Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 7392–7396, December 1984
Biochemistry

Duplicated heavy metal control sequences of the mouse metallothionein-I gene

(transcriptional regulation/repeated DNA sequences/metal-binding proteins)

ANTHONY D. CARTER, BARBARA K. FELBER, MARYJANE WALLING, MARIE-FRANCE JUBIER, CARL J. SCHMIDT, AND DEAN H. HAMER

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, MD 20205

Communicated by Maxine F. Singer, August 13, 1984

ABSTRACT **We present evidence that two distinct regions of the DNA upstream from the mouse metallothionein-I gene contain metal-responsive regulatory sites. This result was obtained by analyzing a systematic series of deletion, insertion, duplication, and clustered point mutations introduced into cultured cells on a simian virus 40 plasmid vector. The two upstream regions contain a duplicated evolutionarily conserved DNA sequence. While either upstream region is sufficient to confer heavy metal responsiveness, both are required to give maximal levels of induced transcription.**

The metallothioneins (MTs) are small cysteine-rich proteins that tightly bind heavy metals. They are present in a broad range of eukaryotes, from yeast to man, and are expressed in many different tissues and cell types. Exposure to heavy metals results in a rapid increase in MT mRNA levels and protein synthesis in both cultured cells and whole animals. This homeostatic regulatory mechanism plays a critical role in protecting cells against toxic ions, such as cadmium and mercury, and may also be involved in the metabolism of essential elements such as zinc and copper (reviewed in ref. 1). The metal response, which occurs in the presence of protein synthesis inhibitors, can be attributed largely if not exclusively to increased transcription rates (2–5).

What MT gene DNA sequences are involved in the rapid transcriptional response to heavy metals? We have shown previously that a cloned mouse MT-I gene, containing 2000 base pairs (bp) of 5′ flanking DNA, retains its ability to be induced by cadmium when introduced into cultured monkey kidney cells on simian virus 40 (SV40) viral and plasmid vectors (4). Appropriate regulation has also been observed for mouse and human MT and MT fusion genes carried on episomal bovine papilloma virus vectors (6, 7), introduced into cultured mouse or rat cells by transformation (5, 8), microinjected into mouse eggs (9), or integrated into the genome in transgenic mice (10). To more precisely localize and characterize the MT regulatory sequences, we have constructed a systematic series of 3′ deletions, 5′ deletions, internal deletions, duplications, insertions, and clustered point mutations in the mouse MT-I gene. These were introduced into cultured cells on an SV40 plasmid vector and analyzed both for their efficiency of transcription and for their ability to be induced by cadmium. We have also compared the 5′ flanking DNA sequences of the mouse MT-I gene with three functional human MT genes. Our results show that the metal-responsive elements lie in at least two distinct regions of the 5′ flanking DNA. These regions share a conserved nucleotide sequence that may serve as a recognition signal for cellular regulatory factors.

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

MATERIALS AND METHODS

Recombinant plasmids were constructed and characterized by standard methods (11). The starting pML-SV40-MT construct, JYMMT(E), was described previously (4). The *Escherichia coli* galactokinase–SV40 transcription unit was obtained from pSVK105Δ, a derivative of pSVK carrying a *Bam*HI site 50 bp upstream from the galactokinase initiation codon (12). Deletion (dl) mutants were isolated by nuclease BAL-31 digestion of appropriately restricted DNA followed by ligation to synthetic oligonucleotide linkers. The linker scanner (ls), duplication (dp), and insertion (in) mutants were constructed by recombining 5′ and 3′ deletion mutants at a common *Cla* I linker site (13). Clustered point (cp) mutants were obtained by insertion of synthetic oligonucleotides (Biologics, Toronto) into ls(−54/−34). Mutant DNA sequences were determined by the method of Maxam and Gilbert (14).

For transfection, monkey kidney CV1 cells were seeded into 100-mm culture dishes and grown in Dulbecco's modified Eagle's medium with 10% fetal calf serum until 50–60% confluent (approximately 2×10^6 cells per plate). Transfection was carried out by the calcium phosphate precipitation plus glycerol shock method (15), using 5–7 μg of recombinant plasmid DNA. The transfected cells were cultured for 20 hr, treated or not treated with 2 μM $CdCl_2$ for an additional 8 hr, and then harvested.

