptides were eluted separately from one another (43-35 kd and 35 to 31 kD) no binding activity was recovered (Data not shown). None of the other fractions showed binding activity (Figure 4, lanes 2, 3, 5 and 6), and the binding activity of fraction C was not altered by addition of any of the other fractions (Figure 4, lanes 7 to 10). Thus, βCP activity can be recovered from a complex of 6 or more polypeptides that fall in the 31 to 43 kD range. More than one of these is required to recover binding activity from a preparative SDS gel. A summary of this purification is given in Table 1; the final purification of βCP activity after two passes over the βCCAAT affinity column was approximately 1700-fold.

To obtain βCP of higher purity, the purification procedure was modified (Figure 3A, scheme 2). After passage over the S-300 column, fractions with βCP activity were applied to a calf thymus

Figure 4. Renaturation of β CP binding activity from an SDS polyacrylamide gel.

Approximately 500 ng of purified βCP was fractionated on a preparative scale SDS polyacrylamide gel. The left panel shows the sections of the gel (Fractions A-E) relative to a silver stained standard that were excised from the preparative gel. Following denaturation-renaturation, the fractions were assayed by mobility shift gel electrophoresis. The result of a binding reaction containing 5 ul (20 ng) of purified β CP is shown in lane 1. In lane 2 the binding reaction contains 30ul (1/20) renatured fraction A; in lane 3 the reaction contains 30ul fraction B; in lane 4 the reaction contains 30ul fraction C; in lane 5 the reaction contains 30ul fraction D and in lane 6, 30ul fraction E. In lane 7, 15ul fraction C was added together with 15ul fraction A in the binding reaction; in lane 8 the binding reaction contains 15ul fraction C and 15ul fraction B added; in lane 9 the binding reaction contains 15ul fraction C with 15ul fraction D added and in lane 10, 15ul fraction C plus 15ul fraction E are in the binding reaction. M is the labelled øX174 HaeIII digest used as size marker.


MW
M 1 2 3 4 5 6 7 8 9 10
A
B
C
D
E
97.4 kD
66.2 kD
42.7 kD
31.0 kD
21.5 kD

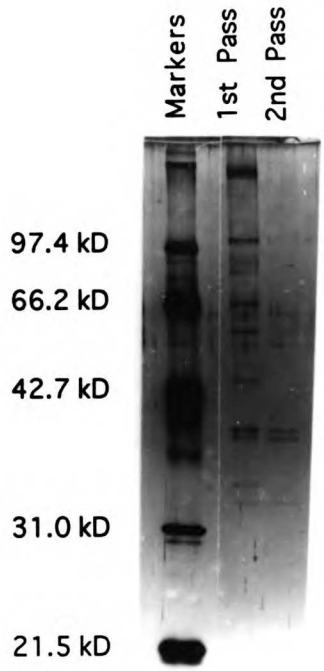
DNA-cellulose column that was then developed with a 0.1 to 0.6 M KCl gradient. βCP activity eluted between 0.4 and 0.6M KCl and was pooled. This pooled DNA-cellulose eluate was subjected to βCCAAT-Sepharose affinity chromatography as in scheme 1. Fractions of this material were separated by SDS-PAGE and visualized by silver staining (Figure 5A). This protocol yielded a preparation of βCP that was considerably purer than that generated in the previous purification scheme. Significantly, this procedure produced 4 prominent bands in the 31 kD to 39 kD molecular weight range; a triplet of bands from 37-39 kD and a single band at 31 kD that co-migrated with bands observed in silver stained βCP fractions produced by the previous protocol.

βCP was also purified by a third protocol (Figure 3A, scheme 3). Because a number of transcriptional regulatory proteins, including the CCAAT binding protein CTF/NF1, are O-glycosylated, we asked if βCP is retained on a wheat germ agglutinin (WGA) column (Jackson and Tjian, 1988). βCP was purified by heparin agarose column chromatography, ammonium sulfate precipitation and sizing on an S-300 column as before. This material was then passed over a WGA column, washed extensively and eluted with N-acetyl-glucosamine. About two thirds of the βCP activity was found in the flow through and wash fractions of this column whereas one third of the βCP activity was retained and eluted by N-acetyl-glucosamine (Figure 5B). When the flow through and wash fractions were reapplied to the WGA column, almost all βCP activity flowed through the column (Figure 5B). Thus it appears βCP can be separated into two pools in which one is glycosylated at a very low density, if at all, whereas the

