An Analysis of Nucleotide Polymorphism in the Human MT-II_A Gene Promoter Region

An Analysis of Nucleotide Polymorphism in the Human MT-II_A Gene Promoter Region

by

Brett J.D. Stevens, B. Sc.

A thesis submitted to the school of graduate studies in partial fulfillment of the requirments for the degree of Master of Science

McMaster University, 1992
Hamilton, Ontario.
Department of Biochemistry

ABSTRACT:

Previous research has shown varying degrees of renal damage on exposure to equal amounts of cadmium in occupationally exposed mining and factory workers. Further work has shown that in vitro exposure of human peripheral lymphocytes to the same cadmium levels resulted in significant variation in Metallothionein (MT) mRNA transcriptional induction over basal MT mRNA expression in a series of individuals. This variation could account for the differences in renal Cd toxicities identified previously. In this study, the human MT-II_A gene was cloned from 12 individuals, and the 5'promoter region was sequenced for each to determine the extent of promoter nucleotide variation. This is of interest since such an analysis has not been done in the past. No study has been done to look at the degree of polymorphism in a particular promoter region. Thus, there are no data on the degree of nucleotide drift or change which can occur in promoter regulatory elements. Such a study could provide insight into whether promoter changes could result in the type of variation described above. It could also give some insight into the degree of variation in sequences in the literature. The results obtained indicated that the human MT-II_A promoter region is highly conserved, with only one polymorphic site identified at position 557, between the glucocorticoid responsive element and the fourth metal regulatory element sequences. This suggests that promoter variation is not likely a significant yfactor in MT mRNA induction variability, although further analysis would be

needed to show this since only 12 people were analyzed. The results were compared against a study of nucleotide polymorphism in Drosophila melanogaster, which is the only other data on nucleotide variation specifically (Kreitman, 1983; Kreitman and Hudson, 1991). As well, a number of discrepancies were noted from the original published sequences in the literature, suggesting that errors are likely published in genomic sequence which are never identified, except through trial and error. This has potential repercussions when considering the use of such sequence in cloning and sequencing projects, like the sequencing of the human genome, since this would depend on the accuracy of previously published data.

ACKNOWLEDGEMENTS

I would like to thank my committee members, Dr. Calvin Harley, Dr. Rick Rachubinski, and Dr. Evert Nieboer for their supervision and support while working on this thesis. I would like to thank Calvin Harley for his advice and support, as well as his patience, while waiting for the completion of the thesis over the last few years. I would like to express my appreciation to Mrs. Jill Regoeczi for her great support and assistance, as well as her attempts to motivate me to work towards completion. I am appreciative of Tracy Arsenault for her advice and assistance while starting the thesis work and working in the laboratory, as well as for her friendship over the years. I would like to express my appreciation to my wife, Gillian, for her support throughout the laboratory work, as well as throughout the completion of this thesis. Without her harrassment this thesis might never have been completed! I would like to express my gratitude to all of those who provided advice throughout the time I worked on my Master's thesis in the various laboratories, and to all of my friends in the department, without whom my stay within Biochemistry might have run the risk of becoming tedious. As well, I would like to thank all of my family and friends for their support over the years, especially Dan Syroid, whom failure to thank specifically could result in bodily harm to myself.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

A_{260}	absorbance at 260 nm
ADH	alcohol dehydrogenase
Ag	silver
AP1	Activating Protein 1 (transcriptional factor)
ATP	adenosine triphosphate
Au	gold
BLE	basal level expression element
bp	base pair
Cd	cadmium, 2+ ion form
cpm	counts per minute
Cu	copper
Cys	cysteine
d	days
Da	daltons
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
DEP	diethylpyrocarbonate
dH_2O	distilled water
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
dTTP	deoxythymidine triphosphate

EDTA	ethylene diamine tetraacetic acid
EPIF	extracellular protein synthesis inducing factor
g	unit gravitational force
Glu	glutamic acid
GRE	glucocorticoid responsive element
GTP	guanosine triphosphate
h	hour(s)
Hg	mercury
HPLC	high pressure liquid chromatography
HSV-TK	herpes simplex virus- thymidine kinase
IL1	interleukin 1
IPTG	isopropylthiogalactoside
IRE	interferon responsive element
kbp	kilobase pairs
LiCl	lithium chloride
M	molar
MES	magnesium chloride, EDTA, saline solution
min	minute(s)
mm	millimeter(s)
mM	millimolar
mmol	millimole(s)

MMTV	mouse mammary tumour virus
Mr	relative molecular mass
MRE	metal responsive element
MRF	metal regulatory factor
mRNA	messenger RNA
MT	metallothionein
uCi	microCurrie(s)
ug	microgram(s)
ul	microlitre(s)
uM	micromolar
umol	micromole(s)
NaCl	sodium chloride
NaOAc	sodium acetate
NaH_2PO_4	monosodium phosphate
NaOH	sodium hydroxide
Pb	lead
PBS	phosphate buffer saline
PEG	polyethylene glycol
PKC	protein kinase C
pmol	picomole(s)
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid

rpm	revolutions per minute
s	seconds
SDS	sodium dodecyl sulfate
SM	simplified medium
SSC	saline sodium citrate solution
SV40	simian virus 40
TBE	tris-borate EDTA solution
TE	tris-EDTA solution
TPA	12-O-tetradecanoylphorbol-13-acetate
TRE	TPA responsive element
UV	ultraviolet
XGAL	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
Zn	zinc

1. INTRODUCTION:

In recent years, the study of the metallothionein (MT) proteins and their gene superfamily has become very productive because of its broad range of interrelated fields and applications. MT is of interest to the cell physiologist and toxicologist because of its potential involvement in trace element homeostasis, metal detoxification and storage, and cellular metabolism (Hamer, 1986; Karin, 1985). Further interest has been generated by evidence that suggests that these proteins may be involved in early development, since some MT genes appear to be modulated at a transcriptional level during embryogenesis (Andrews et al, 1984). As well, MT protein production appears to be stimulated by a number of stressful conditions, including extreme temperature or ultraviolet irradiation, which could implicate them in some kind of response to systemic stress (Bremner, 1987). MT is of interest to the molecular biologist because the genes which encode these proteins are transcriptionally responsive to a wide variety of inducing agents (Hamer, 1986; Karin, 1985). This makes the study of the gene family important as a model for eukaryotic gene regulation, as well as a tool to observe and study the influences of other proteins under the control of the MT promoter in cellular processes under different physiological conditions. Here the usefulness of MT as a model system for gene regulation is considered. Specifically, nucleotide variation in the MT-II_A promoter sequences between different individuals will be

the focus. This is of interest because a significant degree of variation has been found between different individuals with respect to MT mRNA induction in peripheral lymphocytes in culture, on treatment with cadmium (Cd)(Harley et al, 1989). In order to put this study into perspective, a review of the various aspects of the MT proteins and gene family will be given, as well as details of its expression, genetic organization, and the specifics of the MT-II_A promoter structure.

1.1. Rationale and Scope:

The rationale for this study was two fold. Previous research has shown that exposure to Cd occupationally can result in significant renal damage. However, exposure to the same amount of Cd did not necessarily result in the same degree of damage between different people (see section 1.8). MT binds Cd resulting in a decrease in the apparent toxicity of the metal to the exposed individual (see section 1.4). In vitro studies of human peripheral lymphocytes from a series of individuals has shown that a significant variation in MT mRNA transcriptional induction over basal level expression exists between different people, on exposure to Cd. This variation may indeed explain the variability in renal toxicity found on exposure to Cd in such settings. Thus, the human MT-II_A gene was cloned from 12 people, and the promoter sequence was determined for each to see if significant nucleotide variation existed in this

area. This would allow us to determine if significant variation in promoter elements existed and might provide insight into whether or not this could contribute to transcriptional induction variability. It could be argued that such variation is unlikely due to the apparent importance of the MT proteins. These genes are found extensively in the animal kingdom and are even present in yeast species, indicating that they have been conserved evolutionarily. However, no previous analysis of nucleotide promoter polymorphism has ever been done, so no data have been generated to support or deny such a conclusion.

Thus, it was the purpose of this study to assess the frequency of promoter nucleotide polymorphism from the human MT-II_A gene out of basic interest on its own, as well as to develop potential insight into MT induction variation. Such a study would also give the opportunity to reexamine the sequence of this particular sequence, providing insight into the possible degree of error found in published sequences in the literature.

1.2 Metallothionein Classification and Biochemistry:

1.2.1 Occurrence:

Metallothioneins are a low molecular weight, cysteine-rich family of proteins which bind group IB (Cu, Ag, Au) and IIB (Zn, Cd, Hg) metal ions (Kagi and Kojima, 1987; Dunn et al, 1987). These proteins have been studied extensively within the animal kingdom, as well as in certain plants, and in

some species of yeast. The occurrence of MT in these various organisms is covered well by Kagi and Nordberg (1979), Hamer (1986),and Kagi and Kojimi (1987). The MTs vary in respect to the amount produced between species, as well as in respect to which organs in that species express the proteins, and to what extent. Humans have been found to express MT in liver, kidney, brain, heart, testes, and blood tissues, whereas pigs and rabbits only show expression in liver and kidney tissues (Hamer, 1986; Kagi and Kojima, 1987). Its expression has also been studied in various tissue culture cell lines. The induction of synthesis of MT varies greatly between different organ tissues depending on the inducing agent present and the cell type (Dunn et al, 1987). Prokaryotic MTs are similar to mammalian MTs in their physical structures and metal and cysteine compositions, but not all of them have been found to share a high degree of amino acid sequence homology, as is seen in mammalian MTs. The occurrence of MT in so many different organisms from eukaryotic and prokaryotic microorganisms to mammals indicates the evolutionary conservation and importance of the MT gene family.

1.2.2 Definition and Structure:

Metallothioneins were defined in 1978 at the First International Meeting on MT and Other Low-Molecular Weight Metal-Binding Proteins (Kagi and Nordberg,1979) to include any low molecular weight protein (6000-7000)

having a heavy metal content of 7-12 metal ions per molecule of protein, and an amino acid composition of 23-33% cysteine. They are required to have the optical properties of metal thiolate clusters, and lack any disulfide bonds, aromatic amino acids, or histidine. Most of the cysteine residues are in a conserved distribution of Cys-Cys, Cys-X-Cys, or Cys-X-Y-Cys sequences (Kagi and Vallee, 1961). The mammalian MTs are proteins 61 amino acids in length, with 6-8 lysines, 7-10 serines, and 20 cysteines (all of which are invariant), with a highly variable metal content, depending on the tissue-type and environmental exposure to different heavy metals (Hamer, 1986). These proteins consist of two globular domains linked together to give an ellipsoidal shape to the molecule. Residues 1-30 form the amino terminal beta-domain, which contains 9 cysteine residues and has metal binding capacities of 3 atoms of Zn^{2+} or Cd^{2+}, or 6 atoms of Cu^{1+}. The carboxyl-terminal domain consists of amino acids 31-61 with 11 cysteines, allowing for the binding of 4 Zn^{2+} or Cd^{2+} atoms, or 5-6 Cu^{1+} atoms. Cd^{2+} and Zn^{2+} are tetrahedrally coordinated to form cysteine-thiolate ligands, whereas Cu^{2+} coordination is believed to occur trigonally through three thiolate ligands. The result is an overall negative charge to the protein. The alpha and beta domains are 15-20 angstroms in diameter with a spherical shape and contain the metal-thiolate clusters in their centers. The protein consists of two large helical turns which wrap around the metal-thiolate cluster cores (Kagi and Schaffer, 1988).

The binding affinities of MT for different metals vary as a result of the

stability of a given metal-MT complex. In mammals, the order of binding affinities is $Zn^{2+} < Pb^{2+} < Cd^{2+} < Cu^{1+}$, Ag^{1+}, Au^{1+}, where the Cu^{1+} complex has a stability constant 100 times greater than Cd^{2+}, and Cd^{2+} has a constant 10,000 times greater than Zn^{2+} (Kagi and Vallee, 1961). The binding of the metals is highly cooperative indicating the preference for full saturation of the MT proteins present. Although metals show different domain preferences and cooperativities of saturation, MTs are usually found to have a mixed metal population, as opposed to homogeneous metal contents (Kagi and Kojima, 1987; Hamer, 1986). In most higher mammals zinc is found to be bound to MT in levels greater than Cd^{2+} and Cu. In the human liver Zn can often be the sole metal bound by MT. Most physical studies have concentrated on homogeneous metal-protein complexes and the effects of a heterogeneous metal-MT composition on the proteins tertiary structure and cooperative binding characteristics are not known (Kagi and Schaffer, 1988).

The criterion for the classification of MTs presented earlier encompasses a large number of different proteins, which have been subclassified according to their primary structure. The class I MT designation consists of the mammalian MTs, along with those similiar to the mammalian proteins in primary amino acid structure. This includes the Drosophila melanogaster, crab, and Neurospora MTs. The class II designation contains other MTs that have no extended sequence similarities to the class I MTs, but still have Cys-Cys and Cys-X-Cys sequences. Examples of class II proteins

are found in yeast, wheatgerm, and cyanobacteria. The class III subgrouping includes atypical MTs containing gamma-glutamylcysteinyl units and are discussed in more detail in Kagi and Shaffer, 1988. These proteins have been found to have the sequence $(gamma\text{-}Glu\text{-}Cys)_n$, where n=2-8 and X is usually glycine. All of the class I and II proteins have been found to be composed of single-chain proteins, whereas the class III MTs are often made of two or more polypeptides of variable lengths.

The class I-MT proteins, which include the human MTs, usually consist of two or more different isoforms, classified into MT-I and MT-II on the basis of chromatographic behaviour. The isoMTs can again be subdivided by HPLC into subforms such as MT-I_A, I_B, I_E, and so on. These isoforms migrate differently as a result of variation in amino acids outside of the cysteinyl residues, which are highly conserved in their placement (Hamer, 1986). Human MT-I_A differs from II_A in only 7 amino acids and one negative charge at neutral pH, resulting in differential separation of these isoforms. MT-I_B differs from I_A in 9 amino acids, none of which are cysteine. The biochemical properties of these different isoforms are very similiar, with some slight differences between them in metal binding affinities.

1.3 THE MT PROTEINS ARE EXPRESSED BY A MULTIGENE FAMILY:

The human MT gene family is the most complex of the MT gene families studied so far, consisting of at least 12 members. Ten of these have

been shown to be functional, all of which were found to be localized to chromosome 16 (Karin and Richards, 1982). The different functional MT-I genes each encode a specific protein isoform. At present, three non-functional pseudogenes have been identified. MT-II_B is a processed pseudogene located on chromosome 4, which lacks any recognizable promoter region and has no intervening intron sequences (Karin and Richards, 1982). MT-I_C and I_D retain introns, but have in frame termination codons and/or deletions which alter the reading frame (Hamer, 1986; Sadhu and Gadamu, 1988). Chromosomes 1, 18, and 20 have been identified as having MT loci as well, but specific genes have not been identified (Schmidt et al,1984).

The functional isoforms encoded by the MT gene family are expressed differentially, depending on tissue-type, physiological state of the organism, and environmental exposure to MT-inducing agents (reviewed in Dunn,1987). The relative proportions of these isoforms do not only vary between different species, but are also seen to vary between different tissues in the same species. Although some of the human MT isoforms are expressed in a tissue-specific manner, MT-II_A was expressed in all cell types tested. MT-I_B, I_E, I_F, and I_G have all been shown to be expressed in some cell lines, but not others (Heguy et al, 1986; Schmidt and Hamer, 1986). As well, expression of I_F and I_G was variable depending on whether Cu, Cd, or Zn was given. MT-II_A expression in response to these agents was equal and greater overall than for

the others. From these and similiar studies it has been demonstrated that MT induction and regulation are very complex, depending on the metals present and the tissue-type involved. More detailed analysis will be needed to determine the functional significances of the observed induction specificities.

1.4 THE PHYSIOLOGICAL FUNCTION OF THE MT PROTEINS:

1.4.1 Detoxification and Storage:

A number of different possible functions exist for the MT proteins. The most obvious is the detoxification and storage of class IB and IIB metals. MT is induced by Cu, Zn, Cd, and some other metals and binds to most of these metals within the cell (Reviewed in Hamer, 1986, and Kagi and Schaffer, 1988). MT gene transcription was induced on exposure of animals to high Cu levels, and liver tissue under such circumstances was found to contain the Cu in a relatively non-toxic form bound to MT (Johnson et al, 1981). MT has a high affinity for Cu, which is believed to build up as it does in the cell as a result of the low rate of Cu-MT turnover. Menkes' disease which appears to be a defect in Cu-MT turnover, results in the accumulation of Cu in cells, but this Cu cannot be made available when necessary for cell processes and other tissues (Riordan, 1982).

MT has a strong affinity for Cd also, and its induction is highly induced by Cd. Cd can be absorbed into the blood stream by ingestion or inhalation, and is carried to the liver by metal carrier proteins or MTs which are present in

the blood. The Cd enters the cell, binds to MT, and induces MT transcription (Webb, 1987). Degradation of the Cd-MT frees the Cd, which is usually rebound by MT. Evidence has indicated that Cd is much more harmful to cells in its unbound state than in its bound state, supporting a role of detoxification for MT. Cd-MT in the blood stream can also be absorbed by the renal tubular cells where it can again be stored. Degradation of the protein results in the rebinding of the Cd by new protein. This cycle continues until the unbound Cd levels exceed the binding capacity of the MT proteins in the cells, at which point the highly toxic unbound Cd is present, and can begin to cause renal damage (reviewed by Webb, 1987).

Work by Nieboer et al, 1988 has suggested the existence of two mechanisms for renal Cd reabsorption. The first being a complete and saturable process which operates at low levels in non-occupationally exposed individuals. In these people, almost no Cd is found in urine since reabsorption is essentially complete. However, in people occupationally exposed to Cd, a second mechanism appears to become operative, as higher MT-Cd levels occur and the first mechanism is completely saturated. This second mechanism is nonsaturable. In these individuals the urinary Cd level is a constant fraction (approximately 80%) of the blood Cd. Although it would appear that further work to elucidate this mechanism is needed, MT does appear to play an essential role in Cd renal reabsorption and clearance.

Thus, it would appear that MT induction is important to offset renal

damage resulting from exposure to high levels of Cd. The detoxification role is also supported by pretreatment of animals with Zn which results in MT induction, and a protective effect on these animals against lethal doses of Cd (Dunn et al, 1987). Similiar results have been achieved in tissue culture experiments. Cell lines which have additional MT genes or amplified MT genes show increased protection from Cd as well (Compere, 1986).

1.4.2 Cellular Metabolism:

The primary role of MT may not be this detoxification function, since these heavy metals are not normally found in high concentrations in the environment. As well, one would expect to see MT induction only on exposure to these metals, but MT is expressed to a significant level constitutively, in the absence of heavy metals (Karin, 1985). The MTs may also play a role in cellular metabolism, since Cu and Zn are commonly involved in cell processes. The binding and degradation of the MT-metal complexes could allow for the modulation of Cu and Zn levels within the cell, as well as the movement of these metals between cells via individual ion movement or gross movement by membrane fusion processes. Some enzymes are activated in vitro by Zn or Cu-MT complexes. A similiar in vivo scenario could mean that changes in MT-metal degradation could modulate the activity of certain enzymes. Variation in the MT-complex levels and in the free pool of Cu and Zn are likely to have profound effects on cellular

metabolism, so MT could play a major role in such modulation. Evidence collected to date suggests that MT controls cellular Zn levels under normal dietary conditions, and that a decrease in dietary Zn levels will result in decreased MT gene expression (reviewed in Dunn, 1987).

1.4.3 Intestinal Metal Absorption:

MT may also function to regulate the absorption of these metals in the intestine. Studies have indicated that MT is induced in the intestine by Zn administration (Cousins, 1985). When intestinal MT levels are high, Cu and Zn absorption is seen to be decreased, indicating that MT may have a dampening role in the absorption of these metals through the intestine, protecting from large influxes of these metals. Thus, it is very likely that MT is involved somehow in detoxification and storage, as well as cellular metabolism and regulation of systemic and cellular heavy metal ion levels.

1.4.4 Fetal Development:

It is possible that MTs play a role in fetal development, since MT mRNA levels are elevated in certain stages of gestation (Andrews et al, 1984). These studies suggest that MT could play a role in development, possibly by controlling Zn and Cu levels in certain tissues, resulting in variability in enzyme activity or differentiation. However, evidence has not directly implicated MT in developmental regulation. The potential for MT functioning in differentiation

and proliferation has also been studied, but again concrete evidence is lacking.

1.4.5 Cellular Defense Mechanisms:

MT is induced by a number of hormones and other substances, indicating it may be involved in cellular defense mechanisms. These inducers include glucocorticoids, epinephrine and inflammatory agents (Cousins, 1985; Brady et al, 1987). These effects on MT levels can result in changes in plasma Zn levels, as a result of induction not resulting from elevated heavy metal concentrations. The physiological significance of MT induction by these agents is not known. Different types of acute stress also result in MT transcriptional stimulation, such as extreme temperature, exercise, or UV irradiation (reviewed by Bremner, 1987). Injection of bacterial lipopolysaccharides results in induction, which may be modulated partially by inducing hormones like glucocorticoids, but likely involves other factors as well (Karin, 1985). These kinds of stress result in the induction of the acute phase response, resulting in the production of a number of proteins, such as interleukin-1, which has been found to induce MT as well. Interferon also is an inducer (Cousins and Leinart, 1988; Bell et al, 1987). MT may also protect cells from the effects of ionizing radiation, since EPIF (extracellular protein synthesis inducing factor) which is secreted by some cells after UV irradiation, induces the synthesis of MT.