Cell extracts were prepared and assayed for galactokinase by starch gel electrophoresis (12). Enzyme levels were quantitated by laser densitometry of the autoradiogram at several exposure levels. RNA was extracted and analyzed by S1 nuclease mapping as described (4). The single-stranded DNA probe was obtained by primer extension of a phage M13 subclone carrying the mouse MT-I gene 5′ flanking and untranslated region. Autoradiograms were scanned in the densitometer to estimate the quantity of protected 68-base fragment corresponding to the major 5′ end of mouse MT-I mRNA. Several bands of higher molecular weight were protected by the transfected cell RNA and, to a lesser extent, by mouse cell RNA (results not shown). It is not clear whether these are due to read-through transcripts, artefacts of the S1 mapping procedure, or contamination of the RNA samples with DNA.

Each mutant was assayed in at least two, and in most cases four or five, separate transfection experiments. Because there was substantial variation (up to 10-fold) in the efficiency of transfection, results were normalized to give a value of 1 for the uninduced expression of an MTgal1 or MTgal2 control included in the same experiment. After this correction had been made the results varied by a maximum of 2-fold.

Abbreviations: MT, metallothionein; SV40, simian virus 40; bp, base pairs.

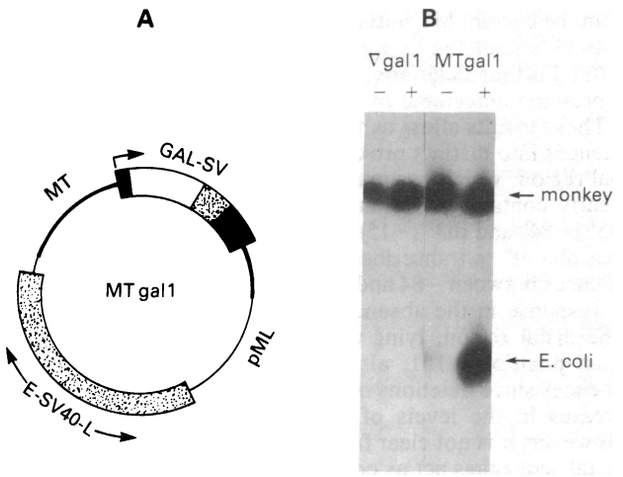

FIG. 1. Regulation of an MT-galactokinase fusion gene. (*A*) Structure of the recombinant plasmid MTgal1. Thick lines represent mouse MT-I flanking sequences, solid bars represent MT gene sequences, hollow bars represent *E. coli* galactokinase coding sequences, stippled bars represent SV40 sequences, and thin lines represent plasmid pML sequences. (*B*) Monkey kidney cells were transfected either with Δgal1 or with MTgal1 and cultured in the absence (−) or presence (+) of cadmium. Cell extracts were electrophoresed through a starch gel and galactokinase activity was detected by incubation with [^{14}C]galactose and ATP. The arrows point to the positions of the endogenous monkey galactokinase produced by the host cells and the *E. coli* galactokinase encoded by the fusion gene. The Δgal1 plasmid is identical to MTgal1 except that the MT gene 5′ flanking and untranslated sequences have been replaced with a *Cla* I linker.

RESULTS

Regulation of an MT-Galactokinase Fusion Gene. We have used gene fusion to an *E. coli* galactokinase–SV40 transcription unit (12) to facilitate our analysis of the mouse MT-I control sequences. As diagrammed in Fig. 1*A*, this *MT–gal* fusion gene consists of 2000 bp of 5′ flanking sequences and 68 bp of 5′ untranslated sequences from the MT gene, the initiation codon and coding sequences for the bacterial galactokinase, a fragment of SV40 DNA that contains splicing and polyadenylylation signals, and the remaining portion of the MT gene and 3′ flanking sequences. To generate the recombinant MTgal1, the fusion gene was inserted into a plasmid pML–SV40 vector (16) containing the origin of viral DNA replication, the early and late promoter and enhancer sequences, and an intact tumor (T)-antigen gene. Cultured CV1 monkey kidney cells were transfected with a calcium phosphate precipitate of this DNA, cultured in the presence or absence of cadmium, then harvested and analyzed for galactokinase activity and fusion mRNA.