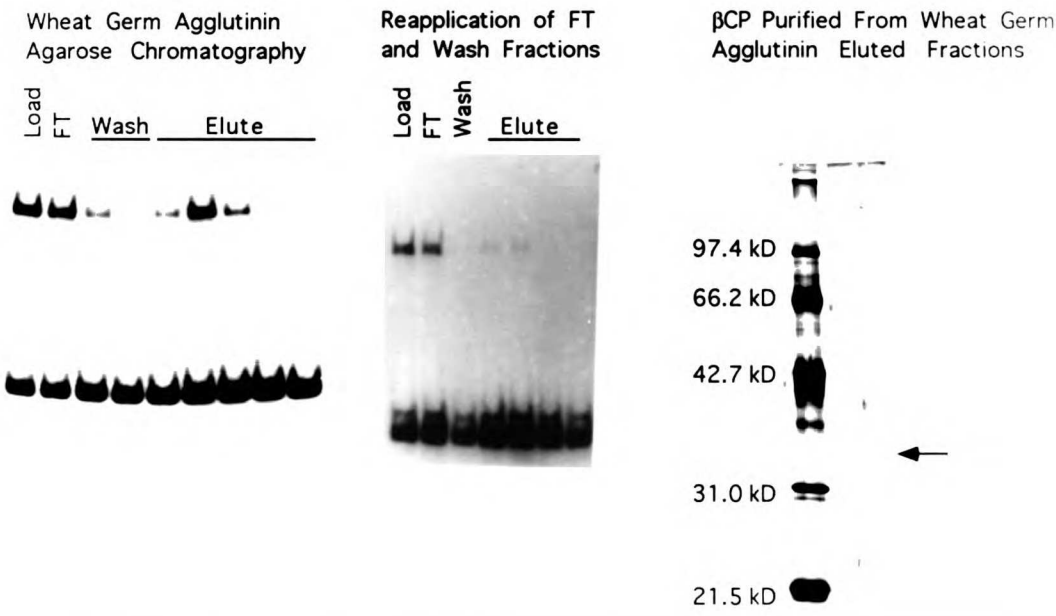
Figure 5. βCP purification schemes 2 and 3.

Panel A. Silver stained SDS-polyacrylamide gel of first and second pass affinity eluates from βCP purification scheme 2. Panel B. β CP purification scheme 3. The first panel shows a gel mobility shift assay of fractions from the wheat germ agglutinin agarose column. In the middle panel, the flow through and wash fractions from the first wheat germ agglutinin column were pooled and reapplied to the column. The right hand panel shows a silver stained, SDS-polyacrylamide gel of the βCCAAT-Sepharose purified wheat germ agglutinin eluate. The arrow at the right marks the position of the faint band at 33 kD.

A


Markers
1st Pass
2nd Pass
97.4 kD
66.2 kD
42.7 kD
31.0 kD
21.5 kD

B


Wheat Germ Agglutinin
Agarose Chromatography
Load
FT
Wash
Elute
Reapplication of FT
and Wash Fractions
Load
FT
Wash
Elute
βCP Purified From Wheat Germ
Agglutinin Eluted Fractions
97.4 kD
66.2 kD
42.7 kD
31.0 kD
21.5 kD

other is glycosylated at a level sufficient to allow its retention on the WGA column. These forms of βCP behanve identically in the gle mobility shift assay. The glycosylated form of βCP was further purified by taking the WGA eluate and passing it over a βCCAAT-Sepharose affinity column. The silver stain gel of this material showed that four bands copurified with βCP activity (Figure 5A). Two of these bands co-migrated with the upper two of the three bands previously observed at 37-39 kD. A third very faint band at 33 kD comigrated with a band observed in the first pass affinity eluates of the first and second purification procedures (Figure 5A). Finally, a fourth band, at > 100 kD, that comigrated with a band seen in scheme 1 affinity eluates was also present. In preliminary experiments, when the material that flowed through the WGA column was further purified by chromatography on βCCAAT-Sepharose, a different pattern of bands were observed. In this case, the 31, 37,38, and 39 kD bands were observed but the 33 kD and 35 kD bands were absent.

In vitro transcription

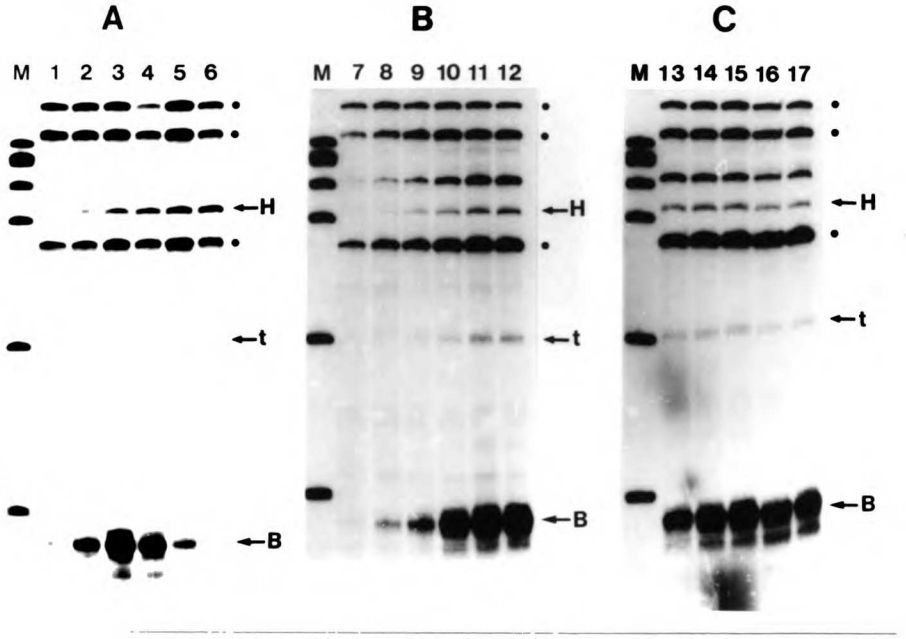
The CCAAT sequence of the β-globin promoter is known to be required for efficient transcription *in vivo* (Charnay, Mellon and Maniatis, 1985; Cowie and Myers, 1988; Dierks, et al., 1983; Myers, Tilly and Maniatis, 1986; Wright, et al., 1984). Therefore, it was of interest to determine whether the CCAAT binding factor, βCP, is functional for *in vitro* transcription from the mouse β-globin promoter. Nuclear extracts from MEL cells are competent for *in vitro* transcription reactions in which the β-globin promoter, the tk

