# The *In Vivo* Examination of Transcriptional Control Mechanisms in Mammalian Cells.

Thesis by
Paul Allen Garrity

In Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy

California Institute of Technology
Pasadena, California

1993

(Submitted December 16, 1992)

© 1993
Paul Allen Garrity
All rights reserved

This work is dedicated to my parents, John and Helen Garrity,
and to Tom, Jane Ann, and Gary (my honorary brother).

**Acknowledgements**

Oh, I just don't know where to begin. First, I want to thank Barbara Wold, for her scientific insight, judgement and enthusiasm, as well as the amusing stories and understanding she has provided over the years I have been in her laboratory. Working for Barbara has been challenging and a lot of fun. I also especially thank Paul Mueller, who has been a friend and scientific elder brother throughout. He has provided much appreciated sage counsel, withering criticism, and lots and lots of ideas. I thank my thesis committee of David Anderson, Norman Davidson, Ellen Rothenberg, Paul Sternberg, Kai Zinn, for their help, enthusiasm, and encouragement. I also thank my collaborators on the IL-2 project, Dan Chen and Ellen Rothenberg, for their essential input and assistance. The people in the Wold lab have been great and I'd like to give special thanks to several. I want to thank Sean Tavtigian for tons of help over the years and for performing the ultimate good deed- offering his computer up to the wilds of my room so I could write this thing. Jeff Miner was also great, with his patience, scientific gifts, and vicious sense of humor. I'd also like to thank Roger and Ruffman for all that beer, Dr. Sam for his stories, Ardem and Sonya for sundry stuff, and Kyuson and Dee for putting up with all the noise.

Chi-Bin Chien, Jim Knierim, and all the people at Prufrock over the years have made life lots more fun. I thank John Bowman for all the camping trips and the late-night runs through San Marino, Lisa Henderson for being a good friend, Eric Arn for all the radio shows, Tom Novak for the BB62 (though I had to give it away), and the Sunday Frisbee bunch. Finally, and most of all, I thank Linda for all her help and understanding (and for reading the unreadable read).

## ABSTRACT

In the investigations described in this thesis I have examined various mechanisms involved in transcriptional control. The first chapter is a general examination of the mechanisms used in transcriptional control. The second chapter addresses issues associated with the selection of transcriptional start site. This work shows that core promoter usage can be altered in a tissue-specific and environmentally-responsive manner. The third chapter of this work describes a significant improvement in the technique of ligation-mediated PCR-aided *in vivo* footprinting and genomic sequencing. This improvement in the quality of *in vivo* footprint data allows the pattern of protein:DNA interactions to be obtained with greater signal to noise and for a larger group of DNA sequences. The fourth chapter of this thesis uses these *in vivo* footprinting techniques to investigate the mechanisms controlling the transcription of the mouse interleukin-2 gene upon T cell stimulation. T cell stimulation was shown to result in the coordinated occupancy of a number of major groove binding proteins to the previously unoccupied IL-2 regulatory region *in vivo*. Finally, the appendix describes *in vivo* footprints at the differentiated-muscle-specific promoter of the delta-subunit of the nicotinic acetylcholine receptor. Multiple protein:DNA interactions were seen at the promoter both before and after muscle cell differentiation, suggesting that transcriptional regulation of this gene occurs at the level of protein replacement or an alteration of the ability of the assembled proteins to activate transcription.

# TABLE OF CONTENTS

# INTRODUCTION

Cellular transcription results in the copying of one strand of the DNA duplex into RNA. The process of transcriptional control determines which portions of the genomic DNA a cell expresses in the form of RNA and the rate at which each RNA species is made. Transcriptional control plays a very important role in the development and function of an organism. Its importance can be seen in the drastic, and often lethal, developmental perturbations caused by the deletion or misexpression of individual transcriptional control factors (Davidson 1991). An enormous amount of information has been obtained on the subject of transcriptional control. In this relatively short review I will try to summarize many of the key concepts and present specific examples to illustrate particular points. My major focus will be on the mechanisms of transcriptional control in mammals. However, many basic structural and functional similarities among transcriptional control mechanisms have been demonstrated for a number of different eukaryotes (Guarente and Bermingham-McDonogh 1992). Thus, I have incorporated specific examples from eukaryotes other than mammals, such as the fruit fly *Drosophila melanogaster* and the budding yeast *Saccharomyces cerevisiae*, in cases where these can serve to illustrate a certain principle particularly clearly.

Transcriptional control of protein-coding genes in eukaryotes is mediated through DNA sequences that flank, and occasionally include, the coding region. These gene-specific cis-acting DNA sequence elements interact with sequence-specific DNA-binding proteins, which in turn, directly or indirectly, control the assembly and activity of the general RNA polymerase II machinery of the cell at the core promoter near the transcriptional start site. Each step along this pathway can be a locus for

transcriptional control. The occupancy of a regulatory region by regulatory proteins involves multiple interactions among sequence-specific DNA-binding proteins, cis-acting DNA sequence elements and the chromatin into which the DNA sequence elements are packaged. Each of these components, such as the set of DNA-binding proteins available or the status of the chromatin are targets for control by the cell. The subsequent transcriptional control effects of these gene-specific, DNA-bound regulatory proteins on the general RNA polymerase II machinery reflects both the intrinsic qualities of the bound proteins, the availability of both gene-specific and general non-DNA-binding regulatory proteins with which the DNA-bound proteins can interact, and the modifications that have been made to all the proteins involved. Each of the components involved in the activation and repression of the general RNA polymerase II machinery are also targets for control by the cell. Finally, even after transcription has been initiated, the elongation properties of RNA polymerase II can also be a target for control. Thus transcription involves the multistep assembly of a large, multiprotein complex (whose molecular weight greatly exceeds one million kD) on a DNA template. The complexity and diversity of the components of this molecular machine provide multiple opportunities for fine-tuned control of transcription.

This discussion of transcriptional control will begin with a description of the structure and function of key components involved in the conversion of DNA-binding by gene-specific regulatory factors to control of RNA polymerase II activity. Transcriptional control domains of gene-specific regulators, the basic RNA polymerase II machinery of the cell, and the proteins that may serve as intermediates between the two will be described,

as well as how the gene-specific regulators control the general machinery. The second section of the review discusses how the transcriptional regulatory activity of the gene-specific regulators is controlled by the cell. In the third section, the mechanisms controlling the occupancy and activity of an entire regulatory region will be discussed, focusing on how the regulatory region integrates multiple regulatory inputs into a single transcriptional output. The final section discusses major experimental approaches to transcription.

## I) TRANSCRIPTIONAL ACTIVATION VIA BOUND REGULATORY FACTORS:

### Gene-specific regulatory factors:

Sequence-specific DNA-binding proteins exert their effects on transcription through both DNA:protein and protein:protein interactions. Often the DNA:protein interaction is necessary but not sufficient for a factor to regulate transcription. In these cases the DNA:protein interaction is required mainly to bring to the DNA regulatory region of interest additional protein domains that control transcription through protein:protein interactions. For example, SP1, a DNA-binding regulator that stimulates transcription, contains a set of zinc-fingers which interact with DNA. However, DNA-binding of this domain alone exerts no control of transcription. Separate non-DNA-binding regions of the protein, including two glutamine-rich domains, are required for SP1 to stimulate transcription (Courey and Tjian 1988). Such separate transcriptional regulatory domains have been found for most DNA-binding transcriptional regulators. Several different classes of transcriptional control domains which stimulate transcription have been identified, including the glutamine-rich domains of

SP1 and Oct 2, the proline-rich domains of CTF and AP2, and the acidic domains of GAL4 (Gill and Tjian 1992). This suggests that not only do most DNA-binding transcriptional regulatory proteins regulate transcription through protein:protein interactions, there may be several distinct classes of proteins that serve as targets of these domains and that each class of regulatory domain may talk to a specific type of target.

Gene-specific transcriptional regulatory proteins need not have the ability to bind DNA on their own. They can associate with a DNA regulatory region through protein:protein interactions with factors that do possess DNA-binding ability. The transcriptional activator VP16, for example, activates transcription of the herpes virus infected cell protein 4 promoter (ICP4) by associating with DNA-bound Oct-1 (recent evidence indicates VP16 can weakly recognize some DNA sites, though it need not do so to activate transcription; Thompson and McKnight 1992). It is important to note that these non-DNA-binding regulatory proteins continue to function when fused to DNA-binding proteins and their transcriptional regulatory domains fall into the same classes as those of the DNA-binding proteins. Thus they do not appear to regulate transcription through a distinct set of mechanisms.

The proteins discussed so far have all been transcriptional activators. Regulators which function as transcriptional repressors are also common and they can be either DNA-binding or non-DNA-binding proteins (Johnson and Krasnow 1992; Keleher et al. 1992). Repressors can function either through competition for an activator's DNA-binding site, competition for the general RNA polymerase machinery's DNA-binding site or through repressive protein:protein interactions. Thus, both transcriptional activators and

repressors exert their effects through both protein:DNA and protein:protein interactions.

**Assembly of general transcriptional machinery at the core promoter**

Core promoters usually contain the TATA and/or initiator (Inr) sequence elements and interact with components of the basic transcriptional machinery of the cell. These sequence elements alone or together are sufficient to drive a basal level of accurately initiated transcription *in vitro* (Roeder 1991). However, a core promoter is insufficient to drive appreciable levels of transcription in the living cell, relying on gene-specific activators to function *in vivo*. Nonetheless, its serves the same basic function *in vitro* as *in vivo*, bringing RNA polymerase II to the promoter, and hence its function has been productively studied *in vitro*.

The general transcription machinery assembles on the core promoter through a precise pathway of interactions (Buratowski et al. 1989; Roeder 1991). The first step is binding of the TATA-binding transcription factor TFIID with the TATA element. TFIIA assists in the assembly of this complex, binding cooperatively with TFIID, though it has no intrinsic sequence-specific-DNA-binding ability on its own. TFIIB binds this complex, assembling across the initiation site though it too has no intrinsic sequence-specific-DNA-binding ability. This TFIIA+TFIID+TFIIB complex recognizes and binds to RNA polymerase associated with TFIIF (which suppresses RNA polymerase's intrinsic affinity for non-specific DNA, which might otherwise cause it to bind non-promoter DNA; Greenblatt 1991). Then TFIIE binds, completing formation of the transcriptional pre-initiation complex (preIC). Addition of nucleotide triphosphates to the preIC results in transcriptional initiation within seconds (Kadonaga 1990).

Genes that do not contain TATA elements, but that do contain initiator (Inr) elements utilize the Inr binding factor TFII-I, which recognizes the degenerate consensus sequence YAYTCYYY (Roy et al. 1991). TFII-I interacts with TFIID to bring it to Inr-containing core promoters, and on promoters containing both TATA and Inr elements, it can compensate for the removal of TFIIA activity from an *in vitro* transcription reaction (Roy et al. 1991). Whether the TFII-I and TFIIA normally collaborate on promoters containing both has not yet been resolved. Subsequent steps in the TFII-I mediated pathway appear to utilize the same factors as the TFIIA-mediated pathway (Roy et al. 1991; Zenzie-Gregory et al. 1992). Some promoters contain neither obvious Inr nor TATA elements. The precise mechanism of transcription complex assembly at these promoters is unclear, but those examined can function using the same crude transcription extracts as Inr-containing and TATA-containing promoters (O'Shea-Greenfield and Smale 1992; Zenzie-Gregory et al. 1992). Thus there is diversity in the structure of core promoters, but most appear to utilize similar sets of general transcription factors. The recent discovery that the core promoters of RNA polymerase I and RNA polymerase III genes can also utilize components of TFIID (see below) in transcriptional activation indicates that all transcription may utilize at least some of the same general factors (White and Jackson 1992).

**Role of core promoter in determining transcriptional start site and direction**

Core promoters also control the selection of start site and direction of transcription, as most elements that interact with gene-specific activators and repressors can function in either orientation upstream or downstream of the initiation site. Both the TATA and Inr elements can specify the transcriptional initiation site and the direction of transcription *in vitro.* In

mammals, the TATA element has been shown to specify initiation at a discrete site approximately 25 bp downstream on one strand, while the initiator specifies unidirectional initiation within its site . When the TATA and Inr elements are separated by more than 30 bp, transcription initiates approximately 25 bp downstream of the TATA element (O'Shea-Greenfield and Smale 1992). This indicates that the TATA element plays a more powerful role in start site selection than the Inr. Similarly, the orientation of the TATA element determined the direction of transcription when the TATA and Inr were of opposite orientation (O'Shea-Greenfield and Smale 1992). However, examination of the Adenovirus IVa2 (Ad IVa2) promoter shows these generalizations are not absolute (Carcamo et al. 1990). The Ad IVa2 promoter contains a functional TATA element in a reversed orientation between 14 and 21 base pairs downstream of the transcriptional start site. Transcription initiates within the Ad IVa2 promoter's Inr element, and the Inr appears to control both the start site and the directionality of transcription. Though no further evidence is available for this promoter, this exception may be caused by the relative positions of Inr, TATA and the relatively uncharacterized NTF-factor binding site located just upstream of the Inr. In a test promoter, several SP1 sites were located 25 bp downstream of a TATA element. This caused the TATA element to facilitate transcriptional initiation approximately 25 bp away from the TATA element, but heading away from the SP1 sites (O'Shea-Greenfield and Smale 1992). Thus the presence of multiple SP1 sites just downstream of the TATA caused a reversal of transcriptional direction and transcriptional start site position. Although this shows that there can be complex interactions between core elements and gene-specific activators in some cases, in general the core elements play the

major role in determining the choice of start site and directionality of transcription.

As noted above, some promoters contain no obvious TATA or Inr elements. These promoters fall into two groups: those that initiate transcription at a discrete start site and those that initiate at multiple start sites over a region that can extend to over a hundred base pairs, often bidirectionally. Many of the former contain elements which show little homology to Inr sequences or each other, but function like an Inr *in vitro* (O'Shea-Greenfield and Smale 1992). One such non-Inr initiator is present in the mouse dihydrofolate reductase gene (Means and Farnham 1990). This sequence element interacts with a sequence-specific DNA-binding protein named HIP1 (Means and Farnham 1990). It is possible that HIP1 and TFII-I may be related activities. Although the Inr sequence recognized by TFII-I is unrelated to the HIP1 recognition sequence, such an interaction is not unprecedented, as TFII-I can bind with high affinity to other Inr-unrelated sequences (Roy et al. 1991). Thus the biochemical properties of the non-Inr initiator core promoters are not well understood, but are potentially similar to the Inr mechanism. Finally, even less well understood are the TATA-less, Inr-less promoters which direct transcription from heterogeneous start sites (Zenzie-Gregory et al. 1992). Such a promoter supports heterogeneously initiated, bi-directional RNA polymerase II transcription *in vitro*, but the mechanism is unknown (Zenzie-Gregory et al. 1992).

**Coactivators facilitate protein:protein interactions between gene-specific regulatory factors and the general machinery at the core promoter.**

To work out the molecular mechanisms by which gene-specific transcriptional activators and repressors function, the effect of these factors

on the general transcriptional machinery have been tested using *in vitro* transcription systems. An important clue to the way in which gene-specific activators work came with the isolation from several species of the gene encoding the TATA-binding component of TFIID, or the TATA-binding protein (TBP) (Ptashne and Gann 1990). When added to an *in vitro* transcription reaction depleted of TFIID, bacterially produced TBP supported the same level of basal transcription as native TFIID purified chromatographically from mammalian or *Drosophila* cells. But when gene-specific transcriptional activator proteins were added to the *in vitro* transcription mixture, reactions containing native TFIID showed increased levels of *in vitro* transcription, whereas those containing recombinant TBP did not. Subsequent experiments showed that native TFIID is a multiprotein complex containing, in addition to TBP, a number of other proteins called TBP-associated factors (TAF's) (Dynlacht et al. 1991) (Tanese et al. 1991). Separating TFIID's components with urea yielded a fraction with TBP activity that could support basal transcription and a fraction that alone supported no transcription but when combined with the other fraction reconstituted the ability of activators to stimulate transcription (Dynlacht et al. 1991). At least some of the TAF's are believed to mediate the effect of gene-specific activators on the general transcriptional apparatus of the cell and thus to serve as "coactivators." In addition to the TAF's, there may be other proteins that can function to mediate the effects of gene-specific activators, but which are not physically associated with components of the basal transcriptional apparatus prior to its assembly on a core promoter (Meisterernst et al. 1991). Although activators apparently require coactivators to function, some activators, such as VP16, can physically associate with the basal transcriptional machinery in

the absence of coactivators. For example, VP16 binds to TBP and to TFIIB *in vitro* (Stringer et al. 1990; Lin and Green 1991). These direct interactions could be functionally significant, as mutations that disrupted VP16's interactions with general factors decreased its activating potential (Lin et al. 1991; Ingles et al. 1991). Thus gene-specific activators (and possibly repressors) can alter the activity of the general transcription apparatus required for basal transcription (TBP, TFIIA, TFIIB, etc..) by working in concert with or through coactivators which are not required for basal transcription.

**Gene-specific activators can work by antagonizing the repressive effects of chromatin on core promoter function.**

The assembly of a transcriptional initiation complex is a complex, multistep process, and each step along this pathway constitutes a potential target of gene-specific activators and repressors. As noted above, a core promoter is sufficient to promote a basal level of transcriptional initiation *in vitro*, but additional elements that interact with gene-specific factors are required for transcription *in vivo*. An important experimental detail is that the *in vitro* transcription reactions used to study basal level transcription above used DNA templates that were free of protein, also known as naked DNA, as starting material. However, DNA in the cell is associated with chromosomal proteins, such as histones, which package the cellular DNA and can exert a powerful repressive effect on transcription *in vivo* (Felsenfeld 1992). A major component of chromatin is the histone octamer which is formed by two of each of the four core histones (H2A, H2B, H3, H4). The histone octamer interacts with approximately 146 bp of DNA, which is wound around the octamer core, creating a nucleosome. Histone H1 recognizes nucleosomes and can bind the DNA between adjacent

nucleosomes, serving to package the nucleosomal "beads on a string" more compactly (Hayes and Wolffe 1992). Thus, chromatin structure is known to have a powerful effect on transcription *in vivo* and enough components of chromatin are known to attempt to study its effect on transcription *in vitro*.

Packaging DNA templates into nucleosomes (using the purified core histones H2A, H2B, H3, and H4) prior to use in an *in vitro* transcription assay mildly repressed basal level transcription two to four fold (Laybourn and Kadonaga 1991). The addition of histone H1 to the nucleosomes resulted in a further ten to twenty-fold repression of basal level transcription (Laybourn and Kadonaga 1991). Repression by the core histones was caused by the formation of a nucleosome on the core promoter, and the H1-mediated increase in repression was postulated to involve the binding of H1 to the fraction of core promoters that were located between nucleosomes (Laybourn and Kadonaga 1991). That chromatin can inhibit transcription by blocking the access of general factors to the core promoter is consistent with the previous finding that binding of TFIID to the core promoter prior to packaging the template into chromatin prevented repression (Workman and Roeder 1987).

When chromatin assembly using H1 and the core histones took place in the presence of a sequence-specific activator such as GAL4-VP16, the level of *in vitro* transcription subsequently observed was 200-fold higher than in the absence of GAL4-VP16 (Laybourn and Kadonaga 1991). This 200-fold activation of transcription is much closer to the large *in vivo* effects seen with GAL4-VP16 than the eight-fold activation seen in the absence of histones (Laybourn and Kadonaga 1991). Thus GAL4-VP16 appears to activate transcription both through counteracting the repressive effects of histone H1-

containing chromatin (an activity termed antirepression) and by stimulating the activity of the general transcriptional machinery using naked DNA templates. Similar experiments using the GAL4-VP16 protein showed that both the DNA-binding and transcriptional activation domains were required for antirepression (Workman et al. 1991). This argues that in addition to DNA-binding, protein:protein interactions between gene-specific activators and the general transcription apparatus and/or between activators and chromatin components play important roles in antirepression. Thus a key component of activator function is to counteract the repressive effects of chromatin on the basic transcriptional apparatus by antirepression. Repressors could work by promoting or stabilizing a repressive chromatin structure.

The activation of a silent gene often involves the remodeling of chromatin structure (Felsenfeld 1992). However, in the above experiments gene-specific factors such as GAL4-VP16 (and SP1) could not reverse the repressive effects of chromatin if they were added after chromatin assembly. This could reflect a limitation particular to the gene-specific factors examined; perhaps GAL4-VP16 and SP1 serve only to keep already activated genes open. Or, it could suggest that these factors do not act alone in counteracting repressed chromatin structure. For example, the assay used in the above experiments might be missing factors which assist gene-specific regulatory proteins in accessing nucleosomal DNA *in vivo*, such as the SNF/SWI products discussed below, which have been implicated in the regulation of chromatin structure (Winston and Carlson 1992).

**Gene-specific activators and repressors can affect the rate and extent of active pre-initiation complex assembly**

Gene-specific activators can work not only by altering the effects of chromatin on the general transcription apparatus (antirepression), but also by changing the nature of particular steps in transcriptional pre-initiation complex (preIC) assembly. Transcriptional regulators could change the rate of preIC formation or alter the equilibrium constant of formation such that a different percentage of accessible templates form active preIC's. A regulator could also alter the rate at which a preIC initiates transcription or the rate at which the next preIC can form after a round of transcriptional initiation. Though their precise targets in the general transcriptional apparatus are still unclear, measurements of the percentage of templates on which preIC complexes form over time indicate that different regulators can have different effects on preIC formation. The transcriptional activator TGA1a increases both the rate and extent of preIC formation *in vitro* (Katagiri et al. 1990), whereas GAL4-VP16 increases the number of preIC's formed without changing the rate at which they form (White et al. 1992). Thus, gene-specific activators can affect either the rate and/or extent of preIC formation *in vitro* .

An important recent insight is that transcriptional initiation can be controlled not only by the number of complexes formed at the core promoter, but by the balance between formation of active (preIC's) versus inactive complexes. The Roeder laboratory has found that more than one type of TFIID containing complex can form on the core promoter. Several activities, including NC1 and NC2, were identified which can compete with TFIIA for assembly with TFIID and result in the formation of TFIID-NC1 or TFIID-NC2 complexes on the core promoter (Meisterernst and Roeder 1991; Meisterernst et al. 1991). The TFIID-NC2 complex does not associate with TFIIB and the TFIID-NC2 complex as well as the TFIID-NC1 complex are transcriptionally

inert. A potential function of gene-specific regulators could be determining the balance between these different complexes.

*In vitro* studies of the gene-specific activator UBX and the gene-specific repressor EVE are consistent with these regulators exerting opposite effects by altering the balance of active versus inactive complex formation (Johnson and Krasnow 1992). These *Drosophila* regulatory proteins do not alter the rate of preIC formation. Instead, they alter the number of preIC's formed. Furthermore, once preIC assembly is complete, neither protein exerts a regulatory effect, suggesting that the production of active versus inactive complexes can be influenced during their formation, but not after they have formed. After addition of nucleotide triphosphates the regulatory proteins can again exert an effect. Thus, a first step in transcription complex assembly can be seen to involve an initial partitioning of core promoters into active versus inactive complexes, in a process subject to the influences of gene-specific regulators. After this initial complex formation, in a step not influenced by either of the regulators examined here, transcription is initiated from the active complex or the inactive complex disassembles in an ATP-dependent step. Then the process of competition between active and inactive complexes starts again (Roeder 1991; Johnson and Krasnow 1992). It will now be important to directly observe the constituents of the active and inactive complexes formed to see if they include TFIIA and NC1/NC2 respectively.

Though the EVE repressor has been shown to alter preIC assembly *in vitro*, gene-specific repressors could also act via other, less direct mechanisms. A repressor could compete with an activator for a DNA-binding site (Keller and Maniatis 1991; Hoch et al. 1992), or mask an activator's transcriptional activation domain (Leuther and Johnston 1992), hence repressing

transcription by depriving the core promoter of an essential activator. Alternatively, a repressor could act by binding to the core promoter and preventing the binding of TFIID (Stromstedt et al. 1992). In a similar vein, an apparent gene-specific activator could also function by antagonizing the function of any of these types of repressors.

**Gene-specific regulators can also alter steps after transcriptional initiation**

After the preIC forms, the RNA polymerase can begin synthesizing RNA. However, this elongation step is also a target for regulatory proteins. In the case of the human immunodeficiency virus type-1 (HIV-1) promoter, most of the elongating RNA polymerases pause just downstream of the promoter and transcription terminates 60 to 80 nucleotides after initiation. In the presence of the viral regulator Tat, which interacts with the first 79 nucleotides of the nascent transcript, transcriptional elongation proceeds (Karn and Graeble 1992). Thus, Tat activates production of HIV-1 RNA by acting after transcriptional initiation. The observation that polymerases can stall just downstream of initiation is not restricted to viral promoters, as many Drosophila promoters contain RNA polymerases paused within 25 bases of the start site, as determined by a combination of *in vivo* crosslinking, nuclear transcriptional run-on, and *in vivo* detection of melted DNA experiments (Rougvie and Lis 1990; Giardina et al. 1992). The regulatory role of RNA polymerase pausing in these promoters is unclear, as heat shock promoters show equivalent numbers of paused polymerases before and after induction of heat shock RNA synthesis (Giardina et al. 1992). Furthermore, it does not appear that RNA polymerase pausing exerts a simple negative effect on transcription, as mutation of the pause sites in the HIV-1 (Karn and Graeble 1992) and hsp70 (Lee et al. 1992) promoters decreases the production of

transcripts from these genes upon induction. In an attempt to reconcile the apparent negative and positive effects of polymerase pausing, Giardini et al., (1992) have postulated that the presence of the paused polymerase complex may assist in the initiation of the next transcript. In this model, the pause site would decrease the basal level of expression by preventing efficient elongation, perhaps through interaction with a repressor of elongation. Positively acting factors would then increase the rate at which the polymerase leaves the pause site without eliminating the pause. If the paused polymerase complex assisted in assembly of the next transcription complex, then the positively acting factors would eliminate the repression of elongation, but leave the activating capacity of the paused polymerase complex intact. Thus paused polymerases could exert both a positive effect on transcriptional initiation and a negative effect on elongation, with the net effect determined by the balance between the two. There is as yet no mechanistic evidence for this model. Nonetheless, the existence of polymerase pausing and the ability of regulators to work at the level of elongation have been established.

**The effect of gene-specific regulators on the choice of RNA start site**

As noted above, in certain cases gene-specific regulators can affect the choice of start site and direction of transcription. The role of gene-specific regulators in the choice of start site *in vivo* was established for the HIS3 promoter in yeast, in which constitutive and inducible regulatory elements directed transcription through alternate TATA boxes (Struhl, 1986). As discussed previously, core promoters that function through TATA and/or initiator elements initiate transcription at relatively discrete start sites, whereas many promoters lacking TATA and initiator elements initiate

transcription at dispersed sites. As described in Garrity and Wold (1990), we found that the choice of core promoter in the mouse metallothionein-I (MT-I) gene was not restricted to decisions between alternate TATA boxes, but could involve using TATA-dependent versus TATA-independent mechanisms. The choice of core promoter showed tissue-specificity, as TATA-independent transcripts predominated only in the meiotic germ cells of the testis. This could be due to tissue-specific regulatory factors. However, a plausible alternative explanation would be that this effect reflects testis-specific changes in the basic transcriptional machinery, a hypothesis supported by the subsequent observation of TATA-independent start site usage in testis by the herpes simplex virus type-1 thymidine kinase promoter (Al-Shawi et al. 1991). In addition to showing different tissue-preference, the two modes of MT-I transcription also differed in their environmental responsiveness (Garrity and Wold 1990). The production of the TATA-dependent transcripts could be stimulated through the MT-I metal responsive enhancer elements, while the TATA-independent transcripts could not be. In addition, site-directed mutation of the TATA-element created a largely metal-insensitive MT-I regulatory region. This differential ability of the TATA-dependent and TATA-independent start sites to be activated through the metal responsive enhancer elements could be due either to the relative positions of the different start sites and the metal responsive enhancer elements or to an inability of the metal responsive enhancer elements to stimulate TATA-independent transcription. The recent isolation of many of the key factors involved in the function of the TATA and initiator elements provide some of the tools for approaching the mechanism behind these observations.

**II) CELLULAR CONTROL OF GENE-SPECIFIC REGULATORS**

In the first section, I discussed the mechanisms by which individual gene-specific regulators exert their effects on the basic transcriptional machinery of the cell. Having established that transcriptional regulation works through these gene-specific regulators, it is important to understand how the cell modulates the activity of the transcriptional regulators as these regulators are its means of creating a developmentally and environmentally responsive pattern of gene expression. As the final pattern of gene expression must take into account a diverse combination of developmental and environmental information, control points where diverse information is integrated are especially important to understand.

Several distinct properties of transcriptional regulators are subject to control. Is a regulator to act as an activator or a repressor of transcription, or neither? The same protein bound to a DNA regulatory region could potentially be used to exert multiple effects. Can the regulator bind to DNA, and if so what is its sequence-specificity? By exerting control over these aspects of a regulator's behavior, the occupancy of particular regulatory sequence elements can be influenced by signals that the cell receives. Finally, for non-DNA binding regulators, with what DNA-binding factors can they associate? Control at this level can be used to control the recruitment of these non-DNA regulators to promoter regions. Though these properties can be controlled in a number of ways, two general strategies can be defined. The first is the use of covalent or non-covalent modification of individual factors. The second is the use of interactions between individual factors. In many cases, both strategies are used simultaneously. For example, covalent modification of individual factors is often used to control interactions between factors. However, for didactic purposes I will discuss the two

separately (focusing on phosphorylation as an exemplary modification of an individual factor), and present particular examples that point out important regulatory implications.

**Control by phosphorylation**

Phosphorylation, like other covalent and non-covalent (e.g., ligand-binding) modifications, can change a protein's function by altering its conformation (perhaps exposing an important domain) or by specifically changing the properties of an active surface (such as a nucleotide binding site or a DNA-binding motif). Of the many examples of phosphorylation control of regulatory factors, two of particular interest concern factors potentially involved in the regulation of the interleukin-2 (IL-2) promoter, which is the subject of Chapter 3 of this thesis. The first factor is NF-κB, whose role in IL-2 transcription is complicated by the presence in activated T cells of a number of apparently distinct factors with related DNA-binding activities (Briegel et al. 1991; Lee et al. 1991; Urban et al. 1991). The second factor is NF-AT, whose combined pattern of T-cell restricted expression, DNA-binding-selectivity, and immunosuppressant drug sensitivity (see below) appear unique and correlate well with the pattern of IL-2 transcription, strongly suggesting that it is a regulator of IL-2 transcription.

The major species of the DNA-binding transcription factor NF-κB is a hetero-oligomer of two members of the c-rel gene family (Fujita et al. 1992; Nolan and Baltimore 1992). Though it is reliant on phosphorylation for activity in many cell types, it is probably only an indirect target of the major phosphorylation event. In these cells, NF-κB is present in the cytoplasm in a complex with members of the I-κB protein family (Baeuerle and Baltimore 1988). This cytoplasmic complex lacks DNA-binding ability *in vitro*, as well

as being in the wrong compartment of the cell to activate transcription of nuclear genes *in vivo*. Treatment of the HeLa cell line or the pre-B cell line 70Z/3 with stimuli such as the phorbol ester TPA, which can activate protein kinase C (PKC), causes the NF-κB/I-κB complex to dissociate. The freed NF-κB relocalizes to the nucleus and can now recognize DNA (Baeuerle and Baltimore 1988). This chain of events has been investigated *in vitro*, where PKC was shown to phosphorylate the I-κB component of the NF-κB/I-κB complex, resulting in the dissociation of the complex and the activation of NF-κB DNA-binding activity (Ghosh and Baltimore 1990). Thus phosphorylation of I-κB by PKC can activate NF-κB *in vitro*, but it has not yet been shown that stimuli such as TPA activate NF-κB through phosphorylation of I-κB *in vivo*. Although the ankyrin repeats in the I-κB family members are involved in this interaction, the molecular details of the interaction and how it is regulated by phosphorylation are unknown (Nolan and Baltimore 1992). Finally, the mechanism by which I-κB sequesters NF-κB in the cytoplasm is also unknown. I-κB could act as a cytoplasmic anchor or it could mask the nuclear transport signal on NF-κB (Hunter and Karin 1992). In summary, *in vitro* studies using purified components have shown a mechanism by which NF-κB can be activated. It appears likely that phosphorylation will also play a role in its activation in at least some situations *in vivo*, though the precise mechanisms used may be different.

The nuclear factor of activated T cells (NF-AT) is a second regulatory factor involved in IL-2 transcription whose activity is controlled by phosphorylation. NF-AT is a multi-component T-cell-specific, DNA-binding factor. It is produced in a protein-synthesis-dependent fashion within 20 minutes after stimulation of T-cells with the combination of TPA and a

calcium ionophore, which together mimic many of the effects of T-cell activation through antigen recognition, including stimulation of IL-2 transcription (Shaw et al. 1988; Crabtree 1989). Through a series of *in vitro* DNA-binding assays using nuclear and cytoplasmic extracts from stimulated and unstimulated cells, NF-AT was found to consist of at least two parts (Flanagan et al. 1991). One is a T cell restricted factor which is present in the cytoplasm of unstimulated cells, but is translocated to the nucleus upon treatment of the cells with TPA and ionophore or with ionophore alone (Flanagan et al. 1991). The other is a factor, containing c-fos and c-jun (Jain et al. 1992), that is synthesized in a number of different cell types in response to treatment with TPA or TPA+ionophore and which is located in the nucleus of the activated cell (Flanagan et al. 1991). Stimulation of T cells with TPA and ionophore in the presence of the immunosuppressant drugs cyclosporin A (CsA) or FK506 does not result in the synthesis of a number of the lymphokines that are normally produced by activated T cells. Among the many effects of CsA and FK506, transcriptional activation of IL-2 is blocked as well as the production of NF-AT (Mattila et al. 1990). When the presence of NF-AT components in different cellular fractions were examined after stimulation in the presence of CsA or FK506, the presence of the TPA-responsive nuclear component was unaltered, but the cytoplasmic component had not translocated to the nucleus (Flanagan et al. 1991). Thus these drugs appear to inhibit NF-AT production by blocking a calcium ionophore-responsive translocation event.