It appears that whatever the true physiological role of MT turns out to be, it will be multifacetted and complex.

1.5 Metallothionein Gene Induction:

1.5.1 Induction by Metals:

Induction of MT by metals was shown to be regulated primarily at the level of transcription by several groups using the mouse MT-I gene. mRNA levels and protein synthesis increased together, peaking four hours after treatment with Cd or Zn in liver or kidney tissues (Durnam and Palmiter, 1981). Maximum MT synthesis occurred 3-10 h after treatment with Cd through cysteine incorporation studies. mRNA levels increased 1.7 fold on treatment with $ZnCl_2$ and 3.7 fold with $CdCl_2$, on the basis of in vitro translation assays (Probst et al, 1977). This increase was transient and MT mRNA levels returned to control levels after 17 h. Maximal levels of MT lag behind those of mRNA by 1-4 days as a result of the proteins longer half-life (Anderson and Weser, 1978). [^{35}S] cysteine incorporation and MT-^{65}Zn binding were both blocked by actinomycin D and cordycepin suggesting that MT biosynthesis is controlled at the level of transcription (Squibb et al, 1977). As well, actinomycin D was found to inhibit the induction of MT mRNA synthesis by $ZnCl_2$ treatment (Shapiro et al, 1978). Durnam and Palmiter (1981) showed a 17 fold increase in the rate of MT transcription in the kidney which resulted in a 40 fold increase in MT-I mRNA transcripts. Any effect of mRNA stability was

not believed to be significant. Because of the variability of MT mRNA in uninduced liver (5-400 molecules per cell), induced levels were difficult to determine but data suggested a 1.5-2.0 fold increase of MT-I mRNA transcription in liver over kidney. Variability in transcription levels between tissues could result from a number of different factors including variability in metal absorption, accessibility to regulatory proteins required for transcription, variation in maximal transcriptional rates between tissues, or differences in mRNA stabilities for various reasons. Other inducing agents have been shown to have their primary effect at the transcriptional level as well (Hamer, 1986, Dunn et al, 1987).

MTs are induced in certain blood cells, but not to the same extent in all types (Hildebrandt and Cram, 1979; Peavy and Fairchild, 1987; Harley et al, 1989). Proteins similiar to HeLa cell MTs can be induced on treatment with Cd in monocytes and lymphocytes (Koizumi et al, 1987; Peavy and Fairchild, 1987; Sone et al, 1988). As well, B and T-lymphocytes have been tested for Cd sensitivity and T-cells were found to be more sensitive, possibly indicating a decreased ability to synthesize MTs (Koizumi et al, 1987). The expression of MTs was shown for monocytes, polymorphonuclear leukocytes and lymphocytes by $CdCl_2$. More recent evidence shows 10 fold induction of MT with 10uM $CdCl_2$, with 3 fold higher absolute levels of MT in monocytes compared to lymphocytes in respect to basal and induced levels. Polymorphonuclear leukocytes contained low levels of MT regardless of $CdCl_2$

treatment (Harley et al, 1989). MT induction appears to be greatly suppressed in some lymphatic cell lines, including primary thymocytes, a thymoma cell line, and a lymphoma cell line, possibly due to DNA methylation effects (Compere and Palmiter, 1981). MT is also found to be excreted into some bodily fluids, including blood, bile, and urine. The amount found appears to depend on the exposure to inducing trace metals. The liver is likely the source of the MTs in the blood and bile, but this is not clearly established (Bremner et al, 1987). The variable expression of MT in blood cells and its presence in given body fluids again emphasizes the complexity of the gene families induction and regulation.

1.5.2 Induction by Glucocorticoids:

Glucocorticoid induction of MTs was first shown by the administration of dexamethasone to rats, resulting in increased MT mRNA levels and reduced plasma Zn levels (Hager and Palmiter, 1981; Mayo and Palmiter, 1981). The induction kinetics were the same for dexamethasone as for the administration of metals, and the steroid induction was shown to be a primary induction response, independent of changes in Zn transport or de novo protein synthesis (Karin et al, 1980). MT mRNA accumulation was dependent on an increase in RNA transcript initiation, as indicated by actinomycin D inhibition, but some evidence indicates a significant involvement of post-transcriptional stabilization. Levels of mRNA transcription and accumulation for

dexamethasone were 1.5-2.0 fold lower than levels for metal induction for the mouse MT-I gene. MT induction was again seen to be variable, depending on the tissue under analysis.

1.5.3 Induction by Other Agents:

Catecholamines, including epinephrine, norepinephrine, or isoproterenol, as well as polypeptide hormones, such as insulin, glucagon, and arg-vasopressin (Cousins et al, 1986; Brady et al, 1987; Nebes et al, 1988) have also been shown to induce hepatic MT gene expression. The beta-agonists glucagon and epinephrine stimulate adenylate cyclase in the plasma membrane leading to the stimulation of cAMP production, activating cAMP-dependent protein kinases (Brady et al, 1987). MT mRNA synthesis was seen with some type-1 interferon-inducing agents to a lesser extent than Cd (Friedman and Stark, 1985; Bell et al, 1987). Interleukin-1 appears to stimulate MT mRNA synthesis in rats by an as yet undetermined mechanism which seems to be a cell receptor-mediated glucocorticoid-independent event (Cousins and Leinart, 1988), but it appears to have a more potent effect than dexamethasone or multi-hormone treatment. Induction of MT synthesis is seen with bacterial endotoxin (lipopolysaccharides) which results in major systemic hormonal changes, although evidence indicates that this is not due to glucocorticoid release. The actual mechanism of induction by these agents is not completely understood but it does indicate the multi-hormonal complexity of MT gene

regulation (Durnam et al, 1984).

Blood serum factors and activators of protein kinase C have been implicated in the regulation of MT gene expression. Imbra and Karin (1987) indicated that some mitogenic agents induced MT gene expression. They hypothesized that the observed effect was controlled by the activation of PKC, possibly through increasing the activity of some regulatory factor involved in MT-II_A transcriptional initiation, although the mechanism of induction by serum factors and TPA were likely not the same, since the induction kinetics were different. Their conclusion that there was a correlation between elevated MT synthesis and cellular proliferation was not supported by studies of MT mRNA levels in the HL-60 human promyelocytic leukemia cell line by Hanke et al (1988). These studies showed that there was no significant difference in Cd induced MT mRNA levels between HL-60 cells differentiated towards neutrophil-like cells by retinoic acid , HL-60 cells differentiated towards macrophage-like cells by TPA, or uninduced HL-60 cells. Because PKC activation in HL-60 cells is down-regulated on extended exposure to TPA and increased after retinoic acid treatment, an involvement of PKC in MT mRNA transcriptional regulation would result in differential mRNA induction between the induced and uninduced HL-60 cell states. The studies of Imbra and Karin and Hanke et al may indicate that PKC activation may be involved in regulatory control of MT expression, but that this control may be tissue-specific, since Imbra and Karin showed induction of MT through PKC

activation in fibroblasts which was not seen by Hanke et al in the HL-60 cell line. It may also be possible that PKC induction is associated with MT induction, but is not causally involved.

1.6 THE MECHANISMS OF MT GENE EXPRESSION:

1.6.1 Mechanisms of Expression in the Mouse:

The mouse MT-I gene (Glanville et al, 1981) promoter region was analyzed by 3', 5', and internal deletions, duplications, insertions, and clustered point mutations by insertion of these promoter mutants into a mammalian expression vector to observe their transcriptional capacity (Brinster et al, 1982; Carter et al, 1984; Searle et al, 1987). Analysis of these point mutations revealed two regions which were necessary for efficient mRNA transcription and induction by Cd. Distal and proximal control regions were identified, each of which allowed for inducibility on its own. Mutation of the distal region alone resulted in a decrease in transcriptional ability, but did not alter the ratios of induced over basal transcription levels. However, mutations in the proximal element did not maintain the same mRNA induction capacity, but did have similiar basal levels. The two regions contain a consensus repeat which was seen to be conserved in the hMT-II_A, I_E, and I_F genes. It is possible that this is a core recognition sequence for potential metal-inducible transcriptional activators (Carter et al, 1984). Progressive 5' deletions towards the transcriptional initiation start site showed a progressive loss of metal

inducibility, also indicating that multiple regulatory elements are involved in the response. Sequence analysis has indicated the occurrence of a series of repeated sequence motifs referred to as metal responsive elements (MREs), which represent the minimal sequences needed to confer metal inducibility to the promoter (Stuart et al, 1984; Carter et al, 1984; Searle et al, 1987) Further analyses were carried out in non-metal-inducible promoters with synthetic MREs corresponding to the different sequences found in the actual mouse MT-I promoter, and induction was found to occur to varying levels depending on which MRE sequence was present (Searle et al, 1987). Because the MREs are not true enhancers, they must be inserted into a functional promoter, and the induced levels over basal level expression can be determined. More recent evidence indicates that one copy of the specific mouse MRE consensus sequence, MREd, gives only 4-5 fold induction, whereas two copies gives 10-20 fold induction. Of the 5 synthetic mouse MREs studied, two induced strongly, one moderately, and two very weakly. One resulted in induction, but it also appeared to result in a very elevated basal level of expression as well. The precise positioning and specific orientation of the MREs does not seem to be essential, but their positioning in respect to the TATA box does have a significant influence on transcriptional efficiency (Carter et al, 1984; Palmiter, 1987). The specific effect of changes at each base in the mouse MRE consensus sequence have now been dissected as have the effects of substitutions outside the consensus which result in less significant

impairment of function. Culotta and Hamer (1989) have shown that the mouse MRE can be dissected into two parts; a highly conserved heptanucleotide core believed to have the regulatory function, and a GC rich region with an Sp1 site believed to have a less important role. The results of the different studies on the mouse MT MREs indicate that all MREs are not created equally, and that the presence of slightly different MREs could result in a greater degree of regulatory capacity of the promoter in response to metals.

1.6.2 Mechanisms of Expression in Humans:

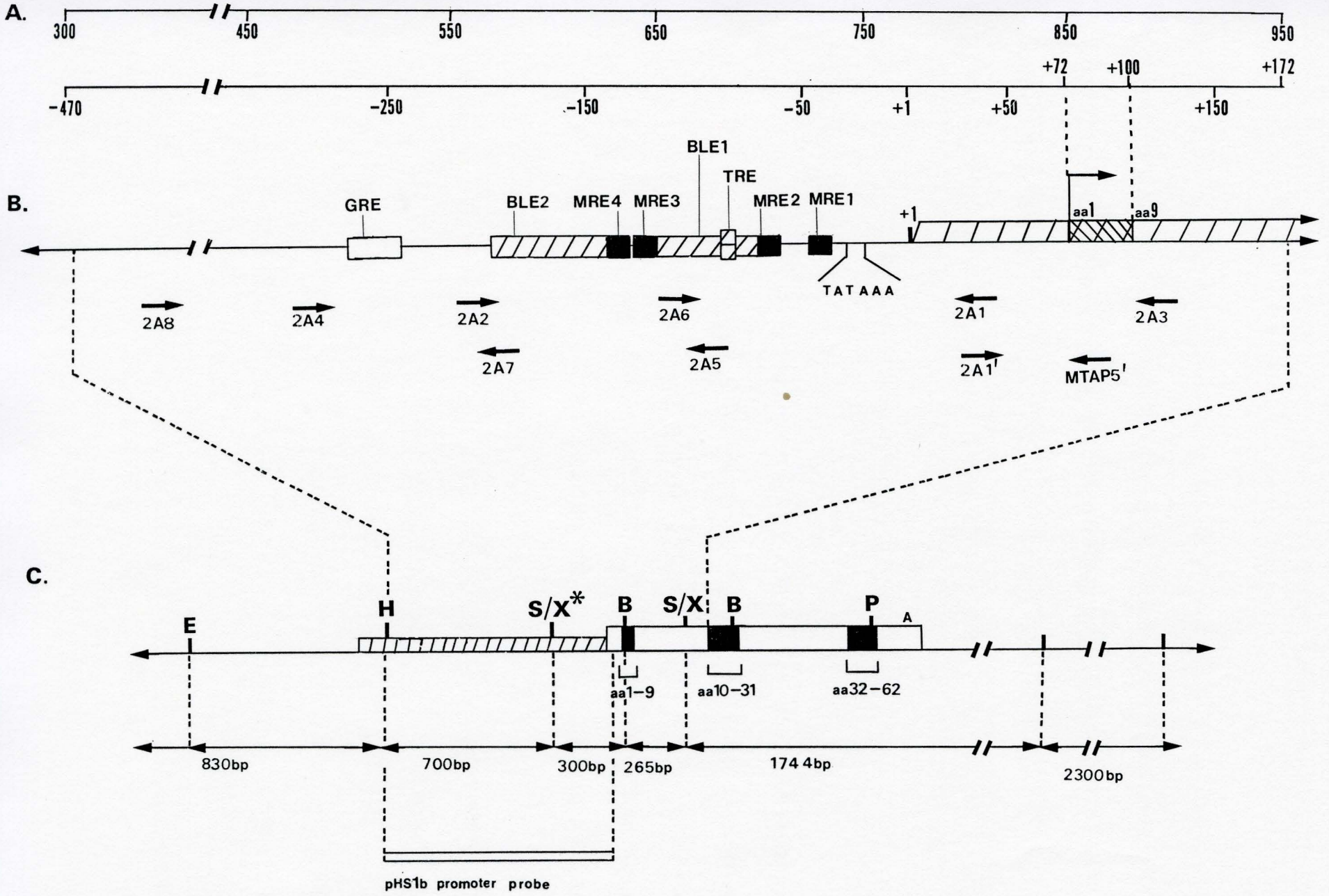
The human MT-II_A gene has also been cloned and dissected in a similiar manner (Karin and Richards, 1982). MT-II_A was introduced into rat fibroblasts and found to be responsive to glucocorticoids and heavy metals (Karin et al, 1984B). 5' deletions of the 5' flanking region of the gene were made and fused to the HSV TK gene. Three separate regulatory regions were determined; one for heavy metal responsiveness, one for glucocorticoid responsiveness, and one for basal level expression. The metal elements were defined here as being between -50 and -166 in respect to the +1 transcriptional start site, and it appeared that only 50 bases were necessary for heavy metal responsiveness. Two MREs were identified here and were seen to be highly conserved with previously identified MREs from other genes (see Karin et al, 1984C, and Figures 1 and 5). Further studies by Karin et al (1987A and 1987B) indicated that not two, but four MREs were present,

located at -38 to -50 (MRE1), -68 to -78 (MRE2), -128 to -139 (MRE3), and -142 to -158 (MRE4). These MREs did not function as enhancers, in that they did not increase transciption on their own, or in multiples, when placed upstream of a heterologous promoter (Karin, 1985). The second regulatory region identified was the basal level expression element, between -66 and -88. Analysis by Haslinger and Karin (1985) showed that the high level of basal transcription was a result of the presence of an enhancer element within the upstream promoter element. This region, which was later narrowed down to approximately -80 to -120 (Karin et al, 1987A), was capable of increasing transcription from a heterologous promoter from as far away as 2.5 kbp in any position or orientation, independently of the MRE sequences. The enhancer element was found to consist of two slightly divergent repeats (see figure 1, sequence in 5). It is currently believed that the activity of the basal level expression elements (BLE) is modulated by the MREs, since the MREs have no inductive effect on their own. The MREs seem to confer inducibility to the BLEs in response to heavy metal ion exposure (Scholer et al, 1986; Karin et al, 1987A). This method of conferring inducibility was illustrated previously with class I transplantation antigen genes in their response to interferon (Israel et al, 1986).

Promoter analysis in the mouse MT-I gene has not allowed for the separation of BLE-like enhancers from the MREs so far, but there is some evidence to support the idea that such an element exists (Searle et al, 1987). It

Figure 1: Nucleotide numbering systems, promoter and gene structure.

The nucleotide numbering systems, gene and promoter structure of MT-II_A are illustrated. (A) The Microgenie (Beckman, Genbank) nucleotide numbering is shown from position 300 to 950, and corresponds to that used in Figure 4. The lower numbering system is in reference to the +1 transcriptional start (Karin et al., 1982; Karin et al., 1987). (B) The human MT-II_A promoter structure is illustrated. The glucocorticoid, metal, TPA, and basal level responsive elements are shown, as well as the +1 transcriptional start site and TATAAA box sequences. Below the promoter structure, the oligonucleotides utilized for sequencing are illustrated. Arrows pointing to the left primed off the (+) strand, giving (-) stand sequence, and those pointing right primed off of the (-) strand, giving (+) strand sequences. (C) The partial restriction map of the MT-II_A EcoRI fragment is shown, with promoter sequences indicated as a hatched box, exons as black boxes, and introns as clear boxes. B= BamHI, E= EcoRI, H= HindIII, P= PvuII, S= SmaI, X= XmaI. Some of the restriction fragment sizes are shown. The polyadenylation signal (A) is indicated, as are the exon amino acid divisions. The pHS1b MT-II_A promoter probe fragment (see section 2.1.3) is also indicated.


A.
300
450
550
650
750
850
950
+72
+100
+172
-470
-250
-150
-50
+1
+50
+150
B.
BLE1
TRE
GRE
BLE2
MRE4
MRE3
MRE2
MRE1
+1
aa1
aa9
TATAAA
2A8
2A4
2A2
2A6
2A1
2A3
2A7
2A5
2A1'
MTAP5'
C.
E
H
S/X*
B
S/X
B
P
A
aa1-9
aa10-31
aa32-62
830bp
700bp
300bp
265bp
174 4bp
2300bp
pHS1b promoter probe

is obvious that DNA sequences other than the MREs are essential for control of transcription in the mouse system as well (Carter et al, 1984). It is most probable that the mechanism for induction of MT-II_A in response to heavy metals in man is analogous to the situation in the mouse.

The human MT-I genes show significant homology in their MRE sequences with the MT-II_A gene, and the spatial arrangement of the MREs is similiar between the two (Foster et al, 1988). The main difference between the two is the spacing between MRE2 and MRE3. In MT-II_A this region consists of the 50 bp containing the first BLE. This region varies between 28 and 35 bp in the MT-I_A, I_B, I_E, I_F, and I_G promoters. Again, the experiments to prove if the mechanism of action of the MT-I genes is the same as that for MT-II_A have not been done, but similarities in the promoter structure and heavy metal induction capacities indicate that it is likely analogous.

1.6.3 Mechanisms of Expression by Glucocorticoids:

In order to determine the mechanism of induction of transcription by glucocorticoid hormones, filter-binding assays were carried out using partially purified hormone-receptor complex and end-labelled DNA fragments containing the MT-II_A 5' flanking deletion mutants (Galas and Schmitz,1978; Scheidereit, 1983). Regions of the promoter were determined to be important in glucocorticoid binding when a promoter deletion showed a loss of hormone-receptor binding (Karin et al, 1984C). In this way, the glucocorticoid

responsive sequences were determined to be at least 268 bp upstream of the +1 site. DNase 1 and methylation protection assays (Ogata and Gilbert, 1978) were used to further pin-point the specific area of hormone-receptor complex-GRE DNA contact down to -241 to -266, with specific protein-DNA contacts between two G residues at -249 and -258. Not only was the GRE responsible for induction, it was also the receptor-DNA binding site. Comparison of the MT-II_A GRE to the MMTV-I and II_A GREs showed a strong consensus (Palmiter, 1987). Experiments done with GRE sequences indicate that unlike the MREs, GREs act as true enhancers, functioning independently to stimulate transcriptional initiation (Chandler et al, 1983). GRE sequences have not been identified in the human MT-I gene family or in the mouse MT-I or MT-II genes (Richards et al, 1984; Searle et al, 1984; Palmiter, 1987). MT-I_A has shown slight glucocorticoid inducibility, whereas I_E shows very strong inducibility similiar to II_A (Hamer, 1986). I_G and I_F appear to be uninduced by dexamethasone (Schmidt et al, 1985; Foster et al, 1988; Sadhu and Gedamu, 1988). Richards et al (Richards et al, 1984) suggested that the lack of strong induction as is seen with MT-II_A is a result of the absence of the GRE consensus. However, this is questionable since MT-I_E and I_F both appear to lack the consensus, but I_E shows induction (Varshney et al, 1986).

A TPA-responsive element (TRE) consensus sequence (Lee et al, 1987; Angel et al, 1987B) responsible for transcriptional activation on treatment with tumour promoting agent (TPA), was found -107 to -95 in the MT-II_A promoter

region. It appears to increase transcriptional activity in the absence of inducing agents, indicating that it may also have basal level effects, as well as TPA induction activity. TREs have been shown to be the binding site of the transcription factor AP-1, which appears to have increased activity on TPA treatment of some cell types (Varshney et al, 1986; Lee et al, 1987). It is believed that the TPA effect is the result of a complex cascade generated by the activation of PKC, but the actual mechanism of TPA induction is not known.

The 5' flanking promoter regions of a number of interferon-alpha inducible genes were also analyzed, and a long stretch of homology (30bp) was found. This has been hypothesized to be an interferon-responsive element (IRE) (Friedman and Stark, 1985). The potential IRE is found to be about 700 bp upstream of the +1 start site in the hMT-II_A promoter. Deletion analysis has not been carried out yet to determine if this is actually a responsive element.

The MT-II_A promoter element also has a GC box (TCCGCCCA) (Kadonaga et al, 1986) which is be necessary for maximum transcriptional efficiency, but has no effect on inductive capacity (Karin et al, 1987A). This GC box is found at -60 to -67 immediately 3' to MRE2 (see figure 1 and 5), and 30bp 5' fo the TATAAA box. Thus, two elements are required for basal level transcription; the BLEs and the GC box. The GC box is believed to be bound directly by Sp1, a promoter specific transcription factor, resulting in the

activation of transcription from these promoters by DNA polymerase II (Dyan and Tijan, 1983).