promoter and histone H4 promoter are used as templates (Fig. 6A). Heparin agarose chromatography separated βCP activity, which elutes at 0.2M KCl, from in vitro transcription activity, which elutes at 0.4M KCl (Fig. 6B; (Dynan and Tjian, 1983)). The latter fraction has RNA polymerase II activity (data not shown), but no βCP binding activity (Fig. 3B, lanes 4). Further fractionation of the 0.4M fraction on a DEAE Sepharose column separated the transcription activity into two fractions: the 0.1M KCl flow-through (D0.1) and the fraction that eluted at 0.25M KCl (D0.25), which has RNA polymerase II activity (data not shown). Neither fraction alone was active for specifically-initiated transcription, but when added together activity was restored (Fig. 6C; (Dynan and Tjian, 1983)).

Addition of 5 to 10 ng of partially pure βCP (From scheme 1, Figure 3C, lane) to the reconstituted transcription reaction containing 6μg of the D0.1 fraction and 2μg of the D0.25 fraction resulted in a two- to five-fold increase in specific transcription from the mouse β-globin promoter (Fig. 7A, lanes 1, 2, 9 and 10). To determine whether the stimulation of transcription was due to specific binding of βCP to the β-globin promoter, an excess of a DNA fragment corresponding to the βCCAAT binding site was included in the *in vitro* transcription reaction. Addition of this DNA fragment to reactions containing purified βCP abolished the stimulatory effect (Fig. 7A, lane 3). Moreover, addition of an excess of DNA fragments containing the CTF/NF1 binding site or the CACCC binding site did not result in the loss of stimulation (Fig. 7A, lanes 4 and 5). Addition of any of these competing DNAs to the *in vitro* transcription reaction in the absence of βCP had no effect (Fig 7A, lanes 6, 7 and 8). A slight

Figure 6. Fractionation of the *in vitro* transcription reaction.

Panel A shows the results of S1 analyses of *in vitro* transcription products from reactions containing 11, 22, 44, 66, 88, and 110 ug nuclear extract (lanes 1 to 6) and 0.1ug each pSPβ MT, pSPtkneo and pBRH4 DNA. B denotes the position of the correct 5' end of the β MT transcript; t denotes the 5' end of the tkneo transcript and H marks the 5' end of the histone H4 transcript. Solid circles indicate positions of full-length protected S1 probes. M is the labelled øX174 HaeIII digest as size marker. Panel B shows the result of S1 analyses of *in vitro* transcription products from reactions containin 2.8, 5.6, 11.2, 16.8 22.4, and 28 ug of the heparin agarose 0.4M KCl fraction (lanes 7 to 12) and the same concentration of DNA templates as in panel A. Designations are also as in A. Panel C shows the results of S1 analyses of *in vitro* transcription reactions containing 3 ug D0.25 fraction and 2, 4, 6, 8, 10 ug D0.1 fraction (lanes 13 to 17), with the same DNA templates as above. Designations are as in A.


A
B
C
M 1 2 3 4 5 6
M 7 8 9 10 11 12
M 13 14 15 16 17
←H
←t
←B

increase in signal over the basal level is observed with the addition of the competitor DNA fragments, possibly due to non-specific binding of inhibitors in the reaction.

We also tested whether purified βCP stimulated transcription *in vitro* from non-erythroid promoters. Addition of βCP to the reconstituted *in vitro* transcription reactions stimulated transcription from the β-globin promoter but not the HSV tk or histone H4 promoters (Fig. 7B, lanes 9 and 10). When the unlabeled βCCAAT probe DNA fragment was included in the reaction, together with βCP, transcription from the β-globin promoter was reduced but no effect was observed on the histone H4 promoter (Fig. 7B, lanes 11 and 12). Addition of the βCCAAT probe DNA fragment resulted in a reduction of the signal from the HSV tk promoter in the absence and presence of βCP, which may indicate that the β-globin CCAAT sequence is competing for a CCAAT binding factor, different from βCP, that interacts with the HSV tk promoter.

