In Vivo Analysis of Interactions Between Trans-Acting Factors and Their Target Genes

Thesis by

Paul R. Mueller

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1990

(Submitted March 1, 1990)

I would like to dedicate this work to my brother Jim, who has shown me from a very early age that science is a world of magic and mystery—something to be explored.

Acknowledgements

To thank all the people who have made my years at Caltech both rewarding and enjoyable is a task that is truly impossible. At the risk of forgetting to mention someone, I will try. First and foremost, I must acknowledge my advisor, Barbara Wold, for her patience and guidance, and for continually challenging me to be better than I am. My thesis committee, David Anderson, Giuseppe Attardi, Mel Simon, Norman Davidson, and most recently Kai Zinn, have been helpful not only with their suggestions but more importantly with their enthusiasm. All of the people in the Wold lab have been wonderful, but a few deserve special mention. Early on, Stuart Kim took me under his wing and showed me everything from where the cheapest movies were to which end of the Pipetman to hold. More recently, Paul Garrity has been the source of many stimulating late-night discussions about science, music, politics and life in general. Also, special thanks go to Jeff Miner for his attempts at correcting my grammar, Roger Wagner for all the beer, and Sam Murray for what must have been thousands of staples. Outside the lab, there have been many people who have helped me get through the last eight years. Jo Topol has been and still is a true friend. Warren Kibbe has always been there whenever I needed him. John Ngai always knew when to give me abuse: a true Californian-New Yorker from New Jersey. Just a few of the other friends and colleagues whom I also need to mention are: Pete Mathers, Paul Herman, Craig Bond, Chris Martin, Tom Novak, Carol Mayeda, Lance Fors, Jane Johnson, Bill Lease, Nancy Gill, Keith Harshman, Mike King, Steve Salser, Kathy Tazmui, and anyone else I forgot. Finally, I would like to give special thanks to Jacquie and Patrick for their love and support.

Abstract

The investigations presented in this thesis use various *in vivo* techniques to understand how trans-acting factors control gene expression. The first part addresses the transcriptional regulation of muscle creatine kinase (MCK). MCK expression is activated during the course of development and is found only in differentiated muscle. Several *in vivo* footprints are observed at the enhancer of this gene, but all of these interactions are limited to cell types that express MCK. This is interesting because two of the footprints appear to represent muscle specific use of general transcription factors, while the other two correspond to sites that can bind the myogenic regulator, MyoD1, *in vitro*. MyoD1 and these general factors are present in myoblasts, but can bind to the enhancer only in myocytes. This suggests that either the factors themselves are post-translationally modified (phosphorylation or protein:protein interactions), or the accessibility of the enhancer to the factors is limited (changes in chromatin structure). The *in vivo* footprinting study of MCK was performed with a new ligation mediated, single-sided PCR (polymerase chain reaction) technique that I have developed.

The second half of the thesis concerns the regulation of mouse metallothionein (MT). Metallothioneins are a family of highly conserved housekeeping genes whose expression can be induced by heavy metals, steroids, and other stresses. By adapting a primer extension method of genomic sequencing to *in vivo* footprinting, I've observed both metal inducible and noninducible interactions at the promoter of MT-I. From these results I've been able to limit the possible mechanisms by which metal responsive trans-acting factors induce transcription. These interpretations correlate with a second line of experiments involving the stable titration of positive acting factors necessary for induction of MT. I've amplified the promoter of MT to 10^2-10^3 copies per cell by fusing the 5' and 3' ends of the MT gene to the coding region of DHFR and selecting cells for methotrexate

resistance. In these cells, there is a metal-specific titration effect, and although it acts at the level of transcription, it appears to be independent of direct DNA binding factors.

Table of Contents

Acknowledgements iii

Abstract iv

Table of Contents vi

Chapter 1: Introduction 1

Development and Transcription Factors: Lessons from Muscle

Chapter 2: 17

In Vivo Footprinting of a Muscle Specific Enhancer by Ligation Mediated PCR

(Appeared in *Science*, **246**: 780-786, 1989)

Chapter 3: 24

Constitutive and Metal Inducible Protein:DNA Interactions at the Mouse Metallothionein-I Promoter Examined by *In Vivo* and *In Vitro* Footprinting

(Appeared in *Genes and Development* **2**: 412-427, 1988)

Chapter 4: 41

In Vivo Competition of Metal Specific Induction Factors Controlling Mouse Metallothionein Transcription

Appendix I: 78

Ligation Mediated PCR for Genomic Sequencing and Footprinting

Appendix II: 91

Genomic Sequencing and Methylation Analysis by Ligation Mediated PCR

(Appeared in *Science*, **246**: 810-813, 1989)

Chapter 1
Introduction

Development and Transcription Factors: Lessons from Muscle

Through the cascade of regulatory events that take place during development, the totipotent cells of the embryo acquire the phenotypes of specialized cells. How this happens for any particular cell type is somewhat of a mystery, but for the myogenic lineage we are beginning to get glimpses of the regulatory processes involved. In Xenopus it has been shown that one of the earliest events in the formation of muscle is the induction of the animal hemisphere cells by the vegetal tissue of the blastula. This leads to the formation of mesoderm, which subsequently differentiates into muscle, notochord, and other tissue types (reviewed in Gurdon et al. 1989). The earliest cells yet found to express muscle specific markers are the somites of the early gastrula (Sassoon et al. 1989). These produce the myotomes and possibly populate the other muscle forming regions of the embryo such as the limb buds. During development, there are at least four waves of skeletal muscle formation: the production of the myotomes in the somites of the gastrula, the primary muscle cells in the embryo, the secondary muscle cells in the fetus, and the satellite or regenerative cells used after birth (Stockdale et al. 1989). The lineage relationship between the early muscle cells and those found later in development is unclear. A further complication is that the individual fibers that make up skeletal muscle can be of different types (fast, slow, or a mixture of both). At least for the embryonic lineages, the fiber type is determined by the muscle precursors and not by external cues such as innervation (Sanes 1987; Miller and Stockdale 1987).

Recent progress in our understanding of how a mesodermal precursor becomes determined or committed to being a myoblast comes from the identification of a family of genes whose ectopic expression can recruit otherwise nonmyogenic cells into the myogenic

lineage. When the pluripotent fibroblast cell line 10T1/2 is treated with the cytosine analog 5-azacytidine, it can be converted to the myogenic, adipogenic, or chondrogenic lineage (Taylor and Jones 1979). 5-azacytidine treatment is thought to function via demethylation of the DNA, which could cause the expression of suppressed regulatory loci (Jones and Taylor 1980). Because a high percentage of cells are converted to the myogenic lineage by this treatment, it was suggested that one or a few loci were sufficient to activate myogenesis (Konieczny and Emerson 1984). Transfection studies with myoblast DNA seemed to confirm this hypothesis because the frequency of conversion was about 1 in 15,000 transfectants, which would be expected of a single genetic locus (Lassar et al.1986). This led to the identification of two myogenic regulatory genes, MyoD and myd, of which MyoD has been cloned (Davis et al. 1987; Pinney et al. 1988). Both of these can convert 10T1/2 cells into the myogenic lineage and seem to play a role in determination. Subsequently, the Wright and Olson laboratories cloned myogenin, a myogenic regulator that is expressed preferentially during differentiation (Wright et al. 1989; Edmondson and Olson. 1989). Surprisingly, the predicted protein product of this gene is similar to that of MyoD. Recently, two additional genes have been isolated that also share protein similarity with MyoD, Myf5 and herculin (MRF4) (Braun et al. 1989a; Rhodes and Konieczny. 1989; Miner and Wold 1990).

The four cloned myogenic regulators share one highly similar protein region that is also found in a number of other regulatory genes such as the myc family of proto-oncogenes, the immunoglobulin enhancer binding genes E12/47 and ITF-2, and the daughterless, twist, and achaete-scute regulatory genes of Drosophila (DePinho et al. 1987; Murre et al. 1989a; Henthorn et al. 1990; Caudy et al. 1988; Thisse et al. 1988; Cronmiller et al. 1988; Villares and Cabrera 1987;). This region can putatively form a helix-loop-helix (HLH) structure. It is preceded by a region of basic amino acids, and together they appear

to be required for the activity of these proteins. Whether myd will also share this HLH/basic region awaits its cloning.

Some of the HLH proteins have been shown to bind to specific sequences of DNA, and in the cases where it has been studied, the HLH/basic region is required for this activity (Murre et al. 1989a; Lassar et al. 1989; Henthorn et al. 1990). The same HLH/basic region allows formation of hetero- and homodimers between these proteins. (However, it appears that in some cases, the HLH can form dimers independent of the basic residues [Berk and Schmidt 1990]. As discussed below, this may be significant in controlling the activities of these factors). A heterodimer of MyoD and E12 binds more strongly to DNA than homodimers of either protein. This and the expression pattern of the various members of the HLH family have led Murre et al. (1989b) to suggest that there are three classes of HLH proteins, of which two can interact to form functional factors. Class A consists of ubiquitous factors such as daughterless and E12/47. These dimerize with the Class B proteins that are tissue specific, for example, MyoD and the achaete-scute family. The third Class, Class C, is represented by the myc proto-oncogene family; it has not yet been shown to dimerize or bind specific sites on DNA.

All five myogenic regulators (MyoD, myd, myogenin, Myf5 and herculin) can convert a nonmyogenic cell in culture into the myogenic lineage. Whether one or various combinations of these can act as determination genes during muscle development in the lifetime of an animal remains to be seen. More is known about the role these factors play at the end point of the myogenesis pathway, the production of terminally differentiated myocytes from replicating myoblasts. Differentiation results in withdrawal of myoblasts from the cell cycle, activation of muscle specific gene expression, assembly of muscle structures, and fusion of single cells into multinucleated muscle fibers. Studies with primary muscle cells and established muscle cell lines have shown that changes in growth

conditions and expression of myogenic regulatory genes are required for this differentiation step.

Myogenic regulators and differentiation.

The transcriptional activation of myocyte specific genes is required for differentiation. Some of the HLH family of proteins play a role in this activation and because they are DNA binding proteins, they may be transcriptional activators. To understand how these factors might be acting, it is helpful to review what is known about the mechanism of transcriptional activation. Transcription by RNA polymerase II is dependent on two sets of factors. One set is directly involved in the initiation and elongation of transcription. Its known members, TFIIA, TFIIB, TFIID, TFIIE, and TFIIF, interact more or less directly with RNA polymerase II to generate the basic transcription apparatus (Parker 1989). This apparatus is in turn controlled by members of the second set, the specific regulatory factors, which act to stimulate or repress it. The HLH proteins are thought to fit in this second set. Such factors are specific for a subset of genes, whereas the basic factors are thought to interact with most Pol II genes (Mitchell and Tjian 1989). Recent studies have shown that many specific factors consist of two functionally distinct domains. One domain binds to a specific sequence of DNA and serves mostly to bring the other domain into proximity of the basic promoter. This second domain is required for the activity of the factor and may modulate transcription by interacting with one or more of the basic transcription factors such as the TATA-binding TFIID or the RNA polymerase.

In a yeast cotransfection system, Henthorn et al. (1990) have shown that the HLH factor E12/47 (ITF-1) can activate transcription after binding to the recognition site μE5/κE2. This site is found in both the heavy chain and κ light chain enhancers, and a similar site is found in muscle specific enhancers (see below). The HLH/basic region,

located in the carboxyl half of E12/47, interacts with the μE5/κE2 sequence, whereas the amino half of E12/47 activates transcription (Henthorn et al. 1990). The activation domain of a number of transcription factors, for example GAL4 and Ap1/Jun, have been shown to function through regions of acidic residues that create areas of net negative charge. It has been suggested that this negative charge region interacts with and stimulates the basic transcription apparatus (reviewed in Ptashne 1988, Mitchell and Tjian 1989) The amino terminus of MyoD, myogenin, and herculin are somewhat acidic, and MyoD can be phosphorylated, which would also generate a net negative charge. Whether the acidic regions of some of the HLH proteins can act like the acidic activating domain of GAL4 remains to be seen.

A number of genes are transcriptionally activated during differentiation. A well-studied example of a myocyte specific gene is muscle creatine kinase (MCK). The expression of MCK increases more than 100-fold during differentiation (Chamberlain et al. 1985). Much of this increase is transcriptional (Jaynes et al. 1986), and has been shown to be dependent on a 200 bp enhancer located approximately 1 kb upstream of the transcription start site (Jaynes et al. 1988; Sternberg et al. 1988). This enhancer contains several sequence similarities found in other promoters and enhancers; however, its activity is limited to muscle cells. *In vivo* and *in vitro* footprinting studies have shown that at least four of these are sites for DNA binding factors (Horlick and Benfield 1989; Buskin and Hauschka 1989; Mueller and Wold 1989; Gossett et al. 1989). Two sites can be occupied by factors present in many cell types, whereas the other two appear to be bound by muscle specific factors. Mutational analysis is limited, but if the two sites that can be occupied by muscle specific factors are removed, the activity of the enhancer is reduced 10- to 25-fold (Buskin and Hauschka 1989). These same sites are bound by homo- and heterodimers of the HLH proteins MyoD and E12/47 *in vitro*, which may contribute to the binding observed *in vivo* (Lassar et al. 1989; Murre et al. 1989b). One of the interesting

observations about this enhancer comes from *in vivo* footprinting studies. These showed that the endogenous MCK enhancer is free of bound factors in both nonmuscle and myoblast cells even though extracts from these cells contain factors that will bind to it *in vitro* (Mueller and Wold 1989). The picture gets even more confusing because the HLH proteins MyoD and E12/47 also bind to the related μE5/κE2 sites of immunoglobulin heavy chain and kappa light chain enhancers (Murre et al. 1989b). From *in vitro* studies, the core binding consensus site ascribed to MyoD and E12/47 is an extremely degenerate CANNTG (Lassar et al. 1989). However, it should be pointed out that studies of E12/47 indicate that certain sequences that match this consensus do not bind this factor *in vitro* (Murre et al. 1989a). Nevertheless, some additional factors must help specify correct *in vivo* binding.

One of the future challenges will be to determine what these other factors are, and how they control transcriptional activity of the HLH proteins. A number of possibilities exist. One possibility makes the assumption that the HLH proteins do not intrinsically have a high affinity for their binding sites, but that the affinity can be increased by additional DNA binding factors. These non-HLH DNA binding factors might stabilize the binding of the HLH factors to the DNA in a cooperative fashion. This idea is consistent with the "all-or-none" binding observed *in vivo* at the MCK enhancer (Mueller and Wold 1989). If either the general factors or the muscle specific factors are missing, none of the sites can be bound. The appeal of this model is that the triggering event that would allow binding, and therefore transcription, might be a modest change in the concentration or activity of any one factor. The affinity for the binding site could also be increased by certain combinations of HLH proteins making heterodimers. A likely candidate for this type of function might be myogenin. On differentiation, myogenin is activated in all skeletal muscle cell lines that have been examined, and it might interact with MyoD or E12/47 to stimulate or specify their activities.

An alternative possibility might be that the binding sites for these factors are physically blocked. This could be accomplished either through a restrictive chromatin structure or by a specific negative DNA binding factor. Developmental changes in the chromatin structure of muscle genes has not been investigated yet, but the idea of a specific negative factor can probably be ruled out, at least for the MCK enhancer. Blocking by specific negative factors has been suggested to take place at the DNA level in the homeodomain gene, fushi tarazu (ftz). Engrailed (en) and even-skipped (eve), two proteins with sequence similarities to ftz, bind and potentially block ftz binding (reviewed in Serfling 1989; Biggin and Tjian 1989). This probably does not happen in muscle differentiation, because *in vivo* footprinting of MCK has shown that there are no specific negative DNA binding factors bound to the enhancer in the absence of transcription.

One last possibility is the existence of a negative factor that acts directly on the HLH proteins, before they bind DNA, via protein:protein interactions. This model has recently received some supportive evidence. Forced overexpression of one of the Class C HLH proteins (myc) inhibits the differentiation of myoblasts into myocytes (Falcone et al. 1985; Miner and Wold, Manuscript in preparation, 1990). An intriguing but unproven idea might be that the Class C proteins block activity of the Class A and B proteins by physically interacting with them and preventing DNA binding. At a recent meeting, Weintraub and colleagues reported that they have identified a previously uncharacterized HLH-like protein, Id, that can dimerize to other HLH proteins but cannot bind DNA (Berk and Schmidt 1990). Id lacks the basic region that accompanies the HLH region in the other HLH proteins, which could explain why it does not bind DNA. Although Id is not a myc protein, it fills the role postulated for the Class C proteins, and, as expected, the level of Id is high in myoblasts and decreases as they differentiate into myocytes. Perhaps the Class C HLH proteins (e.g., myc), that also decrease on differentiation, act as positive activators of this newly discovered negative acting HLH protein.

Myogenic regulators and determination.

Expression of the HLH proteins is required for differentiation, but what role might they play in determination? Although the HLH proteins were originally isolated as determination factors, this function is not as clear-cut as originally thought. The continued expression of MyoD is not required for a cell line to remain determined to the myogenic lineage (Montarras et al. 1989; Vaidya et al. 1989). Furthermore, the cells that populate the limb buds of mouse do not express either MyoD or myogenin even though they are determined to become muscle cells (Sassoon et al. 1989). Perhaps the other myogenic regulators (myd, Myf5, or herculin) are expressed in these situations. Alternatively, there could be a "master regulator" that has yet to be identified.

An interesting property that relates to the stability of the determined state is that MyoD, myogenin, Myf5 and herculin autoregulate their own expression (Braun et al. 1989b; Rhodes and Konieczny 1989; Thayer et al. 1989; Miner and Wold 1990). In many cell lines, the ectopic expression of one of these proteins activates the endogenous counterpart and one or more of the other myogenic HLH genes. As has been postulated for Jun/Ap1 complexes, this type of autoregulation could maintain an otherwise transient signal (Angel et al. 1988). Transient external cues, for example induction, might activate genes like MyoD whose products would then autoregulate the myogenic regulatory genes.

Why are there five or more myogenic regulators? Any one cell type does not need all of these factors, and with the exception of myogenin, they appear to be somewhat interchangeable. For example, the cell line BC_3H1 does not express MyoD but does express Myf5, whereas for the cell line MM14 the situation is reversed (Mueller and Wold 1989). In both lines, the myocyte specific gene MCK is expressed and furthermore, preliminary results show that the *in vivo* footprint pattern at the enhancer is identical (P. Garrity, unpublished observations). However, MCK is a relatively general muscle gene; it is expressed in both skeletal and cardiac muscle tissue. For genes with more limited

expression patterns, such as those that specify muscle fiber type (fast, slow, or mixed), it remains possible that unique combinations of HLH factors could specify different patterns of expression. It is also possible that the different waves of muscle development that take place during the lifetime of an animal might require different HLH proteins. There is some evidence to support this idea: herculin and MyoD are preferentially found in adult muscle tissue, whereas Myf5 expression is highest in fetal tissue (Braun et al. 1989b; Rhodes and Konieczny 1989; Miner and Wold 1990).

Future prospects.

Understanding muscle development has come a long way in recent years. Cell culture model systems have been instrumental in understanding myogenesis. As usual, the more that is learned about a problem, the more there are questions to be answered. How do the HLH proteins, with such an apparently degenerate recognition site, know when and where to activate transcription? Is there a "master regulatory" gene, or is myogenesis simply dependent on the autoregulation of genes like MyoD? What specifies the formation of different muscle fiber types *in vivo*? What factors activate cardiac and smooth muscle genes? These muscle types express many of the same differentiation genes as skeletal muscle, e.g. MCK and actin, but do not express any of the known myogenic regulators. A combination of *in vitro* and *in vivo* analyses will be helpful in answering these questions, but future work will also need to concentrate on the role these myogenic regulators play in the development of different muscle lineages.

References

Angel, P. K. Hattori, T. Smeal, and M. Karin. 1988. The *jun* proto-oncogene is positively autoregulated by its product jun/Ap-1. *Cell* **55**: 875-885.

Berk, A. J. and M. C. Schmidt. 1990. Meeting review: How do transcription factors work? *Genes Dev.* **4**: 151-155.

Biggin, M. D. and R. Tjian. 1989. Transcription factors and the control of *Drosophila* development. *Trends Genetics.* **5**: 377-383.

Braun, T. G. Buschhausen-Denker, E. Bober, E. Tannich and H.H. Arnold. 1989a. A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. *EMBO J.* **8**: 701-709.

Braun, T., E. Bober, G. Buschhausen-Denker, S. Kotz, K. Grzeschik, and H. H. Arnold. 1989b. Differential expression of myogenic determination genes in muscle cells: possible autoactivation by the *Myf* gene products. *EMBO J.* **8**: 3617-3625.

Buskin, J. N. and S. D. Hauschka. 1989. Identification of a myocyte nuclear factor that binds to the muscle specific enhancer of the mouse muscle creatine kinase gene. *Mol. Cell. Biol.* **9**: 2627-2640.

Caudy, M., H. Vassin, M. Brand, R. Tuma, L. Y. Jan, and Y. N. Jan. 1988. *daughterless*, a *Drosophila* gene essential for both neurogenesis and sex determination has sequence similarities to *myc* and the *achaete-scute* complex. *Cell* **55**: 1061-1067.

Chamberlain, J. S., J. B. Jaynes, and S. D. Hauschka. 1985. Regulation of creatine kinase induction in differentiating mouse myoblasts. *Mol. Cell Biol.* **5**: 484-492.

Cronmiller, C., P. Schedl, and T. W. Cline. 1988. Molecular characterization of *daughterless*, a *Drosophila* sex determination gene with multiple roles in development. *Genes Dev.* **2**: 1666-1676.

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

DePinho, R. A., K. S. Hatton, A. Tesfaye, G. D. Yancopoulos, and F. W. Alt. 1987. The human *myc* gene family: structure and activity of L-*myc* and an L-*myc* pseudogene. *Genes Dev.* **1**: 1311-1326.

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

Falcone, G., F. Tato, and S. Alema. 1985. Distinctive effects on the viral oncogenes *myc, erb, fps*, and *src* on the differentiation program of quail myogneic cells. *Proc. Natl. Acad. Sci.* **82**: 426-430.

Gossett, L. A., D. J. Kelvin, E. A. Sternberg, and E. N. Olson. 1989. A new myocyte specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle specific genes. *Mol. Cell. Biol.* **9**: 5022-5033.

Gurdon, J.B., T. J. Mohun, C. R. Sharpe, and M. V. Taylor. 1989. Embryonic induction and muscle gene activation. *Trends Gent.* **5**: 51-56.

Henthorn, P., M. Kiledjian, and T. Kadesch. 1990. Two distinct transcription factors that bind the immunoglobulin enhancer μE5/κE2 motif. *Science.* **247**: 467-470.

Horlick, R. A. and P. A. Benfield. 1989. The upstream muscle specific enhancer of the rat muscle creatine kinase gene is composed of multiple elements. *Mol. Cell. Biol.* **9**: 2396-2413.

Jaynes, J. B., J. S. Chamberlain, J. N. Buskin, J. E. Johnson, and S. D. Hauschka. 1986. Transcriptional regulation of the muscle creatine kinase gene and regulated expression in transfected mouse myoblasts. *Mol. Cell. Biol.* **6**: 2855-2864.

Jaynes, J. B., J. E. Johnson, J. N. Buskin, C. L. Gartside, and S. D. Hauschka. 1988. The muscle creatine kinase gene is regulated by multiple upstream elements, including a muscle specific enhancer. *Mol. Cell. Biol.* **8**: 62-70.

Jones, P. A. and S. M. Taylor. 1980. Cellular differentiation, cytidine analogs and DNA methylation. *Cell* **20**: 85-93.

Konieczny, S. F. and C. P. Emerson, Jr. 1984. 5-Azacytidine induction of stable mesodermal cell lineages from 10T1/2 cells: Evidence for regulatory genes controlling determination. *Cell* **38**: 791-800.

Lassar, A. B., B. M. Paterson, and H. Weintraub. 1986. Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. *Cell* **47**: 649-656.

Lassar, A. B., J. Buskin, D. Lockshon, R. L. Davis, S. Apone, S. D. Hauschka, and H. Weintraub. 1989. MyoD is a sequence-specific DNA binding protein requiring a region of *myc* homology to bind to the muscle creatine kinase enhancer. *Cell* **58**: 823-831.

Miller, J. B and F. E. Stockdale. 1987. What muscle cells know that nerves don't tell them. *Trends in Neurosci.* **10**: 325-329.

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

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

Montarras, D., C. Pinset, J. Chelly, A. Kahn, and F. Gros. 1989. Expression of MyoD1 coincides with terminal differentiation in determined but inducible muscle cells. *EMBO J.* **8**: 2203-2207.

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. Schonleber-McCaw, and D. Baltimore. 1989a. A new DNA binding and dimerization motif in immunoglobulin enhancer binding , *daughterless*, *MyoD*, and *myc* proteins. *Cell* **56**: 777-783.

Murre, C. P. Schonleber-McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub, and D. Baltimore. 1989b. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58**:537-544.

Parker, C. S. 1989. Transcription factors. *Current Opinion in Cell Biology* **1**: 512-518.

Pinney, D. F., S. H. Pearson-White, S. F. Konieczny, K. E. Latham, and C. P. Emerson, Jr. 1988. Myogenic lineage determination and differentiation: Evidence for a regulatory gene pathway. *Cell* **53**: 781-793.

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

Rhodes, S. J. and S. F. Konieczny. 1989. Identification of MRF4: a new member of the muscle regulatory factor gene family. *Genes Dev.* **3**: 2050-2061.

Sanes, J. R. 1987. Cell lineage and the origin of muscle fiber types. *Trends in Neurosci.* **10**: 219-221.

Sassoon, D., G. Lyons, W. E. Wright, V. Lin, A. Lassar, H. Weintraub, and M. Buckingham. 1989. Expression of two myogenic regulatory factors myogenin and MyoD1 during mouse embryogenesis. *Nature.* **341**: 303-307.

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

Sternberg, E. A., G. Spizz, W. P. Perry, D. Vizard, T. Weil, and E. N. Olson. 1988. Identification of upstream and intragenic regulatory elements that confer cell-type-restricted and differentiation-specific expression on the muscle creatine kinase gene. *Mol. Cell. Biol.* **8**: 2896-2909.

Stockdale, F. E., J. B. Miller, J. L. Feldman, G. Lamson, and J. Hager. 1989. Myogenic cell lineages: Commitment and modulation during differentiation of avian muscle. In *Cellular and molecular biology of muscle development*, UCLA Symposium on Molecular and Cellular Biology, (eds. Kedes and F. E. Stockdale), Vol. 93, pp. 3-13. Alan R. Liss, New York.

Taylor, S. M. and P. A. Jones. 1979. Multiple new phenotypes induced in 10T1/2 and 3T3 cells treated with 5-azacytidine. *Cell* **17**: 771-779.

Thayer, N. J., S. J. Tapscott, R. L. Davis, W. E. Wright, A. B. Lassar, and H. Weintraub. 1989. Positive autoregulation of the myogenic determination gene MyoD1. *Cell* **58**: 241-248.

Thisse, B., C. Stoetzel, C. Gorostiza-Thisse, and F. Perrin-Schmitt. 1988. Sequence of the *twist* gene and nuclear localization of its protein in endomesodermal cells of early *Drosophila* embryos. *EMBO J.* **7**: 2175-2183.

Vaidya, T. B., S. J. Rhodes, E. J. Taparowsky, and S. F. Konieczny. 1989. Fibroblast growth factor and transforming growth factor β repress transcription of the myogenic regulatory gene MyoD1. *Mol. Cell. Biol.* **9**: 3576-3579.

Villares, R. and C. V. Cabrera. 1987. The *achaete-scute* gene complex of D. melanogaster: conserved domains in a subset of genes required for neurogenesis and their homology to *myc*. *Cell* **50**: 415-424.

Wright, W. E., D. A. Sassoon, and V. K. Lin. 1989. Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. *Cell* **56**: 607-617.

Chapter 2

In Vivo Footprinting of a Muscle Specific Enhancer by Ligation Mediated PCR

(This chapter was published in *Science*, **246**: 780-786, 1989)

Research Articles

In Vivo Footprinting of a Muscle Specific Enhancer by Ligation Mediated PCR

PAUL R. MUELLER AND BARBARA WOLD

In vivo protein-DNA interactions at the developmentally regulated enhancer of the mouse muscle creatine kinase (MCK) gene were examined by a newly developed polymerase chain reaction (PCR) footprinting procedure. This ligation mediated, single-sided PCR technique permits the exponential amplification of an entire sequence ladder. Several footprints were detected in terminally differentiated muscle cells where the MCK gene is actively transcribed. None were observed in myogenic cells prior to differentiation or in nonmuscle cells. Two footprints appear to correspond to sites that can bind the myogenic regulator MyoD1 in vitro, whereas two others represent muscle specific use of apparently general factors. Because MyoD1 is synthesized by undifferentiated myoblasts, these data imply that additional regulatory mechanisms must restrict the interaction between this protein and its target site prior to differentiation.

THE DEVELOPMENTAL PROGRESSION FROM MESODERMAL precursor cell to determined, proliferating myoblast and from myoblast to postmitotic, differentiated muscle cell (myocyte) involves a cascade of regulatory changes. The determination step that produces myoblasts occurs when the developmental potential of a precursor is restricted to the myogenic lineage. A significant inroad to understanding this developmental decision comes from the recent identification of a family of genes whose ectopic expression in cultured cells can recruit otherwise nonmyogenic cells to function as myoblasts. Several of these genes have been cloned, including MyoD1 (*1*), myogenin (*2*), and Myf5 (*3*). Their products are nuclear proteins, and all share some sequence similarity with other important regulatory molecules. These include the *myc* family of onco-proteins (*4*) and the immunoglobulin enhancer binding proteins (E12 and E47) (*5*) in mammals, and the protein products of the *daughterless*, *twist*, and *achaete-scute* genes (*6*) in *Drosophila*. Myoblasts can, in turn, be triggered to undergo differentiation in cell culture by altering the signals provided by growth factors, hormones, and extracellular matrix components. Differentiation of skeletal muscle ultimately includes cell cycle withdrawal, transcriptional activation of muscle specific genes, assembly of muscle structures, and cell fusion to produce multinucleated myotubes.

P. R. Mueller and B. Wold are in the Division of Biology at the California Institute of Technology, 156-29, Pasadena, CA 91125.

Transcriptional regulation of myoblast and myocyte specific genes is central to the execution of this developmental pathway, but knowledge of how this regulation is achieved is limited. For example, recent studies suggest that MyoD1 can act as a positive transcriptional regulator by binding to sites in several myocyte specific genes (*7*). However, MyoD1 is also expressed in proliferating myoblasts where it can positively regulate expression of its own promoter, but where myocyte specific genes are transcriptionally silent (*8*). This raises the question of how MyoD1 and its relatives act differently in myoblasts versus myocytes. Moreover, genes expressed specifically in myocytes bind some factors in vitro that are widely distributed in nonmuscle cells and myoblasts as well as in myocytes (*9–11*). Whether these factors interact in the cell with myocyte specific genes when they are transcriptionally silent is not yet known.

In vivo footprinting can answer some of these questions by providing information on when and how proteins occupy a given regulatory region of DNA in the living cell. These experiments can be especially useful when taken together with genetic characterization of cis-acting elements and in vitro DNA binding studies of the relevant factors. Sequence inspection, reverse genetic analysis, and gel retardation experiments have identified a number of potentially important cis-acting sequence elements in the upstream enhancer of muscle creatine kinase (MCK) (*9, 12, 13*). Here, we have used genomic footprinting to examine protein-DNA interactions at this region in cells that express MCK (myocytes) and in cells that do not (nonmyogenic cells and myoblasts).

Despite the information provided by in vivo footprinting and the development of several genomic sequencing strategies for this purpose (*14, 15*), application of these techniques has been limited. In vivo footprinting of a single-copy regulatory region in large genomes (mammals) by established strategies is technically challenging. Large cell numbers are required, and experiments often have an unacceptable signal-to-noise ratio. Our genomic footprinting method largely eliminates these problems and enables the in vivo footprinting of relatively small numbers of cells (about 10^5 nuclei) or dissected tissues. It is based on exponential amplification of a genomic sequence ladder by the polymerase chain reaction (PCR) (*16*). Although presented as a footprinting technique, this method should be generally applicable to any PCR problem in which only one end of the region to be amplified is known. For example, Pfeifer *et al.* (*17*) have adapted it to the study of in vivo methylation patterns and genomic sequencing.

Footprinting with ligation mediated, single-sided PCR. In vivo footprints are visualized by comparing samples of DNA that have been exposed to nucleases or alkylating agents in the cell (in vivo) with samples exposed to these agents after the DNA has been

extracted from cells and deproteinized (in vitro or naked) (*18*). Dimethyl sulfate (DMS) is commonly used as the alkylating agent because cell membranes are freely and rapidly permeable to it (*18*). Proteins bound to DNA often alter the accessibility of DMS to guanines at or near the binding site (*19*). After purification of the DNA, both in vitro and in vivo DMS-treated samples are quantitatively cleaved at the methylated guanine residues with piperidine (*20*) and then compared to reveal the footprint.