When Karin et al (1984C) determined the region involved in metal and glucocorticoid induction, they also noted a region of potential secondary structure between -70 and -90. This was an alternating purine-pyrimidine stretch which could be capable of forming a Z-DNA structure. Some studies have shown similiar sequences in the mouse MT-I gene. Visentin and Harley (1987) reversed two bases in the mouse MT-I potential Z-DNA region such that its potential for forming Z-DNA was increased. Two-dimensional chloroquine gel analysis and chemical mapping by DEP sensitivity indicated that when subjected to sufficient torsional stress, the mutated region underwent a B-Z DNA transition. The wild-type segment appeared to undergo a weaker transition showing that this region is capable of forming Z-DNA in vitro. Analysis of transcription from such promoters could allow the role of Z-DNA in promoter function to be determined further. No evidence to date has shown that it does have such a role.

1.7 MODELS FOR INDUCTION OF TRANSCRIPTION:

1.7.1 Models for Induction by Metals:

The mechanism of the transcriptional induction of MT appears to be very complex, involving both the MRE and BLE sequences, as well as other controlling elements, and multiple protein factors. Scholer et al (1986) carried

out in vivo competition assays between MT-II_A and SV40 which indicated that the factors that were competed for by the two promoter regions bound to the BLE and not to the MREs. From their results, they proposed two models to explain heavy metal induction. A rate-limiting transcriptional factor could bind the BLE sequence. The metal regulatory factor (MRF) which binds to the MRE sequence could bind the MRE sequences in the presence of the appropriate metals, allowing for the stabilization of the rate-limiting transcriptional factor, and increased transcriptional efficiency. The second model involved the MRFs as binding to the MREs and inhibiting the rate-limiting transcriptional factor, resulting in a lower transcriptional efficiency. Metal ions like Cd^{2+} could bind to the MRFs, decreasing their affinity for the MRE, allowing transcription to increase. The second model is less likely to be true since studies in which the MRE sequences were deleted and the BLEs were still functional did not result in an increased ability to transcribe from the promoter (Karin et al, 1987B).

In order to identify which proteins were involved in the metal-specific induction of the gene, Seguin and Hamer (1987) used DNA footprinting and bandshift assays in the presence and absence of Cd. They showed that a nuclear factor from mouse L cells was bound to a DNA fragment containing the mouse MREd sequence, and that this binding was regulated by the presence of Cd. It also appeared that the factor was present before exposure to Cd, and that the factor was activated; new synthesis of the factor was not initiated. This was supported by the addition of Cd to noninduced extracts

which resulted in similiar bandshifting patterns to that of induced extracts. These results occurred with Cd levels comparable to intracellular concentrations and induction ratios were comparable to those seen for MT-I in cell culture transfection assays (Mayo and Palmiter, 1981). Results obtained by Westin and Schaffner (1988) indicated that Zn induced binding of a specific protein (MTF-1) from HeLa cell extracts to MREd. Other evidence suggested that this interaction was responsible for transcriptional activation by the heavy metal. As well, the authors showed that the binding of Sp1 and MTF-1-like factors may have a role in MT regulation. Proof that specific protein-MRE DNA interactions occurred in vivo as they do in vitro was acquired using a genomic sequencing technique, in which cells exposed and unexposed to Cd were treated with dimethyl sulfate in order to detect changes in the reactivity of chemical modifying agents to the DNA in treated and untreated cells (Andersen et al, 1987). A trans-acting factor or factors in a rat hepatoma cell line was found to bind specifically to a fragment of the rat MT-I promoter in Cd treated cells. The same study suggested the regulation of the basal level expression was carried out by some other transcriptional factor such as Sp1, although Sp1 is not likely to be the one involved (Westin and Schaffner, 1988). Lee et al (1987) have suggested than an enhancer-binding protein which they have identified (AP2) binds to the distal BLE and activates transcription. Thus, AP2 could be the BLE regulatory protein.

Seguin and Prevost (1988) were first to begin characterizing a metal

regulatory protein specific to MT. A protein specific for the mouse MREd sequence was induced and detected on Cd treatment of mouse L cells, using a protein-blotting procedure. The p108 protein appears to be a unique transcriptional factor with very high affinity for the MREd sequences. Although evidence suggests that p108 is the regulatory factor for mouse MT-I, it is possible that other factors are also involved. Obviously, further studies are needed to gain a full understanding of the protein-DNA interactions governing heavy metal MT induction of transcription.

Experimentation with MT gene regulation in yeast may aid in the understanding of heavy metal MT induction. Thiele et al (1986) showed that even though the mammalian MT proteins are different from those in yeast structurally, they appear to be functionally similiar. Mammalian MTs were expressed at high levels in yeast lacking their own MT genes, resulting in protection from high concentrations of Cu, as well as the ability to carry out the yeast's MT autoregulatory function. It was shown that the trans-activating protein for MT activation in yeast was encoded for by the ACE1 gene (Thiele et al, 1988). This protein was shown to undergo a major conformational change on binding Cu (Furst et al, 1988), and it is believed to be this effect that allows for the activation of the factor and its binding to specific DNA regions. The protein only binds the DNA after the binding of Cu, which coincides with the induced conformational change. Szczypka and Thiele (1989) cloned the ACE1 gene and found 12 cysteine residues in the 105

amino acid protein, with CYS-X-CYS and CYS-X-X-CYS arrangements similiar to the MT protein structure. The implications of this structure and potential metal binding capacity are not known, but it does appear that the protein is nuclear. This provides a reasonable model for the activation of MRFs by the presence of heavy metals, although analogous factors from mice and humans will have to be isolated before the relevance of this data to those systems can be determined.

1.7.2 Models for Induction by Glucocorticoids:

Work done with rodents has resulted in a better understanding of MT glucocorticoid induction than for MT induction by metals. Originally, activation of gene transcription by glucocorticoids was believed to be initiated by the activation of the hormone receptor by hormone binding, and the translocation of the complex into the nucleus (reviewed by Yamamoto, 1985; Becker et al., 1986). However, this has been disputed by evidence which shows that non-hormone bound receptors were found in the nucleus (King and Greene, 1984; Welshons et al, 1984). Unbound receptors have non-specific affinity for DNA, and hormone-receptor binding appears to greatly increase the binding specificity (Yamamoto and Alberts, 1974; Willman and Beato, 1986). One hypothesis (Beato, 1989) suggests that the unbound receptor is complexed with another protein in the cytoplasm and that hormone binding relieves this interaction, allowing translocation. Others feel that hormone binding may

open a nuclear translocation signal within the protein. The receptor involved has been isolated and studied extensively (Danielson, 1987; Danielson et al, 1989) but the actual mechanism by which the glucorticoid-receptor complex activates transcription is not known in terms of the transcriptional machinery involved. The complex binds to the GRE consensus region (Karin et al, 1984C; Beato, 1989)and a number of transcriptional factors have been shown to be involved in the process (reviewed by Beato, 1989). Ptashne (1989) has suggested that the mechanism involves the stabilization of a transcriptional complex by these factors and the hormone-receptor complex. This parallels the MRE situation, in which the MREs are believed to modulate transcriptional initiation by stabilizing a transcriptional complex which has formed on the BLE sequence regions.

It can be seen that transcriptional regulation can occur through the interaction of regulatory factors with specific enhancer elements, as well as through the modulation of transcription by proteins to regulatory elements which affect the stability of the transcriptional machinery. The complexity is further increased by the occurrence of multiple sequence elements and the presence of different enhancers which respond to different agents. Regulation depends on which elements are present in the gene and which regulatory proteins are present within the cell. As well, one protein may have more than one function, and different combinations of proteins may have particular effects. More work must be done before a full understanding of these factors

and influences are understood.

1.8 MT TRANSCRIPTIONAL VARIATION IN RESPONSE TO Cd, AND PROMOTER VARIATION:

The threshold level of Cd required to cause renal damage seems to vary between individuals. Workers who have been exposed to similiar concentrations of Cd in an industrial setting do not necessarily achieve the same degree of kidney damage. The possibility that this variation is a result of differences in MT gene expression have been tested by the determination of MT induction ratios in peripheral lymphocytes from a group of individuals (Harley et al, 1989). Lymphocytes were isolated and treated with Cd. The MT mRNA transcription from cells treated with $CdCl_2$ were compared to untreated cells by isolation of mRNA after incubation with or without Cd, and dot blot analysis. Comparison of the mRNA levels from Cd treated and untreated cells indicated that the induction ratio of MT mRNA ranged from approximately 5-40 fold in different individuals. It is possible that this kind of variation could be responsible for the differences in acquired renal toxicity between workers.

The motivation to pursue the current study consisted of two parts. First, the variation observed in the study discussed above could result from a number of different factors. Differences in MT expression could result from differences in the expression of transcriptional factors, or differences in mRNA

stability or degradation, between individuals. Another possibility is that absorption at the intestinal or cellular level of Cd may be different. Other systemic factors such as exposure to other inducing agents may have effected a particular individuals ability to respond to Cd. Dietary factors may be important, as may be variables at the level of the kidney tubules themselves where reabsorption occurs. The possibility under study here is that variation in the base sequence of the MT gene promoter within or around the MRE and BLE consensus sequences may result in the observed variation in induction ratios. This could occur through differences in the ability of factors to bind to their regulatory regions, or changes in the DNA secondary structure, making the region more or less recognizable to the relevant factors. This is possible, although the apparent importance of MT in many different processes would make it unlikely in an evolutionary sense that significant variation could occur in the promoter region.

This study is also of general interest since variation in DNA sequences in 5' flanking promoter regions has not been determined before. Previous polymorphism studies have consisted of several types. Polymorphism has been examined in respect to electrophoretic migration of the protein gene products after isolation, which reflects the phenotype of the protein, but only allows for analysis of coding region changes which are not silent mutations (Walton et al, 1979). In order to detect a polymorphism with such a technique, the electrophoretic migration of the protein must be altered by a

base change which has resulted in a change in amino acid sequence, and this must result in a net change in charge for the protein. Polymorphisms can also be analyzed by digesting genomic DNAs from a large group of people with different restriction enzymes in the hope of finding restriction fragment length polymorphisms (RFLP) with one or more of the enzymes (Urso et al, 1988; Georges et al, 1987). The limitation of such an analysis is that even if many enzymes were used exhaustively, it is unlikely that one could analyze the complete base sequence in a given region to rule out all potential polymorphisms. Thus, as with the electrophoretic migration studies, RFLP analysis does not allow for actual nucleotide polymorphism determination, as a change may not necessarily result in a new restriction site for the enzymes in use. Analysis of genetic variation has also been done in studies which compared the rates of chemical and radiation induced mutations to rates of spontaneous mutation, but again identification of a base change is dependent on a change in phenotype which can be detected with the particular assay in use.

In 1983 Kreitman analyzed the ADH genes from 11 individuals from five natural populations of Drosophila melanogaster, by cloning gene copies from each and sequencing a 2,721 bp region. At that time only 63 bp of 5' sequence was reported. This study was reported again in 1991 with extension of the data 1243 bp into the 5' flank (Table 1.a) and 1425 bp into the 3' region (Kreitman and Hudson, 1991). Since the initial study, it was found that a new

gene, ADH-dup, which is believed to be an ancient duplication of ADH, was found in the 3' region with its transcript starting 124 bp downstream of the ADH gene (Table 1.b; Kreitman and Hudson, 1991; Schaefer and Aquadro, 1987). Although the precise promoter regulatory elements have not been mapped, for the Drosophila ADH gene, studies have been done to identify the important areas for transcriptional control (Posakony et al, 1985). Two promoters and two mRNA transcripts have been defined (Benyajati et al, 1983). ADH in Drosophila melanogaster is differentially regulated throughout its lifecycle, with the proximal mRNA transcript, produced by the more proximal promoter, found in high levels in late embryos and the larval stages, but in low levels in adults. The distal mRNA transcript, from by the distal promoter, is found at high levels in the adult insect, but is only present transiently in embryos and at low levels in the larval stages. The proximal promoter was found to be in the first intron immediately 5' to exon 2, which contains the translational start for both of the mRNA transcripts. The distal promoter is found in the 5' flank region immediately before the first exon, and its transcriptional start has been defined as the +1 position for the gene. Both promoters have their own TATAA sequences centered around -30 bp to each transcriptional start. Posakony et al (1985) showed that the sequences necessary for normal adult transcription were present between -69 and -660. The sequence necessary for larval transcript specificity was found in intron 1 from position 323 to 708, but transcription to normal levels was

dependent on a region lying over 2 kbp upstream.

The 1983 study identified 43 polymorphic sites, which were distributed throughout the gene as indicated (Table 1). The intron sequences appeared to have the highest polymorphism frequency, followed by the exons, the transcribed, untranslated regions, and the 3' flanking regions, respectively. The 5' flanking region was reported as being very high (4.7 %), but the -1, -2, -3 nucleotides were highly variable with the remainder of the positions being conserved. If the polymorphism frequency of the 60 bp 5' to these were not used, the 5' flanking region frequency of polymorphism was 0 %. This was the only region of the distal promoter presented in this study. The proximal promoter was found to have a low % polymorphism of 0.5.

The 1991 study was not different in respect to the rates of polymorphism found in the regions presented in the 1983 study. The extended 5' franking sequence was found to have a frequency of 2.4 % total. The data was reorganized for the purpose of this study into the two promoter regions, for the sake of comparison. The 660 bp distal adult promoter segment had a rate of change of 3.0 %, as compared to the proximal larval promoter sequence in the first intron which again was found to have rate of only 0.5 %. However, data was not presented on the region present approximately 2 kbp upstream which is believed to be important for this transcript to be produced normally. Different segments of sequence are presented in Table 1 for later comparison to the present study.

The results showed that the total 1243 bp analyzed had a similiar degree of variability as the non-coding regions and was substantially higher that the results found for the flanking regions combined in the 1983 study. The results were similiar in that the -1, -2, and -3 positions were again variable, but no other changes were seen in the next 225 bp of the 5' flank, indicating that the first 225 bp are apparently highly conserved between the five populations studied with the exception of the first three. This variation in -1, -2, and -3 was not seen for the proximal promoter. The sequence 5' of these 225 bp was not highly conserved. Unfortunately, the sequence was not analyzed in respect to the particular functional components of the promoter region, since the study was carried out in an attempt to establish the evolutionary history of the ADH gene in respect to polymorphism and divergence between the five populations. The data for the ADH-dup gene is also presented (1.b). Its rate of polymorphism was found to be significantly lower than that of the ADH gene itself, as was suggested by the 3' flanking sequence reported in the 1983 study.

The Drosophila ADH work is the only other study of nucleotide polymorphism to date, so is useful for comparison to this study. However, differences in the promoter structures and function, as well as differences in the function of the respective proteins in each study, and the fact that the ADH promoter has not been completely mapped out, all result in the limitation of the usefulness of such a comparison. Differences in the populations selected

Table 1: Summary of Drosophila Melanogaster Nucleotide Variation Studies by Kreitman (1983) and Kreitman and Hudson (1991).

Table 1.a):Data for the coding, noncoding and 5' flanking regions of the ADH gene from both studies

Summary of Kreitman's Drosophila Melanogaster studies, discussed in section 1.7. The average number of nucleotides sequenced and compared is presented for the clones isolated from the 11 individual strains. The number of polymorphic sites and % polymorphism are given for each region, as reported in Kreitman (1983) and Kreitman and Hudson (1991). The total nucleotides compared and polymorphic sites found for the coding and noncoding regions are presented to compare each region as a whole. The 5' flanking region data is separated between the two studies because of the difference in the extent of sequence presented between them. Each is presented in respect to the number of nucleotides and % polymorphism for the distal adult and proximal larval promoters as discussed in the text. The distal promoter in the 1983 study has a very high value of polymorphism, but the 3 changes in the region were found in the -1, -2, and -3 positions and only 63 bp were presented. The data is separated into the two promoter regions again for the 1991 study, but is also presented as the sequence and the first 225 bp of sequence for each. This is for the purpose of later discussion, as it was seen that this region of each promoter was relatively low in polymorphic sites. The full region refers to areas defined in deletion studies of the ADH promoter which were found to be required for normal transcription of the gene in the larval or adult stages as is discussed in the text. The data for the total 5' flanking region is also presented. The 3' flanking sequence data is presented in Table 1.b) because this area was identified as the ADH-dup gene after the 1983 study and was compared as such.

Table 1.b): Comparison of the data for the 3' flanking region (1983) and the ADH-dup gene (1991)

This table presents the average number of nucleotides compared and sequenced for the 3' flank in the 1983 study to the ADH-dup gene in the 1991 study which was found to occupy this region on later analysis. The number of polymorphic sites and the % polymorphic sites are presented for each. The ADH-dup gene is presented according to the identified 5' and 3' flanking, coding, and noncoding regions. The first 767 bp for the ADH-dup gene are presented with the total bases sequenced only for the sake of comparison to the 1983 3' flanking region.

Table 1: Summary of Drosophila Melanogaster Nucleotide Variation Studies by Kreitman (1983) and Kreitman and Hudson (1991).

Table 1.a):Data for the coding, noncoding and 5' flanking regions of the ADH gene from both studies

DNA Region		Average # of nucleotides compared	# of Polymorphic sites	% polymorphic sites
1. Coding	A .Exon 1	87	0	0 %
	B.Transl. exon 2	99	1	1.0 %
	C.Exon 3	405	4	1.0 %
	D.Transl. exon 4	264	9	3.5 %
TOTAL:		855	14	1.6 %
2. Non-Coding	A.3' untranslated	178	2	1.1 %
	B. Intron 1	620	11	1.8 %
	C. Intron 2	65	2	3.1 %
	D. Intron 3	70	5	7.1 %
TOTAL:		933	18	1.9 %
3. 5' Flank				
A. -->1983 study				
i) Distal promoter		63	3	4.8 %
ii) Proximal promoter		385	2	0.5 %
B. -->1991 study				
i) Total 5' flank		1243	30	2.4 %
ii) Distal promoter				
--> full region		660	20	3.0 %
--> first 225 bp		225	3	1.3 %
iii) Proximal promoter				
--> full region		385	2	0.5 %
--> first 225 bp		225	2	0.9 %

Table 1.b): Comparison of the data for the 3' flanking region (1983) and the ADH-dup gene (1991).

DNA Region:	Average # of nucleotides compared (bp)	# of polymorphic sites (in bp)	% polymorphic sites
1.3'Flank (1983 study data)	767	5	0.6 %
2.Adh-dup Gene (1991 study) i)Total bp ii)First 767 bp	 1425 767	 18 5	 1.3 % 0.6 %
a)5' Flank	124	3	2.4 %
b)Coding Regions i)Exon 1 ii)Exon 2 iii)Exon 3	 96 405 315	 0 2 3	 0 % 0.5 % 1.0 %
c)Noncoding Regions i)Intron 1 ii)Intron 2	 376 51	 2 7	 0.5 % 13.7 %
d)3' Flank	58	1	1.7 %

from for the current study and the Drosophila ADH studies also limit the usefulness of this comparison. The current study is done from the regional population, but whether it can be called an individual population is difficult, since the people in the study were from a diverse range of backgrounds. Thus, only one study has been done in the past on nucleotide polymorphism in particular, and these studies were carried out in insects. The present study would be the first of its kind in humans.

Another study analyzed the glucose-6-phosphate dehydrogenase locus in humans by restriction fragment length polymorphism, using 14 unique probes and 18 restriction enzymes on four main populations. Analysis of a 100 kbp region resulted in the identification of only one polymorphism (D'Urso et al, 1988). However, it is ironic that only one was found since over 300 variants of the enzymes exist (Beutler and Yoshida, 1988). This shows the limitation of RFLP analysis and again indicates the usefulness and interest in comparing a portion of 5' flanking promoter region from a group of individuals to see if significant variation exists, or if the region is highly conserved.

Thus, in order to determine nucleotide polymorphism in a promoter region, the MT-II_A gene was cloned from ten individuals and sequenced from the first exon to the GRE promoter sequences.

2. MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Enzymes:

The restriction endonucleases utilized were obtained from Pharmacia, Gibco/BRL, and New England Biolabs. T4 DNA ligase was purchased from Gibco/BRL and DNA polymerase I Klenow fragment was obtained from Gibco/BRL and Pharmacia. The sequencing kit was originally obtained from United States Biochemicals, but most of the work was done with Pharmacia's T7 DNA polymerase system.

2.1.2 Cloning Vectors and Host Strains:

EcoRI digested, dephosphorylated lambda gt11 arms were obtained from Promega Biotech as part of the Protoclone system. In vitro lambda-bacteriophage packaging extracts were purchased from Promega Biotech as the Packagene system and Stratagene as the Gigapack Gold system. The host strain utilized for lambda gt11 bacteriophage was the Y1090 strain received from Promega Biotech with the Protoclone system.

The vector pUC118 which was utilized for subcloning was obtained from Dr. A.B. Futcher (Department of Biochemistry, McMaster University), who originally acquired them from Dr. J. Viera (Viera and Messing, 1987), as was the host strain MV1193 which was used for the pUC plasmid transformations.

2.1.3 The MT-II_A Promoter Region:

The plasmid pHS1b was obtained from Dr. M. Karin (Dept. of Microbiology, University of Southern California, School of Medicine) which consists of the MT-II_A promoter region from the 5' HindIII site to the 3' BamHI site, cloned into the HindIII and NcoI sites of pUC8. See Figure 2 for an illustration of this construct.