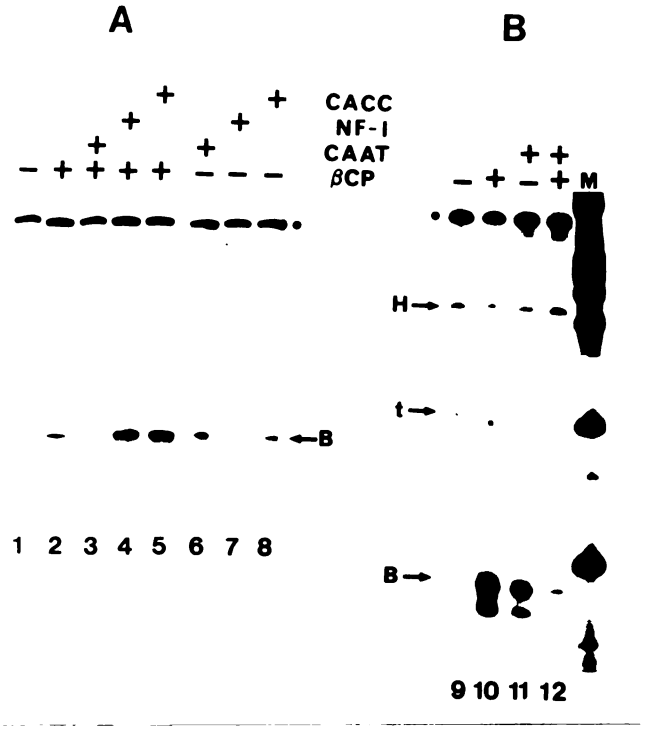
Discussion

We report here the identification and purification of a CCAAT binding factor from MEL cells, which we have named βCP. This factor binds specifically to the β-globin CCAAT element and protects a region around the sequence GGCCAATCTG from DNase I digestion (Figures 2A and 2C). Also, methylation of any of the A and G residues within this sequence interferes with binding (Figures 2B and 2C). In three different purification schemes, multiple peptides co-purified with βCP activity (Figures 3C, 5A and 5B). Two peptides, at 38 and 39 kD, co-purified in all three preparations. One peptide at

Figure 7. The effect of βCP in the *in vitro* transcription reaction.

Panel A shows the results of S1 analyses of reconstituted ***in vitro*** transcription reactions containing 2 ug D0.25 fraction and 6 ug D0.1 fraction, with pSPβMT. The addition of purified βCP and competing DNA fragments is as indicated above the lanes. B marks the position of the correctly initiated βMT transcript. Solid circle denotes the full-length protected S1 probe.

Panel B is a composite, showing the result of S1 analyses of ***in vitro*** transcription reactions, as above, containing either pSPβMT, pSPtkneo or pBRH4 DNA with addition of purified βCP and competing DNA fragments as indicated above the lanes. B is the 5' end of the βMT transcript; t is the 5' end of the tkneo transcript and H is the position of the 5' end of the histone H4 transcript. Solid circles denote full-length protected S1 probes.


A
B
CACC
NF-I
CAAT
βCP
M
H
t
B
1 2 3 4 5 6 7 8
9 10 11 12

31 kD and another at 37 kD copurified in schemes 1 and 2. In scheme 3, which included the wheat germ agglutinin column, a peptide at 33 kD copurified with βCP activity. This band also appeared in the first purification scheme. When βCP from scheme 1 was separated on a preparative SDS-polyacrylamide gel, the material from the 31 to 43 kD range could be denatured, renatured and demonstrated to have CCAAT binding activity. Material from other regions of the preparative gel did not give rise to βCP activity in this assay, nor did subfractions of the 31 to 43 kD molecular weight range. When run on an S-300 column, βCP activity elutes in the range of 200 kD proteins. Furthermore, when applied to a glycerol gradient the bands in the 31-39 kD range co-sediment (Alison Cowie and RMM, unpublished data). We interpret these results to indicate that βCP is a large hetero-oligomeric DNA binding protein (See below).

Chromatography of partially purified βCP across a wheat germ agglutinin column demonstrated that βCP consists of a densely glycosylated form and a form that is not or is only sparsely glycosylated (Figure 5B). The 38 and 39 kD bands are present in all preparations of βCP. The bands at 31 and 37 kD are exclusively associated with the unglycosylated forms of βCP and the 33 kD band fractionates entirely with the glycosylated form. Since the upper two bands at 38 and 39 kD are present in both the glycosylated and unglycosylated forms of βCP, these bands must not differ greatly in their glycosylation state. Likewise the lowest band of the triplet, the 37 kD band, which does not purify with glycosylated βCP, is unlikely to be responsible for the differential fractionation of βCP on WGA. If

this band had a glycosylated form that bound to the WGA column, a band unique to the WGA eluted material that migrated at a lower mobility than the 37 kD band would be expected. The 33 kD band, however, could represent a glycosylated form of the 31 kD band. If correct, this could explain the differing fractionation properties of βCP on wheat germ agglutinin. This model of βCP glycosylation can be tested by using conditions that chemically cleave O-linked sugars or by blotting βCP and probing with labeled wheat germ agglutinin. Also, demonstration of relatedness between the proteins of the 31 and 33 kD bands is required (See below).

βCP binding activity, as assayed on mobility shift gels, does not appear to be exclusive for the β-globin promoter. Competition with DNA fragments from other promoters revealed that both the α-globin and Friend virus promoters competed strongly for the binding to the β-globin sequence, whereas the HSV tk promoter competed weakly and the histone H4 and polyomavirus early promoters did not compete for binding to the β-globin CCAAT sequence. It is of interest to note that the promoters with the strongest affinity for this binding activity are all expressed in erythroid cells. However, βCP is not necessarily specific to this cell type. Preliminary binding experiments with nuclear extracts from HeLa cells and mouse L cells detected the presence of a gel mobility shift activity that retards the βCCAAT probe to a similar mobility as that seen with βCP (Alison Cowie and RMM unpublished data).