The PCR consists of repeated cycles of template denaturation, primer annealing, and DNA polymerase extension to exponentially amplify a segment of DNA located between two primers. Each cycle doubles the number of templates, and after 25 to 30 cycles a single-copy gene can be amplified more than 10^6-fold (*16*). Conventional PCR is not immediately applicable to sequencing or footprinting because it requires two defined ends. A sequence or footprint ladder is composed of a population of related nucleic acid fragments. One end of each fragment is fixed by a primer or restriction cut and is therefore the same for all, whereas the other end is determined by variable chemical cleavage or chain termination and is therefore unique for each fragment. To apply PCR to a sequence ladder, we have introduced a simple ligation step that adds a common oligonucleotide sequence to the unique end of each member. A primer complementary to this new common sequence is then used, together with a primer complementary to the original fixed end, for simultaneous exponential amplification of all members of the sequence ladder. The procedure has high selectivity and specificity that are derived from the design of the ligation step and the choice of primers (Fig. 1). It also has high fidelity; a footprint consists of subtle differences in the starting concentrations of particular members of a sequence ladder, and these differences are reproducibly retained through the amplification.

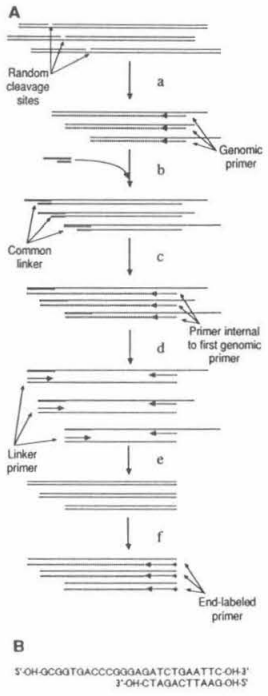


Fig. 1. Schematic of ligation mediated PCR footprinting or sequencing. (**A**) Starting material is genomic DNA that has been treated with Maxam and Gilbert sequencing chemistry (*20*) . This leaves 5′ and 3′ phosphates (*20*). The first step (a) defines the fixed end of the sequence ladder by denaturing the cleaved genomic DNA and annealing a gene specific primer. Extension (dotted line) of this primer to the variable cleavage site generates a family of blunt-ended duplex molecules, which are substrates for T4 DNA ligase–catalyzed addition (b) of a specially designed linker (heavy lines; see below) to each member of the sequence ladder, thereby providing each with a common, defined end. The genomic DNA provides the 5′ phosphate used in the ligation. Specificity is provided by the fact that the bulk of the genomic DNA does not serve as a substrate in the ligation reaction because it lacks a blunt double-stranded end. The DNA is then denatured (c), and a second gene specific primer is annealed and extended. To increase specificity, the second gene specific primer is positioned so that its extending end is 3′ to that of the first primer. The original genomic DNA is used again as a template in this reaction, only now it has the longer strand of the linker covalently attached to it, and the extension product reads through this added sequence. Each member of the sequence ladder now has two defined ends (the common linker primer and the second gene specific primer), and is suitable for PCR (*16*). After 16 rounds of PCR (d and e), the sequence ladder is amplified ~10^4-fold. It is visualized by primer extension of an end-labeled, third primer (f). Its appearance is that of the corresponding sequence ladder, except that it is uniformly longer by the additional length of the linker. The third primer should overlap the second primer, and also be positioned so that its extending end is 3′ to that of the second primer. The sequence ladder can also be visualized by filter blotting (*17*). (**B**) Structure of the common linker. The linker shown is but one example of the possible sequences that could be used. It is important that: (*i*) the linker contains no 5′ phosphates and is staggered to eliminate self-ligation and assure directionality in ligation, respectively; (ii) the duplex between the long and short oligomers is stable at ligation conditions, but not at PCR temperatures; and (iii) the longer oligomer should have a comparable T_m (melting temperature) to that of the second gene-specific primer (see above).

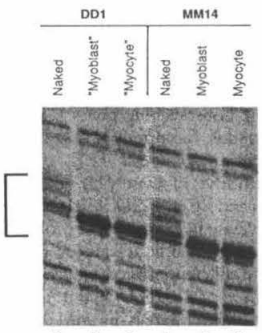


Fig. 2. In vivo footprint of metallothionein-I promoter visualized by ligation mediated, single-sided PCR. A footprint is ~185 bp upstream of transcription start on the coding strand of the MT-I promoter (*15*) in both DD1 (lanes 1 to 3) and MM14 (lanes 4 to 6) cells. The Sp1 consensus site is bracketed on left. Naked DNA's are genomic control samples from DD1 or MM14 cells that were treated with DMS in vitro. Myoblast and myocyte DNA's are from the same cells grown under proliferation or differentiation conditions, respectively, and treated with DMS in vivo. In vivo labels for DD1 cells are in quotations because these cells are differentiation defective and therefore do not form true myoblasts or myocytes. Cell culture and DNA preparations as in Figs. 3 and 4. All primers and oligomers were gel purified. For first strand synthesis, 3 μg of DMS-piperidine treated DNA (*15*) and 0.3 pmol of primer 1 (*44*) were suspended in 15 μl of 40 m*M* tris, *p*H 7.7, 50 m*M* NaCl. The sample was heated at 95°C for 2 minutes and then incubated at 60°C for 30 minutes. Hybridization was stopped by transferring to ice; a solution of 7.5 μl of 20 m*M* $MgCl_2$, 20 m*M* dithiothreitol (DTT), and 0.02 m*M* of each deoxynucleoside triphosphate (dNTP) was added, then 1.5 μl of a 1:4 dilution of Sequenase version 1.0 (USB) [diluted in 10 m*M* tris (*p*H 7.5), 1 m*M* EDTA] was added, and the sample was incubated at 47°C for 5 minutes. The reaction was stopped by heating at 60°C for 5 minutes, then adding 6 μl of 310 m*M* tris (*p*H 7.7), and then heating for 10 minutes at 67°C. For ligation of linker, the sample was transferred to ice, and a solution of 20 μl of 17.5 m*M* $MgCl_2$, 42.3 m*M* DTT, and BSA at 125 μg/ml was added, then 25 μl of ligation mixture [10 m*M* $MgCl_2$, 20 m*M* DDT, 3 m*M* ATP, BSA at 50 μg/ml, with 5 μl of PCR linker mix (20 pmol of linker per microliter in 250 m*M* tris, *p*H 7.7) and 3 Weiss units of T4 ligase per 25 μl] was added. The linker was prepared as described (*45*). After incubation overnight at 15°C, the reaction was stopped by heating to 70°C for 10 minutes. The sample was precipitated in the presence of 10 μg of yeast carrier tRNA. For the PCR reaction; the precipitated samples were washed once with 75 percent ethanol and then suspended in water. 20 μl of 5× Taq buffer (200 m*M* NaCl, 25 m*M* tris *p*H 8.9, 25 m*M* $MgCl_2$, 0.05 percent w/v gelatin) was added along with 20 nmol of each dNTP, 10 pmol of a primer 2 (*44*), 10 pmol of the longer oligomer of the linker (Fig. 1B), and 5 units of Taq polymerase (Cetus). The volume was adjusted to 100 μl with H_2O. Samples were covered with 90 μl of mineral oil, heated to 94°C for 1 minute, and then manually cycled (denatured for 1 minute at 94°C, hybridized for 2 minutes at 63°C, and extended for 3 minutes at 76°C) 16 times. Samples were placed on ice, 1 to 5 pmol of an end-labeled (*46*) primer 3 (*44*) was added, along with 2.5 units of Taq polymerase and 20 nmol of each dNTP. Samples were heated to 94°C for 2 minutes, hybridized at 66°C for 2 minutes, and extended at 76°C for 10 minutes. Polymerase activity was stopped by chilling on ice, adding 295 μl of 260 m*M* sodium acetate, 10 m*M* tris *p*H 7.5, and 4 m*M* EDTA, and, extracting with a mixture of 50 parts phenol, 50 parts chloroform, and 1 part isoamyl alcohol. The samples were precipitated, resuspended in loading dye, and half of each sample was placed on each lane of a standard sequencing gel (*20*).

We tested this PCR footprinting technique on the mouse metallothionein I (MT-I) promoter, which has been well characterized both in vivo and in vitro. Previously, we had used cells containing more than 100 copies of the MT-I promoter to observe in vivo interactions and found a prominent footprint at the upstream Sp1 binding site (Sp1-A) (*15*). Using PCR footprinting, we reexamined this region in MM14 and DD1 cells that contain only a single copy of the MT-I gene per haploid genome. The expected footprint is apparent in both MM14 and DD1 lines, as indicated by comparison of the naked DNA control sample (Fig. 2, lanes 1 and 4) with the in vivo DNA sample from cells grown under either proliferation (lanes 2 and 5) or differentiation (lanes 3 and 6) conditions. This result illustrates the sensitivity of the technique; the data shown are from a 9-hour, screened exposure on Kodak XAR-P film. These footprints correlate with the observed basal expression of MT-I in these cells (Fig. 3), and establish that the data obtained from ligation mediated PCR footprinting are consistent with data from more conventional methods.

Expression of muscle regulatory genes during differentiation. Differentiation of myoblasts into myocytes is accompanied by complex changes in the expression of muscle regulatory and structural gene products as well as some housekeeping genes (Fig. 3) (*21*). Levels of RNA's relevant to this footprinting study were examined in MM14 cells, a permanent myogenic cell line (*22*), DD1 cells, a differentiation defective derivative of these (*23*), and two other myogenic lines, aza-myoblasts and BC_3H1 cells. Under differentiation conditions, there was a decrease in expression of proliferation related genes, such as *c-myc*, and general housekeeping genes, such as MT-I, in both myogenic and differentiation defective lines. Myocyte specific gene products such as MCK (Fig. 3) and myosin heavy chain (*24*) increased during differentiation in MM14 but remained absent from DD1 cells. Two of the three known myogenic regulators are expressed in the MM14 cells; MyoD1 remains constant before and after differentiation, whereas myogenin is activated upon differentiation. The third regulator, Myf5, could not be detected in MM14 cells, and all three are absent in DD1 cells.

In vivo footprinting of the muscle creatine kinase enhancer. MCK expression is activated during muscle differentiation (Fig. 3) (*25*). The MCK transcriptional control elements that have been identified include an intronic enhancer region, a proximal promoter region, and an upstream enhancer region (*12, 13, 26*). We have focused on the upstream enhancer region because it confers high-level stage and tissue specific expression on a reporter gene in both cell culture (*12, 13, 27*) and transgenic mice (*28*). This enhancer is located about 1 kb upstream of the transcription start and contains sequence motifs similar to recognition sites of several putative general and muscle specific transcription factors (see below). Four of these potential binding sites are occupied in MM14 myocytes (Fig. 4), though none are detectably occupied in MM14 myoblasts or DD1 cells. Multiple, independent DNA preparations from each cell type were tested, and all interactions were highly reproducible (summarizd in Fig. 5).

A myocyte specific in vivo footprint is near the upstream end of an adenine-rich sequence (CTAAAAATAACCC) located at −1077 (Fig. 4, lane 3). The A-T–rich character of this sequence prevents us from observing additional interactions in this region employing the DMS-guanine reaction. Factors from extracts of both muscle and nonmuscle cells bind to this sequence in vitro (*9*), suggesting that the in vivo interaction observed here results from myocyte specific use of factors present in many other cell types, including myoblasts. Evidence that this A-rich sequence may be acting as a positive, myocyte specific regulatory element comes from studies of the chicken myosin light chain 2-A (MLC2-A) promoter (*10*) and the rat MCK enhancer (*11*). In both cases, genetic and biochemical studies show that it contributes to muscle specific expression and binds one or more proteins present in extracts both from muscle and nonmuscle cells.

The enhancer contains an exact match to the in vitro binding site of transcription factor AP-2 (*29, 30*). Although AP-2 is not a myocyte specific factor (*29*), the in vivo footprint is restricted to differentiated MM14 cells (Fig. 4, lane 9, and Fig. 5). The protections observed are partial, an indication that the site may be occupied only part of the time, on average, or that the nature of the physical interaction between protein and DNA only partially occludes DMS accessibility. In either case, the reproducibility of this observation in multiple experiments indicates that it is a genuine footprint. There were no footprints at this site in proliferating MM14 myoblasts, differentiation defective DD1 cells (Fig. 4, lanes 8 and 11 and 12), Balb/c 3T3 fibroblasts, or L cells (*24*). This myocyte specific pattern contrasts with the in vitro interactions detected with gel-mobility shift assays, in which factors from both muscle and nonmuscle cell extracts bind in this region (*9*).

The AP-2 site and the adenine-rich element flank a 110–base pair (bp) central core that, by itself, retains most of the activity of the upstream enhancer (*9*). Two sequences similar to elements found in the immunoglobulin kappa (κ) (*31*) and heavy (H) chain (*32*) cellular enhancers are present in the core, and both are critical for the activity of the murine MCK enhancer. Deletion of the κ chain enhancer–like sequence results in an ~10-fold decrease in enhancer activity (*33*), and mutation of the H chain enhancer–like sequence results in decrease of about 25-fold (*9*). Buskin and Hauschka have reported a myocyte specific binding activity, MEF-I (myocyte enhancing factor one) (*9*), that interacts with the H chain enhancer–like sequence in vitro. We have therefore identified the H chain enhancer–like site as MEF-I in Figs. 4 and 5. In vivo footprints where found at both the κ chain enhancer–like and MEF-I (H chain enhancer–like) sites in MM14 myocytes, but not in the other cell types tested (Fig. 4, compare lanes 3 and 9 to lanes 2, 5, 6, 8, 11, and 12).

Based on similarity to cis-acting sequence motifs, it has been suggested that the CarG and *sph1* elements present in the MCK

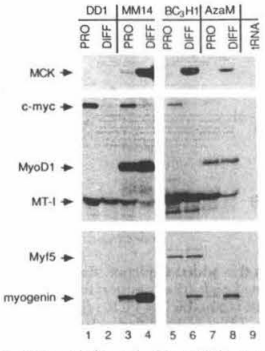


Fig. 3. (lower right) RNA analysis of muscle and nonmuscle specific genes. Total RNA's from MM14 (*22*) (lanes 3 and 4), BC_3H1 (*47*) (lanes 5 and 6), aza-myoblasts (*48*) (AzaM, lanes 7 and 8), and DD1 (*23*) (lanes 1 and 2), were examined. Lane 9 contains yeast tRNA (30 μg). Cells were grown under proliferation (PRO) or differentiation (DIFF) conditions. Each left (DD1 and MM14) and right (BC_3H1 and AzaM) set of panels are from the same gel. The weak band above myogenin in lane 4 is independent of Myf5. The top set of panels shows the extension of 2 μg of RNA with an MCK primer to create a 72-nt fragment. The middle set of panels shows RNase protection of 10 μg of RNA with c-*myc*, MyoD1, and MT-I probes simultaneously to create 159-, 92-, and 67-nt fragments, respectively. The bottom set of panels shows RNase protection of 6 μg of RNA with Myf5 and myogenin probes simultaneously to create 197- and 175-nt fragments, respectively. The DD1 and MM14 RNA's were purified by the guanidinium-CsCl method (*15*). Cells were immunostained with myosin heavy chain antibodies (*24*) to determine the percentage of cells that had differentiated (*23*). There were no detectable myocytes in the DD1 cultures in either type of media. MM14 cells in proliferation and differentiation media were 6 and 86 percent myosin positive, respectively. The procedure for the MCK primer extension was as described (*28*) and for RNase protection was as described (*15*).

enhancer may be important for its function (*12, 13*). In the context of other muscle specific genes, the CArG element has been shown to be important for expression (*34*). The *sph1* element has been shown to be important for the activity of the SV40 enhancer (*35*). However, neither of these sites in the MCK enhancer were detectably occupied in any of the cell lines we tested (Fig. 5) (*24*).

With the exception of MM14 myocytes, the cell types tested had no convincing protein-DNA interactions at the MCK enhancer. There was, however, a reproducible difference between in vivo and in vitro DMS treated DNA samples. It consists of two in vivo hypersensitivities at adenines −1154 and −1152, just upstream of the MEF-I site (Fig. 4, lanes 8 and 9 and 11 and 12; Fig. 5) (*24*). Although the piperidine cleavage reaction used favors strand scission at alkylated guanines, it may reveal, with reduced sensitivity, adenine residues that are particularly reactive with DMS. Hypersensitivity to DMS alkylation may be caused by torsional strain on the DNA or by proteins closely interacting with DNA to create local hydrophobic pockets (*19*). We now favor the former possibility, in part because there are no associated protected residues that, in our experience, tend to be a better general indicator of protein-DNA interactions. If this does represent protein binding, it differs from the other footprints observed because it occurs in all cell types and does not change when MCK is expressed.

In vivo footprinting of DD1, the differentiation defective derivative of MM14, permitted us to ask whether any interactions observed in fully differentiated myocytes are actually dependent on the switch from growth factor–rich proliferation medium to growth factor–poor differentiation medium, irrespective of differentiation itself. In particular, it seemed possible that in vivo binding by putative general factors like AP-2 might be regulated in response to growth signals without requiring overt muscle differentiation. Although this would be an elegant mechanism for linking withdrawal from the cell cycle with expression of differentiation specific genes, we found no support for this possibility. DD1 cells displayed no in vivo footprints under either culture condition (Fig. 4, lanes 4 to 6, and 10 to 12).

Regulation of DNA binding activities during myogenesis. All in vivo protein-DNA interactions detected at the MCK enhancer were confined to differentiated myocytes in which MCK is actively transcribed. Although the correlation of in vivo footprints with gene

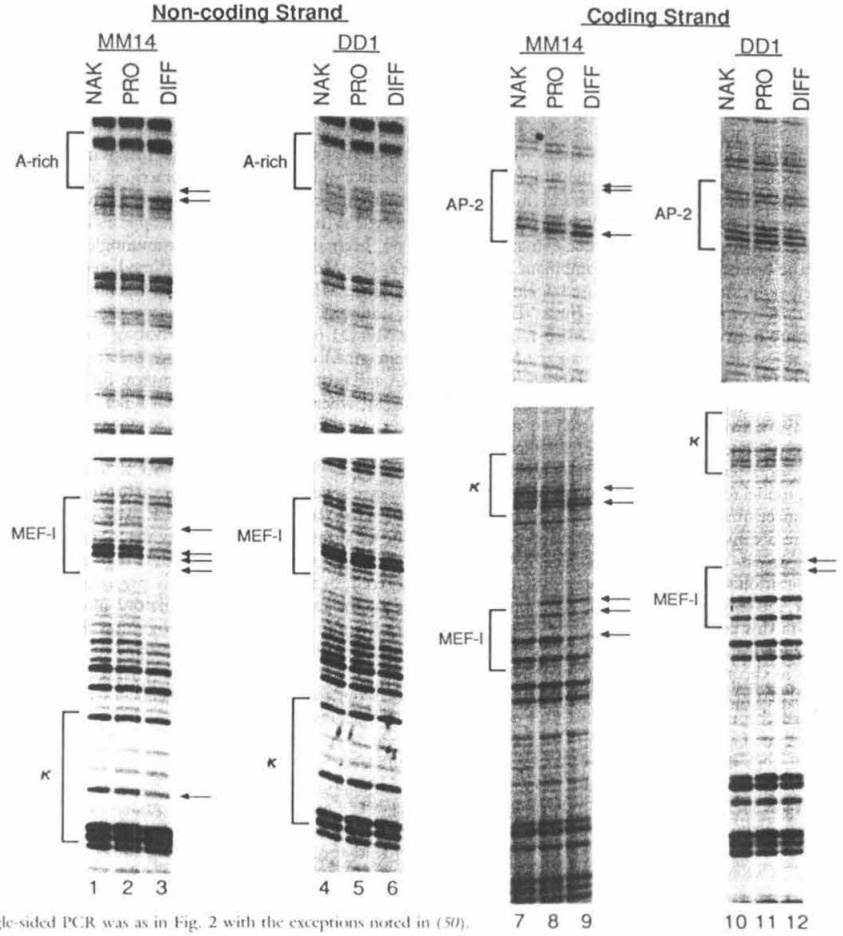


Fig. 4. In vivo footprinting of MCK enhancer in MM14 muscle cells and a differentiation defective derivative, DD1 cells. Noncoding and coding strands were visualized by ligation mediated PCR footprinting (Figs. 1 and 2). Cell lines are labeled at the top of each set of footprint ladders. In vitro DMS treated "naked" DNA (NAK, lanes 1, 4, 7, or 10) are compared to in vivo DMS treated DNA from cells grown in proliferation (PRO, lanes 2, 5, 8, or 11) or differentiation (DIFF, lanes 3, 6, 9, or 12) media. To conserve space, irrelevant parts of the footprint ladder are not shown; they contained no footprints (*24*). Brackets on the left of each footprint ladder identify the location of consensus sequences (see text for identity). Arrows on the right of each ladder mark bases that were consistently protected or hypersensitive in multiple experiments on independent preparations of DNA. These data are summarized in Fig. 5. Occasionally, we have observed spurious fluctuation in the intensity of an individual band. For example, in lane 3 between the MEF-I and κ chain enhancer–like sites there is an apparent protection, this was seen only in the experiment shown, but not in several others, and is therefore not considered a legitimate footprint. Therefore multiple experiments are necessary; in all cases shown here, multiple experiments confirmed the authenticity of the indicated footprints. MM14 and DD1 cells were grown (*27*) and treated with DMS in vivo (*15*) as described. DNA for in vitro and in vivo DMS treatment was harvested from cells as per (*15*) or by adaptation (*49*) of the RNA isolation procedure used in Fig. 3. In vitro DMS treatment (naked DNA) was as described (*15*) except that 2.5 μl of DMS per 1 ml of DNA solution was used. Piperidine treatment of DNA was as in (*15*). Ligation mediated single-sided PCR was as in Fig. 2 with the exceptions noted in (*50*).

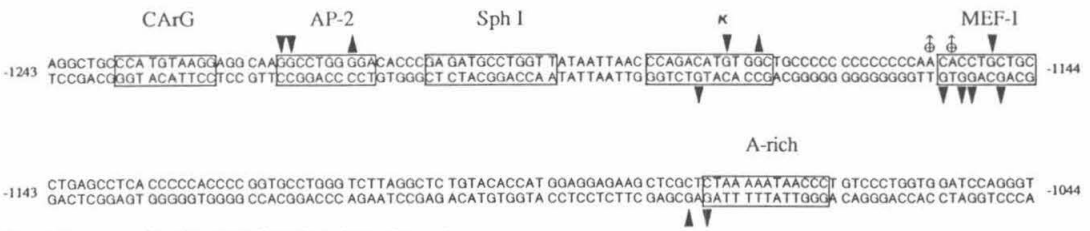


Fig. 5. Summary of in vivo DMS footprints observed over the upstream MCK enhancer. Upstream sequence (*12*) of MCK from −1243 to −1044 is shown. The enhancer as defined by (*12*) is from −1256 to −1050, and the core enhancer as defined by (*9*) is from −1207 to −1097. Sequence similarities to known factor binding sites are boxed. Changes in sensitivity to DMS (Fig. 4) are indicated; all interactions were reproducible. Identification of protected and hypersensitive bases was made by counting bases on overexposed gels and alignment with marker DNA ladders. Hypersensitivities (▲) and protections (▼) were observed in MM14 myocytes only, with the exception of the two adenine hypersensitivities (δ) at −1153 and −1155, which were observed in all cell types.

activity is direct, their relation to specific in vitro binding activities from various cell types is not so straightforward (Fig. 6). For example, even though the A-rich and AP-2–like sites are occupied only in myocytes in vivo, they can be occupied in vitro by factors present in many cell types (*9*). This suggests developmentally restricted use of general factors. An alternative possibility is that factors that ultimately occupy these sites in differentiated muscle are different from those in extracts from nonmyogenic cells. In either case, a clear understanding of how this enhancer is regulated should take into account the differential use of these recognition sites.

The MEF-I and κ chain enhancer–like sites appear to be recognition elements for myogenesis-specific factors. However, a detailed comparison of several in vitro assays with the in vivo data suggests additional regulation beyond the simple presence or absence of the factors. The MEF-I site is bound in vitro only when myocyte extracts are used, and no binding activity has been detected at the κ chain enhancer–like site in these extracts (*9*). However, both sites can be bound in vitro by recombinant MyoD1 protein (*7*). Moreover, polyclonal antibodies raised against MyoD1 recognize MEF-I–DNA complexes (*7, 36*). These data suggest a similarity or identity between MyoD1 and MEF-I, and imply that MyoD1 may be, at least in part, responsible for the MEF-I or κ chain enhancer–like in vivo footprints (Fig. 6). However, MyoD1 RNA and protein are present in both myoblasts and myocytes (Fig. 3) (*8*), whereas MEF-I activity appears restricted to myocyte extracts. Thus, the mere presence of MyoD1 is not sufficient to produce the in vivo footprints observed or to activate the MCK enhancer.

A model for the developmental regulation of this myocyte specific enhancer must accommodate both the in vivo and the in vitro data. There are two substantially different possibilities. One of these is that access of sequence specific DNA binding proteins to the recognition sites is restricted until differentiation is triggered. In this view, the failure of factors present in nonmuscle cells or myoblasts to act on this enhancer in vivo would be governed not by changes in their intrinsic activities, but by the availability of the MCK binding sites in chromatin. The idea of restricted access to developmentally regulated genes has often been discussed in the context of open and closed chromatin configurations (*37*) or covalent modifications of DNA such as methylation (*38*). Our data on the absence of in vivo binding to the MCK enhancer by factors present in all cell types is consistent with a model in which accessibility is limited. Accessibility may also be restricted if potential binding sites are already occupied by competing, sequence specific DNA binding proteins (*39*). We did not, however, detect occupancy of any myocyte specific sites in vivo in myoblasts or nonmuscle cells, nor did we find evidence of any additional sites occupied in myoblasts that become unoccupied in myocytes. While there may be undetected interactions, the complete absence of footprints in myoblasts and nonmuscle cells is striking.

An alternative possibility is that the intrinsic activity of the enhancer binding factors themselves is regulated, rather than the accessibility of their binding sites. This regulation of activity could occur by post-translational modification or via interplay among factors. If the interaction of several factors with the MCK enhancer is highly cooperative, the triggering event may be a modest change in the concentration or activity of only one of these factors. Specifically, MEF-I binding activity appears to be completely restricted to differentiating myocytes (*9*), and may therefore regulate enhancer activity by nucleating cooperative binding of general factors. MEF-I itself might be subject to further regulation. Protein-protein interactions within the *myc*-MyoD1 regulatory family could generate different binding affinities and specificities. For example, dimerization appears to be important for activity of the related E12 and E47 immunoglobulin enhancer binding proteins (*5*). This interaction is mediated through the structural similarity also found in MyoD1, myogenin, and Myf5, and by analogy, such interactions among the myogenic regulators could provide an elegant mechanism for activating and modulating muscle specific enhancers.

The picture presented for the MCK enhancer in MM14 skeletal myocytes cannot fully account for its known activity in other myogenic cell types. This enhancer can drive muscle specific transcription of a reporter gene in cardiac muscle where Myf5, myogenin, and MyoD1 expression is absent (*28*). In addition, some established MCK positive myogenic cell lines do not express MyoD1, but do express myogenin or Myf5 (Fig. 3, lanes 5 to 8) (*2, 3*). How MCK expression is activated in these MyoD1 negative cells is not yet known. One possibility is that more than one member of the *myc*-MyoD1 regulatory family could interact at the same binding site. There is precedent for such overlapping binding specificities among regulatory molecules in other systems (*40*). Alternatively, different recognition sites within the enhancer segment may be used in different types of myocytes. Comparison of the MM14 case presented here with studies of similar design in MyoD1 negative myocytes should reveal how different combinations of regulatory molecules can activate this enhancer in related but distinct myogenic cells.

Applications of single-sided ligation mediated PCR. We have developed a ligation-based, single-sided PCR strategy and applied it to genomic sequencing. Its sensitivity is high. From 1 μg of cellular DNA, the sequence of a single copy segment of a mammalian genome (about 3.3 pg per haploid genome) can be obtained from an overnight autoradiographic exposure. This opens the possibility of in vivo footprinting from small cell numbers and specific dissected tissues. In view of the high sensitivity and specificity of the

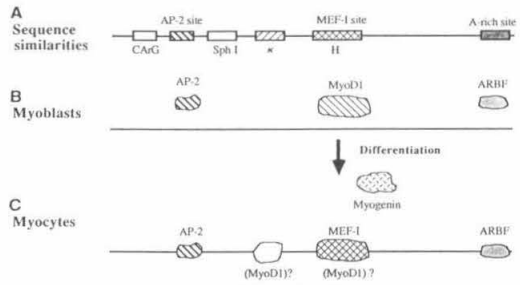


Fig. 6. Schematic representation of interactions between protein and DNA at the MCK enhancer. See text for discussion. (**A**) Location of sequence motifs similar to those of known binding factors. (**B**) Myoblasts contain factors that can bind the MCK enhancer in vitro, but no interactions are observed in vivo. (**C**) On differentiation, the MCK enhancer is occupied at four of these sequence similarities. In addition, myogenin (*2*), a factor with substantial protein similarities to MyoD1 is expressed. ARBF, A rich binding factor.

technique, we expect that the lower limit of the number of nuclei per sample will now be governed by statistical considerations, which are quite complex for this procedure (*17*). In principle, if the number of founder molecules representing a given member of the sequence ladder before amplification is too low, fluctuation among samples could give artifactual variation in the intensity of the corresponding band at the end of the amplification procedure. This might be mistaken for a legitimate footprint. In practice, we have not worked with cell numbers below 3×10^5 per reaction. We have empirically determined that under these conditions multiple analyses of the same sequence ladder are generally free of detectable under- or over-representation of individual bands. For example, the similar intensity of bands in all DNA samples from DD1 cells illustrates typical reproducibility (Fig. 4, lanes 4 to 6 and 10 to 12). The few spurious variations observed were resolved by comparing several independent experiments, and only protections or hypersensitivities that appear in each experiment are identified as footprints in Figs. 4 and 5.

A second limitation on in vivo footprinting is heterogeneity of the starting cell population. If the cells are not uniform with respect to expression of the gene of interest, footprints may be obscured by background from physiologically distinct nuclei. In our experiments, we have studied clonally derived cell lines, and care was taken to achieve physiologic uniformity: cell populations of myoblasts were as free of prematurely differentiated myocytes as possible (less than 6 to 8 percent myocytes), and differentiated myocyte preparations were harvested when very few undifferentiated precursors remained (less than 12 to 14 percent myoblasts), as determined by immunostaining for myosin heavy chain (*24*).

Ligation mediated PCR can be adapted for uses other than in vivo footprinting. For example, it has been used to determine the in vivo methylation pattern for genes subject to differential methylation during development (*17*). It can also be used to clone a new segment of genomic DNA beginning, for instance, with a primer positioned near the 5′ end of a known mRNA sequence to produce a nested series extending into an unknown promoter (*41*). The procedure has been adapted for sequencing all four bases and has been used with multiplexing (*17*), which permits several different sequences to be determined simultaneously (*42*). It may therefore be possible to conduct a genomic sequence walk through a single copy gene, beginning from a region of known sequence and proceeding through new sequence by successive steps of one-sided PCR, thereby entirely bypassing cloning steps.

For applications that require single-base resolution, this method differs significantly from another one-sided PCR strategy reported recently in which terminal transferase is used to add an oligo-dG (deoxyguanylate) or -dA (deoxyadenylate) tail to one end of each substrate molecule (*43*). This homopolymeric tail is expected to be somewhat variable in length, and terminal transferase will make additions to random single-stranded ends as well as to blunt-end duplexes. By contrast, the ligation method adds a uniform, defined sequence to the end of each molecule, and takes advantage of the high specificity of DNA ligase for a blunt-end duplexed substrate. These design features, together with other details of the ligation based procedure, reduce nonspecific background and provide the resolution required for genomic footprinting, sequencing, and methylation studies.