2.1.4 Oligonucleotides:

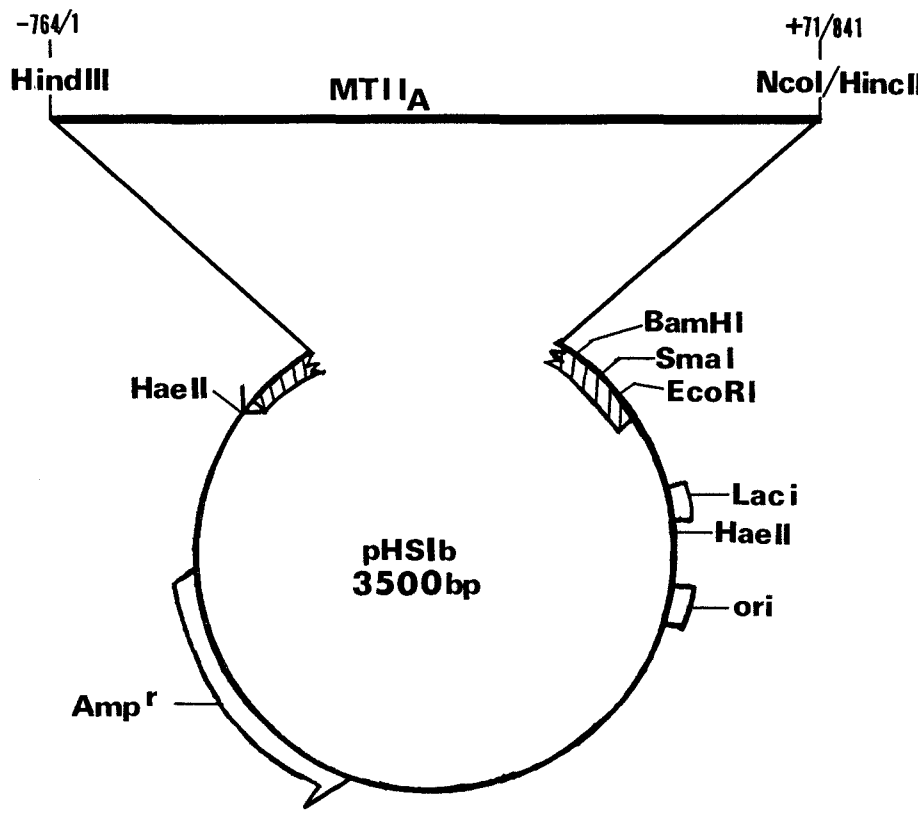
The primers utilized for sequencing were generated by the Central Facility of the Institute of Molecular Biology and Biotechnology, McMaster University, using an Applied Biosystems Model 380B DNA synthesizer. Oligonucleotides were desalted and purified on an FPLC gel prior to their acquisition, and ranged in size from 17 to 23 nucleotides in length. All of the primers are listed in Table 2. Their annealing positions within the MT-II_A promoter region are illustrated in figure 1 and 4.

2.2 METHODS

2.2.1 Nucleic Acid Digestion and Analysis:

Unless otherwise indicated methods carried out were as described in Maniatis et al (1983), Davis et al (1986), or Ausubel et al (1987). Product profile sheets and kit manuals were followed according to the suppliers. All buffers, solutions, and reactions conditions used can be found in the above sources.

Figure 2: pHS1b plasmid



The plasmid was constructed and generously supplied by Dr. M. Karin (Dept. of Microbiology, University of Southern California, School of Medicine). This construct contains the MT-II$_A$ promoter region from -764 (1) to +71 (841), from the 5' HindIII site to the 3' NcoI site which is located at the beginning of exon 1. As a result, it includes 71 bp of transcribed, untranslated MT-II$_A$ (see figure 1). This fragment was ligated with HindIII and HincII digested pUC8 to make pHS1b. The extrapolated area in the figure indicated the MT-II$_A$ DNA sequence. The hatched region indicates the remaining Lac Z' region which contains the multiple cloning sites of pUC8. The area containing the Lac Z' region between the two HaeII sites was derived from M13mp sequence. The remainder of the sequence was derived from pBR322.

Table 2: List of oligonucleotides

All of the MT oligonucleotides utilized for DNA sequencing of the isolated clones are defined, as well as their strand in which they are found, and their sequence (see section 2.1.4). The non-bracketted numbers indicate the position of the oligonucleotide ends in respect to Microgenie (Genbank) nucleotide sequences. The numbers in brackets refer to the nucleotide ends in respect to the +1 transcriptional start site. These numbering systems, as well as the oligonucleotide binding sites and orientations, are illustrated in figure 1.

oligo/strand		sequence	ends
MTAP	(-)	5'-GGGCAGCAGGAGCAGCAGCT-3'	1440 1459 (+689 +670)
MTAP5'	(-)	5'-CGCAGGAGCAGTTGGG-3'	846 862 (+76 +93)
MT2A1	(-)	5'-CAGGTTGCACGCGGGTTCGCTGG-3'	788 810 (+19 +41)
MT2A1'	(+)	5'-GAACCCGCGTGCAACCTG-3'	793 810 (+24 +41)
MT2A2	(+)	5'-GCGTCCCCGAGGCGCAAGTGGGC-3'	548 570 (-223 -200)
MT2A3	(-)	5'-GTCTACACTGGGCATCCCCAGCC-3'	875 897 (+134 +96)
MT2A4	(+)	5'-ACGGCGGAGGCGCACGGCGTGGG-3'	469 491 (-302 -279)
MT2A5	(-)	5'-CGCGCTGAGTCACTTG-3'	661 676 (-110 -95)
MT2A6	(+)	5'-TTCGCTTGGAGCCGCA-3'	647 662 (-124 -109)
MT2A7	(-)	5'-ACTTGCGCCTCGGGGACG-3'	559 566 (-222 -205)
MT2A8	(+)	5'-ATTAAGCCCTGGCTGC-3'	330 345 (-441 -426)

2.2.2 Southern Analysis:

Southern analysis was typically carried out with 10-15 ug of human genomic DNA isolated from peripheral lymphocytes as described in section 2.2.5. DNA was digested with the appropriate restriction endonuclease and electrophoresed on 0.8-1.5 % 1 X TBE agarose gels at 20-30 V for 12-24 h. The gel was soaked in 0.5 ug/ml ethidium bromide and exposed to ultraviolet irradiation to nick the DNA fragments, before transfer to a nitrocellulose (Schleicher and Schuell, Gelman Sciences) or nylon membrane (Zeta-probe, Biorad) in 6 X SSC according to the method described by Maniatis et al (1982). Blots were baked at 80°C under vacuum for 2.5 h before probe hybridization as described in section 2.2.3 and 2.2.4.

2.2.3 Labelling and Hybridization of DNA Probes:

DNA fragments were generated by restriction endonuclease digestion using the appropriate enzymes. Restriction digests were electrophoresed on 1 % agarose gels in 0.5 X TBE buffer required to separate the expected bands. Gels were stained in 0.5 ug/ml ethidium bromide and the required DNA fragments were excised for electroelution from the agarose. Agarose strips were placed in dialysis tubing and electrophoresed in 0.5 X TBE buffer for 2 h at 200 V. The buffer from the dialysis bag was then extracted with

phenol: chloroform: isoamyl alcohol (25: 24: 1) precipitated in ethanol and 0.3 M NaOAc (pH 4.5), and dissolved in TE buffer. Fragment isolation was also carried out using the Geneclean kit (Bio/Can Scientific) as described by Vogelstein and Gillespie, 1979, but electroelution was used preferentially.

Purified DNA fragments were radiolabelled using the random primer method of Feinberg and Vogelstein, 1983. Labelling was usually achieved to 10^8-10^9 Cerenkov cpm/ ug fragment, depending on the DNA preparation used.

2.2.4 Hybridization:

Nitrocellulose filters used for the purposes of genomic screening or Southern DNA analysis were typically prehybridized for 3-18 h at 48°C with a solution containing 50 % formamide, 5 X SSC, 5 X Denhardt's solution, 250 ug/ml sheared and denatured salmon sperm DNA, and 0.1 % w/v SDS. Zeta-probe filters were prehybridized with 50 % formamide, 0.25 M $NaHPO_4$ (pH 7.2), 0.25 M NaCl, 7 % SDS, 1 mM EDTA, and 250 ug/ml salmon sperm DNA as indicated in the Zeta probe instruction manual (Biorad). Hybridization was carried out with approximately 1-2 x 10^5 cpm of probe/ 82 mm diameter, 0.45 um nitrocellulose filter for bacteriophage-lambda gt11 genomic screening or 10^6-10^7 cpm probe/ filter for Southern DNA analysis. Probes for hybridization were denatured by boiling in an adequate volume of hybridization buffer before addition to the filters. Hybridization was carried out for 18-24 h at 48°C, after which the filters were washed twice in 2 X SSC, 0.1 % SDS for 10 min at

room temperature, and twice in 0.1 X SSC, 0.1 % SDS for 20 min at 48°C. Washed filters were allowed to dry at room temperature on 3MM paper and were exposed for 2-5 days to Kodak XAR-5 film at -70°C with an intensifying screen, depending on the experiment. Zeta-probe filters were wrapped in Saran Wrap immediately after washing if re-probing was necessary. The stripping of the probes from filters was carried out by boiling the filters for 10 min in 0.1 X SSC. The filters were usually exposed to film before reprobing to ensure successful removal removal of the probe used before.

2.2.5 Isolation and Purification of Genomic DNA from Peripheral Lymphocytes:

Approximately 30-50 ml of blood was obtained from different individuals by venipuncture using heparinized Vacutainer tubes. Fifteen ml of blood was diluted with an equal volume of PBS buffer in 50 ml Falcon 2098 tubes and underlaid with 20 ml of a Ficoll-Hypaque solution. This was spun for 30 min at 2200 rpm in a Sorvall GLC-2B centrifuge and allowed to stop without braking to avoid disturbing the gradient. The white, opaque peripheral lymphocyte layer was removed into a new 50 ml Falcon 2098 tube using a pasteur pipette, and was diluted with three volumes of PBS buffer. The cells were pelleted by spinning for 10 min at 2000 rpm. This washing procedure was repeated and the cells were finally resuspended in 10 ml of 0.1 M NaCl, 0.05 M Tris-HCl (pH 7.5), 0.001 M EDTA, 0.5 mg/ml Proteinase K (BRL), and 0.5 % SDS. This was

incubated overnight at 37°C to ensure complete protein degradation, after which the mixture was extracted three times with phenol: chloroform: isoamyl alcohol (25: 24: 1) for 10 min and twice with chloroform: isoamyl alcohol (24: 1) for 5 min, to remove cell debris and peptides. Sodium acetate was added to the final aqueous phase to a final concentration of 0.3 M and 2.5 volumes of absolute ethanol were added to precipitate the DNA. After centrifugation of the sample at 12,000 g for 30 min, the DNA pellet was washed with 70 % ethanol and resuspended in 0.5 ml of TE. With some preparations it was necessary to re-extract the DNA with phenol: chloroform and re-precipitate several times in order to obtain DNA clean enough for restriction digestion and cloning.

2.2.6 Preparation of Genomic DNA for Bacteriophage-lambda gt11 Library Construction:

Genomic DNA was purified from a series of human individuals as described in Section 2.2.5. Approximately 100-150 ug was digested with 1 U EcoRI/ug of DNA for 18-24 h at 37°C. Digested DNA was then placed on top of a 10-40 % sucrose density gradient and spun in a Sorvall TST 41.14 rotor at 27,000 rpm for 20-24 h at 20°C in order to separate the DNA fragments on the basis of size. Gradient tubes were then punctured in the bottom and 36 fractions of approximately 0.3 ml each were taken. Fractions containing DNA fragments in the 4-10 kbp range were pooled, diluted with TE buffer to ensure

a sucrose concentration of less than 10 %, and precipitated with sodium acetate and absolute ethanol. The purified DNA was dissolved to a concentration of 0.3-1.0 ug/ ul for the purpose of cloning into the bacteriophage lambda gt11 vector.

2.2.7 Construction of Human Genomic Bacteriophage-lambda gt11 Libraries:

The gt11 bacteriophage-lambda vector was used for the cloning of the MT-II_A genes in order to analyze the promoter regions. One ug of EcoRI digested, dephosphorylated lambda gt11 arms (Promega Biotech) were ligated with 1-2 ug of sucrose gradient enriched EcoRI digested human genomic DNA (see section 2.2.6), under the conditions described in the T4 DNA ligase product data sheets (Gibco/BRL, Pharmacia). Ligations were left overnight at 4°C with 2-3 U of ligase in 5-10 ul total volume and were then packaged into virulent phage using either the Packagene (Promega Biotech) or the Gigapack Gold (Stratagene) in vitro packaging extract systems. 1-1.5 X 10^4 phage were added to 200 ul of the bacterium E. coli strain Y1090, which was grown to an optical density of 0.7-0.9 at 550 nm. Phage adsorption was carried out for 30 min at 37°C, the cells were added to 3 ml of YT + agarose at 46°C, and plated on YT agar plates in the presence of 100 ug/ul ampicillin. The plates were incubated at 37°C for 6-8 h to allow the phage plaques to become nearly confluent, were cooled to 4°C, and were transferred to nitrocellulose in duplicate. Platings were done with 10 ul of 100 mM IPTG and

10 % X-gal for blue/white lacZ colour selection when the background of non-insert lambda-gt11 phage were to be determined. Plaque transfers were denatured in 1.5 M NaCl, O.5 M NaOH for 2 min, neutralized in 1.5 M NaCl, O.5 M Tris-HCl (pH 8.0) for 5 min, and washed in 0.2 M Tris-HCl (pH 7.5), 2 X SSC for 1 min. Filters were dried on 3MM paper under a heating lamp and baked at 80°C under vaccuum for 2.5 h. The plates were kept at 4°C until positive plaques could be picked from the screening results.

2.2.8 Screening of Human Genomic Bacteriophage-lambda gt11 Libraries:

Screening for MT-II_A genomic clones was carried out using the HindIII-EcoRI fragment of the plasmid pHS1b (see section 2.1.3). This fragment contains 835 bp of the MT-II_A promoter region, as illustrated in figure 2. The fragment was labelled by the random primer labelling technique (see section 2.2.3). The prehybridization, hybridization, and filter washing steps were carried out as described. Positive clones were identified by autoradiography, plugs containing the appropriate agar regions were picked from the plates, and the phage were eluted into 1 ml of SM buffer. Plaque elutions were plated as above and screened until single positive plaques were isolated. Once single positive phage were isolated, DNA was isolated from each using a small-scale plate-lysate method (Davis et al, 1987). DNA preparations were digested with EcoRI to define the insert size, as well as for the purpose of

subcloning. Southern analysis was also carried out on some of these digested lambda gt11 clones to ensure that MT-II_A sequences were present within the clone.

2.2.9 Subcloning of MT-II_A Bacteriophage-Lambda gt11 Clones:

Lambda gt11 EcoRI-digested cloned DNA was ligated with the EcoRI cut plasmid vector pUC118 at a molar ratio of 3:1 insert to vector in 20 ul. Ligations were carried out at 4°C overnight, according to the product data sheets for T4 DNA ligase (Gibco/BRL).

Bacterial transformations were carried out according to the method of Hanahan (1983). Prior to transformation 2 ul of 1 M MES (pH 6.3) was added to the ligation mixture in order to alleviate the problem of inhibition of transformation seen with ligation solutions having too alkaline a pH (Dr. A.B. Futcher, personal communication). A fraction (1/10 to 1/2) of the transformed cells was plated on YT + ampicillin agar plates with 10 ul of 100 mM IPTG and 10 ul of 10 % X-gal for blue/ white colour selection. White colonies were grown up in 2 ml liquid cultures and plasmids were isolated from each using an alkaline-lysis mini-screen protocol (Maniatis et al, 1982). EcoRI insert sizes were determined for each and cultures containing cells with plasmids containing 6 kbp inserts were grown-up on a large scale. Large scale plasmid preparations were originally carried out according to Pulleyblank's method (Pulleyblank et al, 1983), which utilized a pronase/ SDS cell lysis and DNA

purification through LiCl and PEG precipitations. Later, an alternate method was utilized (Personal communication, D. Andrews, McMaster University). Alkaline-lysis was used for cell lysis, followed by isopropanol, LiCl, and PEG precipitations, as well as phenol/chloroform extractions and ethanol precipitations. This alternate protocol was used for the majority of the large scale plasmid preparations that were done, since it gave greater yields with greater consistency. It also served to produce cleaner templates for sequencing purposes.

2.2.10 Sequencing of MT-II_A Bacteriophage-lambda gt11 Clones:

DNA sequencing was carried out according to the Sanger dideoxynucleotide incorporation method. Two to three ug of plasmid DNA was mixed with 1 ul of 5 N NaOH and 1 ul of 4 mM EDTA to a total volume of 20 ul. The DNA was precipitated in absolute ethanol and 5 M ammonium acetate and resuspended for annealling according to the T7 DNA polymerase sequencing kit manual (Pharmacia). Originally the USB T7 DNA sequencing kit was used, but the Pharmacia kit was found to give superior sequencing quality with the templates used. The majority of the sequencing was done with the Pharmacia system. Annealing reactions were carried out with 1-5 pmol of primer. The eight MT primers used allowed for the determination of the sequence from approximately nucleotide 320-370 to nucleotide 830-870 in both strands spanning the region from the GRE to the +1 transcriptional

initiation site, which includes the MRE regions (see Results Section 4).

Sequencing reactions were carried out according to the sequencing manual (Pharmacia) and were electrophoresed on 8 % acrylamide, 7 M urea denaturing gels after denaturation. Between 150 and 300 nucleotides of sequence were determined for each reaction depending on the template and primer in question. Sequencing gels were dried and exposed to Kodak XK-1 film for 18-24 h.

The sequencing gels were read and recorded on the Microgenie Sequence Software System (Beckman). To determine nucleotide changes in the MT-II_A promoter sequence, the MT-II_A sequence from the Microgenie data base was compared by hand to the sequencing gels. Changes were checked in both strands and against the published sequence to identify base changes in the individual clones, and any mistakes in the published sequences.

2.2.11 Verification of the Identified Polymorphisms by Restriction and RFLP Analysis:

In order to verify the identified changes, restriction analysis of the pUC118 clones was first carried out with the appropriate enzymes to see if the new restriction site expected from the mutuation was actually present. This confirmed the polymorphisms in the cloned DNA fragments.

Attempts were then made to analyze DNA from the individuals via RFLP analysis. Genomic DNA purified according to section 2.2.5 was digested with

the appropriate restriction endonuclease, selected from the restriction analysis carried out above in accordance with the expected polymorphism. The DNA was then processed according to the southern analysis method described in section 2.2.3 and 2.2.4. The DNA was transferred and hybridized with the pHS1b MT-II_A promoter fragment (see section 2.1.3), and exposed to film for the required time.

3. RESULTS

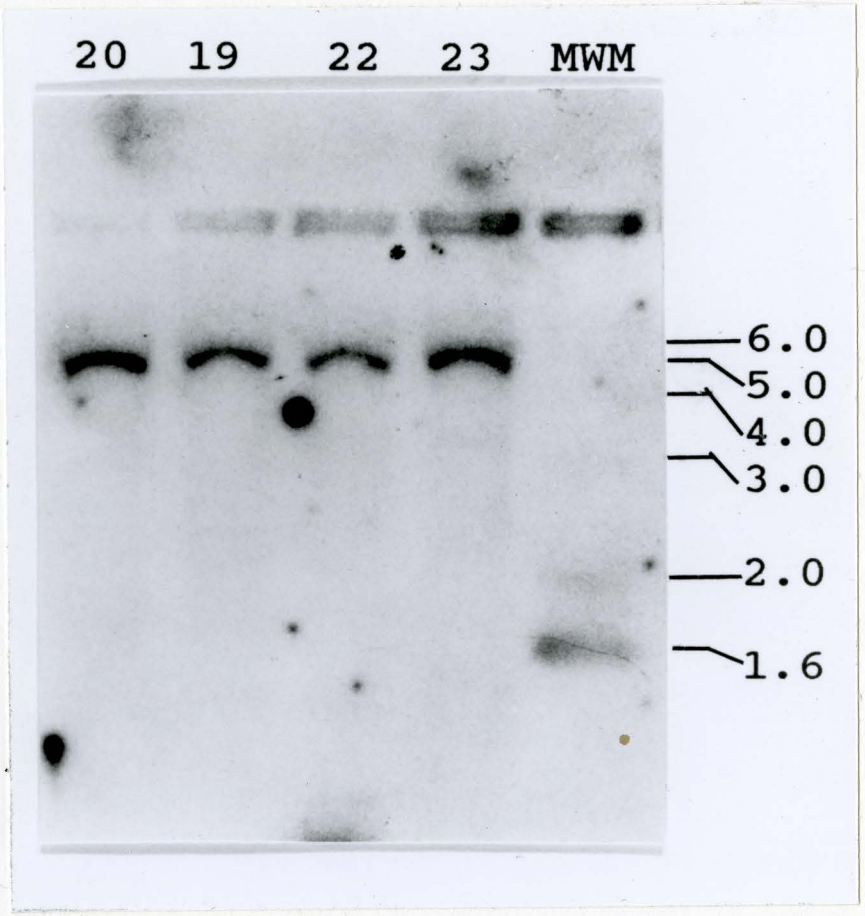
MT-II_A genomic sequences were obtained from 12 different individuals and sequenced accordingly. The sequencing results which are discussed here indicated the existence of discrepancies between the MT-II_A promoter region sequence published in papers previously and the sequence identified here. As well, the promoter region was found to be highly conserved, with only one polymorphic site identified in two individuals in this study.