Numerous CCAAT binding factors have been described biochemically and a subset of these have been purified and cloned. These factors belong to a minimum of three structural families. βCP

appears to be distinct from two of these families. CTF/NF1 type CCAAT proteins differ in sequence specificity (Figure 1A, lanes 5 and 10) from βCP and at 52-66kD CTF/NF1 family members are all larger than the proteins that co-purify with βCP (Rosenfeld and Kelley, 1986). CTF/NF1 peptides can also bind DNA as homodimers whereas the results of the gel elution experiment suggest that βCP requires two or more subunits to bind DNA (Figure 4; Mermod, et al., 1989). Also, in preliminary experiments, we find that βCP is not recognized by anti-CTF/NF1 anti-sera (GH and RM unpublished data). The C/EBP family of proteins also differs in sequence specificity from βCP and at least some of its members bind DNA as homodimers (Johnson, et al., 1987; Landschulz, et al., 1989).

A third family of proteins shares a number of similarities with βCP. This family includes CP1, a factor from HeLa cells (Chodosh, et al., 1988), NF-Y, a factor that binds the major histocompatibility class II genes (Dorn, et al., 1987), CBF, a factor that binds the α2(I) collagen promoter (Hatamochi, et al., 1988; Maity, et al., 1988), and HAP2/HAP3/HAP4, a heteromeric factor from yeast (Chodosh, et al., 1988). The similarities between βCP and CP1 include: first, the rank order of affinities of βCP and CP1 for the tk, α-globin, β-globin and CTF/NF1 type CCAAT elements is the same (Figure 1A; Chodosh, et al., 1988). Second, both factors are hetero-oligomers (Figure ;(Chodosh, et al., 1988)). Third, the estimated molecular weights of the two chromatographically separated components of CP1 are 25 kD and 45 kD, values near those of peptides that copurify with βCP activity (Chodosh, et al., 1988). Also, CP1 sediments as a 140-150 kD complex in a glycerol gradient while βCP migrates as a similarly large

complex, ~ 200 kD, on a sizing column. A gene that most likely codes for one subunit of CP1 has been recently cloned from HeLa cells. The sequence of its cDNA codes for a 27 kD protein, a mass consistent with that of the peptides observed in our silverstain gels (Becker, et al., 1991).

A second protein of interest, also a probable member of the CP1 family, was purified from MEL cells on the basis of its ability to bind to the α-globin CCAAT box (Kim, Barnhart and Sheffery, 1988). This protein, αCP1, shows the same relative affinities for the α-globin, β-globin and NF1 type CCAAT boxes as βCP and more importantly, when purified, gives a very similar pattern of bands on a silverstain gel (Kim, Barnhart and Sheffery, 1988; Kim and Sheffery, 1990). Specifically, αCP1 appears to be composed of a set of peptides whose number, relative mobilities and staining intensities mirrors that of βCP. The mobilities measured for the peptides of αCP1 are uniformly 4-5 kD less than those measured for βCP. We believe this is most likely due to small differences in gel conditions rather than differences between the proteins. For clarity, we will refer to αCP1 protein mobilities in terms of our estimates of kD here.

Purified αCP1 is composed of seven or more proteins (Kim, Barnhart and Sheffery, 1988; Kim and Sheffery, 1990). These include, a triplet of bands that appears to correspond to the triplet we observe at 37-39 kD, three bands that appear to correspond to the bands we observe at 31, 33 and 35 kD, and a doublet of bands that may correspond to a tight spaced doublet of bands we see at 40 and 42 kD in less purified preparations of βCP (Figure 3C). Similar to βCP, when αCP1 was cut from a preparative SDS-polyacrylamide gel,

denatured and renatured, DNA binding was regained only if material eluted from the 25 and 30 kD region of the gel was combined with that from 30 to 43 kD region of the gel(Kim, Barnhart and Sheffery, 1988). Also, a protein that migrates at 33 kD, as does that observed in our scheme 3, has been observed in αCP1(Kim and Sheffery, 1990). Finally, αCP1 stimulates transcription in vitro to levels similar to those observed in this work (Kim, et al., 1990). Given the great many similarities between the two, we conclude that βCP is related or identical to αCP1.

Proteolytic mapping of αCP1 suggested that the 37, 38,40 and 42 kD peptides are related to one another, the 31 and 35 kD peptides form another class of protein and the 39 kD protein is distinct from the others (Kim and Sheffery, 1990). These data, chemical cross linking and various separations in high salt were all used to argue that these three classes of proteins each contribute a monomer to a heterotrimeric CCAAT protein (Kim and Sheffery, 1990). If this model is correct, and assuming that the 33kD protein is related to the 31 and 35 kD proteins, then all of our preparations of βCP contain at least one member of each class of protein in the heterotrimer. Finally, the class of protein that included the 40 and 42 kD proteins was reported to be only loosely associated with the other members of this complex (Kim and Sheffery, 1990). Perhaps our purification schemes 2 and 3, in which we did not observe purification of these proteins, resulted in loss of these two peptides but retention of the related 37 and 38 kD peptides.

In figure 7 of this paper, we demonstrate that βCP, purified in scheme 1, will stimulate transcription in vitro in a CCAAT site

dependent manner. Although it is possible that this partially pure protein preparation contains another CCAAT binding protein that is not detected in the SDS-gel elution denaturation-renaturation assay, our result is consistent with results obtained with purified αCP1. Because purified βCP is not stable, we have not repeated the gel purification-denaturation/renaturation experiment or in vitro transcription using βCP derived from purification schemes 2 and 3. In vitro transcription with these preparations may be particularly enlightening since they will allow observation of the effects of glycosylation on transcription activity.