REFERENCES AND NOTES

1. R. L. Davis, H. Weintraub, A. B. Lassar, *Cell* **51**, 987 (1987).
2. W. E. Wright, D. A. Sassoon, V. K. Lin, *ibid.* **56**, 607 (1989); D. G. Edmondson and E. N. Olson, *Genes Dev.* **3**, 628 (1989).
3. T. Braun, G. Buschhausen-Denker, E. Bober, E. Tannich, H. H. Arnold, *EMBO J.* **8**, 701 (1989).
4. R. A. DePinho, K. S. Hatton A. Tesfaye, G. D. Yancopoulos, F. W. Alt, *Genes Dev.* **1**, 1311 (1987).
5. C. Murre, P. S. McCaw, D. Baltimore, *Cell* **56**, 777 (1989).
6. M. Caudy *et al.*, *ibid.* **55**, 1061 (1988); C. Cronmiller, P. Schedl, T. Y. Cline, *Genes Dev.* **2**, 1666 (1988); B. Thisse, C. Stoetzel, C. Gorostiza-Thisse, F. Perrin-Schmitt, *EMBO J.* **7**, 2175 (1988); R. Villares and C. V. Cabrera, *Cell* **50**, 415 (1987).
7. A. B. Lassar *et al.*, *Cell* **58**, 823 (1989).
8. S. T. Tapscott *et al.*, *Science* **242**, 405 (1988).
9. J. N. Buskin and S. D. Hauschka, *Mol. Cell. Biol.* **9**, 2627 (1989).
10. T. Braun, E. Tannich, G. Buschhausen-Denker, H. H. Arnold, *ibid.*, p. 2513.
11. R. A. Horlick and P. A. Benfield, *ibid.*, p. 2396.
12. J. Jaynes, J. E. Johnson, J. N. Buskin, C. L. Gartside, S. D. Hauschka, *ibid.* **8**, 62 (1988).
13. E. A. Sternberg *et al.*, *ibid.*, p. 2896 (1988).
14. G. M. Church and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1991 (1984); J. M. Huibregtse and D. R. Engelke, *Gene* **44**, 151 (1986); P. D. Jackson and G. Felsenfeld, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2296 (1985).
15. P. R. Mueller, S. J. Salser, B. Wold, *Genes Dev.* **2**, 412 (1988).
16. For reviews, see: R. K. Saiki *et al.*, *Science* **239**, 487 (1988); T. J. White, N. Arnheim, H. A. Erlich, *Trends Gen.* **5**, 185 (1989).
17. G. P. Pfeifer, S. D. Steigerwald, P. R. Mueller, B. Wold, A. D. Riggs, *Science* **246**, 810 (1989).
18. E. Giniger, S. M. Varnum, M. Ptashne, *Cell* **40**, 767 (1985); A. Ephrussi, G. M. Church, S. Tonegawa, W. Gilbert, *Science* **227**, 134 (1985).
19. W. Gilbert, A. Maxam, A. Mirzabekov, in *Control of Ribosome Synthesis*, Alfred Benzon Symposium IX, N. O. Kjeldgaard and O. Maaloe, Eds. (Academic Press, New York, 1976), pp. 139–148; L. Johnsrud, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5314 (1978).
20. A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
21. For example, see: P. Gunning *et al.*, *Mol. Cell. Biol.* **7**, 4100 (1987).
22. T. A. Lindhart, C. H. Clegg, S. D. Hauschka, *J. Supramol. Struct.* **14**, 483 (1980).
23. R. W. Lim and S. D. Hauschka, *Dev. Biol.* **105**, 48 (1984).
24. P. R. Mueller and B. Wold, unpublished results.
25. J. S. Chamberlain, J. B. Jaynes, S. D. Hauschka, *Mol. Cell. Biol.* **5**, 484 (1985).
26. J. B. Jaynes, J. S. Chamberlain, J. N. Buskin, J. E. Johnson, S. D. Hauschka, *ibid.* **6**, 2855 (1986).
27. J. E. Johnson, C. L. Gartside, J. B. Jaynes, S. D. Hauschka, *Dev. Biol.* **134**, 258 (1989).
28. J. E. Johnson, B. Wold, S. D. Hauschka, *Mol. Cell. Biol.* **9**, 3393 (1989).
29. P. J. Mitchell, C. Wang, R. Tjian, *Cell* **50**, 847 (1987).
30. M. Imagawa, R. Chiu, M. Karin, *ibid.* **51**, 251 (1987).
31. D. Picard and W. Schaffner, *Nature* **307**, 80 (1984).
32. J. Banerji, L. Olson, W. Schaffner, *Cell* **33**, 729 (1983).
33. J. E. Johnson, thesis, University of Washington, Seattle (1989).
34. For review, see A. Taylor, H. P. Erba, G. O. Muscat, L. Kedes, *Genomics* **3**, 323 (1988).
35. I. Davidson *et al.*, *Nature* **323**, 544 (1986).
36. J. Buskin, A. B. Lassar, R. L. Davis, H. Weintraub, S. D. Hauschka, *J. Cell Biol.* **107**, 98a (1988).
37. For review, see H. Weintraub, *Cell* **42**, 705 (1985).
38. For reviews, see A. D. Riggs and P. A. Jones, *Adv. Cancer Res.* **40**, 1 (1983); R. Holliday, *Science* **238**, 163 (1987).
39. K. Zinn and T. Maniatis, *Cell* **45**, 611 (1986).
40. C. Desplan, J. Theis, P. H. O'Farrell, *ibid.* **54**, 1081 (1988); T. Hoey and M. Levine, *Nature* **332**, 858 (1988); J. B. Jaynes and P. H. O'Farrell, *ibid.* **336**, 744 (1988).
41. L. Fors, R. Saavedra, L. Hood, unpublished results.
42. G. M. Church and S. Kieffer Higgens, *Science* **240**, 185 (1988).
43. M. A. Frohman, M. K. Dush, G. R. Martin, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998

(1988); E. L. Loh, J. F. Elliott, S. Cwirla, L. L. Lanier, M. M. Davis, *Science* **243**, 217 (1989).
44. MT-I primers: 1, CGGAGTAAGTGAGGAGAAGGTACTC; 2, GGAGAAGGTACTCAGGACGTTGAAG; 3, GAAGTACTCAGGACGTTGAAGTCGTGG.
45. For hybridization, the oligomers of the linker (see Fig. 1B for sequence) were brought to a final concentration of 20 pmol/μl in 250 m*M* tris, *p*H 7.7, heated to 95°C for 5 minutes, transferred to 70°C, and then slowly cooled (~3 hours) to 4°C. The hybridized linker was stored at −20°C and thawed on ice as needed.
46. The labeled oligomer was phosphorylated with crude [γ-^{32}P]ATP (DuPont) and T4 polynucleotide kinase (NEB), and unincorporated ^{32}P was removed by Nensorb-20 columns (DuPont). Specific activity of the primer was 4×10^6 to 9×10^6 cpm/pmol.
47. D. Schubert *et al.*, *J. Cell. Biol.* **61**, 398 (1974).
48. A. B. Lassar, B. M. Paterson, H. Weintraub, *Cell* **47**, 649 (1986).
49. E. Meese and N. Blin, *Gene Anal. Tech.* **4**, 45 (1987).
50. For footprinting the MCK enhancer, we used 2 μg of DNA per reaction. The Sequenase reaction was performed at 45°C with 0.1 m*M* of each dNTP at and 0.1 pmol of primer. The final concentration of $MgCl_2$ in the Taq polymerase reactions was 10 m*M*. Although the conditions we used in this figure are acceptable, those in Fig. 2 are recommended; also see (*17*) for additional improvements. In lanes 1 to 3, ligation of the linker was allowed to proceed for only 2 hours; for all other lanes, ligation was for the usual 10 to 12 hours. This shortened ligation time did not change the pattern of footprints (*24*), but did result in a reduced intensity of particular regions of the sequence such as the stretch of guanines between the κ chain enhancer–like and MEF sites [for example, compare lane 1 with lane 4, also see (*17*)]. Experiments with longer ligation times confirmed that there was no observable footprint in this region (*24*). MCK coding strand primers used were: 1, CAAACCTGCGGCTGAGGGGAAGTGG; 2, CTGCCCCTCACCCTGGATCCACCAG; 3, CTGCCCCTCACCCTGGATCCACCAGGG. MCK noncoding strand primers used were: 1, GCTCTGGTCTGCCTTCCACAGCTTG; 2, CATATTGTGTCCTGCTCTGGTCTGC; 3, GCTCTGGTCTGCCTTCCACAGCTTGGG. Because of the different length and G·C content of primers, different temperatures were used during the Taq polymerase hybridizations: coding strand primer 2, 66°C; coding strand primer 3, 72°C; noncoding strand primer 2, 66°C; noncoding strand primer 3, 69°C.
51. We thank J. Miner for providing template plasmids used in the RNase protection assays and aza-myoblast RNA's; S. Hauschka for the MM14 and DD1 cell lines; S. Sharp for BC_1H1 RNA's; P. Garrity, J. Johnson, U. Landegren, H. Weinhard, and P. Mathers for helpful discussions; S. Hauschka, J. Buskin, H. Arnold, E. Olsen, A. Lassar, and H. Weintraub for supplying information before publication; and N. Davidson, D. Anderson, K. Zinn, A. Riggs, G. Pfeifer, S. Steigerwald, and members of the Wold group for critical reading of this manuscript. Supported by NIH (General Medicine) grants RR07003 and GM35526 (B.W.). B.W. would like to dedicate this work to D.E.W.

9 July 1989; accepted 20 September 1989

"Tit for tat, I suppose. He was a vegetarian."

Chapter 3

Constitutive and Metal Inducible Protein:DNA Interactions at the Mouse Metallothionein-I Promoter Examined by *In Vivo* and *In Vitro* Footprinting

(This chapter was published in *Genes and Development*, **2**: 412-427, 1988)

Constitutive and metal-inducible protein:DNA interactions at the mouse metallothionein I promoter examined by in vivo and in vitro footprinting

Paul R. Mueller, Stephen J. Salser,[1] and Barbara Wold[2]

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

A method of high resolution in vivo footprinting has been developed and used to survey the mouse metallothionein I (MT-I) promoter for protein : DNA interactions associated with basal-level transcription and with high-level metal-induced transcription. This promoter and its associated regulatory region is structurally complex. It contains multiple potential binding sites for metal regulatory factors and for other transcription factors, including SP1 and MLTF. In several cases potential recognition sites overlap, and the experiments reported here provide a view of which sites are utilized in vivo. These data also show how the pattern of protein : DNA contacts changes when cells are shifted from basal-level expression to metal-induced expression. The noninduced footprint pattern consists of interactions at basal elements that are thought to be responsible for the moderate transcription of this gene in the absence of added metals. These interactions remain unchanged upon metal induction. When MT-I expression is increased by exposing cells to zinc or cadmium, a new footprint pattern is observed. It includes the basal interactions and a new set of metal-dependent footprints that are positioned over all five genetically defined metal responsive elements (MREs), MRE-A–MRE-E. In addition, these data identify a sixth probable MRE, MRE-F, which displays a dimethylsulfate (DMS) footprint similar to that at other MREs.

[*Key Words*: Metallothionein gene; in vivo footprinting; transcription; SP1; MLTF]

Received November 16, 1987; revised version accepted March 8, 1988.

The structure of most eukaryotic promoters and associated regulatory regions is complex, consisting of an array of multiple, interdigitated sequence elements that act as recognition sites for the proteins that govern transcription (Dynan and Tjian 1985; reviewed by Maniatis et al. 1987). The task of identifying physiologically relevant sites and dissecting their individual functions is complicated by several factors: Functional binding sites are not readily discriminated from nonfunctional sites by primary sequence alone; a given segment of DNA sometimes serves as a binding site for two or more different factors; and it is often difficult to generate mutants that ablate all copies of a recognition site without also distorting the remaining sequences and their spatial relationships. In vivo footprinting studies can circumvent some of these problems by providing information on when and how recognition sites are occupied in the cell (Ephrussi et al. 1985; Giniger et al. 1985; Jackson and Felsenfeld 1985; Zinn and Maniatis 1986; Kemper et al. 1987). In this work we use in vivo footprinting of the mouse metallothionein-I (MT-I) promoter to begin to characterize DNA : protein contacts as they occur in the cell under noninducing and inducing conditions.

Metallothionein genes (MT) code for small, cysteine-rich proteins that bind heavy metals, and they are present in all higher eukaryotes surveyed (reviewed by Karin 1985; Hamer 1986). Expression of the murine MT-I gene is regulated in response to several different physiological cues. Mouse MT-I is typically expressed in cultured cells at a significant basal level of about 10^2 mRNAs per cell (P. Mueller, unpubl.); in the presence of added zinc or cadmium the mRNA level increases 5- to 20-fold. This metal-dependent accumulation of MT-I RNA is primarily due to increased transcription (Durnam and Palmiter 1981; Hamer and Walling 1982). A segment of the gene containing the promoter and flanking sequences has been shown to direct both basal and induced transcription of adjacent heterologous sequences (Brinster 1982; Mayo et al. 1982; Pavlakis and Hamer 1983). DNA transfection studies of deletion mutations and synthetic promoters, together with DNA sequence data from many MT genes, have led to the identification of sequence elements that are responsible for metal induction (metal responsive elements or MREs) (Carter et al. 1984; Stuart et al. 1984, 1985; Searle et al. 1985). Other sequence elements influence the levels of

[1]Present address: Department of Biochemistry, University of California, San Francisco, California, USA.
[2]Corresponding author.

both basal and induced expression. In vitro transcription of the related human MT-IIA gene showed that basal transcription can be increased by the addition of SP1 (Lee et al. 1987a) and other factors including AP1 and AP2 (Lee et al. 1987b; Mitchell et al. 1987), but metal-dependent transcription has not yet been established in vitro, and metal dependent in vitro footprints have been confined to a single cadmium-specific site near MRE-D (Seguin and Hamer 1987).

The in vivo footprints presented here define two patterns of apparent protein : DNA contacts associated with basal level transcription and metal induced transcription of the murine MT-I gene. The footprinting strategy uses cells containing transfected, amplified MT genes to increase sensitivity and a genomic primer extension reaction to provide high resolution footprints. Our observations, taken together with existing genetic data, suggest that proteins bound constitutively at basal elements support transcription in the absence of metal induction, and also amplify the increase in transcription mediated by MREs in the presence of metals. These results focus attention on the potential for in vivo competition among factors at overlapping recognition sites, and also provide an explicit set of predictions for DNA-binding patterns expected from in vitro studies.

Results

In vivo footprinting strategy

In this work two different methods have been used to visualize in vivo protein : DNA contacts on the mouse MT-I promoter. The highest-resolution picture of these interactions was obtained by treating intact cells with dimethylsulfate (DMS). DMS methylates guanine residues at the N7 position, rendering them susceptible to subsequent cleavage with piperidine (Maxam and Gilbert 1980). Cellular proteins bound at or near specific guanines can either enhance or reduce the frequency of DMS methylation relative to the same residues in naked DNA (Gilbert et al. 1976). Because cell membranes are freely and rapidly permeable to DMS, this experiment permits detection of DNA : protein contacts in intact cells that have undergone a minimum of physiological disruption. DMS was originally used for this purpose by Giniger et al. (1985) and Ephrussi et al. (1985), and in the work presented here the procedure was adapted for mouse L cells.

The second method used to detect protein : DNA interactions is DNase I treatment of nuclei. Because DNase is a large molecule and does not freely cross the cell membrane, cells must be lysed to expose nuclei for digestion. Although such nuclei are probably not initiating new rounds of transcription (Groudine et al. 1981), it is clear from prior DNase studies of several different promoters that many specific protein : DNA contacts can be detected (Jackson and Felsenfeld 1985, 1987; Zinn and Maniatis 1986; Gimble and Max 1987). Moreover, DNase I footprints identify protein : DNA interactions at all four residues. Therefore we use this method to complement data obtained by DMS treatment.

DNase and DMS footprints were visualized using a modification of the genomic sequencing procedure of Huibregtse and Engelke (1986). Cleaved genomic DNA was hybridized with a vast excess of a specific, end-labeled oligonucleotide primer under conditions that minimize self-renaturation of genomic DNA. The majority of the unhybridized primer was removed, and the bound primer was extended with AMV reverse transcriptase. Finally, the extension products were separated on a sequencing gel.

The in vivo footprint experiments were performed on L-cell lines transfected with a mouse MT-I–dihydrofolate reductase (DHFR) fusion gene. The gene contains 1775 bp of the MT-I promoter and 5′-flanking sequence fused to the coding region of DHFR cDNA (Fig. 1a). This segment of the MT-I gene has been shown to direct both basal and metal-induced gene expression when transfected into cultured cells or mouse eggs (Brinster et al. 1982; Mayo et al. 1982; Pavlakis and Hamer 1983). The cell lines used here, A-0.3 and A-60, were selected because they contain multiple copies of the fusion gene, thus increasing the sensitivity of the footprint experiments. Expression of the fusion genes follows the pattern expected for MT-I regulated genes (Fig. 1b and P. Mueller, S. Salser, and B. Wold, unpubl.). In the absence of added metals, A-60 and A-0.3 cell lines display characteristic basal-level expression; upon stimulation with zinc this level increased by 5- and 10-fold, respectively.

A DMS footprint of the MT-I promoter in A-60 cells under both noninducing and inducing conditions is shown in Figure 2. Pairwise comparison of in vitro DMS-treated DNA, ('naked' DNA, Fig. 2, lanes 5 and 10) with companion in vivo DMS-treated samples from cells grown in the presence or absence of heavy metal induction (Fig. 2, + zinc, lanes 4 and 9; no zinc, lanes 3 and 8) reveals several footprints in the region between −35 and −190, and these are discussed individually in the sections that follow. By contrast, there are no detectable interactions between −30 and +20, except a single, small, but reproducible, hypersensitivity localized just before the transcription start site. The 25-base oligonucleotides used to prime reverse transcription of cleavage products are complementary to residues −227 to −203 on the noncoding strand and to residues +47 to +23 on the coding strand. With these primers we can visualize both strands of the metal-responsive promoter sequence, as defined previously in studies of 5′ and 3′ deletion mutations (Carter et al. 1984; Stuart et al. 1984; Searle et al. 1985). Footprint experiments were performed several times on both cell lines, and DNA preparations from these DMS treatments were often assayed in multiple independent footprints. All interactions detected were highly reproducible. Densitometric traces of the autoradiograms were made, and the results are summarized in Figure 3, in which the height of the arrows is proportional to the degree of protection or enhancement at a given site.

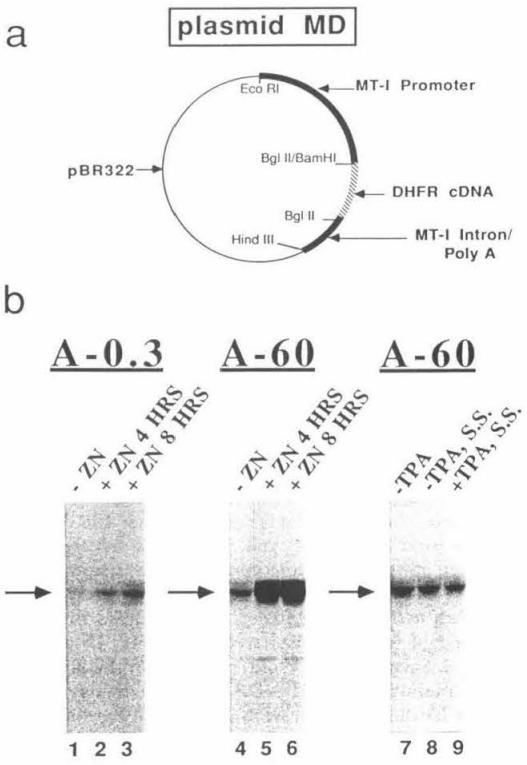


Figure 1. Zinc induction of MT/DHFR. (*a*) Map of plasmid MD which contains ~7458 bp. Thin lines are pBR322, thick solid lines are mouse MT-I, and thick striped lines are mouse DHFR. Construction is described in Materials and methods, (*b*) MT-I/DHFR RNA measurements. RNA (15 μg total was hybridized with a uniformly labeled, 126-base riboprobe that was complementary to 80 bases (→) of the MT-promoted DHFR mRNA as per Zinn et al. (1983) and Material and methods. The endogenous DHFR is complementary to 65 bases of this riboprobe, but can not be seen on these short exposures (≤30 min). (Lanes *1–3*), A-0.3 RNA; (lane *1*), without zinc; (lane *2*) 130 mM zinc sulfate for 4 hr; (lane *3*) 130 mM zinc sulfate for 8 hr. (Lanes *4–9*) A-60 RNA; (lane *4*) without zinc; (lane *5*) 130 mM zinc sulfate for 4 hr; (lane *6*) 130 mM zinc sulfate for 8 hr; (lane *7*) same as lane 4; (lane *8*), 0.5% serum for total of 36 hr; (lane *9*), 0.5% serum for total of 36 hr and 60 ng/ml TPA for the last 12 hr.

SP1 binds to the MT-I promoter in vivo and in vitro

A sequence search of the MT-I promoter reveals two good matches to the consensus recognition site for the transcription factor SP1 (Kadonaga et al. 1986; Jones et al. 1986) that are centered at −182 and −139, but it is not known whether either actually functions as an SP1 binding site. Comparison of the in vitro (Fig. 2a, lane 10) and in vivo (Fig. 2a, lanes 8 and 9) guanine ladders near base −182 on the coding strand shows a prominent in vivo footprint. Figure 2b shows an enlargement of this footprint together with densitometric scans; the data are summarized in Figure 3. The sequence within this region matches the SP1 GGCGGG core perfectly and matches the extended consensus site at 8 of 10 bases (Jones et al. 1986). For convenience, this is referred to as the SP1-A site, while the more proximal sequence at −139 is referred to as the SP1-B site. The similarity between the pattern of in vivo DMS sensitivity at SP1-A and that observed using purified SP1 in vitro is striking. The consensus in vitro pattern is predominantly on one strand, and shows protection of the first 4 bases and hypersensitivity at the 6th and/or 7th bases (Gidoni et al. 1984). The in vivo footprint at the SP1-A (−182) site is virtually identical; the first four guanines are protected ~70%, and the guanine at position 7 is ~threefold hypersensitive. Also in agreement with in vitro experiments (Jones et al. 1986) are the less prominent interactions detected on the noncoding strand, with 35–40%

Figure 2. (*a*) In vivo DMS footprinting of the mouse MT-I promoter. Noncoding and coding strands are visualized by primer extension. The first and second loadings of the reaction products on a 60-cm gel are placed on top of each other with a slight space between them. Distance from the start of transcription (TXN) is shown to the left of each strand with '-' indicating 5' from start and '−' indicating 3' from start. Potential sites for binding of transcription factors (based on DNA sequence and genetic data, see text) are shown by brackets to the right of each strand. A-60 cells were the source of the DNAs. (DMS) dimethylsulfate; (PIP) piperidine; (NA) naked DNA; (−ZN), DNA methylated in vivo from noninduced cells; (+ZN) DNA methylated in vivo from zinc-induced cells. (Lanes *1* and *5*) Control samples containing in vitro purified (naked) DNA that was not treated with DMS or piperidine. (Lanes *2* and *6*) Control samples containing in vitro purified (naked) DNA that was only treated with piperidine. (Lanes *3* and *7*) DNA from noninduced cells treated with DMS in vivo and piperidine in vitro. (Lanes *4* and *9*) DNA from cells induced with 130 mM zinc sulfate for 4 hr before being treated with DMS in vivo and piperidine in vitro. (Lanes *5* and *10*) In vitro purified DNA (naked) treated with DMS and piperidine in vitro. See Materials and methods for experimental details and Figure 3 for summary of data. (*b*) Detail of SP1-A coding strand footprint. (*Enlarged inset*) Shorter exposure of coding strand over SP1-A region enlarged to show detail and constitutive nature of the SP1 in vivo footprint. (NA) Naked DNA methylated in vitro; (−ZN) DNA methylated in vivo from noninduced cells; (+ZN), DNA methylated in vivo from zinc induced cells. Vertical lines beneath photo identify guanines in sequence. Densitometry plots show the initial, normalized densitometry traces of the same region enlarged above, and are superimposed on each other as indicated. Difference plots illustrate regions of footprinting in noninduced and induced cells, and were obtained by subtracting the superimposed densitometry plots from each other as indicated. Horizontal baseline represents zero difference or no footprint; traces below it represent hypersensitivity; and traces above it represent protection. These data are quantitated in Figure 3.

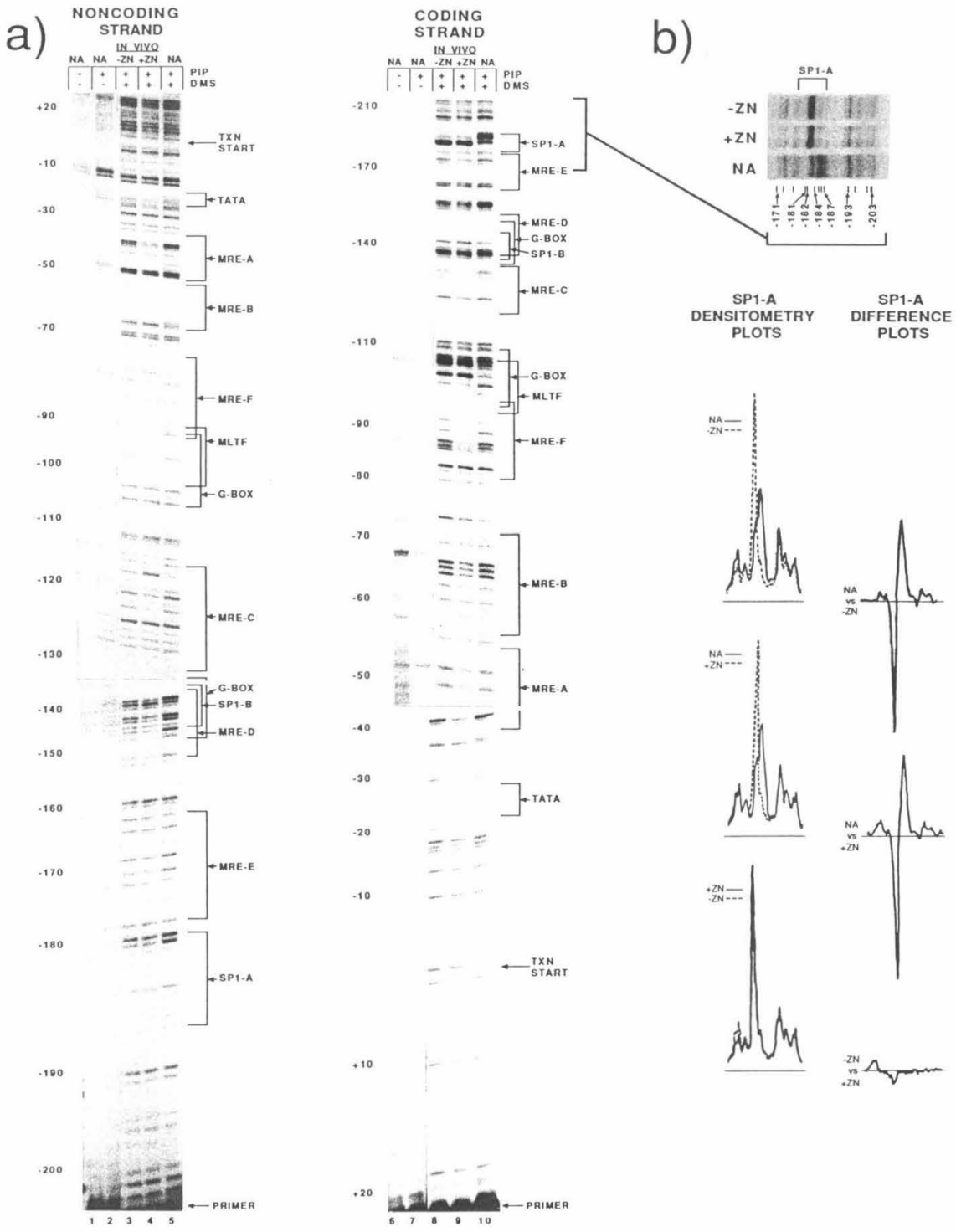


Figure 2. *(See facing page for legend.)*

protection at the fifth and tenth bases. Specific guanine residues are protected from DMS attack when those residues are closely associated with a protein. If binding of a particular protein results in complete exclusion of DMS from guanines in its recognition site, 100% protection will be observed when all copies of that site are occupied continuously. Guanines −184 to −187 of the SP1-A site are 70% protected in vivo, suggesting that at least 70% of the MT-I genes are occupied throughout the 2-min DMS treatment or, alternatively, all copies in all cells are occupied 70% of the time on average. In contrast to protection from DMS attack, quantitation of hypersensitivity cannot be interpreted in a similarly straightforward fashion. It is thought that hypersensitivity to DMS

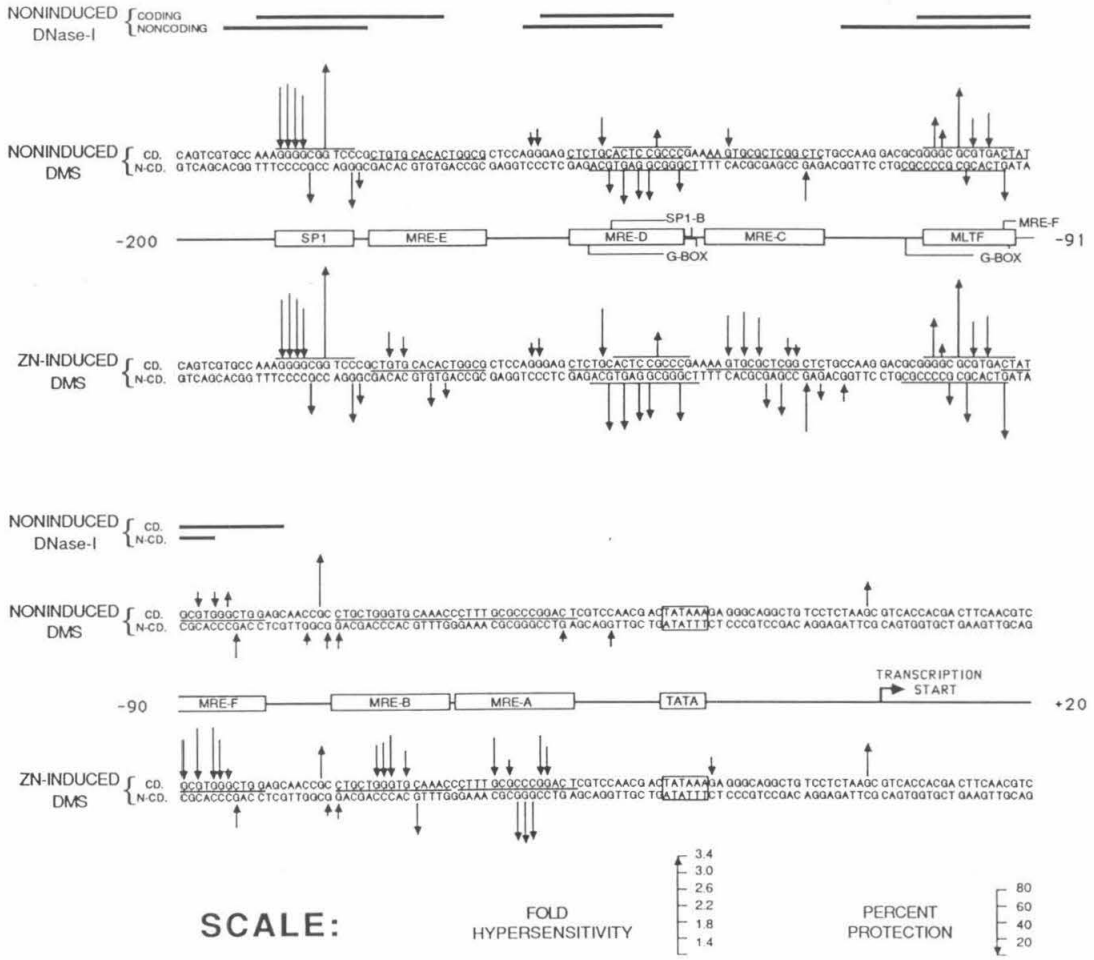


Figure 3. Summary of in vivo DMS and DNase footprinting over the mouse MT-I promoter. The sequence of both strands of the mouse MT-I promoter are shown. (N-CD) Noncoding strand; (CD) coding strand. Base pairs −200 through −91 are in the top half of the figure, and base pairs −90 through +20 are in the bottom half of the figure. Potential recognition sites are shown schematically between the noninduced and induced sequences. These sites are also localized directly on the sequences by horizontal lines: (SP1 and MLTF) a line over the coding strand; (MRE) a line between the strands; (G element) a line below the noncoding strand; (TATA box) a box around both strands. Changes in methylation patterns (Fig. 2) are shown directly over the coding strand and under the noncoding strand; protections are indicated by downward pointing arrows (↓), and hypersensitivities are indicated by upward pointing arrows (↑). See the scale at the bottom of the figure for quantitation. All protections ≥15% and hypersensitivities ≥1.3-fold are shown. Regions of protection from DNase digestion under noninduced conditions (Fig. 4) are shown for both strands by a set of thick double lines above the noninduced DMS footprints. The sequence of this promoter is as per Glanville et al. (1981), except that an additional adenine was found at −112 (S.J. Salser, unpubl.) and is added to the sequence and numbering throughout this paper.

results when protein : DNA interactions create a local hydrophobic environment around a target guanine (Gilbert et al. 1976). It is also possible that local changes in DNA conformation might alter DMS sensitivity (Johnsrud 1978). Given the nature of hypersensitivity, it is not possible to deduce the degree of occupancy by comparison with DMS-treated naked DNA, but it is informative to compare hypersensitivity at a given site under noninducing and inducing conditions.