3.1. Cloning:

Human genomic DNA was isolated from peripheral lymphocytes according to section 2.2.5, digested with EcoRI, and analyzed by southern analysis (section 2.2.2) to confirm the presence of sequences homologous to the MT-II_A promoter. Hybridization was done with the pHS1b HindIII-EcoRI fragment (section 2.2.3 and Figure 1). The expected size of the EcoRI fragment containing the entire MT-II_A gene promoter and coding regions was 6.1 kbp, according to the restriction map published by Karin et al (1984; Figure 1). A strong signal was seen between 5 and 6.0 kbp, indicating the presence of the expected MT-II_A genomic EcoRI fragment in the purified DNA (Figure 3).

EcoRI digested DNA was size fractionated according to section 2.2.6, and utilized for lambda- gt11 genomic cloning (section 2.2.7). Two or more independent clones were isolated for 9 individuals, and one genomic clone

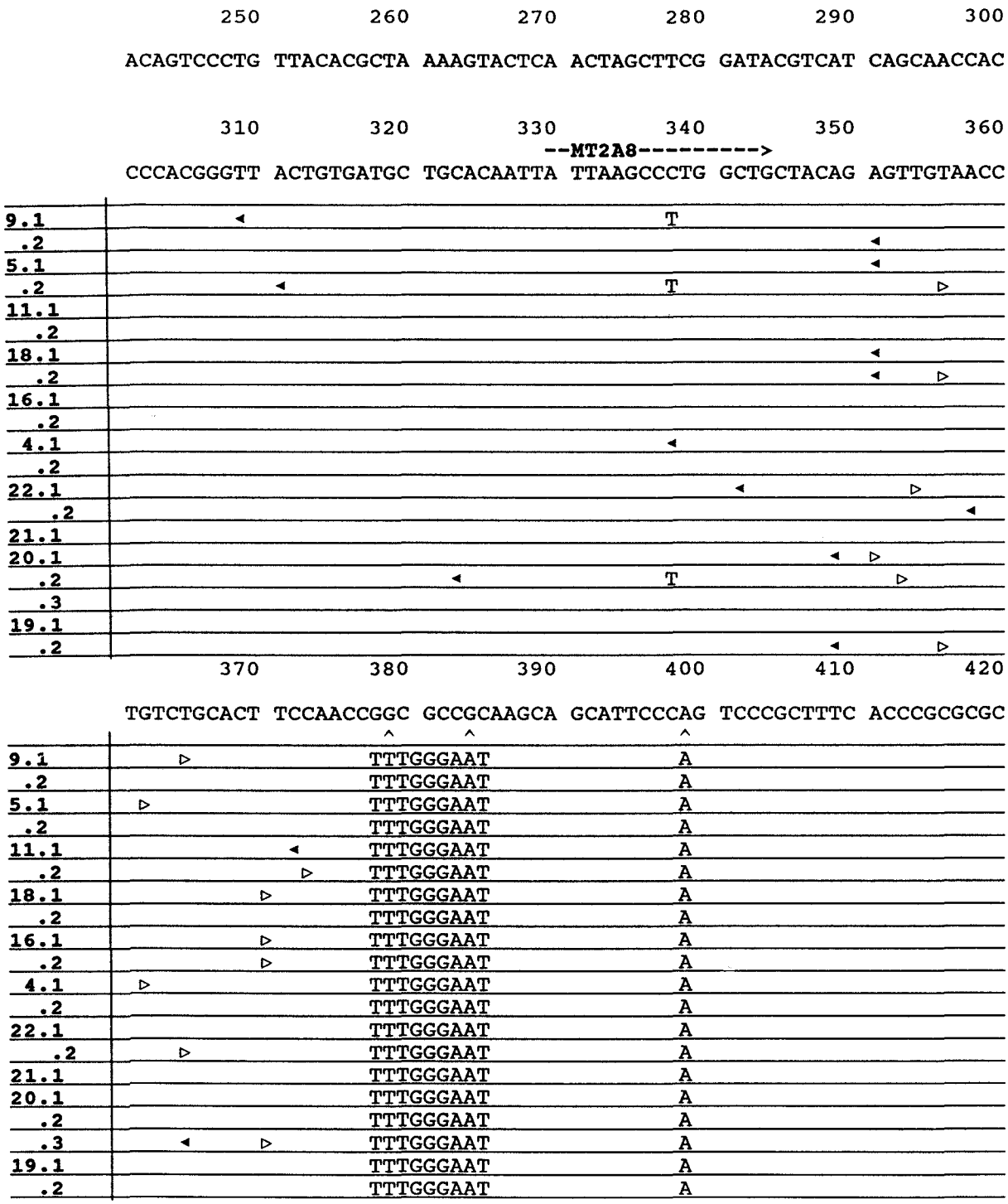
Figure 3: EcoR1 Genomic Southern Blot Analysis

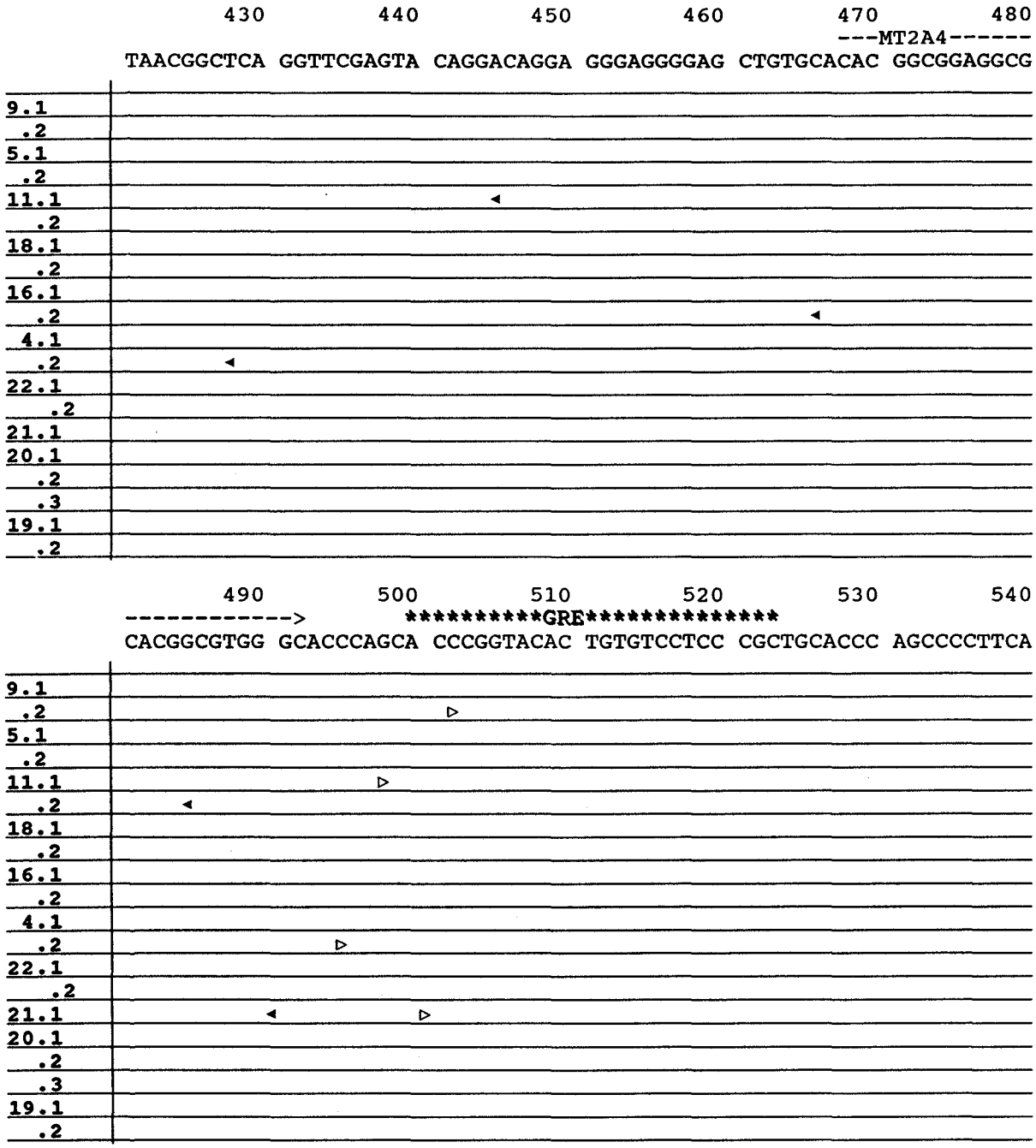


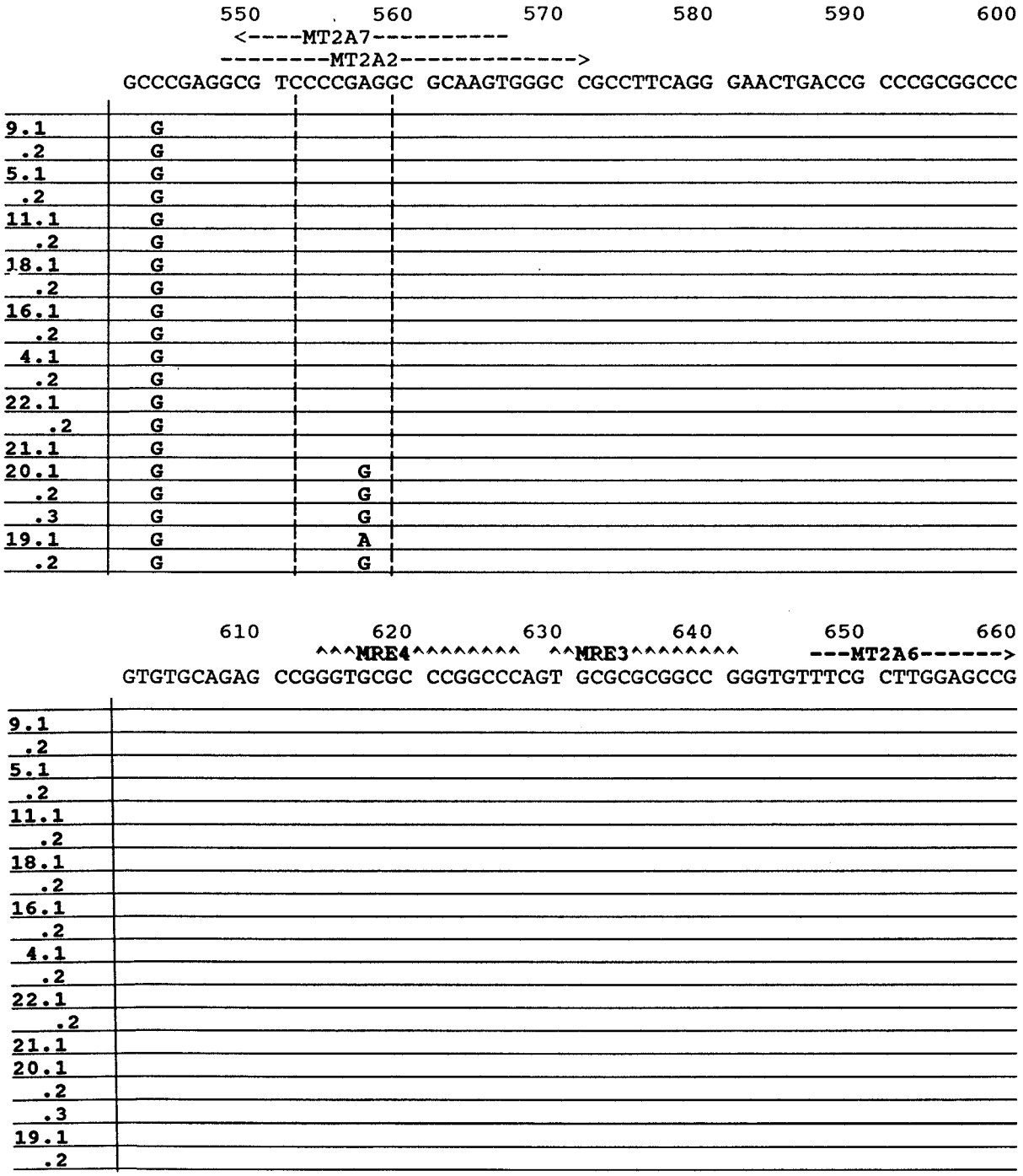
Genomic DNA from donors 19, 20, 22, and 23 were digested with EcoR1, electrophoresed on a 1 %, 1 X TBE agarose gel, transfered to nitrocellulose, and probed with the pHS1b MT-II_A promoter fragment. Hybridization is seen with a 5-6 kbp band, corresponding to the expected EcoR1 genomic MT-II_A fragment of 6 kbp. The 1 kbp ladder size markers (MWM) are as indicated.

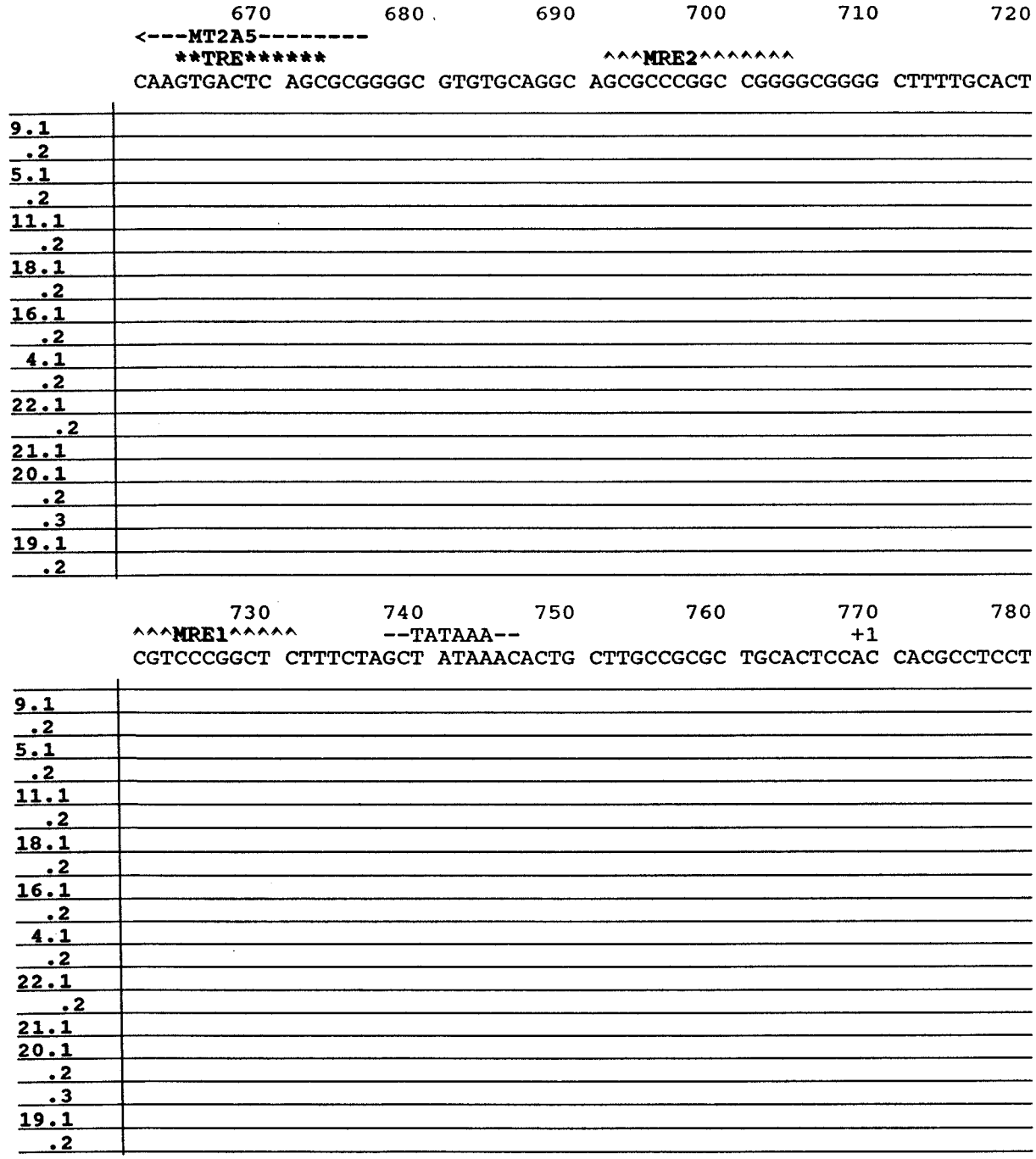
Figure 4: Sequencing summaries for all individuals.

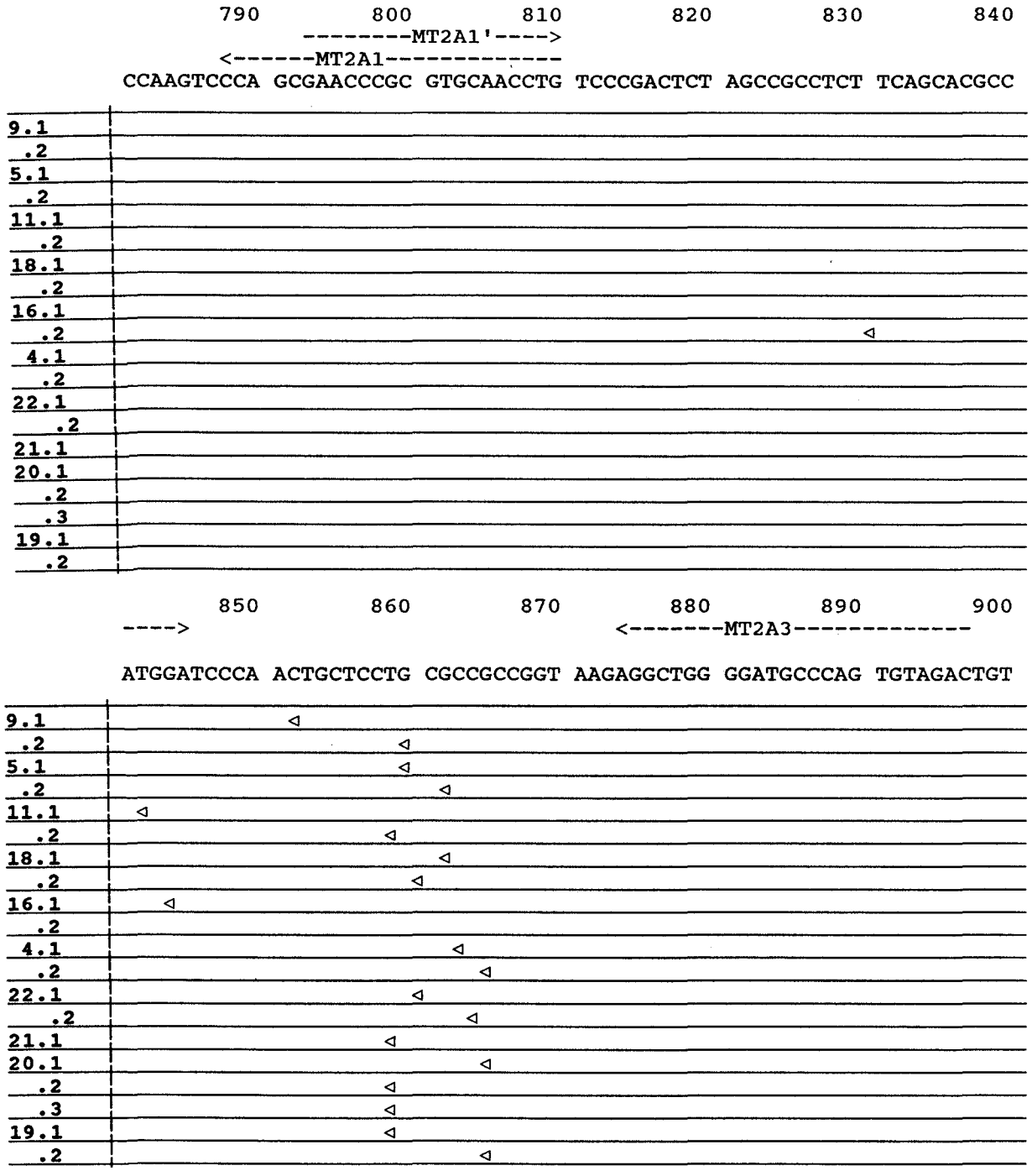
The MT-II_A nucleotide sequence is seen from position 241 to 1080 (see Microgenie (Genbank) nucleotide numbering system, figure 1). The positions of the responsive elements and other regulatory sequences are shown, as are the positions of the sequencing oligonucleotides listed in table 2, and illustrated in figure 1. Each clone is listed to the left of the sequence, and changes found for each from comparison with the published sequences (Karin et al, 1984; Karin et al, 1987) are noted below the original sequence. Single bp changes are found directly below the specific position. Nucleotides in the original sequence, but not found in the obtained sequence are indicated by an X. Nucleotides found in the sequence obtained in this study but not found in the orginal sequence are indicated by the presence of arrows found beneath the sequence in the figure. Double arrows indicate an inserted stretch of nucleotides. Thus, the sequence at position 378 to 385 was read as GGTTTGGGAATGC in the sequences obtained in this study. The sequence at 987 to 988 should read GGGG as opposed to GG as indicated in the original sequence. The total extent of sequencing in each strand is also shown for each promoter cloned. Open arrows pointing to the right indicate the start of obtained sequence in the (+) strand and the black arrows in the same direction indicate the end of the sequence obtained in the same strand. The arrow system is the same for the (-) strand, except they point to the left. From this, areas sequenced in only one strand, or areas not sequenced, can be determined for each clone.