A final point of interest comes from the observation of βCP's relative affinities for other promoters. Recall that the β-globin CCAAT element is required, along with the CACCC element, for the far upstream regulatory element, the LCR, to have an effect (Antoniou and Grosveld, 1990). The LCR stimulates erythroid specific high level transcription of the β-globin genes as well as of other non-erythroid specific genes including the HSV tk gene (van Assendelft, et al., 1989). Interestingly, the LCR has no effect on transcription of a linked histone H4 gene (van Assendelft, et al., 1989). The histone H4 promoter, unlike the β-globin and tk promoters, failed to compete for βCP binding (Figure). Therefore, the lack of an effect of the LCR on the histone H4 promoter could reflect the absence of a protein, βCP, required to mediate a functional interaction between the LCR and a promoter. Although the tk promoter binds βCP with a lower affinity than the β-globin promoter, this interaction may be sufficient to drive tk promoter/LCR interactions when they are linked in cis. This model could be tested by inserting a βCP binding site into the histone

REFERENCES

Antoniou, M., and Grosveld, F. (1990). β-globin dominant control region interacts differently with distal and proximal promoter elements. Genes and Dev. *4*, 1007-1013.

Ausubel, F.K., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1991). Current Protocols in Molecular Biology. Greene Publishing Associates/Wiley-Interscience.

Barberis, A., Superti-Furga, G., and Busslinger, M. (1987). Mutually exclusive interaction of the CAAT-binding factor and of a displacement protein with overlapping sequences of a histone promoter. Cell *50*, 347-359.

Barnhart, K.M., Kim, C.G., Banerji, S.S., and Sheffery, M. (1988). Identification and characterization of multiple erythroid cell proteins that interact with the promoter of the murine α-globin gene. Molec. Cell. Biol. *8*, 3215-3226.

Becker, D.M., Fikes, J.D., and Guarente, L. (1991). A cDNA encoding a human CCAAT-binding protein cloned by functional complementation in yeast. Proc. Natl. Acad. Sci. USA *88*, 1968-1972.

Charnay, P., Mellon, P., and Maniatis, T. (1985). Linker scanning mutagenisis of the 5'-flanking region of the mouse β-major-globin gene: sequence requirements for transcription in erythroid and non-erythroid cells. Mol. Cell. Biol. *5*(6), 1498-1511.

Chodosh, L.A., Baldwin, A.S., Carthew, R.W., and Sharp, P.A. (1988). Human CCAAT-binding proteins have heterologous subunits. Cell *53*, 11-24.

Chodosh, L.A., Olesen, J., Hahn, S., Baldwin, A.S., Guarente, L., and Sharp, P.A. (1988). A yeast and a human CCAAT-binding protein have heterologous subunits that are functionally interchangeable. Cell *53*, 25-35.

Choi, O.-R.B., and Engel, J.D. (1988). Developmental regulation of β-globin gene switching. Cell *55*, 17-26.

Cohen, R.B., Sheffery, M., and Kim, C.G. (1986). Partial purification of a nuclear protein that binds to the CCAAT box of the mouse α1-globin gene. Molec. Cell. Biol. *6*(821-832),

Cowie, A., and Myers, R.M. (1988). DNA sequences involved in transcriptional regulation of the mouse β-globin promoter in murine erythroleukemia cells. Mol. Cell. Biol. *8*(8), 3122-3128.

Curtin, P.T., Liu, D., Liu, W., Chang, J.C., and Kan, Y.W. (1989). Human β–globin gene expression in transgenic mice is enhanced by a distant DNase I hypersensitive site. . Proc. Natl. Acad. Sci. USA *86*, 7082-7086.

Deisseroth, A., and Hendrick, D. (1978). Human alpha-globin gene expression following chromosomal dependent gene transfer into mouse erythroleukemia cells. Cell *15*, 55-63.

Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J., and Weissman, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit b-globin gene in mouse 3T6 cells. Cell *32*, 695-706.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. *11*, 1475-1489.

Dorn, A., Bollekens, J., Staub, A., Benoist, C., and Mathis, D. (1987). A

multiplicity of CCAAT box-binding proteins. Cell *50*, 863-872.

Dynan, W.S., and Tjian, R. (1983). Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell *32*, 669-680.

Favalaro, J., Treisman, R., and Kamen, R. (1980). Transcription maps of polyoma virus-specific RNA: analysis by two dimensional nuclease S1 gel mapping. Methods Enzymol. *65*, 718-749.

Foley, K.P., and Engel, J.D. (1992). Individual stage selector element mutations lead to reciprocal changes in β- vs. ε-globin gene transcription: genetic confirmation of promoter competition during globin gene switching. Genes and Dev. *6*(730-744),

Forrester, W.C., Epner, E., Driscoll, C., Enver, T., Brice, M., Papayannopoulou, T., and Groudine, M. (1990). A deletion of the human β-globin locus activation region causes a major alteration in chromatin structure and replication across the entire β-globin locus. Genes Dev. *4*, 1637-1649.