In agreement with the DMS data, the SP1-A site in MT-I is protected from DNase I digestion of nuclei, indicating that protein is bound at this site. A DNase I footprint of A-0.3 nuclei is shown in Figure 4. Comparison of the DNase I pattern on naked DNA (Fig. 4, lanes 2 and 6) and on nuclei of noninduced cells (Fig. 4, lanes 4 and 8) reveals a protected region of about 20 nucleotides at the SP1-A site in nuclei. This pattern of DNase protection is in agreement with the pattern observed for SP1 binding to the SV40 promoter in vitro (Gidoni et al. 1984, 1985; and see below) and in vivo (Buchanan and Gralla 1987).

The second prospective SP1 site, SP1-B, is centered at −139. It is a 9 of 10 match with the consensus element, includes a perfect match to the GGCGGG core, and is inverted in orientation relative to the site at −182. A DMS footprint is observed at this site, but it is much less intense than the one at the SP1-A site, and the pattern of protection and hypersensitivity is quite different from the SP1 DMS consensus pattern defined in vitro (Gidoni et al. 1984; Jones et al. 1986) and observed for the SP1-A site in vivo. The DNase I experiment corroborates the DMS data, showing a region of protection surrounding the −139 sequence (Fig. 4).

The data presented above suggest that the site at −182 is a bonafide SP1 site that is occupied in mouse L cells. The proximal site, on the other hand, shows an in vivo footprint that bears little resemblance to a simple SP1 interaction. Although the footprint at the B site does not look like a typical SP1 interaction, it remains possible that in the living cell multiple factors, including SP1, compete for this site. To determine directly if, in the absence of competing factors, bona fide SP1 binds anywhere on the MT-I promoter, an in vitro footprint experiment was done using purified SP1 (a gift of J. Kadonaga and R. Tjian.). Varying amounts of homogeneous

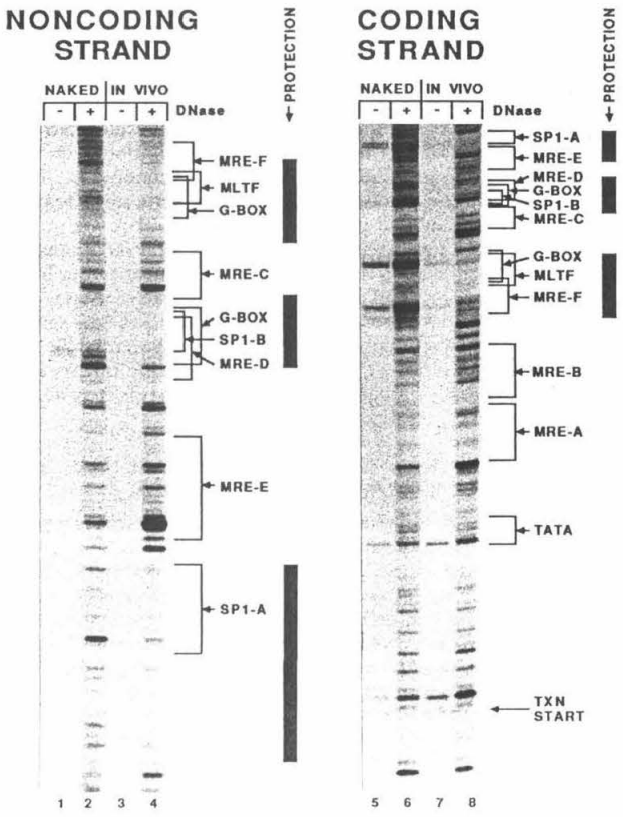


Figure 4. In vivo DNase I footprinting of the mouse MT-I promoter. The MT promoter was footprinted using DNase I as a cleaving agent as in Zinn and Maniatis (1986) and the footprint was detected by primer extension (see Fig. 2 legend and Materials and methods). The noncoding strand is shown on the right and the coding strand is shown on the left, and potential factor binding sites are indicated by brackets next to the photo. Regions of DNase protection are shown by the thick vertical lines to the left of the brackets. These were assigned by comparison of naked and in vivo patterns. The boundaries of protected regions are approximate due to typical sequence preferences in the DNase I digestion. (Lanes *1* and *5*) In vitro purified (naked) DNA not treated with DNase I; (lanes *3* and *7*) DNAs purified from nuclei without the addition of exogenous DNase I; (lanes *2* and *6*) in vitro purified (naked) DNA treated with DNase I in vitro; (lanes *4* and *8*) DNAs form nuclei treated with DNase I in vivo. These data are summarized in Fig. 3.

SP1 were used to protect MT-I and SV40 promoter fragments from DNase I digestion (Fig. 5). The only sites protected in MT-I are the two identified above by in vivo footprinting and sequence similarity. The distal site is protected at lower SP1 concentrations than is the proximal site. The size of the region protected in both cases is 18–20 bases, in agreement with the in vivo DNase footprint and with other in vitro SP1 footprints (Gidoni et al. 1984; Jones and Tjian 1985; Jones et al. 1986). In this experiment, the parallel SV40 footprint provides a convenient, well-characterized standard for SP1 binding. The distal murine MT-I site binds human SP1 with an affinity between that of strong and intermediate SV40 sites, whereas binding at the proximal MT site is similar to that at weak SV40 sites. In summary, all in vivo and in vitro data suggest that the sequence at −182 of the MT promoter is occupied by a murine SP1 protein in vivo. This is consistent with the fact that deletion analyses of MT-I indicate that sequence within this region contributes to both basal and induced expression (Carter et al. 1984; Searle et al. 1985). By contrast, the proximal SP1-B site at −139 presents a more complicated picture. Apparently it is a binding site for one or more proteins in vivo, and it can serve as an SP1 site in vitro, albeit a weak one. Whether SP1 is among the proteins that bind there in vivo remains uncertain (see below).

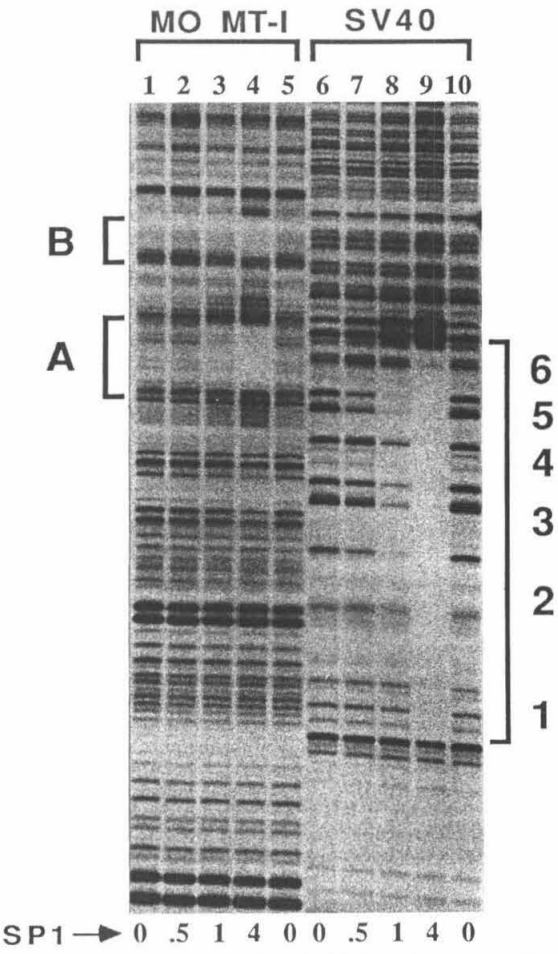


Figure 5. In vitro binding of purified SP1 on the mouse MT-I promoter. The indicated amounts of SP1 (in microliters) were incubated with end-labeled MT (lanes *1–5*) and SV40 (lanes *6–10*) promoter sequences and the subjected to DNase I treatment in vitro (see Materials and methods). Regions of footprinting are indicated by brackets on the side of the photo. (A) MT-I SP1-A site; (B) MT-I SP1-B site (see text), and SV40 sites 1–6 as per Gidoni et al. (1984).

Some genetic analyses suggest that sequences which influence basal expression also amplify metal induction (Carter et al. 1984; Searle et al. 1985; Karin et al. 1984; Karin and Holtgreve 1984; Stuart et al. 1984). Is the SP1-like in vivo footprint in Figure 2 affected, either qualitatively or quantitatively, by metal induction? The data in Figures 2 and 3 show clearly that there is no major change in the footprint at the SP1-A site upon zinc induction. This is best visualized by inspection of the enlarged footprint in the inset (Figure 2b, top), and is illustrated in the densitometric scans and difference plots (Figure 2b, bottom); and is summarized in Figure 3. The difference plots of noninduced and naked DNA or induced and naked DNA pairs show clearly the SP1 footprint, and the difference plot between induced and noninduced samples reveals the striking similarity of the SP1-A footprint in cells under both conditions. The same is true for cadmium induction (data not shown). We conclude that if SP1 binding at the SP1-A site influences the level of expression during metal induction, it does not do so by simply increasing occupancy nor by a steric change that is strongly DMS sensitive. Metal induction does have a detectable effect on the DMS- sensitivity of some guanines in the second potential SP1-B site located at −139, but the juxtaposition of this site with possible binding sites for other proteins may be responsible (see Discussion).

MLTF binds the murine MT-I gene in vivo

Another prominent set of hypersensitive and protected guanine residues comprise a clear in vivo DMS footprint between positions −94 and −105 (Figs. 2a and 3). This region is also protected from DNase I digestion in nuclei of noninduced cells (Fig. 4). Inspection of the sequence in this region revealed a 9 of 12 bp match to the recognition site of the major late transcription factor (MLTF), a cellular factor that was first identified by its activity at the major late promoter of adenovirus 2 (Ad 2) (Carthew et al. 1985; Miyamoto et al. 1985; Sawadogo and Roeder 1985; Chodosh et al. 1986). Comparison of our in vivo DMS footprint with the in vitro DMS footprint for purified MLTF at the Ad2 promoter (Miyamoto et al. 1985)

shows that the patterns of protection from methylation are very similar (Fig. 6). Moreover, recent in vitro footprinting experiments show that purified MLTF binds at this site in the murine MT-I promoter (Carthew et al. 1987). This region also contains a weaker sequence similarity (5 of 7) to the recognition site of transcription factor AP1, which has been shown to function in the human MT-IIA gene (Lee et al. 1987a,b). AP1 activity can be modulated by the phorbol ester TPA (Angel et al. 1987), and genes containing an AP1 element can be induced by treating cells with TPA (Angel et al. 1986, 1987; Imbra and Karin 1987). In an initial experiment, we tested whether expression of the MT-I genes in our A-60 cells is elevated in response to TPA. The data in Figure 1b show that the steady-state level of RNA was not affected by TPA treatment. This experiment is not decisive, but when the result is considered together with divergence from the consensus site, it suggests that AP1 is not responsible for the footprint centered around −100. We conclude that murine MLTF is bound to the MT-I gene in vivo at the site between −93 and −104. Our data also show that there is little change in occupancy at this site upon metal induction (Fig. 2a and summarized in Figs. 3 and 6).

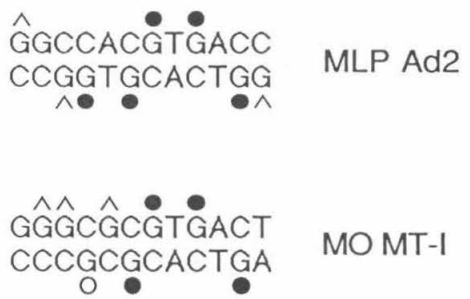


Figure 6. Comparison of in vitro and in vivo MLTF DMS footprints. The in vitro DMS footprint pattern over the MLTF site between −63 and −52 in the major late promoter of adenovirus 2 (Miyamoto et al. 1985) is compared to the in vivo DMS footprint pattern over the MLTF site between −104 and −93 in the mouse MT-I promoter (Fig. 2a and 3). Carrots (^), hypersensitivity; filled circles (●), protection under noninduced and Zn-induced conditions; and open circles (○), protection only observed under induced conditions.

Metal dependent protection of MRE sequences

All MT genes contain multiple copies of MREs (Carter et al. 1984; Karin et al. 1984; Stuart et al. 1984, 1985; Anderson et al. 1986; Maroni et al. 1986). When two or more MREs are inserted upstream from a basal promoter, they can confer metal responsiveness upon that promoter (Searle et al. 1985; Stuart et al. 1985). MREs are a family of moderately similar sequence of about 15 bp that are related to a rather loosely defined consensus sequence: CTNTCC[A/G]CNCGGCCC, in which the underscored bases comprise the core consensus element. The five MREs of the mouse MT-I gene have been designated MRE-A through MRE-E, with MRE-A being closest to the start of transcription and MRE-E most distant (Stuart et al. 1984).

In noninduced cells, MREs A–E display a DNase I pattern that is essentially identical with naked DNA (Fig. 4), and they show very little protection from in vivo DMS treatment (Figs. 2, 3, 7), with the exception of MRE-D, which is discussed below. Upon stimulation with zinc, all of the MREs show protection of most guanines in vivo. The degree of protection ranges from 20% to 65% and corresponds well with the ability of each individual MRE to confer metal response upon a heterologous promoter (Stuart et al. 1985). For example, MRE-E is unable to confer metal inducibility on the HSV-TK promoter, and it shows very little DMS protection here. The converse is true for elements A and C, which are more highly protected and are also potent metal response elements in synthetic promoters. Detectable protection from DMS is dependent on zinc induction at all sites except MRE-D, where metal treatment enhances a preexisting low-level protection. Slight hypersensitivity at the last three or four positions of most MREs is also observed. Figure 7 contains a summary of DMS footprints at MREs together with representative densitometry. A simple interpretation of the footprint pattern is that in the noninduced state MREs are not occupied, but upon metal induction they are protected due to the binding of positive-acting metal-responsive factors (MRF). This agrees with data from in vivo competition studies which show that the limiting factors in metal induction act in a positive fashion (Seguin et al. 1984; Scholer et al. 1986; Mueller et al., in prep.). While this is the simplest view consistent with all data, the absence of a footprint under noninducing conditions must be interpreted cautiously, because it is expected that some interactions will escape detection, especially if they are relatively unstable or transitory.

Virtually all metal-dependent DMS protections can be accounted for by MREs A–E, with the noteworthy exception of a cluster of protected residues centered around −87 (Figs. 2 and 3). This region contains a sequence that is moderately related to the MRE consensus: It matches at 9 of 15 positions overall, but possesses greater similarity in the MRE 'core', in which 4 of 5 bases agree. The DMS pattern is also quite similar to that at the other MREs (Fig. 7). Some deletion analyses of this gene are consistent with the sequence functioning as a hitherto unidentified MRE (Searle et al. 1985). Although these mutation data are not decisive, the sequence similarity, footprint pattern, and deletion data lead us to identify it provisionally as MRE-F. One potentially interesting distinction between MRE-F and the other MREs is that it appears to be less efficiently protected from DMS when cells are treated with cadmium than when they are induced with zinc (data not shown), whereas the others are affected identically by both inducers.

Footprinting amplified and transfected genes

In these experiments we have shown that in vivo footprinting studies can be executed successfully on transfected, amplified genes, and this approach should be generally applicable to promoters other than MT. It requires ≥15 functional copies of the promoter per mammalian genome (S. Salser, unpubl.), although single-copy genes should be readily accessible in smaller genomes. In this study, cell lines containing sets of active genes were de-

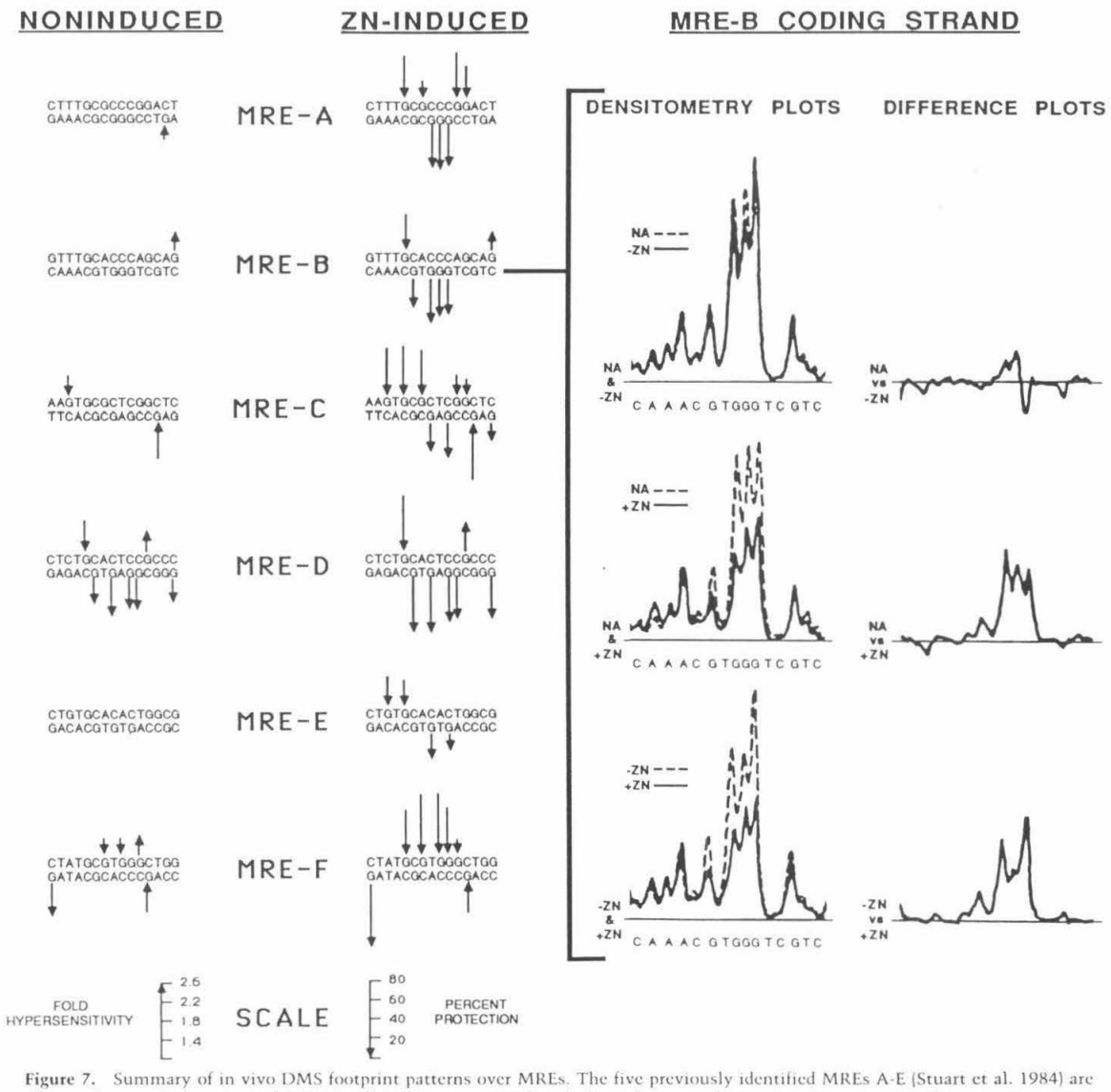


Figure 7. Summary of in vivo DMS footprint patterns over MREs. The five previously identified MREs A-E (Stuart et al. 1984) are shown along with the potential MRE-F. All MREs are oriented with the coding strand on top as they appear in the promoter except MRE-B, which is inverted here to allow comparisons of the footprint patterns. Quantitation of DMS footprints under noninduced and Zn-induced conditions was performed as described in Materials and methods and is indicated by downward pointing arrows (↓) for protection, and by upward pointing arrows (↑) for hypersensitivities. Scale is at bottom of figure. An example of MRE densitometry is shown for the coding strand of MRE-B on the right half of the figure. The initial, normalized plots are superimposed pairwise as indicated and the corresponding difference plots are shown on the right. Beneath the densitometry plots is the coding strand sequence of MRE-B. (NA) In vitro purified DNA (naked); (− ZN) in vivo DNA, noninduced; (+ ZN) in vivo DNA, Zn-induced.

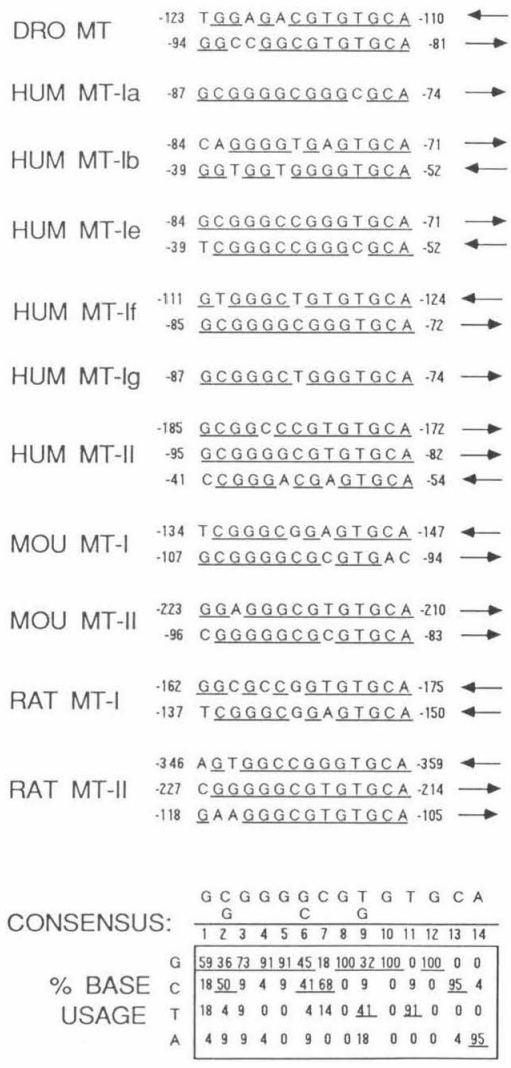


Figure 8. Evolutionary conservation of G elements in higher eukaryotic metallothioneins. Shown are the 22 G elements found in 11 MT promoters, and the consensus we have derived. General conservation of G rich regions has been noted previously (Karin 1984) and designated variously as a basal level enhancer (BLE) (Scholer et al. 1986), a part of the BLE (Karin et al. 1987), or a G-rich sequence (Seguin et al. 1986). Numbers on either side of the G elements indicate their location in the promoters relative to the transcriptional start site. Arrows to the left of the sequences indicate the orientation of the elements in their promoter. All of the elements have been displayed in the orientation of the consensus for comparative purposes. The consensus along with the percent base usage is indicated at the bottom of the figure. Sources: *Drosophila* MT (Maroni et al. 1986); human MT-Ia (Richards et al. 1984); human MT-Ib (Heguy et al. 1986); human MT-Ie (Schmidt et al. 1985); human MT-If and human MT-Ig (Varshney et al. 1986); human MT-II (Karin et al. 1987); mouse MT-I (S.J.S. unpubl., see Fig. 3 legend and Glanville et al. 1981); mouse MT-II (Searle et al. 1984); rat MT-I (Andersen et al. 1986); rat MT-II (R.D. Andersen, unpubl. 1986, NIH data base).

rived by amplification of a chimeric MT–DHFR construct, but multiple-copy transformants could be equally useful. As in other transfection studies, it is necessary to select for footprinting transformants that show proper expression. Our cell lines were checked to ensure metal induction (Fig. 1b), and basal-level expression characteristic of the native MT gene was maintained by continuous selection for DHFR expression (see Materials and methods). From the 70% and 65% protection at SP1 and MRE sites, respectively, it is clear that a large fraction of the transfected MT promoters in these cells is accessible to binding by cellular factors. It is worth noting that transfected target genes sometimes present specific limitations. For example, a peculiarity of mouse MT-I is that the endogenous gene is inducible by glucocorticoids, but no cloned DNA segment that retains this property upon transfection into cultured cells or into transgenic mice has been identified (Mayo et al. 1982; Pavlakis and Hamer 1983; Palmiter et al. 1982). In agreement with these observations, we found that the MT–DHFR genes are not induced by dexamethasone, and therefore we did not attempt to footprint following hormone treatment.

Although in vivo footprinting is an effective method for visualizing a subset of protein : DNA interactions on a given gene, it is not expected that all protein : DNA contacts will be detected by this assay. For example, we do not observe significant footprinting over or near the TATA box (Figs. 2 and 3), even though these genes are being transcribed at a physiologically significant rate. This is also the case in several other in vivo footprinting studies of actively transcribing genes (Zinn and Maniatis 1986; Kemper et al. 1987). Nevertheless, it is clear that a set of proteins in nuclear extracts do bind in vitro in other TATA containing genes (Davison et al. 1983; Parker and Topol 1984; Sawadogo and Roeder 1985), and previous genetic experiments have shown these sequences to be important for efficient MT-I transcription (Carter et al. 1984). Perhaps the interactions that are observed in vivo at upstream elements are significantly more stable than are those that occur at the TATA sequence. In spite of these constraints, it is clear that we have successfully identified interactions at MRF, SP1, and MLTF binding sites in the murine MT-I promoter, and genetic analyses (Carter et al. 1984; Stuart et al. 1984; Searle et al. 1985) support the functional significance of these interactions. Based on these results, we suggest that it should now be efficient to survey other complex control regions in this way at an early stage in their characterization to focus subsequent mutagenesis or in vitro protein binding studies.

Discussion

We have used in vivo genomic footprinting with DMS and DNase I to observe DNA : protein interactions at the MT-I promoter (Figs. 3 and 7). These experiments led to definition of two sets of probable protein : DNA interactions as they occur in cells under noninducing and inducing conditions. The results obtained allowed us to identify a new metal-responsive element, to limit the possible mechanisms for interaction between basal-level elements and metal-responsive elements, and to begin to resolve how overlapping, and thus potentially competing, recognition sites are actually used.

It is interesting to compare the pattern of in vivo protection at MREs with in vitro studies of Seguin and Hamer (1987). They observed a cadmium-dependent Exo III footprint at MRE-D when MT-I DNA was incubated with a nuclear extract. Our results differ from their in vitro data in two respects. First, they detect binding only in the presence of cadmium but not in the presence of zinc, whereas we find similar in vivo footprints with both metals at MREs A-E. Second, they observe binding at just one MRE, whereas we see footprints at MREs A-E and F. The general pattern of in vivo interactions we have observed is, however, in close agreement with in vivo studies reported recently by Herschman and colleagues (Anderson et al. 1987) for the related rat MT-I gene. We do not know the basis for the differences between in vivo and in vitro results, although it is interesting that our data find MRE-D to be the only metal response element showing a detectable footprint in the absence of added metal (Fig. 7). It may also be relevant that MRE-D (−150 to −136) overlaps the potential binding site SP1-B (−143 to −134), as discussed below.

Overlapping recognition sites and possible competition among factors

The presence of overlapping potential binding sites at the MRE-D/SP1-B region raises the possibility of competition among several different factors. DNA sequence data alone do not provide enough information to predict which factors, if any, will be bound at this site in the cell. The in vitro SP1 footprint data (Fig. 5) establish that this sequence can serve as a SP1 binding site in the absence of competing factors, which leaves open the possibility that it may function as a SP1 site in vivo. The in vivo DNase (Fig. 4) and DMS (Fig. 2) footprints show that this region is occupied by factors in L cells, but the absence of the characteristic SP1 DMS footprint pattern makes it unlikely that SP1 is responsible. Another possibility is that, in both the induced and noninduced state, there is significant binding of metal-responsive factors at this site. This alternative is supported by the pattern of DMS footprinting. The DMS pattern at MRE-D in noninduced cells is qualitatively similar to that at other functional MREs after metal induction. The induced pattern at MRE-D is quantitatively increased upon metal induction but not qualitatively changed from its basal pattern. It is interesting that MRE-D is the most effective MRE when present in multiple copies in a synthetic promoter (Stuart et al. 1985) and is also the only MRE to show metal-dependent protein binding in vitro (Seguin and Hamer 1987).

For the case of the MRE-D/SP1-B region, our working model is that in L cells the balance between active SP1 and MRFs favors binding by MRF, but this may not be the case in other cell types or under different physiological conditions. The notion of in vivo competition for overlapping sites has some appeal because it provides a potentially sensitive mechanism for regulating gene expression in different cell types or tissues in response to modest differences in relative levels of participating factors. Small genetic changes can also tip the balance between two binding sites, and such changes may be important in the evolution of different patterns of expression. An inadvertent experimental example was provided when Stuart et al. (1985) placed oligomers containing the MRE-D overlap region in front of a truncated HSV-TK promoter containing only a TATA element. The wild-type MRE-D region made TK expression metal inducible. Surprisingly, they found that if the oligomer was modified so as to make it a better match to the SP1 consensus without disrupting the MRE, the gene was no longer inducible by metals, but exhibited a 10-fold elevation in basal expression. Competition among factors for binding in this region may also contribute to the multiplicity of protein : DNA complexes observed as distinct electrophoretic species when DNA is incubated with crude nuclear extracts (Seguin and Hamer 1987).

The phenomenon of overlapping, potentially competing recognition sites is probably not confined to the MRE-D/SP1-B region. All higher eukaryotic MT genes for which sequence data are available contain one or more copies of a conserved 14-bp G-rich sequence, and from 22 examples we have derived a consensus element G[C/G]GGG[C/G]CG[T/G]GTGCA (Fig. 8). A function has not yet been defined for this sequence feature, so it is simply referred to here as the G element or G box. Murine MT-I contains two of these elements: one centered around −140, the other centered around −100. Like other MT G-boxes, these overlap other potential recognition sites. The site centered at −140 overlaps the MRE-D and SP1-B sites, and the element centered at −100 overlaps the MLTF site. In other MT genes (for example, human MT-IIA, human MT-IA, mouse MT-II, and *Drosophila* MT) G elements overlap potential SP1 and MRE sites. Our footprinting data did not detect interactions specifically attributable to G elements, and existing mutational studies assayed in various cell lines have also failed to detect a specific G-box function. On the other hand, the presence of G-elements in all MT genes, together with their substantial sequence similarity, suggest that they are important enough to demand evolutionary conservation, probably because they serve as recognition sites for one or more factors. It seems possible that the G elements may be required in a physiological context that has not been tested by the transfection assays and footprinting studies.

Interaction of MREs and basal elements

Although multiple MREs are sufficient to direct metal-inducible transcription from heterologous promoters, the absolute level of expression depends on additional non-MRE upstream elements. For example, Stuart et al. (1985) found that in the context of a synthetic promoter, SP1 elements amplify the MRE-mediated response, and promoter mutations generally support this view. The effect is not simply additive, but is synergistic. A survey of different MT regulatory regions shows that the theme of multiple MREs combined with several potential basal elements is common, but there is considerable flexibility in basal element identity and sequence organization. This suggests that the capacity of basal element factors to enhance MRE activity is a general property shared by several factors including SP1 (mouse MT-I, human MT-IIA), AP1 (human MT-IIA), and MLTF (mouse MT-I).

How do basal elements amplify metal induction? The constitutive occupancy of SP1 and MLTF sites in vivo rules out one straightforward mechanism for amplifying metal- dependent expression: Active MREs do not function by simply increasing the efficiency of factor binding at basal element sites. The possibilities that remain fall into two classes. MREs and basal elements both may act by increasing directly the rate of initiation. Alternatively, basal elements might alter the rate of metal-induced transcription by increasing the accessibility of DNA to factors that do govern initiation—possibly MREs. These different functions for basal elements are not easily distinguished by standard cell transfection assays or by current in vitro transcription systems, although it should now be possible to introduce appropriately mutated promoters and observe how ablation of one or more of the basal elements affects the in vivo protein : DNA interactions at MREs.

Materials and methods

Plasmid construction

Plasmid MD was constructed from the mouse MT-I genomic sequences of pEE-3.8 (a gift from R. Palmiter) and the DHFR sequences of SV3–DHFR (Southern and Berg 1982). The MT-I promoter fragment used in this plasmid contains ~1775 bp of 5′ nontranslated sequences between *Eco*RI and *Bgl*II. The *Bgl*II site ends 8 bp on the 5′ side of the MT translation start. The 3′ MT fragment contains the complete second intron and poly(A) addition site of MT-I as an *Alu*I–*Hin*dIII fragment, with the *Alu*I site converted to a *Bgl*II site. These fragments were inserted into pBR322 to create pMT-33. The coding region of DHFR, along with 60 bp on the 5′ side of and 80 bp on the 3′ side of the translation start and stop respectively, was removed from SV3–DHFR as a *Fnu*DII–*Bgl*II fragment and inserted into the *Bgl*II site of pMT-33 after addition of a *Bam*HI linker at the *Fnu*DII site.

Cell culture

Murine cell lines A-0.3 (formerly KTO-A, Kim and Wold 1985) and A-60 were gifts from S. Kim, and were adapted for growth in dialyzed calf serum. These lines were maintained in Dulbecco's modified Eagle's medium (DME) supplemented with 10% dialyzed calf serum (Gibco), 100 U/ml penicillin G, 7.5 U/ml streptomycin, 4 μg/ml azaserine, 15 μg/ml adenine, and 3× nonessential amino acids. Dialysis of the serum increased the induction ratio of MT genes ~twofold by lowering the basal level of expression (P.R. Mueller, unpubl.). A-0.3 and A-60 were further supplemented with 0.3 μM and 60 μM methotrexate (MTX), respectively. The presence of MTX did not influence the expression or induction of the MT genes (data not shown). A-0.3 and A-60 contain 100 and 700 copies of pMD, respectively (J. Jong, unpubl.).