Fig. 1*B* shows a typical starch gel separation of the bacterial and mammalian galactokinase produced in transfected monkey cells. Cells transfected with the *MT–gal* fusion gene clearly produced *E. coli* galactokinase, and the enzyme was induced 12-fold by cadmium. In contrast, cells transfected with the plasmid Δgal1, from which the MT promoter region has been removed, produced no detectable *E. coli* galactokinase. The 5′ end of the fusion mRNA was analyzed by S1 nuclease mapping. Fig. 4 shows that RNA from cells transfected with the fusion gene protected a major DNA fragment of 68 bases, the length expected for RNA starting at the normal initiation site of the mouse MT-I gene (4, 17). Furthermore, this RNA was strongly induced by cadmium. We conclude from these experiments that the *MT–gal* fusion gene is appropriately transcribed and is specifically inducible by a heavy metal.

Deletion Mutants. We first wished to determine whether any sequences within the MT gene itself are required for heavy metal regulation. For this purpose we constructed a series of 3′ deletion mutants in which the galactokinase gene is fused to the MT gene at positions ranging from +65 to −15 and tested them for galactokinase expression as described above. (MT gene sequences are numbered from position +1 at the transcription initiation site, negative numbers for 5′ flanking sequences and positive numbers for gene sequences.) As summarized in Fig. 2*A*, all of these mutants were strongly stimulated by cadmium, indicating that the necessary regulatory information is located in the upstream flanking region. Moreover, the similar galactokinase expres-

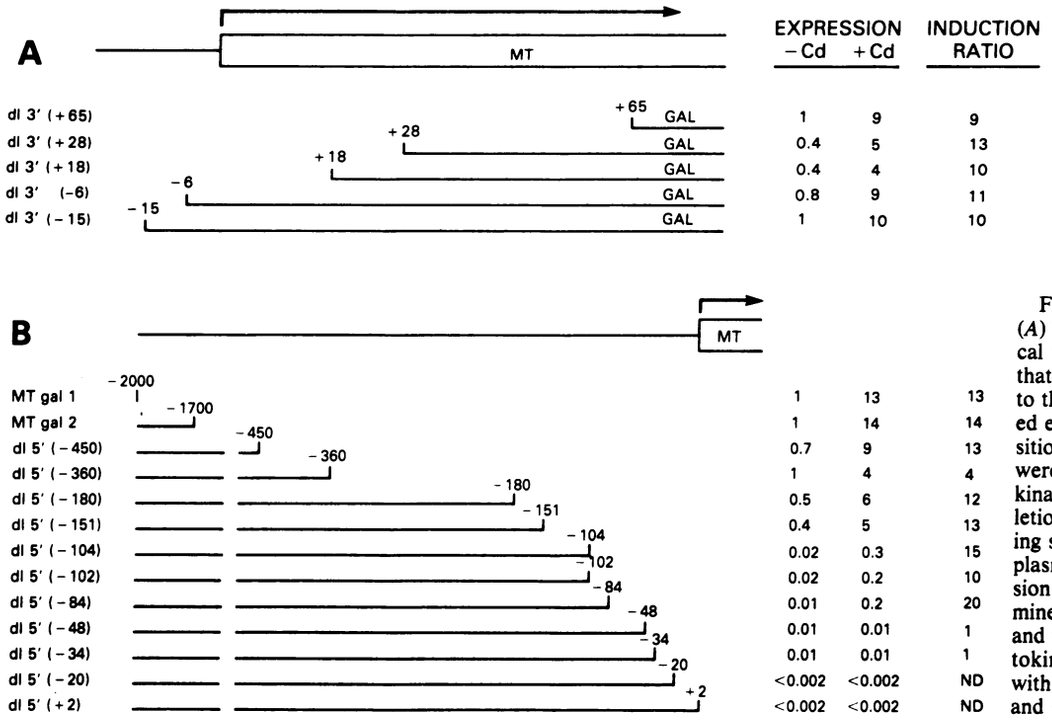

FIG. 2. Deletion mutants. (*A*) The 3′ deletions are identical to MTgal2 (see *B*) except that the *gal*–SV40 unit is fused to the MT gene at the indicated endpoint rather than at position +68. Expression levels were determined by galactokinase assays. (*B*) In the 5′ deletions the MT gene 5′ flanking sequences are replaced by plasmid pML DNA. Expression levels were initially determined by nuclease S1 mapping and were confirmed by galactokinase assays for mutants with endpoints between −151 and +2.

sion levels of all the mutants suggests that neither the exact length nor the sequence of the 5′ untranslated region is critical for translational efficiency.