Forrester, W.C., Novak, U., Gelinas, R., and Groudine, M. (1989). Molecular analysis of the human β-globin locus activation region. Proc. Natl. Acad. Sci. USA *86*, 5439-5443.

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

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

Grosveld, F., van Assendelft, G.B., Greaves, D.R., and Kollias, G. (1987). Position-independent, high level expression of the human β-globin

gene in transgenic mice. Cell *51*, 975-985.

Hager, D.A., and Burgess, R.R. (1980). Elution of proteins from sodium dodecyl sulfate, and renaturation of enzymatic activity: results with sigma subunit of E. coli RNA polymerase, wheat germ topoisomerase, and other enzymes. Anal. Biochem. *109*, 76-86.

Hatamochi, A., Golmbek, P.T., van Schaftingen, E., and de Crombrugghe, B. (1988). A CCAAT DNA binding factor consisting of two different components that are both required for DNA binding. J. Biol. Chem. *263*, 5940-5947.

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

Jane, S.M., Ney, P.A., Vanin, E.F., Gumucio, D.L., and Nienhuis, A.W. (1992). Identification of a stage selector element in the human γ-globin gene promoter that fosters preferential interaction of the 5' HS2 enhancer when in competition with the β-promoter. EMBO J. *11*, 2961-2969.

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

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

Kim, C.G., Barnhart, K.M., and Sheffery, M. (1988). Purification of multiple erythroid cell proteins that bind the promoter of the α-globin gene. Molec. Cell. Biol. *8*, 4270-4281.

Kim, C.G., and Sheffery, M. (1990). Physical characterization of the purified CCAAT transcription factor, alpha-CP1. J. Biol. Chem. *265*(22), 13362-13369.

Kim, C.G., Swendeman, S.L., Barnhart, K.M., and Sheffery, M. (1990). Promoter elements and erythroid cell nuclear factors that regulate α-globin gene transcription in vitro. Molec. Cell. Biol. *10*, 5958-5966.

Krieg, P.A., and Melton, D.A. (1987). In Vitro RNA synthesis with SP6 RNA polymerase. *155*, 397.

Landschulz, W.F., Johnson, P.F., and McKnight, S.L. (1989). The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science *243*, 1681-1688.

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

Maddon, P.J., Littman, D.R., Godfrey, M., Maddon, D.E., Chess, L., and Axel, R. (1985). The isolation and nucleotide sequence of a cDNA encoding the T cell surface protein T4: a new member of the immunoglobulin superfamily. Cell *42*, 93-104.

Maity, S.N., Golumbek, P.T., Karsenty, G., and de Crombrugghe, B. (1988). Selective activation of transcription by a novel CCAAT binding factor. Science *241*, 582-585.

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

Mellon, P., Parker, V., Gluzman, Y., and Maniatis, T. (1981). Identification of DNA sequences required for transcription of the human α-globin gene in a new SV40 host-vector system. Cell *27*,

279-288.

Mermod, N., O'Neill, E.A., Kelly, T.J., and Tjian, R. (1989). The proline-rich transcriptional activator of CTF/NF1 is distinct from the replication and DNA binding domain. Cell *58*, 741-753.

Myers, R.M., Tilly, K., and Maniatis, T. (1986). Fine structure genetic analysis of a β-globin promoter. Science *232*, 613-618.

Raymondjean, M., cereghini, S., and Yaniv, M. (1988). Several distinct "CCAAT" box binding proteins coexist in eukaryotic cells. Proc. Natl. Acad. Sci. USA *85*, 757-761.

Reinberg, D., and Roeder, R.G. (1987). Factors involved in specific transcription by mammalian RNA polymerase II. J. Biol. Chem. *262*, 3310-3321.

Rosenfeld, P.J., and Kelley, T. (1986). Purification of nuclear factor 1 by DNA recognition site affinity chromatography. J. Biol. Chem. *261*, 1398-1408.

Santoro, C., Mermod, N., Andrews, P.C., and Tjian, R. (1988). A family of human CCAAT-box binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. nature *334*, 218-225.

Stuve, L.L., and Myers, R.M. (1990). A directly repeated sequence in the β-globin promoter regulates transcription in murine erythroleukemia cells. Mol. Cell. Biol. *10*(3), 972-981.

Townes, T., and Behringer, R.R. (1990). Human globin locus activation region (LAR): role in temporal control. Trends Genet. *6*, 219-223.

Tuan, D.Y.H., Solomon, W.B., London, I.M., and Lee, D.P. (1989). An erythroid specific, developmental-stage-independent enhancer far

upstream of the human "β-like globin" genes. Proc. Natl. Acad. Sci. USA *86*, 2554-2558.

van Assendelft, G.B., Hanscombe, O., Grosveld, F., and Greaves, D.R. (1989). The β-globin dominant control region activates homologous and heterologous promoters in a tissue specific manner. Cell *56*, 969-977.

Veldman, G.M., Lupton, S., and Kamen, R. (1985). Polyomavirus enhancer contains multiple redundant sequence elements that activate both DNA replication and gene expression. Molec. Cell. Biol. *5*, 649-658.

Wright, S., Rosenthal, A., Flavell, R., and Grosveld, F. (1984). DNA sequences required for regulated expression of β-globin genes in murine erythroleukemia cells. Cell *38*, 265-273.