RNA preparation and measurement

Cells were ~70% confluent before induction. Four or 8 hr prior to harvesting the RNA, fresh media were added to the plates with or without 130 μM zinc sulfate. Prior to TPA (12-O-tetradecanoyl-phorbol-13-acetate) inductions, cells were washed with PBS and transferred to media containing 0.5% dialyzed calf serum, without azaserine or methotrexate (MTX) for 24 hr. TPA (in ETOH) or an equivalent volume of ETOH (14.4 μl) was then added to the serum-starved media for an additional 12 hr. Final concentration of TPA was 60 ng/ml. Total RNA for each condition was prepared from two 150-cm plates according to the method of Chirgwin et al. (1979), as modified by Ngai et al. (1984). RNA was further purified by extraction with phenol/Sevag and precipitation with isopropanol and then ethanol. Subsequent treatment with RNase-free DNase had no effect on RNA quantitation (data not shown), so this was not done routinely.

RNase protection was performed as described by Zinn et al. (1983) and modified by Kim and Wold (1985). Fifteen micrograms of total RNA was hybridized with a molar excess (12 ng, or 280 fmoles) of labeled probe of specific activity 7×10^8 cpm/μg. The probe used protects the DHFR sequences in the construct and the endogenous DHFR, but the endogenous signal is much weaker than the strong construct signal and cannot be seen on short exposures. Longer exposures of RNA experiments performed on the parental cell lines show that the protected endogenous band is ~15 bases shorter than the protected construct band (P. Mueller, unpubl.). This is consistent the with major transcriptional start site of mouse DHFR (Farnham and Schimke 1986). Gel slices containing the protected probe were excised and counted for quantitation.

In vivo and in vitro DMS/piperidine cleavage of DNA

Cells for in vivo footprinting were treated identically to those used for RNA measurements. After 4 hr with or without 130 μM zinc sulfate, the media was replaced with media containing 1 μl/ml DMS (DMS was added immediately before use). This replacement media also contained 130 mM zinc sulfate, if appropriate, and was prewarmed to 37°C in a 4% CO_2 environment. The DMS was allowed to react with the cells for 2 min, at which time the DMS-containing media was removed and the plates immediately were rinsed once with 37°C PBSA and subsequently washed 3× with 37°C PBSA for 30 sec with gentle shaking. Cells were lysed and scraped from each plate in 1.5 ml of DNA harvest buffer [1 mM Tris (pH 7.5), 400 mM NaCl, 2 mM EDTA, 0.2% SDS, 0.2 mg/ml proteinase K], and this mixture was incubated with occasional mixing for an additional 3–5 hr

at 37°C. DNA was extracted once each with phenol, phenol/Sevag, and Sevag; and twice with ether. The DNA was then precipitated once with isopropanol and once with ethanol before resuspending it in 10 mM Tris (pH 7.5), 1 mM EDTA. The final nucleic acid concentration was 1–1.5 mg/ml, of which 50–60% was DNA. To precipitate the in vivo methylated DNA in parallel with the in vitro methylated DNA (see below), 200 μl of this mixture was mixed with 50 μl of 1.5 M sodium acetate (pH 7.0), 1 M β-mercaptoethanol, 100 μg/ml yeast tRNA, and 750 μl of ethanol. The pelleted DNA was washed once with 75% ethanol before being suspended in 200 μl of 1 M piperidine, and then heated to 90°C for 30 min. Piperidine treatment cleaves DNA at methylated guanines and eliminates contaminating RNA by base hydrolysis. Piperidine was removed by vacuum followed by two cycles of washing and precipitation.

DNA for in vitro DMS treatment (naked) was prepared identically, except the in vivo DMS treatment step was omitted. This DNA was methylated in vitro by adding 0.9 μl of DMS to 200 μl of DNA (~1 mg/ml) for 30 sec at room temperature. This DMS condition was empirically determined to match the in vivo DMS conditions. Under methylation would show a general underrepresentation of lower-molecular-weight bands and overrepresentation of high-molecular-weight bands. The converse is true for overmethylation. Such mismatches in methylation conditions do not, however produce specific local patterns of protection or hypersensitivity that would be easily mistaken for protein : DNA footprints. The reaction was stopped with 50 μl ice-cold stop buffer [1.5 M sodium acetate (pH 7.0), 1 M β-mercaptoethanol, 100 μg/ml of yeast tRNA] and 750 μl of ethanol chilled to −70°C. The samples were then piperidine treated in parallel with the in vivo samples. For in vitro controls that were not reacted with DMS or piperidine, the DNA was first cleaved with *Eco*RI and RNase A to reduce viscosity and remove contaminating RNA, respectively, and then processed in parallel with the other samples.

In vivo and in vitro genomic DNase treatment

Cells for in vivo DNase I footprinting were treated identically to those used for RNA measurements. Preparation of nuclei and in vivo DNase treatment was performed as described by Zinn and Maniatis (1985) with slight modifications. Naked DNA (in vitro) was prepared by the conditions described above, and digested with DNase I (Boehringer Mannheim Biochemicals) to the same extent as the nuclear samples (empirically determined). Underdigestion would show a general underrepresentation of lower-molecular-weight bands and overrepresentation of high-molecular-weight bands. The converse is true for overdigestion. Such mismatches in DNase I conditions do not, however, produce specific local patterns of protection that would be easily mistaken for protein : DNA footprints. Both the in vitro and the in vivo DNAs were also digested with *Eco*RI and RNase A to reduce viscosity and remove contaminating RNA.

Primer extension of DMS and DNase-cleaved genomic DNA

Visualizing the footprint patterns by primer extension eliminates the technically challenging blotting of sequencing gels and subsequent hybridization (Church and Gilbert 1984). In addition, extension of an end-labeled primer ensures that single base resolution is not compromised by either radioactive decay of a multiply end-labeled probe or imprecise action of a single-stranded nuclease (Jackson and Felsenfeld 1985). Because loading more than ~75 μg of DNA per lane (lane size 8 mm × 0.8 mm) leads to decreased resolution, our procedure required ~15 copies per mammalian genome (P. Mueller, unpubl.). Specifically, 65 μg A-0.3 DNA or 14 μg of A-60 DNA were mixed with 0.13 pmole labeled primer (sp. act. 6–14 × 10^6/pmole) and suspended in 55% deionized formamide, 6× SSPE [900 mM NaCl, 60 mM NaH_2PO_4 : H_2O, 6 mM EDTA (pH 7.4)]; final volume was 70 μl. The primers used were complementary to −227 to −203 on the noncoding strand and to +47 to +23 on the coding strand. These 25-mers have a GC content of 60% and 52%, respectively. Samples were heated to 95°C for 5 min and incubated for 30–60 min at 30°C. These conditions proved to maximize the amount of primer hybridized while minimizing the self-hybridization of the genomic DNA and nonspecific background hybridization of the primers (P. Mueller, unpubl.). Different length or GC content of the primers might require different hybridization conditions. To remove excess primer and change the reaction buffer, one of two methods was used. For the DNase and early DMS experiments, the hybridization mixture was run over a quick-spin column of Sephadex G-50 equilibrated in low KT buffer (12.5 mM KCl, 12.5 mM Tris, pH 8.3 at 50°C), and the volume of the efflux was adjusted to 87.5 μl with low KT buffer. In other experiments, the hybridization mixture was diluted to a volume of 670 μl with 2.25 M ammonium acetate and then precipitated with isopropanol. The pelleted DNA was then dissolved in 87.5 μl of low KT buffer. In either case the DNA was then mixed with 12.5 μl 8× AMV reverse transcription buffer so that the final concentrations were 50 mM KCl, 50 mM Tris, pH 8.3 at 50°C, 8 mM $MgCl_2$, 0.8 mM dNTP, and 4 mM DTT. Two microliters (28 units) AMV reverse transcriptase (Life Sciences, Inc.) were added and primers were extended for 30 min at 50°C. The reaction was terminated by adding 3 μl 500 mM EDTA, 35 μl 8 M ammonium acetate, and 10–20 μg yeast tRNA. Samples were precipitated with 1 volume isopropanol, suspended in 12 μl loading buffer [80% deionized formamide, 40 mM Tris-borate (pH 7.5)] and run on 8% denaturing polyacrylamide gels 0.8-mm thick. Gels were dried and exposed for 1–10 days without an intensifying screen.

In vitro footprinting of SP1

In vitro DNase I footprints were performed, as described by Briggs et al. (1986), with 95% homogeneous, human SP1 that was ~10 ng/ml (a gift from J. Kadonaga and R. Tjian). The mouse 400-bp MT-I promoter fragment was labeled at the *Bam*HI site of plasmid MT-IΔi (a gift from R. Palmiter), and the SV40 326 bp promoter fragment was labeled at the *Hin*dIII site of plasmid SV2-DHFR (Southern and Berg 1982); 6.5 fmoles of each of these end-labeled DNA fragments were used and 8 ng/ml was the final DNase I concentration (Boehringer Mannheim Biochemicals). Gels were run as above.

Quantitation of DMS footprints

Several exposures of footprinting gels were prepared using Kodak XAR-5 film without an intensifying screen. Densitometry was performed with a LKB UltroScan XL laser densitometer, which has a beam size of 800 μm by 50 μm. Each lane was scanned nine times with a 200-μm overlap between scans and a 120-μm vertical step. These scans were analyzed using the Turbo-Densitometry program (S.J.S.). Averaging of the nine scans produced a single composite scan per lane. To compare the various lanes of each loading (i.e., naked, in vivo noninduced, and in vivo induced), the composite scans were aligned, and the signal normalized. This normalization was to sequences in the guanine ladder that lacked both binding sites and observable footprints; this compensated for minor variations in sample loading, and had only a slight effect (≤10%) on the final quantitation of protection and hypersensitivity. To

quantitate individual protections and hypersensitivities, the integrated area for each base in the original plots was subtracted from the corresponding area in the naked DNA plot, and the difference was divided by the area of the naked DNA base. Difference plots are useful for illustrating footprints and comparing occupancy under noninduced and induced conditions. These were obtained by subtracting the digitized and normalized scan of one lane from that of another lane on a point-by-point basis. These plots indicate regions of protection above the baseline and regions of hypersensitivity below the baseline.

Acknowledgments

The authors with to thank Drs. R. Palmiter and R. Schimke for supplying DNAs used in plasmid constructions, R. Tjian and J. Kadonaga for purified SP1 factor, and J. Huibregtse and D. Engelke for communicating results before publication. We thank Prof. Peter Dervan for use of the scanning densitometer; Jan Hoh, J. Daniels, and Dr. John Ngai for helpful discussion; and Profs. David Anderson, Norman Davidson, Mel Simon, and members of the Wold group for interesting discussions and critical reading of the manuscript. This work was supported by grants from the Rita Allen Foundaton and the National Institutes of Health to B.W., a National Research Services Award (1 T32 GM 07616) from the National Institute of General Medical Sciences to P.R.M., and a California Institute of Technology SURF Fellowship to S.J.S.

References

Anderson, R.D., B.W. Birren, S.J. Taplitz, and H.R. Herschman. 1986. Rat metallothionein-I structural gene and three pseudogenes, one of which containing 5′-regulatory sequences. *Mol. Cell. Biol.* **6:** 302–314.

Andersen, R.D., S.J. Taplitz, S. Wong, G. Bristol, B. Larkin, and H.R. Herschman. 1987. Metal-dependent binding of a factor in vivo to the metal-responsive elements of the metallothionein I gene promoter. *Mol. Cell. Biol.* **7:** 3574–3581.

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.

Angel, P., A. Poting, U. Mallick, H.J. Rahmsdorf, M. Schorrp, and P. Herrlich. 1986. Induction of metallothionein and other mRNA species by carcinogens and tumor promoters in primary human skin fibroblast. *Mol. Cell. Biol.* **6:** 1760–1766.

Briggs, M.R., J.T. Kadonaga, S.P. Bell, and R. Tjian. 1986. Purification and biochemical characterization of the promoter-specific transcription factor, SP1. *Science* **234:** 47–52.

Brinster, R.L., H.Y. Chen, R. Warren, A. Sarthy, and R.D. Palmiter. 1982. Regulation of metallothionein-thymidine kinase fusion plasmids injected into mouse eggs. *Nature* **296:** 39–42.

Buchanan, R.L. and J.D. Gralla. 1987. Factor interactions at simian virus 40 GC-box promoter elements in intact nuclei. *Mol. Cell. Biol.* **7:** 1554–1558.

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 gene. *Proc. Natl. Acad. Sci.* **81:** 7392–7396.

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.

———. 1987. The major late transcription factor binds to and activates the mouse metallothionein I promoter. *Genes Dev.* **1:** 973–980.

Chirgwin, J.M., A.E. Przybyla, R.J. MacDonald, and W.J. Rutter. 1979. Isolation of biologically active ribonuclei acid from sources enriched in ribonuclease. *Biochemistry* **18:** 5294–5299.

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.

Church, G.M. and W. Gilbert. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci.* **81:** 1991–1995.

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

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.

Dynan, W.S. and R. Tjian. 1985. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature* **316:** 774–778.

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.

Farnham, P.J. and R.T. Schimke. 1986. In vitro transcription and delimitation of promoter elements of the murine dihydrofolate gene. *Mol. Cell. Biol.* **6:** 2392–2401.

Gidoni, D., J.T. Kadonaga, H. Barrera-Saldana, K. Takahashi, P. Chambon, and R. Tjian. 1985. Bidirectional SV40 transcription mediated by tandem SP1 binding interactions. *Science* **230:** 511–517.

Gidoni, D. W.S. Dynan, and R. Tjian. 1984. Multiple specific contacts between a mammalian transcription factor and its cognate promoters. *Nature* **312:** 409–413.

Gilbert, W., A. Maxam, and A. Mirzabekov. 1976. Contacts between the LAC repressor and DNA revealed by methylation. In *Control of ribosome synthesis*, Alfred Benzon Symposium IX (ed. N.O. Kjeldgaard and O. Maaloe), pp. 139–148. Academic Press, New York.

Gimble, J.M. and E.E. Max. 1987. Human immunoglobulin κgene enhancer: Chromatin structure analysis at high resolution. *Mol. Cell. Biol.* **7:** 15–25.

Giniger, E., S.M. Varnum, and M. Ptashne. 1985. Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell* **40:** 767–774.

Glanville, N., D.M. Durnam, and R.D. Palmiter. 1981. Structure of mouse metallothionein-I gene and its mRNA. *Nature* **292:** 267–269.

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

Hamer, D.H. 1986. Metallothionein. *Annu. Rev. Biochem.* **55:** 913–951.

Hamer, D.H. and M. Walling. 1982. Regulation in vivo of a cloned mammalian gene: Cadmium induced the transcription of a mouse metallothionein gene in SV40 vectors. *J. Mol. Appl. Genet.* **1:** 273–288.

Heguy, A., A West, R.I. Richards, and M. Karin. 1986. Structure and tissue-specific expression of the human metallothionein I_B gene. *Mol. Cell. Biol.* **6:** 2149–2157.

Huibregtse, J.M. and D.R. Engelke. 1986. Direct identification of small sequence changes in chromosomal DNA. *Gene* **44:** 151–158.

Imbra, R.J. and M. Karin. 1987. Metallothionein gene expres-

sion is regulated by serum factors and activators of protein kinase C. *Mol. Cell. Biol.* **7:** 1358–1363.

Jackson, P.D. and G. Felsenfeld. 1985. A method for mapping intranuclear protein-DNA interactions and its application to a nuclease hypersensitive site. *Proc. Natl. Acad. Sci.* **82:** 2296–2300.

———. 1987. In vivo footprinting of specific protein-DNA interactions. In *Methods enzymol.* (ed. S. Berger and A.R. Kimmel), vol. 152, pp. 735–755. Academic Press, New York.

Johnsrud, L. 1978. Contacts between *Escherichia coli* RNA polymerase and a lac operon promoter. *Proc. Natl. Acad. Sci.* **75:** 5314–5318.

Jones, K.A. and R. Tjian. 1985. Sp1 binds to promoter sequences and activates herpes simplex virus 'immediate-early' gene transcription in vitro. *Nature* **317:** 179–182.

Jones, K.A., J.T. Kadonaga, P.A. Luciw, R. Tjian. 1986. Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. *Science* **232:** 755–759.

Kadonaga, J.T., K.A. Jones, and R. Tjian. 1986. Promoter-specific activation of RNA polymerase II transcription by Sp1. *Trends in Biochem.* **11:** 20–23.

Karin, M. 1985. Metallothioneins: Proteins in search of function. *Cell* **41:** 9–10.

Karin, M., A. Haslinger, A. Heguy, T. Dietlin, and T. Cooke. 1987. Metal-responsive elements act as positive modulators of human metallothionein-II_A enhancer activity. *Mol. Cell. Biol.* **7:** 606–613.

Karin, M., A. Haslinger, H. Holtgreve, R.I. Richards, P. Krauter, H.M. Westphal, and M. Beato. 1984. Characterization of DNA sequences through which cadmium and glucocorticoid hormones induce human metallothionein-II_A gene. *Nature* **308:** 513–519.

Karin, M. and H. Holtgreve. 1984. Nucleotide sequence requirements for transient expression of human metallothionein-IIA–thymidine kinase fusion genes. *DNA* **3:** 319–326.

Karin, M. and R.I. Richards. 1982. Human metallothionein genes–primary structure of the metallothionein-II gene and a related processed gene. *Nature* **299:** 797–802.

Kemper, B., P.D. Jackson, and G. Felsenfeld. 1987. Protein-binding sites within the 5′ DNase I-hypersensitive region of the chicken α^D-globin gene. *Mol. Cell. Biol.* **7:** 2059–2069.

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.

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

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

Maniatis, T., S. Goodbourn, and J.A. Fischer. 1987. Regulation of inducible and tissue-specific gene expression. *Science* **236:** 1237–1245.

Maroni, G., E. Otto, and D. Lastowski-Perry. 1986. Molecular and cytogenetic characterization of a metallothionein gene of *Drosophila*. *Genetics* **112:** 493–504.

Maxam, A.M. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. In *Methods enzymol.* (ed. L. Grossman and K. Moldave), vol. 65, pp. 499–560. Academic Press, New York.

Mayo, K.E., R. Warren, and R.D. Palmiter. 1982. The mouse metallothionein-I gene is transcriptionally regulated by cadmium following transfection into human or mouse cells. *Cell* **29:** 99–108.

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

Miyamoto, N.G., V. Moncollin, J.M. Egly, and P. Chambon. 1985. Specific interacton between a transcription factor and the upstream element of the adenovirus-2 major late promoter. *EMBO J.* **4:** 3563–3570.

Ngai, J., Y.G. Capetanaki, and E. Lazarides. 1984. Differentiation of murine erythroleukemia cells results in the rapid repression of vimentin gene expression. *J. Cell. Biol.* **99:** 306–314.

Palmiter, R.D., H.Y. Chen, and R.L. Brinster. 1982. Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring. *Cell* **29:** 701–710.

Parker, C.S. and J. Topol. 1984. A *Drosophila* RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. *Cell* **37:** 273–283.

Pavlakis, G.N. and D.H. Hamer. 1983. Regulation of a metallothionein-growth hormone hybrid gene in bovine papilloma virus. *Proc. Natl. Acad. Sci.* **80:** 397–401.

Richards, R.I., A. Heguy, and M. Karin. 1984. Structural and functional analysis of the human metallothionein-I_A gene: Differential induction by metal ions and glucocorticoids. *Cell* **37:** 263–272.

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.

Schmidt, C.J., M.F. Jubier, and D.H. Hamer. 1985. Structure and expression of two human metallothionein-I isoform genes and a related pseudogene. *J. Biol. Chem.* **260:** 7731–7737.

Scholer, H., A. Haslinger, A. Heguy, H. Holtgreve, and M. Karin. 1986. In vivo competition between a metallothionein regulatory element and the SV40 enhancer. *Science* **232:** 76–80.

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.

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.

Seguin, C. and D.H. Hamer. 1987. Regulation in vitro of metallothionein gene binding factors. *Science* **235:** 1383–1387.

Seguin, C., B.K. Felber, A.D. Carter, and D.H. Hamer. 1984. Competition for cellular factors that activate metallothionein gene transcription. *Nature* **312:** 781–785.

Southern, P.J. and P. Berg. 1982. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J. Mol. Appl. Genet.* **1:** 327–341.

Stuart, G.W., P.F. Searle, and R.D. Palmiter. 1985. Identification of multiple metal regulatory elements in mouse metallothionein-I promoter by assaying synthetic sequences. *Nature* **317:** 828–831.

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.* **81:** 7318–7322.

Varshney, U., N. Jahrousi, R. Foster, and L. Gedamu. 1986. Structure, organization, and regulation of human metallothionein IF gene: Differential and cell-type-specific expression

in response to heavy metals and glucocorticoids. *Mol. Cell. Biol.* **6:** 26–37.

Zinn, K. and T. Maniatis. 1986. Detection of factors that interact with the human β-interferon regulatory region in vivo by DNase I footprinting. *Cell* **45:** 611–618.

Zinn, K., D. DiMaio, and T. Maniatis. 1983. Identification of two distinct regulatory regions adjacent to the human β-interferon gene. *Cell* **34:** 865–879.

Chapter 4

In Vivo Competition of Metal Specific Induction Factors Controlling Mouse Metallothionein Transcription.

In Vivo Competition of Metal Specific Induction Factors Controlling Mouse Metallothionein Transcription.

Paul R. Mueller and Barbara Wold

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

Abstract

Stable cell lines containing up to 1,400 copies of mouse metallothionein-I (MT-I) controlling regions were generated by gene amplification of a transfected MT-DHFR fusion construct. These additional copies acted as competitors for transcription factors required for metal induced expression of the endogenous MT-I and MT-II genes. Metal induced expression was reduced by ~70%, while basal expression and glucocorticoid induction remained normal. Among several genes assayed, the titration effect was specific for MT genes. *In vivo* footprints of the endogenous MT-II gene indicate that DNA binding factors that interact with the MT promoter are not titrated by this treatment. This suggests that additional factors besides those that bind directly to the well-characterized metal responsive elements (MREs) are required for metal induction of MT.

Introduction

The expression of eukaryotic genes can be controlled by transcriptional, post-transcriptional, or translational mechanisms. In many cases all of these mechanisms function in concert to control the expression of a gene product. A good example of this is found in the metallothionein (MT) gene family. Metallothionein genes code for highly conserved, cysteine-rich heavy metal binding proteins that are found in all higher eukaryotes (reviewed in Karin 1985; Hamer 1986). They appear to function in zinc and copper homeostasis and to provide protection from heavy metal toxicity. Metallothioneins have an appreciable basal level of expression that can be induced 2- to 40-fold by several environmental signals including heavy metals, glucocorticoids, interferon, and stress (Durnam and Palmiter 1981; Hamer and Walling 1982; Hager and Palmiter 1981; Mayo and Palmiter 1981). In addition, they appear to be developmentally regulated. Transcriptional control of MT has been shown to be dependent on basal, hormonal, and metal responsive trans-acting factors that bind to cis-acting elements in the MT promoter. Basal factors bind to the promoter constitutively and are thought to be responsible for basal level expression (Anderson et al. 1987; Mueller et al. 1988). The additional binding of environmentally responsive factors increases transcription by up to 20-fold. However, this does not account entirely for the observed increase in MT expression. At least some of the induction of MT mRNA takes place at the post-transcriptional level. Even after the metallothionein protein is made, it can be stabilized up to 4-fold by the presence of metal (Karin et al. 1981a; Durnam and Palmiter 1981). In fact, the only level at which control has not been observed is translational efficiency (Durnam and Palmiter 1981).

In this work, we have studied the MT-I and MT-II genes of mouse. These are chromosomally linked, have similar sequence motifs in their promoters and untranslated RNA sequences, and differ by only 4 amino acids in their coding regions. Their expression is coordinate (Searle et al. 1984; Yagle and Palmiter 1985). The steady-state

level of MT-I mRNA is 1.5- to 2-fold higher than that of MT-II under both basal and induced conditions. To further understand the regulation of MT, we have titrated trans-acting regulatory factors by stably integrating multiple copies of the mouse MT-I promoter and untranslated sequence elements into mouse cells. In cell lines that contain >500 copies of cis-acting MT-I sequences, the endogenous MT-I and -II genes are induced by metals to only 30% of their wild-type levels. Basal level expression and induction by glucocorticoids are unaffected. This is observed even in lines that have up to 1,400 copies of MT-I cis-acting elements. The kinetics of mRNA induction suggests that this titration is affecting transcription, but *in vivo* footprinting analyses show that the binding of metal responsive trans-acting factors to the promoter is unaltered.

Results

In vivo titration between excess copies of control-sequence DNA and limited amounts of trans-acting regulatory factors can be used to gain insight into the mechanisms of gene regulation in the absence of detailed biochemical knowledge of the regulatory factors (Scholer and Gruss 1984). In mammalian cell culture systems, such studies are usually performed by transient cotransfection of constant amounts of an indicator gene with a variable amount of a competitor gene. Because only 1 to 10% of the cells actually take up DNA in a transfection assay, only the cotransfected gene is detectably responsive to the titration effect. In the 90 to 99% of the cells that are not transfected, the endogenous gene cannot be affected. In addition, because these transfected cells are not stable, it is difficult to study the molecular interactions that are being affected by the titration. We have stably transfected mouse cells with a unique competitor gene that can be amplified to generate a high copy number of competitor sequences. This allows titration of an endogenous indicator gene and permits us to assay the level at which its regulation is titrated.

A MT-I-DHFR fusion gene (pMD) was used as the competitor and the endogenous MT genes were used as the indicator in these experiments (Fig. 1). pMD uses the mouse MT-I promoter to drive expression of mouse DHFR protein. DHFR provides a selectable marker for generating a graduated series of cell lines with increasing copies of pMD. This is accomplished through the use of the folate analog, methotrexate (MTX), which acts as a competitive inhibitor of DHFR. Cells can become resistant to MTX by overproducing DHFR protein, usually by means of gene amplification (reviewed in Stark and Wahl 1984).

The competitor gene pMD includes the complete MT-I genomic gene with 5' and 3' flanking regions except for sequences encompassing half of the first exon, all of the first intron, and half of the second exon (Fig 1). DHFR protein coding sequence replaces these missing segments. The regions of MT included in the construct have been shown to direct basal and metal induced expression, but not glucocorticoid induced expression (Mayo et al. 1982; Brinster et al. 1982; Pavlakis and Hamer 1983). The endogenous indicator genes are also shown in Figure 1. The easily identifiable sequence motifs shared by the competitor and indicator genes are indicated. In the promoter region, metal responsive elements (MREs) and several general factor binding sites, such as the Sp1 and TATA elements, are potential competitive elements. In the 3' half of the genes, the remaining MT coding DNA and a conserved 8 bp element of unknown function might serve the same purpose, although only the promoter region has been shown to affect MT expression.

Three independent cell lines, A, B, and C, were made by transfection of the competitor pMD fusion gene into a derivative of mouse L-cell fibroblasts (Kim and Wold 1985). These were all found to be resistant to concentrations of MTX 30-fold greater than the parental line (Table 1). This resistance is the result of increased levels of DHFR produced from the multiple copies of the transfected MT-DHFR gene (Mueller et al. 1988, and P. Mueller, data not shown). Cells with additional copies of the pMD competitor gene were generated from the original transfected lines by repeated rounds of selection using

increasing levels of MTX. This was done in the absence of any metal or glucocorticoid induction. Lines resistant to MTX levels 2 x 10^5 greater than the parental cell line's resistance were eventually obtained. This correlates with up to 1,400 copies of the competitor pMD per cell (Table 1).

Metal-specific titration of MT expression. We initially assayed titration of the endogenous MT-I indicator gene (Fig. 2). Total RNA from zinc induced and uninduced cells was examined by nuclease protection. This was done for the three independently derived cell lines at different stages of pMD amplification and for the parental cell line. In lines containing 500 or more copies of the competitor, the endogenous MT-I gene cannot be induced with zinc to the same level as that observed in the parental line or lines containing less than 100 copies (Fig. 2, lanes 8-14). The basal level of expression remains largely unaffected (Fig.2, lanes 1-7). The reduction in the metal stimulated expression of MT shows that the dominant mechanism of zinc induction is by the action of positively acting regulatory molecules that are present in limiting quantities. If a negative acting mechanism were involved, titration of negative factors should lead to an increase of MT expression in the absence of inducer as the copy number of competitor increase. This was not observed, so either there is no negative regulatory factors or they are in sufficient abundance so as not to be titrated. These results are in agreement with those of Seguin et al. (1984), who showed in a transient assay system that MT induction by cadmium is controlled by positively acting factors. Nevertheless, our data do not rule-out the possibility that MT metal regulation might use negative regulatory factors in addition to positive ones.

Is the titration specific to MT genes? All cell lines used in this work were previously transfected with HSV TK, and its expression was monitored to determine if the metal induced titration effect also affects expression of other genes. Levels of HSV TK

RNA remain largely unaffected by zinc induction and by the increased copies of pMD (Fig. 2). Furthermore, an endogenous cellular gene, glucose responsive protein 78 (GRP78), which is a housekeeping gene that shares similarity with members of the heat shock gene family (Munro and Pelham 1986; reviewed in Pelham 1986), is also unaffected by the increased copies of competitor (Fig. 3c). It is expressed in the parental cell line at an appreciable level, which is decreased several-fold upon zinc stimulation (Lanes 1 and 5, Fig 3c). In the A-series of cells, this pattern of expression remains unchanged (Lanes 3-4 and 6-8, Fig. 3c). We conclude from these results that the cellular factors titrated are not general ones such as RNA polymerase II, Sp1, or TATA binding factor that are used by HSV TK, GRP78, and MT.

The lack of metal inducibility in three independently transfected and amplified cell lines indicates that the titration is not the result of clonal variation. Further examination of MT titration was performed on the A-series of cells. Mouse MT-I and MT-II share many cis-acting sequence motifs such as Sp1, TATA, and MREs, yet each also possesses sites present in only one such as MLTF in MT-I and AP1 in MT-II (Fig. 1). If titration is dependent on shared elements, both genes should be titrated. MT-I and -II mRNA levels were measured in A-cells harvested under inducing and noninducing conditions at various stages of competitor amplification (Fig. 3a). As the copy number of competitor increases, the zinc induced expression of both genes decreases, whereas the basal expression remains unchanged. Similar results were obtained with cadmium induction (data not shown). In Figure 3d, MT expression is presented in terms of fold-induction. Without competing DNA, MT-I and -II are ~9- and 12-fold zinc inducible, respectively. As the copy number of competitor increases, the fold-inducibility by metal decreases. This trend continues until 700 copies of competitor are present. Further increases in the competitor copy number does not lead to any further decrease in inducibility. One possible reason for the residual 3-fold induction is that there is more than one species of factor contributing to the overall

metal response: one that is titrated completely by ~700 copies of competitor, and another that remains in excess at ~1,400 copies. In addition, the factors that drive basal expression appear to be in sufficient abundance, as this high number of competitor molecules does not titrate basal expression.

An unusual feature of the mouse MT genes is that even though the endogenous MTs are induced by glucocorticoids, no cloned DNA segment that retains this property upon transfection into cultured cells or transgenic mice has yet been identified (Mayo et al. 1982; Palmiter et al. 1982; Pavlakis and Hamer 1983). As expected from those prior studies, the transfected pMD construct is not inducible by dexamethasone (DEX), a synthetic analog of glucocorticoid (data not shown). If the mechanism of DEX induction is entirely independent of that used by metal induction, no titration of DEX induced MT should be observed. If, on the other hand, the two types of induction use a common factor, titration of DEX induced expression might be seen. The parental cell line shows a small, but reproducible 2- to 2.5-fold induction with DEX (Fig. 3b,d). This remains unchanged by increased copies of pMD, confirming that metal and DEX induction of MT are independent.

The titration of MT genes is limited to conditions of heavy metal induction, while basal level expression and induction by other means is unaffected. Furthermore, expression of other cellular and foreign genes is also unaffected by excess MT cis-acting elements. Taken together, the data suggest that multiple copies of pMD titrate limiting quantities of metal specific regulatory factors, although they do not reveal the mechanism of action for the limiting factors.

The metal specific regulatory factors titrated are transcriptional. In addition to MT promoter elements, pMD contains 5' and 3' untranslated MT flanking sequences, and these are present in the RNA transcript made from this competitor.