CsA and FK506 are known to function through their interactions with cellular proteins called immunophilins, CsA interacting with the cyclophilin family and FK506 interacting with FK506-binding protein (FKBP) family

(Schreiber and Crabtree 1992). Though the cyclophilins are not homologous to the FKBP's, both families encode proteins with *cis-trans* prolyl isomerase activity, and thus have the potential to be involved in protein folding. Though both CsA and FKBP inhibit the prolyl isomerase activity of their respective immunophilin targets, this is not thought to be the mechanism by which they exert their immunosuppressant effects, as molecules related to FK506 inhibit this activity just as efficiently but do not have immunosuppressive effects (Schreiber and Crabtree 1992). Instead, CsA and FK506 each function, at least in part, as a component of a drug/immunophilin complex which has been demonstrated to bind and inhibit the activity of the calcium- and calmodulin-dependent protein phosphatase calcineurin *in vitro* (Liu et al. 1991). The *in vivo* validity of this *in vitro* model was assessed by overexpressing native calcineurin or a constitutively active form of calcineurin in a human T cell line (Clipstone and Crabtree 1992; O'Keefe et al. 1992). This overexpression increased the concentrations of CsA and FK506 required to inhibit transcription from the IL-2 promoter as well as to inhibit transcription from a reporter plasmid regulated by several NF-AT recognition sites. In addition, the overexpression of calcineurin decreased the need for calcium ionophore to stimulate transcription, whereas the overexpression of a different phosphatase had no effect. Thus, calcineurin can play an important role in T cell activation and CsA and FK506 exert at least some of their immunosuppressive effects by inhibiting its activity. Combining this data with the effects of these drugs on NF-AT synthesis and activity strongly suggests that the nuclear translocation of the T-cell-specific cytoplasmic component of NF-AT depends on a dephosphorylation event. Furthermore, the synthesis of the other component

of NF-AT is stimulated by TPA, suggesting that its production can be regulated by PKC. In both cases the precise targets of the phosphorylation and dephosphorylation events are unknown, however phosphorylation can control both the production and cellular localization of the components of NF-AT.

In addition to altering the cellular localization of regulatory factors, phosphorylation can alter the ability of proteins to bind to DNA and to regulate transcription. One example that illustrates both principles is the control of c-jun function (a component of AP-1) by phosphorylation. Phosphorylation of threonine and serine residues near its DNA-binding domain can inhibit c-jun's ability to bind DNA *in vitro* (Lin et al. 1992). Activation of PKC results in the dephosphorylation of these residues *in vivo* and an increase in AP-1 binding activity (Boyle et al. 1991). Phosphorylation can also alter the ability of c-jun to activate transcription when bound to DNA, as phosphorylation of serine residues in its activation domain (distinct from the serines near its DNA-binding domain) correlates with an increase in activation by AP-1 and mutation of these serines results in a decrease of transcriptional activating ability (Pulverer et al. 1991; Smeal et al. 1991; Hunter and Karin 1992; Smeal et al. 1992). In addition, phosphorylation not only can regulate whether or not a factor can bind DNA and what its transcriptional activation properties are, it can also regulate the sequence-specificity of DNA-binding. The binding of the GHF1/Pit1 factor to some sites was decreased by phosphorylation, while the binding to other sites was unaffected or enhanced (Kapiloff et al. 1991). Thus the cell can use phosphorylation not only to regulate whether or not a factor can bind DNA, it can affect to which site a factor binds as well. Though no examples were

specifically mentioned above, phosphorylation can also be used in the control of repressors of transcription (Hunter and Karin 1992).

The effects of phosphorylation on DNA-binding and transcriptional activation vary depending on the regulators (both DNA-binding and non-DNA-binding) being modified and the precise modifications that are made (Hunter and Karin 1992). In cases where increased phosphorylation inhibits DNA-binding or enhances transcriptional activation, the increased negative charge could function by electrostatically decreasing a DNA-binding domain's affinity for DNA or increasing an activation domain's acidic character. But these effects will apply to only a subset of such cases. Many of the regulatory effects of phosphorylation will result from conformational changes that expose or occlude DNA-binding surfaces, activation/repression domains, or regions of the protein involved in other sorts of protein:protein interactions. Changes in the chemical properties of surfaces of the protein that recognize other proteins, creating or destroying adhesive interactions, for example, could also be involved. And, as demonstrated for NF-κB regulation through phosphorylation of I-κB, the effects can be indirect. Despite the many clear examples of such regulation, the physical details of how phosphorylation affects regulatory factor activity remain largely a matter of speculation in most cases.

Control through changes in the phosphorylation state of proteins is a way of rapidly altering the activity of regulatory factors in response to a signal. The abundance of different kinases and phosphatases with different substrate specificities ensure that combinations of signals can be sent separately through this cellular messenger system. NF-AT synthesis provides an example how this system can be used to integrate information. The signal

that stimulates translocation of the cytoplasmic component is different than the signal that stimulates production of the fos/jun containing component. Only the temporally coincident transmission of these two signals will result in the creation of the single transcriptional activator NF-AT. Systems such as this which allow for the integration of information are extremely important in the regulation of transcription.

**Oligomerization of regulatory factors**

Many transcriptional regulators contain domains that mediate protein:protein interactions and readily form oligomers with specific sets of partners (which often include themselves). Often these proteins function, for example, bind to DNA, only when oligomerized. Many of the regulatory proteins which possess the ability to oligomerize are members of large gene families, such as the helix-loop-helix (HLH), leucine zipper and NF-kB/c-rel families (Murre et al. 1989; Lamb and McKnight 1991; Nolan and Baltimore 1992). The members of individual families usually oligomerize with a limited subset of their own group, and some may oligomerize with apparently unrelated proteins (Bengal et al. 1992). This ability to create many different oligomeric species by combining a few different individual species in a number of different ways has been termed combinatorial complexity. Such combinatorial complexity can be used to create multiple species with distinct DNA-binding properties and transcriptional activation/repression properties.

Combinatorial mechanisms not only increase the potential diversity of transcriptional regulators, they can be used to integrate multiple signals into a single output. Different proteins that can oligomerize with one another can each be controlled by different signals. Thus the creation of a key oligomeric

combination might rely on a number of different signal components. One situation in which such a scheme has been proposed is skeletal muscle differentiation (Benezra et al. 1990).

Undifferentiated skeletal muscle precursor cells called myoblasts express members of the skeletal muscle-specific MyoD family (hereafter referred to generically as MyoD) of HLH proteins (Mueller and Wold 1989; Bober et al. 1991). These proteins can form DNA-binding oligomers with themselves as well as DNA-binding oligomers with other partners, such as the more ubiquitously expressed E2A proteins (Lassar et al. 1991). These MyoD-containing DNA-binding species appear to regulate the transcription of a number of genes expressed in differentiated skeletal muscle cells, including muscle creatine kinase (MCK) (Weintraub et al. 1991). But despite the presence of both MyoD and E2A proteins in myoblasts, differentiation-specific genes such as MCK are not expressed. A combination of *in vivo* and *in vitro* analysis suggested one possible reason that differention-specific genes are not expressed in myoblasts is that the MyoD/E2A oligomers that recognize the transcriptional regulatory regions of MCK fail to form in the myoblast (Benezra et al. 1990). A possible mechanism explaining these observations came with the discovery of a variant form of HLH protein called Id (Benezra et al. 1990). Id can oligomerize with MyoD and with E2A family members, but neither the MyoD/Id nor E2A/Id oligomers bind to DNA. The levels of Id RNA were found to decrease upon muscle differentiation *in vitro* and overexpression of the Id protein inhibited activation of transcription through the MCK regulatory region (Benezra et al. 1990). Benezra et al. used these pieces of information to create a model in which the expression of differentiated skeletal muscle-specific genes was dependent on the formation

of MyoD/E2A family oligomers. The cell was marked as a determined myoblast by its expression of the MyoD family. However, until the circumstances were appropriate for it to stop proliferating and differentiate, high levels of Id would be expressed. High levels of Id would promote the accumulation of MyoD/Id or E2A/Id oligomers and inhibit the production of MyoD/E2A family oligomers. The signal to differentiate would downregulate Id levels, allow the production of MyoD/E2A family oligomers and thus allow expression of differentiated skeletal-muscle-specific genes such as MCK. Though recent observations complicate this simple picture (Bengal et al. 1992), it clearly illustrates the concept of using combinatorial complexity to integrate multiple signals into a single output and emphasizes how the relative levels and not just the presence or absence of different interacting proteins can exert powerful regulatory effects.

In the previous example the existence of a key DNA-binding heteromeric combination was assumed and the effects of combinatorial complexity were described in terms of creating functional versus non-functional oligomers. Though this case is a useful starting point, it is important to consider how combinatorial complexity could contribute to important alterations in DNA-binding specificity. In the case of the leucine zipper family regulatory proteins fos, jun, and ATF, different DNA-binding specificities have been described for different dimeric pairs. Fos-jun dimers preferentially bind AP-1 sites over CRE sites while jun-ATF dimers prefer to bind to CRE sites over AP-1 sites (Lamb and McKnight 1991). One could imagine that in a situation in which levels of jun were limiting, the relative levels of fos and ATF could determine whether a promoter containing AP-1 sites or a promoter containing CRE sites functioned. If fos and ATF were

responsive to different signals, the two might function as a molecular balance, measuring the intensity of two inputs and turning on only one gene in response. This strategy of exploiting changes in specificity to create differential gene expression could also be utilized for combinations with differential transcriptional activation or repression effects or by DNA-binding pair-specific interactions with non-DNA-binding regulators. In summary, oligomeric interactions between proteins adds to the diversity of regulatory species available and can be used to help integrate information.

**Additional mechanisms that can control the activity of transcriptional regulators**

A number of other cellular mechanisms that can control the activity of transcriptional regulators have been identified. A number of these can considered covalent/non-covalent modification of regulatory factors, including ligand binding (Evans 1988) and proteolysis (Nolan and Baltimore 1992). In addition to modification of the regulators, there can be modification of their DNA targets. Methylation of cytosine residues is the most common example of covalent modification of DNA and, as it is a heritable modification, it has been implicated in controlling such diverse heritable regulatory phenomena as X-chromosome inactivation and chromosomal imprinting (Riggs and Pfeiffer 1992). Methylation of a target binding site can limit interactions with regulatory factors in a selective fashion. For example methylation of a potential target site for both the myc/max and MLTF DNA-binding regulatory factors inhibits the binding of myc/max, but not of MLTF to this site (Prendergast et al. 1991; Prendergast and Ziff 1991). Methylation can also exert a relatively non-specific repressive effect, mediated perhaps through repressor proteins that bind to methylated DNA such as CP-1

(Meehan et al. 1989). Thus the cell can exert regulatory influences through controlling both the activity of regulatory factors and, potentially, the templates upon which they assemble.

## III) PUTTING IT ALL TOGETHER: HOW A DNA REGULATORY REGION FUNCTIONS

An essential property of a DNA regulatory region is that it has the capacity to integrate multiple inputs, diverse developmental and environmental information, into a single output, RNA transcript production. As described in the preceding sections it does so through interactions with DNA-binding factors that in turn, directly and indirectly, control the assembly and activity of the transcriptional machinery of the cell at the transcriptional start site. Some of the information it receives has already undergone integration, for example by using the oligomerization strategies mentioned above. However, much of the integrative process takes place right at the DNA regulatory region. The function of a DNA regulatory region can be separated into two parts: the occupancy of the DNA regulatory region by DNA-binding regulatory factors and the activation or repression of the basic transcriptional machinery by the bound regulators and their associates. Each part could conceivably function as an integrator of information. In the following section I discuss the mechanisms that may underlie these integrative functions.

### Regulatory region accessibility

In order for gene-specific regulatory factors to recognize their DNA binding sites, the DNA in the site must be, to some extent, free from other interactions. DNA in the cell nucleus interacts with a variety of chromosomal proteins, such as histones, which can package and compact the DNA into structures, for example, the nucleosome described above. One effect of the

packaging of DNA into chromatin can be to inhibit the ability of gene-specific regulatory factors to gain access to their DNA binding sites. This inhibition could be through the masking of important chemical contacts or through steric hindrance (Hayes and Wolffe 1992). Regulation of the accessibility of the DNA regulatory region can be important in controlling regulatory region occupancy.

Regulatory region accessibility plays an important role in controlling transcription of the mouse mammary tumor virus promoter. Prior to glucocorticoid treatment a nucleosome is present just upstream of the core promoter region and the promoter is inactive (Richard-Foy and Hager 1987; Archer et al. 1991). The addition of glucocorticoid stimulates the binding of glucocorticoid receptor (GR) to some of its recognition sites in the promoter, even though those sites are wrapped around the histone octamer. The binding of GR stimulates the disruption of the nucleosome (Richard-Foy and Hager 1987; Archer et al. 1991; Archer et al. 1992). Then other factors, such as the NF1 DNA-binding protein, which unlike GR could not bind to its target site until the nucleosome was disrupted, bind the exposed DNA regulatory region and transcription is stimulated (Cordingley et al. 1987; Archer et al. 1992). Attempts to reconstitute this process *in vitro* have shown that although GR can bind its sites in nucleosomal DNA, it does not cause the disruption of the nucleosome seen *in vivo* nor does it facilitate the binding of NF1 (Archer et al. 1991). Thus the GR, through its ability to recognize its target within nucleosomal DNA, facilitates the binding of additional gene-specific regulatory factors to the MMTV promoter. But though it is essential, additional factors are required to mediate this chromatin remodeling. Perhaps some these additional factors are related to the products of group of

genes, originally identified in yeast but also present in mouse and humans, that have been implicated in the control of chromatin accessibility. The SNF and SWI genes behave genetically as activators of the transcription of many different genes, though the proteins they encode do not appear to bind DNA (Winston and Carlson 1992). Genes isolated as suppressors of SNF and SWI mutations include genes that encode histones as well as the non-histone chromosomal protein HMG1, suggesting that a function of SNF/SWI products may be to reverse the repressive effects of chromatin (Winston and Carlson 1992). These examples suggest that the reversal of chromatin-based repression is a common component of transcriptional activation. In addition, the MMTV example also suggest that chromatin can repress transcription by inhibiting the binding of gene-specific regulators to a DNA regulatory region as well as by inhibiting the assembly of the general transcription machinery at the core promoter.

**Mechanisms for specificity of DNA-binding**

Once a DNA regulatory region has been rendered accessible, regulatory factors must occupy it in order to control transcription (although the arguments made below can also be applied to situations in which accessibility and occupancy are functionally linked). DNA regulatory regions that drive transcription in a very restricted pattern are often composed of multiple factor recognition sites that are each important for expression. Many of these essential elements contribute to very different patterns of transcription when they are present in other DNA regulatory regions. An excellent example is the muscle-specific human α-cardiac actin promoter which contains individual potential binding sites for the MyoD, SP1, and serum response (SRF) regulatory factors (Sartorelli et al. 1990). Mutation of

any one of these factors binding sites completely extinguished the transcription of this gene. Though they are essential for muscle-specific expression of the α-cardiac actin gene, when located in other regulatory regions adjacent to different regulatory elements the SP1 and SRF binding sites are each capable of contributing to transcriptional activation in a variety of other cell-types. Thus DNA regulatory regions provide a series of sites for individual protein:DNA interactions, and the function of the regulatory region relies extensively on how these regulatory factors interact.

The biochemical properties of the regulatory factors provoke as many questions as they answer. *In vitro* experiments have established that cells can contain multiple factors that recognize a single binding site and that a single factor can recognize more than one binding site (Lamb and McKnight 1991). In addition, a cell's nucleus may contain multiple factors capable of recognizing an isolated DNA regulatory element *in vitro*, yet that site may be unoccupied *in vivo* (Becker et al. 1987). What determines which proteins occupy a given site, and what determines when and where a given site is to be occupied?

**Protein:protein interactions to add specificity**

One possible mechanism for achieving regulatory specificity using factors which in isolation appear promiscuous is to require protein:protein interactions between the species that can bind the regulatory region to establish occupancy of a DNA regulatory region (Ptashne 1986). The need for protein:protein interactions could enforce two levels of specificity. First, individual regulatory factors would only bind to DNA regulatory elements that had adjoining sites for compatible factors. Second, the DNA regulatory region would be occupied only when a sufficient number of compatible

proteins were available. Utilizing both protein:protein and DNA:protein interactions would increase the effective sequence-specificity of a factor by increasing its binding affinity for appropriate (specific) sites relative to inappropriate (nonspecific) sites (Ptashne 1986). Cooperative interaction could also create thresholds where small changes in binding factor concentrations could lead to large changes in DNA regulatory region occupancy, analogous to the cooperative binding of oxygen to hemoglobin (Stryer, 1988). Thus, cooperative interactions on a DNA template would help establish a situation in which sufficient quantities of diverse inputs would be required before any regulatory occupancy occurred. This type of mechanism could potentially underlie the phenomena of coordinated occupancy triggered by stimuli as seen *in vivo* at the muscle creatine kinase (Mueller and Wold 1989) and IL-2 DNA regulatory regions (chapter 4 of this thesis).

**Accessibility as an integrative mechanism**

The control of regulatory region accessibility could also contribute to such integrative control. In the case of the glucocorticoid responsive enhancer element of the liver-specific tyrosine-amino-transferase gene (TAT), genomic footprinting studies suggest that the GR may function transiently to remove a nucleosome from the enhancer (Rigaud et al. 1991). The enhancer then becomes occupied by liver-specific factors, potentially through the cooperative mechanisms suggested above. Thus one could imagine this two-step mechanism functioning to integrate two pieces of information, glucocorticoid exposure and cell-type identity. In the absence of either piece of information the enhancer would not be occupied.

**Physical basis of cooperativity**

What could be the physical basis of cooperativity between adjacent DNA-binding proteins? Involvement of the DNA-binding domain in protein:protein interactions has been established for a number of proteins. In one example (Thompson and McKnight 1992), the DNA-binding protein GABPα weakly recognizes the "cigar" motif DNA-binding site in the ICP4 promoter through an ETS family DNA-binding domain. It recognizes this cigar motif much more avidly if it is mixed with the protein GABPβ, a protein which does not bind DNA on its own. The two proteins interact through the ETS domain of GABPα and a TPLH repeat domain in the GABPβ protein. This TPLH repeat also appears to interact with DNA in the GABPα/GABPβ complex. The protein GABPβ interacts only with GABPα, and not with other ETS domain proteins. This could serve to make the cigar motif specific for this particular combination of factors (Thompson and McKnight 1992). Such bifunctional DNA-binding regions could also contribute to the cooperative binding interactions observed on other DNA elements which contain adjacent binding sites for factors whose individual patterns of expression are different than the regulatory function of the element, such as the OAP40/OCT site of the IL-2 promoter (Ullman et al. 1991).

Mutational studies of DNA regulatory regions such as IL-2 demonstrate that multiple individual factor binding sites can collaborate despite being dispersed over hundreds of base pairs (summarized in chapter 4 of this thesis). Cooperativity between DNA-binding factors is still possible even when the binding sites are separated by multiple turns of the DNA helix, often involving bending the intervening DNA (Ptashne 1986). In cases where binding sites are not adjacent, the interactions might involve greater contributions from protein:protein interaction domains distinct from the

DNA-binding region. In fact, a number of DNA-binding proteins contain multiple oligomerization domains. For example, the E2A proteins contain leucine zipper domains in a region of the protein separate from the well-characterized DNA-binding and HLH domains (Murre et al. 1989). Such separate domains could serve to link a DNA-bound E2A protein to a protein bound elsewhere, though this has not been demonstrated. In addition, the GABPβ protein which binds at the cigar motif of the ICP4 promoter in a complex with GABPα, also has a separate oligomerization domain through which it dimerizes with itself (Thompson and McKnight 1992). Thus the GABP complex forms a tetramer, which is especially suited to the two cigar motifs located near one another at the ICP4 promoter. Cooperative interactions between separate gene-specific DNA-bound factors could also be through non-DNA-binding proteins. These proteins could either be gene-specific regulators or, potentially, more general factors such as coactivators that might link together multiple proteins by simultaneously interacting with their activation domains. Thus cooperativity can be mediated through a variety of different particular strategies, but all involve DNA:protein and protein:protein interactions.

**DNA regulatory region as template for assembly of a series of interdependent regulatory factor:regulatory factor:DNA complexes**

In summary, one might propose occupancy is governed by a series of cooperative assembly events across a DNA regulatory region. An initial level might involve intimate contact between proteins that recognize adjacent sites. The stability of these complexes would then be dependent on interactions with other, more distantly bound regulators. The first level of interactions would be very sensitive to changes in spacing between regulatory elements,

whereas the subsequent interactions would be more tolerant of changes in spacing between individual regulatory sites (Ptashne 1986). The DNA would thus serve as a template for the assembly of the complex. As noted by Lamb and McKnight (1991), many of the protein:protein interactions involved in such an assembly would be relatively weak and require the DNA template in order form. One example of such DNA-mediated protein:protein interaction has been shown for Pit-1 which dimerizes only when bound to its DNA recognition site (Ingraham et al. 1990). Such DNA facilitated protein:protein interaction could be a major way in which a DNA regulatory region integrates information.

**Cooperativity of transcriptional activation**

Occupancy of the DNA regulatory region is only the first part of transcriptional regulation. The activation or repression of the core promoter is the second part and is an equally likely target for control. The potential for cooperative activation was tested in an experiment using GAL4-VP16 (Carey et al. 1990). As the number of GAL4 binding sites was increased (in the presence of sufficient amounts of GAL4-VP16 to saturate all binding sites), the amount of transcription increased in a greater than linear fashion. This apparent cooperativity of activation was called synergism, and was also observed between multiple different regulatory factors (Lin et al. 1990). Synergism could provide an additional way in which multiple inputs could be integrated at a regulatory region. In this case, multiple activation domains would need to be assembled to effectively activate transcription. Models in which combinations of different regulatory factors activate transcription more effectively than a similar number of copies of a single factor have also been proposed (Han et al. 1989). Thus, even though multiple factors could be

bound to a DNA regulatory region, unless a sufficient number or combination of functional activating domains were present no transcriptional activation might occur. In this case as well, the heart of the integrative effect would be that the output is different than the sum of its parts.

**Interaction between repressors and activators**

This discussion has addressed only positive levels of control. There are a number of DNA regulatory regions that are known to be negatively as well as positively controlled. This has been especially well studied in the *even-skipped* (*eve*) stripe two enhancer DNA regulatory region of *Drosophila* (Small et al. 1991). This DNA regulatory region senses position along the anterior/posterior axis of the embryo based on the relative levels of multiple regulatory factors. This enhancer integrates the relatively gradual concentration gradients of several factors into a very sharply bounded domain of expression, creating large differences in gene activity between adjacent nuclei in the embryo. A single gradient could be used to create sharp boundaries through a relatively simple cooperative binding or activation mechanism, but judging the relative levels of different regulatory factors is more complicated. As indicated above, oligomerization of the different factors can assist in integrating this information at the individual factor level. However, this strategy might not provide sufficiently tight regulation to form precise boundaries of gene activity. For example, relying on Id to inhibit MyoD function works well only if the differences in the levels of the factors involved are relatively large. In the case of the *eve* stripe 2 enhancer where subtle differences need to be acted on precisely, combining cooperative interactions with competition between negative and positively acting factors could provide the necessary sensitivity. Small et al. have

suggested a system for the formation of the posterior boundary of stripe 2 (Small et al. 1991). The binding of the *Kruppel* gene product to one or two sites that overlap sites for the activator *bicoid* was hypothesized to disrupt the potentially cooperative binding of the activators *bicoid* and *hunchback* to their multiple sites. However, an alternative model in which repression is exerted not through a competition for binding sites, but through antagonism of activation cannot be excluded. The net effect of a scheme of interaction like that proposed for the *eve* stripe 2 element is that mild differences in individual regulators are amplified into large changes in transcription.

**Modulation of transcriptional level**

DNA regulatory regions control not only whether there is transcription, but the level of transcription as well. Therefore, once a threshold of stable occupancy and transcriptional initiation has been established, additional interactions need to be accommodated to provide more subtle alterations in level of expression. The transcriptional upregulation of the constitutively active mouse metallothionein-I promoter upon exposure to metal provides one example of how the level of transcription can be controlled. Genomic footprinting showed that this upregulation is accompanied by the binding of additional factors to its regulatory region (Mueller et al. 1988). A quite different mechanism is used in the glucocorticoid-dependent upregulation of TAT promoter activity (Rigaud et al. 1991). In TAT the upregulation involves the occupancy of a previously unoccupied glucocorticoid-responsive enhancer. Thus modulation of ongoing transcription can involve the modest remodeling of protein:DNA and protein:protein interactions on an already occupied DNA

template as well as the activation of additional DNA regulatory elements that can cooperate with those already present.

## IV) TAKING IT ALL APART: FIGURING OUT HOW REGULATORY REGIONS WORK

The question of how information is integrated by DNA regulatory regions is clearly one of the most important questions to be understood in how transcription occurs in response to environmental and developmental cues. In the above section I discussed several integrative strategies that could be used together for appropriate regulation: DNA regulatory region accessibility, cooperative and competitive DNA occupancy, and synergistic transcriptional activation. *In vivo* footprinting experiments such as those described below suggest that while some DNA regulatory regions respond to signals with a change in the occupancy of factors, other regulatory regions respond by changing the activity of factors already bound to the regulatory region. To achieve an understanding of how particular DNA regulatory regions function, a number of different types of experimental approaches must be pursued. I will briefly describe the basic experimental approaches, discuss their limitations, explain why I have worked on the development of *in vivo* footprinting techniques, and finally, what has been learned through my investigations of *in vivo* protein:DNA interactions.

### Cis-element analysis

An essential component of understanding transcriptional regulation is the definition and dissection of DNA regulatory regions through cis-element analysis. Cis-element analysis is often performed by fusing potential regulatory DNA regions to DNA sequences encoding reporter proteins whose synthesis can be easily analyzed. These DNA sequences are created

using recombinant DNA techniques and the promoter/reporter fusions are introduced into cells through DNA transfection techniques. The production of the reporter protein is analyzed and used to gauge the regulatory capacity of a DNA sequence. Once a DNA regulatory region possessing the desired regulatory properties has been defined, this region can be further dissected and individual binding sites for regulatory factors can be identified. Mutational analysis of individual elements can give an idea of which are essential for the proper regulation of a regulatory region and which may serve only to modulate the level of transcription. Once defined, individual elements can be studied in isolation. Such experiments can give an idea of the properties of regulatory factors that function through an element. However, it is often necessary to multimerize an individual element for it to support detectable transcription *in vivo*. This could change the character of the element as the protein:protein interactions that underlie this need for multimerization are probably different from the protein:protein interactions that govern its activity in the context of the original regulatory region. The major limitation of such analyses is that a regulatory region's properties are determined by the collaboration of individual elements that often behave very differently from one another when studied in isolation. Mutational analysis of the intact DNA regulatory region can help establish the existence of collaboration between individual elements, but it does not allow one to determine the biochemical mechanisms used in that collaboration.

Additional interpretational complications are inherent in the nature of most cis-element analysis. DNA regulatory regions must be removed from their native chromosomal locations in order to study their components. Chromatin structure, which can play an important role in transcriptional

regulation, may differ, for example, between an endogenous gene and a recently transfected gene, especially in transient transfection analyses where most copies of the regulatory region are not integrated into a chromosome. Clear differences in chromatin structure at the MMTV promoter were observed when a transiently transfected MMTV DNA regulatory region was compared with the identical regulatory region in the chromosome of the transfected cell (Archer et al. 1992). The chromosomal promoter contained a nucleosome (as described above) which prevented the binding of NF1 to its recognition site. Upon glucocorticoid treatment, the glucocorticoid receptor bound to its site, the nucleosome was disrupted, and NF1 bound to its site. However, the transiently transfected promoter never showed evidence of a bound nucleosome and NF1 was bound to its site constitutively. Thus, although cis-element analysis is indispensable in the study of transcriptional control, it is limited in what mechanistic insight it can provide into the control of endogenous promoters.

***In vitro* biochemistry**

*In vitro* biochemistry, involving cell-free experiments using cell extracts and purified factors (e.g., *in vitro* transcription reactions and electrophoretic mobility shift assays), is another essential tool for gaining an understanding of transcriptional control mechanisms. *In vitro* biochemistry can be used to describe in detail the interaction of regulatory factors with DNA regulatory elements and the mechanisms by which regulatory factors affect the activity and assembly of the basic transcriptional machinery of the cell. As we have discussed above, regulatory factors function *in vivo* in the context of a multicomponent protein:DNA complex. Despite the inherent difficulty of trying to accurately simulate this process *in vitro*, an enormous

amount of mechanistic insight has been obtained. However, the selection of reaction components, their concentrations and the post-translational modifications that are made to them will all influence the results obtained. This makes it essential to combine *in vitro* results with *in vivo* observations to confirm that the *in vitro* results accurately reflect *in vivo* processes. *In vitro* observations are indispensable for a mechanistic understanding of transcription *in vivo*, but they cannot stand alone.

**Genetic analysis**

Genetic analysis of transcriptional control can identify key components that may be biochemically difficult to obtain and can allow the pathways in which different components interact in the living organism to be defined. For example, much of the understanding of the *eve* stripe 2 enhancer element is based on genetic data (Small et al. 1991). However, genetic analysis alone is somewhat limited in its ability to determine the detailed biochemistry of the transcriptional process.

***In vivo* footprinting**

Cis-element analysis, *in vitro* biochemistry, and genetic analysis each provide essential information, but each have their own limitations in what information they can provide about the biochemical mechanisms at work *in vivo*, in an intact cell. An additional approach which complements these is the *in vivo* biochemical approach of *in vivo* footprinting. *In vivo* footprinting provides information about when and where regulatory factors are bound to a DNA regulatory region in the intact cell. Knowing when and where regulatory factors are bound to DNA can provide substantial information about such questions as whether DNA occupancy or transcriptional activation is likely to be the limiting step at a particular DNA regulatory

region. Furthermore, it can also provide information about the function of the core promoter, such as whether a complex is assembled there and whether a paused polymerase is likely to be present (Giardina et al. 1992).

**LMPCR**

*In vivo* footprinting in organisms with large genomes (such as mammals) was initially introduced by Ephrussi et al. in 1985. However, it was not widely utilized as it was technically challenging and even when successful resulted in a low level of signal and an unfavorable signal to noise level. Several years ago, Paul Mueller and I devised the ligation-mediated polymerase chain reaction (LMPCR) strategy for performing genomic footprinting. Paul Mueller's original LMPCR procedure substantially improved the absolute level of signal, the signal to noise ratio, and decreased the amount of starting material needed for *in vivo* footprinting (Mueller and Wold 1989). However, the quality of the result varied in a regulatory region-dependent fashion. Some DNA regions could not be visualized, while many others contained spurious bands which destroyed much of the *in vivo* footprint data. The third chapter of this thesis describes alterations made to LMPCR which improved the quality of genomic footprinting and substantially expanded the range of DNA regulatory regions that could be successfully examined through *in vivo* footprinting.

***In vivo* footprinting of IL-2**

The fourth chapter of this thesis describes the use of this LMPCR *in vivo* footprinting procedure to investigate the transcriptional regulatory mechanism involved in the induction of IL-2 RNA synthesis in stimulated mouse T cells. Stimulation of the EL4 mouse thymoma cell line (EL4 T cells) was shown to trigger the coordinated occupancy of multiple regulatory

elements in the IL-2 DNA regulatory region *in vivo*. Further evidence for the integration of multiple inputs at the IL-2 regulatory region by the mechanism of coordinated occupancy was obtained through the use of the immunosuppressant drug CsA. Through the use of electrophoretic mobility shift assays to characterize the presence of regulatory factors in nuclear extracts, CsA was shown to block the DNA-binding activity of some, but not all of the regulatory factors activated in stimulated EL4 T cells. However, CsA blocked the binding of all these factors to the IL-2 regulatory region *in vivo*. Thus CsA was shown to block IL-2 transcriptional induction by blocking the coordinated occupancy of the IL-2 regulatory region by regulatory factors. IL-2 is synthesized as a result of the T cell integrating multiple internal and external signals. The observed multi-component, coordinated assembly appears to be an essential component of this integrative process.

***In vivo* footprinting of the promoter of the acetylcholine receptor δ-subunit**

The use of LMPCR *in vivo* footprinting to observe the pattern of *in vivo* protein:DNA interactions at the δ-AChR regulatory region gives a result very different from that seen for IL-2. The δ-AChR regulatory region was occupied by multiple factors in undifferentiated myoblasts, though this regulatory region is not transcriptionally active in these cells until they have differentiated into myocytes. Upon differentiation, the pattern of *in vivo* interactions at the δ-AChR regulatory region was unaltered. Therefore, activation of δ-AChR transcription may be due to alterations in the ability of bound regulatory factors to activate transcriptional initiation, the replacement of previously bound factors with other factors at the same sites, or alterations in transcriptional elongation. However, the MCK enhancer region, which has

the same pattern of transcriptional activation in the cells examined, was unoccupied until it activated transcription. Therefore, the δ-AChR and MCK enhancer use quite different activation mechanisms even though their pattern of expression is similar.