The heavy metal control sequences were further localized by sequential deletions of the 5′ flanking DNA. These mutants were initially constructed in an intact MT gene and screened by S1 mapping of transfected cell RNA as shown in Fig. 4. Deletions in the region of most interest, between positions −151 to +2, were transferred into the Δgal1 vector and further analyzed for galactokinase expression. In general, there was good agreement between the values obtained by these two assay methods. As summarized in Fig. 2*B*, the 5′ deletion mutants fell into four classes:

(*i*) Deletion mutants of the 5′ flanking sequences between positions −1700 and −151 retained the ability to be induced by cadmium and had basal and induced expression levels within 30% of the starting gene.

(*ii*) Deletion mutants of the DNA between positions −151 and −84 also retained cadmium inducibility but had expression levels only 1–3% of the starting gene. The removal of nucleotides −151 to −104 caused the most significant decreases in both basal (20-fold) and induced (17-fold) transcription but did not alter the major transcription initiation site.

(*iii*) Mutants with deletions to positions −48 or −34 had readily detectable basal expression levels but were not induced by cadmium. Nuclease S1 mapping showed that the RNA from these mutants had the same 5′ end as the RNA from the starting gene. Thus, these mutants are transcribed from the correct MT initiation site but have lost the capability to be stimulated by a heavy metal.

(*iv*) Further deletions, to positions −20 or +2, gave no expression detectable by our assays.

These results allow us to divide the upstream flanking sequences into distinct proximal and distal regions. The proximal region, which has maximum boundaries of −84 to −15, clearly contains a cadmium-responsive element since both dl5′ (−84) and dl3′ (−15) are inducible. Moreover, the fact that dl(−48) is transcribed but not inducible shows that sequences between −84 and −48 are critical for the heavy metal response in the absence of further upstream sequences. The distal region, lying upstream from position −84 to at least position −151, also contains important control sequences since deletions of this region lead to substantial decreases in the levels of basal and induced transcription. However, it is not clear from these experiments whether the distal sequences act as constitutive or regulated control elements.

Fine Mapping Mutants. To resolve these possibilities, we analyzed a series of internal deletions, clustered point mutations, duplications, and insertions that independently alter the sequences of either the proximal or distal control regions. Fig. 3 summarizes the structures and expression levels of these mutants and Fig. 4 shows typical nuclease S1 analyses. The following results were obtained:

(*i*) Two internal deletions in the distal region, dl(−200/−106) and dl(−190/−88), resulted in reduced levels of basal and induced expression but normal induction ratios. The mu-