Although most of the MT metal induction takes place at the transcriptional level (Durnam and Palmiter 1981; Hamer and Walling 1982), it remains possible that the limiting factors are acting at a post-transcriptional level by interacting with these untranslated regions. MT-DHFR RNA levels reach 4 x 10^4 copies per cell upon induction in the A-60 line (data not shown), and this might titrate a putative factor regulating RNA stability. For example, a candidate cis-acting RNA recognition element for such a regulatory factor might be the sequence, TGTAAATA, found 18 to 27 bases 3' of translation stop in mouse, rat, chicken, monkey, and in most human MT RNAs. No function has yet been assigned to this element. To determine whether it is transcription or turnover that is affected by titration, we examined the kinetics of MT-I and -II mRNA accumulation over time in the parental and A-60 cell lines that contain 0 and 700 copies of pMD, respectively (Fig. 4a, data not shown). A change in the half-life of these transcripts between the parental cells and the A-60 cells would point to a stability factor.

The kinetics of mRNA accumulation after induction can be used to estimate the half-life of an transcript. The half-life of the mRNA is equal to the time needed for it to reach half-maximal induction. In the parental cells, MT mRNA reaches maximal accumulation by ~8 hours (Fig 4a, lanes 1-6, data not shown). When allowance is made for initiating metal response (which is very short, see below) and the nuclear dwell time, we can estimate the half-life of MT mRNA to be slightly less than 4 hours. The same kinetic result is obtained with the A-60 cells (Fig. 4a, lanes 7-12, data not shown), indicating that the stability of MT RNA is not significantly changed between the parental and titrated cells. If the titration effect were due entirely to post-transcriptional destabilization of MT mRNA, the half-life of the message would have had to decrease from ~4 hours to 90 minutes, and that difference would have been readily detected. Therefore, we conclude that the titration effect is acting at a mechanistic level prior to mRNA stabilization, either at the level of transcription or an early intranuclear processing event. This does not exclude a role for mRNA stabilization

during metal induction, but any such mechanism is not overwhelmed by the 4 x 10^4 transcripts of MT-DHFR carrying MT-I sequences.

The induction of MT by metals has been shown to be a rapid, primary response (Karin et al. 1980; Karin et al. 1981b; Mayo and Palmiter 1981). At the level of protein:DNA interaction, this has been confirmed by *in vivo* footprinting studies which show that the MREs of mouse MT-II are almost completely occupied 30 minutes after zinc treatment is initiated (P. Mueller and B. Wold, manuscript in preparation). Moreover, full occupancy in the footprint assay shows that the metal induction is highly uniform in the population of cells. Within 2 hours of metal induction, hnMT-I and -II RNA have reached new steady-state levels, where they remain constant for more than 4 hours (Fig. 4b, data not shown). In the A-60 cells, the hnMT RNAs also obtain new study-state levels within 2 hours after zinc induction, although the absolute amount is lower (see below, data not shown). Thus, in both the parental and titrated cells, the half-life of MT hnRNA is less than 2 hours and probably considerably less, as would be expected of a nuclear precursor.

The rapid, homogeneous response of the MT promoter to metal, the short half-life of MT hnRNA, and the fact that unspliced MT RNA is and early product of MT transcription allowed us to use the levels of hnRNA to compare the kinetics of metal induced MT transcription in the parental and titrating cell lines. In both cell types, the constant amount of precursor RNA after induction shows that transcription is quickly increased to a new constant level. Zinc induced and noninduced unspliced MT RNA levels were measured in parental cells and in various members of the A-series of cells (Fig. 4c, data not shown). As the copy number of pMD competitor DNA increases, the amount of zinc induced MT hnRNA decreases, suggesting that the rate of metal induced transcription also decreases with increasing competitor. This parallels the decrease that was observed using probes for the mature message (Fig 3a). As expected, the increase in competitor has no effect on the basal expression of the MT hnRNA (data not shown). Taken together with

the conclusion that RNA turnover is not affected by titration, these results show that metal specific transcriptional factors are in limiting quantities in these cells and can be titrated with greater than 500 copies of DNA competitor.

***In vivo* DNA binding of metal responsive factors is unaltered by the presence of titrating quantities of competitor sequences.** Transcriptional induction by metals has been shown to be dependent on cis-acting metal responsive promoter elements (MREs) (Carter et al. 1984; Stuart et al. 1984; Stuart et al. 1985; Searle et al. 1985; Searle et al. 1987; Culotta and Hamer 1989). These elements can be loosely defined by the sequence CTNTGC(A/G)CNCGGCCC, in which the underlined residues are the most highly conserved. Most MT genes have several MREs. For example, mouse MT-I has 6 and MT-II has 9 MREs that can be footprinted *in vivo* (Mueller et al. 1988; P. Mueller and B. Wold, manuscript in preparation). Mutagenesis studies show that multiple MREs are needed for proper induction. This has been confirmed by promoter fusion experiments that show that a basal level promoter requires multiple MREs for metal induced expression (Searle et al., 1985; Stuart et al. 1985). Specific, metal induced changes in DMS sensitivity at MREs *in vivo* support the idea that there is a unique metal responsive trans-acting factor (MRTF) that binds to MREs and induces transcription of MT genes (Anderson et al. 1987; Mueller et al. 1988). Factors in cellular extracts (Seguin and Hamer 1987; Westin and Schaffner 1988) and partially purified proteins of 74 (Imbert et al. 1989) and 108 kDa (Seguin and Prevost 1988) bind to MREs *in vitro*. The relationship of the 74 and 108 kDa polypeptides is not known, and neither completely mimics the induction or binding activity of the MRTF observed *in vivo* .

It seemed reasonable that the MREs present on the competitor gene could successfully compete for limited, metal responsive binding factors and that this would explain the lack of transcriptional induction in titrated cells. Titration of binding activity should cause a reduction in the occupancy of the MREs *in vivo*: either a large reduction in

occupancy at a subset of MREs or a small but equal reduction at all MREs. To test this, we performed *in vivo* DMS footprinting of the endogenous single-copy MT-II gene in the parental, A-0.3, and A-60 cell lines containing 0, 600, and 4,200 additional copies of the MREs, respectively (recall that each competitor gene has 6 MREs). There was no detectable difference in the intensity or pattern of footprints over any of the MREs of MT-II (Fig. 5, and data not shown). Furthermore, occupancy of the putative binding sites for Sp1, Ap1, and TFIID in MT-II do not show any difference at the footprint level (Fig. 5, and data not shown). A representative example of these comparative footprints is in Figure 5, in which the DMS protection patterns over the coding strand of MRE-A and the TATA site are shown. TATA shows moderate occupancy without zinc stimulation, which increases with the occupancy of the MREs. There is no apparent difference in this pattern or its intensity between the three cell types (Fig. 5, compare lanes 3, 6, and 9). The same result was observed in the other 300 bp of the MT-II promoter that was examined in these cell lines (data not shown). Neither the metal responsive binding factors nor any other apparent DNA binding factor within the proximal MT-II promoter are titrated. It is possible that the DMS footprint has failed to identify some metal responsive non-MRE interaction. However, extensive genetic, deletion, and mutational studies have clearly shown that MREs are both necessary and sufficient for metal induced transcription of MT genes together with TATA (Carter et al. 1984; Stuart et al. 1984; Searle et al. 1985). The same genetic studies show that no additional cis-acting element in this region is metal responsive. Thus it seems unlikely that there is an independent species of metal responsive cis-acting element that has escaped detection by all of these assays.

Discussion

Transcriptional regulation of metallothionein is controlled by two or more factors. In this work we have used a titration analysis to show that metal induction of metallothionein expression is partially dependent on limiting, positive acting factors. This is consistent with results of previous titration studies of metallothionein (Seguin et al. 1984; Scholer et al. 1986). The titrated factors are independent of those involved in MT basal expression or in MT induction by glucocorticoids, and they appear to be specific to metallothionein genes. This was also found to be the case for transient titration of mouse MT-I (Seguin et al. 1984), but not for human MT-II_A (Scholer et al. 1986). In that work, it was shown that SV40 and human MT compete for a common factor under conditions of metal stimulation. This is not surprising because SV40 and human MT-II_A share a number of sequence similarities. Because the experiments presented in this work used stably transfected cells, we were able to address the mechanism by which metal induction is effected. The metal specific trans-acting factors that are limiting are transcriptional in nature, but do not bind to the cis-acting metal response sequences (MREs) or any other DNA motif in the promoter of MT, as assayed by *in vivo* footprinting. This last conclusion was quite surprising, because the MREs have been shown to be responsible for transcriptional induction of MT by metals. There are two possible explanations for this apparent paradox.

The first is that there may be an additional cis-acting metal responsive DNA element that is distinct from the well-characterized MRE. The factor that would bind to this element would be essential for full metal response and has become limiting in the MT titration lines. The remaining 3-fold induction would be due to a different species of factor, probably those that interact directly with MREs. However, the fact that extensive deletion and mutational analysis has shown that the proximal ~300 bases of MT genes are all that is required for complete metal induction, argues against this possibility (Carter et al. 1984;

Stuart et al. 1984; Searle et al. 1985). Nevertheless, in these cell lines an additional cis-element may be especially important. If this model is correct, the *in vivo* footprints data suggest that such a discrete element would not be found in the proximal promoter, but instead would reside in a less well-studied region. A possible candidate is the conserved 8 bp element found in the 3' untranslated region of many MT genes.

An alternative possibility is that the titration effect is caused by a factor that acts through the MREs, but does not interact with them directly. In this case, the active metal responsive trans-acting factor would consist of a complex of two proteins. One protein would bind the DNA at the loosely conserved MRE on metal stimulation. This could be designated as the MRE binding protein (MRB). A second protein would interact with the bound MRB and would stimulate some early step in transcription. It would act as a metal responsive activator (MRA). The net effect would be to generate a metal responsive trans-activating factor (MRTF) in which the sequence-specific DNA binding function is joined to the transcriptional activating region through protein:protein interactions, instead of being present as two domains on a single protein. This latter situation is observed with many specific transcription factors such as SP1 and GAL4. These are one protein with two functionally distinct domains. One domain binds to a specific sequence of DNA and serves mostly to bring the other domain into proximity of the basic promoter. This second domain is required for the activity of the factor and may modulate transcription by interacting with one or more of the basic transcription factors such as the TATA binding TFIID or the RNA polymerase (reviewed in Mitchell and Tjian 1989). If the this model of MRTF is correct, the MRB, but not the MRA, would be in sufficient excess, so it would not be titrated. The remaining 3-fold induction of MT in the titrated cell lines could be caused by bound MRB activating transcription independently of MRA. The advantage of such a two-tier regulatory system might be that it allows fine tuning of the induction response.

Alternatively, the two proteins that make up the complete MRTF might interact with factors in addition to each other to modulate other cellular responses.

Regardless of which model explains the metal induction, the footprinting results allow us to estimate the minimum number of factors present in a cell that can bind to the MRE. The data shown in Figure 5 and our unpublished results indicate that the MREs are on average at least 50% occupied after metal stimulation. Therefore, at any given time 2,100 MREs are bound by factors in A-60 cells. Because this number is based on protection from DMS, this is undoubtably an underestimate and may be as high as 4,200 (see Mueller et al., 1988 for discussion). There must be at least 2,100 factors present per cell, and because the cell lines used in this study do not contain sufficient numbers of MREs to titrate binding activity, the actual number of factors could be significantly higher.

Titration and *in vivo* footprinting data have led us to propose that there is an additional species of factor, distinct from that binding directly to MRE, that is responsible for metal induction. We favor the model that proposes a trans-acting metal responsive complex composed of two or more proteins acting together to stimulate metal induced transcription of MT. This is consistent with the results of a transient MT-I titration study that used replicating SV40 based plasmids for competitors and indicators (Seguin et al. 1984). Because that study was done transiently, the authors were not able to address whether titration involved direct protein:DNA interactions. However, using various deletion constructs, they found that the cis-acting elements responsible for titration are in MRE containing sequences. Either the MREs themselves or sequences near them are responsible for the titration. Because we did not detect any differences in protein:DNA interactions in this region, the most likely explanation is that the titrated factor is dependent on the MRE for its activity, but acts through protein:protein interactions. This would be consistent with the accumulated genetic data on the cis-acting regions of MT responsible for metal induction, although it remains possible that a previously unidentified metal dependent

cis-acting element is instead responsible for our observations. *In vitro* purification of the metal responsive trans-acting factor/complex, along with further mapping of cis-acting MT sequences will be helpful in identifying the additional factors or elements responsible for MT metal induction.

MATERIALS AND METHODS

Competitor construction. Plasmid MD (pMD) was constructed as per Mueller et al. (1988). The MT-I promoter fragment used in this plasmid contains ~1775 bp of 5' non-transcribed and 68 bp of non-translated sequences between Eco RI and Bgl II. The Bgl II site ends 8 base pairs 5' of the MT translation start. The 3' metallothionein fragment contains the complete second intron and poly A addition site of MT-I as an Alu I-Hind III fragment with the Alu I site converted to a Bgl II site. These fragments were inserted into pBR322 to create pMT-33. The coding region of DHFR, along with 60 base pairs 5' and 80 base pairs 3' of the translation start and stop, respectively, was removed from SV3-DHFR as a FnuD II-Bgl II fragment and inserted into the Bgl II site of pMT-33 after addition of a Bam HI linker at the FnuD II site.

Construction of plasmids for RNA probes. The plasmids used for probe synthesis were pSP6-MT-I, pMT-I INT, pT7-MT-II (a gift from P. Garrity), pSP6-αHSV TK (Kim and Wold 1985), and pT7-GRP78 (a gift from P. Garrity). pSP6-MT-I was constructed by inserting an Eco RI/Xba I genomic mouse MT-I fragment into an SP64 vector. *In vitro* transcription with SP6 polymerase generates an antisense MT-I probe that is complementary to part of the first intron and all of the the first exon, if cut with Bst EII or part of the first intron and part of the first exon, if cut with Bgl II. pMT-I INT was constructed by inserting a Bgl II/Hin PI fragment of pMT-IΔi (an intron-less MT-I cDNA plasmid, a gift of R. Palmiter) into a Bam HI/Acc I cut pT3/T7 vector. *In vitro* transcription with T7 polymerase generates an antisense MT-I mRNA probe that is complementary to the spliced MT-I sequences not present in pMD. *In vitro* transcription of pT7-MT-II with T7 polymerase generates a antisense MT-II RNA probe complementary to the first exon and part of the first intron of mouse MT-II. *In vitro* transcription of Ban I cut SP6-αHSV TK with SP6 polymerase generates an antisense probe that is complementary to 293 bases of HSV TK mRNA. *In vitro* transcription of pT7-GRP78 with T7

polymerase generates a probe that is complementary to the 3' untranslated region of the mouse GRP78 gene.

Cell Culture. The parental cell line, A-O.3, B-0.3, and C-0.3, (formerly P, KTO-A, KTO-B, and KTO-C, Kim and Wold 1985) and A-60, B-60, and C-60 were gifts from S. Kim, and were grown as per Mueller et al. (1988). Cell lines A-5, A-20, and A-1000 were generated as in Kim and Wold (1985), except that dialized serum was used. Growth media for all cell lines except the parental line were supplemented with methotrexate (MTX) as indicated in Table 1. The presence of MTX did not influence the expression, induction, or titration of the MT genes (data not shown). The copy number of the pMD was determined by Southern blotting as per Maniatis et al. (1982) (J.Jong and S. Salser, unpublished data).

RNA preparation and measurement. Cells were ~70% confluent before induction. Prior to harvesting the RNA, fresh media were added to the plates with or without 90 μM zinc sulfate or 100 nM dexamethasone (DEX). Induction was for 8 hours unless otherwise indicated. Total RNA for each condition was prepared from two 150 cm plates by the method of Chirgwin et al. (1979) as modified by Mueller et al. (1988). RNase protection was performed as described by Zinn et al. (1983) as modified by Kim and Wold (1985). Gel slices containing the protected probe were excised and counted for quantitation.

***In vivo* DMS footprinting.** *In vivo* and *in vitro* DMS/piperidine cleavage of DNA was as per Mueller et al. (1988) as modified by Mueller and Wold (1989). Cells for *in vivo* footprinting were treated identically to those used for RNA measurements. Ligation mediated PCR was as per Mueller and Wold (1989). Primers used were: #1:CGGTTTGA-AGAGTTCTAGGAGCGTG, #2: AGGAGCGTGATGGAGAGAAGCACGC, #3: AGG-AGCGTGATGGAGAGAAGCACGCGG.

ACKNOWLEDGEMENTS

The authors wish to thank R. Palmiter and P. Garrity for supplying DNAs used in plasmid and probe constructions, and J. Jong and S. Salser for assistance in DNA and RNA measurements. We thank members of the Wold group for interesting discussions and critical reading of the manuscript.

REFERENCES

Andersen, R.D., S.J. Taplitz, S. Wong, G. Bristol, B. Larkin, and H.R. Herschman. 1987. Metal dependent binding of a factor *in vivo* to the metal responsive elements of the metallothionein-I gene promoter. *Mol. Cell. Biol.* **7**: 3574-3581.

Brinster, R.L., H.Y. Chen, R. Warren, A. Sarthy, and R.D. Palmiter. 1982. Regulation of metallothionein-thymidine kinase fusion plasmids injected into mouse eggs. *Nature* **296**: 39-42.

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 gene. *Proc. Natl. Acad. Sci.* **81**: 7392-7396.

Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**: 5294-5299.

Culotta, V. C., and D. Hamer. 1989. Fine mapping of a mouse metallothionein gene metal response element. *Mol. Cell. Biol.* **9**: 1376-1380.

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.

Glanville, N., D. M. Durman, and R. D. Palmiter. 1981. Structure of mouse metallothionein-I gene and its mRNA. *Nature* **292**: 267-269.

Hager, L. J., and R. D. Palmiter. 1981. Transcriptional regulation of mouse liver metallothionein-I gene by glucocorticoids. *Nature* **291**: 340-342.

Hamer, D.H., and M. Walling. 1982. Regulation *in vivo* of a cloned mammalian gene: Cadmium induced the transcription of a mouse metallothionein gene in SV40 vectors. *J. Mol. Appl. Genet.* **1**: 273-288.

Hamer, D.H. 1986. Metallothionein. *Ann. Rev. Biochem.* **55**: 913-951.

Imbert, J., M. Zafarullah, V. C. Culotta, L. Gedamu, and D. Hamer. 1989. Transcription factor MBF-I interacts with metal regulatory elements of higher eucaryotic metallothionein genes. *Mol. Cell. Biol.* **9**: 5315-5323.

Karin, M., R. D. Anderson, E. Slater, K. Smith, and H. R. Herschman. 1980. Metallothionein mRNA induction in HeLa cells in response to zinc or dexamethasone is a primary induction response. *Nature* **286**: 295-297.

Karin, M., E. P. Slater, and H. R. Herschman. 1981a. Regulation of metallothionein synthesis in HeLa cells by heavy metals and glucocorticoids. *J. Cell. Physio.* **106**: 63-74.

Karin, M., R. D. Anderson, and H. R. Herschman. 1981b. Induction of metallothionein mRNA in HeLa cells by dexamethasone and by heavy metals. *Eur. J. Biochem.* **118**: 527-531.

Karin, M. 1985. Metallothioneins: Proteins in search of function. *Cell* **41**: 9-10.

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.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A laboratory manual. Cold Spring Harbor.

Mayo, K. E., and R. D. Palmiter. 1981. Glucocorticoid regulation of metallothionein-I mRNA synthesis in cultured mouse cells. *J. Biol. Chem.* **256**: 2621-2624.

Mayo, K.E., R. Warren, and R.D. Palmiter. 1982. The mouse metallothionein-I gene is transcriptionally regulated by cadmium following transfection into human or mouse cells. *Cell* **29**: 99-108.

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

Mueller, P. R., S. J. Salser, 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.

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

Munro, S., and H. R. B. Pelham. 1986. An Hsp70-like protein in the ER: Identity with the 78 kd glucose regulated protein and immunoglobulin heavy chain binding protein. *Cell* **46**: 291-300.

Palmiter, R.D., H.Y. Chen and R.L. Brinster. 1982. Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring. *Cell* **29**: 701-710.

Pavlakis, G.N., and D.H. Hamer. 1983. Regulation of a metallothionein-growth hormone hybrid gene in bovine papilloma virus. *Proc. Natl. Acad. Sci.* **80**: 397-401.

Pelham, H. R. B. 1986. Speculations on the functions of the major heat shock and glucose regulated proteins. *Cell* **46**: 959-961.

Scholer, H. R., and P. Gruss. 1984. Specific interactions between enhancer-containing molecules and cellular components. *Cell* **36**: 403-411.

Scholer, H., A. Haslinger, A. Heguy, H. Holtgreve, and M. Karin. 1986. *In vivo* competition between a metallothionein regulatory element and the SV40 enhancer. *Science* **232**: 76-80.

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.

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.

Searle, P. F., G. W. Stuart, and R. D. Palmiter. 1987. Metal regulatory elements of the mouse metallothionein-I gene. *Experientia Suppl.* **52**: 407-414.

Seguin, C., B.K. Felber, A.D. Carter, and D.H. Hamer. 1984. Competition for cellular factors that activate metallothionein gene transcription. *Nature* **312**: 781-785.

Seguin, C., and D.H. Hamer. 1987. Regulation *in vitro* of metallothionein gene binding factors. *Science* **235**: 1383-1387.

Seguin, C., and J. Prevost. 1988. Detection of a nuclear protein that interacts with a metal regulatory element of the mouse metallothionein-I gene. *Nucl. Acids Res.* **16**: 10547-10560.

Stark, G. R., and G. M. Wahl. 1984. Gene amplification. *Ann. Rev. Biol.* **53**: 447-492.

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.* **81**: 7318-7322.

Stuart, G.W., P.F. Searle, and R.D. Palmiter. 1985. Identification of multiple metal regulatory elements in mouse metallothionein-I promoter by assaying synthetic sequences. *Nature* **317**: 828-831.

Westin, G., and W. Schaffner. 1988. A zinc responsive factor interacts with a metal regulated enhancer element (MRE) of the mouse metallothionein-I gene. *EMBO J.* **7**: 3763-3770.

Yagle, M. K., and R. D. Palmiter. 1985. Coordinate regulation of mouse metallothionein-I and -II genes by heavy metals and glucocorticoids. *Mol. Cell.Biol.* **5**: 291-294.

Zinn, K., D. DiMaio, and T. Maniatis. 1983. Identification of two distinct regulatory regions adjacent to the human β-interferon gene. *Cell* **34**: 865-879.

Table 1. Copy number of competitor pMD per cell and resistance to MTX.

Cell line	µM MTX[¥]	Copies of pMD[*]
Parental (3.3)	0.01	0
A-0.3	0.3	100
A-5	5.0	300
A-20	20.0	650
A-60	60.0	700
A-1000	2000.0	1,400
B-0.3	0.3	70
B-60	60.0	600
C-0.3	0.3	70
C-60	60.0	1,100

(¥) Concentration of methotrexate in the growth media.

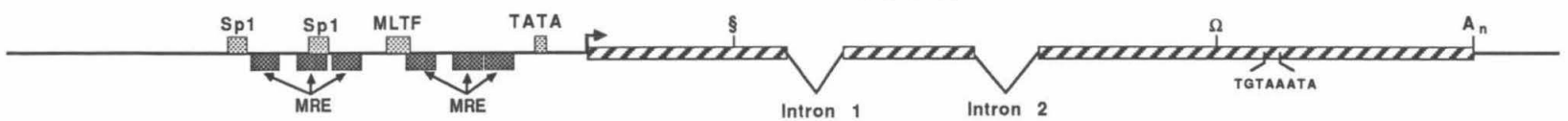
(*) Copy number of competitor gene pMD per cell.

Figure 1. Schematic map of competitor and indicator genes. These maps are based on the published sequences of mouse MT-I (Glanville et al. 1981) and mouse MT-II (Searle et al.1984), and construction of pMD (Mueller et al. 1988). The competitor gene and MT-I share regions of sequence identity (see Methods). Thin lines represent sequence that are not present in the mRNA. Thick lines represent sequences that are present in the mRNA: light-striped lines: MT-I mRNA coding sequence; thick heavy-striped lines MT-II mRNA coding sequence. Boxes represent sites where factors were observed to bind *in vivo* (Mueller et al. 1988, Mueller and Wold, unpublished observations). Light stippled boxes: basal elements occupied with and without metal induction; heavy stippled boxes: (MREs) metal responsive elements, elements occupied with metal induction. Most MREs were previously identified by Stuart et al. (1985). Arrow: transcription start; §: translation start; Ω: translation stop; A_n poly-adenine addition site. TGTAAATA: sequence found in mouse, rat, chicken, monkey, and most human MT RNAs 18 to 27 bases 3' of translation stop.

Competitor gene pMD

Endogenous Indicator genes MT-I and MT-II

MT-I



MT-II

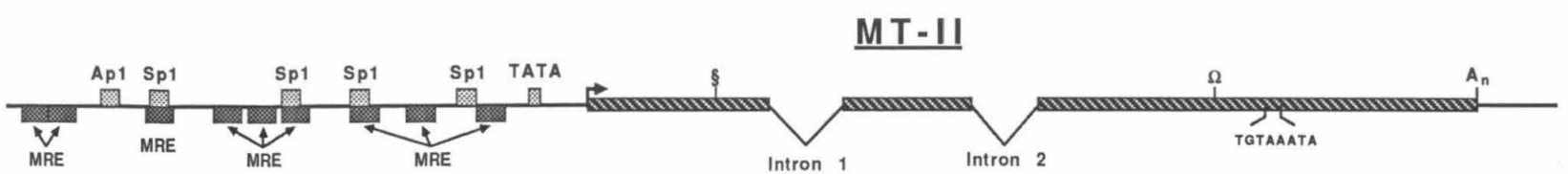


Figure 2. RNA levels of the endogenous MT-I indicator gene and HSV TK in three independently derived lines. MT-I measurement was made using 15 μg of total RNA from the indicated cell lines and a uniformly labeled pMT-INT riboprobe. HSV TK measurement was made using 10 μg of the same RNA and a uniformly labeled pSP6-αHSV TK riboprobe. For copy number of competitor pMD, see Table 1. See Methods for experimental details.

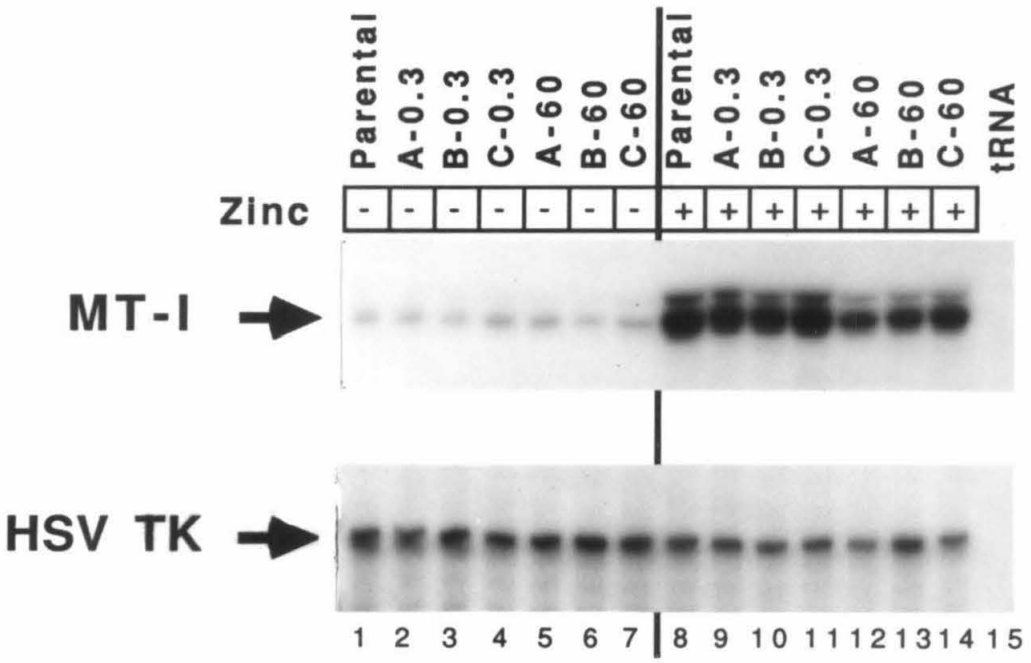
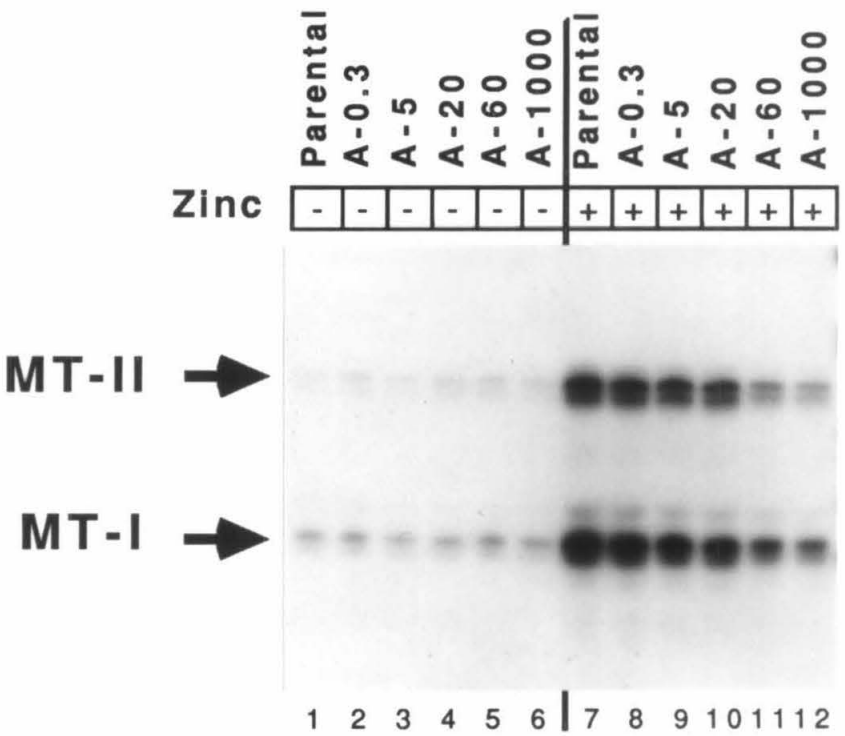
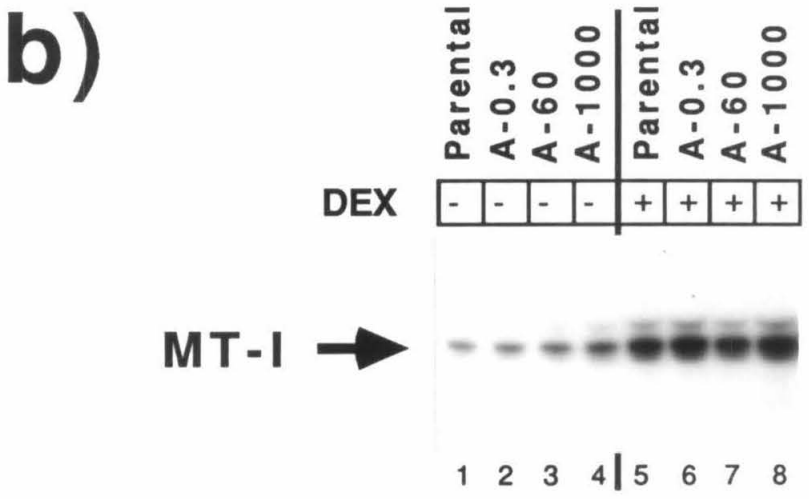
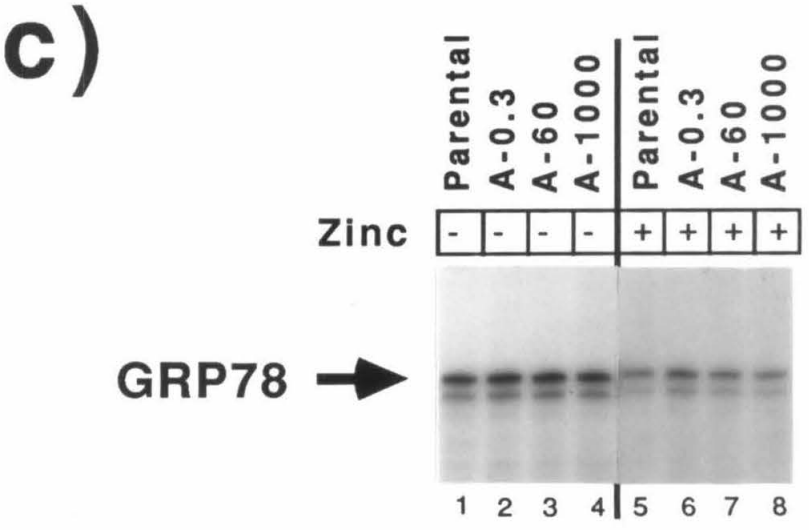