**REFERENCES:**

Al-Shawi, R., J. Burke, H. Wallace, C. Jones, S. Harrison, D. Buxton, S. Maley, A. Chandley, and J. O. Bishop. 1991. The herpes simplex virus type 1 thymidine kinase is expressed in the testes of transgenic mice under the control of a cryptic promoter. *Mol. Cell. Biol.* **11:** 4207-4216.

Archer, T. K., M. G. Cordingley, R. G. Wolford, and G. L. Hager. 1991. Transcription factors access is mediated by accurately positioned nucleosomes on the mouse mammary tumor virus promoter. *Mol. Cell. Biol.* **11:** 688-698.

Archer, T. K., P. Lefebvre, R. G. Wolford, and G. L. Hager. 1992. Transcription factor loading on the MMTV promoter: a bimodal mechanism for promoter activation. *Science* **255:** 1573-1576.

Baeuerle, P. A., and D. Baltimore. 1988. Activation of DNA-binding activity in an apparently cytoplasmic precursor of the NF-κB transcription factor. *Cell* **53:** 211-217.

Becker, P. B., S. Ruppert, and G. Schutz. 1987. Genomic footprinting reveals cell-type specific DNA binding of ubiquitous factors. *Cell* **51:** 435-445.

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

Bengal, E., L. Ransone, R. Scharfmann, V. J. Dwarki, S. J. Tapscott, H. Weintraub, and I. M. Verma. 1992. Functional anatagonism between

c-jun and myoD proteins: a direct physical association. *Cell* **68:** 507-519.

Bober, E., G. E. Lyons, T. Braun, G. Cossu, M. Buckingham, and H. H. Arnold. 1991. The muscle regulatory gene, Myf-6, has a biphasic pattern of expression during early mouse development. *J. Cell Biol.* **113:** 1255-1265.

Boyle, W. J., T. Smeal, L. H. K. Defize, P. Angel, J. R. Woodgett, M. Karin, and T. Hunter. 1991. Activation of protein kinase C decreases phosphorylation of c-jun at sites that negatively regulate its DNA-binding activity. *Cell* **64:** 573-584.

Briegel, K., B. Hentsch, I. Pfeuffer, and E. Serfling. 1991. One base pair change abolishes the T cell -restricted activity of a κB-like proto-enhancer element from the interleukin-2 promoter. *Nucl. Acids Res.* **19:** 5929-5936.

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

Carcamo, J., E. Maldonado, P. Cortes, M. H. Ahn, I. Ha, Y. Kasai, J. Flint, and D. Reinberg. 1990. A TATA-like sequence located downstream of the transcription initiation site is required for expression of an RNA polymerase II transcribed gene. *Genes and Dev.* **4:** 1611-1622.

Carey, M., Y. S. Lin, M. R. Green, and M. Ptashne. 1990. A mechanism for the synergistic activation of a mammalian gene by GAL4 derivatives. *Nature* **345:** 361-364.

Clipstone, N. A., and G. R. Crabtree. 1992. Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation. *Nature* **357:** 695-697.

Cordingley, M. G., A. Tate-Riegel, and G. L. Hager. 1987. Steroid-dependent interaction of transcription factore with the inducible promoter of mouse mammary tumour virus *in vivo*. *Cell* **48:** 261-270.

Courey, A. J., and R. Tjian. 1988. Analysis of SP1 *in vivo* reveals multiple domains, including a novel glutamine-rich activation motif. *Cell* **55:** 887-898.

Crabtree, G. R. 1989. Contingent genetic regulatory events in T lymphocyte activation. *Science* **243:** 355-361.

Davidson, E. H. 1991. Spatial mechanisms of gene regulation in metazoan embryos. *Development* **113:** 1-26.

Dynlacht, B. D., T. Hoey, and R. Tjian. 1991. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* **66:** 563-576.

Ephrussi, A., G.M. Church, S. Tonegawa, and W. Gilbert. 1985. B lineage specific interactions of an immunoglobulin enhancer with cellular factors *in vivo*. *Science* **227:** 134-140.

Evans, R. M. 1988. The steroid and thyroid hormone receptor superfamily. *Science* **240:** 889-895.

Felsenfeld, G. 1992. Chromatin as an essential part of the transcriptional mechanism. *Nature* **355:** 219-224.

Flanagan, W. M., B. Corthesy, R. J. Bram, and G. R. Crabtree. 1991. Nuclear association of a T-cell transcription factor blocked by FK-506 and cylclosporin A. *Nature* **352:** 803-807.

Fujita, T., G. P. Nolan, S. Ghosh, and D. Baltimore. 1992. Independent modes of transcriptional activation by the p50 and p65 subunits of NF-κB. *Genes and Dev.* **6:** 775-787.

Garrity, P. A., and B. J. Wold. 1990. Tissue-specific expression from a compound TATA-dependent and TATA-independent promoter. *Mol. Cell. Biol.* **10:** 5646-5654.

Ghosh, S., and D. Baltimore. 1990. Activation *in vitro* of NF-κB by phosphorylation of its inhibitor I-κB. *Nature* **344:** 678-682.

Giardina, C., M. Perez-Riba, and J. T. Lis. 1992. Promoter melting and TFIID complexes on *Drosophila* gene *in vivo*. *Genes and Dev.* **6:** 2190-2200.

Gill, G., and R. Tjian. 1992. Eukaryotic coactivators associated with the TATA box binding protein. *Current Opinion in Genetics and Development* **2:** 236-242.

Greenblatt, J. 1991. RNA polymerase-associated transcription factors. *Trends Biochem. Sci.* **16:** 408-411.

Guarente, L., and O. Bermingham-McDonogh. 1992. Conservation and evolution of transcriptional mechanisms in eukaryotes. *Trends Genet.* **8:** 27-32.

Han, K., M. S. Levine, and J. L. Manley. 1989. Synergistic activation and repressin of transcription by *Drosophila* homeobox proteins. *Cell* **56:** 573-583.

Hayes, J. J., and A. P. Wolffe. 1992. The interaction of transcription factors with nucleosomal DNA. *BioEssays* **14:** 597-603.

Hoch, M., N. Gerwin, H. Taubert, and H. Jackle. 1992. Competition for overlapping sites in the regulatoryregion of the *Drosophila* gene *Kruppel*. *Science* **256:** 94-97.

Hunter, T., and M. Karin. 1992. The regulation of transcription by phosphorylation. *Cell* **70:** 375-387.

Ingles, C. J., M. Shales, and W. D. Cress. 1991. Reduced binding of TFIID to transcriptionally compromised mutants of VP16. *Nature* **351:** 588-590.

Ingraham, H. A., S. E. Flynn, J. W. Voss, V. R. Albert, M. S. Kapiloff, L. Wilson, and M. G. Rosenfeld. 1990. The POU-specific domain of Pit-1 is essential for sequence-specific, high-affinity DNA binding and DNA-dependent Pit-1:Pit-1 interactions. *Cell* **61:** 1021-1033.

Jain, J., P. G. McCaffrey, V. E. Valge-Archer, and A. Rao. 1992. Nuclear factor of activated T cells contains Fos and Jun. *Nature* **356:** 801-804.

Johnson, F. B., and M. A. Krasnow. 1992. Differential regulation of transcription preinitiation complex assembly by activator and repressor homeo domain proteins. *Genes and Dev.* **6:** 2177-2189.

Kadonaga, J. T. 1990. Assembly and disassembly of the *Drosophila* RNA polymerase II complex during transcription. *J. Biol. Chem.* **265:** 2624-2631.

Kapiloff, M. S., Y. Farkash, M. Wegner, and M. G. Rosenfeld. 1991. Variable effects of phosphorylation of Pit-1 dictated by the DNA response elements. *Science* **253:** 786-789.

Karn, J., and M. A. Graeble. 1992. New insights into the mechanism of HIV-1 transactivation. *Trends Genet.* **8:**

Katagiri, F., K. Yamazaki, M. Horikoshi, R. G. Roeder, and N. H. Chua. 1990. A plant DNA-binding protein increases the number of active preinitiation complexes in a human *in vitro* transcription system. *Genes and Dev.* **4:** 1899-1909.

Keleher, C. A., M. J. Redd, J. Schultz, M. Carlson, and A. D. Johnson. 1992. Ssn6-Tup1 is a general repressor of transcription in yeast. *Cell* **68:** 709-719.

Keller, A. D., and T. Maniatis. 1991. Identification and characterization of a novel repressor of B-interferon gene expression. *Genes and Dev.* **5:** 868-879.

Lamb, P., and S. L. McKnight. 1991. Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. *Trends Biochem. Sci.* **16:** 417-422.

Lassar, A. B., R. L. Davis, W. E. Wright, T. Kadesch, C. Murre, A. Voronova, D. Baltimore, and H. Weintraub. 1991. Functional activity of myogenic HLH proteins requires heter-oligomerization with E12/E47-like proteins *in vivo*. *Cell* **66:** 305-315.

Laybourn, P. J., and J. T. Kadonaga. 1991. Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. *Science* **254:** 238-245.

Lee, H., K. W. Kraus, M. F. Wolfner, and J. T. Lis. 1992. DNA sequence requirements for generating paused polymerase at the start of *hsp70*. *Genes and Dev.* **6:** 284-295.

Lee, J. H., Y. C. Li, S. Doerre, P. Sista, D. W. Ballard, W. C. Green, and R. B. Franza. 1991. A member of the set of κB binding proteins, HIVEN86A, is a product of the human *c-rel* proto-oncogene. *Oncogene* **6:** 665-667.

Leuther, K. K., and S. A. Johnston. 1992. Nondissociation of GAL4 and GAL80 *in vivo* after galactose induction. *Science* **256:** 1333-1335.

Lin, A., J. Frost, T. Deng, T. Smeal, N. Al-Alawi, U. Kikkawa, T. Hunter, D. Brenner, and M. Karin. 1992. Casein kinase II is a negative regulatory of c-jun DNA binding ad AP-1 activity. *Cell* **70:** 777-789.

Lin, Y., M. Carey, M. Ptashne, and M. Green. 1990. How different eukaryotic transcriptional activators can cooperate promiscuously. *Nature* **345:** 359-361.

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

Lin, Y. S. I., E. Maldonado, D. Reinberg, and M. R. Green. 1991. Binding of general transcription factor TFIIB to an acidic activating region. *Nature* **353:** 569-571.

Liu, J., J. D. Farmer, W. S. Lane, J. Friedman, I. Weissman, and S. L. Schreiber. 1991. Calcineurin is a common target of cylcophilin-cyclosporin A and FKBP-FK506 complexes. *Cell* **66:** 807-815.

Mattila, P. S., K. S. Ullman, S. Fiering, E. A. Emmel, M. McCutcheon, G. R. Crabtree, and L. A. Herzenberg. 1990. The actions of cyclosporin A and FK506 suggest a novel step in the activation of T lymphocytes. *EMBO J.* **9:** 4425-4433.

Means, A. L., and P. J. Farnham. 1990. Transcription initiation from the dihydrofolate reductase promoter is positioned by housekeeping initiation protein 1 binding at the initiation site. *Mol. Cell. Biol.* **10:** 653-661.

Meehan, R. R., J. D. Lewis, S. McKay, E. L. Kleiner, and A. P. Bird. 1989. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58:** 499-507.

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

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

Mueller, P. R., S. J. Salser, and B. Wold. 1988. Constitutive and metal-inducible protein:DNA interactions at the mouse metallothionein-I promoter examined by *in vivo* and *in vitro* footprinting. *Genes and Dev.* **2:** 412-427.

Mueller, P. R., and B. Wold. 1989. *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* **246:** 780-786.

Murre, C., P. S. McCaw, and D. Baltimore. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *myoD*, and *myc* proteins. *Cell* **56:** 777-783.

Nolan, G. P., and D. Baltimore. 1992. The inhibitory ankyrin and activator Rel proteins. *Current Opinion in Genetics and Development* **2:** 211-226.

O'Keefe, S. J., J. Tamure, R. L. Kincaid, M. J. Tocci, and E. A. O'Neill. 1992. FK506 and CsA-sensitive activation of the IL-2 promoter by calcineurin. *Nature* **357:** 692-694.

O'Shea-Greenfield, A., and S. T. Smale. 1992. Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. *J. Biol. Chem.* **267:** 1391-1402.

Prendergast, G. C., D. Lawe, and E. B. Ziff. 1991. Association of myn, the murine homolog of max, with c-myc stimulates methylation-sensitive DNA binding and ras cotransformation. *Cell* **65:** 395-407.

Prendergast, G. C., and E. B. Ziff. 1991. Methylation-sensitive sequence-specific DNA-binding by the c-myc basic region. *Science* **251:** 186-189.

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

Ptashne, M., and A. A. F. Gann. 1990. Activators and targets. *Nature* **346:** 329-331.

Pulverer, B. J., J. M. Kyriakis, J. Avruch, E. Nikolakaki, and J. R. Woodgett. 1991. Phosphorylation of c-jun mediated by MAP kinases. *Nature* **353:** 670-674.

Richard-Foy, H., and G. L. Hager. 1987. Sequence-specific positioning of nucleosomes over the stroid-inducible MMTV promoter. *EMBO J.* **6:** 2321-2328.

Rigaud, G., J. Roux, R. Pictet, and T. Grange. 1991. *In vivo* footprinting of rat TAT gene: dynamic interplay between the glucocorticoid receptor and a liver-specific factor. *Cell* **67:** 977-986.

Riggs, A. D., and G. P. Pfeiffer. 1992. X-chromosome inactivation and cell memory. *Trends Genet.* **8:** 169-174.

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

Rougvie, A. E., and J. T. Lis. 1990. Post-initiation transcriptional control in *Drosophila melanogaster*. *Mol. Cell. Biol.* **10:** 6041-6045.

Roy, A. L., M. Meisterernst, P. Pognonc, and R. G. Roeder. 1991. Cooperative interaction of an initiator-binding transcription initiation factors and the helix-loop-helix activator USF. *Nature* **354:** 245-248.

Sartorelli, V., K. A. Webster, and L. Kedes. 1990. Muscle-specific expression of the cardiac a-actin gene requires myoD1, CArG-box binding factor, and SP1. *Genes and Dev.* **4:** 1811-1822.

Schreiber, S. L., and G. R. Crabtree. 1992. The mechanism of action of cyclosporin A and FK506. *Immun. Today* **13:** 136-142.

Shaw, J.-P., P. J. Utz, D. B. Durand, J. J. Toole, E. A. Emmel, and G. R. Crabtree. 1988. Identification of a putative regulator of early T cell activation genes. *Science* **241:** 202-205.

Small, S., R. Kraut, T. Hoey, R. Warrior, and M. Levine. 1991. Transcriptional regulation of a pair-rule stripe in *Drosophila*. *Genes and Dev.* **5:** 827-839.

Smeal, T., B. Binetruy, D. Mercola, M. Birrer, and M. Karin. 1991. Oncogenic and transcriptional cooperation with Ha-Ras requires phosphorylation of c-jun on serines 63 and 73. *Nature* **354:** 494-496.

Smeal, T., B. Binetruy, D. Mercola, G. Heidecker, U. R. Rapp, and M. Karin. 1992. Oncoprotein mediated signaling cascade stimulates c-jun activity by phosphorylation of serines 63 and 73. *Mol. Cell. Biol.* **12:** 3507-3513.

Stringer, K. F., C. J. Ingles, and J. GreenblatT. 1990. Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. *Nature* **345:** 783-786.

Stromstedt, P.-E., L. Poellinger, J. Gustafsson, and J. Carlstedt-Duke. 1992. The glucocorticoid receptor binds to a sequence overlapping the TATA box of the human osteocalcin promoter: a potential mechanism for negative regulation. *Mol. Cell. Biol.* **11:** 3379-3383.

Struhl, K. 1986. Constitutive and inducible *Saccharomyces cerevisiae* promoters: evidence for two distinct molecular mechanisms. *Mol. Cell. Biol.* **6:** 3847-3853.

Stryer, L. 1988. *Biochemistry.* W.H. Freeman and Company, New York.

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

Thompson, C. C., and S. L. McKnight. 1992. Anatomy of an enhancer. *Trends Genet.* **8:** 232-236.

Ullman, K. S., W. M. Flanagan, C. A. Edwards, and G. R. Crabtree. 1991. Activation of early gene expression in T lymphocytes by oct-1 and an inducible protein, $OAP^{40}$. *Science* **254:** 558-562.

Urban, M. B., R. Schreck, and P. A. Baeuerle. 1991. NF-κ B contacts DNA by a heterodimer of the p50 and p65 subunit. *EMBO J.* **10:** 1817-1825.

Weintraub, H., R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benezra, T. K. Blackwell, D. Turner, R. Rupp, S. Hollenberg, Y. Zhuang, and A. Lassar. 1991. The myoD gene family: nodal point during specification of the muscle cell lineage. *Science* **251:** 761-766.

White, J., C. Brou, J. Wu, Y. Lutz, V. Moncollin, and P. Chambon. 1992. The acidic transcriptional activator GAL-VP16 acts on preformed template-committed complexes. *EMBO J.* **11:** 2229-2240.

White, R. J., and S. P. Jackson. 1992. The TATA-binding protein: a central role in transcription by RNA polymerases I, II, III. *Trends Genet.* **8:** 284-288.

Winston, F., and M. Carlson. 1992. Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *Trends Genet.* **8:** 387-391.

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

Workman, J. L., I. C. A. Taylor, and R. E. Kingston. 1991. Activation domains of stably bound GAL4 derivatives alleviate repression of promoters by nucleosomes. *Cell* **64:** 533-544.

Zenzie-Gregory, B., A. O'Shea-Greenfield, and S. T. Smale. 1992. Similar mechanisms for transcription inititation mediated through TATA box or an initiator element. *J. Biol. Chem.* **267:** 2823-2830.

# Chapter 2

## Tissue-specific expression from a compound TATA-dependent and TATA-independent promoter.

(This chapter was published in Molecular and Cellular Biology **10**: 5646-5654.)

MOLECULAR AND CELLULAR BIOLOGY, Nov. 1990, p. 5646–5654
0270-7306/90/115646-09$02.00/0
Copyright © 1990, American Society for Microbiology
Vol. 10, No. 11

# Tissue-Specific Expression from a Compound TATA-Dependent and TATA-Independent Promoter

PAUL A. GARRITY AND BARBARA J. WOLD*

*Division of Biology, 156-29, California Institute of Technology, Pasadena, California 91125*

Received 7 June 1990/Accepted 7 August 1990

**We have found that the mouse metallothionein-I (MT-I) gene promoter functions in an unusual, compound manner. It directs both TATA-dependent and TATA-independent modes of transcription in vivo. The TATA-dependent message is initiated at the previously characterized +1 transcription start site and is the predominant species in most tissues. In many cell types it is metal inducible. The TATA-independent initiation sites are distributed over the 160 bp upstream of the previously characterized +1 start site, and the RNA products are present in all tissues examined. Only in testis, however, do the TATA-independent transcripts predominate, accumulating to highest levels in pachytene-stage meiotic cells and early spermatids. Unlike the TATA-dependent +1 transcript, these RNAs are not induced by metal, even in cultured cells in which the +1 species is induced. Transfection studies of site-directed mutants show that destruction of the TATA element drastically alters the ratio of the two RNA classes in cells in which the +1 transcripts normally dominates. In TATA-minus mutants, the TATA-independent RNAs become the most prevalent, although they remain refractory to metal induction. Thus, the MT-I promoter utilizes two different types of core promoter function within a single cell population. The two different types of core promoters respond very differently to environmental stimuli, and the choice between them appears to be regulated in a tissue-specific fashion.**

Two different classes of promoter structure and function have been described for RNA polymerase II transcription in eucaryotes. The first class includes the majority of genes and is characterized by the presence of a TATA element which specifies the site of transcription initiation, usually about 30 nucleotides downstream. Extensive genetic and biochemical evidence has shown that the TATA element plays a significant role in start site selection and in specifying the frequency of transcriptional initiation (3, 5, 12, 15, 17, 19, 22, 25, 40). By itself, a TATA element is sufficient to stimulate specific transcription in vitro, though it does so with low efficiency in vivo in the absence of auxilliary elements (32, 45). Based on these properties, the TATA element can reasonably be considered to constitute a type of core promoter. The activity of such a core promoter is dependent in vivo on interactions with other protein-DNA complexes. These modulators include transcriptional enhancers, which can act at considerable distances from the core promoter, as well as more proximal upstream elements that are less flexible than enhancers in their orientation and distance requirements (38). In the current view, enhancers and upstream elements mediate environmental and developmental regulation of transcription initiation by positively or negatively modulating the activity of the core promoter complex (25, 38).

A second class of core promoters lacks a detectable TATA element. Some of these initiate transcription at a single position (50), while others initiate transcription at multiple sites over an extended region (16, 26, 31, 35, 37, 44, 47). The nature and extent of mechanistic differences between TATA-dependent and TATA-independent modes of transcription initiation are not presently known. In genes described to date, a promoter has been considered either TATA dependent or TATA independent, with developmental regulation determined by the action of separable positive or negative modulators on this core promoter structure (or in some cases more than one core promoter of the same type, e.g., two TATA elements [51]). Here we describe an interesting departure from this view of fundamental promoter classes and their developmental regulation based on the study of promoter structure and function for the mouse metallothionein-I (MT-I) gene.

The MT-I gene is constitutively expressed in many tissues and can be induced to higher levels by exposure to heavy metals (13). This induction has a major transcriptional component (13) and can be attributed in large part to the positive action of upstream metal-regulatory elements (MREs) on the TATA-dependent core promoter (7, 49, 52). Here, closer examination shows that there is also a non-metal-inducible population of MT-I RNAs that are initiated upstream of the TATA element and, at least in cultured fibroblasts, are TATA independent. Thus, the MT-I promoter simultaneously uses both TATA-dependent and TATA-independent transcription regimes which are divergent in their response to environmental stimuli. Moreover, there appears to be profound tissue-specific regulation in the choice of core promoter mechanism. Upstream RNAs similar to those seen in fibroblasts and many other cell types are the predominant species in the testis. Testis has previously been noted as unique among the tissues which have been studied in its regulation of MT-I RNA. Despite possessing a high basal level of MT-I RNA, it does not respond to cadmium (Cd) accumulation with a rapid increase in MT-I transcripts (13). Thus, MT-I is a naturally occurring case of compound core promoter function, and the choice between the modes of transcriptional initiation appears to be regulated in a tissue-specific fashion. Previously, this had not been considered a choice but rather an intrinsic property of gene architecture. These observations contribute to a revised view of the interaction of regulatory elements, including MREs, with the two classes of initiation complex.

* Corresponding author.

## MATERIALS AND METHODS

**Constructs and probes.** Standard procedures and manufacturer's instructions were used in constructing plasmids and synthesizing RNA transcripts. Four MT-I probes were used. The 5′-end genomic probe was generated from pSP6-MT-I (gift of P. Mueller), which contains mouse MT-I genomic sequences in the pSP64 vector. pSP6-MT-I was cut with *Bst*EII and transcribed with SP6 RNA polymerase. pMT-IΔi (gift of R. D. Palmiter) was used to create pMT-ISH and pMT-IBH. pMT-ISH was created by inserting a *Sac*I-*Hin*dIII fragment of pMT-IΔi into *Sac*I- and *Hin*dIII-cut pT3/T7-19 (Bethesda Research Laboratories, Inc.). pMT-ISH was cut with *Eco*RI and transcribed with T7 RNA polymerase to generate the full-length cDNA probe. pMT-IBH was created by inserting a *Bgl*II-*Hin*dIII fragment of pMT-IΔi into *Bam*HI- and *Hin*dIII-cut pBluescriptSK$^+$ (Stratagene). pMT-IBH was cut with *Xba*I and transcribed with T3 RNA polymerase to generate the 3′ cDNA probe. pMT-INT (gift of P. Mueller), a *Bgl*II-*Hin*PI fragment of pMT-IΔi inserted in *Bam*HI- and *Acc*I-cut pT3/T7-19, was cut with *Eco*RI and transcribed with T7 RNA polymerase to generate the internal cDNA probe. pBS$^-$ (Stratagene) was a gift of S. Tavtigian. pMTCAT (30) is a derivative of pBluescriptKS$^-$ (Stratagene), which contains the MT-I promoter *Eco*RI-*Bgl*II fragment, the chloramphenicol acetyltransferase (CAT) sequence of pSV2cat, and the simian virus 40 splicing and poly(A) addition signals. Oligonucleotide-directed mutagenesis was used to create pMCSPA (gift of L. J. Maher and S. J. Salser), the Sp1 site mutant derivative of pMTCAT, and pMCT, the TATA mutant derivative of pMTCAT. The substitution of TCGAGA for the wild-type TATAAA sequence in pMCT was anticipated to disable TATA element function. The double mutant TAGAGA has previously been shown, in studies in which the TATA-binding transcription factor TFIID from *Saccharomyces cerevisiae* was used to complement a mammalian transcriptional extract, to drastically reduce transcription from the adenovirus type 2 major late promoter in vitro (6). Furthermore, the binding affinity of yeast TFIID for the double mutant adenovirus major late promoter is similar to its affinity for nonspecific DNA (21).

**Cell culture and induction.** P cells (derived from L cells as described in reference 29) were maintained in dialyzed calf serum (39). Eight hours prior to harvesting for RNA, fresh medium with or without 6 μM $CdSO_4$ was added. Cells for transfection were plated at $3.7 \times 10^5$ per 10-cm dish and transfected by a standard calcium phosphate coprecipitation 27 h later. Each plate received 10 μg of test plasmid (pMTCAT, pMCT, pMCSPA, or pBS$^-$), 10 μg of L cell DNA, and 1 μg of pY3, which confers resistance to hygromycin B. After 11 h of DNA exposure, cells were refed with fresh medium. After 24 h, cells were refed with fresh medium containing 200 μg of hygromycin B per ml. The cells were refed every 3 days with fresh selection medium until they had spent 17 days in selection medium. More than 90 separate colonies were obtained on each plate. Each test plasmid was used to transfect three plates, and the pools obtained were maintained and analyzed separately. No viable colonies were recovered from a control plate which had received no DNA. Cells were split three times prior to RNA analysis and were maintained in selection medium. Eight hours prior to harvesting for RNA, fresh hygromycin-free medium with or without 6 μM $CdSO_4$ was added.

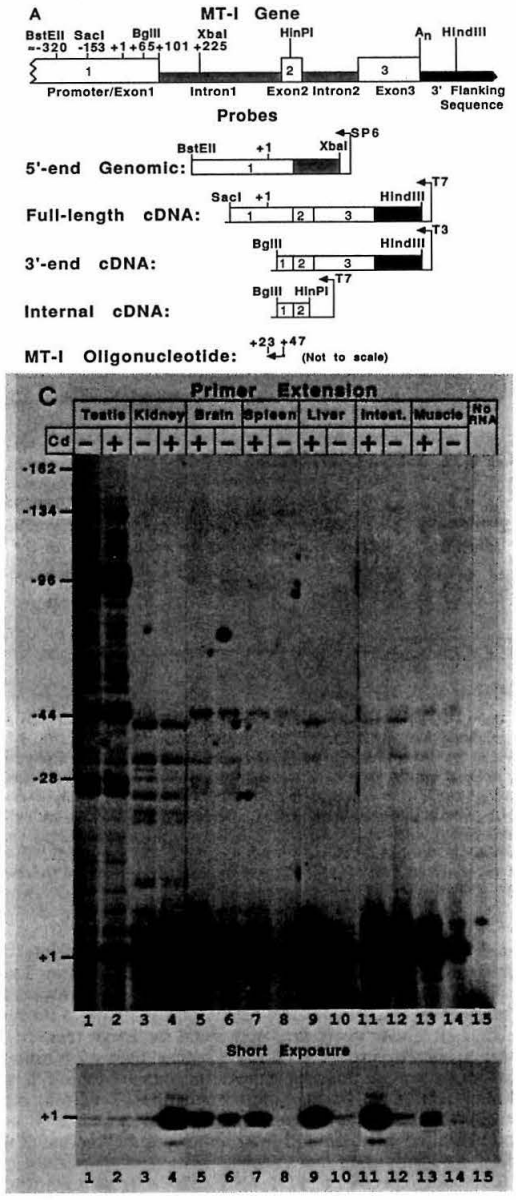
**Animals and induction.** Male mice (8 to 15 weeks old) of strains DBA/2J, C57BL/6J (both obtained from Jackson Laboratories, Bar Harbor, Maine) and $BDF_1$ (offspring of DBA × C57BL matings) were used. No strain or age variation in MT RNA expression was noted, and 15-week-old DBA/2J mice were used for quantitation. Mice were injected subcutaneously with 200 μl of $H_2O$ (or 0.9% NaCl) with or without $CdSO_4$ (10 mg of Cd per kg of mouse body weight) 3 h before the mice were sacrificed by cervical dislocation. Tissues were flash-frozen in liquid nitrogen and stored at −80°C until RNA was isolated.

Pretreatment of mice with Cd was as described in reference 42, using $CdSO_4$ instead of $CdCl_2$. Bodies, testes, and kidneys were weighed after the mice were sacrificed. No body or organ weight differences were noted between mice pretreated with daily injections of saline ($n = 5$) and those receiving saline with $CdSO_4$ (0.25 mg of Cd per kg) ($n = 6$) after 12 days. The testes of mice given a single injection of 1 mg of Cd per kg after pretreatment with saline ($n = 6$) and examined 13 days later weighed approximately half as much as those from mice injected with saline after the saline pretreatment ($n = 5$). There was no effect on body or kidney weight. Mice pretreated with $CdSO_4$ showed no such decrease in testicle weight ($n = 4$).

**RNA preparation.** RNA was prepared from L cell derivatives (P cells) as described before (11, 39). RNA was prepared from mice as described in reference 11, with the addition of a second organic extraction. Purified testis cell type and mutant mouse RNAs were provided by Kelly Thomas and Mel Simon (sample set 1 [reference 53]) and by Debra Wolgemuth (sample set 2 [reference 54]). These RNAs were prepared by LiCl precipitation (9). RNA concentrations were determined by the $A_{260}$. DNase treatment of RNAs did not alter the results.

**RNase protection.** RNase protection assays were performed as described before (29, 56) with the following modifications. RNA probes were isolated prior to use from either *N,N′*-bis-acrylylcystamine acrylamide (Bio-Rad Laboratories) or conventional acrylamide gels. The type III preparation of small yeast RNAs (Sigma catalog no. R-7125) was used as carrier RNA and in lanes labeled tRNA control. Total MT-I RNA levels were determined by using the 5′ genomic probe and the internal cDNA probe. Protected bands were excised from the gel, and the radioactivity was counted. Counting efficiency was determined by counting gel-isolated dilutions of probe. Transcript abundances were also determined from autoradiographs of 5′ genomic probe RNase protection experiments and MT-I oligonucleotide primer extensions (see below) with an LKB XL densitometer. The densitometric results were generally consistent with the results from scintillation counting.

**Primer extensions.** Primer extensions were performed with avian myeloblastosis virus reverse transcriptase (Life Sciences) (27). Reverse transcription of endogenous MT-I RNA was primed with an oligonucleotide (5′-CGGAGTAAGT GAGGAGAAGGTACTC-3′) complementary to nucleotides +45 through +23. A 0.1-pmol amount of gel-isolated, 5′-end-labeled primer was used in each reaction mixture. Signal linearity was confirmed by analyzing dilutions of a single RNA sample. The results of RNase protection experiments carried out on all samples in Fig. 1C (Fig. 1B and data not shown) indicate that the primer extension products in Fig. 1C accurately reflect MT-I RNA termini. Primer extension products were quantitated by densitometry as above. Reverse transcription of MT-CAT fusion RNAs was primed with an oligonucleotide (5′-ATATCAACGGTGGTATATCC AGTGA-3′) complementary to CAT sequences. It lies 143 to



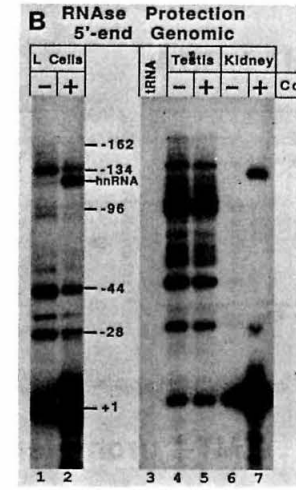




FIG. 1. (A) Structure of the mouse MT-I gene and the antisense probes used in RNase protection and primer extension assays. Sequence is from reference 18, with an additional A at −112 (39). Unshaded boxes, Exon and promoter sequences; stippled boxes, introns; black boxes, sequences downstream of 3′ cleavage and polyadenylation site ($A_n$); thin horizontal lines, vector sequences. Restriction sites are numbered according to complementary nucleotides retained in derived probes. (B) RNase protection analysis performed with 5′-end genomic probe. Sample RNA (8 μg) plus tRNA (30 μg) were used in lanes 1, 2, and 4 to 7; 30 μg of tRNA alone was used in lane 3. Products were initially sized by using DNA markers and subsequently assigned to specific start sites by comparison with primer extension results. Longer exposures show upstream RNAs present in the kidney. The smear near −28 in lane 7 is artifactual. Please note that because RNase protection probes are continuously labeled, the specific activity (disintegrations per minute per mole) of each protected product increases with product size. hnRNA, heterogeneous nuclear RNA (unspliced +1 RNA). (C) Primer extensions performed with the MT-I oligonucleotide. An 8-μg amount of RNA was used in each lane (except lane 15, to which no RNA was added). Product identities were determined by comparison with adjacent lanes generated by dideoxy sequencing of MT-I with the same oligonucleotide. Subsequent analysis of multiple independent preparations of RNA indicated that the apparent metal induction of upstream RNA in liver shown in this figure was anomalous and that upstream RNAs in liver were actually slightly repressed by metal. Intest., Intestine.

119 bp downstream of the MT-I +1 site in the pMTCAT plasmid.