250 260 270 280 290 300
ACAGTCCCTG TTACACGCTA AAAGTACTCA ACTAGCTTCG GATACGTCAT CAGCAACCAC
310 320 330 340 350 360
--MT2A8--------->
CCCACGGGTT ACTGTGATGC TGCACAATTA TTAAGCCCTG GCTGCTACAG AGTTGTAACC
9.1
.2
5.1
.2
11.1
.2
18.1
.2
16.1
.2
4.1
.2
22.1
.2
21.1
20.1
.2
.3
19.1
.2
370 380 390 400 410 420
TGTCTGCACT TCCAACCGGC GCCGCAAGCA GCATTCCCAG TCCCGCTTTC ACCCGCGCGC
TTTGGGAAT
A


430
440
450
460
470
480
---MT2A4------
TAACGGCTCA GGTTCGAGTA CAGGACAGGA GGGAGGGGAG CTGTGCACAC GGCGGAGGCG
9.1
.2
5.1
.2
11.1
.2
18.1
.2
16.1
.2
4.1
.2
22.1
.2
21.1
20.1
.2
.3
19.1
.2
490
500
510
520
530
540
---------->
**********GRE**********
CACGGCGTGG GCACCCAGCA CCCGGTACAC TGTGTCCTCC CGCTGCACCC AGCCCCTTCA
9.1
.2
5.1
.2
11.1
.2
18.1
.2
16.1
.2
4.1
.2
22.1
.2
21.1
20.1
.2
.3
19.1
.2

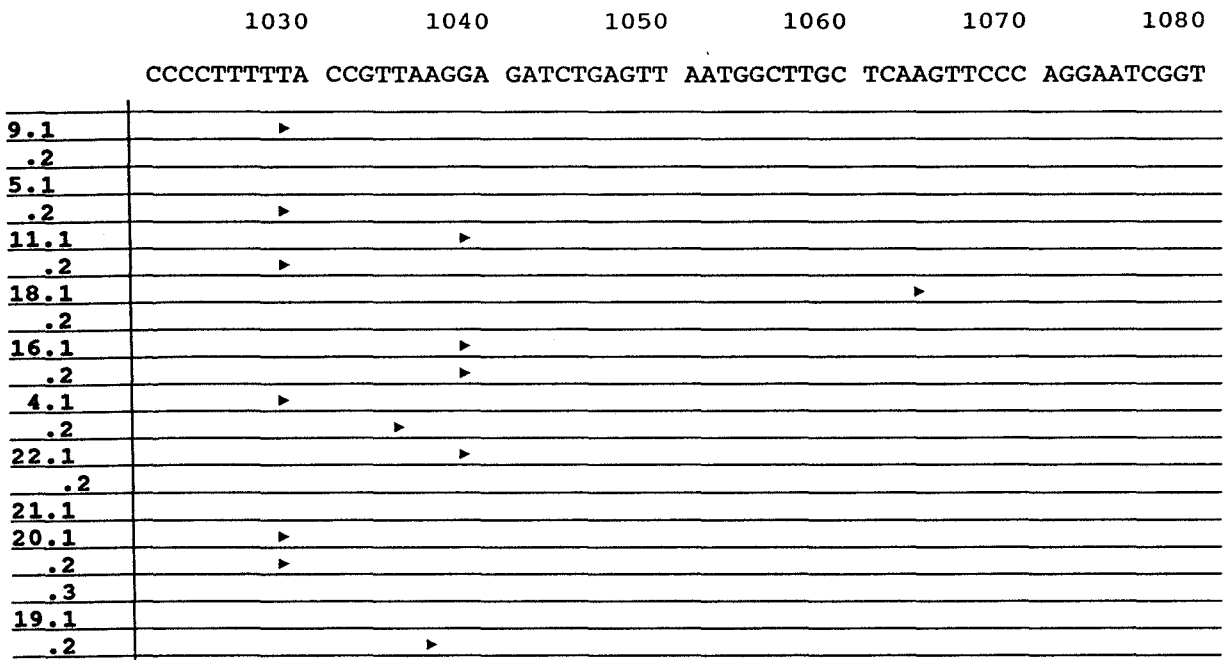

550 560 570 580 590 600
<----MT2A7----------
--------MT2A2------------>
GCCCGAGGCG TCCCCGAGGC GCAAGTGGGC CGCCTTCAGG GAACTGACCG CCCGCGGCCC
9.1 G
.2 G
5.1 G
.2 G
11.1 G
.2 G
18.1 G
.2 G
16.1 G
.2 G
4.1 G
.2 G
22.1 G
.2 G
21.1 G
20.1 G G
.2 G G
.3 G G
19.1 G A
.2 G G
610 620 630 640 650 660
^^^MRE4^^^^^^^^ ^^MRE3^^^^^^^^ ---MT2A6------>
GTGTGCAGAG CCGGGTGCGC CCGGCCCAGT GCGCGCGGCC GGGTGTTTCG CTTGGAGCCG
9.1
.2
5.1
.2
11.1
.2
18.1
.2
16.1
.2
4.1
.2
22.1
.2
21.1
20.1
.2
.3
19.1
.2


670 680 690 700 710 720
<---MT2A5--------
TRE****
^^^MRE2^^^^^^^
CAAGTGACTC AGCGCGGGGC GTGTGCAGGC AGCGCCCGGC CGGGGCGGGG CTTTTGCACT
9.1
.2
5.1
.2
11.1
.2
18.1
.2
16.1
.2
4.1
.2
22.1
.2
21.1
20.1
.2
.3
19.1
.2
730 740 750 760 770 780
^^^MRE1^^^^^
--TATAAA--
+1
CGTCCCGGCT CTTTCTAGCT ATAAACACTG CTTGCCGCGC TGCACTCCAC CACGCCTCCT
9.1
.2
5.1
.2
11.1
.2
18.1
.2
16.1
.2
4.1
.2
22.1
.2
21.1
20.1
.2
.3
19.1
.2


790
800
810
820
830
840
--------MT2A1'---->
<------MT2A1------------
CCAAGTCCCA GCGAACCCGC GTGCAACCTG TCCCGACTCT AGCCGCCTCT TCAGCACGCC
9.1
.2
5.1
.2
11.1
.2
18.1
.2
16.1
.2
4.1
.2
22.1
.2
21.1
20.1
.2
.3
19.1
.2
850
860
870
880
890
900
---->
<-------MT2A3------------
ATGGATCCCA ACTGCTCCTG CGCCGCCGGT AAGAGGCTGG GGATGCCCAG TGTAGACTGT
9.1
.2
5.1
.2
11.1
.2
18.1
.2
16.1
.2
4.1
.2
22.1
.2
21.1
20.1
.2
.3
19.1
.2

	910	920	930	940	950	960
	AGCGCTAGAG	AAGCAATTTC	TGACCCCTCT	TTCTTTCTCT	GGTCACTCAA	TTTCAGGCAC
						^
9.1						X
.2						X
5.1						X
.2						X
11.1						X
.2						X
18.1						X
.2						X
16.1						X
.2						X
4.1						X
.2						X
22.1						X
.2						X
21.1						X
20.1						X
.2						X
.3						X
19.1						X
.2						X

	970	980	990	1000	1010	1020
	AGGAGTTGCT	CTTCCCAAAG	AGTTTTGGTA	TCTTTCTCTC	CATTCTAGGT	TATTCGGAGC
	^		^^			
9.1	C		GG			
.2	C		GG ▸			
5.1	C		GG			▸
.2	C		GG			
11.1	C		GG			
.2	C		GG			
18.1	C		GG			
.2	C		GG		▸	
16.1	C		GG			
.2	C		GG			
4.1	C		GG			
.2	C		GG			
22.1	C		GG			
.2	C		GG			
21.1	C		GG			▸
20.1	C		GG			
.2	C		GG			
.3	C		GG			
19.1	C		GG			
.2	C		GG			▸


1030
1040
1050
1060
1070
1080
CCCCTTTTTA CCGTTAAGGA GATCTGAGTT AATGGCTTGC TCAAGTTCCC AGGAATCGGT
9.1
.2
5.1
.2
11.1
.2
18.1
.2
16.1
.2
4.1
.2
22.1
.2
21.1
20.1
.2
.3
19.1
.2

was isolated for a tenth individual (Figure 4). Two independent clones were subcloned into pUC118 (section 2.2.9) for the purposes of sequencing, with the exception of donor 20 for which three independent clones were subcloned, and donor 21 for whom only one independent clone was isolated. In total, twenty independent clones were subcloned for sequencing.

3.2. Sequencing:

pUC118 subcloned DNA was prepared (section 2.2.9) and sequencing was carried out using the different primers indicated in Table 2 (section 2.1.4), according to the protocols given in section 2.2.10. These primers were arranged such that the sequence could be obtained for each clone in both strands from approximately 140 bp 5' to the GRE consensus to the +1 transcriptional initiation site (Figure 1). The MT2A1' primer was utilized to sequence from the promoter through the +1 site, the first exon, and into the first intron, to give some indication of the variation in these regions. These sequences were not confirmed in the opposite strand.

The extent of sequencing, primer binding sites and regulatory element consensus sequences can be seen in Figure 4, along with the sequence changes determined in this study. All subclones were sequenced in both strands from the 5' end of the GRE consensus to the MT2A3 primer binding site (from position 500 to 860, page E), spanning the MREs, TRE, TATAAA box, and transcriptional and translational start sites. All of the subclones,

except for donor 21, were sequenced in one strand from the MT2A3 primer binding site to position 1020 to 1070 using MT2A1'. The region from MT2A8 to the GRE consensus (page A and B) was sequenced in both strands for subclones 4.1, 5.1, 5.2, 9.1, 18.1, 18.2, 19.1, 19.2, 20.1, 20.2, 22.1, and 22.2. Adequate sequencing results were never obtained with 21.1 or 4.2 with the MT2A7 or MT2A8 primers, even though several different large scale preparations were used. The reason for this is not known. However, the sequencing reactions with these primers provided poorer quality sequence data as compared to the other DNA samples. It could be that their hybridizations required more exact conditions, making them more sensitive to contaminating proteins or DNA in the plasmid preparations. The sequencing data obtained in the MT2A8 binding region suggested that there was a single base pair discrepancy in this area, which could account for the poor sequence quality with this primer. However, this was only sequenced in three clones in which the MT2A7 sequence had extended far enough, and was not confirmed in the opposite strand, as no primer was used to sequence more 5' of MT2A8. Thus, there is a possibility that this could account for the poor sequencing results obtained in these areas. This will be discussed further in section 3.3. The MT2A7 binding region was sequenced for all clones analyzed and was found to be conserved in all cases. From this it would seem that poor sequencing quality with this primer was not due to errors in the primer binding region.

Ambiguous sequence was found in the data obtained for 9.2 and 11.1 with MT2A8, and for 11.1, 16.1, or 16.2, with MT2A7. However, in most cases the sequencing patterns were adequate to verify any changes found in these subclones with other fully determined subclone sequences. Clone 20.3 was only sequenced to confirm the sequences obtained from subclones 20.1 and 20.2, between the GRE and MRE4 consensus sequences (page C). The rest of the sequence was not determined.

3.3. The consensus MT-II_A promoter sequence:

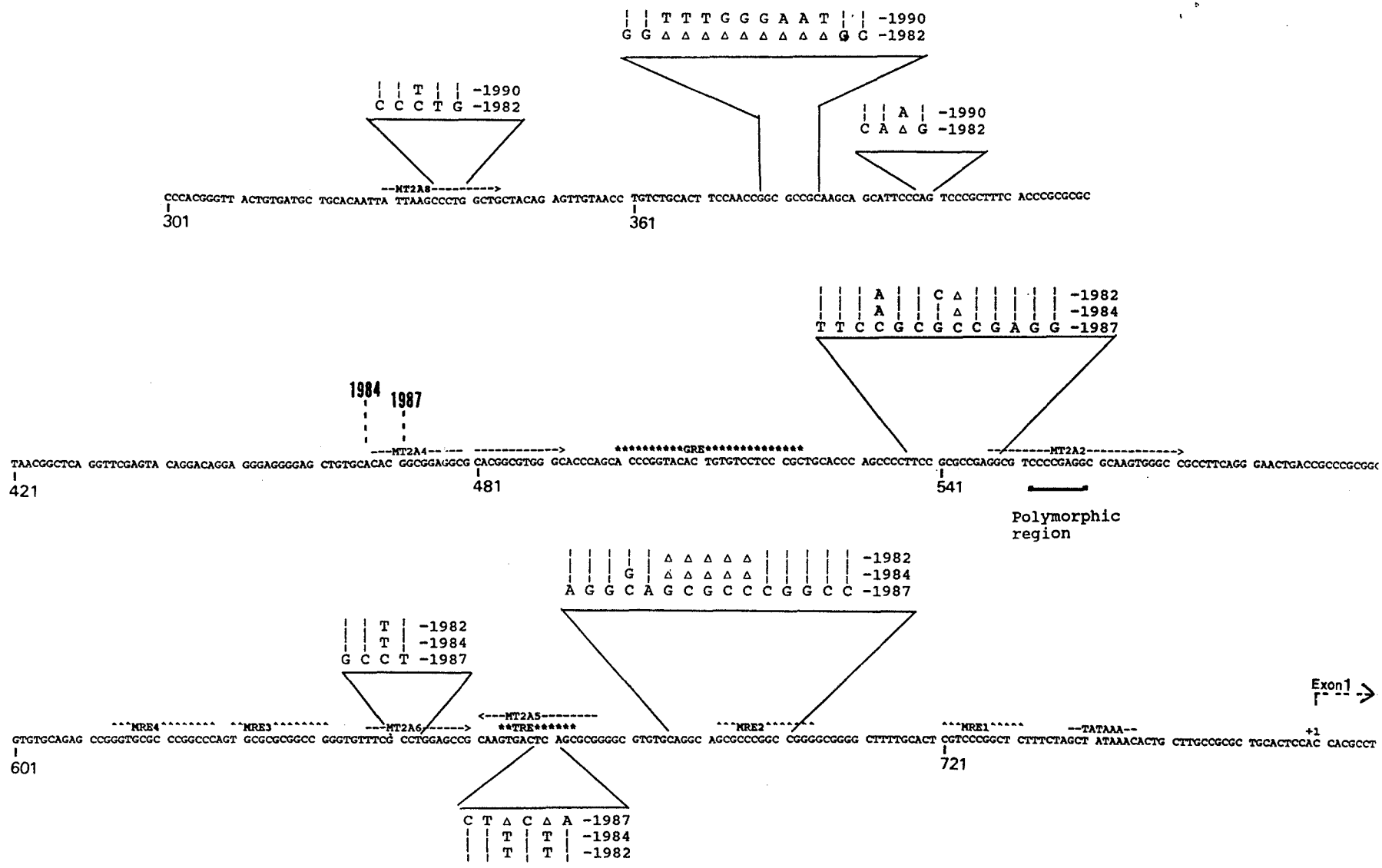
The sequencing data obtained from the 20 independent clones indicated that there were several regions within the originally published sequence (Karin et al, 1982), which were not indicative of the MT-II_A sequence in the general population. The MRE regions from position -50 to -140 in the original publication had significant differences from the 1987 publication of the promoter region (Karin et al, 1987A). A MT-II_A sequence in a 1984 publication also differed from the 1987 sequence, and was in agreement with the original 1982 sequence for the most part. Figure 5 shows the 1987 sequence from 301 to 1071, with the areas of difference in the 1982 and 1984 sequences indicated. Four areas of difference are present, one at 540 to 553, and three between 640 and 710. The sequences obtained in this study were in agreement with the 1987 sequence in these regions. Examples of the sequencing data obtained in this analysis are given in Figure 6, in order to

verify these changes. The differences between the 1982, 1984 and 1987 sequences could be due to differences within the MT-II_A sequence in the population. However, the 1984 and 1987 publications did not refer to the changes noted. It is possible that these changes were errors, due either to mistakes in sequence reading or recording. The Microgenie sequence was in accordance with the 1982 sequence, with the exception of an A at position 836, which is a T in the 1982 and 1984 sequences, as well as the obtained sequence, likely indicating an error in entering the sequence into the listings.

Several other areas in the clones analyzed had changes from those noted in the published sequence. The sequence published for positions 378 to 384 was GG/CGCC/GC in the (+) strand. Twelve independent clones in this study contained GG/TTTGGGAAT/GC (shown in the (-) strand in Figure 7), in both strands, representing seven different people. As well, position 397 to 402 in the 1982 sequence was CCAGTC, but was CCAAGTC ((-) strand; figure 6) for these clones. Sequencing gel patterns confirmed these results for the other from the 1982 sequence, which have not been republished. These are indicated in Figure 5, and the actual sequences are presented in Figure 7. At position 378 to 384, the sequence numbering corresponds to the sequence seen in figures 1 and 4. individuals cloned, even though the entire sequence for the particular region was not definitive. Position 336 to 342 was published as GG/ACCG but sequencing data here showed it to be GG/A/ACCG. Changes were seen between 950 and 1000 for all 19 clones sequenced with

Figure 5: Summary of changes in previously published MT-II_A promoter sequences.

The MT-II_A promoter sequence is shown from position 301 to 1070, spanning from the MT2A8 primer binding site to the first intron, and the sequencing primers and promoter elements are indicated. The sequence presented corresponds to the corrected sequence up to 1987. The sequence 5' to the MT2A4 oligo primer was published in 1982, and not republished in the 1984 or 1987 sequences. The dates above the sequence at 461 to 471 indicate where each publication started its sequence. The 1984 and 1987 sequences both stopped at the beginning of translation (position 851). The expanded regions between MT2A4 and exon 1 show the 1987 sequence and indicate the sequence differences in this region in the 1982 and 1984 sequences. The expanded regions 5' to the MT2A4 primer site and 3' to exon 1 indicate the 1982 published sequence, and changes identified in the current work. These are differences identified in all of the clones sequenced in these regions, as discussed in the Results (section 3). Deleted sequences are indicated by a triangle above or below the respective nucleotide position. Nulceotides that are unchanged in the expanded regions have a straight line above or below them. The polymorphic position at 557 is also indicated. The actual sequencing results from these regions are presented in the accompanying sequencing gel figures.


-1990
-1982
-1990
-1982
-1990
-1982
--MT2A8--------->
CCCACGGGTT ACTGTGATGC TGCACAATTA TTAAGCCCTG GCTGCTACAG AGTTGTAACC TGTCTGCACT TCCAACCGGC GCCGCAAGCA GCATTCCCAG TCCCGCTTTC ACCCGCGCGC
301
361
-1982
-1984
-1987
1984
1987
---MT2A4---
---------->
**********GRE***************
--------MT2A2---------->
TAACGGCTCA GGTTCGAGTA CAGGACAGGA GGGAGGGGAG CTGTGCACAC GGCGGAGGCG CACGGCGTGG GCACCCAGCA CCCGGTACAC TGTGTCCTCC CGCTGCACCC AGCCCCTTCC GCGCCGAGGCG TCCCCGAGGC GCAAGTGGGC CGCCTTCAGG GAACTGACCGCCCGCGG
421
481
541
Polymorphic region
-1982
-1984
-1987
-1982
-1984
-1987
Exon1
***MRE4^^^^^^^^
**MRE3^^^^^^^^
---MT2A6------>
<---MT2A5--------
TRE****
***MRE2^^^^^^^
***MRE1^^^^^
--TATAAA--
+1
GTGTGCAGAG CCGGGTGCGC CCGGCCCAGT GCGCGCGGCC GGGTGTTTCG CCTGGAGCCG CAAGTGACTC AGCGCGGGGC GTGTGCAGGC AGCGCCCGGC CGGGGCGGGG CTTTTGCACT CGTCCCGGCT CTTTCTAGCT ATAAACACTG CTTGCCGCGC TGCACTCCAC CACGCCT
601
721
-1987
-1984
-1982