CHAPTER 5

SUMMARY AND CONCLUSIONS

In this work, I have described the regulation of transcription as being based on three strategies: first, the arrangement of cis elements in a promoter specifies the set of all potential regulatory decisions that can be made at a promoter. Second, the complement of proteins in a particular cell and their state of activity determines which regulatory decisions can or will be implemented at that promoter. As described in the introduction, the above two mechanisms are employed both by prokaryotes and eukaryotes. Eukaryotes, because of their diversity of different cell types versus one cell type for prokaryotes, face a greater task of regulation. Many eukaryotic genes are expressed in only a subset of cell types or only during a subset of developmental stages. It appears that eukaryotes use a third strategy to deal with this additional regulatory burden: chromatin mediated repression of transcription. By specifically and dominantly inactivating a gene via repressing chromatin structures, the cis elements that specify that gene's regulatory characteristics can be masked. This strategy has the advantage that an elaborate system of site specific DNA binding proteins that repress specific genes is not required and that negative regulation of genes combined with positive regulatory mechanisms can increase the dynamic range of gene expression.

In the globin system, the cis elements of the adult β-globin genes are well understood. In addition to experimental evidence supporting the role of the CACCC, CCAAT, βDRE and TATA elements, naturally occurring mutations in the CACCC and TATA elements of

the human β-globin gene affect transcription in vivo. Although a number of other elements near the β-globin gene have been shown to affect β-globin transcription in transfection assays and in transgenic mice, these elements are not required for regulation of β-globin transcription when assayed in the presence of the LCR. Therefore, the only sequences near the β-globin gene that are required for transcriptional control fall in the +1 to -100 region of the promoter. Although the LCR has not been dissected in as great detail as the β-globin promoters, it is becoming clear that it is composed of reiterated CACCC, GATA-1 and NF-E2 elements. The repeated occurrence of GATA-1 sites near CACCC elements suggests that the relative spacing of these elements is also important. Finally, evidence is mounting that the interaction of the LCR with the promoter of the β-like globin genes is regulated as well. First, genes expressed at earlier stages of development can outcompete later developmental variants of β-globin and second, the fetal and embryonic genes appear to be specifically silenced to prevent their expression in later developmental stages. For the most part, the cis elements and trans factors involved in these interactions remain to be discovered.

In contrast to the wealth of information available for the cis regulatory elements of the β-globin genes, the identities of the proteins that perform transcriptional regulation of the β-globin genes are largely unknown. With the exception of the GATA-1 gene, none of the genes encoding upstream regulators of β-globin have been cloned. My thesis work has been directed at determining what proteins bind two of these elements: the CCAAT and CACCC. During

the course of these studies, I confronted two issues that must be resolved before many transcription regulatory units can be understood.

First, I found that the CACCC element is bound by multiple proteins in vitro. To discriminate between these, I developed a methodology of combining a site selection assay with genetic and biochemical correlations of a protein's affinity for a mutant site and the activity of that mutant site in transcription. Because many DNA-binding proteins have overlapping sequence specificities, this sort of discrimination is an important one. Furthermore, because it is possible that transcriptional activator proteins function in more than one way, merely demonstrating that one protein has the potential to affect transcriptional regulation of a promoter is not enough evidence to conclude that protein actually does so in normal in vivo situations. For example, in the case of the CACCC element, Sp1 may be sufficient to bring a transcriptional activation domain to the β-globin promoter in vivo, but the CACD protein may specifically mediate promoter-LCR interactions. Therefore, while my efforts have suggested that CACD is most likely to activate β-globin in vivo, more experiments are required to prove this.

In the case of the β-globin CCAAT element, it appears that βCP activity is much more abundant than any other CCAAT box binding activity in MEL cells. Therefore, βCP is most likely the protein responsible for activity from the CCAAT element in vivo. Study of βCP raised a second issue of general importance to the understanding of transcriptional regulation. Because multiple peptides copurify with βCP activity, and because we believe β CP to be a multipartite

activity, it is possible that different isoforms of βCP with different activities exist in vivo. Investigation of this issue requires a higher resolution knowledge of βCP's composition antibodies specific to the particular peptides that copurify with βCP would be helpful in studying this issue as would the cloning of cDNAs specific to these subunits. If indeed, multiple isoforms of βCP do exist, then pure preparations of particular isoforms would facilitate in vitro studies of their function.

In addition to βCP, CACD and GATA-1, at least two other proteins appear to be involved in regulating β–globin gene expression. These are NF-E2 and the protein(s) that binds the βDRE. Both of these have only been described as site specific DNA-binding activities in the gel mobility shift assay. Additional proteins thought to be involved in regulating other genes of the β–globin locus include Sp1 and an octamer binding protein. Clearly, more biochemical characterization of these various proteins is required before a meaningful analysis of protein-protein interactions in the β–globin locus can be carried out. Another unresolved issue in β-globin regulation is that of chromatin structure. Although the LCR appears to regulate the chromatin structure of the β-locus, the mechanism by which this occurs is totally unknown. Of particular interest is the fact that the two cell type restricted factors, GATA-1 and NF-E2, both appear to be expressed in megakaryocytes yet β–globin is not expressed. Because CACD, Sp1 and βCP appear to be widely expressed, this suggests that other factors regulating β–globin gene expression may yet be found.