## RESULTS

**Expression pattern of the multiple species of MT-I RNA.** MT-I RNA was examined in tissues from DBA/2J, C57BL/6J, and $BDF_1$ mice, in NIH 3T3 cells, and in an L cell derivative (P cells [29]) by RNase protection (Fig. 1B and data not shown) and primer extension (Fig. 1C). The profile of MT-I RNA in testis (Fig. 1B, lanes 3 and 4, and Fig. 1C, lanes 1 and 2) was dramatically different from that found elsewhere. The expected transcript, with its 5′ end at the previously described +1 site (18), constituted only 15% of the total MT-I RNA in testis. The remainder were transcripts that extended well upstream. (To facilitate quantitation, we have operationally grouped the many species of transcripts into six major groups, as indicated in Fig. 2B. Total MT-I and individual start site RNA data are presented in the Fig. 2A histograms.) Although this 5′ heterogeneity was most striking in the testis, close inspection of both primer extension and RNase protection experiments showed that all sources tested possessed a complex array of MT-I 5′ ends. For example, L cells, while producing high levels of the +1

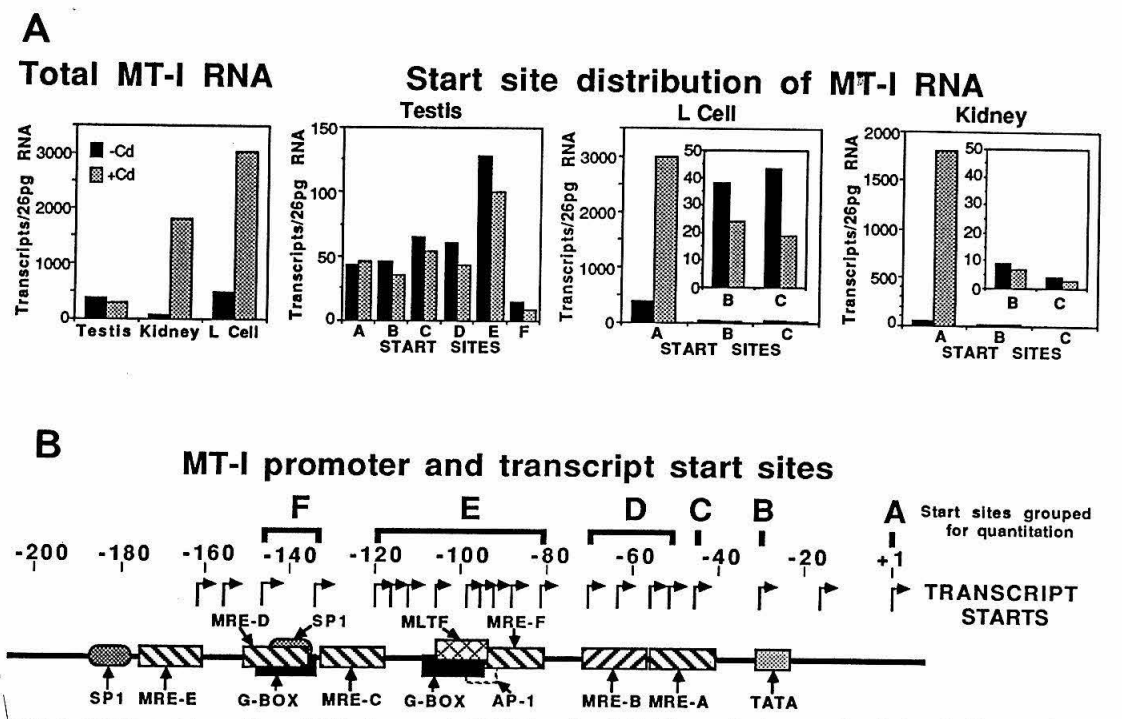



FIG. 2. (A) Transcripts per 26 pg of RNA, the amount of RNA in an L cell (4). The quantitation procedure is described in Materials and Methods. Transcriptional start sites were grouped into six clusters to ease analysis (see panel B). RNA termination sites by group: A, +3 to −1; B, −25 to −35; C, −39 to −45; D, −51 to −75; E, −82 to −119; F, −133 to −146. (B) Map of MT-I promoter, with arrows indicating the position of each discrete start site (judged by its appearance as a unique entity in both RNase protection and primer extension experiments) and with the groupings used in quantitation indicated. Groupings reflect the smallest regions which could be reliably excised from a gel and/or densitometrically integrated. Boxes indicate potential DNA-protein interaction sites (39). The MREs show metal-inducible dimethyl sulfate (DMS) protection in L cells in vivo. The distal Sp1 site binds Sp1 in vitro and shows an Sp1-like footprint in L cells in vivo. The proximal Sp1 site binds Sp1 weakly in vitro but does not shown an Sp1-like footprint in L cells in vivo. The MLTF site shows MLTF-like DMS protection in L cells in vivo. The G element (G-BOX) is a conserved sequence feature of the MT gene family with no known function. The potential AP-1 site (1) has not been shown to interact with this factor.

transcript, expressed essentially the same set of upstream RNAs as the testis, although many of these species (e.g., −96) were expressed at much reduced levels (Fig. 1B, lanes 1 through 5, and Fig. 2), while kidney expressed lower levels of upstream RNA than either testis or L cells and used a different set of upstream initiation sites (Fig. 1B, lanes 6 and 7; Fig. 1C, lanes 3 and 4; and Fig. 2). Inspection of the MT-I promoter sequence revealed no other TATA homologies in this region and no concordance between the upstream start sites and potential "initiator" elements of the type shown to fix the start site in the TATA-less mouse terminal deoxynucleotidyltransferase gene (50). The MT-I upstream DNA also lacked any identifiable HIP1-binding sequences, which direct start site selection at one of the several start sites in the mouse dihydrofolate reductase (*dhfr*) gene (33).

Cd administration induced only the +1 RNA species. The +1 RNA levels increased by at least 40% and commonly over 25-fold in all cells and tissues surveyed except testis (Fig. 1C, short exposure), in agreement with previous determinations of total MT-I RNA content (13). Steady-state levels of the upstream MT-I RNAs were relatively unaffected by Cd. The greatest change in the upstream RNAs was detected in L cells, where the longer transcripts decreased 40 to 60% after metal treatment (Fig. 1B, lanes 1 and 2; Fig. 2); testis and kidney showed smaller but reproducible diminutions (Fig. 1B, lanes 4 and 5; Fig. 1C, lanes 1 through 4; Fig. 2). These upstream, non-metal-inducible transcripts were a specific property of MT-I rather than a ubiquitous feature of metallothionein genes. In contrast to MT-I, the closely related mouse metallothionein-II (MT-II) (48) transcripts showed little upstream heterogeneity at the 5′ end and exhibited metal induction behavior paralleling that of the MT-I +1 RNA (data not shown).

**The many species of MT-I RNA differ only at their site of initiation.** We examined MT-I RNA from various sources to see whether the upstream initiation sites produced RNAs substantially different from the traditional +1 site RNAs. RNA gel blots, containing samples from testis and kidney, were probed with an oligonucleotide specific for upstream RNA and with an oligonucleotide which would hybridize to both upstream and +1 RNAs. In each experiment only one general size species of MT-I RNA was detected (data not shown), indicating no gross structural differences in the body of the RNA. More sensitive assays for structural heteroge-

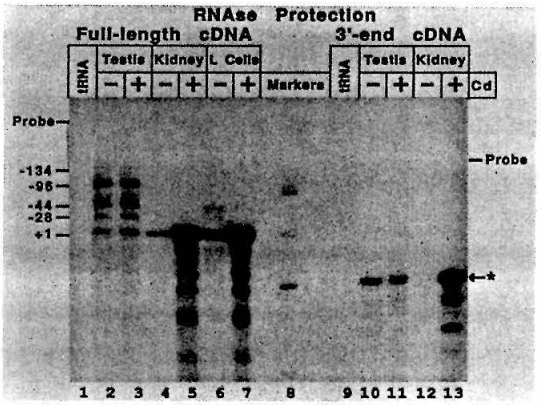



FIG. 3. RNase protection assays performed with the full-length cDNA probe (lanes 1 to 7) and 3'-end cDNA probe (lanes 9 to 13). Sample RNA (3.4 μg) plus tRNA (30 μg) was used in lanes 2 to 7 and lanes 10 to 13; 30 μg of tRNA alone was used in lanes 1 and 9. Products were sized by using *Hpa*II-cut pBR322 and pUC19 (lane 8). Products were assigned to specific start sites based on comparison with Fig. 1. Longer exposures showed upstream RNAs present in the kidney. Note that the complementarity of the full-length cDNA probe extended only to −153. The asterisk denotes the predicted protection product of the 3'-end cDNA probe.

neity were also performed. 3'-End cDNA and full-length cDNA probes (Fig. 1A) were used in RNase protection assays. The 3'-end cDNA probe (Fig. 3, lanes 9 to 13) required appropriate splicing and 3' cleavage of the RNA to yield the product indicated by the asterisk in Fig. 3. The amounts of this product in testis and kidney were consistent with our total MT-I RNA measurements and so represented the vast majority of MT-I RNA species. Furthermore, the full-length cDNA probe (Fig. 3, lanes 1 to 7), which is sensitive to the transcription initiation site as well as splicing and 3' cleavage, recapitulated the start site distribution seen in RNase protection with the 5'-end genomic probe (Fig. 1B, lanes 1 to 7) and in primer extension with the MT-I oligonucleotide (Fig. 1C, lanes 1 to 4). These data, combined with previous work establishing that all MT-I RNA in testis is polyadenylated (55), show that, within our limits of detection, all species of MT-I transcripts are spliced and 3'-end processed identically and therefore differ from one another only at their 5' ends.

On the basis of the data presented thus far, it was formally possible that the upstream RNAs could be hitherto-undetected precursors of the +1 species that accumulate to higher steady-state levels in testis. Such a pattern of 5'-end processing would be unprecedented, however, and appears unlikely to be the case for two reasons. First, the longer RNAs persisted in L cells over the course of a 6-h dactinomycin (actinomycin D) transcriptional block instead of being chased into the +1 form (or any other detectable form) by subsequent processing (data not shown). Also, as detailed below, site-directed mutation of the MT-I TATA element virtually eliminated the +1 product while leaving upstream RNAs intact, an observation inconsistent with a precursor-product relationship.

**Testis sensitivity to Cd can be altered without changing MT RNA behavior.** Metallothionein proteins bind heavy metals (like Cd) and can protect cells from their potentially toxic effects (23). Compared with other tissues, testis shows an unusual sensitivity to Cd, suffering widespread and irreversible damage at Cd doses that have no other major effect on the animal (34, 43). However, treatment of susceptible strains with low doses of Cd protects these mice from testicular damage by a subsequent high dose of Cd (42). We reexamined these phenomena in the light of our present knowledge of MT RNA behavior in testis. It is possible that testis hypersensitivity might be due to the noninducibility of MT-I and MT-II RNAs in testis and that this noninducibility might be related to the unusual behavior of MT-I RNA there. A mouse strain reported to be "less susceptible" to testicular damage by Cd (20), C57BL/6J, was examined and showed no difference in basal or induced expression of MT-I and MT-II (data not shown) compared with the susceptible strain DBA/2J. A low-level Cd pretreatment regime did protect DBA/2J mice from Cd-induced testis damage but had no discernible effect on the basal level of MT-I RNA or MT-II RNA in testis, nor did it render these genes metal inducible in this tissue (data not shown). Thus, it is possible to change Cd sensitivity without modifying MT RNA expression. Although the failure of MT RNA to be induced by metal in testis may contribute to the organ's susceptibility to Cd poisoning, additional mechanisms are involved in generating resistance by the Cd pretreatment regime.

**Upstream MT-I RNAs are the predominant species in germ cells of the testis.** The testis is a complex organ containing specialized somatic accessory cells, such as Sertoli and Leydig cells, along with premeiotic germ cells, cells in various stages of meiosis, and maturing sperm. To find out which of these very different cell types are responsible for the predominance of upstream MT-I transcripts in total testis, RNAs from fractionated cell populations from the testes of mice not exposed to metal were first examined. The upstream RNAs were most prominent in late meiotic prophase (pachytene-stage cells) and remained prominent in the early stages of sperm maturation (round spermatids) (Fig. 4, lanes 3 to 7). Thus, in germ cells, the level of upstream RNA per unit mass of total RNA increased late in meiosis and then declined in the residual bodies and cytoplasmic fragments cast off by condensing spermatids (Fig. 4, lanes 1 to 7). Conversely, the level of the +1 RNA gradually declined as the germ cell population proceeded from spermatogonia to round spermatids and was even lower in the residual bodies and cytoplasmic fragments.

Another method, independent of cell separation techniques, can distinguish expression in germ cells from expression in accessory cells. Mutant mice of the W/W$^v$ genotype lack germ cells but possess all the somatic cell types of the testis (36, 41, 55). The testes of W/W$^v$ mice were found to contain substantially lower levels of upstream RNA than the phenotypically wild-type testes of +/W$^v$ mice (Fig. 4, lanes 9 and 10). This argues that high-level upstream MT-I transcript expression is limited to germ cells.

Can the different species of MT-I RNA encode variant proteins? Inspection of the 5' nucleotides of the longer transcripts indicated that they contained an open reading frame that could code for a 22-amino-acid protein (showing none of the cysteine-richness of the MTs) after initiation at the −91 ATG. However, there were stop codons in this and all other reading frames between −4 and +40, upstream of the previously recognized initiation codon. It therefore appears unlikely that the extra nucleotides contribute any information toward a larger than usual MT protein. Possible effects of the longer 5' leader on RNA processing efficiency,

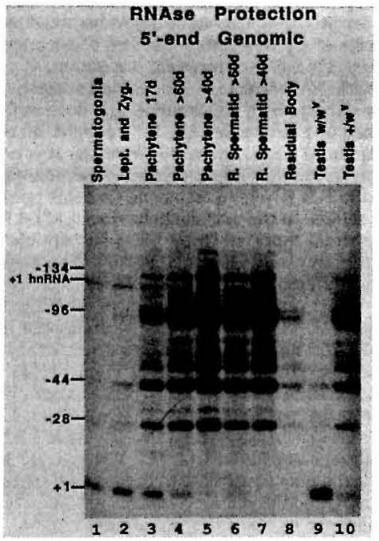



FIG. 4. RNase protections performed with the 5'-end genomic probe. Sample RNA (0.5 μg) plus tRNA (30 μg) was used in each lane. The tRNA lane (not shown) gave no signal. Mouse age at preparation of sample: type A and B spermatogonia, 8 days; leptotene (Lept.) and zygotene (Zyg.) cells, 17 days; pachytene and round (R.) spermatids as indicated (in days); residual body and mutant mouse testes, >40 days. Lanes 1, 2, 3, 4, and 6 are from sample set 1. Lanes 5, 7, 8, 9, and 10 are from sample set 2. The practical necessity of using mice of different ages to prepare the various cell populations did not appear to compromise our conclusions about stage specificity. The MT-I RNA profiles of pachytene cells from 17-, >60-, and >40-day-old mice were qualitatively similar, while the leptotene, zygotene, and pachytene cells from 17-day-old mice were quite different. hnRNA, heterogeneous nuclear RNA (unspliced +1 RNA).

message stability, or translational regulation remain to be examined.

**Site-directed mutation separates MT-I RNA species into TATA-dependent and TATA-independent classes.** The presence of a group of RNAs that initiate upstream of the TATA homology element and are regulated differently than the +1 RNA suggested that the MT-I promoter might be directing transcription in two different ways, one TATA dependent and the other TATA independent. We directly tested this possibility by constructing appropriate mutants of the MT-I promoter and analyzing the RNA products of each in transfection assays. Because L cells express the same overall repertoire of upstream start sites as germ cells (albeit with quantitative variations) (Fig. 1B), they were used as transfection hosts. The constructs contained MT-I sequence extending from 1.7 kb upstream to 65 bp downstream of the MT-I +1 site fused to the CAT gene (30). Site-directed mutagenesis was used to disrupt the TATA element by substitution of the sequence TCGAGA for the wild-type TATAAA. A construct containing mutations in the distal Sp1 site (GGGGCGG changed to GCCCGGG) was also created. This Sp1 site binds Sp1 in vitro, is occupied in L cells in vivo (39), and is of interest because Sp1 activates TATA-independent transcription from the *dhfr* promoter in vitro (14).

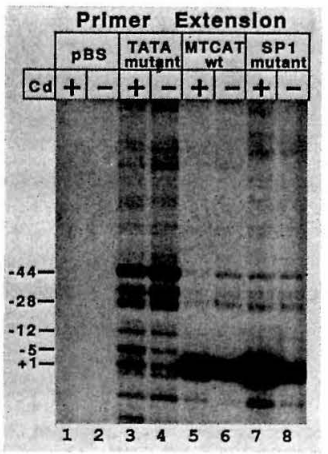



FIG. 5. Primer extensions performed with the transgene-specific MT-CAT oligonucleotide. A 9-μg amount of RNA from each transfected cell line was used per lane. Overexposure of the autoradiograph necessary to see upstream RNA and mutant +1 RNA caused the induction of the +1 RNA in the wild-type (wt) and Sp1 mutant lanes to be underrepresented. Product identities were assigned by comparison with adjacent lanes generated by dideoxy sequencing of pMCT with the MT-CAT oligonucleotide.

Polyclonal pools of stably transfected L cells containing each construct (and controls containing the plasmid pBS$^-$) were made. Primer extension analysis was performed with a transgene-specific CAT primer on total RNA from three independent transformant pools for each construct (Fig. 5 and data not shown). As expected, the wild-type MT-I-promoted transgene produced hybrid MT-CAT RNAs which paralleled the endogenous MT-I RNAs in 5'-end heterogeneity and +1 RNA predominance (Fig. 5, lanes 5 and 6). The +1 RNA species was induced by metal, and the quantitatively minor upstream RNAs were slightly repressed. In sharp contrast, the TATA mutant transgene produced predominantly the upstream RNAs (Fig. 5, lanes 3 and 4). Some RNA still arose from the +1 region, but it exhibited 5'-end heterogeneity, a common result of eliminating the TATA element from a TATA-dependent promoter. These data confirm that, in L cells, the +1 RNA of the wild-type promoter is TATA dependent and show that the upstream RNAs are TATA independent. In addition to this clear qualitative change in RNA expression pattern, mutation of the TATA element led to an apparent increase in the absolute amount of upstream RNA produced in each transfected pool. This may imply that the TATA-dependent and TATA-independent transcriptional modes compete with each other and that ablation of the TATA element altered that competition. However, such quantitative differences between stably transfected polyclonal cell pools are sometimes seen even when separate pools that contain the same construct are compared. At present we favor the former explanation but cannot rule out the latter.

The central result of these experiments is that, in L cells, the upstream RNAs are independent of the MT-I TATA element, while the +1 start is TATA dependent. Until such mutations are also analyzed in transgenic mice, we cannot be certain that this is true in other cell types, although it seems likely that it will be. Subtle results of the TATA mutation

were also observed. The first concerns the upstream RNAs of the TATA mutant, which, like those of the parental gene, were slightly repressed by metal. In contrast, the residual +1 region RNAs of the TATA mutant retained some metal inducibility (Fig. 5, lanes 3 and 4, and data not shown), but the effect was small compared with the degree of metal induction seen in the wild-type construct. Another question was whether the TATA mutation alone would switch the L cell pattern to a precise copy of the germ cell pattern. It did not do so. Given the differences already observed in the details of upstream RNA populations in different tissues, it seemed unlikely that TATA dependence alone dictates the pattern of upstream start site utilization. Close inspection of the TATA mutant RNAs in L cells supported this view; start sites adjacent to +1 (e.g., −5 and −12) were much more pronounced than in germ tissue, and the shorter upstream RNAs (e.g., −28 and −44) were much more prevalent than the longer upstream RNAs, the usual L cell pattern. Since the TATA mutation does not result in a complete recapitulation of the germ cell RNA pattern in L cells, the unique germ cell pattern must be created by more than an inability of the TATA element to function. Nonetheless, it is likely that the upstream RNAs in germ cells will share the independence of the TATA element of their L cell counterparts.

Mutation of the distal Sp1 site did not have a major impact on RNA production (Fig. 5, lanes 7 and 8). Therefore, in L cells, this element is not essential for basal expression of either the +1 or the upstream MT-I RNAs and is not required for metal induction of the MT-I +1 RNA. In the transfected pools, expression from the endogenous MT-I gene was not affected by the presence of the various transgenes (data not shown), eliminating the possibility that the behavior of the introduced genes could be due to a promoter titration effect.

## DISCUSSION

We have shown that the mouse MT-I promoter directs RNA transcription via both TATA-dependent and TATA-independent mechanisms, that it functions in both ways even in apparently homogeneous populations of cells, and that the choice between modes is developmentally regulated. This type of compound promoter behavior has not, to our knowledge, been recognized previously. Whether the two modes coexist temporally within a single cell or even at a single promoter is an interesting but unanswered question. A different type of compound promoter has been described for *S. cerevisiae* (2, 24, 51). Those promoters are compound in the sense that they possess multiple, functionally distinct TATA elements and therefore operate via multiple TATA-dependent mechanisms.

We were led to this discovery by the observation that the MT-I promoter initiates transcription at many sites and thus produces multiple species of MT-I transcript. The well-characterized +1 transcript (group A in Fig. 2) is initiated 23 nucleotides downstream of the TATA element and is TATA dependent. The newly identified MT-I RNA species (groups B through F in Fig. 2) are initiated at a multitude of sites upstream of the +1 site (most of these are initiated upstream of the TATA element as well) and are, in contrast, TATA independent. Both classes of RNA are properly spliced and polyadenylated. These characteristics strongly argue that both upstream and +1 species are RNA polymerase II products, although this has not yet been determined directly.

The RNAs produced from the two classes of transcription have distinct patterns of tissue-specific expression, implying significant developmental regulation of the choice between TATA-independent and TATA-dependent transcription. While the +1 transcript constitutes the bulk of MT-I RNA in most cell types, in the late meiotic and postmeiotic cells of the testis the TATA-independent transcripts are the predominant MT-I RNA species. In addition, cell types differ in the specific sets of TATA-independent RNAs that they contain, suggesting an additional level of tissue specificity beyond the choice of TATA-independent versus TATA-dependent initiation. For example, approximately 60% of the TATA-independent RNAs in testis are group D, E, and F RNAs (Fig. 2), but these RNAs contribute only a minor fraction of TATA-independent RNAs in kidney and L cells. The precise locations of the start sites differ as well and fall into two patterns, one epitomized by testis (Fig. 1C, lanes 1 and 2, and data not shown) and the other by kidney (Fig. 1C, lanes 3 and 4, and data not shown).

The TATA-dependent and TATA-independent classes of RNA are also regulated differently by metal, which is known to increase the overall rate of transcription of the MT-I gene in many cell types (13). Thus, the choice of core promoter mechanism appears to be intimately linked with the gene's metal inducibility. The +1 RNA is metal inducible in all cellular contexts examined except for the testis, which does not exhibit significant metal induction of any MT-I or MT-II RNAs. The TATA-independent RNAs are not induced by metal in any cellular context tested, nor do they become metal responsive in the case of a TATA-minus mutant assayed in cells in which the wild type MT-I gene is metal responsive. We favor the view that these differences in expression of MT-I RNAs are generated mainly at the transcriptional level. However, in the absence of direct measurement of transcription initiation, we cannot exclude the possibility that any of these observations are substantially due to posttranscriptional effects. For example, it is formally possible that preferential stabilization of the upstream RNAs in testis combined with efficient degradation of +1 RNAs in the same cells could give rise to the observed distribution of transcripts. Conversely, preferential destabilization of upstream RNAs in response to metal could be invoked to explain the metal-refractory nature of the upstream RNAs. In the latter case, our studies of MT-I fusion genes show that such differential posttranscriptional effects would have to rely on nucleotides in the −1 to −12 region. Furthermore, these sequences would have to perform any posttranscriptional, metal-responsive function independently of the MT-I sequences replaced by CAT in the transgenes. Such a scenario seems unduly perverse.

The extent of interrelation between TATA-dependent and TATA-independent transcriptional modes is an issue raised but not resolved by our observations. Closely related DNA-protein complexes could activate both transcriptional modes, with a subset of components determining the balance and separate characteristics of TATA dependence and TATA independence. Alternatively, mutually exclusive complexes containing entirely different accessory factors could be used. The ability of promoters which normally appear to support little or no TATA-independent transcription to function after TATA element removal (3, 5, 12, 17, 19) and the ability of Sp1 to stimulate both TATA-dependent and TATA-independent transcription (14) argue that the switch to TATA-independent transcription may not require wholesale remodeling of the transcription complex.

It is interesting to consider this new information about MT-I promoter function in the context of what is known of its structure. The MT-I proximal regulatory region contains

a number of *cis*-acting elements identified either genetically or by homology to known transcription factor-binding sites (Fig. 2B). These include several metal-responsive enhancer elements (MREs) (7, 49, 52) and potential binding sites for general transcription factors such as MLTF (8, 10, 46) and Sp1 (28). In vivo footprinting in L cells has detected DNA-protein interactions at many of these sites, with only the MRE interactions dependent on metal treatment (39) (Fig. 2B legend). The relevance of this current picture of MT-I DNA-transcription factor interactions to the TATA-independent mode of transcription is unknown, but it is of interest that the MREs directly overlap a number of the upstream start sites. Metal-dependent occupancy of these sites activates transcription only from the +1 region, suggesting that they are either inappropriately positioned to activate upstream RNA synthesis or incompatible with the upstream initiation complexes. The MREs could also be responsible for the observed metal-induced decrease in upstream RNA in L cells. Factors bound to the MREs could suppress the use of the upstream start sites directly, by steric occlusion, or indirectly, by activation of the TATA-dependent transcriptional mode at the expense of the TATA-independent mode.

Further investigation of these phenomena is likely to be difficult using L cells alone, considering the small amount of upstream RNA that they contain. However, an in vivo footprint study of the MT-I gene in the pachytene cells of the testis together with analysis of site-directed mutants in the testes and other tissues of transgenic mice may provide further insight into regulation of this compound promoter. Ultimately, the MT-I compound promoter should serve as a model system for in vitro reconstruction of promoter choice and for characterization of the specific protein interactions that account for the strongly preferential interaction of the metal response with TATA-dependent initiation.

## ACKNOWLEDGMENTS

We thank Kelly Thomas, Mel Simon, and Debra Wolgemuth for generous gifts of purified testis cell-type RNAs and mutant RNAs, Richard Palmiter for pMT-IΔi, Jim Maher for assistance with mutagenesis, Sean Tavtigian for his initial observations, Tom Novak for the BB62, Paul Mueller, Jeff Miner, and members of the Wold lab for materials and invaluable discussion, and David Anderson, Richard Axel, and Mel Simon for reading the manuscript.

This work was supported by grants from the National Institutes of Health, Rita Allen Foundation, and Lucille P. Markey Charitable Trust to B.W. and a National Research Service Award (5 T32 GM 07616-11) from the National Institute of General Medical Sciences to P.G.

## LITERATURE CITED

1. **Angel, P., M. Imagawa, R. Chiu, B. Stein, R. J. Imbra, H. J. Rahmsdorf, C. Jonat, P. Herrlich, and M. Karin.** 1987. Phorbol ester-inducible genes contain a common *cis* element recognized by a TPA-modulated *trans*-acting factor. Cell **49**:729–739.
2. **Arndt, K. T., C. Styles, and G. R. Fink.** 1987. Multiple global regulators control *HIS4* transcription in yeast. Science **237**:874–880.
3. **Benoist, C., and P. Chambon.** 1981. In vivo sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–315.
4. **Brandhorst, B. P., and E. H. McConkey.** 1974. Stability of nuclear RNA in mammalian cells. J. Mol. Biol. **85**:451–463.
5. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
6. **Buratowski, S., S. Hahn, P. A. Sharp, and L. Guarente.** 1988. Function of a yeast TATA element-binding protein in a mammalian transcription system. Nature (London) **334**:37–42.
7. **Carter, A. D., B. K. Felber, M. Walling, M.-F. Jubier, C. J. Schmidt, and D. H. Hamer.** 1984. Duplicated heavy metal control sequences of the mouse metallothionein-I gene. Proc. Natl. Acad. Sci. USA **81**:7392–7396.
8. **Carthew, R. W., L. A. Chodosh, and P. A. Sharp.** 1985. An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell **43**:439–448.
9. **Cathala, G., J.-F. Savouret, B. Mendez, B. L. West, M. Karin, J. A. Martial, and J. D. Baxter.** 1983. Laboratory methods—a method for isolation of intact, translationally active ribonucleic acid. DNA **2**:329–335.
10. **Chodosh, L. A., R. W. Carthew, and P. A. Sharp.** 1986. A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. Mol. Cell. Biol. **6**:4723–4733.
11. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**:156–159.
12. **Dierks, P., A. van Ooyen, M. D. Cochran, C. Dobkin, J. Reiser, and C. Weissman.** 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.
13. **Durnam, D. M., and R. D. Palmiter.** 1981. Transcriptional regulation of the mouse metallothionein-I gene by heavy metals. J. Biol. Chem. **256**:5712–5716.
14. **Dynan, W. S., S. Sazer, R. Tjian, and R. T. Schimke.** 1986. Transcription factor Sp1 recognises a DNA sequence in the mouse dihydrofolate reductase promoter. Nature (London) **319**: 246–248.
15. **Eisenmann, D. M., C. Dollard, and F. Winston.** 1989. *SPT15*, the gene encoding the yeast TATA binding factor TFIID, is required for normal transcription in yeast. Cell **58**:1183–1191.
16. **Farnham, P. J., and R. T. Schimke.** 1986. Murine dihydrofolate reductase transcripts through the cell cycle. Mol. Cell. Biol. **6**:365–371.
17. **Ghosh, P. K., P. Lebowitz, R. J. Frisque, and Y. Gluzman.** 1981. Identification of a promoter component involved in positioning the 5′ termini of simian virus 40 early mRNAs. Proc. Natl. Acad. Sci. USA **78**:100–104.
18. **Glanville, N., D. M. Durnam, and R. D. Palmiter.** 1981. Structure of mouse metallothionein-I gene and its mRNA. Nature (London) **292**:267–269.
19. **Grosschedl, R., and M. L. Birnstiel.** 1980. Identification of regulatory sequences in the prelude sequences of an HSA histone gene by the study of specific deletion mutants in vivo. Proc. Natl. Acad. Sci. USA **77**:1432–1436.
20. **Gunn, S. A., T. C. Gould, and W. A. D. Anderson.** 1965. Strain differences in susceptibility of mice and rats to cadmium-induced testicular damage. J. Reprod. Fertil. **10**:273–275.
21. **Hahn, S., S. Buratowski, P. A. Sharp, and L. Guarente.** 1989. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. Proc. Natl. Acad. Sci. USA **86**:5718–5722.
22. **Hahn, S., S. Buratowski, P. A. Sharp, and L. Guarente.** 1989. Isolation of the gene encoding the yeast TATA-binding protein TFIID: a gene identical to the *SPT15* suppressor of Ty element insertions. Cell **58**:1173–1181.
23. **Hamer, D. H.** 1986. Metallothionein. Annu. Rev. Biochem. **55**:913–951.
24. **Harbury, P. A. B., and K. Struhl.** 1989. Functional distinctions between yeast TATA elements. Mol. Cell. Biol. **9**:5298–5304.
25. **Horikoshi, M., C. K. Wang, H. Fujii, J. A. Cromlish, P. A. Weil, and R. G. Roeder.** 1989. Cloning and structure of a yeast gene encoding a general transcription initiation factor TFIID that binds to the TATA box. Nature (London) **341**:299–303.
26. **Ishii, S., Y. Xu, R. H. Stratton, B. A. Roe, G. T. Merlino, and I. Pastan.** 1985. Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene. Proc. Natl. Acad. Sci. USA **82**:4920–4924.
27. **Jones, K. A., K. R. Yamamoto, and R. Tjian.** 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. Cell **42**:559–572.