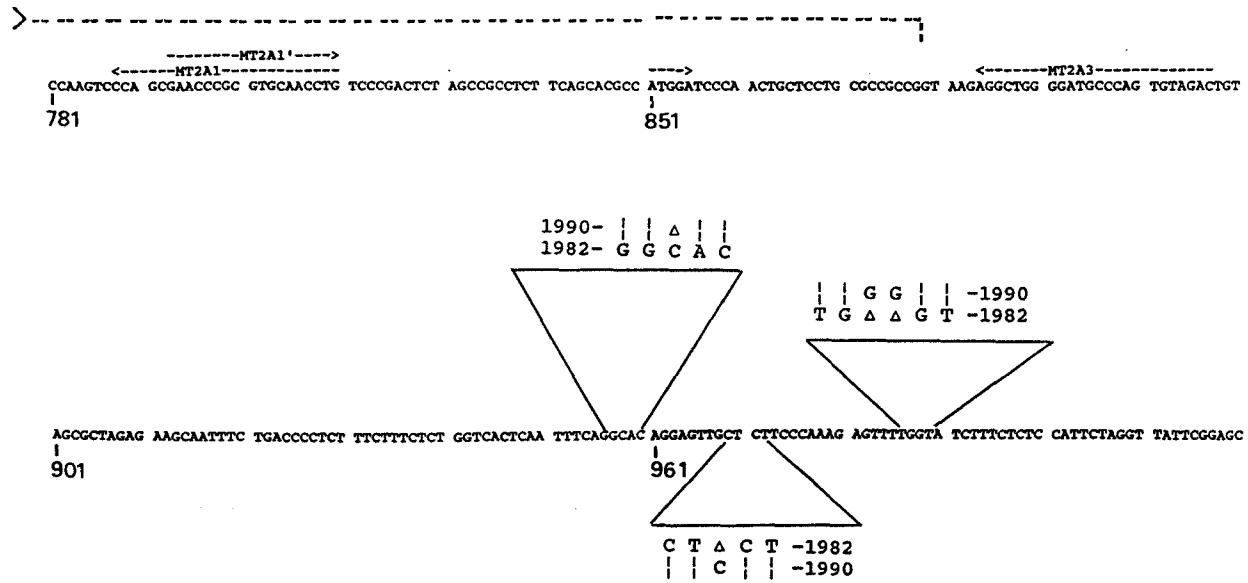
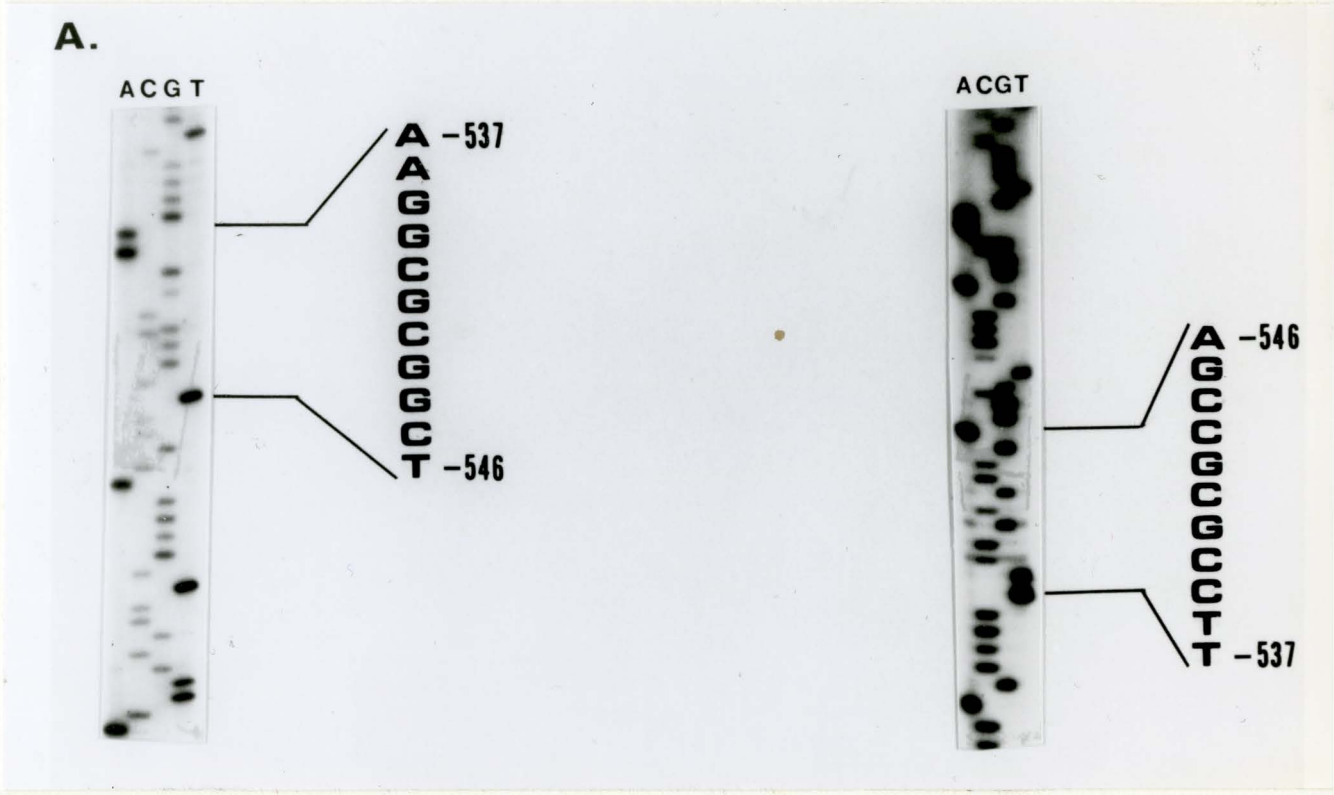
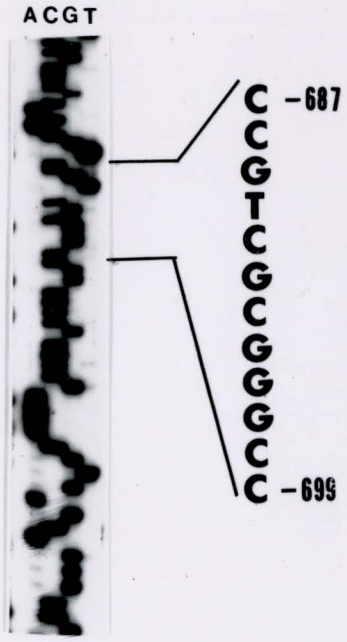

--------MT2A1'---->
<------MT2A1-------------
---->
<-------MT2A3-------------
CCAAGTCCCA GCGAACCCGC GTGCAACCTG TCCCGACTCT AGCCGCCTCT TCAGCACGCC ATGGATCCCA ACTGCTCCTG CGCCGCCGGT AAGAGGCTGG GGATGCCCAG TGTAGACTGT
781
851
1990- | | Δ | |
1982- G G C A C
| | G G | | -1990
T G Δ Δ G T -1982
AGCGCTAGAG AAGCAATTTC TGACCCCTCT TTCTTTCTCT GGTCACTCAA TTTCAGGCAC AGGAGTTGCT CTTCCCAAAG AGTTTTGGTA TCTTTCTCTC CATTCTAGGT TATTCGGAGC
901
961
C T Δ C T -1982
| | C | | -1990
CCCCTTTTTA CCGTTAAGGA GATCTGAGTT AATGGCTTGC TCAAGTTCCC AGGAATCGGT
1021

Figure 6: Sequencing Results

Sequence data confirming the sequence changes indicated in the 1987 publication from the 1982 and 1984 sequences are shown (see figure 5, position 461 to 721). (A) illustrates the region from 537 to 546 in both strands, which corresponds to the region 5' to the MT2A2 primer binding site and the polymorphic site at 557 (seen lower in the sequence shown, but not indicated). (B) shows the sequence from 687 to 699 around the MRE2, in the (-) strand, and (C) shows the sequence containing the differences in the MT2A6 and the MT2A5 binding regions, in the (-) strand again. Both (B) and (C) sequences were obtained using the MT2A1 primer.



B.


ACGT
C -687
CGTCGGGGCC
C -699

C.

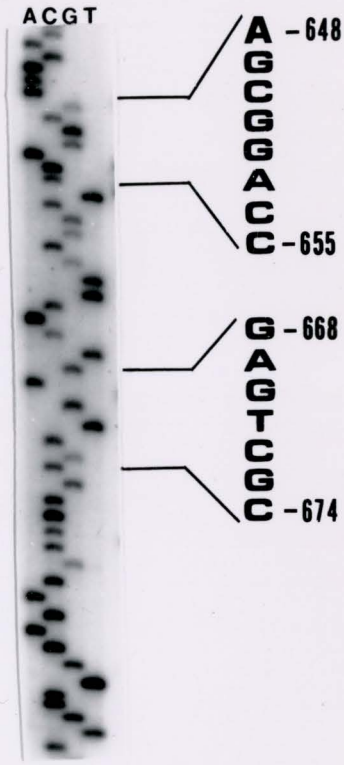
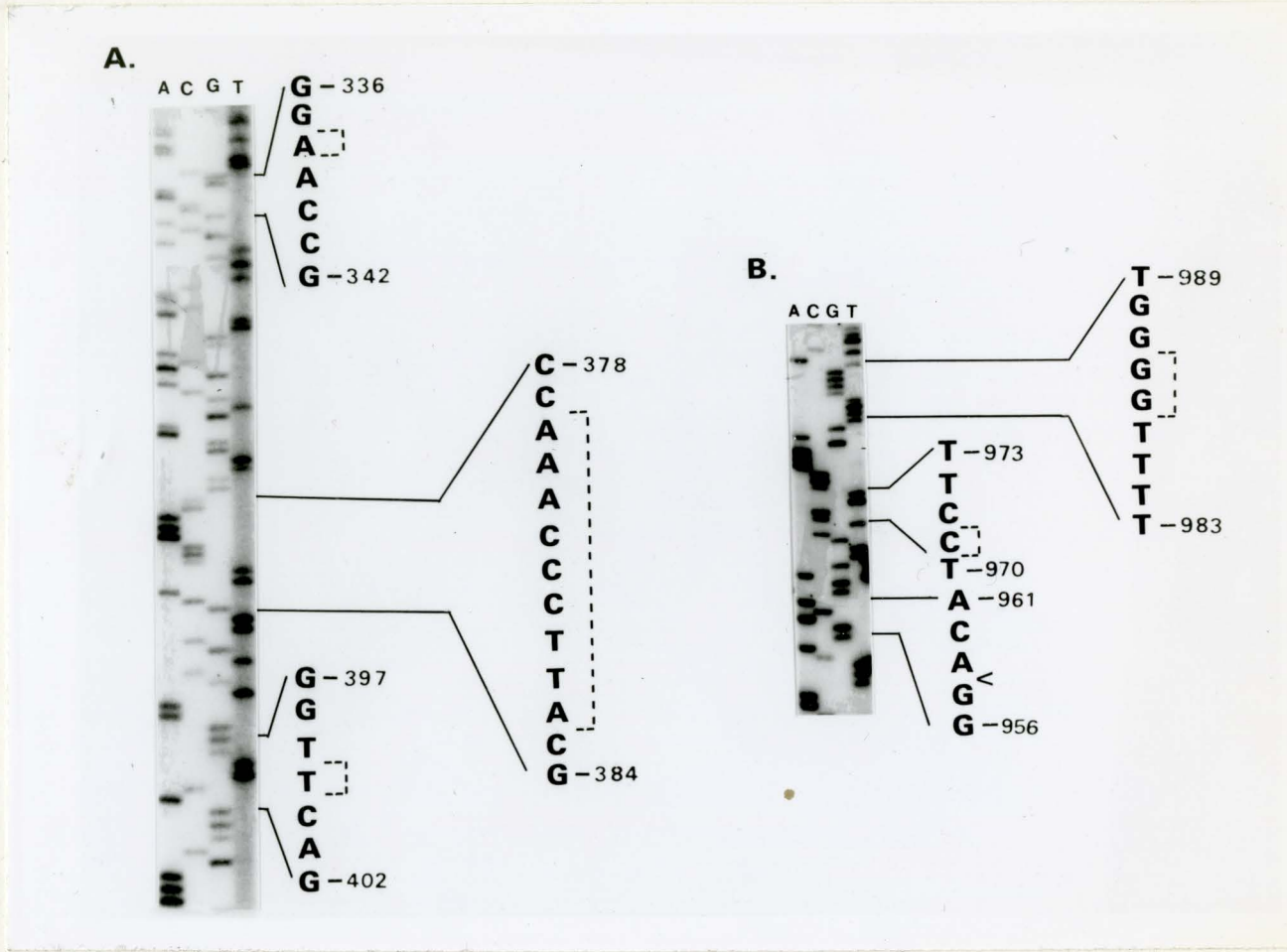

ACGT
A -648
AGCGGACC
C -655
G -668
GAGTCGC
C -674

Figure 7: Sequencing Results



A. Sequence is indicated from position 336 to 402 in the (-) strand, corresponding to sequence derived from MT2A7, and spanning the region of the MT2A8 primer, and the three regions indicated in figure 5 as being different from the 1982 sequence. The specific nucleotides that are changed for the 1982 sequence are contained within the broken lines.

B. Sequence is indicated for position 956 to 989, corresponding to sequence derived from MT2a1', in the (+) strand, spanning the regions sequenced 3' to exon 1. The three regions illustrated are again seen in the sequencing summary figure 5, in respect to the 1982 sequence publications. The sequence numbering corresponds to the sequence seen in figures 1 and 4.

MT2A1'. The C at 958 in the 1982 sequence was not present in this analysis. Position 970 to 973 was published as TCTT, but was seen to be TCCTT here. The GG at 987 to 988 in the 1982 publication was GGGG for the clones analyzed. The area from 950 to 1000 was only sequenced in the (+) strand, so it was not confirmed in the other strand. However, the sequence in this area was clean, with no compressions, so it is believed to be correct (Figure 7, sequence B). All of these changes were found within the first intron and would therefore not affect the amino acid sequence or the reading frame.

The sequencing data obtained suggested that a change from the original sequence was also present in the MT2A8 primer binding site, where the C at position 338 was found to be a T. However, this was found in only three clones in which sequence in the (-) strand from MT2A7 extended far enough (Figure 7, sequence A). Although the clarity of the bands in this area are less than ideal, it does appear that an AA doublet exists at 338-339. The sequence found here indicated 336-340 to be CCTTG (shown in (-) strand), not CCCTG as seen in the 1982 publication. This could explain why sequencing difficulties occurred with the MT2A8 reactions, which proved much harder to obtain than with the other primers. This region was not verified in the (+) strand, so is not proven definitively.

As with the changes found between the previously published MT-II_A promoter sequences, the changes found in this study could be due to actual differences in the sequence of the MT-II_A gene between the individuals

sequenced in 1982 (Karin et al, 1982) and the individuals sequenced in this study. With the number of changes, and the fact that the changes are located in two areas of less than 100 bp each in size, it seems more likely that these are errors due to experimental difficulties with sequencing, in sequencing gel reading, or transposition of the sequence during recording and publication of the earlier reports.

3.4. Human MT-II_A promoter nucleotide polymorphism:

Other than finding the changes listed above in the previously published MT-II_A sequences, the MT-II_A promoter sequence was found to be invariant between the twenty independent clones isolated. Only one position was found to be variable. Nucleotide 557 was found to be a G for all three donor 20 clones studied, and for one of two donor 19 clones analyzed (19.2). All of the other subclones were determined to have an A at this position, in accordance with the published sequence. Figure 8 shows the sequence between 552 and 559 in both strands for two of the clones with nonpolymorphic sequences and one polymorphic clone. Two examples of the wild-type sequences are shown, because of the compression observed at position 560 to 562. Two patterns were observed for these clones. Wt1 in the (+) strand shows a CGC, but the bands are compressed, whereas wt2 in the (+) strand shows a sequence resembling a GC, with the C being more diffuse, but not obviously two bands. The wt1 and wt2 (-) strand sequences are

obviously GCG at 560 to 562, confirming the sequence as that in the original publication. No compression is seen at 560 to 562 for the polymorphic clones, likely due to the change in the sequence 4 bp away.

The A to G transition at 557 should change the sequence at 552 to 559 from CCCCGAGG to a CCCCGGGG, generating a SmaI restriction site (CCC^GGG). This was confirmed by digesting the respective pUC118/MT-II_A subclones with SmaI as seen in figure 9. Two orientations were possible as seen in figure 9, since the MT-II_A EcoRI was cloned into the EcoRI site in pUC118. The presence of the polymorphic site was confirmed by the generation of a 565 bp fragment in either orientation, which spanned from position 557 to a SmaI site in the 5' third of the transcribed region. The two possible orientations are indicated in part B) of the figure. Samples 20.1, 20.2, 20.3, and 19.2 were seen to have the polymorphic fragment, confirming the presence of the site. Sample 19.2 did not have the fragment, as was the case for samples 18 and 22, indicating that the sequence at 557 was in agreement with the previously published sequence. Orientation i) is seen in donor samples 20.1 and 19.2 indicating SmaI fragments of 565 bp, 1.5 kbp, and 7.2 kbp, approximately. The opposite orientation ii) is seen in samples 20.2 and 20.3 indicating fragments of 565 bp, as well as 2 bands of approximately 4.6 and 4.7 bp. Samples 20.1, 18 and 22 are seen to only have 2 bands. These are the expected sizes of orientation ii) without the polymorphic site present. The expected fragments would be approximately 4.6 and 5.3 kbp in size.

Thus, it was confirmed that the pUC118 subclones for samples 20.1, 20.2, 20.3, and 19.2 contained the polymorphism at position 557.

In total, four of the twenty independent clones contained the identified nucleotide polymorphism, representing two different people, one of which was heterozygous for the change. Using the data from all ten people (Table 3), the rates of polymorphism can be determined for the area sequenced, as well as for each DNA region, as was done in Kreitman's Drosophila studies (Kreitman, 1983; Kreitman and Hudson, 1991). A variable amount of sequence was obtained for each individual in the 5' promoter region. For all clones, at least 270 bp of 5' promoter region was sequenced completely in both strands. However, for seven people, an area of 430 bp was sequenced in both strands, so a potential range of the rate of polymorphism can be generated for this region of 0.23 to 0.37 bp/ 100 bp or 0.23 to 0.37 % polymorphism, representing one change in 270 to 430 bp. This means that the rate of polymorphism is 0.37 %, but could be as low as 0.23 % if all clones were to be sequenced to 430 bp in both strands and no further changes were identified. Because no other sequence changes were identified, the rate of polymorphism was 0/ 100 bp for the transcribed, untranslated region (70 bp; 20/ 20 clones), the translated portion of exon 1 (28 bp; 20/ 20 clones), and the intron 1 region sequenced (150 to 195 bp; 19/ 20 clones, in 1 strand only).

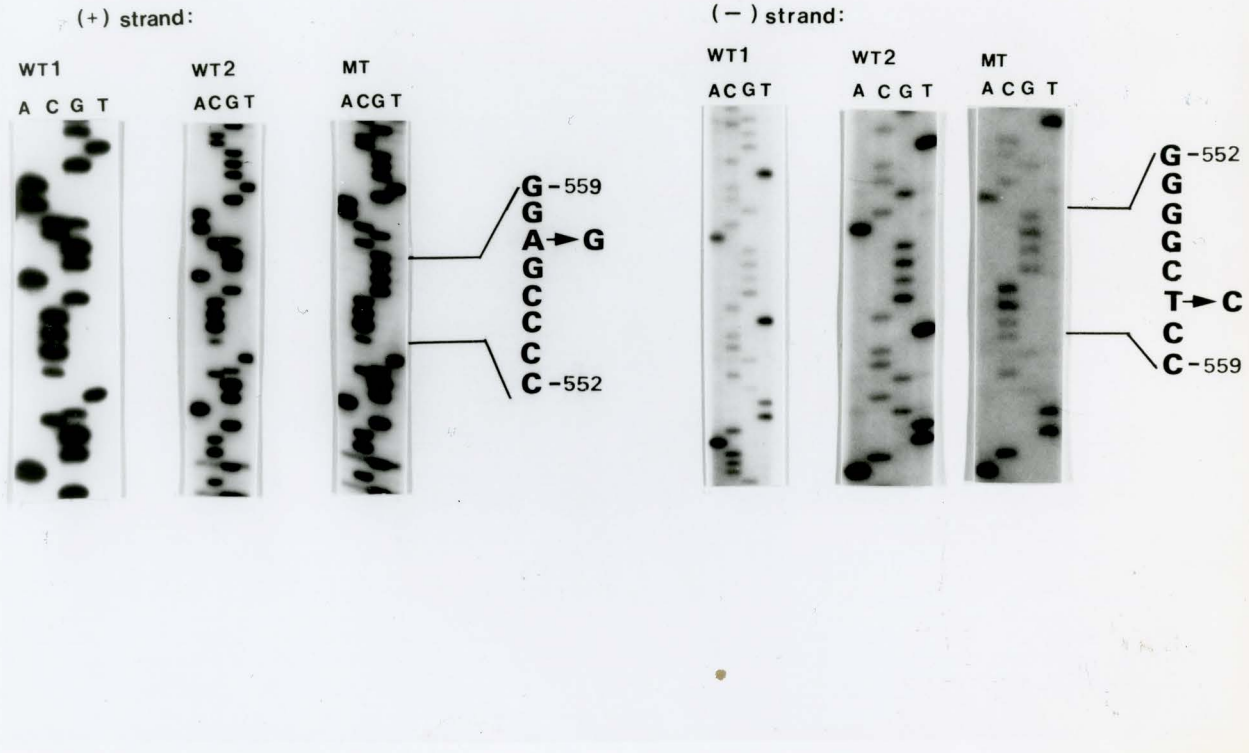
The rate of polymorphism for the whole segment sequenced was determined in two ways as seen in Table 3, box 4. A range is again

presented for each to include the minimum value with the 270 bp obtained for all clones in both strands, as well as the maximum value using the 430 bp obtained for the clones from seven of the people analyzed. The maximum and minimum values are presented with and without the intron sequence results, since this was not confirmed in both strands. However, because of the clarity of the sequencing data in this area, the single stranded data was believed to be reliable. The range of polymorphism for the complete sequence obtained with the intron sequence was 0.14 to 0.21 % with the 0.21 % being the value derived from the region for all clones in both strands. The 0.14 % is a hypothetical value which would be true if the sequence for the remaining clones not sequenced to 430 bp were found to be invariant as well. The range without the intron sequence was higher at 0.18 to 0.28 %. These results indicate that the promoter regions analyzed, as well as the first exon and part of the first intron appear to be very highly conserved between the different individuals studied here. It should be noted however that the full MT promter was not sequenced. This will be dealt with further in the discussion, and the data here will be compared with the Drosophila studies presented earlier.

3.5 Comparison of the MT-II_A promoter to the Drosophila ADH Nucleotide Polymorphism Study:

The frequencies of polymorphism determined above for MT-II_A were

Figure 8: Comparison of wild-type and polymorphic sequences.