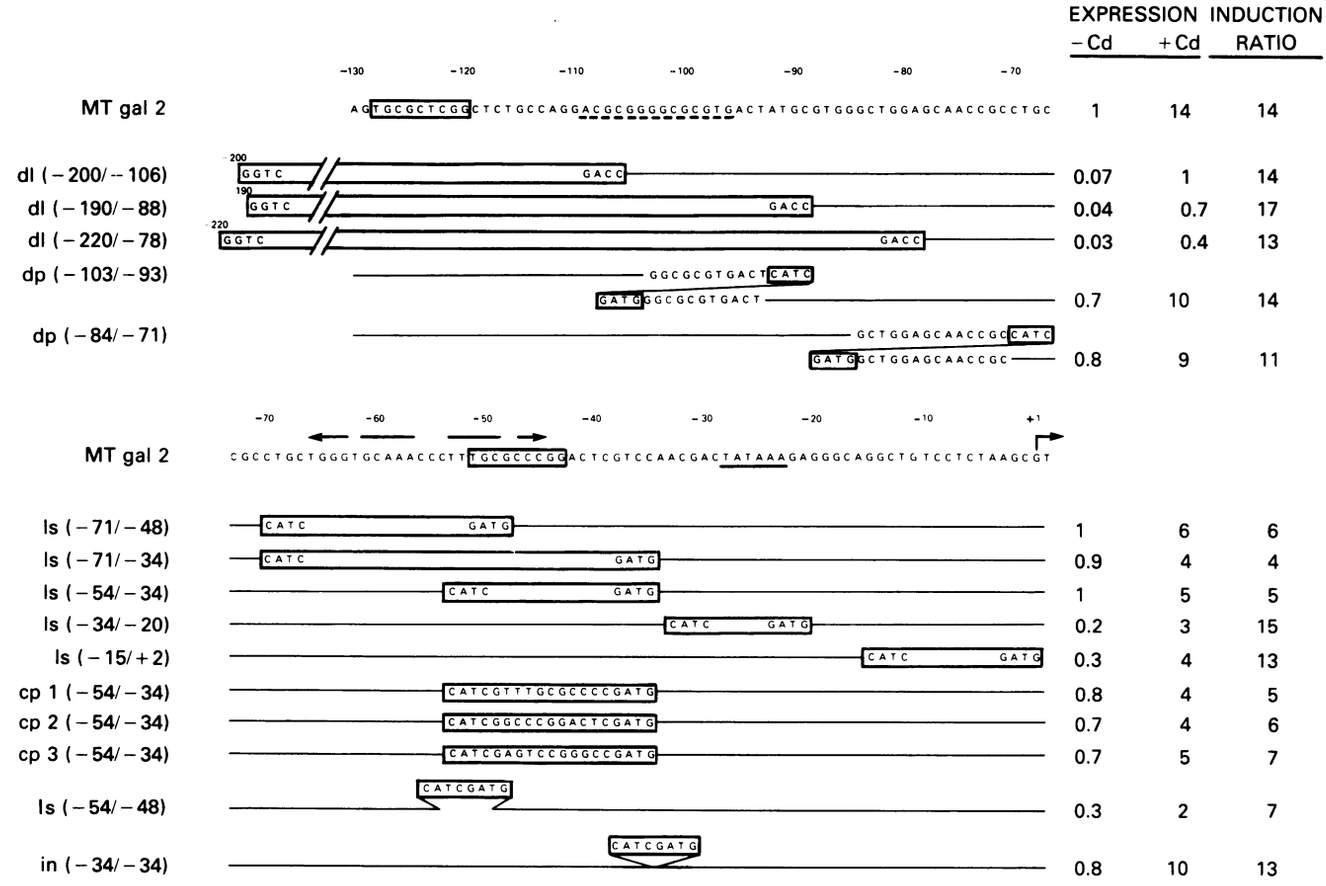

FIG. 3. Fine mapping mutants. The line marked MTgal2 shows the 5′ flanking sequence of the mouse MT-I gene determined by Glanville *et al.* (17) and confirmed in our laboratory. Shown below are the structures of the deletion (dl), linker-scanner (ls), clustered point (cp), duplication (dp), and insertion (in) mutants. The top part of the figure shows mutants in the distal region, the bottom part mutants in the proximal region. The altered sequences are enclosed in boxes; when no nucleotides are shown there is a deletion. These mutants are identical to MTgal2 (see Fig. 2*B*) except for the indicated alterations. Expression levels were determined by galactokinase assays and, in most cases, confirmed by nuclease S1 mapping.