28. **Kadonaga, J. T., K. A. Jones, and R. Tjian.** 1986. Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends Biochem. Sci. **11**:20–23.
29. **Kim, S. K., and B. J. Wold.** 1985. Stable reduction of thymidine kinase activity in cells expressing high levels of anti-sense RNA. Cell **42**:129–138.
30. **Maher, L. J., B. Wold, and P. B. Dervan.** 1989. Inhibition of DNA-binding proteins by oligonucleotide-directed triple helix formation. Science **245**:725–730.
31. **McGrogan, M., C. C. Simonsen, D. T. Smouse, P. J. Farnham, and R. T. Schimke.** 1985. Heterogeneity at the 5′ termini of mouse dihydrofolate reductase mRNAs. J. Biol. Chem. **260**: 2307–2314.
32. **McKnight, S. L., E. R. Gavis, R. Kingsbury, and R. Axel.** 1981. Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. Cell **25**:385–398.
33. **Means, A. L., and P. J. Farnham.** 1990. Transcription initiation from the dihydrofolate reductase promoter is positioned by housekeeping initiation protein 1 binding at the initiation site. Mol. Cell. Biol. **10**:653–661.
34. **Meek, E. S.** 1959. Cellular changes induced by cadmium in mouse testis and liver. Br. J. Exp. Pathol. **40**:503–507.
35. **Melton, D. W., C. McEwan, A. B. McKie, and A. M. Reid.** 1986. Expression of the mouse HPRT gene: deletional analysis of the promoter region of an X-chromosome linked housekeeping gene. Cell **44**:319–328.
36. **Mintz, B., and E. S. Russell.** 1957. Gene-induced embryological modifications of primordial germ cells in the mouse. J. Exp. Zool. **134**:207–230.
37. **Mitchell, P. J., A. M. Carothers, J. H. Han, J. D. Harding, E. Kas, L. Venola, and L. A. Chasin.** 1986. Multiple transcription start sites, DNase I-hypersensitive sites, and an opposite-strand exon in the 5′ region of the CHO *dhfr* gene. Mol. Cell. Biol. **6**:425–440.
38. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA-binding proteins. Science **245**:371–378.
39. **Mueller, P. R., S. J. Salser, and B. Wold.** 1988. Constitutive and metal-inducible protein-DNA interactions at the mouse metallothionein I promoter examined by in vivo and in vitro footprinting. Genes Dev. **2**:412–427.
40. **Nakajima, N., M. Horikoshi, and R. G. Roeder.** 1988. Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID. Mol. Cell. Biol. **8**:4028–4040.
41. **Nocka, K., S. Majumdar, B. Chabot, P. Ray, M. Cervone, A. Bernstein, and P. Besmer.** 1989. Expression of c-*kit* gene products in known cellular targets of W mutations in normal and W mutant mice: evidence for an impaired c-*kit* kinase in mutant mice. Genes Dev. **3**:816–826.
42. **Nordberg, G. F.** 1971. Effects of acute and chronic cadmium exposure on the testicles of mice. Environ. Physiol. **1**:171–187.
43. **Parizek, J.** 1960. Sterilization of the male by cadmium salts. J. Reprod. Fertil. **1**:294–309.
44. **Reynolds, G. A., S. K. Basu, T. F. Osborne, D. J. Chin, G. Gil, M. S. Brown, J. L. Goldstein, and K. L. Luskey.** 1984. HMG CoA reductase: a negatively regulated gene with unusual promoter and 5′ untranslated regions. Cell **38**:275–285.
45. **Sassone-Corsi, P., J. Corden, and P. Chambon.** 1981. Promotion of specific in vitro transcription by excised "TATA" box sequences inserted in a foreign nucleotide environment. Nucleic Acids Res. **9**:3941–3958.
46. **Sawadogo, M., and R. G. Roeder.** 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell **43**:165–175.
47. **Sazer, S., and R. T. Schimke.** 1986. A re-examination of the 5′ termini of mouse dihydrofolate reductase RNA. J. Biol. Chem. **261**:4685–4690.
48. **Searle, P. F., B. L. Davison, G. W. Stuart, T. M. Wilkie, G. Norstedt, and R. D. Palmiter.** 1984. Regulation, linkage, and sequence of mouse metallothionein I and II genes. Mol. Cell. Biol. **4**:1221–1230.
49. **Searle, P. F., G. W. Stuart, and R. D. Palmiter.** 1985. Building a metal-responsive promoter with synthetic regulatory elements. Mol. Cell. Biol. **5**:1480–1489.
50. **Smale, S. T., and D. Baltimore.** 1989. The "initiator" as a transcription control element. Cell **57**:103–113.
51. **Struhl, K.** 1986. Constitutive and inducible *Saccharomyces cerevisiae* promoters: evidence for two distinct molecular mechanisms. Mol. Cell. Biol. **6**:3847–3853.
52. **Stuart, G. W., P. F. Searle, H. Y. Chen, R. L. Brinster, and R. D. Palmiter.** 1984. A 12-base-pair DNA motif that is repeated several times in metallothionein gene promoters confers metal regulation to a heterologous gene. Proc. Natl. Acad. Sci. USA **81**:7318–7322.
53. **Thomas, K. H., T. M. Wilkie, P. Tomashefsky, A. R. Bellve, and M. I. Simon.** 1989. Differential gene expression during mouse spermatogenesis. Biol. Reprod. **41**:729–739.
54. **Wolgemuth, D. J., E. Gizang-Ginsberg, E. Engelmyer, B. J. Gavin, and C. Ponzetto.** 1985. Separation of mouse testis cells on a Celsep™ apparatus and their usefulness as a source of high-molecular-weight DNA or RNA. Gamete Res. **12**:1–10.
55. **Zakeri, Z. F., and D. J. Wolgemuth.** 1987. Development-stage-specific expression of the hsp70 gene family during differentiation of the mammalian male germ line. Mol. Cell. Biol. **7**:1791–1796.
56. **Zinn, K. A., D. DiMaio, and T. Maniatis.** 1983. Identification of two distinct regulatory regions adjacent to the human β-interferon gene. Cell **34**:865–879.

# Chapter 3

# Effects of different DNA polymerases in ligation-mediated PCR: Enhanced genomic sequencing and *in vivo* footprinting.

(This chapter was published in *Proc. Natl. Acad. Sci. USA* **89**: 1021-1025.)

*Proc. Natl. Acad. Sci. USA*
Vol. 89, pp. 1021–1025, February 1992
Biochemistry

# Effects of different DNA polymerases in ligation-mediated PCR: Enhanced genomic sequencing and *in vivo* footprinting

**(methylation/Vent DNA polymerase/terminal transferase/DNase I/transcription)**

PAUL A. GARRITY AND BARBARA J. WOLD*

Division of Biology, California Institute of Technology, Pasadena, CA 91125

*Communicated by Norman Davidson, October 8, 1991 (received for review September 4, 1991)*

**ABSTRACT** **We have developed a simplified procedure for the ligation-mediated polymerase chain reaction (LMPCR) using *Thermococcus litoralis* DNA polymerase (Vent DNA polymerase). We show that Vent DNA polymerase produces correct, blunt-ended primer extension products with substantially higher efficiency than *Thermus aquaticus* (*Taq*) DNA polymerase or modified T7 DNA polymerase (Sequenase). This difference leads to significantly improved genomic sequencing, methylation analysis, and *in vivo* footprinting with LMPCR. These improvements include representation of all bands with more uniform intensity, clear visualization of previously difficult regions of sequence, and reduction in the occurrence of spurious bands. It also simplifies the use of DNase I cut DNA for LMPCR footprinting.**

Footprinting experiments are commonly and productively used to study protein–DNA interactions and DNA configuration *in vitro*. Analogous *in vivo* experiments done on genes in the living cell can bring a different and useful data set to the problem of gene expression, but they require special methods for visualizing the result. Direct genomic sequencing techniques, which permit the examination of single-copy genes in large genomes, are being used increasingly for this purpose (1–4). Ligation-mediated PCR (LMPCR) is a recently introduced method that substantially increases the absolute signal and the signal-to-noise ratio obtained for genomic sequencing (2, 5, 6). It does so by coupling PCR with genomic sequencing to provide specific amplification of a sequence "ladder," while preserving the identity and relative quantitative representation of each rung in the original cleaved genomic DNA preparation. Its application has made *in vivo* footprinting (2) and chromosomal methylation analysis (6) more readily accessible for organisms with large genomes (e.g., mammals).

While LMPCR has been used successfully by a number of investigators to obtain high quality *in vivo* footprint and methylation information (2, 6, 7, 8), it has had two problems that can significantly compromise data quality. These effects are minor in some regions of sequence but can be problematic in others. First, certain bands are consistently weak or missing in the genomic ladders. Second, "extra" bands occasionally appear in the genomic ladders. These bands, which aren't predicted from the sequence as independently determined from cloned DNA, are usually adjacent to expected bands and therefore convert some triplets into quartets, some doublets into triplets, and so on. We present here a solution for these problems that also permits simplification of the LMPCR procedure. These improvements stem from the use of *Thermococcus litoralis* DNA polymerase (Vent polymerase). This thermostable polymerase possesses no detectable terminal deoxynucleotidyltransferase activity under our conditions, and this characteristic dramatically improves LMPCR genomic sequencing. For *in vivo* footprinting and genomic sequencing applications, Vent polymerase yields substantially superior results, improving overall signal and, most importantly, the quality of sequence in difficult regions. We also show that starting material possessing 3′-hydroxyl ends (in this case DNase I-cut DNA), which had required modification of template ends with dideoxynucleotides in the older form of LMPCR (9), can now be used for *in vivo* footprinting purposes without modification.

## MATERIALS AND METHODS

**Cell Culture and DNA Preparation.** L cells were grown in Dulbecco's modified Eagle's medium with 10% undialyzed calf serum (Irvine) and 2 mM glutamine. Naked and *in vivo* dimethyl sulfate (DMS)-treated MM14 DNA was provided by P. Mueller. DNA samples for genomic sequencing and DMS footprinting were prepared as in refs. 10 and 11. *In vivo* DNase I treatment was as in ref. 12, except that cells were permeabilized on ice with lysolecithin (0.25 mg/ml) for 60 sec. Addition of dideoxynucleotides prior to LMPCR where noted was as in ref. 9.

**LMPCR.** LMPCR using T7 DNA polymerase (Sequenase version 1.0; United States Biochemical) and *Thermus aquaticus* (*Taq*) DNA polymerase (AmpliTaq; Cetus) was done as in refs. 2 and 5. LMPCR using *Thermococcus litoralis* DNA polymerase (Vent; New England Biolabs) was done as below. All solutions were chilled and manipulations were performed on ice except as noted. The pH values given are for room temperature. To 5 μl (2 μg) of DNA in TE (10 mM Tris·HCl, pH 7.5/1 mM EDTA) was added 25 μl of first-strand mix [1.2× first-strand buffer (48 mM NaCl/12 mM Tris·HCl, pH 8.9/6 mM $MgSO_4$/0.012% gelatin) with 0.3 pmol of gene-specific primer 1, 240 μM each dNTP, and 1 unit of Vent polymerase]. First-strand synthesis used a thermal cycle of 5 min at 95°C, 30 min at 60°C, and 10 min at 76°C. The samples were immediately iced. (It is important to minimize Vent polymerase activity during the ligation step by keeping the sample cold.) Twenty microliters of dilution solution (110 mM Tris·HCl, pH 7.5/18 mM $MgCl_2$/50 mM dithiothreitol/0.0125% bovine serum albumin) and 25 μl of ligation solution [10 mM $MgCl_2$/20 mM dithiothreitol/3 mM ATP/0.005% bovine serum albumin with 100 pmol of unidirectional linker in 250 mM Tris·HCl (pH 7.7) (thawed and added on ice) and 4.5 units of T4 DNA ligase (Promega)] were added. After incubation for 12–16 hr at 17°C, samples were iced and 9.4 μl of precipitation solution (0.1% yeast tRNA/2.7 M sodium acetate, pH 7.0) and 220 μl of ethanol were added. The samples were placed at −20°C for ≥2 hr and then spun for 15 min at 4°C in a microcentrifuge. The pellets were washed with 75% ethanol and dried in a Speed-Vac rotary evaporator (Savant). Samples were resuspended in 70 μl of water at room

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: LMPCR, ligation-mediated polymerase chain reaction; DMS, dimethyl sulfate; MCK, muscle creatine kinase.

*To whom reprint requests should be addressed.

temperature and placed on ice. After addition of 30 μl of amplification mix [3.33× amplification buffer (133 mM NaCl/67 mM Tris, pH 8.9/17 mM $MgSO_4$/0.03% gelatin/0.3% Triton X-100) with 670 μM each dNTP, 10 pmol of gene-specific primer 2, and 10 pmol of linker primer LMPCR.1)] and 3 μl (3 units) of Vent polymerase, samples were overlaid with 90 μl mineral oil and subjected to PCR using 18 cycles of 1 min at 95°C, 2 min at 66°C, and 3 min at 76°C, with these modifications: (*i*) first-round denaturation was 3.5 min at 95°C; (*ii*) 5 sec was added to the 76°C step with each successive cycle (e.g., second round, 3 min 5 sec at 76°C); (*iii*) for cycle 18, the 76°C step was 10 min. Samples were then placed on ice and 5 μl of labeling mix [1× amplification buffer with 2 mM each dNTP, 2.3 pmol of gene-specific primer 3 (end-labeled as in refs. 2 and 5), and 1 unit of Vent polymerase] was added. The labeling cycle was 3.5 min at 95°C, 2 min at 69°C, 10 min at 76°C, 1 min at 95°C, 2 min at 69°C, and 10 min at 76°C. The reaction was stopped by placing the samples on ice and adding 300 μl of stop solution (10 mM Tris·HCl, pH 7.5/4 mM EDTA/260 mM sodium acetate, pH 7.0, containing tRNA at 67 μg/ml). Samples were shifted to room temperature and extracted with 400 μl of phenol/chloroform/isoamyl alcohol (25:24:1, vol/vol). The aqueous phase was split into four aliquots of 94 μl, and 235 μl of ethanol was added to each. Before electrophoresis, samples were precipitated, washed, and dried as above. After resuspension in 7 μl of load dye (2, 5) and heating at 85°C for 5 min, samples were iced and then loaded on 6% polyacrylamide sequencing gel (2, 5). Loading one-fourth of an LMPCR mixture per lane yielded a strong signal on X-AR film (Kodak) after 3 hr with an intensifying screen at −80°C or 12 hr with no screen at −20°C. The unidirectional linker, linker primer, and muscle creatine kinase (MCK) oligonucleotides were as in ref. 2. The sequences (5′ to 3′) of the metallothionein I oligonucleotides were GAGTTCTCGTAAACTCCAGAGCAGC (primer 1), CAGAGCAGCGATAGGCCGTAATATC (primer 2), and AGCGATAGGCCGTAATATCGGGGAAAGC (primer 3).

## RESULTS AND DISCUSSION

LMPCR (Fig. 1) relies on creation of a blunt end in the initial primer extension reaction to serve as a ligation substrate. Later, in the labeling reaction, precise blunt-end termination of the extension product is required. If the final labeling extensions stop short or add extra nontemplated bases, the result will be extraneous, inappropriate bands. In general, imperfect extension products may result from DNA polymerases adding a nontemplated additional base after creating a blunt end (referred to as terminal transferase activity) (13). Both polymerases commonly used for LMPCR display some terminal transferase activity. Sequenase, used in the first-strand synthesis reaction, adds an extra base to ≈50% of its products. *Taq*, used in the PCR amplification and labeling steps, adds an extra base to ≈95% of its products (P. Mueller and B.J.W., unpublished data). Such activity during the first-strand synthesis creates molecules unable to participate in the blunt-end ligation. Should this activity show sequence preference, it would lead to underrepresentation or even complete loss of specific bands in the final LMPCR product. Terminal transferase activity might also explain the origin of spurious "extra" bands in an LMPCR ladder. For example, a single band in a genomic sequencing ladder would appear as a doublet if some products of the labeling reaction acquired the extra base. We hypothesized that the terminal transferase activities of Sequenase and *Taq* were the major source of imperfect regions in LMPCR ladders and sought a DNA polymerase that lacks appreciable terminal transferase activity.

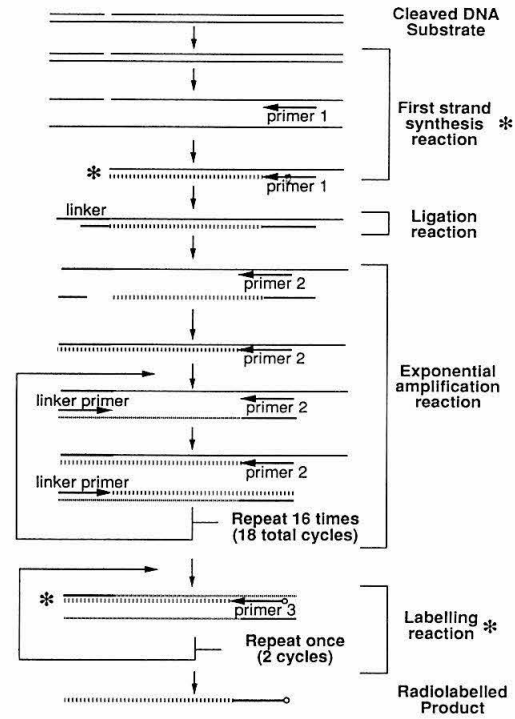



FIG. 1. LMPCR. Gene-specific primer 1 is hybridized to appropriately cleaved genomic DNA and extended using a DNA polymerase creating a blunt end (first-strand synthesis). This blunt end is ligated to a unidirectional linker of defined length and sequence with one blunt end (ligation reaction). This product is a substrate for PCR using gene-specific primer 2, which hybridizes internally, and a linker primer (LMPCR.1), which hybridizes to the ligated sequence (exponential amplification reaction). An end-labeled gene-specific primer, primer 3, is used to visualize the LMPCR product. Asterisks mark steps requiring efficient blunt-end generation. Open circle represents radioactive 5′-labeling of primer 3.

Vent DNA polymerase was tested in side by side comparisons with Sequenase and *Taq* DNA polymerase. The noncoding strand of the mouse MCK enhancer was used (14, 15) because it contains runs of G residues that have been problematic in Sequenase/Taq-based LMPCR. Mouse genomic DNA treated with DMS *in vitro* was used in the initial tests. Subsequent piperidine treatment gave G-specific cleavage (16), and LMPCR with MCK primers yielded the G-specific MCK sequence ladder. The activities of *Taq* DNA polymerase and Vent DNA polymerase were compared in the amplification and labeling stages of LMPCR. The products of Vent-catalyzed amplification and labeling consistently migrated more rapidly than those of *Taq*-catalyzed companion reactions by a one-base increment (Fig. 2). Since *Taq* is known to add an extra base to most of its products, we interpreted the migration shift as an indication that Vent lacked detectable terminal transferase activity and might therefore be an excellent candidate to replace both Sequenase and *Taq* in LMPCR.

When Vent was compared with Sequenase in the first-strand synthesis, the most obvious effect was that the yield of LMPCR product increased severalfold (Fig. 2). This is consistent with the creation of more blunt-ended molecules by Vent in the first-strand synthesis and confirms that this

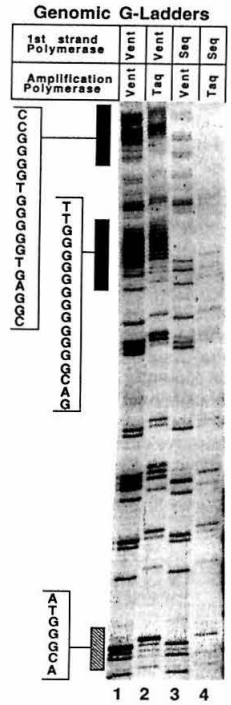



FIG. 2. LMPCR polymerase comparison. Each reaction used MCK primers and 2 μg of MM14 cell DNA treated with DMS *in vitro* and piperidine cleaved. Sequence at left is taken from Fig. 3. Hatched and solid bars are discussed in text. Seq, Sequenase.

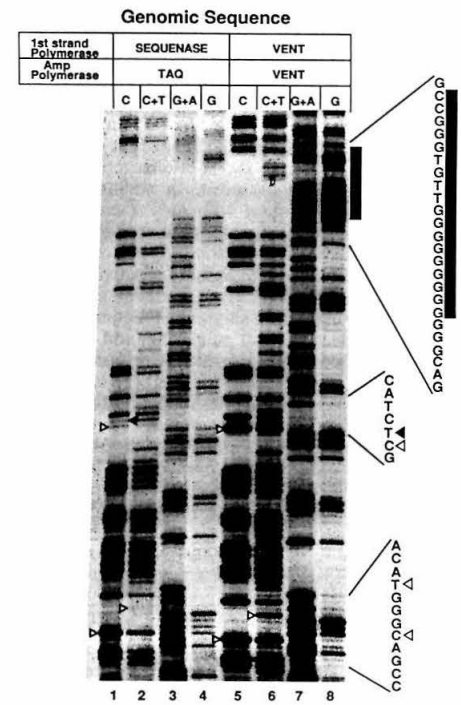



FIG. 3. LMPCR direct genomic sequencing method comparison. Each reaction used MCK primers and 2 μg of MM14 cell DNA treated with the appropriate Maxam and Gilbert sequencing reagent *in vitro* and then cleaved with piperidine. Black bar and triangles are discussed in text. Sequence at right was determined in ref. 15 from cloned DNA. An A residue reported in ref. 15 is here of near equal intensity in G and G+A lanes, so the sequence below the bar is given as GACGGG in agreement with determination from cloned DNA in ref. 14.

step is suboptimal with Sequenase. More important, Vent first-strand synthesis allowed visualization of bands that were underrepresented or absent in Sequenase-based LMPCR (note especially the several runs of G in the MCK enhancer denoted by the black bars in Fig. 2). Also, all bands in the ladder were more even in intensity. This is consistent with Vent creating blunt ends at many different sequences with similar efficiency and suggests that Sequenase shows sequence bias in blunt-end production. In addition to the shift in migration, the use of Vent instead of *Taq* in the amplification and labeling steps generated a sequencing ladder in which fewer spurious bands appeared. A good example is noted by the hatched bar in Fig. 2, where a known G triplet (15) appeared to be a quartet when *Taq* was used but resolved into the expected triplet when Vent was used. Thus, the use of Vent polymerase in LMPCR improved both the first-strand synthesis step and the amplification and labeling steps.

**Sequencing and Methylation.** In LMPCR-based genomic sequencing, total genomic DNA is cleaved by the Maxam and Gilbert base-specific reactions (11, 16) and this cleaved DNA serves as the substrate for LMPCR. Fig. 3 shows comparative genomic sequencing of the MCK enhancer done by either Sequenase/*Taq*-based or Vent-based LMPCR. Regions are noted where sequence determinations could not be made using Sequenase/*Taq*-based LMPCR (the published sequence is given at right). In the region noted at the bottom of Fig. 3, not only is a T residue missing (upper open triangle), but a C residue is rendered ambiguous by the presence of spurious bands in the G+A and G lanes (lower open triangle). In the middle region, the lower two bases of a CTC triplet are of equal intensity in the C lane (open triangle, correct band) due to a spurious band in the C lane (filled triangle). In each of these cases the Vent-based results are unambiguous. The extended G-rich regions (noted by the black bar) are even more dramatically improved by the use of Vent, though some ambiguity remains.

The analysis of chromosomal methylation [which has been implicated in such phenomena as chromosomal imprinting (17) and X chromosome inactivation (18)] is also improved. The Maxam and Gilbert C and C+T reactions require that C residues be unmethylated to participate. Because only unmethylated C residues appear in the sequence ladder, genomic sequencing provides both sequence and methylation information. The two adjacent cytosines in the G-rich region are represented well only with Vent-based LMPCR and the CTC triplet noted above appears to be a CCC triplet unless Vent-based LMPCR is used. Thus methylation information concerning these residues, which would have been either ambiguous or nonexistent using the older method, can now be clearly determined. It can be concluded from this experiment that the C residues of the MCK enhancer are not entirely methylated in MM14 cells, because they do appear in the ladder. However, the degree of partial methylation of any C residue can be determined only by a side-by-side comparison with unmethylated control DNA; hence, methylation of some copies of the MCK enhancer would not be detected in this experiment.

**DMS Footprinting.** *In vivo* footprinting is an especially sensitive application of genomic sequencing because it focuses directly on the relative intensities of individual bands in different samples. This demands that the intensity of each band in the final sequence ladder reflect in a consistent manner its relative abundance in the population of starting material. Thus, spurious "extra" bands that comigrate with genuine bands can obscure the subtle quantitative changes that usually comprise a footprint. In addition, some very useful cleavage agents react with only a subset of bases. Visualization of all legitimate cleavage products is therefore important, because an area of protein–DNA interaction may contain only one or two differentially reactive sites.

*In vivo* footprinting using DMS involves exposing intact cells to DMS, terminating the alkylation reaction, purifying the alkylated DNA, cleaving with piperidine, and comparing the resulting G-specific sequence ladder with one generated by exposing purified, naked DNA to DMS *in vitro* (4). Band intensity changes between samples reflect protein binding and any other changes in DNA structure that alter reactivity with DMS. Fig. 4 shows *in vivo* footprinting of the MCK enhancer, which is active in differentiated muscle cells (myocytes) but not in undifferentiated muscle precursor cells (myoblasts) (2, 14, 15). G-specific sequence ladders were derived from *in vivo* DMS treatment of undifferentiated MM14 myoblasts and differentiated MM14 myocytes and from *in vitro* DMS treatment of naked MM14 DNA. LMPCR was performed with either Sequenase or Vent first-strand synthesis followed by either *Taq* or Vent amplification and labeling. The footprint information derived from this experiment was consistent with that of previous LMPCR footprints of the MCK enhancer (2). In the region shown, myocyte-specific footprints are noted at three regulatory elements that have previously been defined as important for function (reviewed in ref. 2). The biological implications of this pattern have been discussed (2), and we focus here on how the new methods affect *in vivo* footprint analysis. Vent-based LMPCR gives greater absolute signal, and interactions that are sometimes difficult to see using Sequenase/*Taq*-based LMPCR, such as those at MEF-2 and near MEF-1, are now more obvious. This improvement is the combined result of acquiring previously missing bands and eliminating extraneous bands. For footprinting purposes, it is vital that identical DNA samples yield identical LMPCR results so that quantitative differences between different DNA samples can be interpreted. Vent-based LMPCR yields highly reproducible results, as shown by the exact match between the

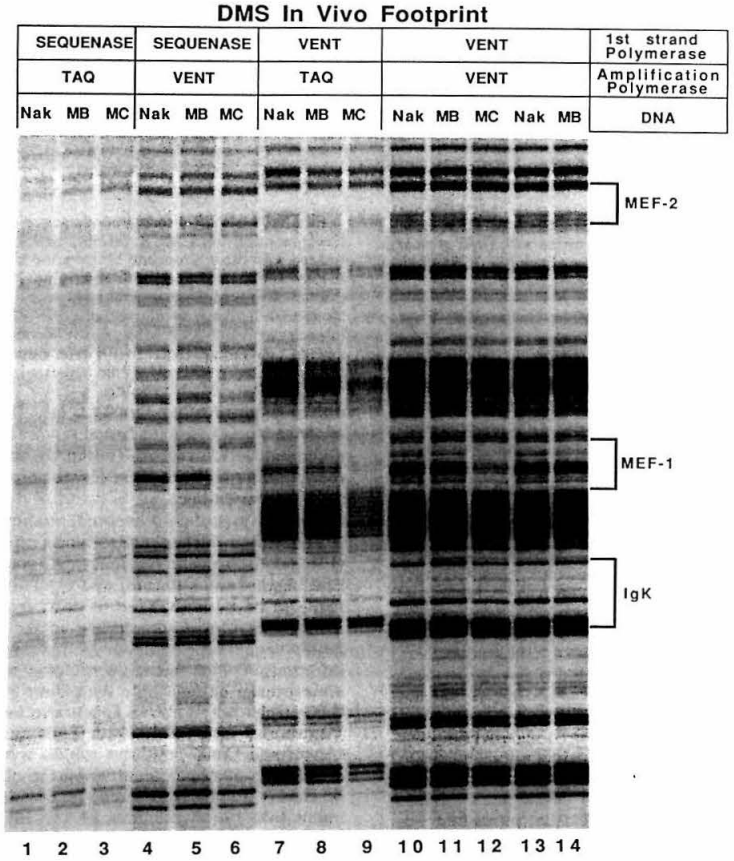



FIG. 4. LMPCR DMS *in vivo* footprinting method comparison. Each reaction used MCK primers and 2 μg of MM14 cell DNA treated with DMS *in vivo* or *in vitro* prior to piperidine cleavage. Naked (Nak) DNA was purified and then DMS-treated *in vitro*. Myoblast (MB) and myocyte (MC) DNA samples were from cultured cells treated with DMS. Binding sites indicated at right are described in ref. 2, where IgK is called κ and MEF-2 is called A-rich. Overexposure of lanes 10–14, required to see lanes 1–3, obscures footprints apparent at nonsaturating exposures.

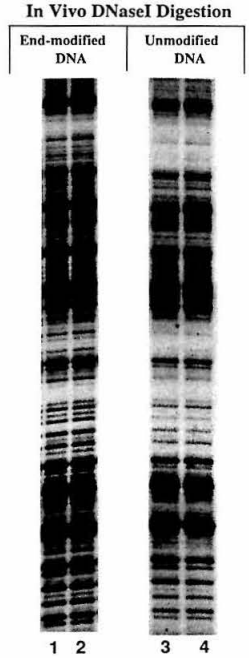



FIG. 5. Comparison of LMPCR of DNase I-digested DNA with (lanes 1 and 2) and without (lanes 3 and 4) prior dideoxynucleotide modification. Each reaction used metallothionein I primers and 2 $\mu$g of *in vivo* DNase I-digested L-cell DNA (lanes 1 and 3, DNase I at 37.5 $\mu$g/ml; lanes 2 and 4, 25 $\mu$g/ml). All lanes are from the same autoradiograph.

ladders from independent LMPCRs using duplicate DNA samples (Fig. 4, lanes 10 and 13, lanes 11 and 14). The LMPCR improvements observed were not restricted to the MCK enhancer region shown. Similar results were obtained for the mouse metallothionein I promoter (unpublished data).

**DNase I-Cleaved Substrates.** Beyond the issues of ladder quality, a separate limitation of LMPCR had been an inability to use DNA possessing 3′-hydroxyl ends for *in vivo* footprinting from organisms with large genomes. Although DMS/piperidine cleavage does not leave 3′-hydroxyl ends, other useful footprinting agents, such as DNase I, do. Riggs and coworkers were able to circumvent this problem by the addition of a dideoxynucleotide to the 3′-hydroxyl ends of DNase I-digested material prior to Sequenase/*Taq*-based LMPCR (9) and have obtained *in vivo* footprints with DNase I (19). Though effective, the additional manipulations are time-consuming and in our hands have resulted in low recovery of input DNA. With the Vent-based protocol, unmodified *in vivo* DNase I-digested DNA samples yield ladders similar in clarity and intensity to those from dideoxynucleotide-modified DNA (Fig. 5). Thus the addition of a dideoxynucleotide prior to LMPCR of 3′-hydroxyl-containing DNA is no longer necessary. The basis for this is not certain, but increased temperature (76°C versus 47°C) in the Vent-catalyzed first-strand synthesis reaction may be important. It may inhibit the priming of DNA synthesis by the enormous numbers of genomic DNA 3′-hydroxyl ends present in the first-strand reaction, while still providing efficient extension from the hybridized, gene-specific oligonucleotide.

The more uniform and efficient use of starting material in Vent-based LMPCR should allow the use of less sample DNA while still obtaining statistically significant results. Fluctuations in band intensity due to sampling error occur when the population sampled is small (founder effect, as discussed in ref. 5). They can obscure sequence or be mistaken for a footprint. Improved efficiency in LMPCR reduces the potential for founder-based artifacts in formerly problematic sequences. Although the favorable effects of the Vent-based LMPCR procedure can be readily explained by an absence of terminal transferase activity, that is not formally proved here. Whatever their mechanistic origin, the properties of Vent extension reactions reported here suggest that Vent polymerase may also be superior for other applications in which blunt-ended products are desired.

We thank Paul Mueller for generous gifts of DNA, comments on the manuscript, and many helpful discussions; Linda Huang and Jeff Miner for advice and comments on the manuscript; and Joe Hacia and Pete Mathers for assistance and discussions. This work was supported by grants from the Muscular Dystrophy Association and National Institutes of Health to B.J.W. and by Predoctoral Training Award H600021.

1. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
2. Mueller, P. R. & Wold, B. (1989) *Science* **246**, 780–786.
3. Mirkovitch, J. M. & Darnell, J. E. (1991) *Genes Dev.* **5**, 83–93.
4. Ephrussi, A., Church, G. M., Tonegawa, S. & Gilbert, W. (1985) *Science* **227**, 134–140.
5. Mueller, P. R. & Wold, B. (1991) *Methods* **2**, 20–31.
6. Pfeifer, G. P., Steigerwald, S. D., Mueller, P. R., Wold, B. & Riggs, A. D. (1989) *Science* **246**, 810–813.
7. Kara, C. J. & Glimchĕr, L. H. (1991) *Science* **252**, 709–712.
8. Rideout, W. M., Coetzee, G. A., Olumi, A. F. & Jones, P. A. (1990) *Science* **249**, 1288–1290.
9. Tanguay, R. L., Pfeifer, G. P. & Riggs, A. D. (1990) *Nucleic Acids Res.* **18**, 5902.
10. Mueller, P. R., Salser, S. J. & Wold, B. (1988) *Genes Dev.* **2**, 412–427.
11. Saluz, H. P. & Jost, J. P. (1987) *A Laboratory Guide to Genomic Sequencing* (Birkhauser, Boston).
12. Zhang, L. & Gralla, J. D. (1989) *Genes Dev.* **3**, 1814–1822.
13. Clark, J. M. (1988) *Nucleic Acids Res.* **16**, 9677–9686.
14. Jaynes, J. B., Johnson, J. E., Buskin, J. N., Gartside, C. L. & Hauschka, S. D. (1988) *Mol. Cell. Biol.* **8**, 62–70.
15. Sternberg, E. A., Spizz, G., Perry, W. M., Vizard, D., Weil, T. & Olsen, E. N. (1988) *Mol. Cell. Biol.* **8**, 2896–2909.
16. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
17. Solter, D. (1988) *Annu. Rev. Genet.* **22**, 127–146.
18. Lyon, M. F. (1991) *Trends Genet.* **7**, 69–70.
19. Pfeifer, G. P. & Riggs, A. D. (1991) *Genes Dev.* **5**, 1102–1113.