Parental
A-0.3
B-0.3
C-0.3
A-60
B-60
C-60
Parental
A-0.3
B-0.3
C-0.3
A-60
B-60
C-60
tRNA
Zinc
- - - - - - - + + + + + + +
MT-I
HSV TK
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Figure 3. RNA levels of the A series cell line. (a) RNA levels of the endogenous MT-I and MT-II indicator genes with and without zinc. MT-I and MT-II RNA measurement were simultaneously made using 5 μg of total RNA from the indicated cell lines and a mixture of uniformly labeled pMT-INT and pT7-MT-II riboprobes. The induction ratio is quantitated in panel d. (b) RNA levels of the endogenous MT-I indicator gene with and without DEX. MT-I RNA measurement was made using 15 μg of total RNA from the indicated cell lines and a uniformly labeled pMT-INT riboprobe. The induction ratio is quantitated in panel d. (c) RNA levels of the endogenous GRP78 gene with and without zinc. GRP78 RNA measurement was made using 30 μg of total RNA from the indicated cell lines and a uniformly labeled pT7-GRP78 riboprobe. (d) Fold induction of MT-I and MT-II versus copy number of pMD per cell. Data from gels used in panels a and b, and unpublished results . See Methods for experimental details.

a)


Parental
A-0.3
A-5
A-20
A-60
A-1000
Parental
A-0.3
A-5
A-20
A-60
A-1000
Zinc
- - - - - - + + + + + +
MT-II
MT-I
1 2 3 4 5 6 7 8 9 10 11 12


b)
Parental
A-0.3
A-60
A-1000
Parental
A-0.3
A-60
A-1000
DEX
- - - - + + + +
MT-I
1 2 3 4 5 6 7 8


c)
Parental
A-0.3
A-60
A-1000
Parental
A-0.3
A-60
A-1000
Zinc
- - - - + + + +
GRP78
1 2 3 4 5 6 7 8

d)

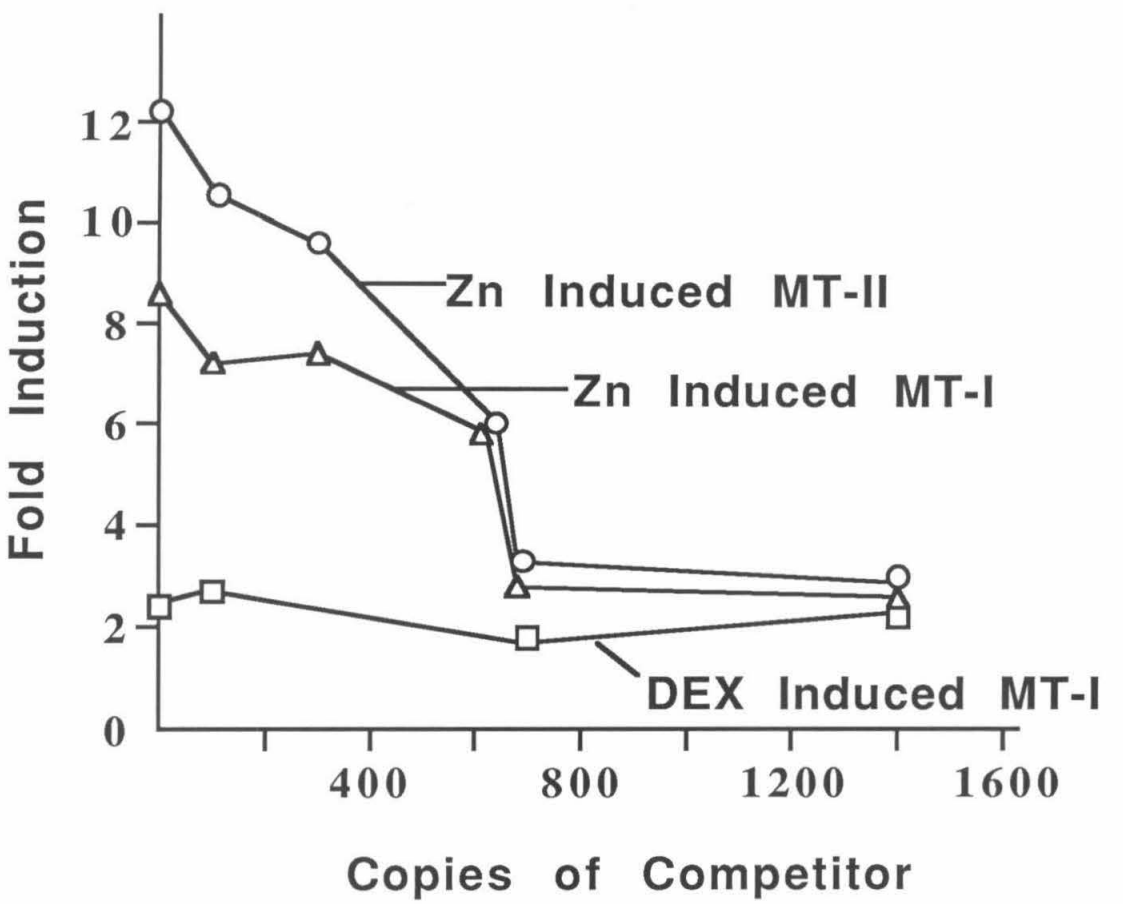
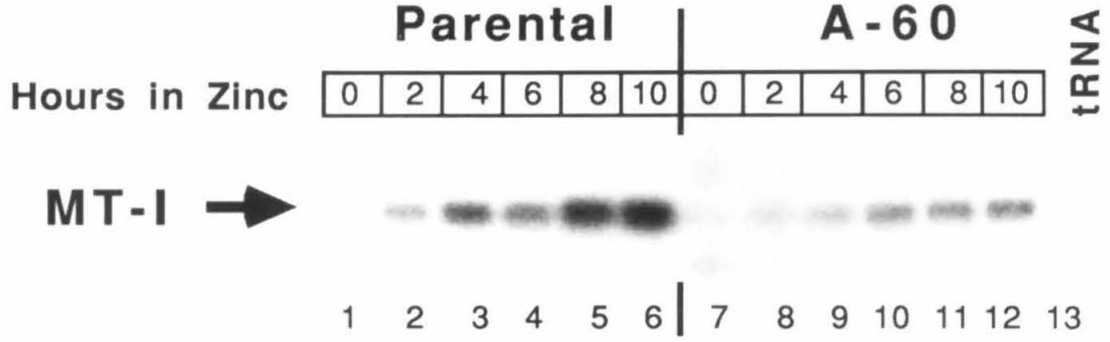
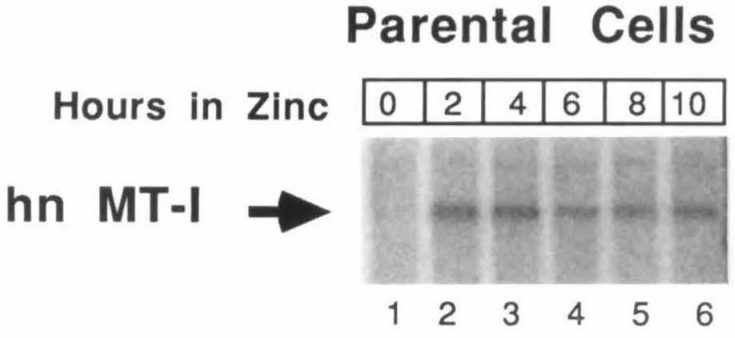

Fold Induction
12
10
8
6
4
2
0
Zn Induced MT-II
Zn Induced MT-I
DEX Induced MT-I
400
800
1200
1600
Copies of Competitor

Figure. 4. (a) Stability of MT-I mRNA in parental and A-60 cells is identical. (b) MT-I hnRNA levels reach a peak level 2 hours after induction. (c) Decrease in the level of MT-I hn RNA suggests that titration is acting at the level of transcription. Riboprobe use in these experiments was pSP6-MT-I cut with Bst EII (panels a and b) or Bgl II (panel c). The probe generated from pSP6-MT-I cut with Bst EII (panels a and b) will protect all of the first exon and half of the first intron, whereas the probe generated from pSP6-MT-I cut with Bgl II (panel c) will protect only 37 bases of the first exon and half of the first intron. The band identified as hn MT RNA in panels b and c is the unspliced precursor RNA containing both exon and intron sequences. See Methods for experimental details.

a)


Parental
A-60
Hours in Zinc
0 2 4 6 8 10 0 2 4 6 8 10
tRNA
MT-I
1 2 3 4 5 6 7 8 9 10 11 12 13

b)


Parental Cells
Hours in Zinc
0 2 4 6 8 10
hn MT-I
1 2 3 4 5 6

c)

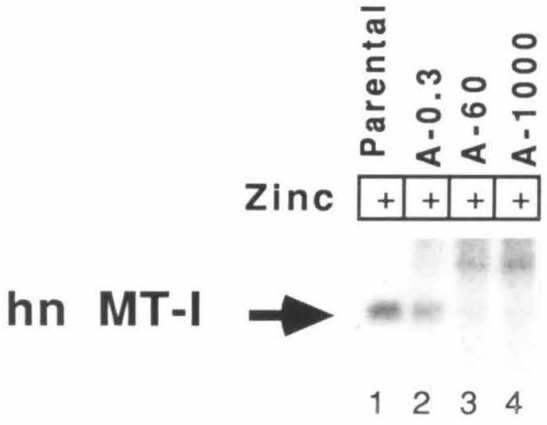
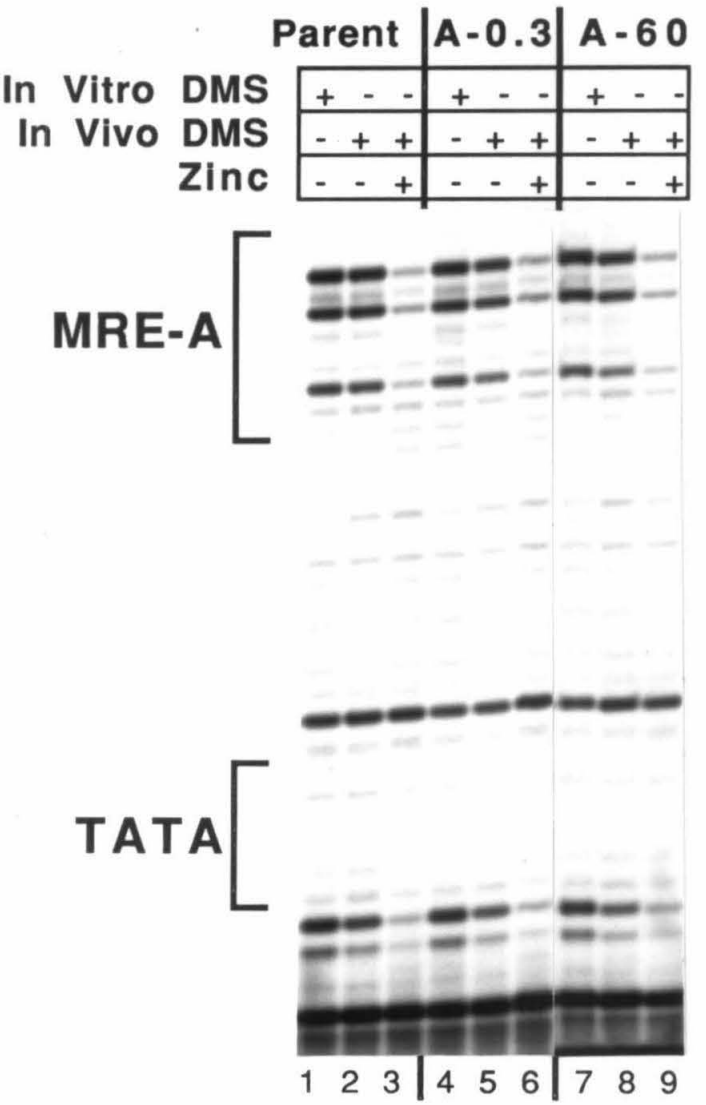

Parental
A-0.3
A-60
A-1000
Zinc
+ + + +
hn MT-I
1 2 3 4

Figure 5. *In vivo* occupancy over promoter elements is not affected by titration. *In vivo* DMS footprints over the TATA element and MRE-A of the coding strand of MT-II. Cell lines and conditions are indicated. Each lane contains equal amounts of genomic DNA. Ligation mediated footprinting was performed as described in Mueller and Wold (1989) and in Methods. Zinc induction was for 4 hours.


Parent
A-0.3
A-60
In Vitro DMS
In Vivo DMS
Zinc
+ - - + - - + - -
- + + - + + - + +
- - + - - + - - +
MRE-A
TATA
1 2 3 4 5 6 7 8 9

Appendix I

Ligation Mediated PCR for Genomic Sequencing and Footprinting

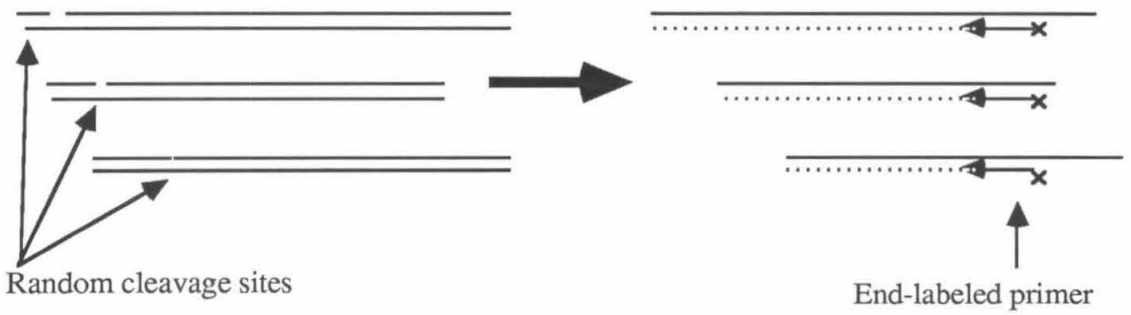
This is a detailed description of the ligation mediated PCR technique that I developed. It was used in Chapters 2 and 4, and in Appendix II.

PCR Genomic Sequencing for Footprinting

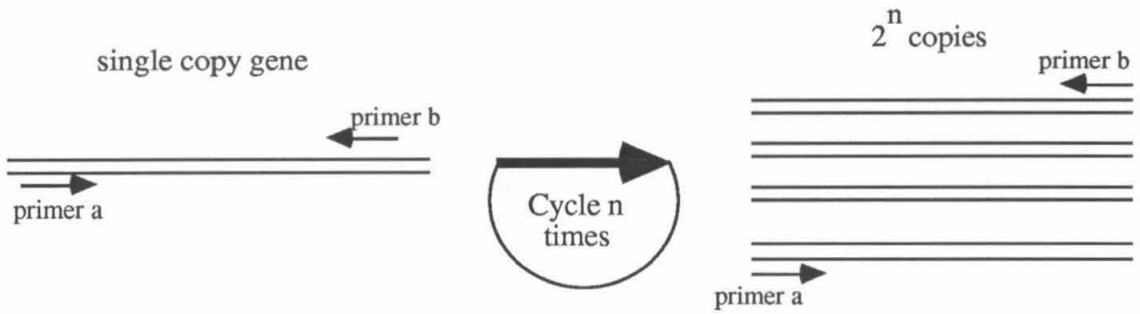
This procedure has appeared recently in *Science,* where it was used to footprint a muscle specific enhancer *in vivo* (Mueller and Wold 1989). This is a more detailed description of the technique. Because of the number of inquiries about the methodology, we felt this handout might prove helpful. We have footprinted five genes using this procedure, and we know of a number of labs that have had success footprinting, sequencing and cloning genes with this technique. If you have any questions or suggested improvements, please call us. Paul Mueller and Barbara Wold; California Institute of Technology; Division of Biology; Pasadena , CA 91125; (818) 356-4923 or (818) 356-4916.

Introduction:

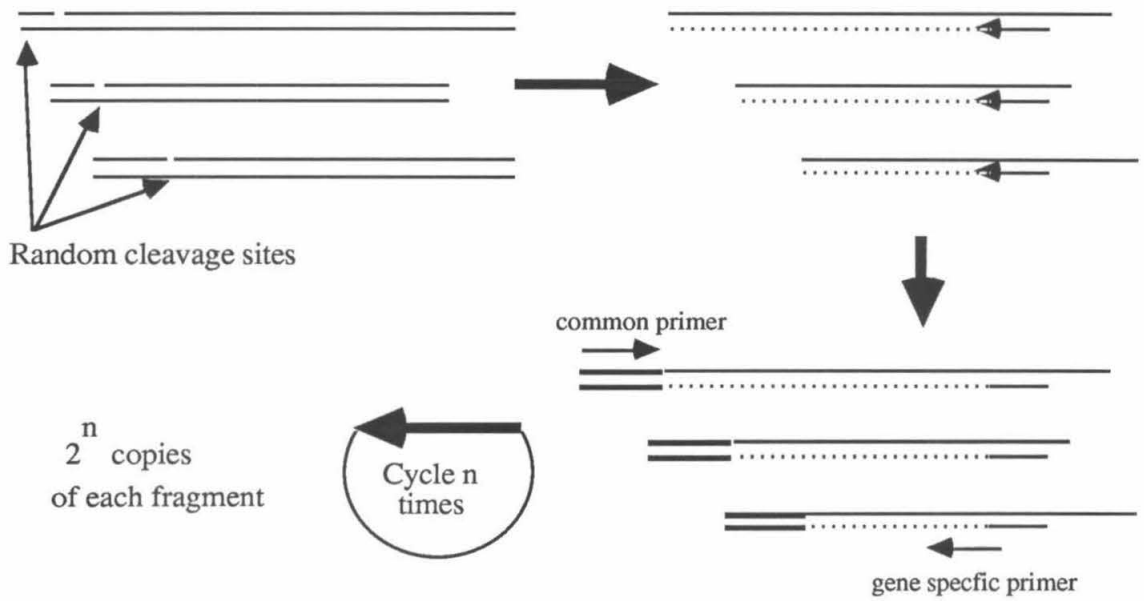
We have developed this technique to visualize footprints of single-copy genes in mammalian cells. Previously, we had used primer extension to footprint multiple copies of the metallothionein promoter (Mueller et al. 1988). In that method, randomly cleaved DNA was visualized by extension of an end-labeled primer specific to the metallothionein



promoter. Because of limitations in sensitivity, it was applicable only to small genomes or to genomes that contained multiple copies of the gene of interest (> 20 per genome). PCR or the polymerase chain reaction (reviewed in Saiki et al. 1988; White et al. 1989) has been used to amplify specific fragments of single-copy genes 10^6 times . It relies on two primers that flank the specific fragment of DNA to be amplified. Repeated cycles of



denaturation, annealing and extension exponentially amplify the DNA, each cycle doubling the number of templates. This technique is not immediately applicable to sequencing or footprinting because one end of the template is random. The procedure outlined here is based on the idea that by attaching a common sequence to this random end it would create suitable substrates for PCR, and therefore, exponentially amplify each representative of the sequence ladder. The labeling step itself is essentially the same as the indirect labeling



method used previously, except that now it is done on 10^4 more templates. In practical terms, starting with as little as 1 μg of mammalian DNA (equivalent to ~10^5 cells), a genomic sequence ladder of a single-copy gene can be seen with an overnight exposure. Additional uses of this method are: sequencing genomic DNA (Pfeifer et al. 1989),

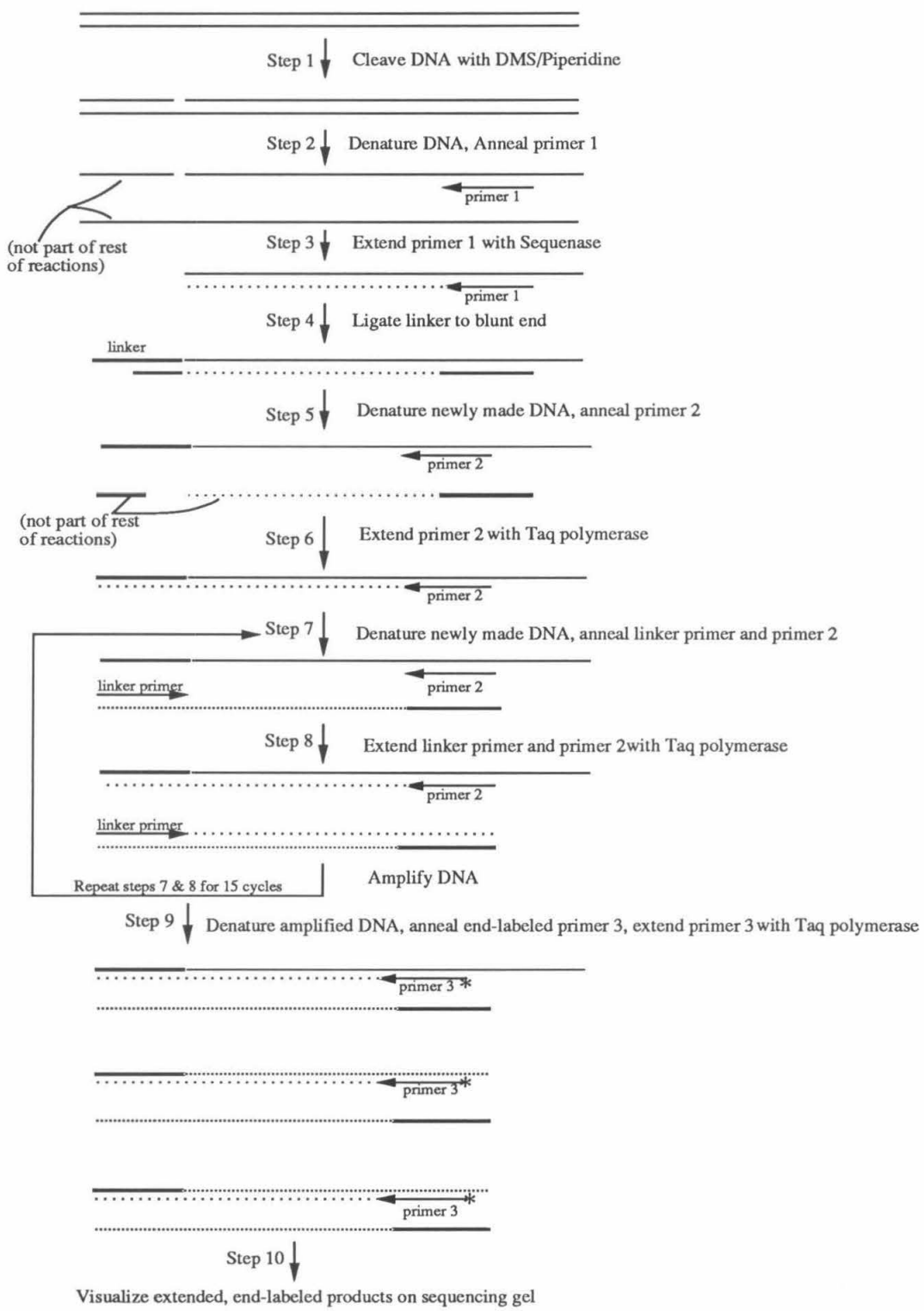
studying *in vivo* methylation patterns of C residues (Pfeifer et al. 1989), or cloning promoter elements. The latter can be done by picking primers from known sequence of a cDNA and extending those primers into the promoter (see below).

The details of the method are outlined below, however, two underlying technical points deserve emphasis at outset. First, in order to keep background to a minimum, the hybridizations and extensions should be performed at the highest temperatures that the polymerases and annealed primers will allow. The Sequenase extension is performed at 47°C and the Taq hybridizations are incubated at 63-69°C for this reason. These temperatures may need to be adjusted on a case-by-case basis, depending on the T_m of the primers used. The second important point is that background can be greatly reduced by using multiple primers, each primer having an extending end 3' to the previous one. This adds a level of specificity that can not be obtained with a single 25 base primer.

Outline of Method:

(see figure on next page) **Step 1.** The starting material is genomic DNA that has been cleaved in such a way that it retains a 5' phosphate at the cleavage site. This can be done at a specific site by using a restriction enzyme, or at random bases by Maxam and Gilbert chemistry (see Saluz and Jost 1987 for applications to genomic DNA) or DNase-I treatment. I have worked mainly with DMS/piperidine chemistry to create G ladders, and will describe the method from that viewpoint. (Please see Limitations section for DNase-I or restriction enzyme cleavages.) For clarity, I've shown only one cleavage product at a specific site, but in practice there would be a population of random cleavage products. **Step 2.** The cleaved DNA is denatured and a specific primer is annealed to a region of interest. **Step 3.** This primer is extended with a processive polymerase such as Sequenase to create a blunt end at the random cleavage site (in the example of DMS/piperidine treated DNA, at the former location of a guanine). **Step 4.** A common linker is ligated to this blunt end. Because the linker is staggered on one end, it will ligate

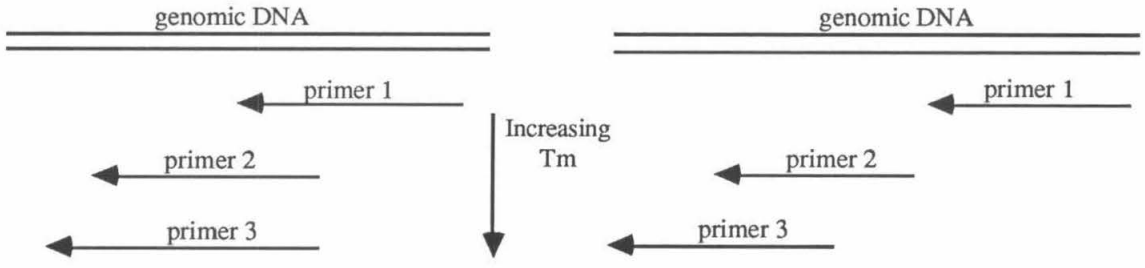
PCR Genomic Sequencing Scheme



to the blunt-ended genomic Sequenase products in only one orientation. In addition, the lack of 5' phosphates on the linker prevents linker:linker ligation products from forming. These could interfere with subsequent steps in the procedure. **Step 5.** The DNA is denatured and a second primer is annealed to the genomic DNA. In order to minimize background, it is important that the extending end of the second primer be 3' to the first. **Step 6.** This second primer is annealed to the genomic DNA and then extended through the ligated linker region. **Step 7 & 8.** This is now a suitable substrate for a PCR reaction. On one end (left in the figure) there is a linker sequence which a linker primer can anneal to, and on the other end (right in the figure) there is a genomic sequence which a gene-specific primer can anneal to. Only molecules that have both of these sequences will be exponentially amplified (~15,000-fold); molecules with only one of the sequences will be linearly amplified (~15-fold). The difference between the exponential amplification and background from linear amplification provides a favorable signal-to-noise ratio. Note: in practice, steps 5 to 8 are done simultaneously. (The exponentially amplified fragment is now suitable for cloning—if sequencing or footprinting continue.) **Step 9 &10.** For the last extension, a third overlapping primer is used to label the DNA indirectly, and these end-labeled extension products are visualized on a standard sequencing gel. Once again, this primer should have an extending end that is 3' to the last primer, but in addition it is must overlap the last primer (see discussion of primers below). In order to simplify the reaction, the second primer could be used to label the DNA indirectly (use a hot primer in steps 7 & 8). Doing so will result in higher background and greater exposure to radiation by the user. For the best results, a third primer is highly recommended. Alternativly, visualization could be performed by blotting as in Pfeifer et al. (1989).

Design and position of primers: Shown on the next page is a schematic of two possible positioning arrangements of the primers. I have tested a number of different combinations and can make the following suggestions. The first and second primer may or

may not overlap, but the extending end of the second primer must be 3' to the first primer. If they do overlap, the overlap should be less than ~12 bases. The third primer, on the other hand, can completely overlap the second primer and extend a few extra bases 3' to it (as shown on the left), or it can be positioned with an overlap of ~15 bases (as shown on the right). We have found, somewhat surprisingly, that if the second and third primers do



not overlap, the labeling extension will not work. They must compete for the same binding site. The exact reason for this is unclear, but we believe it may involve the extension of primer 2 during the hybridization phase excluding the hybridization of primer 3. (Because of the high temperatures used during the hybridization phase of the PCR cycle, extension probably begins immediately after the primer binds.) Alternatively, extension of of an upstream primer by Taq polymerase may displace a downstream primer—perhaps through an exonuclease activity. If primer 3 has a higher T_m than primer 2 it can compete better for the overlapping binding site. In general, the T_m of the primers should increase from primer 1→ 2→ 3. This way the Sequenase primer (primer 1) will be unstable under the amplification conditions and the amplification primer (primer 2) will be unstable under the labeling conditions. In practice, this doesn't appear to be that critical, but it would probably be a bad idea to have the relative T_ms go the other way. The T_ms can be adjusted by changing the GC content and/or the length of the primers, and will be discussed in more detail in the comment section at the end of this handout. The linker primer is part of the linker itself and is explained below.

Linker: The linker I have used is of the following structure and sequence:

```
5'-GCGGTGACCCGGGAGATCTGAATTC-3' 25mer, 60% GC
                CTAGACTTAAG-5' 11mer, 36% GC
```

The longer oligomer (which will be ligated to the genomic DNA) contains restriction sites for Bst EII, Sma I, Bgl II, and Eco RI. This oligomer will also function as the linker primer in the PCR amplification. Restriction sites are important only for cloning purposes, and are not used in sequencing or footprinting. The sole function of the 11mer is to create a blunt end so that T4 DNA ligase can ligate the 25mer to the 5' end of the genomic DNA. By making the short oligomer small and of low GC content, it cannot serve as a primer in subsequent Taq polymerase reactions. Neither of the oligomers has 5' phosphates. The 3' end of the 25mer is ligated to 5' end of the genomic DNA, and the 11mer does not ligate to anything. The exact sequence, GC content, length, or type of restriction sites in this linker is not that important. What is important is that: 1) the short oligomer should be able to bind to the long oligomer at ligation conditions, but not Taq conditions; 2) the long oligomer should have a T_m in Taq buffer comparable to primer 2 (see comment section); and 3) the linker is a ligatable structure.

Limitations, Cautions, and Disclaimers: This method has been working well for the purpose of genomic Maxam and Gilbert sequencing (Pfeifer et al. 1989) and genomic DMS footprinting (Mueller and Wold 1989). The only caution we have at this time is that some unique sequence structures are not evenly represented in the sequence ladder. Specifically, one of the promoters we are working with has a run of 13 G's in a row. These G's do not have uniform intensity in the genomic DMS/piperidine ladder. We are investigating the cause of this, but preliminary results suggest that the reason is either that the Sequenase extension is stopping before the end of this sequence, or that it is difficult to ligate a blunt-ended linker to such a structure. (See comment sections on Sequenase

reactions and ligation conditions) The rest of the G ladder looks fine, so we feel this problem will be rare.

Because of statistical limitations, we recommend that $\geq 10^5$ nuclei be used in each reaction if sequencing or footprinting. Using fewer nuclei can result in missing bands or fluctuations in the intensity of bands. See Pfeifer et al. (1989) for discussion.

Preliminary attempts to adapt this technique to DNA that has been cleaved with restriction enzymes or DNase-I have been only partially successful. Enzymatic cleavage of DNA leaves a 5' phosphate and a 3' OH. Piperidine treatment, on the other hand, leaves DNA with a 5' phosphate as well as a 3' phosphate. (Piperidine is used to cleave DNA that has been chemically modified by the Maxam and Gilbert sequencing or DMS footprinting chemistry.) A 3' phosphate is enzymatically dead in that it can not be extended or ligated, and such a 3' end may be important in reducing background. We are currently investigating this possiblility.

Methods:

The procedure will be discussed from the point of view of genomic sequencing.

Solutions:

1) Linker soln.(prepare ahead of time):

The oligomers for the linker are gel-purified before they are hybridized to each other. I have not played with the hybridization conditions for the linkers, but the following conditions seem to work well for me. The Tris that the ligase will eventually need is used as a salt to anneal the two oligomers together. Add the 11mer and 25mer together in 250 mM Tris pH 7.7 to a final concentration of 20 pmole/µl each. Heat to 95°C for 5 min. Transfer to 70°C and gradually cool over a period of ~ 1 hour to room temperature. Leave at room temperature for ~1 hour then gradually cool over a period of ~1 hour to 4°C. Leave at 4°C for ~24 hours and then store at -20°C. I typically thaw and keep the linkers on ice when I use them.

2) 5x Mg-free Sequenase buffer:

200 mM Tris, pH 7.7 (at room temp.)

250 mM NaCl

3) Mg/DTT/dNTP soln.(make fresh, keep on ice):

20 mM $MgCl_2$

20 mM DTT

0.2 mM of each dNTP (Pharmacia)

4) 310 mM Tris, pH 7.7 (at room temp.)

5) Dilution soln. (make fresh, keep on ice):

17.5 mM $MgCl_2$

42.3 mM DTT

125 µg/ml BSA

6) Ligation mix (Make fresh, keep ice-cold)

Final concentrations in ligation mix are:

10 mM $MgCl_2$

20 mM DTT

3 mM rATP (Pharmacia)

50 µg/ml BSA

50 mM Tris pH 7.7 (at room temp.)

100 pmoles linker per reaction

3 Weiss units of T4 DNA ligase per reaction (Promega and Pharmacia have been tried and both work fine)

Make the ligation mix as a cocktail. First mix the Mg, DTT, rATP, BSA, and any needed water together. Chill this on ice. Then add 5µl of the ice-cold linker soln. per reaction. Recall that the linkers are in 250 mM Tris. Finally, just before use, add the ligase.