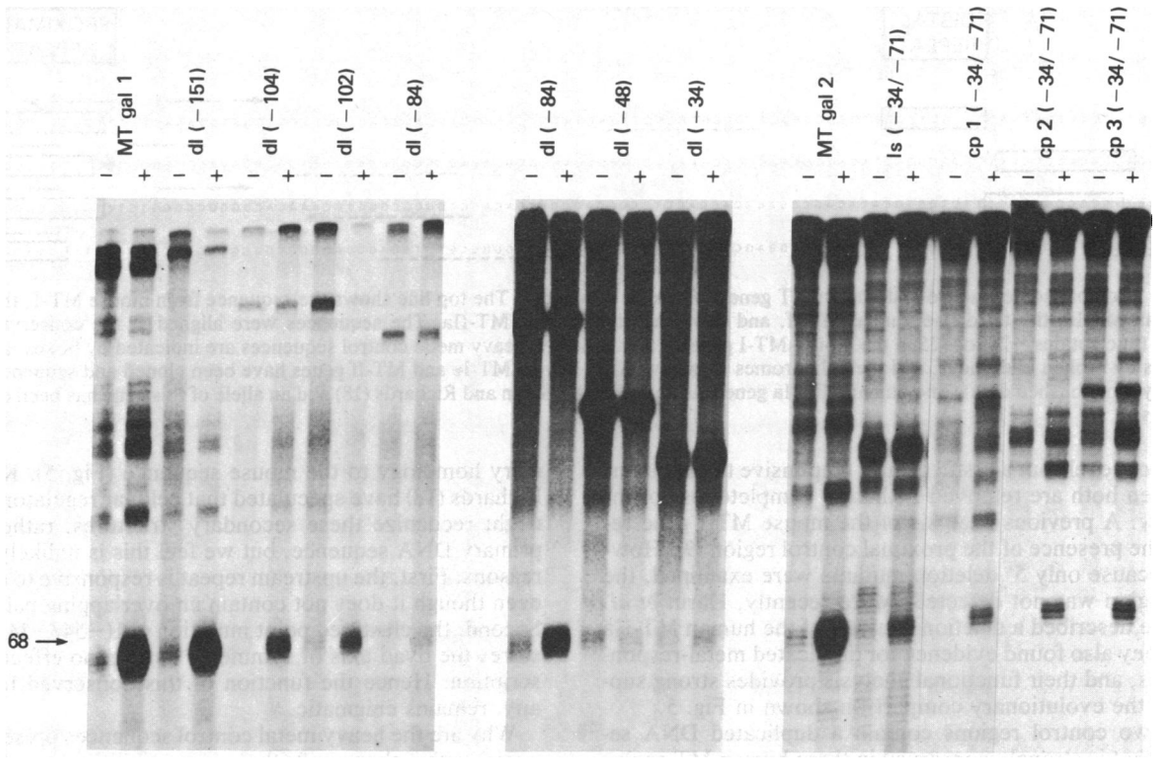

FIG. 4. Mapping the 5′ ends of transfected cell RNA. Cells were transfected with the indicated DNA and cultured in the absence (−) or presence (+) of cadmium. RNA was extracted and analyzed by S1 nuclease mapping as described (4). The single-stranded probe extends from the *Bgl* II site (position +68) to the *Sac* I site (position −151) and was uniformly labeled by primer extension of an M13 subclone. The band at 68 bases represents the major 5′ end of mouse MT-I mRNA (4, 17). The additional bands may represent read-through transcripts, artefacts of the S1 mapping procedure, or contamination of the RNA samples with DNA.

tant dl(−220/−78) had a similar phenotype, indicating that the upstream boundary of the proximal region actually lies at or 3′ to position −78.

(*ii*) The linker scanner mutants ls(−71/−48), ls(−71/−34), ls(−54/−34), and ls(−54/−48) alter or delete the proximal DNA sequences known to contain a metal-responsive element. Each of these mutants is expressed ½ to ¼ as efficiently as the starting gene in the presence of cadmium but, nevertheless, shows a clear 4- to 7-fold induction over the basal level. Nuclease S1 mapping of the RNA from the most extensive mutant, ls(−71/−34), confirmed its inducibility. The ability of these mutants to respond to cadmium indicates that there must be a second metal-responsive site located in the upstream sequences distal to position −71.

(*iii*) The relatively low levels of induced transcription of the linker scanner mutants described above could be due to either the altered DNA sequence in the proximal region or the altered spacing between the distal region and the gene. To distinguish between these possibilities we restored the normal spacing by inserting appropriate oligonucleotides into the mutant ls(−54/−34) to generate the clustered point mutations cp1(−54/−34), cp2(−54/−34), and cp3(−54/−34). Each of the resulting constructs had basal and induced transcription levels close to the level of the starting deletion mutant and encoded RNA with the normal 5′ end. Thus the alteration of the sequence, not the spacing, causes this behavior.

(*iv*) The effect of the spacing between the control elements was further tested by the construction of duplication and insertion mutants. Mutants dp(−103/−93) and dp(−84/−71) increase the distance between the proximal and distal regions by 11 to 14 bp, whereas in(−34/−34) inserts 8 bp between the upstream control regions and the "TATA box." Despite these alterations, all of the mutants had expression levels and induction ratios close to those of the starting gene.

(*v*) Linker scanner mutants that delete the TATA box, ls(−34/−20), or the 5′ end of the gene, ls(−15/+2), are expressed ⅓ to ⅕ as efficiently as the starting gene but are strongly induced by cadmium.