# Chapter 4

## Activation of IL-2 transcription: Coordinated *in vivo* assembly of differentially regulated factors.

**Thesis version**

**Paul A. Garrity, Dan Chen, Ellen V. Rothenberg, Barbara Wold**

Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125

**Abstract**

The interleukin-2 gene is transcribed exclusively in T cells in response to antigen stimulation. Despite its developmentally-restricted signal-dependent expression, its regulatory region contains multiple individual elements that could potentially drive IL-2 transcription in a much wider range of cell types and conditions. Here, *in vivo* footprinting has been used to determine the pattern of protein:DNA interactions at the IL-2 locus in EL4 T cells, which transcribe IL-2 in response to appropriate stimuli, and in an hematopoietic cell population which cannot be induced to make IL-2, the pre-mast cell line 32D. Parallel *in vitro* analyses show that all cell types examined possess nuclear factors that recognize elements within the IL-2 regulatory region. However, *in vivo* analysis shows no evidence of assembly of a pre-complex at the IL-2 locus in EL4 cells prior to stimulation, which might serve as a lineage-specific marker at the IL-2 locus. We also see no evidence of protein:DNA interactions at the IL-2 locus in the 32D pre-mast cells, and hence no evidence of active repression in a non-expressing cell type. Similarly no *in vivo* protein:DNA interactions were detected at IL-2 in the non-expressing L cell fibroblast line. Upon stimulation of EL4 T cells, a multi-factor complex assembled at the IL-2 locus and transcription of IL-2 was induced. The sites recognized by proteins *in vivo* also recognized activities present in nuclear extracts from these cells. Stimulation of EL4 T cells in the presence of the immunosuppressant drug cyclosporin A (CsA) prevented a subset of the DNA-binding activities seen in nuclear extracts, however the *in vivo* effect of CsA-treatment was to entirely block establishment of the enhancer complex. Taken together these data indicate that proteins do not stably bind to the IL-2 promoter/enhancer *in vivo* until T-cell-specific, signal-dependent, CsA-sensitive events initiate the

coordinated binding of general and cell-specific factors. These observations suggest that a significant amount of the signal integration involved in IL-2 induction is carried out at the level of DNA occupancy and that the observed multi-component, coordinated assembly is an essential component of this integrative process.

**Introduction**

Interleukin-2 (IL-2) is an important growth factor for T cells and B cells which is synthesized by a subset of T helper cells upon their activation by antigen recognition. The regulation of IL-2 production has been extensively studied not only because of its importance in mounting an immune response, but also as a paradigm for the induction of new gene expression in response to signal transduction pathways involved in T cell activation. Furthermore, IL-2 induction can be specifically blocked by treatment with immunosuppressant drugs such as cyclosporin A (CsA). This aspect of IL-2 regulation has been extensively examined as a way of gaining a molecular understanding of the activity of these drugs.

Much of the control of IL-2 production takes place at the transcriptional level (Kronke et al. 1985; Brorson et al. 1991; Ullman et al. 1990). The recognition of antigen by the T helper cell in the context of the appropriate antigen presenting cell initiates a signal transduction cascade which, among many other things, results in a rapid rise in intracellular calcium and an activation of protein kinase C (Rothenberg 1992). Within twenty minutes, in a process requiring protein synthesis, IL-2 transcription is induced (Shaw et al. 1988). CsA blocks aspects of this signal transduction cascade and in so doing blocks IL-2 transcriptional induction. Recent evidence suggests that it does so, at least in part, through an inhibition of the calcium-calmodulin-dependent phosphatase calcineurin (O'Keefe et al. 1992; Clipstone and Crabtree 1992). Thus, IL-2 transcription is both developmentally-restricted (T cell specific) and signal-dependent (taking place only in activated T cells not treated with CsA).

The mechanisms used to control a gene's transcription can vary greatly, but share a basic feature. All depend on multiple regulatory elements

flanking, and sometimes within, the coding region which interact with cellular sequence-specific DNA binding proteins. In turn, these factors directly or indirectly control the assembly and activity of the general RNA polymerase II machinery of the cell near that gene's transcriptional initiation site. Previous studies of the *in vivo* assembly of transcription complexes at the muscle creatine kinase (MCK) (Mueller and Wold 1989) and c-fos (Herrera et al. 1989) genes have shown that the mechanisms by which transcription is turned on in a signal-dependent fashion can be very different. In the case of the myocyte-specific enhancer of the MCK gene, that enhancer is not occupied by major groove DNA-binding factors in the non-expressing myoblast. Upon the differentiation of myoblasts to myocytes, which is triggered by decreasing the levels of growth factors such as fibroblast growth factor in the culture medium, multiple regulatory factors bind to the enhancer and MCK transcription is activated. c-fos induction by epidermal growth factor treatment of serum-starved fibroblasts is mediated through a serum responsive element (SRE). In stark contrast to the MCK case, the SRE is indistinguishably occupied by DNA-binding regulatory factors both before and after induction. Thus, in the MCK case the first step in activation was the binding of proteins to the DNA whereas in the c-fos case proteins were already bound and it was their activities that were altered. In both of these cases determining the pattern of protein:DNA interactions at these regulatory regions in the living cell was important in determining the mechanisms used to control their gene's transcription. Similar mechanistic diversity can be imagined for controlling transcription in a cell-type-restricted fashion. However, in most cases studied, such as MCK (Mueller and Wold 1989) and the immunoglobulin heavy chain enhancer

(Ephrussi et al. 1985), no protein:DNA interactions were detected in non-expressing cell types.

An additional commonality among diverse genes is that a functional transcriptional enhancer or silencer is composed of a number of individual factor binding sites (Thompson and McKnight 1992). These individual cis-acting elements often differ from one another in their regulatory capacities when examined by themselves, e.g., one may function as a cell-type specific activator while another as a ubiquitous activator. However, when combined as they are in naturally occurring regulatory regions, they collaborate to generate large and specific effects which are quite distinct from a simple summing of parts. As shown in the MCK and c-fos examples above, knowing the degree to which occupancy of the regulatory region or the activity of bound factors is limiting is key to understanding the mechanisms controlling gene activation and inactivation. Thus, in addition to identifying and categorizing individual cis-elements by classical mutagenesis approaches and characterizing the presence of potential binding activities by *in vitro* biochemical assays, direct measurement of *in vivo* occupancy of these binding sites is needed, but is available for only a few genes.

The 300 base pairs immediately upstream of the IL-2 transcriptional start site contain a minimal promoter/enhancer region sufficient to drive a reporter gene in an appropriate IL-2-type expression pattern (Ullman et al. 1990). This DNA regulatory region contains multiple individual DNA regulatory elements of diverse character. The NF-AT element, of which two are present, binds a factor found only in activated T cells (Shaw et al. 1988) and when present in several copies confers expression largely restricted to activated T cells upon a reporter gene in a transgenic mouse (Verweij et al.

1990). The NF-κB and AP-1 elements confer activation upon a reporter gene in stimulated, but not unstimulated T-cells (Brabletz et al. 1991). However, these elements recognize factors found in many cell types and are signal-responsive, but not cell-type-specific activating elements (Hentsch et al. 1992). The IL-2 regulatory region also contains two CACCC motifs which bind a number of factors which are expressed constitutively and are not T-cell-restricted, and they likely act as constitutive, general elements (Chen and Rothenberg 1992). Finally, the IL-2 regulatory region contains a site, OCT/OAP40, which binds both the relatively ubiquitous octamer family proteins (Kamps et al. 1990; Hentsch et al. 1992), as well as the more restricted, stimulation-dependent OAP40 protein (Ullman et al. 1991). The OCT/OAP40 unit functions as a stimulated-T-cell-specific activator site, though the octamer component is capable of binding to DNA constitutively (Kamps et al. 1990; Brabletz et al. 1991; Ullman et al. 1991). As a whole the IL-2 regulatory region drives expression in a stimulation-dependent, T-cell-restricted fashion. The reliance on interactions between multiple sites to create this pattern of transcription is evident in a number of mutational studies. Gross alterations, such as splitting the regulatory region approximately in half, left two halves that each contained multiple factor binding sites. However, these halves were unable to inducibly activate transcription alone, but dimerizing either half largely restored activity (Durand et al. 1988). Finer mutations which altered individual cis-elements, such as the NF-AT, AP-1 or OCT/OAP40 sites, led to between 4 and 20 fold decreases in expression (Novak et al. 1990; Jain et al. 1992; Durand et al. 1988). Taken together, these data suggest that there exists extensive functional collaboration between the diverse individual regulatory elements within the IL-2 regulatory region. But it does not indicate whether these collaborative

events occur largely at the level of DNA regulatory element occupancy or the ability to activate transcription.

The IL-2 minimal promoter/enhancer region also confers upon a reporter gene the complete CsA sensitivity of the IL-2 gene (Mattila et al. 1990; Randak et al. 1990). The individual regulatory elements show a diversity of behavior (Mattila et al. 1990; Randak et al. 1990). The ability of the NF-AT and OCT/OAP40 elements to activate transcription of a reporter gene when multimerized is entirely blocked by CsA, whereas the ability of NF-κB to function in this way is only partially sensitive. The ability of the AP-1 element to function in these assays is insensitive to CsA. The factors which bind to the CACCC motifs *in vitro* appear insensitive as well (see below), however this element's transcriptional activating potential has not been examined.

In this work, we investigate the mechanism used to control IL-2 transcription, including the way in which the individual regulatory elements collaborate, by examining the pattern of protein:DNA interactions at the IL-2 locus in living cells. We examine this in different cell-types: lineally related cells which can and cannot be induced to transcribe IL-2 as well as unrelated cell types which do not make IL-2. We also examine cells in whom induction of IL-2 transcription has been blocked by treatment with CsA.

**RESULTS:**

To examine the mechanisms used to turn on IL-2 transcription in T cells as well as mechanisms that may be used to keep it off in non-expressing cell types, we have determined the pattern of protein:DNA interactions at the IL-2 locus in three different types of cells. EL4.E1.F4 (EL4) thymoma cells were used as a model IL-2 producer cell type. Upon stimulation with the phorbol

ester 12-O-tetradecanoyl phorbol 13-acetate (TPA) and the calcium ionophore A23187, which mimic many of the physiological effects of antigen stimulation, EL4 T cells show many of the same gene induction events as antigen-activated T helper cells (Novak et al. 1990). This includes strong IL-2 transcriptional induction which is sensitive to CsA treatment (Randak et al. 1990). In addition, much of the extensive work on IL-2 regulation has been done in this cell type. To examine why the IL-2 gene is not expressed in other hematopoietic cell types, 32D pre-mast cells were chosen because they respond to TPA and ionophore stimulation with some of the same gene-induction responses as EL4 T cells (Novak et al. 1990; Novak and Rothenberg 1990). However, they fail to induce IL-2. Finally, L cell fibroblasts were studied to examine mechanisms that may be used to keep the IL-2 gene off in non-hematopoietic cells (data not shown). RNase protection analysis was performed to confirm the expression pattern of IL-2 in the various cell populations (Fig. 1), using the same cell preparations analyzed in the *in vitro* DNA binding and *in vivo* footprinting experiments described below.

*Multiple factors that can bind sites in the IL-2 promoter/enhancer in vitro are present in the nuclei of cells that do and cells that do not express IL-2.*

The presence of nuclear factors that can bind to the IL-2 promoter/enhancer is a prerequisite for the assembly of a transcription complex on the IL-2 gene *in vivo*. To establish the availability of such factors, nuclear extracts were prepared from the same EL4 T cell and 32D pre-mast cell populations examined above. These nuclear extracts were then tested to determine whether they contained factors capable of binding to isolated sites in the IL-2 promoter/enhancer *in vitro* by using gel-mobility shift assays. The

binding activities detected in EL4 T cells (figure 2) can be put into three categories: constitutive, stimulation-dependent (inducible), and stimulation-dependent but CsA sensitive. Oligonucleotides spanning the CACCC sites in the IL-2 promoter bound nuclear factors in extracts from all cell preparations (lanes 1 to 6). The most prominent CACCC activities were neither stimulation-dependent nor CsA-sensitive. Similarly, the octamer binding activities in these cells were neither stimulation-dependent nor CsA sensitive (lanes 7 to 9). The proximal AP-1 site was bound by factors which were inducible, but were not CsA sensitive (lanes 10 to 12). By contrast, both the distal and proximal NF-AT binding sites, as seen previously by others, showed stimulation-dependent binding activity that was highly CsA sensitive (lanes 19 to 24). The NF-κB site bound inducible factors whose binding activity, under our conditions, was also much reduced by CsA (lanes 16 to 18). Finally, the newly identified TGGGC element (see below) bound inducible activities with partial CsA-sensitivity (lanes 13 to 15). Thus multiple DNA-binding factors are present in the nuclei of unstimulated, stimulated, and CsA-treated stimulated EL4 T cells, but the precise combination of nuclear factors that were present differed in a signal-dependent and CsA-sensitive fashion.

Similar assays were done to determine whether factors recognizing IL-2 regulatory elements were also present in the 32D pre-mast cells (data not shown). Constitutive activities capable of binding the CACCC and octamer motifs, and TPA/A23187-responsive activities that bound the NF-κB and TGGGC sites were present at levels comparable to those in EL4 T cells. However, factors recognizing the AP-1 site were much reduced in abundance and inducible activities recognizing NF-AT were undetectable. Taken together, these results show that the nuclei of all EL4 and 32D cell populations

examined contain multiple sequence-specific factors capable of binding to elements within the IL-2 regulatory region *in vitro*. However, the precise combination of nuclear factors that were present differed in a cell-type-specific, as well as signal-dependent, and CsA-sensitive fashion.

*Determining the pattern of protein:DNA interactions at the IL-2 locus in vivo*

To detect *in vivo* protein:DNA interactions, *in vivo* footprinting was performed by treating intact cells with the membrane permeable alkylating agent dimethyl sulfate (DMS) (Ephrussi et al. 1985). DMS detects proteins bound to DNA at G-residues in the major groove. Protein occupancy can protect individual G-residues within its recognition site from reaction with DMS or, occasionally, render additional G-residues within that site hypersensitive to DMS. After this *in vivo* alkylation with DMS, subsequent treatment with piperidine was used to produce DNA cleavage at the methylated G's. When appropriately visualized, this results in an *in vivo* G-specific cleavage ladder. In parallel, unmodified DNA was purified from cells, deproteinized and this naked DNA was then treated with DMS *in vitro* and subsequently cleaved with piperidine. DMS footprints were revealed by comparing the *in vivo* G-specific cleavage ladders from different cell populations with each other and with the pattern of *in vitro* G-specific cleavage of naked DNA. The DMS footprints, consisting of protected G-residues and associated G-hypersensitivities, were visualized by ligation mediated PCR (LMPCR) genomic footprinting (Mueller and Wold 1989) as described previously (Garrity and Wold 1992).

*Stimulation of EL4 T cells results in the coordinated binding of multiple proteins to the IL-2 promoter/enhancer*

The pattern of *in vivo* protein:DNA interactions at the IL-2 locus was assessed by quantitatively comparing band intensities in different *in vivo* DMS-treated and naked DNA samples (Fig. 3A, 3B). First, note that many sites in the IL-2 promoter/enhancer showed footprints in EL4 T cells upon induction. Comparison of the G ladder generated from IL-2 transcribing, induced EL4 T cells (Fig. 3A lane 5, Fig. 3B lane 6) with the G ladder from the same cells prior to induction (uninduced EL4 T cells, Fig. 3A lane 4, Fig. 3B lane 5) revealed many differences in DMS reactivity. Each interaction, indicated at the right margin of the figures 3A and 3B and summarized in figure 4, involves just a few specific G residues, as expected for a DMS footprint. These footprints were highly reproducible in multiple, independent experiments.

Each interaction corresponds to a sequence-specific binding activity that can be detected *in vitro* in an electrophoretic mobility shift assay, with the exception of one site not yet tested. *In vivo* footprints were found at previously identified NF-AT (both -280 and -135), AP-1 (-150), octamer/OAP40 (-70 to -85), and NF-κB (-200) recognition sites, all of which have been shown to be functionally important by *cis*-element mutagenesis experiments (Serfling et al. 1988; Ullman et al. 1991; Durand et al. 1988; Briegel et al. 1991; Jain et al. 1992). The *in vivo* footprinting also confirmed the occupancy of a CACCC site (-294) described elsewhere (Chen and Rothenberg 1992) and revealed interactions at two previously unidentified elements, a proximal CACCC motif (-60) and a TGGGC site (-225). Each of these elements bound sequence-specific factors from nuclear extracts *in vitro* as demonstrated in figure 2. An additional

unanticipated interaction was detected at an ATGG site (-175), which has not yet been examined *in vitro*. Finally, an additional interaction was seen between the two IL-2 TATA boxes, which may reflect the assembly of the basic RNA polymerase II machinery in that region. Occupancy of all these sites was detectable after one hour of induction, was maximal by two hours, and persisted at maximal levels at least 10 hours (figure 3 and data not shown). Footprints at the distal NF-AT site, the NF-κB site, and the proximal AP-1 site after two and five hours of induction are shown in figure 5.

Conversely, several other potential regulatory elements were not detectably occupied in the major groove *in vivo*. We failed to detect contacts at the G residues in the distal octamer motif at -255, the distal AP-1 motif at -180 (16, 18), and the region 3' of the distal NF-AT core at -280 (3' NF-AT site). The counterparts of these elements in the human gene, though not perfectly homologous to the murine elements, have been shown to contribute to its function (Kamps et al. 1990; Jain et al. 1992; Thompson et al. 1992). In these cells, no clear evidence was found for the binding of factors to the CD28 response element (CD28RE) either, although protection was seen at the junction of this site with the proximal AP-1 site. This interpretation is consistent with the observation that the CD28RE plays no role in IL-2 induction in response to TPA+A23187 induction in the human Jurkat T cell line (Fraser et al. 1992). Though *in vivo* footprints were performed up to 24 hours after induction (data not shown), no interactions were detected at the G residues in the putative negative element at -105, proposed to decrease induced transcription of the human IL-2 gene by functioning to shut off IL-2 transcription at later time points (Williams et al. 1991). Thus, for several

potential regulatory elements we see no *in vivo* evidence that they play a role in IL-2 transcriptional regulation in mouse EL4 T cells.

*No protein:DNA interactions can be detected in unstimulated EL4 T cells, 32D pre-mast or L cells*

In principle, cells that are not transcribing IL-2 might have stably bound repressors at the IL-2 locus or partially assembled enhancer complexes. Furthermore, a developmentally committed, but as yet uninduced IL-2 producer, such an EL4 T cell, might reveal its committed status in the form of a different pattern of protein:DNA interaction than an IL-2 non-producer, such as a 32D cell or L cell. However, we detected no *in vivo* protein:DNA interactions in any of the non-expressing cells examined (unstimulated EL4 T cells, both stimulated and unstimulated 32D cells, and L cells [data not shown]). This conclusion was drawn by comparing the G ladders derived from naked DNA samples with those from *in vivo* samples. The naked DNA samples (Fig. 3A lane 3, Fig. 3B lanes 1 and 4) were very similar to the *in vivo* samples from cells not expressing IL-2 (Fig. 3A, lanes 1, 2, and 4, Fig. 3B lanes 2, 3, and 5) (20). Particularly illuminating is inspection of the sites that showed extensive protein:DNA interactions in the induced EL4 cells. This clearly shows that the set of protein:DNA interactions present in that IL-2 transcribing cell are completely absent from the cells that do not express IL-2. The absence of detectable *in vivo* interactions contrasts with the presence of corresponding DNA-binding activities in the nuclei of these cells.

Several subtle differences in DMS reactivity between naked DNA samples and *in vivo* DMS-treated samples were observed on the coding strand near -100 and -200. Several G's, only on that strand, were 20% to 25% hyper-

reactive in all the *in vivo* DMS-treated samples from non-expressing cells, but were not associated with the DMS protections that usually accompany protein:DNA interactions Such isolated hypersensitivities are often encountered in comparisons of naked DNA versus *in vivo* DMS-treated DNA and do not seem to correspond with any protein:DNA interactions (Mueller and Wold 1989; Kara and Glimcher 1991).

*CsA treatment entirely blocks protein:DNA interactions in stimulated EL4 cells*

IL-2 transcription is completely sensitive to CsA, whereas the individual regulatory elements essential for its transcription differ in their CsA-sensitivity. This effect could occur either through a complete block of DNA occupancy or through the elimination of a sufficient subset of DNA-bound activators that transcription could not occur. When EL4 T cells were stimulated in the presence of CsA, all protein:DNA interactions detected in induced EL4 T cells were absent (lanes 6 and 10). Thus, CsA blocks the *in vivo* binding of all the factors whose binding to the IL-2 regulatory region could be detected, despite affecting the *in vitro* DNA-binding activities of only a subset of factors. Therefore, at least one CsA-sensitive component of the EL4 response is required for the coordinated binding of all factors. Furthermore, when EL4 T cells were induced for 2 hours, long enough to form a full complex (figure 5, data not shown), and then treated with CsA, all footprints were gone when protein:DNA interactions were examined 7 hours later (data not shown), a time point when the complex would otherwise have been present. Thus, CsA blocks both the establishment and the maintenance of the protein:DNA complex at the IL-2 promoter/enhancer.

*The absence of in vivo protein:DNA interactions in cells that do not express IL-2 is gene specific*

The conclusion that unstimulated EL4 T cells, CsA-treated stimulated EL4 T cells, and 32D cells show none of the *in vivo* interactions characteristic of induced T cells required a positive control to show that the absence of IL-2 footprints is gene specific and not a simple artifact of cell handling or footprinting manipulations. This was done by using the same DNA preparations to footprint another gene, metallothionein-I (MT-I), which is expressed in all of these cell populations (Fig. 1). Figure 6 shows that the characteristic *in vivo* footprint at the MT-I locus SP1 site is present in all samples (compare *in vivo* lanes 2, 3, 5, 6 and 7 with *in vitro* DMS lanes 1 and 4).

**DISCUSSION:**

We have shown that EL4 T cell induction results in the coordinated assembly of a protein:DNA complex at the IL-2 promoter/enhancer *in vivo* (see figure 7). We saw no preassembly of the available subset of factors on the regulatory region prior to the cell receiving the inducing signal. Such preassembly would be indicative of transcriptional activation acting as the limiting step. In contrast, upon EL4 T cell stimulation, we saw the coordinated binding of multiple factors to the previously unoccupied IL-2 regulatory region, implicating the regulation of DNA occupancy as an important limiting step in IL-2 transcriptional activation. CsA has been shown previously to block the *in vitro* binding of factors sites which are important for the transcriptional activity of IL-2 *in vivo*. Our data show that CsA works through these sites by preventing the occupancy of not just these, but all the other factor binding sites that have been detected *in vivo*.

No *in vivo* protein:DNA interactions were detected in cells which were not transcribing IL-2, whether the cells were of hematopoietic or non-hematopoietic lineage. Our observations do not rule out the possibility of repressive interactions to which DMS was insensitive, such as binding of proteins in the minor groove or to recognition sites devoid of G residues. Nonetheless, in non-expressing cells we see no *in vivo* evidence of proteins bound to sites that become detectably occupied by activator proteins in IL-2 expressing cells. This argues quite strongly against competition between activator and repressor proteins for binding site contacts in the major groove (of the type prominent in the *eve* enhancer elements of Drosophila, for example) (Small et al. 1992) as a way of enforcing cell-type and signal-dependent restriction of IL-2 in these cells.

Multiple factors capable of recognizing individual elements in the IL-2 enhancer/promoter were detected in the nuclei of all cell types and physiological states examined (Figure 7). It is clear that the presence of a subset of factors alone is not sufficient to allow them to stably occupy their binding sites in the IL-2 gene *in vivo*. For example, the CACCC and octamer sites are occupied *in vivo* only in the IL-2 transcribing cells, though corresponding binding activities are present in nuclear extracts constitutively. Upon EL4 T cell stimulation binding activities appear which recognize additional cis-elements. Even when the stimulation is performed in the presence of CsA some of these inducible DNA-binding activities, such as AP-1, still appear. However, the binding sites for these CsA-resistant factors remain unoccupied. One interpretation of this is that inducible, CsA-sensitive factors such as NF-AT (and possibly the OAP40/OCT complex) play the limiting roles in nucleating DNA:protein complex assembly in EL4 T cells. In considering

the absence of *in vivo* footprints in 32D cells, it is formally possible that the IL-2 locus is permanently shut down in a cell line destined never to make IL-2 using a mechanism that may not reflect normal developmental processes. However, the presence in these cells after induction of most, but not all (e.g., AP-1, NF-AT) of the activities found in induced EL4 T cells is consistent with the idea that key factors such as NF-AT are required for DNA:protein complex assembly.

The specific regulatory activities attributable to the newly identified TGGGC and proximal CACCC elements can be addressed in part by reexamining previously published deletion mutant studies. A series of 5' deletion constructs of the mouse IL-2 gene experienced a 5 fold drop in inducibility in EL4 cells when 36 bp including the TGGGC motif were eliminated (Serfling et al. 1988). This region contains no other element which showed an *in vivo* interaction. An internal deletion across the proximal CACCC motif in the human IL-2 gene decreased activity of the otherwise intact enhancer 5-fold, though this deletion is positioned next to the TATA boxes, and alterations in spacing could have contributed to this effect (Durand et al. 1988). Thus the newly identified elements appear to contribute to IL-2 expression.

Two models, which are not mutually exclusive, could explain the coordinated binding seen at IL-2 *in vivo*: cooperative binding or limited site accessibility. Cooperative binding interactions between the multiple DNA-binding proteins that can interact with individual cis-elements, perhaps anchored by a few key factors, could stabilize the binding of each member of the complex. Such cooperativity could operate through direct interactions of the DNA-binding factors with each other as well as through additional, as yet

unidentified proteins which would act through protein:protein interactions with the DNA-binding factors. Inducible, T-cell specific, CsA sensitive factors would be limiting for stable complex assembly. For IL-2, one cooperative interaction has been documented at the OAP40/OCT element where cooccupancy by OAP40 and oct-1 stabilizes the binding of oct-1 (Ullman et al. 1991). It will now be important to look for such interactions between the multiple elements in IL-2.

Regulated site accessibility is envisioned to involve the masking of potential binding sites near IL-2 in a repressed chromatin configuration. For example, in the mouse mammary tumor virus promoter it has been shown that the binding of glucocorticoid receptor to DNA assembled in a nucleosome facilitates the binding of NF-1 to its nearby binding site by rendering the NF-1 site accessible to its factor (Archer et al. 1991; Archer et al. 1992). In the case of IL-2, the action of inducible factors could similarly serve to open the region allowing all the proteins to interact with their binding sites. For either model, our observations that CsA treatment apparently blocks both the establishment and the maintenance of the enhancer complex highlight the CsA-sensitive factors as candidates for a chromatin opening function and/or for nucleating cooperative occupancy.

IL-2 is synthesized as a result of the T cell integrating multiple internal and external signals. We have demonstrated that stimulation of EL4 T cells triggers the coordinated occupancy of the IL-2 regulatory region and that CsA totally blocks this coordinated occupancy of the complete regulatory region *in vivo* despite affecting the ability of only a subset of factors to recognize their isolated binding sites *in vitro*. Together these observations both provide evidence that a significant amount of the signal integration involved in IL-2

induction is carried out at the level of DNA occupancy and that the observed multi-component, coordinated assembly is an essential component of this integrative process.

## MATERIALS AND METHODS:

*Cell culture*

EL4.E1.F4 thymoma cells and 32D pre-mast cells were grown as described in (Novak and Rothenberg 1990) and L cells as described in (Garrity and Wold 1992). EL4 cells and 32D cells were induced using 10 ng/ml TPA and 180 nM A23187. When inductions were performed in the presence of CsA, it was used at a concentration of 0.5 µg/ml.

*RNA preparation and measurements*

Total RNA was prepared by the guanidinium isothiocyanate/organic extraction method of (Chomczynski and Saachi 1987) and RNase protection analysis performed as previously described (McGuire and Rothenberg 1987; Miner and Wold 1991). RNA probes were as previously described for IL-2 (McGuire et al. 1988), MT-I (Mueller et al. 1988), and GAPDH (Tavtigian 1992). 5 µg of total RNA were used in the IL-2 analysis, 1 µg in the MT-I analysis, and 1 µg in the GAPDH analysis.

*In vivo and in vitro DMS/piperidine treatment of DNA*

Adherent L cells were treated as described in (Mueller et al. 1988). EL4 and 32D suspension cells in media ($10^7$ to $10^8$ in 50 ml) were pelleted at 500 g for 5 min at room temperature. Sufficient media was left behind to allow resuspension of the cell pellet in a final volume of 1 ml. Cells were transferred to a 1.5 ml microcentrifuge tube and placed in a 37°C water bath. 10µl of a freshly made 10% DMS / 90% ethanol solution were added and the sample was mixed by gentle inversion. After incubation at 37°C for 1 minute, cells were transferred to a tube containing 49 ml of ice-cold phosphate buffered

saline (PBS) and centrifuged at 500g for 5 min. at 4°C. The cell pellet was resuspended by gentle pipetting in 1 ml ice-cold PBS, 49 ml of additional ice-cold PBS were added, and the sample was centrifuged at 500g for 5 min. at 4°C. The cell pellet was resuspended in 0.3 ml ice-cold PBS, then added to 2.7 ml of lysis solution (300mM sodium chloride, 50mM Tris pH 8.0, 25 mM EDTA pH 8.0, 200µg/ml proteinase K, 0.2% SDS). DNA was prepared as described in (Mueller et al. 1988). Naked DNA preparation and *in vitro* DMS treatment was as in (Mueller et al. 1988) except 0.125% DMS for 2 minutes at room temperature was used. Subsequent piperidine cleavage was as in (Mueller et al. 1988).

*Ligation-mediated PCR visualization of genomic footprints*

LMPCR-aided DMS *in vivo* footprinting was carried out as detailed in (Garrity and Wold 1992). Oligonucleotides used in the LMPCR to detect interactions on the non-coding strand of IL-2 oligonucleotides were: primer 1) CTATCTCCTCTTGCGTTTGTCCACC; primer 2) TGTCCACCACAACAGG-CTGCTTACAGGT; primer 3) CACCACAACAGGCTGCTTACAGGTTCAG-GATG. Coding strand IL-2 oligonucleotides were: primer 1) GGACTTGAG-GTCACTGTGAGGAGTG; primer 2) CAAGGGTGATAGGCAGCTCTTCAG-CATG; primer 3) CAAGGGTGATAGGCAGCTCTTCAGCATGGGAG. LMPCR hybridization temperatures for both primer sets were: primer 1, 60°C; primer 2, 69°C; primer 3, 72°C. The coding strand MT-I oligonucleotides were: primer 1) CGGAGTAAGTGAG-GAGAAGGTACTC; primer 2) GGAGAAG-GTACTCAGGACGTTGAAG; primer 3) GAAGGTACTCAGGACGTTGA-AGTCGTGG. LMPCR hybridization temperatures were: primer 1, 60°C; primer 2, 66°C; primer 3, 69°C.

Quantitation of band intensity was performed using an LKB UltroScan XL laser densitometer and recording peak heights. Fixed and dried footprinting gels were exposed to XAR-5 film without Saran-Wrap or an intensifying screen. Multiple exposures of two entirely separate *in vivo* footprint experiments were examined, using three separate scans for each lane. Protections and hypersensitivities indicated ranged in intensity from 25% to 65% protection and from 25% to two-fold hypersensitivity in the induced EL4 cell samples compared to both naked DNA samples and uninduced EL4 cell samples and were observed in both experiments. Several G residues on the coding strand near -100 and -200 were between 20% and 25% hyper-reactive in all *in vivo* samples from non-expressing cells.

*Electrophoretic mobility shift assays*

Nuclear extracts were prepared as described (Stein et al. 1989) from the same cell populations used in the *in vivo* footprinting experiments. The mobility shift assay was performed as described in (Chen and Rothenberg 1992) using 5 μg nuclear extract, 0.5 μg poly (dI-dC), and approx. 5 fmol of the appropriate $^{32}$P-labeled oligonucleotide. Oligonucleotides were labeled by end-filling with Klenow polymerase. The sequence of one strand is given for each oligonucleotide used (bases not present in the IL-2 gene are lower case): distal CACCC (-300/-278): gatcTCTCCACCCCAAAGAGGAAAATTgatc; proximal CACCC (-72/-52): gatcACATCGTGACACCCCCATATTgatc; octamer (-100/-69): gatcTCTTTGAAAATATGTGTAATATGTAAAACATgatc; AP-1 (-161/-143): AATTCCAGAGAGTCATCAG; TGGGC (-237/-218): gatcCCACCTAAGTG-TGGGCTAACgatc; NF-κB (-211/-192): AAGAGGGAT-TTCACCTAAAT; distal NF-AT (-289/-260): AAGAGGAAAATTTGTTTC-

ATACAGAAGGCGAAT; proximal NF-AT (-147/-120): gATCAGAAGAGGA-AAAACAAAGGTAATGCgatc.

**ACKNOWLEDGEMENTS:**

The authors thank Paul Mueller, Linda Huang, Sarah Fashena, Jeff Miner, Ardem Patapoutian, Julia Yang-Snyder, Paul Boyer, and Sonya Zabludoff-Palmer for helpful comments on the manuscript.

**REFERENCES:**

Archer, T. K., M. G. Cordingley, R. G. Wolford, and G. L. Hager. 1991. Transcription factors access is mediated by accurately positioned nucleosomes on the mouse mammary-tumour virus. *Mol. Cell. Biol.* **11:** 688-698.

Archer, T. K., P. Lefebvre, P. Wolford, and G. L. Hager. 1992. Transcription factor loading on the mouse mammary-tumour virus promoter- a bimodal mechanism for promoter activation. *Science* **255:** 1573-1576.

Brabletz, T., I. Pietrowski, and E. Serfling. 1991. The immunosuppressives FK 506 and cylcosporin A inhibit the generation of protein factors binding to the two purine boxes of the interleukin 2 enhancer. *Nucl. Acids Res.* **19:** 61-67.

Briegel, K., B. Hentsch, I. Pfeuffer, and E. Serfling. 1991. One base pair change abolishes the T cell-restricted activity of a kB-like proto-enhancer element from the interleukin 2 promoter. *Nucl. Acids Res.* **19:** 5929-5936.