7) 5x Taq buffer (store at -20°C):

200 mM NaCl
50 mM Tris pH 8.9 (at room temp.)
25 mM $MgCl_2$
0.05% (w/v) gelatin

8) Labeling mix (make fresh, keep on ice)

Final concentrations in labeling mix are:

40 mM NaCl
10 mM Tris pH 8.9 (at room temp.)
5 mM $MgCl_2$
0.01% (w/v) gelatin
2 mM for each dNTP
1-10 pmoles end-labeled primer 3 per reaction, (see comment section)
2.5 units of Taq polymerase per reaction (add immediately before use)

Make labeling mix as a cocktail. Dilute 5x Taq buffer with H_2O, add dNTPs,and primer. Chill on ice. Immediately before use, add 2.5 units Taq polymerase per reaction. Final volume needed per reaction is 5 µl.

9) Taq stop soln.

260 mM NaOAc
10 mM Tris pH 7.5
4 mM EDTA

Procedure:

(a) Sequenase reaction.

Mix:

3 µg cleaved DNA (Because we are using mouse genomic DNA, this is equivalent to 4.5 x 10^5 nuclei)

0.3 pmole primer 1, (see comment section)

3 µl 5x Mg-free Sequenase buffer

adjust final volume to 15 µl

I use 1.5 ml siliconized tubes. In addition, to prevent tubes from "popping" during the various heating steps, I use Cap-Locks (Intermountain Scientific).

Heat at 95°C for 2-4 min.

Transfer to 60°C for 30 min.

Transfer to ice, quick spin at 4°C (to remove condensation), keep on ice

Add 7.5 µl of ice-cold Mg/DTT/dNTP soln.

mix by pipet, keep chilled

Add 1.5 µl of ice-cold, freshly diluted Sequenase (1 part Sequenase:3 parts ice-cold TE)

mix by pipet, keep chilled

Transfer to 47°C for 5 min.

Transfer to 60°C for ~5 min.

Add 6 µl of room temp.310 mM Tris pH 7.7, mix by pipet, and immediately

Transfer to 67°C for 10 min.

Transfer to ice, quick spin at 4°C (to remove condensation), keep on ice

(b) Ligation Reaction.

Add 20 µl of ice-cold dilution soln.
mix by pipet, keep chilled

Add 25 µl of ice-cold ligation mix
Mix by pipet, keep chilled.

Transfer to 15-20°C for overnight (see comment section on ligation)
Heat kill ligase by transferring to 70°C for 10 min.
Quick spin to remove condensation
Add 8.4 µl 3 M NaOAc and 1 µl of 10 mg/ml tRNA, mix
Add 220 µl EtOH, mix
Transfer to -20°C > 2 hours

(c) Taq Reactions (for sequencing or footprinting)

Spin samples 10-15 min. at 4°C
Pellet will be somewhat large because of co-precipitated proteins—don't worry
Wash pellet with 75% EtOH
remove last traces of EtOH with pipet, bring up pellet in 70 µl H_2O
Pellet should go into soln. in mins.

add 20 µl 5x Taq buffer
add 20 nmoles of each dNTP (2 µl of 10 mM soln.)
add 10 pmoles primer 2, (see comment section)
add 10 pmoles linker primer (the 25mer from the linker) (this may be unnecessary)
(The above can be added as a freshly-made, 30 µl per reaction cocktail)
Final volume should be 100 µl, mix by vortex

Cover with ~90 µl mineral oil
add 5 units Taq Polymerase to aqueous phase, mix, and transfer to ice
(When working with multiple samples, I routinely dilute the Taq polymerase 1:4 in 1X Taq buffer so that the final concentration of Taq

polymerase is 1 unit per µl. I then add 5 µl of diluted Taq polymerase per reaction. Adjust the other volumes accordingly so that the final volume is still 100 µl.)

Quick spin to separate phases

Cycle ~16 times

denature 1 min. at 94-95°C (first denaturing is for 2-4 min.)
hybridize 2 min. at 66°C (see comment section on T_m of primers)
extend 3 min. at 76°C, add 5 sec. per cycle. (For the last extension, allow 10 minutes).

I have done this manually in water-filled temp-blocks. Theoretically, this could be done in an automated thermal cycling machine, but be careful that all the wells have uniform tempertures.

Transfer to ice
Add 5 µl of ice-cold labeling cocktail, mix, and keep on ice
Quick spin to separate phases

Transfer to 94-95°C for 2-4 min.
Transfer to 69°C for 2 min. (see comment section on T_m of primers)
Transfer to 76°C for 10 min.
Transfer to ice

Add 295 µl Taq stop soln.
Add 10 µg tRNA (Can be freshly added to stop soln. and added together)

Phenol/Sevag extract to remove mineral oil. Phenol/sevag is: 40-50 parts phenol [equibrated in 150 mM NaCl, 50 mM Tris pH 7.5, 1 mM EDTA], 50-60 parts chloroform, and 1 part iso-amyl alcohol.
Precipitate aqueous with 2.5 volumes of EtOH
-20°C for > 2 hours
Spin, wash with 75% EtOH, bring pellet up in loading dye (80% formamide, 0.5x TBE, and dye markers BØB and X.C.), denature, and load on 6% sequencing gel.

The results are usually better if the sample is run on a thicker sequencing gel, e.g., ~0.56 mm instead of the usual 0.2 mm. In order to read 200-300 bases,

you will probably need to do a double-loading. Just split the sample before precipitating it.

Comments:

1) T_m of the Primers. Background can be kept to a minimum if the reactions are run under temperature conditions that provide as much specificity as possible. For instance, I run the Sequenase extension at 45 to 47°C so that the formation of imperfect duplexes between the genomic DNA and the primer are minimized. (It would be better to use an even higher temperature except that Sequenase does not function very well much above 50°C.) Any primer that is still stable at this temperature in Sequenase buffer could be used. I use a 25mer of ~50% GC content that has a T_m of ~60°C in this buffer. For this reaction, this might not be the best primer condition. I think that something with a lower T_m, like a 17mer of the same GC content (T_m ~50°C), might work better because it could reduce background. I have not yet tested this idea. The amplification primer (primer 2) should have an equal or higher T_m than the Sequenase primer, and should match the T_m of the linker primer. The conditions I suggest work well for a 25mer of ~60% GC content. The labeling primer (primer 3) should have an even higher T_m. During the labeling extension, the linker primer does not need to extend, and the labeling primer (#3) can better compete with the amplification primer (#2) for occupancy if it has a higher T_m. Recall that the amplification primer and the labeling primer must overlap or the labeling primer is excluded from binding. The conditions I suggest work well for a 27mer of ~63% GC content. I do not wish to leave you with the idea that the primers used must be like what I've suggested here. The exact GC content or length of the primers is probably not that important. What is important is their relationship to each other and to the temperature of the reactions. (Of course, they must be specific to the gene of interest. For example, genome size might influence the length of the primers you choose.) If you do use primers of different T_ms, adjust the hybridization conditions accordingly. I use the following formula

to calculate the Tm of the primers: 81.5 + 16.6(log M) + 0.41(%GC) - 500/n. (n = length of primer, M=molarity of salt in buffer. For 1x Taq buffer assume 40 mM NaCl and 0.66(10 mM Tris) = 47 mM salt.) Hybridizations work best if done ~ 2°C above this calculated T_m. You may wish to determine this empirically for each of your primers. I gel-purify my primers.

2) Sequenase reaction. Sequenase is used for the initial extension for two reasons. First, it is a highly processive polymerase, and second, Sequenase buffer can be easily modified to T4 DNA ligation buffer. The extension is done at 45 to 47°C instead of 37°C in order to minimize background. In addition, the higher temperature seems to help the Sequenase get through certain end structures. Other enzmes could be used for this purpose. I have try only Sequenase version 1.0.

3) Heat kill of Sequenase. If the Sequenase is not destroyed, it may extend the short oligomer of the linker. Although 60°C seems to be sufficient to kill the Sequenase, I routinely increase the heat kill temperature to 67°C as an extra precaution. The Tris that the ligase will need is added before this higher temperature heat kill so that the shorter extension products do not become denatured. (The Tris will act as a salt to increase the T_ms of the extension products.)

4) Ligation conditions. Most of the sequencing products appear to be ligated after ~45 min. at 15°C. Some regions seem to require longer periods of time to reach full intensity (for example, long runs of guanines), so I suggest an overnight ligation. The ligation can be done at room temperature (which would require less time), but this is close to the calculated T_m of the linker.

5)Taq conditions. I've used only Taq polymerase from Cetus. I have found that NaCl works about as well as KCl except when the region being amplified contains long tracks of G's, in which case NaCl is clearly superior. I do not recommend the buffer that New England Biolabs suggests because the salt they use (NH_4SO_4), is difficult to remove before loading the samples on the gel and therefore the bands are not as well resolved. I've

tested several pH and $MgCl_2$ levels and found that those higher than Cetus recommends work best for this procedure. I've also found that temperatures higher than those usually used in PCR reactions can reduce the background without losing the signal in this procedure.

6) **Rounds of Amplification.** I do only ~16 rounds of amplification. This seems like a good compromise number. Additional rounds of amplification would require more primer 2. (For 16 rounds, 10 pmoles are empirically in excess.) Additional primer 2 decreases the efficiency of the labeling step because, primer 3 and 2 partially compete for the same binding site. Furthermore, the polymerase starts to die as the cycling continues. For direct comparison of two sequence lanes (as in footprint analysis), lane-to-lane variability that is due to polymerase death could be a problem. 16 rounds of amplification do not show variability. For cloning or simple sequencing experiments, more rounds could be done, but additional primer should be used during the amplification. (This has not yet been tested.)

7) **Labeling step.** The labeling can be done with primer 2 during the amplification cycles. Background is higher if this is done, but for simple sequencing this might not be a problem. For footprint analysis or high quality sequencing, I recommend using a third primer. At the time primer 3 is added, additional polymerase and dNTPs are added. This increases the signal several-fold. (I suspect that the polymerase and dNTPs are beginning to break down from all of the cycling at high temperatures.) If primer 3 has a higher Tm than primer 2, the hybridization can be done under conditions that favor primer 3. This also increases the signal.

8) **Efficiency.** I have not yet determined the efficiency at each step in this procedure. I can, however, make a practical comparison to the method used in Mueller et al. (1988). In those experiments, genomic sequencing was done by one round of primer extension, on a multi-copy gene, using AMV reverse transcriptase. Theoretically, if each step in the PCR sequencing (Sequenase extension, ligation, 16 cycles of amplification, and labeling) was as

efficient as the RT extension, then its signal should be ~60,000 times stronger. In practice, I find the signal only ~15,000 times stronger. Therefore, the signal is ~25% of its expected maximum. This, of course, ignores the differences between the polymerases and the regions being sequenced, so it is a very soft number.

9) Background problems It is very important that the DMS and the piperidine used in the reactions be fresh. Old reagents give high levels of background. In addition, all of the piperidine must be removed from the DNA before any enzymatic step (Saluz and Jost 1987). This can be done by using a speed-vac and multiple precipitations. A low signal-to-noise level can sometimes be improved by increasing the denaturation temperature during the Taq reaction to 95°C. Changing primers can also improve results.

References

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.

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

Pfeifer, G. P., S. D. Steigerwald, P. R. Mueller, B. Wold, A. D. Riggs. 1989. Genomic sequencing and methylation analysis by ligation mediated PCR. *Science* **246**: 810-813.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.

Saluz, H.P., and J. P. Jost. 1987. A Laboratory guide to genomic Sequencing: the direct sequencing of native uncloned DNA. Birkhauser.

White, T. J., N. Arnheim, and H. A. Erlich. 1989. The polymerase chain reaction. *Trends Gent.* **5**: 185-189.

Appendix II

Genomic Sequencing and Methylation Analysis by Ligation Mediated PCR

This appendix is included because it discusses some of the statistical limitations of ligation mediated PCR as it is applied to genomic sequencing and footprinting.

It was published in *Science*, **246**: 810-813, 1989.

Genomic Sequencing and Methylation Analysis by Ligation Mediated PCR

GERD P. PFEIFER, SABINE D. STEIGERWALD, PAUL R. MUELLER, BARBARA WOLD, ARTHUR D. RIGGS*

Genomic sequencing permits studies of in vivo DNA methylation and protein-DNA interactions, but its use has been limited because of the complexity of the mammalian genome. A newly developed genomic sequencing procedure in which a ligation mediated polymerase chain reaction (PCR) is used generates high quality, reproducible sequence ladders starting with only 1 microgram of uncloned mammalian DNA per reaction. Different sequence ladders can be created simultaneously by inclusion of multiple primers and visualized separately by rehybridization. Relatively little radioactivity is needed for hybridization and exposure times are short. Methylation patterns in genomic DNA are readily detectable; for example, 17 CpG dinucleotides in the 5′ region of human X-linked PGK-1 (phosphoglycerate kinase 1) were found to be methylated on an inactive human X chromosome, but unmethylated on an active X chromosome.

METHYLATION OF CPG DINUCLEOTIDES in critical regions of many vertebrate genes may be part of a gene silencing mechanism involved in cell differentiation, X chromosome inactivation, and genomic imprinting (*1, 2*). Methylation-sensitive restriction endonucleases are commonly used to determine in vivo methylation patterns, but this limits the analysis to a small subset of all CpG dinucleotides. Another method for methylation analysis is genomic sequencing (*3*), a method that retains information normally lost during cloning, such as the location of 5-methylcytosines (*3*) and DNA-protein interactions (*4*). Genomic sequencing has, however, been difficult, requiring large amounts of radioactivity and long autoradiographic exposures (*5*). Primer extension has been used to simplify genomic sequencing, but these procedures still require the special preparation of primers labeled to extremely high specific activity and up to 50 μg of DNA per sequencing lane (*6*).

We now describe a genomic sequencing method in which we use a ligation mediated polymerase chain reaction (PCR) procedure [see figure 1 in (*7*)]. Briefly, step 1 of our genomic sequencing procedure is base-specific chemical cleavage of DNA samples (*8*) at either G, G + A, T + C, or C (*9*), generating 5′ phosphorylated molecules. Step 2 is gene-specific primer extension of an oligonucleotide (primer 1) by a DNA polymerase to give molecules that have a blunt end on the side opposite the primer (*10*). Step 3 is the ligation of an unphosphorylated linker to the blunt ends (*11*). Step 4 is the exponential amplification of the linker-ligated fragments with the use of the longer oligonucleotide of the linker (as a linker-primer) and a second gene-specific primer (primer 2) in a PCR reaction (*12*). After undergoing 15 to 18 amplification cycles, the DNA fragments are separated on a sequencing gel, transferred by electroblotting to nylon membranes (*13*), and hybridized with a single-stranded gene-specific probe (*14*). This procedure works well for all bases, sensitivity is improved, and the background is minimized by the transfer and hybridization steps. Moreover, several different sequences can be analyzed in a single experiment by rehybridization of the membrane.

The human X-linked phosphoglycerate kinase (PGK-1) gene is a housekeeping gene that is subject to X inactivation. The 5′ region is a CpG-rich island (*15*), but, unlike most autosomal CpG islands that are characteristically unmethylated, the Hpa II sites in the region shown (Fig. 1) are methylated on the inactive X chromosome (*16, 17*).

In an experiment with HeLa cell DNA, two different primer sets (Fig. 1, D and E) were included simultaneously in the primer extension and amplification reactions. The sequence defined by primer set D was visualized first (Fig. 2A) by hybridization with an Eco RI–Dde I hybridization probe. After stripping of the first probe from the membrane and rehybridization with an Xma III–

G. P. Pfeifer, S. D. Steigerwald, A. D. Riggs, Molecular Biology Section, Beckman Research Institute of the City of Hope, Duarte, CA 91010.
P. R. Mueller and B. Wold, California Institute of Technology, Division of Biology, Pasadena, CA 91125.

*To whom correspondence should be addressed.

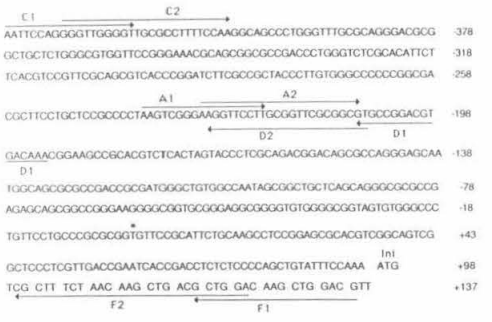


Fig. 1. Sequence of the CpG island containing the promoter and first exon of the human PGK-1 gene. Arrows indicate the oligonucleotide primer positions and the direction of reading. The asterisk indicates the major transcription start site. "Ini" marks the translation start point.

Pvu II probe, the sequence defined by primer set F became visible (Fig. 2B). As shown, both sequences could be read unambiguously; the correct calls would be made for a total of 336 nucleotides, even if the sequence were not known.

5-Methylcytosine appears as a gap in the cytosine ladder of a Maxam-Gilbert sequencing gel. To test the procedure as an assay for methylated cytosines, we studied human-hamster hybrid cell lines that contained either an inactive or an active human X chromosome (*17*). Primer set A (Fig. 1) allowed reading of the sequence toward the transcription start site, and the data indicate that the human inactive X chromosome is methylated at all 17 CpG dinucleotides in the investigated region (Fig. 3, lane X_i), whereas the active X_i chromosome is unmethylated at the same 17 sites (Fig. 3, lane X_a). By comparison with cloned DNA, we found that HeLa DNA has no detectable methylation at 52 CpG dinucleotides analyzed in the PGK-1–associated CpG-rich island, a result consistent with this cell line not having an inactive X chromosome.

Although the band patterns shown in Figs. 2 and 3 are reproducible, band intensities vary from fragment to fragment within a single lane; a few bands are weak or even missing on short exposures. For example, three bands in a region of about 100 nucleotides are not seen in the sequence shown in Fig. 3. Band intensity may vary because (i) the chemical reactivity of individual bases may differ; (ii) small effects of sequence on amplification efficiency may play some role, although PCR can be very quantitative and reproducible [±4% SD through 25 cycles (*18*)]; (iii) certain sequences may not be extended to a blunt end as efficiently in primer extension prior to ligation; (iv) the ligation step may introduce some bias. Overexposed autoradiograms can often be used to obtain sequence information even from weak bands. This is the case for most of the sites that we have studied, such as the CpG site at nucleotide position −119. The intensities of only the weak bands were increased by increasing the time for ligation or the temperature of the first primer extension reaction (*19*). Thus, the steps before the PCR may be very efficient for strong bands.

Exponential amplification works so well that band detection probably no longer limits its sensitivity. Comparing band intensities with the standard genomic sequencing method applied to the same PGK-1 sequence studied with primer set D, we estimate an amplification of 3000- to 6000-fold after 15 PCR cycles (10 to 20% overall efficiency). One microgram of HeLa cell DNA per reaction gave a strong signal with short exposures (2 hours, Fig. 2B). However, a statistical limit should be considered. The lowest limit of the DNA needed for sequence analysis is a single molecule for each band in the sequence ladder. For optimum efficiency of molecule usage, there should be one chemical cleavage per length of sequence to be analyzed, for example one cut per 200 nucleotides (nt). To avoid band intensity variation of more than 10% standard deviation that is due only to statistical sampling fluctuation, about 100 founder molecules are needed for each band. If ligation and first primer extension for an unfavorable fragment is only about 10% efficient, then 2×10^5 molecules ($200 \times 100 \times 10$) should give a band of invariant intensity. This then would be a statistical limit for experiments where visualization of this band is important. Efforts at further increasing the sensitivity of the method should focus on weak bands and on the steps that precede PCR, such as ligation and first primer extension. Increasing ligation time was helpful; however, attempts to increase

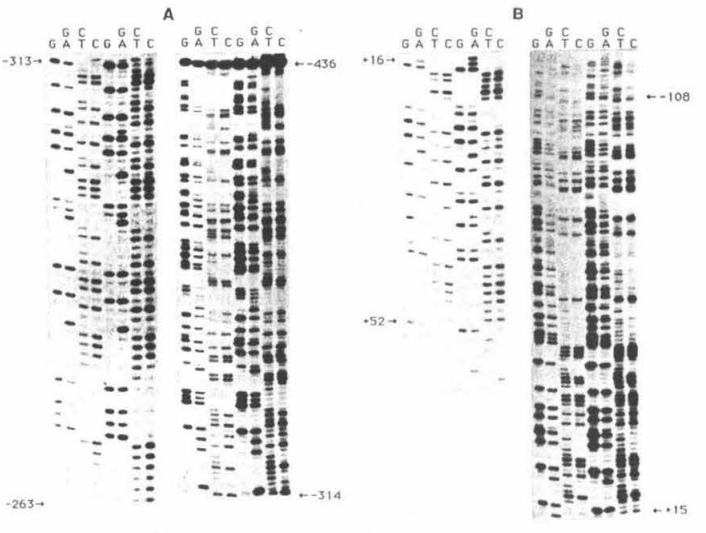


Fig. 2. Genomic sequence data for the 5′ end of PGK-1 in HeLa DNA. Primer sets D and F were used and two sequences were simultaneously amplified from 1 μg of DNA per base-specific reaction. The amplified DNA was then split for gel analysis with the four rightmost lanes of each panel receiving 2.5 times more DNA than the first four left lanes in order to visualize the weakest bands. (**A**) A single-stranded Eco RI–Dde I fragment was used as a hybridization probe to visualize the sequence defined by primer set D. (**B**) Rehybridization of the nylon membrane with a single-stranded Xma III–Pvu II fragment allowed visualization of the sequence defined by primer set F. The strong signal in all lanes at position −436 of A corresponds to the Eco RI site at which the DNA had been cut. The sequence corresponds to the sequence in (*15*), except at position −430, where there is an extra A.

ligation efficiency by adding polyethylene glycol (*20*) or hexamine cobalt (*21*) to the ligation mixtures were unsuccessful. The optimum temperature and Mg^{2+} concentrations for PCR depend on the primer and the gene region to be studied. Optimization of conditions was more critical for certain primers than for others. We obtained good sequence ladders for each of the four primer sets in Fig. 1, even though the region is more than 70% G+C.

Instead of directly labeling the fragments of the sequence ladder (*7*), we transferred to nylon membranes and, to visualize the sequence, hybridized with a single-stranded probe located 3′ to the gene-specific primers. Transfer to membranes makes it possible to use relatively low levels of radioactivity, results in minimal radiation exposure to laboratory workers, and permits analysis of multiple sequences by rehybridization of the membrane (Fig. 2, A and B). We do not yet know what the limit is for simultaneous ladder formation and sequential visualization, but a membrane containing a sequence ladder can be rehybridized up to 50 times (*22*).

By improving the quality of the data and by decreasing the amount of DNA needed, genomic sequencing by ligation mediated PCR should aid in vivo footprinting studies (*7*) and studies on DNA cytosine methylation. The procedure also appears adequate for sequencing unknown DNA regions adjacent to any known region by primer extension into the unknown region. Step by step genomic sequencing, in which the primers are selected from the previous sequence determination, would avoid any cloning procedure.

Most methylation studies to date have used DNA 5-methylcytosine–sensitive restriction endonucleases such as Hpa II. Methylation analysis with PCR after Hpa II cleavage is more sensitive than genomic sequencing, needing only a few molecules (*18*), but the Hpa II–PCR procedure assays only about 6% of all methylated sites. Genomic sequencing allows analysis of all sites, and we found that the inactive human X chromosome in Chinese hamster–human hybrid cells faithfully retains a high level methylation at 17 CpG sites in the PGK-1 upstream region, whereas an active X chromosome is unmethylated at these sites. It has been suggested that more than 50% of methylated cytosines are not at CpG sites (*23*), and in vitro DNA methyltransferase will occasionally methylate cytosine at other than CpG sites (*24*). It is therefore of interest that we have not seen methylated cytosines in dinucleotide sequences other than CpG, even in the heavily methylated inactive X chromosome. As additional information obtained by genomic sequencing accumulates, this question about the in vivo specificity of DNA methyltransferase should be answered.

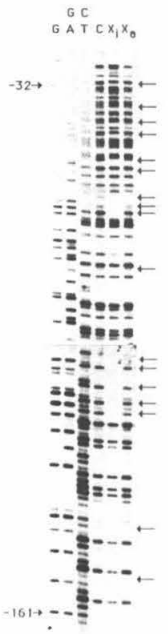


Fig. 3. Genomic sequence data for HeLa cell DNA and Chinese hamster–human hybrid cell DNA. The lanes labeled G, G+A, T+C, and C show results obtained with 2 μg of HeLa DNA per lane. Oligonucleotides A1 and A2 were used as primers and a single-stranded Mae III–Apa I fragment was the hybridization probe. From the bottom of the gel to the top, the sequence reads the lower strand 3′ to 5′. Also shown are C lanes for DNA (2 μg) from hamster-human hybrid cells containing either an active (X_a) or an inactive (X_i) human X chromosome. Methylated cytosines, which appear as a gap in the sequence ladder, are indicated by arrows.

REFERENCES AND NOTES

1. For reviews on DNA methylation, see A. Razin and A. D. Riggs, *Science* **210**, 604 (1980); A. D. Riggs and P. A. Jones, *Adv. Cancer Res.* **40**, 1 (1983); W. Doerfler, *Annu. Rev. Biochem.* **52**, 93 (1983); R. Holliday, *Science* **238**, 163 (1987); H. Cedar, *Cell* **53**, 3 (1988); S. Grünwald and G. P. Pfeifer, *Prog. Clin. Biochem. Med.* **9**, 61 (1989).
2. For reviews on genomic imprinting, see M. Monk, *Genes Dev.* **2**, 921 (1988); D. Solter, *Annu. Rev. Genet.* **22**, 127 (1988); W. Reik and M. A. Surani, *Nature* **338**, 112 (1989).
3. G. M. Church and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1991 (1984).
4. A. Ephrussi, G. M. Church, S. Tonegawa, W. Gilbert, *Science* **227**, 134 (1985).
5. H. P. Saluz and J. P. Jost, *A Laboratory Guide to Genomic Sequencing* (Birkhauser, Boston, 1987).
6. ———, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2602 (1989); M. M. Becker, Z. Wang, G. Grossmann, K. A. Becherer, *ibid.*, p. 5315.
7. P. R. Mueller and B. Wold, *Science*, **780** (1989).
8. DNA from human-hamster hybrid cell lines that contained either an inactive (X8-6T2) or an active (Y162-11C) human X chromosome (*17*) was provided by R. S. Hansen and S. M. Gartler. Nuclei were isolated from HeLa (S315) cells by published procedures [J. Wijnholds, J. N. J. Philipsen, G. Ab, *EMBO J.* **7**, 2757 (1988)], and DNA was purified from the isolated nuclei as described (*5*). DNA's were usually cleaved with Eco RI to reduce viscosity.
9. Base-specific cleavage: DNA, usually a 10-μg sample, was subjected to the Maxam-Gilbert reactions [A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980)] as modified by Saluz and Jost (*5*). After piperidine cleavage, DNA was precipitated with 2.5 volumes of ethanol, washed twice with 80% ethanol, and dissolved in 200 μl of water. Traces of piperidine were removed by vacuum drying the sample overnight. The DNA pellet was dissolved in water (1 μg/μl).
10. First primer extension: 1 to 2 μg of chemically cleaved DNA was mixed (final volume, 15 μl) with 0.6 pmol of a gene-specific primer (A1, D1, or F1, in Fig. 1) in 40 m*M* tris-HCl (*p*H 7.7) and 50 m*M* NaCl, heated at 95°C for 3 min, and then at 45°C for 30 min. The solution was adjusted to 25 m*M* tris-HCl (*p*H 7.7), 30 m*M* NaCl, 6 m*M* $MgCl_2$, 6 m*M* dithiothreitol, 80 μ*M* dNTP's, and a final volume of 24 μl. Sequenase™ (4.5 units, USB) was added, and the reaction mixture was incubated at 45°C for 15 min. The reaction was stopped by adding 6 μl of ice-cold 310 m*M* tris-HCl (*p*H 7.7), and the enzyme was inactivated at 67°C for 15 min. The ligation step (*11*) followed without further treatment.
11. Ligation: The structure of the linker and the annealing and ligation conditions were as described in (*7*).
12. PCR amplification: Primer extended, ligated molecules were amplified with Taq polymerase (Amplitaq™, Perkin-Elmer Cetus) in the presence of the longer oligonucleotide of the linker and a gene-specific primer (A2, D2, or F2; Fig. 1). The fragments were amplified in 100 μl of 10 m*M* tris-HCl (*p*H 8.9), 40 m*M* NaCl, 2 m*M* $MgCl_2$, gelatin (0.01%), 0.25 m*M* dNTP's, 10 pmol of each primer, with 3 units Taq polymerase, according to the Perkin-Elmer Cetus protocol. Conditions for the 15 to 18 thermal cycles were 1 min at 95°C, 2 min at 65°C, and 3 min at 75°C. Fresh enzyme (1 unit) was added and incubation was continued for 10 min at 74°C; EDTA and sodium acetate (*p*H 5.2) were added to 20 m*M* and 300 m*M*, respectively, and the reaction mixture was extracted with phenol and $CHCl_3$, and subsequently precipitated with ethanol and centrifuged. The DNA was then dissolved in electrophoresis sample dye (*5*), and, just before being placed on the gel, it was heated to 95°C for 2 min.
13. Gel electrophoresis and electroblotting: DNA fragments were separated on a sequencing gel (95 by 0.08 or 0.04 cm) consisting of 8% polyacrylamide (acrylamide-bisacrylamide, 29:1) and 7*M* urea; and then transferred to GeneScreen nylon membrane by electroblotting (*5*). A simple transfer apparatus was constructed with porous steel plates of a Bio-Rad gel dryer as electrodes (Bio-Rad 200/2.0 power supply). The details for this and other procedures are available on request. After the transfer, the membranes were air-dried, baked at 80°C for 20 min in a vacuum oven, then subjected to ultraviolet crosslinking [P. B. Becker, S. Ruppert, G. Schütz, *Cell* **51**, 435 (1987)].
14. Hybridization: To make probes the cDNA procedure [F. Weih, A. F. Stewart, G. Schütz, *Nucleic Acids Res.* **16**, 1628 (1988)] was used. Small (100 to 200 bp) restriction fragments from the 5′ region of PGK-1 (see below) were subcloned into Bluescript SK+ (Stratagene), and then RNA was produced from the inserts on a preparative scale; either T3 or T7 RNA polymerase was used. Labeled cDNA was made from the RNA (0.5 μg) by reverse transcriptase with the use of T7 or T3 primers and 100 μCi of [^{32}P]dCTP (6000 Ci/mmol). After ethanol precipitation, the single-stranded cDNA probe (1×10^9 to 5×10^9 dpm/μg) was used directly for hybridization. Hybridization probes were made from three subclones that contained fragments Eco RI–Dde I (197 bp; positions −436 to −240); Mae III–Apa I (181 bp; −199 to −19); or Xma III–Pvu II (150 bp; −69 to +81). Hybridization and washing buffers were as described (*5*). Preliminary hybridization (25 ml of hybridization buffer, 10 min) and hybridization (50 μCi per gel in 5 ml of hybridization buffer, overnight) were performed in rolling cylinders at 68°C in a hybridization oven. Nylon filters were exposed for 2 to 7 hours to Kodak XAR-5 film with intensifying screens at −70°C. For rehybridization experiments, probes were stripped from the membrane by incubation in 0.2*M* NaOH at 45°C for 30 min.
15. J. Singer-Sam *et al.*, *Gene* **32**, 409 (1984).
16. D. H. Keith, J. Singer-Sam, A. D. Riggs, *Mol. Cell. Biol.* **6**, 4122 (1986).

17. R. S. Hansen, N. A. Ellis, S. M. Gartler, *ibid.* **8**, 4692 (1988).
18. J. Singer-Sam *et al.*, in *Nucleic Acid Methylation*, G. Clawson, D. Willis, A. Weissbach, P. A. Jones, Eds. (Liss, New York, in press); this is volume 128 of *UCLA Symposia on Molecular and Cellular Biology, New Series*.
19. G. P. Pfeifer, S. D. Steigerwald, P. R. Mueller, B. Wold, A. D. Riggs, unpublished results.
20. B. H. Pheiffer and S. B. Zimmerman, *Nucleic Acids Res.* **11**, 7853 (1983).
21. J. R. Rusche and P. Howard-Flanders, *ibid.* **13**, 1997 (1985).
22. G. M. Church and S. Kieffer-Higgens, *Science* **240**, 185 (1988).
23. D. M. Woodcock, P. J. Crowther, W. P. Diver, *Biochem. Biophys. Res. Commun.* **145**, 888 (1987).
24. D. Simon *et al.*, *Nature* **304**, 275 (1983); G. P. Pfeifer, E. Spiess, S. Grünwald, T. L. J. Boehm, D. Drahovsky, *EMBO J.* **4**, 2879 (1985).
25. Supported by National Institute of Aging grant AG08196 (A.D.R.), NIH grants RR07003 and GM355262BW (B.W.), and a fellowship from the Deutsche Forschungsgemeinschaft (Pf212/1-1) (G.P.P.).

11 July 1989; accepted 26 September 1989