The Two Control Regions Contain a Conserved Duplicated DNA Sequence. The observation that both the proximal and distal upstream control regions respond to cadmium suggested that they might interact with similar or identical transcriptional factors. If that were true, we would anticipate the two regions to contain related DNA sequences. Fig. 5 shows that this is the case; a nonanucleotide sequence at positions −51 to −43 in the proximal region is almost perfectly (one mismatch) repeated upstream at positions −128 to −120 in the distal region.

The significance of this duplication was further assessed by comparing the 5′ flanking sequences of the mouse MT-I gene to those of three functional human MT genes (ref. 18 and unpublished data). As shown in Fig. 5, each of the human genes contains a duplicated 12-bp sequence closely related to the 9-bp sequence found in the mouse gene. The proximal repeat is centered 14–17 bp upstream of the TATA box and is preceded by a pyrimidine-rich sequence. The position of the distal repeat, centered 95 to 115 bp upstream of the TATA box, is more variable. The overall homology between these sequences is 91%, a possible consensus version being 5′ T-G-C-G-C-C-C-G-C-$\frac{C}{T}$-C 3′.

DISCUSSION

Our major finding is that at least two regions of the DNA upstream of the mouse MT-I gene contain metal-responsive regulatory sites. Mutants in either region have decreased

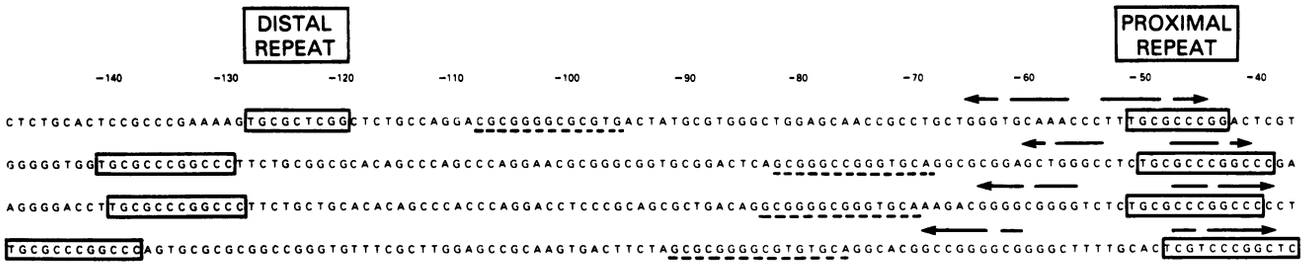

FIG. 5. Comparison of mouse and human MT gene 5′ flanking sequences. The top line shows the sequence from mouse MT-1, the second line human MT-Ie, the third line human MT-If, and the fourth line human MT-IIa. The sequences were aligned at the conserved TATA sequence (nucleotides −28 to −25 in the mouse MT-I gene). The duplicated heavy metal control sequences are indicated by boxes, the G-rich sequences by broken underlines, and the palindromes by arrows. The human MT-Ie and MT-If genes have been cloned and sequenced in our laboratory (unpublished data). The human MT-IIa gene was described by Karin and Richards (18) and an allele of this gene has been characterized by us.

expression levels but are still clearly responsive to cadmium; only when both are removed is there a complete loss of inducibility. A previous analysis of the mouse MT-I gene revealed the presence of the proximal control region (9). However, because only 5′ deletion mutants were examined, the distal region was not detected. More recently, Karin *et al.* (19) have described a deletion analysis of the human MT-IIa gene. They also found evidence for duplicated metal-responsive sites, and their functional analysis provides strong support for the evolutionary comparison shown in Fig. 5.

The two control regions contain a duplicated DNA sequence that is strongly conserved in three human MT genes. It is tempting to speculate that these homologous sequences act as core recognition sites for the binding of metal-responsive cellular regulatory factors. Alternatively, they could serve as binding sites for polymerase or other common transcription factors that recognize the sequence only in the presence of metal. Partial homologies to the repeat are found elsewhere in the MT-I flanking sequences, and it is possible that these also contribute to the overall regulation. Although both upstream regions function in the heavy metal response, they appear to differ in detail. Mutants with internal deletions of the distal region have severely reduced transcription levels but induction ratios as high as the intact gene. In contrast, mutants with deletions of the proximal region have almost normal levels of basal expression but lowered induction ratios. These quantitative differences may be mediated by the sequences surrounding the repeated nonanucleotide. Varying the distance between duplicated control sequences by 11–14 bp, or between the control sequences and the gene by 8 bp, had little or no effect on transcription or induction. Similar flexibility in the spacing of promoter elements has been observed for several other eukaryotic genes (reviewed in ref. 20).