Brorson, K. A., B. Beverly, S. M. Kang, M. Lenardo, and R. H. Schwartz. 1991. Transcriptional regulation of cytokine genes in nontransformed-T cells: Apparent constitutive signals in run-on assays can be caused by repeat sequences. *J. Immunol.* **147:** 3601-3609.

Chen, D., and E. V. Rothenberg. 1992. Molecular basis for developmental changes in interleukin-2 gene inducibility. *Submitted* .

Chomczynski, P., and N. Saachi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162:** 156-159.

Clipstone, N. A., and G. R. Crabtree. 1992. Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation. *Nature* **357:** 695-697.

Durand, D. B., J.-P. Shaw, M. R. Bush, R. E. Replogle, R. Belagaje, and G. R. Crabtree. 1988. Characterization of antigen receptor response elements within the interleukin-2 enhancer. *Mol. Cell. Biol.* **8:** 1715-1724.

Ephrussi, A., G. M. Church, S. Tonegawa, and W. Gilbert. 1985. B lineage specific interactions of an immunoglobulin enhancer with cellular factors *in vivo*. *Science* **227:** 134-140.

Fraser, J.D., B.A. Irving, G.R. Crabtree, and A. Weiss. 1992. Regulation of interleukin-2 gene enhancer activity by the T cell accessory molecule CD28. *Science* **251:** 313-316.

Garrity, P. A., and B. J. Wold. 1992. Effects of different DNA polymerases in ligation-mediated PCR: Enhanced genomic sequencing and *in vivo* footprinting. *Proc. Natl. Acad. Sci. USA* **89:** 1021-1025.

Hentsch, B., A. Mouzaki, I. Pfeuffer, D. Rungger, and E. Serfling. 1992. The weak, fine-tuned binding of ubiquitous transcription factors to the IL-2 enhancer contributes to its T cell-restricted activity. *Nucleic Acids Research* **20:** 2657-2665.

Herrera, R. E., P. E. Shaw, and A. Nordheim. 1989. Occupation of the *c-fos* serum response element *in vivo* by a multiprotein complex is unaltered by growth factor induction. *Nature* **340:** 68-70.

Jain, J., V. E. Valge-Archer, and A. Rao. 1992. Analysis of the AP-1 sites in the IL-2 promoter. *J. Immunol.* **148:** 1240-1250.

Kamps, M. P., L. Corcoran, J. H. LeBowitz, and D. Baltimore. 1990. The promoter of the human interleukin-2 gene contains two octamer-binding sites and is partially activated by the expression of Oct-2. *Mol. Cell. Biol.* **10:** 5464-5472.

Kara, C. J., and L. H. Glimcher. 1991. *In vivo* footprinting of MHC class II genes: bare promoters in bare lymphocyte syndrome. *Science* **252:** 709-712.

Kronke, M., W. J. Leonard, J. M. Depper, and W. C. Greene. 1985. Sequential expression of genes involved in human T lymphocyte growth and differentiation. *J. Exp. Med.* **161:** 1593-1598.

Mattila, P. S., K. S. Ullman, S. Fiering, E. A. Emmel, M. McCutcheon, G. R. Crabtree, and L. A. Herzenberg. 1990. The actions of cyclosporin A and FK506 suggest a novel step in the activation of T lymphocytes. *EMBO J.* **9:** 4425-4433.

McGuire, K. L., and E. V. Rothenberg. 1987. Inducibility of interleukin-2 RNA expression in individual mature and immature T lymphocytes. *EMBO J.* **6:** 939-946.

McGuire, K. L., J. A. Yang, and E. V. Rothenberg. 1988. Influence of activating stimulus on functional phenotype: interleukin-2 mRNA accumulation differentially induced by ionophore and receptor ligands in subsets of murine T cells. *Proc. Natl. Acad. Sci. USA* **85:** 6503-6507.

Miner, J. H., and B. J. Wold. 1991. *c-myc* inhibition of myoD and myogenin-initiated myogenic differentiation. *Mol. Cell. Biol.* **11:** 2842-2851.

Mueller, P. R., S. J. Salser, and B. Wold. 1988. Constitutive and metal-inducible protein:DNA interactions at the mouse metallothionein-I

promoter examined by *in vivo* and *in vitro* footprinting. *Genes and Dev.* **2:** 412-427.

Mueller, P. R., and B. Wold. 1989. *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* **246:** 780-786.

Novak, T. J., and E. V. Rothenberg. 1990. cAMP inhibits induction of interleukin 2 but not of interleukin 4 in T cells. *Proc. Natl. Acad. Sci.* **87:** 9353-9357.

Novak, T. J., P. M. White, and E. V. Rothenberg. 1990. Regulatory anatomy of the murine interleukin-2 gene. *Nucl. Acids Res.* **18:** 4523-4533.

O'Keefe, S. J., J. Tamura, R. L. Kincaid, M. J. Tocci, and E. A. O'Neill. 1992. FK-506 and CsA-sensitive activation of the interleukin-2 promoter by calcineurin. *Nature* **357:** 692-694.

Randak, C., T. Brabletz, M. Hergenrother, I. Sobotta, and E. Serfling. 1990. Cyclosporin A suppresses the expression of the interleukin 2 gene by inhibiting the binding of lymphocyte-specific factors to the IL-2 enhancer. *EMBO J.* **9:** 2529-2536.

Rothenberg, E. V. 1992. The development of functionally responsive T cells. *Adv. Imm.* **51:** 85-214

Serfling, E., R. Barthelmas, I. Pfeuffer, B. Schenk, S. Zarius, R. Swoboda, F. Mercurio, and M. Karin. 1988. Ubiquitous and lymphocyte-specific factors are involved in the induction of the mouse interleukin-2 gene in T lymphocytes. *EMBO J.* **8:** 465-473.

Shaw, J.-P., P. J. Utz, D. B. Durand, J. J. Toole, E. A. Emmel, and G. R. Crabtree. 1988. Identification of a putative regulator of early T cell activation genes. *Science* **241:** 202-205.

Small, A., A. Blin, and M. Levine. 1992. Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *EMBO J.* **11:** 4047-4057.

Stein, B., H. J. Ramsdorf, A. Steffen, M. Liffin, and P. Herrlich. 1989. UV-induced DNA damage is an intermediate step in UV-induced expression of human immunodeficiency virus type 1, collagenase, c-fos, and metallothionein. *Mol. Cell. Biol.* **9:** 5169-5181.

Tavtigian, S. V. Thesis. California Institute of Technology, 1992.

Thompson, C. B., C.-Y. Wang, I.-C. Ho, P. R. Bohjanen, B. Petryniak, C. H. June, S. Miesfeldt, L. Zhang, G. J. Nabel, B. Karpinski, and J. M. Leiden. 1992. Cis-acting sequences required for inducible interleukin-2 enhancer function bind a novel Ets-related Protein, Elf-1. *Mol. Cell. Biol.* **12:** 1043-1053.

Thompson, C. C., and S. L. McKnight. 1992. Anatomy of an enhancer. *Trends Genet.* **8:** 232-236.

Ullman, K. S., W. M. Flanagan, C. A. Edwards, and G. R. Crabtree. 1991. Activation of early gene expression in T lymphocytes by Oct-1 and an inducible protein, $OAP^{40}$. *Science* **254:** 558-562.

Ullman, K. S., J. P. Northrop, C. L. Verweij, and G. R. Crabtree. 1990. Transmission of signals from the T lymphocyte antigen receptor to the genes responsible for cell proliferation and immune function: The missing link. *Annu. Rev. Immunol.* **8:** 421-452.

Verweij, C. L., C. Guidos, and G. R. Crabtree. 1990. Cell type specificity and activation requirements for NFAT-1 (Nuclear Factors of Activated T-cells) transcriptional activity determined by a new method using transgenic mice to assay transcriptional activity of an individual nuclear factor. *J. Biol. Chem.* **265:** 15788-15795.

Williams, T.M., D. Moolten, J. Burlein, J. Romano, R. Bhaerman, A. Godillot, M. Mellon, F.J. Rauscher III, J. A. Kant. 1991. Identification of a zinc finger protein that inhibits IL-2 gene expression. Science **254:** 1791-1794.

**FIGURE 1:** RNase protection analysis of interleukin-2 (IL-2), metallothionein-I (MT-I), and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) transcripts in RNA from the *in vivo* footprinted cell populations. RNA from unstimulated EL4 cells (lane 1), EL4 cells stimulated for 7 hours with 10 ng/ml TPA and 180 nM A23187 (lane 2), EL4 cells stimulated for 7 hours in the presence of 0.5 μg/ml CsA (lane 3), unstimulated 32D cells (lane 4), and 32D cells stimulated with 10 ng/ml TPA and 180 nM A23187 for 2 hours (lane 5) was used. IL-2 RNA was also easily detected using RNA from EL4 cells stimulated for 2 hours (data not shown). Lane 6 contained only 30μg of tRNA. 5 μg of total RNA plus 30μg of tRNA were used in the IL-2 analysis, 1 μg plus 30μg of tRNA in the MT-I analysis, and 1 μg plus 30μg of tRNA in the GAPDH analysis.

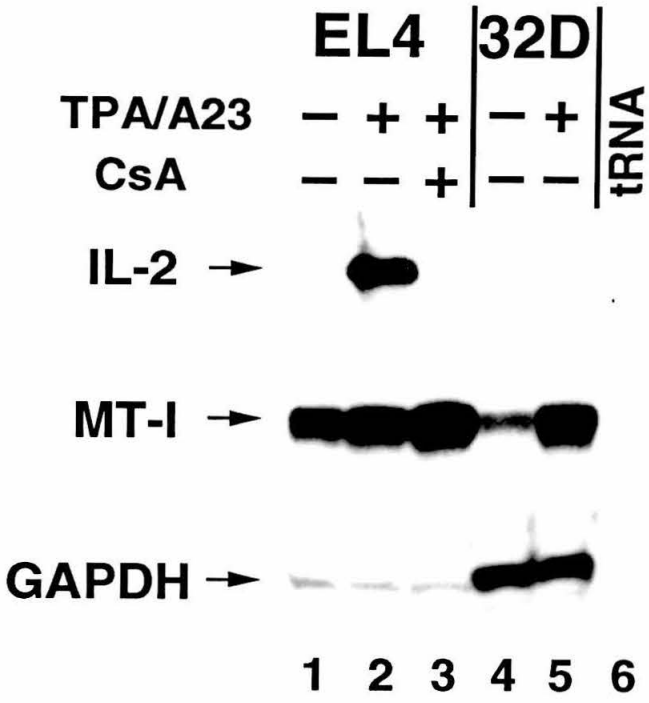

EL4
32D
TPA/A23 – + + – +
CsA – – + – –
tRNA
IL-2
MT-I
GAPDH
1 2 3 4 5 6

**FIGURE 2:** Characterization of nuclear factors present in EL4 cells by electrophoretic mobility shift assay. Nuclear extracts were prepared from the same cell populations used in the *in vivo* footprinting experiments. The $^{32}$P-labeled oligonucleotide used in each experiment is indicated at the top of each panel. Arrowheads and brackets mark the major binding activities, as discussed in the text. The major complex for either CACCC motif can be competed by the oligonucleotide containing the other CACCC motif or, even more efficiently, by a consensus SP1 oligonucleotide (8, data not shown). The unmarked, inducible lower complex in lane 2 can be competed by the distal NF-AT oligonucleotide and likely represents binding of factors to the portion of the distal NF-AT site present in the distal CACCC oligonucleotide. Additional, faster migrating complexes was detected in lanes 4 to 6. None were CsA sensitive, and the most prevalent species was constitutively expressed. In lanes 7 to 9, OAP40 binding activity could not be detected in these experiments, likely due to the abundance of octamer binding species in EL4 cell extracts. In lanes 16 to 18, a faster migrating, minor NF-κB site binding species was detected which was inducible and partially CsA-sensitive like the major species. In lanes 19 to 24, in addition to the well-characterized, inducible, CsA-sensitive NF-AT complex (filled arrowheads), a faster migrating species with similar regulatory features (open arrowheads) was observed for both NF-AT sites. Additional, constitutive NF-AT motif binding activities were also detected in these EL4 cells.

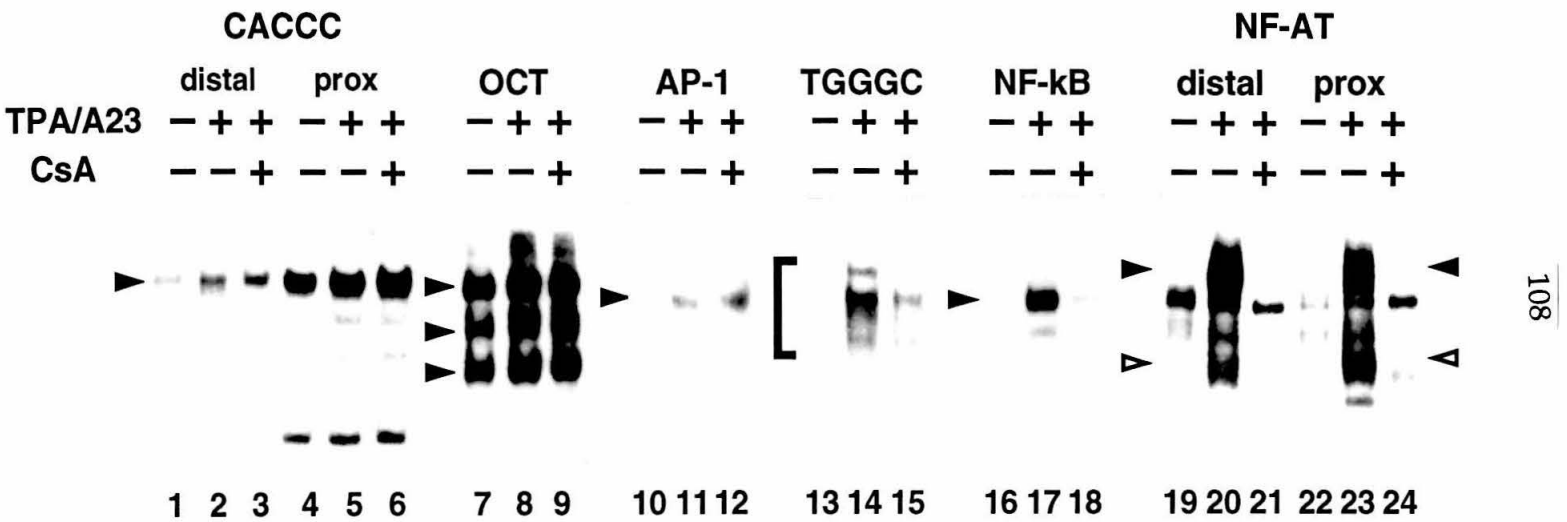
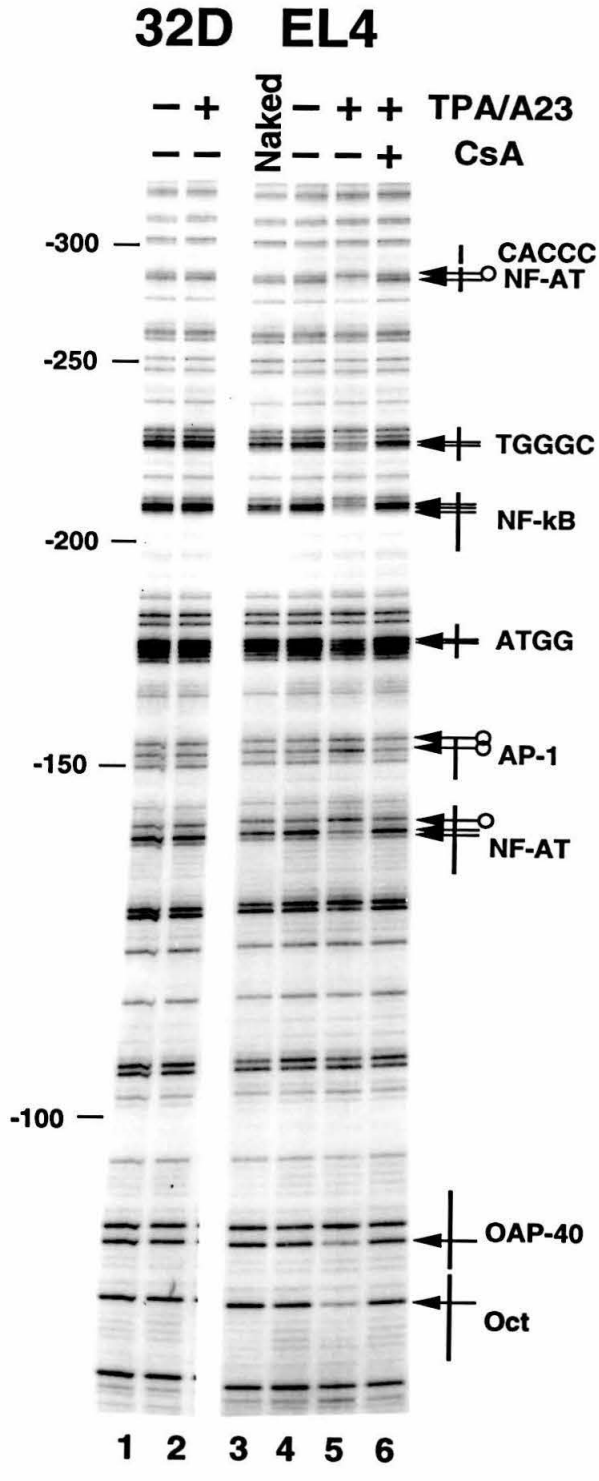

CACCC
distal
prox
OCT
AP-1
TGGGC
NF-kB
NF-AT
distal
prox
TPA/A23
CsA
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

**FIGURE 3:** DMS *in vivo* footprint of the IL-2 regulatory region. DNA from *in vivo* DMS-treated unstimulated 32D cells (panel A, lane 1; panel B, lane 2), stimulated 32D cells (panel A, lane 2; panel B, lane 3), unstimulated EL4 cells (panel A, lane 4; panel B, lane 5), stimulated EL4 cells (panel A, lane 5; panel B, lane 6), and EL4 cells stimulated in the presence of CsA (panel A, lane 6; panel B, lane 7) was used. Naked 32D DNA (panel B, lane 1) and naked EL4 DNA (panel A, lane 4, panel B lane 3), each treated with DMS *in vitro*, were also used. **A)** shows the *in vivo* footprint pattern of the coding strand and **B)** the *in vivo* footprint pattern of the non-coding strand. The position in the sequence ladder with respect to the major start site of IL-2 transcription is indicated on the left hand side of each panel. Noted on the right-hand side are bands which were reproducibly more than 25% protected or more than 25% hypersensitive in induced EL4 cells when compared to uninduced EL4 cells and to naked EL4 DNA. Arrows mark protections and arrows with a circle at the end mark hypersensitivities. Certain weak interactions apparent in the experiment shown were not marked (e.g., a protection near -100 on the coding strand) as they were not reproducibly observed over the multiple independent experiments performed. An artifactual band not predicted from the primary sequence of IL-2 is marked with an asterisk and is specific for the non-coding strand primer set. It became 5 to ten-fold more intense than any other band in the ladder under LMPCR conditions which gave generally high background (using excess Vent polymerase or hybridization temperatures that were too low, data not shown). Its intensity correlated in no way with the identity of the DNA being used, but, instead, with the intensity of an apparent oligonucleotide primer dimer band (data not shown).


32D EL4
Naked
TPA/A23
CsA
-300
-250
-200
-150
-100
CACCC
NF-AT
TGGGC
NF-kB
ATGG
AP-1
NF-AT
OAP-40
Oct
1 2 3 4 5 6

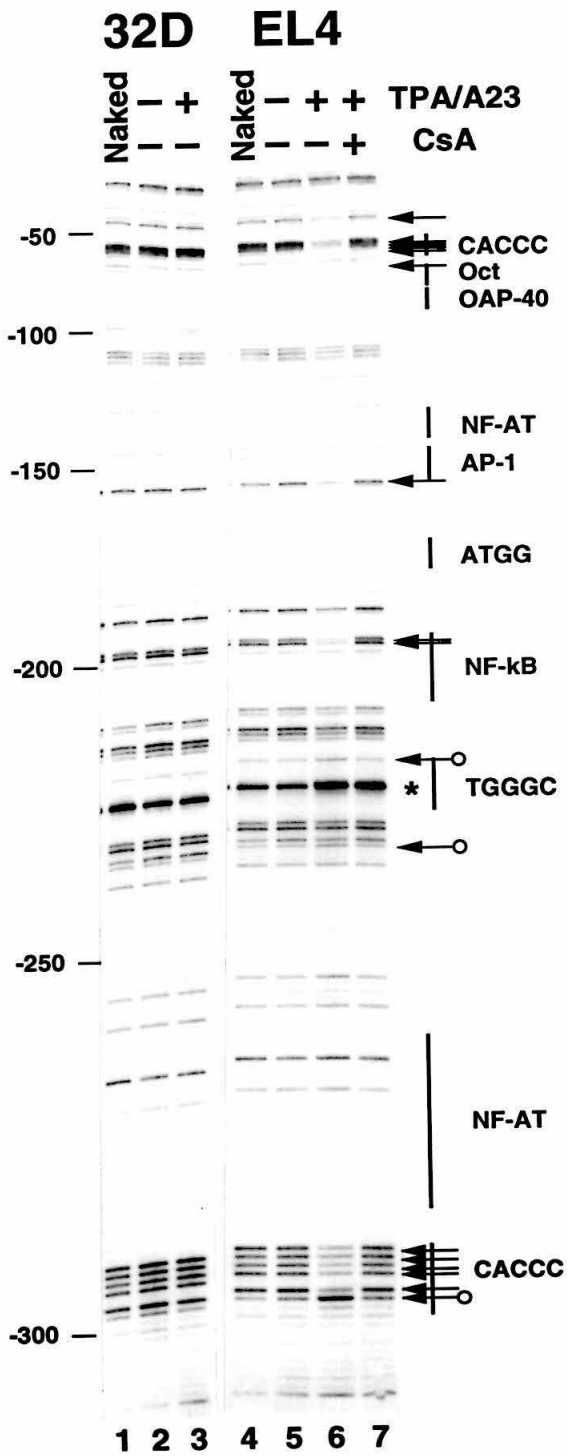

32D
EL4
Naked
– + TPA/A23
– – CsA
Naked
– + +
– – +
-50
-100
-150
-200
-250
-300
CACCC
Oct
OAP-40
NF-AT
AP-1
ATGG
NF-kB
*
TGGGC
NF-AT
CACCC
1 2 3 4 5 6 7

**FIGURE 4:** Summary of *in vivo* protein:DNA interactions detected in stimulated EL4 T cells. Guanines protected from reaction with DMS in stimulated EL4 T cells are indicated by downward pointing arrows and guanines made hypersensitive to DMS in stimulated EL4 T cells are indicated by upward pointing arrows.

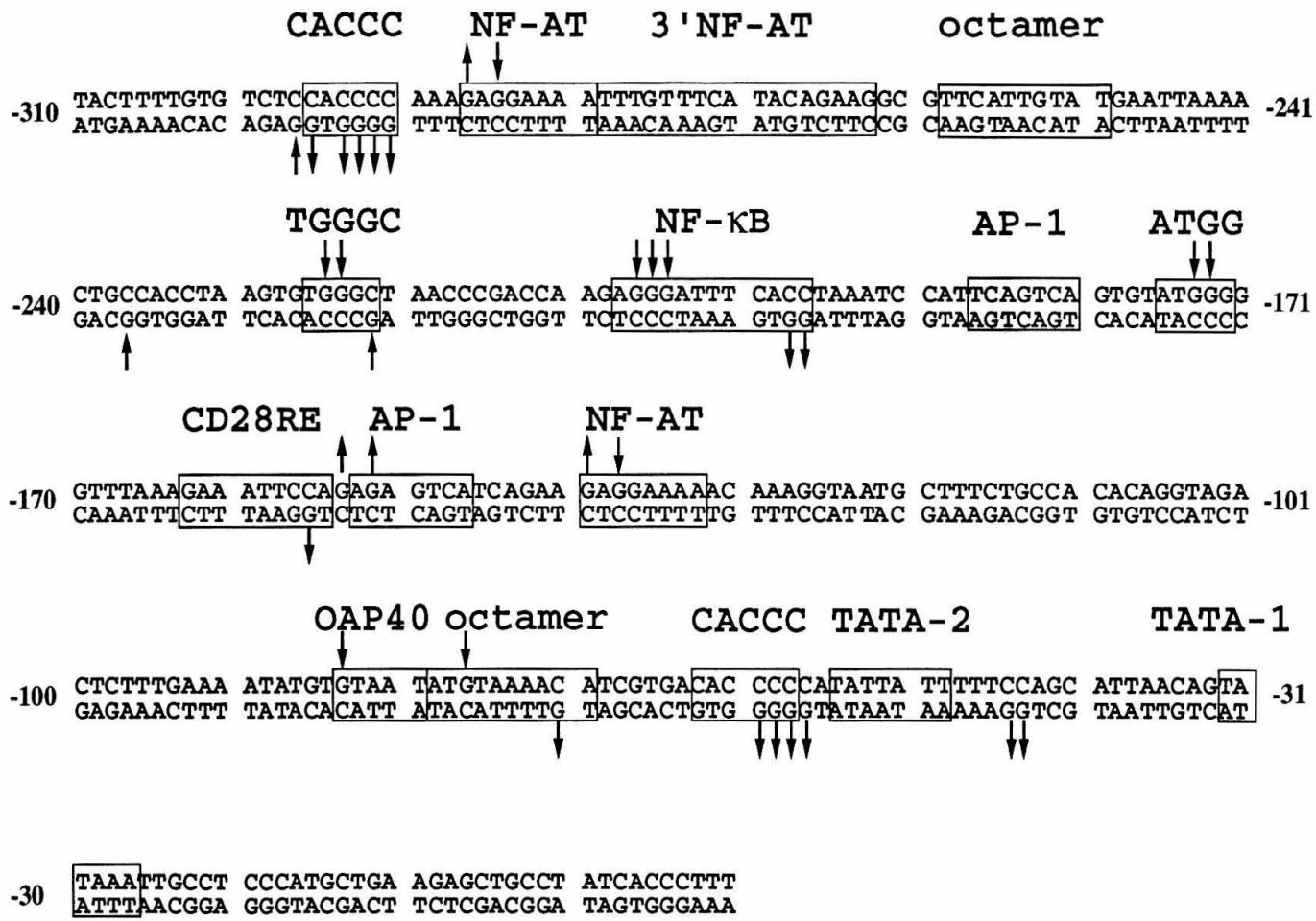

CACCC
NF-AT
3'NF-AT
octamer
-310
TACTTTTGTG TCTCCACCCC AAAGAGGAAA ATTTGTTTCA TACAGAAGGC GTTCATTGTA TGAATTAAAA
ATGAAAACAC AGAGGTGGGG TTTCTCCTTT TAAACAAAGT ATGTCTTCCG CAAGTAACAT ACTTAATTTT
-241
TGGGC
NF-κB
AP-1
ATGG
-240
CTGCCACCTA AGTGTGGGCT AACCCGACCA AGAGGGATTT CACCTAAATC CATTCAGTCA GTGTATGGGG
GACGGTGGAT TCACACCCGA TTGGGCTGGT TCTCCCTAAA GTGGATTTAG GTAAGTCAGT CACATACCCC
-171
CD28RE
AP-1
NF-AT
-170
GTTTAAAGAA ATTCCAGAGA GTCATCAGAA GAGGAAAAAC AAAGGTAATG CTTTCTGCCA CACAGGTAGA
CAAATTTCTT TAAGGTCTCT CAGTAGTCTT CTCCTTTTTG TTTCCATTAC GAAAGACGGT GTGTCCATCT
-101
OAP40 octamer
CACCC
TATA-2
TATA-1
-100
CTCTTTGAAA ATATGTGTAA TATGTAAAAC ATCGTGACAC CCCCATATTA TTTTTCCAGC ATTAACAGTA
GAGAAACTTT TATACACATT ATACATTTTG TAGCACTGTG GGGGTATAAT AAAAAGGTCG TAATTGTCAT
-31
-30
TAAATTGCCT CCCATGCTGA AGAGCTGCCT ATCACCCTTT
ATTTAACGGA GGGTACGACT TCTCGACGGA TAGTGGGAAA

**FIGURE 5:** DMS *in vivo* footprints over representative binding sites at 2 hours and 5 hours of induction. DNA from *in vivo* DMS-treated uninduced EL4 cells (lane 1), EL4 cells induced for 2 hours (lane 2), and EL4 cells induced for 5 hours (lane 3) was used. The entire set of interactions shown in figure 2 were seen in this experiment. All were of equal intensity in both induced samples, but only a subset are shown here to conserve space. Lane 2 was exposed for 20% longer than lanes 1 and 3 to compensate for underloading of that lane.

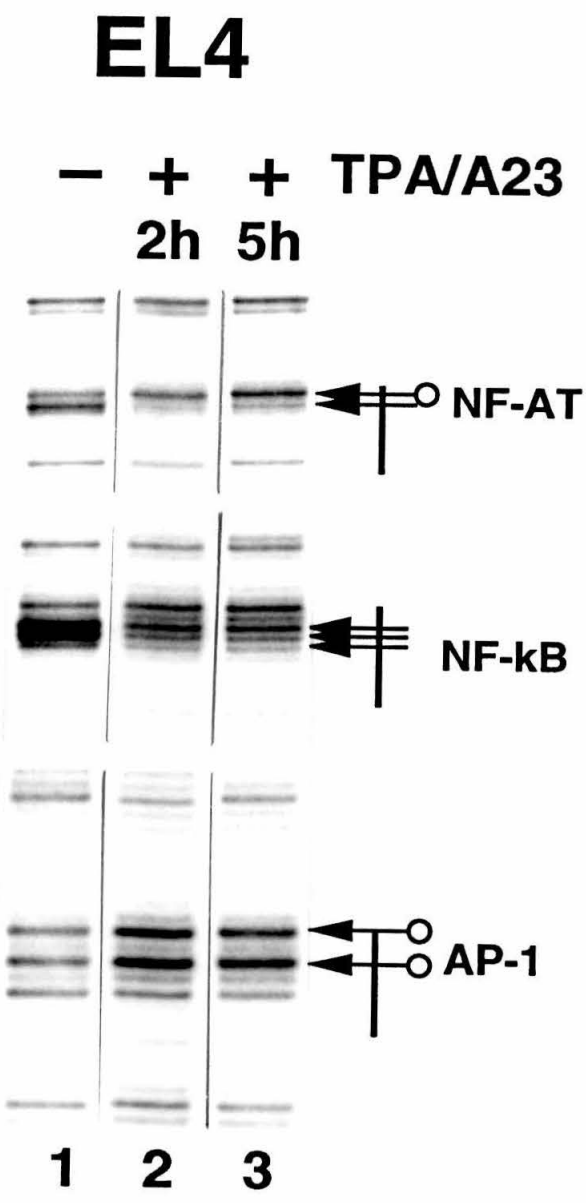

EL4
− + + TPA/A23
2h 5h
NF-AT
NF-kB
AP-1
1 2 3

**FIGURE 6:** DMS *in vivo* footprint of the SP1 binding site in the metallothionein-I (MT-I) gene. DNA from *in vivo* DMS-treated unstimulated 32D cells (lane 2), stimulated 32D cells (lane 3), unstimulated EL4 cells (lane 5), stimulated EL4 cells (lane 6), and EL4 cells stimulated in the presence of CsA (lane 7) was used. Naked 32D DNA (lane 1) and naked EL4 DNA (lane 3), each treated with DMS *in vitro*, were also used.

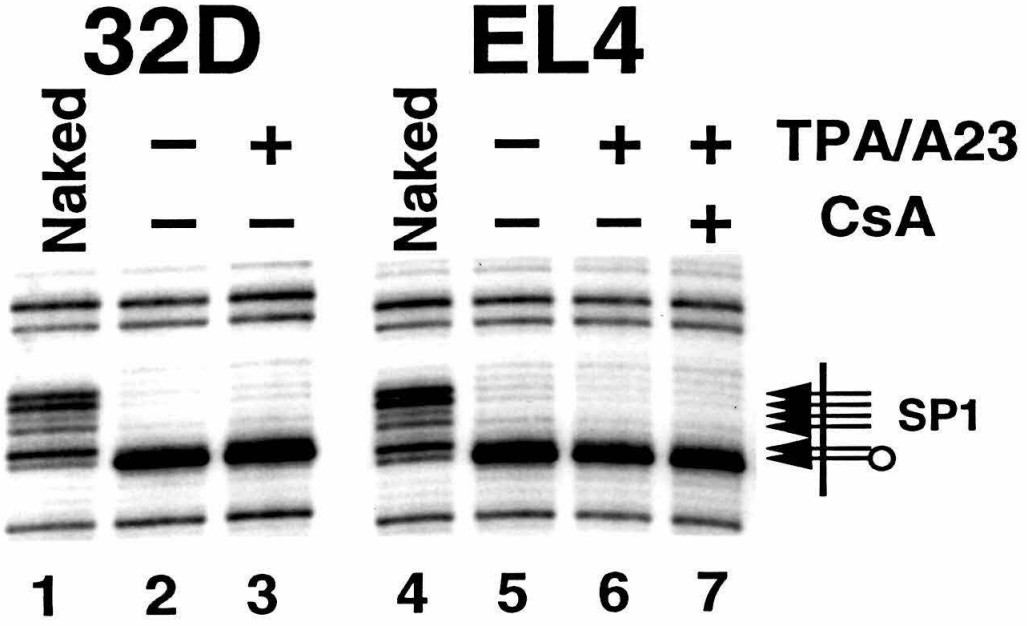
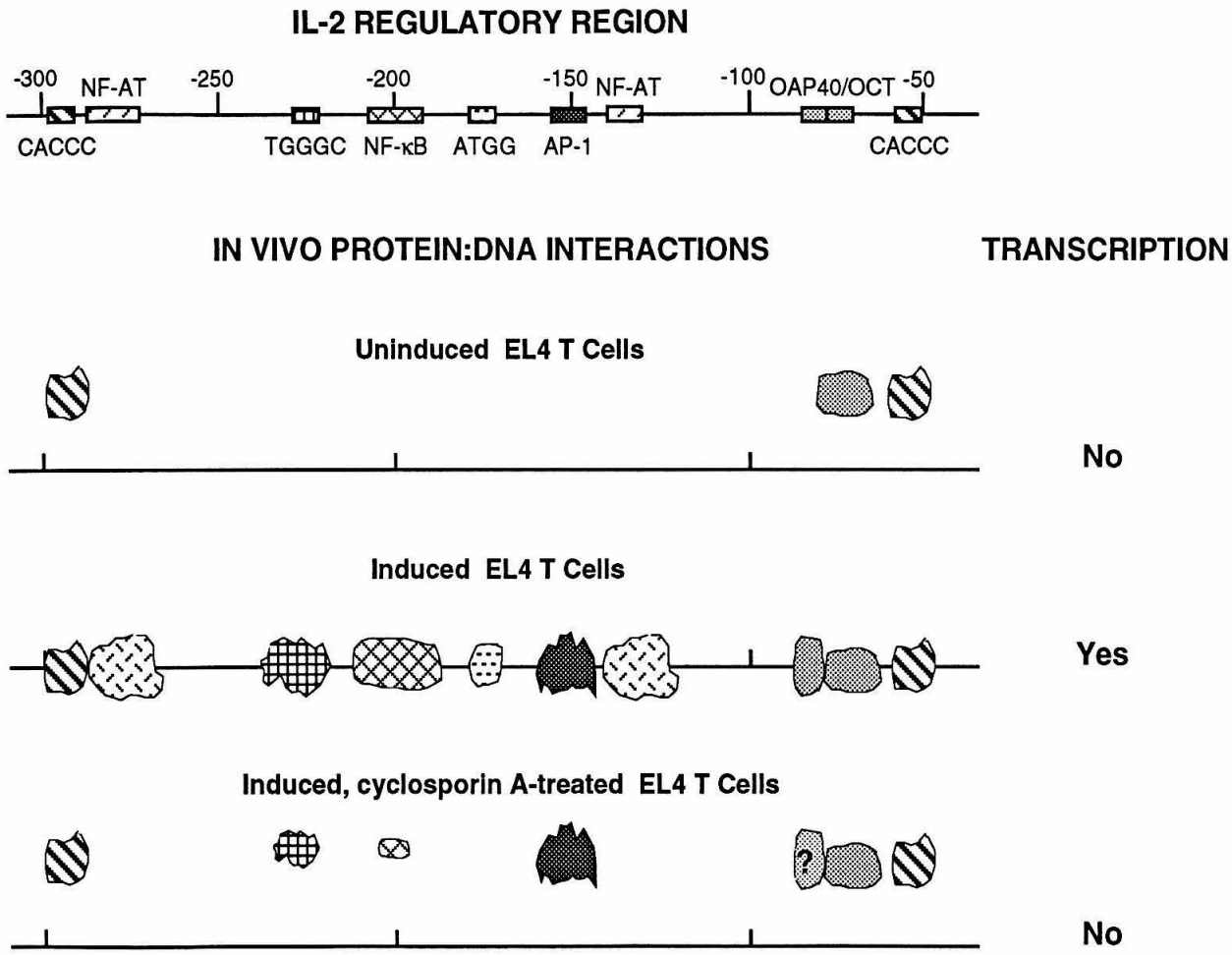

32D
EL4
Naked
Naked
−
+
−
+
+
TPA/A23
−
−
−
−
+
CsA
SP1
1
2
3
4
5
6
7

**FIGURE 7:** Summary of the protein:DNA interactions detected EL4 T cells by *in vivo* footprint analysis and the presence of nuclear factors in these cells as determined through *in vitro* DNA-binding analysis. The decreased size of the TGGGC and NF-κB recognizing activities in stimulated, CsA-treated EL4 T cells signifies the quantitative decrease in *in vitro* DNA-binding activity seen in nuclear extracts from these cells in figure 2. Though OAP40 binding activity could not be detected in figure 2, its presence in stimulated EL4 cells was inferred because of the *in vivo* methylation protection of a guanine residue adjacent to the octamer site which specifically disrupts OAP40 binding when methylated (Ullman et al. 1991). Stimulation-dependent OAP40 binding activity has been detected in the human Jurkat T cell line, but its CsA sensitivity was not examined (Ullman et al. 1991). Binding of a specific-factor to the ATGG motif in stimulated EL4 cells was also inferred from *in vivo* footprint data.


IL-2 REGULATORY REGION
-300
NF-AT
-250
-200
-150
NF-AT
-100
OAP40/OCT
-50
CACCC
TGGGC
NF-κB
ATGG
AP-1
CACCC
IN VIVO PROTEIN:DNA INTERACTIONS
TRANSCRIPTION
Uninduced EL4 T Cells
No
Induced EL4 T Cells
Yes
Induced, cyclosporin A-treated EL4 T Cells
?
No

# Appendix

## *In vivo* footprinting of the myocyte-specific promoter of the nicotinic δ-AChR

**INTRODUCTION:**

The delta-subunit of the nicotinic acetylcholine receptor (δ-AChR) is specifically expressed in differentiated skeletal muscle cells. The region between 148 bp upstream and 24 bp downstream of the start site of transcription of the δ-AChR gene contains a minimal enhancer/promoter region which is sufficient to drive a reporter gene in the appropriate pattern of cell-type-specific expression (Baldwin and Burden 1988). C2C12 myoblasts stably transformed with this promoter/reporter fusion show a greater than 100-fold increase in expression of the reporter upon differentiation (Baldwin and Burden 1988). This parallels the increase in δ-AChR RNA in these cells upon differentiation (Yu et al. 1986). As the promoter/reporter fusion RNA contained 24 nucleotides of δ-AChR RNA sequence, there was still a possibility of a large post-transcriptional component to this regulation. However, this region of the rat δ-AChR promoter, which is approximately 90% homologous to the mouse promoter, supports differentiated skeletal muscle-specific expression in C2C12 cells when only 4 nucleotides of δ-AChR RNA sequence are included in the reporter RNA (Chahine et al. 1992). Thus, it seems certain that the expression pattern of the reporter constructs are regulated at the transcriptional level and that the δ-AChR promoter activates transcription only in differentiated myocytes.

The structure of the δ-AChR DNA regulatory region has several characteristics that make it an interesting target for an analysis of the mechanism used in its transcriptional induction *in vivo*. It contains one binding site recognized *in vitro* by MyoD (the site is termed an E-box), but this site is located in the core promoter region 17 bp upstream of the transcriptional initiation site (Baldwin and Burden 1989). Most potential binding sites for

positively acting gene-specific regulatory factors are not located in the core promoter region, though in some cases repressors bind there and inhibit transcriptional initiation by directly competing with the general transcriptional apparatus of the cell for DNA-binding. However, the E-box, based on its role in a number of other regulatory regions functional in skeletal muscle (Weintraub et al. 1991), would naively be considered a site which would work to activate the expression of a gene such as δ-AChR. I wanted to determine whether this element was occupied *in vivo*, and if so, whether it was occupied only in the myoblast, perhaps acting as a repressor, or only in the myocyte, where it might substitute for a TATA box. This was conceivable as the δ-AChR core promoter possessed only a poor TATA homology, TAAA, in the TATA region. Instead of a clear TATA, however, the δ-AChR core promoter region contained an initiator (Inr) element which is homologous at 8 of 8 positions to the Inr found in the terminal deoxynucleotidyl transferase (TdT) gene. The Inr element of the TdT can function by itself, without a TATA box, to fix the start site and direction of transcriptional initiation (Smale and Baltimore 1989). The Inr element functions by binding the general transcription factor TFII-I, which then recruits TFIID to the core promoter even in the absence of a functional TATA element (Roy et al. 1991). In addition, TFII-I has also been shown to be capable of interacting with E-boxes (Roy et al. 1991). I was interested in studying how the components of the rest of the DNA regulatory region would interact with the core promoter and what the pattern of occupancy of an Inr driven core promoter would be *in vivo*. Finally, I wanted to compare mechanisms by which this regulatory region was regulated with that of the muscle creatine kinase (MCK) enhancer, which shows the same pattern of transcriptional activity in these cells. Occupancy of the DNA regulatory region

by gene-specific regulatory factors appears to be the rate-limiting step in MCK enhancer function in these cells (Mueller and Wold 1989). It is show here that the δ-AChR promoter is regulated in an entirely different fashion.

**RESULTS AND DISCUSSION:**

The pattern of *in vivo* protein:DNA interactions at the δ-AChR promoter was examined by dimethyl sulfate (DMS) *in vivo* footprinting as previously described (Garrity and Wold 1992) Two different mouse skeletal myoblast cell lines were examined, C2C12 and MM14. These two lines express many of the same proteins upon differentiation, though they express different combinations of the MyoD family of myogenic regulatory proteins. Proliferating C2C12 myoblasts express myf-5 and MyoD and, when differentiation is stimulated, myogenin and MRF-4 are also expressed (Mueller and Wold 1989; Miner and Wold, 1990). Proliferating MM14 myoblasts express only MyoD, and upon differentiation express myogenin as well (Mueller and Wold 1989; Miner and Wold, 1990). Thus, although both express many of the same skeletal muscle-specific proteins, the mechanism used to initiate their synthesis might be different in the two cell types.

The pattern of *in vivo* protein:DNA interactions in these cell were first examined at the regulatory region of one constitutively expressed gene, the mouse metallothionein-I promoter, and one regulatory region that is only active in differentiated muscle cells, the MCK enhancer. Footprints should be observed in all samples at the constitutively expressed MT-I gene. This serves as a positive control to ensure that should a different regulatory region lack detectable interactions, the absence of interactions is gene-specific and hence not an artifact of the procedure. The examination of the MCK enhancer, which

functions only in differentiated C2C12 and MM14 cells, helps gauge the ability to see differences in interactions between the undifferentiated and differentiated cell types. The MM14 samples were previously examined at both the MT-I and MCK regulatory regions (Mueller and Wold 1989; Garrity and Wold 1992). As previously described, interactions were detected at the MT-I promoter in both myoblasts and myocytes, while interactions at the MCK enhancer were only detected in differentiated myocytes. The MT-I promoter and the MCK enhancer were also examined in the C2C12 samples and the patterns of protein:DNA interactions were both identical to those seen in the MM14 samples (data not shown). This indicates that protein:DNA interactions can be detected in all samples, and that the overall mechanism of activation of the MCK enhancer and the use of individual regulatory elements within the MCK enhancer are the same in both C2C12 and MM14 cells, even though they contain different combinations of MyoD family myogenic regulators. The MCK enhancer contains two E-boxes which are occupied by each of the members of the MyoD family *in vitro* and which are occupied *in vivo* (Mueller and Wold 1989). Either a common member of the MyoD family, such as myogenin, recognizes these elements in both cell types or different regulatory factors occupy the site indistinguishably *in vivo*.

The pattern of *in vivo* protein:DNA interactions at the δ-AChR promoter were then characterized in these two cell types. G-residues on the coding strand from 25 bp downstream to beyond 250 bp upstream of the start site of transcription were examined for interactions as well as G-residues on the non-coding strand from 164 bp upstream to beyond 100 bp downstream of the start site of transcription. This covers the entire minimal promoter region and additional regions both upstream of the promoter and in the coding region

of the gene. Thus all the interactions at the minimal promoter and in the flanking regions that can be detected with DMS have been examined. Of the 95 G-residues present in the 172 bp minimal promoter, *in vivo* interactions were detected at 26. Surprisingly, and in stark contrast to the pattern at the MCK enhancer in these cells, both MM14 myoblasts and MM14 myocytes showed multiple regions of occupancy (figure 1). The pattern of protein:DNA interactions was identical in intensity as well as in character in both myoblasts and myocytes in MM14 cells (figure 1). Identical results were obtained in C2C12 myoblasts and myocytes (data not shown). Furthermore, the pattern of interactions in C2C12 and MM14 cells were indistinguishable (figure 1 and data not shown). Thus the δ-AChR promoter activates transcription only in myocytes, though it appears to be identically occupied in both myoblasts and myocytes. Thus the induction of δ-AChR transcription is mediated by changing the activity of a complex that has already formed on its promoter. The observed protections are for the most part extremely strong, indicating that the complex has formed on the large majority of δ-AChR promoters in the population. No changes in the complex formed on the δ-AChR promoter were detected upon differentiation. However, there may be alterations in the pattern of protein:DNA interactions to which DMS is insensitive, such as minor groove interactions or binding of proteins to sites in which there are no G residues (though the latter is unlikely as the δ-AChR promoter contains no area with more than four consecutive A's and T's and only one with more than three). In addition, myocyte-specific proteins might replace myoblast-specific proteins without altering the DMS footprint. However, it is clear that there is a multiprotein complex assembled on the δ-AChR promoter prior to its activation.

Inspection of the pattern of DMS protection and enhancement seen in the comparison of naked muscle cell DNA treated with DMS *in vitro* and muscle cell DNA treated with DMS *in vivo* revealed three regions of protein:DNA interaction (MM14 in figure 1, C2C12 data not shown). One region of 20 bp extended from the downstream edge of the E-box, up to the TAAA site, and, though more weakly, to a position 4 bp upstream of the TAAA. A second set of interspersed interactions were detected over a 17 bp region centered near -60. A third set of interactions was detected near -100. This element contains a 9 for 9 match with a site in the human α-cardiac actin promoter. That site can recognize SP1 *in vitro* and *in vivo* (as shown by activation through SP1 expression in *Drosophila* cells), and is necessary but not sufficient for the muscle-specific activation of that promoter in collaboration an E-box and an element recognizing the serum response factor (Sartorelli et al. 1990). This element is also homologous to the AP-2 element occupied in the MCK enhancer (Mueller and Wold 1989). Though there are six G-residues in the 10 bp surrounding the Inr element, no interactions can be detected at this site. TFII-I could recognize this site *in vivo* and escape detection in this assay if it bound exclusively in the minor groove or if by virtue of dissociating from the core promoter after every initiation event it effectively occupied this site only a small fraction of the time. Nonetheless, there is no evidence of its binding. In addition, there is no evidence for the binding of factors to the three TGGCCT motifs present between -108 and -148 which were shown to bind a myocyte-specific factor *in vitro* (Baldwin and Burden 1989). No other interactions can be detected, either upstream of the minimal promoter nor in the coding region.

These results show that the δ-AChR promoter is occupied by regulatory factors both prior to and after it is transcriptionally active. The pattern of

occupancy in the myoblast is indistinguishable from that in the myocyte. Thus the transcription of the δ-AChR gene is not limited by the repressive effects of chromatin structure or the stable binding of factors to its regulatory region. Instead it is limited by the activity of the assembled DNA:protein complex. The E-box which overlaps its probably non-functional TATA-box is constitutively occupied. The large size of the region of contiguous occupancy suggests that multiple proteins bind adjacent to one another at this site. The nature of this complex and whether it may be functioning as a component of the core promoter, perhaps by interacting with TFII-I, or as an upstream regulator of initiation through the Inr, through binding of gene-specific factors such as MyoD, is an interesting, but currently unexplored question. The cell-type distribution of this complex is currently unknown. Investigations in which the δ-AChR locus is examined during myogenic conversion of non-muscle cells with MyoD might suggest whether the assembly of this complex is a marker of myogenic commitment. More technically difficult, but ultimately most illuminating, would be the examination of this promoter in different populations of developing skeletal muscle cells from a developing mouse. We have examined the pattern of protein:DNA interactions in two cultured skeletal myoblast cell lines, C2C12 cells, in which transcription of the δ-AChR gene has been shown to be activated in response to differentiation, and in MM14 cells, in which such experiments have not yet been performed. The identical *in vivo* footprint patterns in the two cell lines suggests that the promoter is functioning in a similar fashion in MM14 cells as well. However, only in the C2C12 cells can we be certain that the promoter is occupied before transcription is activated.

The absence of transcription from the assembled complex might reflect either an inability for the core promoter to assemble the general transcriptional machinery or the formation of inactive complexes of general factors, such as the inert TFIID/NC2 complexes observed *in vitro* (Meisterernst and Roeder 1991). A paused polymerase is an unlikely explanation, as sequences in the transcribed region appear to mediate this function in other genes and the transcribed region of δ-AChR is not required for the differentiation-specificity of this promoter (Giardina et al. 1992; Chahine et al. 1992). Therefore, we have shown that the transcriptional activation of δ−AChR upon muscle differentiation occurs either through a replacement of factors bound at its promoter or through alterations in the ability of bound factors to activate transcription of this gene.

## MATERIALS AND METHODS:

*Cell culture*

C2C12 cells were grown on collagen-coated tissue culture dishes in Dulbecco's modified Eagle's medium (DMEM) with 20% fetal bovine serum and 2 mM glutamine. Differentiation was initiated by feeding cells with switch medium (DMEM containing 2% horse serum and 2 mM glutamine). After 24 hours in switch medium cells were refed switch medium containing $10^{-5}$M cytosine arabinoside and allowed to differentiate 48 hours more prior to *in vivo* footprinting. The MM14 DNA's were the gift of Paul Mueller.

*LMPCR*

LMPCR was performed using the conditions described in Garrity and Wold, 1992. δ-AChR primers used were as follows: Non-coding strand primers: 1) AGAGGATTCGCCCCCCACCCTCT; 2)ACCCTCTGCCACCCTACCCCACC-AG; 3)CCTACCCCACCAGGCCCCACCCTACCA. Coding strand primers: 1) TGTGAGCACAGGCCCTGCCAT; 2)CCTGCCATCCCCCACCTCTGACT; 3)C-CTGCCATCCCCCACCTCTGACTGAC.

## ACKNOWLEDGEMENT:

I wish to thank Paul Mueller for his generous gift of MM14 DNA.

**REFERENCES:**

Baldwin, T. J., and S. J. Burden. 1988. Isolation and characterization of the mouse acetylcholine receptor delta subunit gene: identification of a 148-bp cis-acting region that confers myotube-specific expression. *J. Cell. Biol.* **107:** 2271-2279.

Baldwin, T. J., and S. J. Burden. 1989. Muscle-specific gene expression controlled by a regulatory element lacking a MyoD1-binding site. *Nature* **341:** 716-720.

Chahine, K. G., W. Walke, and D. Goldman. 1992. A 102 base pair sequence of the nicotinic acetylcholine receptor delta-subunit gene confers regulation by muscle electrical activity. *Devt.* **115:** 213-219.

Garrity, P. A., and B. J. Wold. 1992. Effects of different DNA polymerases in ligation-mediated PCR: Enhanced genomic sequencing and *in vivo* footprinting. *Proc. Natl. Acad. Sci. USA* **89:** 1021-1025.

Giardina, C., M. Perez-Riba, and J. T. Lis. 1992. Promoter melting and TFIID complexes on *Drosophila* gene *in vivo*. *Genes and Dev.* **6:** 2190-2200.

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

Miner, J. H., and B. J. Wold. 1990. Herculin, a fourth member of the MyoD family of myogenic regulatory genes. *Proc. Natl. Acad. Sci.* **87:** 1089-1093

Mueller, P. R., and B. Wold. 1989. *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* **246:** 780-786.

Roy, A. L., M. Meisterernst, P. Pognonc, and R. G. Roeder. 1991. Cooperative interaction of an initiator-binding transcription initiation factors and the helix-loop-helix activator USF. *Nature* **354:** 245-248.

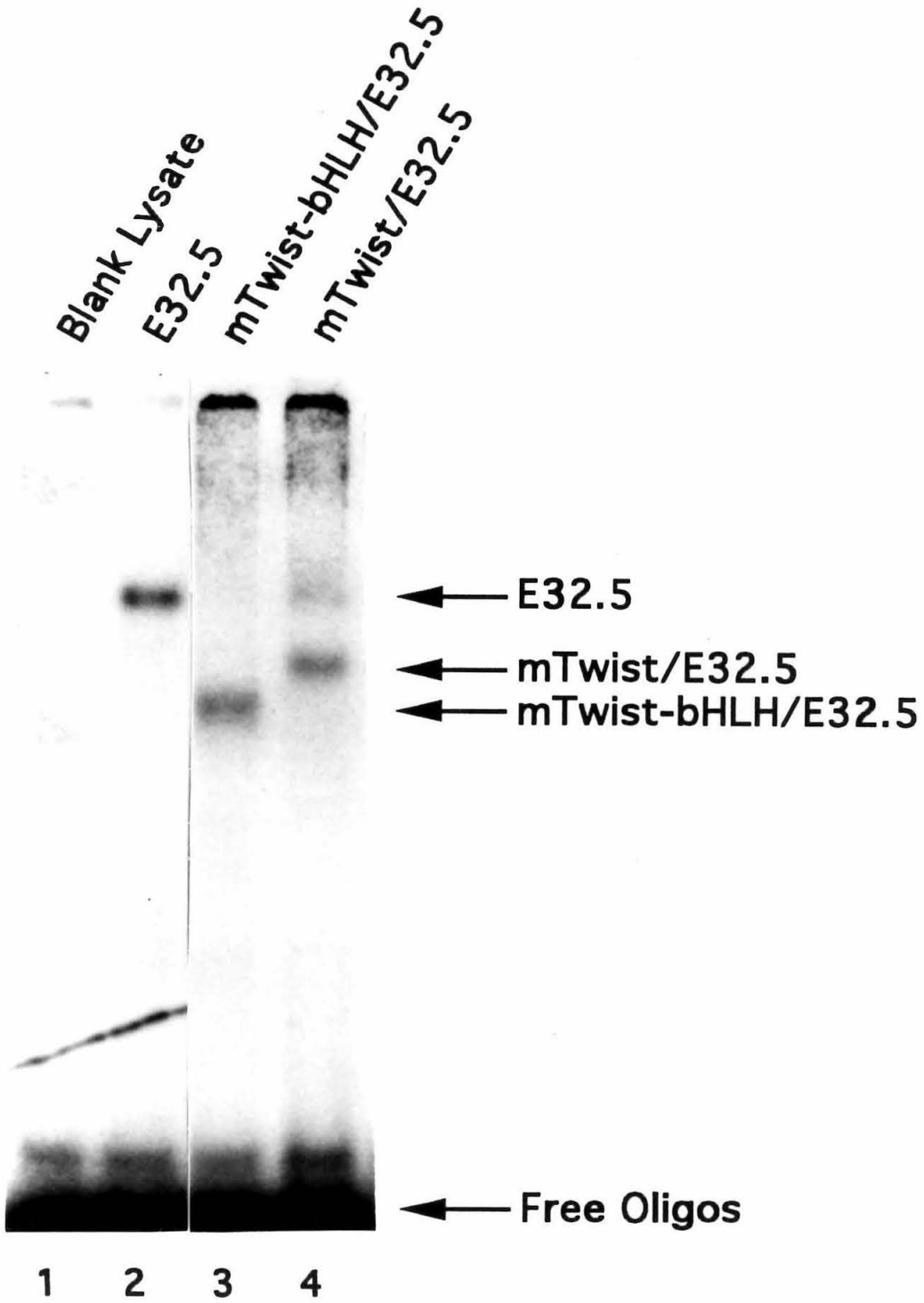
Sartorelli, V., K. A. Webster, and L. Kedes. 1990. Muscle-specific expression of the cardiac a-actin gene requires MyoD1, CArG-box binding factor, and SP1. *Genes and Dev.* **4:** 1811-1822.

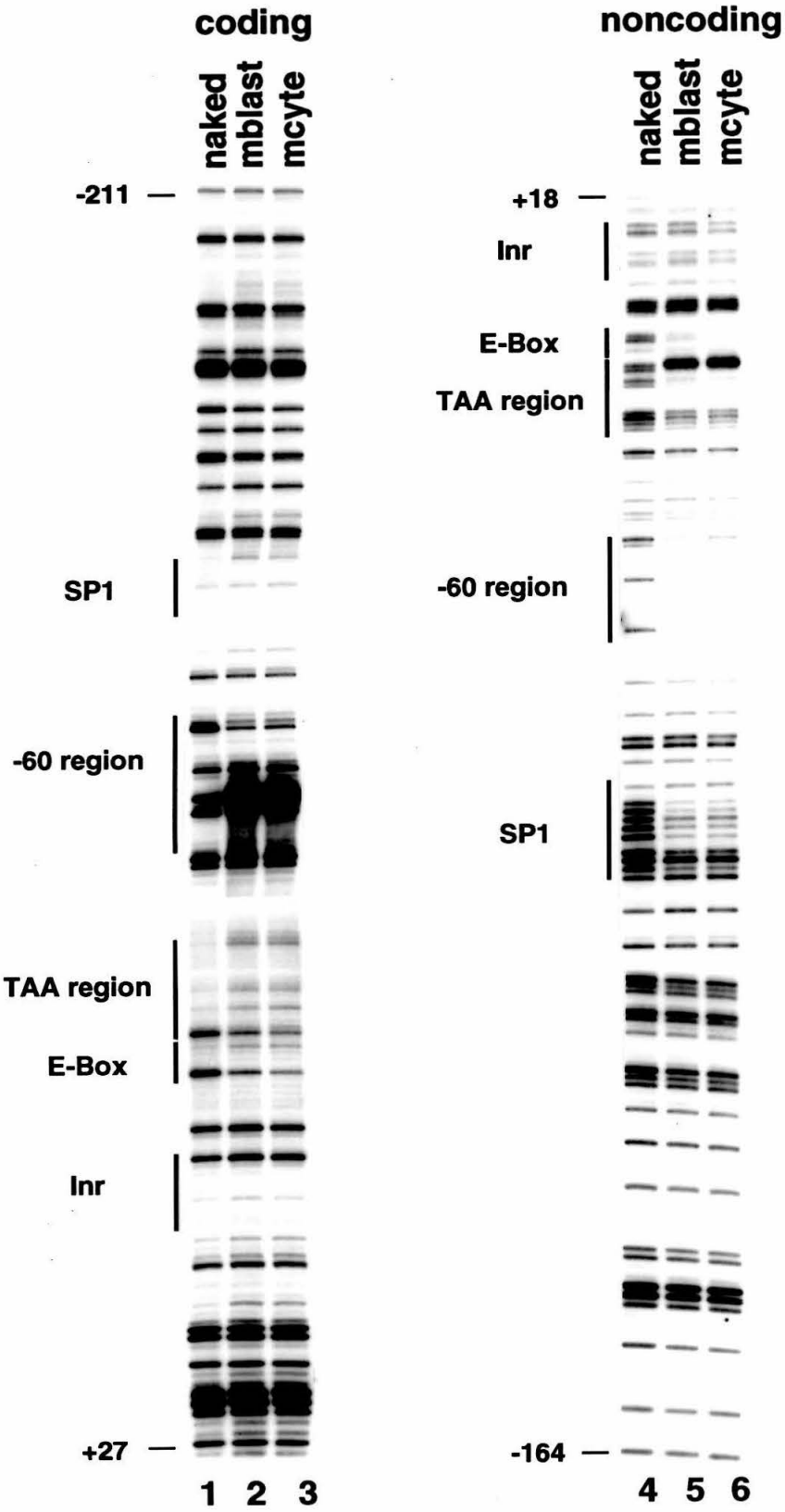
Smale, S. T., and D. Baltimore. 1989. The initiator as a transcriptional control element. *Cell* **57:** 103-113.

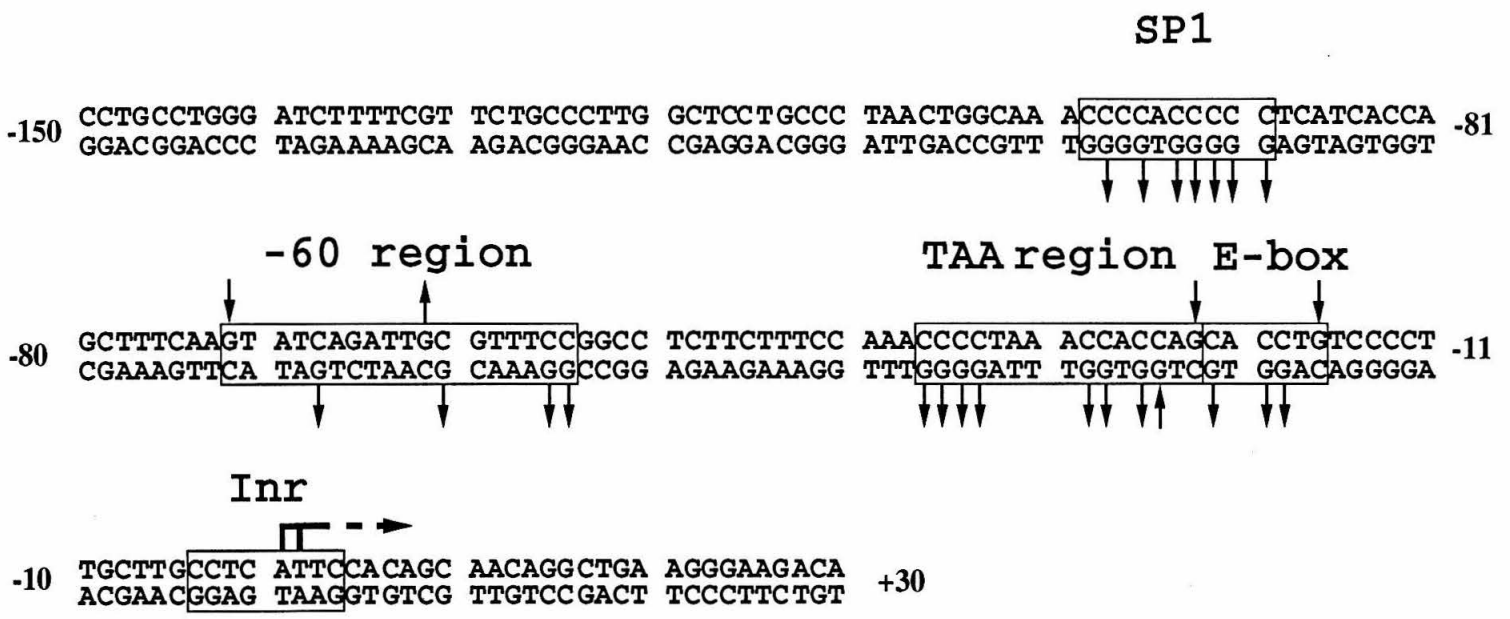
Weintraub, H., R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benezra, T. K. Blackwell, D. Turner, R. Rupp, S. Hollenberg, Y. Zhuang, and A. Lassar. 1991. The MyoD gene family: nodal point during specification of the muscle cell lineage. *Science* **251:** 761-766.

Yu, L., R. J. LaPolla, and N. Davidson. 1986. Mouse muscle nicotinic acetylcholine receptor γ–subunit: cDNA sequence and gene expression. *Nucl. Acids Res.* **14:** 3539-3554.

**FIGURE 1:** Dimethyl sulfate (DMS) *in vivo* footprint of the δ-AChR regulatory region visualized by ligation mediated PCR performed on DMS-treated DNA samples. **PANEL A**: Lanes 1 through 3 show the pattern of DMS reactivity on the coding strand and lanes 4 through 6 show the pattern of DMS reactivity on the non-coding strand. Lanes 1 and 4 contained naked DNA, which was deproteinized and purified from MM14 cells prior to *in vitro* DMS treatment (Mueller and Wold, 1989). Lanes 2 and 5 contained DNA from MM14 myoblasts treated with DMS *in vivo*. Lanes 3 and 6 contained DNA from MM14 myocytes treated with DMS *in vivo*. Sites of protein:DNA interaction are marked at the panel edge. **PANEL B:** Summary of *in vivo* protein:DNA interactions. Guanines protected from reaction with DMS *in vivo* are indicated by downward pointing arrow and guanines made hypersensitive to DMS *in vivo* are indicated by upward pointing arrows. Several adenines also appear to be hypersensitive *in vivo*. Though reproducible, the hyperreactive adenines are not included in the summary figure, as their relation to specific protein:DNA interactions has not been established (Mueller and Wold, 1989). Regions of protein:DNA interaction have boxes around them and were identified as described in the text. The potential Inr element is also indicated with a box. The dashed arrow indicates the two adjacent transcriptional start sites of this gene.


Blank Lysate
E32.5
mTwist-bHLH/E32.5
mTwist/E32.5
E32.5
mTwist/E32.5
mTwist-bHLH/E32.5
Free Oligos
1
2
3
4


coding
naked
mblast
mcyte
-211
SP1
-60 region
TAA region
E-Box
Inr
+27
1 2 3
noncoding
naked
mblast
mcyte
+18
Inr
E-Box
TAA region
-60 region
SP1
-164
4 5 6


SP1
-150
CCTGCCTGGG ATCTTTTCGT TCTGCCCTTG GCTCCTGCCC TAACTGGCAA ACCCCACCCC CTCATCACCA
GGACGGACCC TAGAAAAGCA AGACGGGAAC CGAGGACGGG ATTGACCGTT TGGGGTGGGG GAGTAGTGGT
-81
-60 region
TAA region
E-box
-80
GCTTTCAAGT ATCAGATTGC GTTTCCGGCC TCTTCTTTCC AAACCCCTAA ACCACCAGCA CCTGTCCCCT
CGAAAGTTCA TAGTCTAACG CAAAGGCCGG AGAAGAAAGG TTTGGGGATT TGGTGGTCGT GGACAGGGGA
-11
Inr
-10
TGCTTGCCTC ATTCCACAGC AACAGGCTGA AGGGAAGACA
ACGAACGGAG TAAGGTGTCG TTGTCCGACT TCCCTTCTGT
+30