While we believe that the repeated elements are critical for heavy metal regulation, it is clear that additional DNA sequences play an important role in controlling the overall efficiency of transcription. Thus deletions with end points upstream, between, or downstream of the repeats all resulted in decreased basal and induced expression levels. The region between the repeats contains a G-rich sequence, flanked by short stretches of alternating pyrimidines and purines, that is partially conserved in the human MT genes (Fig. 5), and sequence homologies are also present in the region upstream of the repeats (unpublished data).

An additional interesting feature of the mouse MT-I sequence, noted by Glanville *et al.* (17), is the hyphenated dyad axis of symmetry or palindrome that overlaps the proximal repeat. The human MT genes contain sequences that have similar potential secondary structures yet bear little primary homology to the mouse sequence (Fig. 5). Karin and Richards (18) have speculated that cellular regulatory factors might recognize these secondary structures, rather than a primary DNA sequence, but we feel this is unlikely for two reasons. First, the upstream repeat is responsive to cadmium even though it does not contain an overlapping palindrome. Second, the clustered point mutation cp1(−54/−34) fully restores the dyad axis of symmetry but has no effect on transcription. Hence the function of this conserved feature, if any, remains enigmatic.

Why are the heavy metal control sequences present in two copies rather than one? Obvious possibilities include protection against genetic damage, binding of a multimeric protein, or cooperative interaction between related but distinct cellular factors. To test these ideas it will be essential to develop methods to detect the interaction between the regulatory molecules and their target DNA sequences.

We thank Cary Queen, Carl Wu, George Khoury, and Maxine Singer for comments on the manuscript and Alice Middleton and Beverly Miller for excellent typing.

1. Foulker, E. C., ed. (1982) *Biological Roles of Metallothionein* (Elsevier, New York).
2. Karin, M., Anderson, R. D., Slater, E., Smith, K. & Herschman, H. R. (1980) *Nature (London)* **286,** 295–297.
3. Durnam, D. M. & Palmiter, R. D. (1981) *J. Biol. Chem.* **256,** 5712–5716.
4. Hamer, D. H. & Walling, M. J. (1982) *J. Mol. Appl. Genet.* **1,** 273–288.
5. Karin, M., Haslinger, A., Holtgreve, H., Cathala, G., Slater, E. & Baxter, J. D. (1984) *Cell* **36,** 371–379.
6. Pavlakis, G. N. & Hamer, D. H. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 397–401.
7. Karin, M., Cathala, G. & Nguyen-Huu, M. C. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 4040–4044.
8. Mayo, K. E., Warren, R. & Palmiter, R. D. (1982) *Cell* **29,** 99–108.
9. Brinster, R. L., Chen, H. Y., Warren, R., Sarthy, A. & Palmiter, R. D. (1982) *Nature (London)* **296,** 39–42.
10. Palmiter, R. D., Chen, H. Y. & Brinster, R. L. (1982) *Cell* **29,** 701–710.
11. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
12. Schumperli, D., Howard, B. & Rosenberg, M. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 257–261.
13. McKnight, S. L. & Kingsbury, R. (1982) *Science* **217,** 316–324.
14. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
15. Gorman, C., Padmanabhan, R. & Howard, B. H. (1983) *Science* **221,** 551–553.
16. Lusky, M. & Botchan, M. (1981) *Nature (London)* **293,** 79–81.
17. Glanville, N., Durnam, D. M. & Palmiter, R. D. (1981) *Nature (London)* **292,** 267–268.
18. Karin, M. & Richards, R. I. (1982) *Nature (London)* **299,** 797–802.
19. Karin, M., Haslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Nestphal, H. M. & Beata, M. (1984) *Nature (London)* **308,** 513–519.
20. Hamer, D. H. & Khoury, G. (1983) in *Enhancers and Eukaryotic Gene Expression*, eds. Gluzman, Y. & Shenk, T. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 1–15.