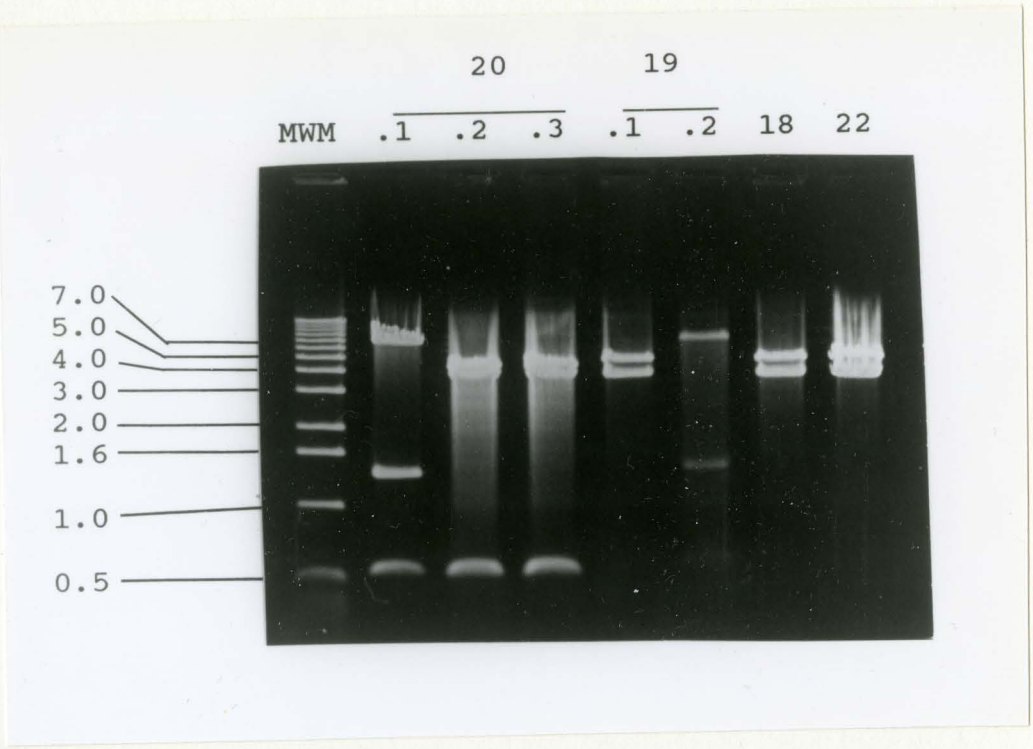
Comparison of the nonpolymorphic and polymorphic sequences, in the (+) and (-) strands from position 552-559. An example of one of the clones from one of the polymorphic individuals is seen, represented by the MT heading. The change is an A to G transition at position 557. Two nonpolymorphic sequences are shown in both the (+) and (-) strand to illustrate the two resulting patterns for the compression seen from position 559 to 562, obtained with the nonpolymorphic sequence only. The nonpolymorphic sequence is indicated by the heading WT. The sequence from 560 to 562 should read CGC as seen in the polymorphic sequence, but this is not obvious from the nonpolymorphic sequence. In WT1, two Cs are seen, but the spacing is altered. In WT2, the two Cs are not obvious and appear as one more diffuse C. However, as seen in the WT (-) strand, the sequence does read as CGC in the (+) strand, in accordance with the published sequence. Two examples were given to clarify this point. See figure 4 for the sequencing summaries of each clone, and see figure 1 for the position of the change in the gene.

Figure 9: SmaI digests of pUC118/MT-II_A subclones for confirmation of the polymorphic site.

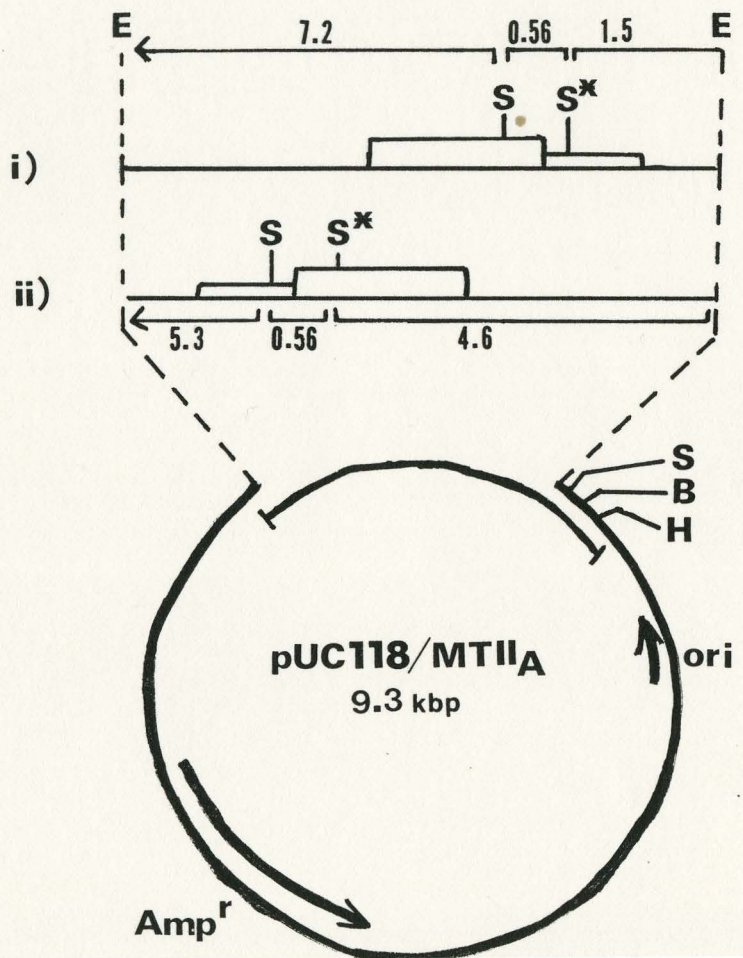
Part A) indicates the SmaI digested subclones for donor samples 20.1, 20.2, 20.3, 19.1, 19.2, 18 and 22, with the 1 kbp ladder as a standard. The 6 kbp EcoRI MT-II_A fragment was subcloned into the EcoRI site of the multiple cloning segment of pUC118. The samples were electrophoresed on a 1.0 % agarose/1 X TBE gel and stained with ethidium bromide.

Part B) shows the pUC118/MT-II_A restriction map for the MT-II_A fragment inserted into the plasmid in both orientations. Both orientations were found in the digests. Orientation i) is seen in donor samples 20.1 and 19.2 indicating SmaI fragments of 565 bp, 1.5 kbp, and 7.2 kbp, approximately. The opposite orientation ii) is seen in samples 20.2 and 20.3, indicating fragments of 565 bp, as well as 2 bands of approximately 4.6 and 4.7 kbp each. Samples 20.1, 18, and 22 are seen to only have 2 bands. These are the expected sizes of orientation ii) without the polymorphic site present. The expected fragments would be approximately 4.6 and 5.3 kbp in size. The MT-II_A restriction map is indicated to scale in figure 1. pUC 118 is illustrated in Sambrook et al, 1989.

The smaller boxed area indicates the promoter region and the larger boxed area indicates the transcibed region. The fragment sizes are noted below each orientation in kbps. The origin of replication (ori) and ampicillin resistance gene (Amp^r) are indicated. The restriction sites are indicated as E=EcoRI, S=SmaI, B=BamHI, and H=HindIII. The polymorphic SmaI site is indicated by the S^*.


20
19
MWM
.1
.2
.3
.1
.2
18
22
7.0
5.0
4.0
3.0
2.0
1.6
1.0
0.5

B.


E
7.2
0.56
1.5
E
S
S*
i)
S
S*
ii)
5.3
0.56
4.6
S
B
H
pUC118/MTIIA
9.3 kbp
ori
Ampr

compared to the distal and proximal promters from the Drosophila ADH gene (section 1.8). The Drosophila ADH data is presented in Table 1 and the MT promoter data is presented in sections 3.4 and Table 3. The 1991 study by Kreitman and Hudson showed that the extended 5' flank sequence had a total frequency of 2.4 %, with the distal adult promoter having a frequency of 3.0 % over 660 bp. The first 225 bp had a rate of 1.3 %. The larval proximal promoter had a frequency of 0.5 % over 385 bp, with a frequency of 0.9 % if the first 225 bp only were analyzed. Thus, it would appear that the frequency of polymorphism is greater for the distal promoter and is relatively low for the proximal promoter (see Table 1.a). The % polymorphism for MT described in section 3.4 above (Table 3) are apparently considerably lower than those found for the ADH promoters. These rates of nucleotide polymorphism were subjected to statistical analysis to see if the differences were indeed significant (Table 4).

The chi-squared statistical analysis for the comparison of different promoter regions is presented in Table 4, using the null hypothesis for each that the areas in question do have the same degree of polymorphism. The maximum and minimum areas of MT sequence were both compared to the two ADH promoters together giving X^2 values of 16.54 and 12.35 respectively, with two degrees of freedom. This means that the null hypothesis can be rejected for both at the 0.005 level, and the three regions have statistically different frequencies of polymorphism. However, comparison of the data for

the first 270 bp of the MT promoter to each ADH promoter region separately (comparison 7 and 8, Table 4), gave X^2 of 6.04 and 0.095 for the distal and proximal promoters respectively. Thus, for this region of the MT promoter, the null hypothesis is rejected for the comparison with the distal adult promoter (at the level of $p= 0.025$), but cannot be rejected for the proximal larval promoter (at the level of $p= 0.5$), meaning the degree of change is not statistically different between the MT and proximal promoters, but is not for the other ADH promoter. Comparisons of the two ADH promoters indicated that the null hypothesis could be rejected for for the comparison of the full regions as well as the first 385 bp. Thus, the rates of change in the two ADH promoters are significantly different. The 385 bp comparison was done because the area of the proximal promoter found in the first intron was 385 bp but this does not contain the full sequence for normal larval mRNA transcription. Therefore, this comparison may be more appropriate than comparing this region with the entire 660 bp of the distal promoter.

Comparing the first 270 bp of each promoter showed these regions to be significantly different with rejection of the null hypothesis at the 0.05 level. However, when the first 225 bp of each are compared, the hypothesis cannot be rejected and so the regions are not significantly different (see boxes 3 and 4, Table 4).

From these analyses, it appears that the proximal larval ADH promoter and the MT promoter regions analyzed were not statistically different.

However, this was not the case for the comparison of the MT and distal adult promoter regions. Analysis of the two ADH promoters showed that these regions also were statistically different in respect to their frequencies of polymorphism, with the distal promoter having a greater degree of change. Comparison of the first 225 bp for all three promoters showed that these areas were not statistically different for their relative rates of polymorphism.

3.6 Restriction Fragment Length Polymorphism Analysis:

After the results were verified in the individual clones, attempts were made to show the polymorphisms by restriction fragment length polymorphism analysis. The human genomic DNAs were digested with the enzymes listed above used for the restriction enzyme verification, as indicated in section 2.2.11, electrophoresed, transferred, and probed with the pHS1b HindIII-EcoRI fragment. Figure 1 shows the partial restriction map of the MT-II_A genomic EcoRI fragment as determined by Karin et al (1984). SmaI and XmaI southern analysis of the donor 20 genomic DNA should result in two fragments to which the probe should hybridize (as indicated in figure 1); one of 556 bp, and another of at least 1.5 kbp, spanning from the SmaI/ XmaI site at position 557 across the promoter and part of the 5' EcoRI site. For an individual without this change, only one fragment of at least 2 kbp should exist, depending on where the next SmaI/ XmaI site occurs 5' to the EcoRI site. For the donor 19 DNA, a combination of both would be expected if

Table 3: Summary of Polymorphism in the Metallothionein-II_A gene regions analyzed in this study.

This table summarizes the sequencing data for nucleotide polymorphism frequency for MT-II_A in this study. The numbers of bp sequenced, as well as the number of polymorphic sites and the % polymorphism for each region are presented. Values are given with maximum and minimum numbers because not all of the clones were sequenced to the same extent due to sequencing template variability, as described in the text. A maximum value indicates the result that would be obtained if all clones had been sequenced to 430 bp, as was the case for the clones from 7 individuals. A minimum result indicates the "truest" value in that it is the extent to which all clones in the study were actually sequenced in both strands. In box 4, the maximum (4.A) indicates the bases sequenced for the clones in which the most sequence was obtained. The minimum in box 4.B represents the minimum amount of sequence which was obtained for all clones. This would again represent the "truest" value since the maximum value would assume that no other polymorphisms would be found if the remainder of the sequence for all the clones was obtained. Each of these is given with and without the intron sequence since this was only determined in one strand. However, it is believed to be accurate for the sequence obtained. As a result, ranges of % polymorphism are obtained for the intron and promoter regions, as well as for the MT-II_A region sequenced in total. (ss=single stranded DNA sequence).

Table 3: Summary of Polymorphism in the Metallothionein-II_A gene regions analyzed in this study.

DNA Region:	# of nucleotides sequenced (bp)	# of polymorphic sites found (bp)	% polymorphism
1. exon 1:			
i)Translated portion	99	0	0 %
ii)Transcribed, untranslated region	71	0	0 %
2. Intron 1: Maximum bp sequenced	195	0	0 %
Minimum bp sequenced	150	0	0 %
3. Promoter: Maximum bp sequenced	430	1	0.23 %
Minimum bp seq-	270	1	0.37 %
4. TOTAL:			
A) i)Maximum bp seq. with ss intron seq.	720	1	0.14 %
ii)Maximum bp seq. without ss intron seq.	559	1	0.18 %
B) i)Minimum bp seq. with ss intron seq.	488	1	0.21 %
ii)Minimum bp seq. without ss intron seq.	358	1	0.28 %

Legend to Table 4: Statistical analysis to determine whether the MT-II_A and Drosophila melanogaster ADH 5' promoter regions are significantly different.

Chi-squared analysis was used with the null hypothesis that the promoter regions being compared were essentially the same in respect to their degree of nucleotide polymorphism. This hypothesis was then accepted or rejected according to the calculated outcome. An example of the chi-squared analysis for one comparison is given in Table 4.a). Table 4.b) follows this example and gives the calculated values for each comparison done.

Comparisons were done with the two ADH promoters and the MT promoter for both the maximum and minimum sequence, as described above. The 3 promoters were compared for the first 225 bp as well, because of the relative paucity of changes in these areas for all three regions. Comparison was also done with the first 270 bp of each to see if this relative degree of conservation was maintained to the extent to which all of the MT clones were sequenced. The two ADH regions were compared for the sake of discussion. Each ADH promoter region was compared on its own to the data for the minimum 270 bp sequenced for all MT clones, to see if these were statistically different independently, in respect to their degree of polymorphism.

The chi-squared results are reported for each in Table 4.b), along with the degree of freedom for each comparison. The percentile of the chi-squared distribution needed to acheive greater than 95 % statistical significance for each is indicated. In brackets after this value is the degree to which the null hypothesis can be rejected, or the probability. For those in which the hypothesis could not be rejected, the value indicates the required result for rejection of the hypothesis at 95 %, or p= 0.5. Below each of these values, the outcome of whether the hypothesis was rejected or accepted is indicated (Remington and Schork, 1985; von Fraunhofer and Murray, 1976).

Example of the Calculation:

Chi-squared values were calculated using the equation:

$$X^2 = \frac{(O-E)^2}{E}$$

where O is the observed frequency of changes and E is the calculated expected frequency.

The following indicates the data table for comparison 1 with the 430 bp of MT promoter sequence and the two ADH promoter sequences:

Table 4.a)

	MT	ADH 1	ADH 2	Total
1. # of bases changed	1	20	2	23
2. # of bases not changed	429	640	383	1452
3. Total # of bases for each region compared	430	660	385	1475

From this, the expected frequency of polymorphism if the null hypothesis is true (all three regions have the same degree of change) would be:

$= 23/ 1475= 0.016,$

and the expected % of base positions not changed would be:

$= 1452/ 1475= 0.984$

Using these values, the value of X^2 can be determined:

$$X^2= \frac{(1-(430x\ 0.016))^2}{(430x\ 0.016)} + \frac{(20-(660x\ 0.016))^2}{(660x\ 0.016)} + \frac{(2-(385x\ 0.016))^2}{(385x\ 0.016)} +$$

$$\frac{(429-(430x\ 0.984))^2}{(430x\ 0.984)} + \frac{(640-(660x\ 0.984))^2}{(660x\ 0.984)} + \frac{(383-(385x\ 0.984))^2}{(385x\ 0.984)}$$

$= 5.025 + 8.439 + 2.810 + 0.082 + 0.137 + 0.046$
$= 16.54$, with 2 degrees of freedom

This value is then compared against a table of chi-squared distributions. It exceeds the value of 10.6 at $X^2(0.995)$, so the null hypothesis can rejected at the 0.005 level, meaning that the three regions compared are statistically different in respect to their relative rates of polymorphism.

The degrees of freedom (df) are indicated with each of the calculated X^2 values.

From this analysis, the only comparisons where the null hypothesis could not be rejected were those of the first 225 bp of each promoter, and the comparison of the MT and proximal ADH promoter. Thus, these are the only sequences that are not statistically different. The results are discussed further in the text. Whether or not the null hypothesis could be accepted or rejected is indicated in Table 4.b) for each comparison.

Table 4.b): Statistical analysis to determine whether the MT-II_A and drosophila melanogaster Adh 5' promoter regions are significantly different.

Comparison done	Chi-squared calcuated result	Percentile of the chi-squared distribution for significance
1. 430 bp of MT with both promoters	16.45, 2 df	10.6 (0.995), hypothesis rejected
2. 270 bp of MT with both promoters	12.35, 2 df	10.6 (0.995), rejected
3. First 270 bp of all three promoters	6.28, 2 df	5.99 (0.95), rejected
4. First 225 bp of all three promoters	1.01, 2 df	5.99 (0.95), accepted
5. Comparison of the 2 ADH promoters	7.46, 1 df	6.63 (0.99), rejected
6. Comparison of first 385 bp of each ADH	9.05, 1 df	7.88 (0.995), rejected
7. Comparison of MT with distal ADH promoter	6.04, 1 df	5.02 (0.975), rejected
8. Comparison of MT and proximal ADH promoter	0.095, 1 df	3.84 (0.95), accepted

the individual was actually heterozygous.

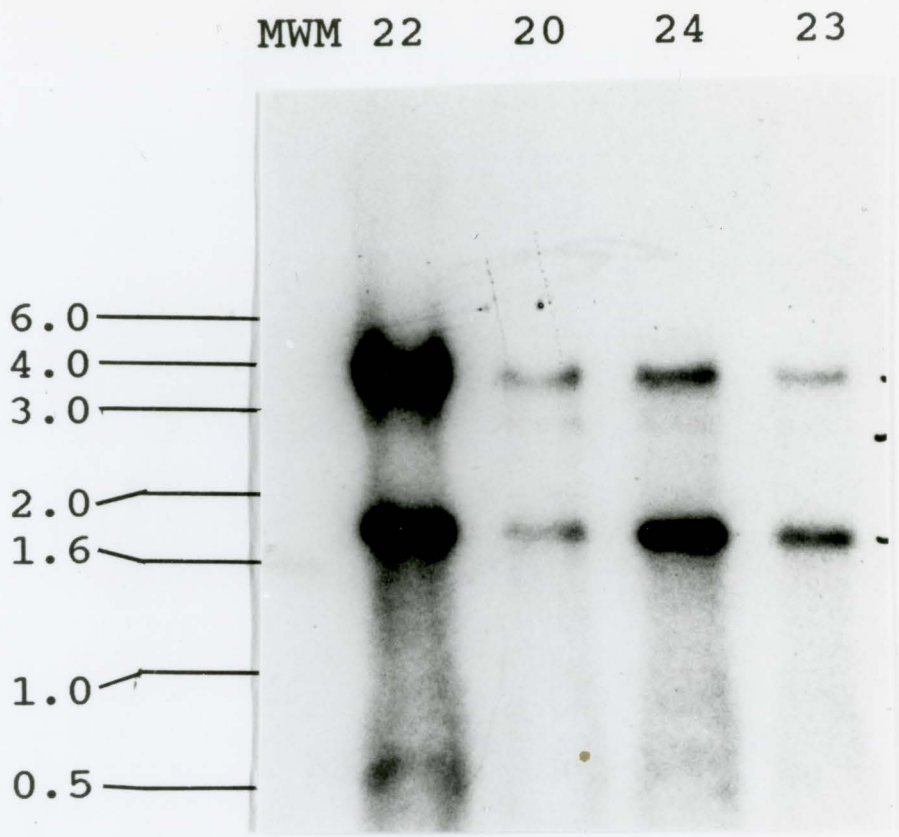
The Southern results obtained were more complicated than expected. Analysis of Figure 10 shows SmaI Southern analysis of the donor 20 and 22 individuals, along with two other individuals for which no cloning was done. No significant difference can be seen between samples. Bands can be seen at approximately 600 bp and 1.8 kbp, along with a series of bands above 3 kbp which could represent genomic fragments with cross-hybridization with the MT-II_A probe. No banding is seen in the 1.5 kbp region. The band running at 1.8 to 1.9 kbp could be the SmaI fragment uncut at position 557, although this appears to be running low for this to be the case. This could be due to some migration effect due to salt or protein in the sample. However, the bands at 0.8 kbp are unexplained, as is the presence of a 600 bp fragment in all samples.

Figure 11 shows the Southern analysis of the donor 19, 20, and 22 individuals. Again, a band at 500-600 bp is seen in all three lanes, as is a band at 1.4 kbp. Further Southern analysis showed similiar patterns to those indicated here. No results were obtained which could be explained according to the MT-II_A restriction map. Southern analysis was also carried out with MspI digested genomic DNA. However, because of the small size of the restriction fragments in the area of interest, they could not be separated effectively by acrylamide or agarose gel electrophoresis. The data for these experiments

was inconclusive and is not shown.

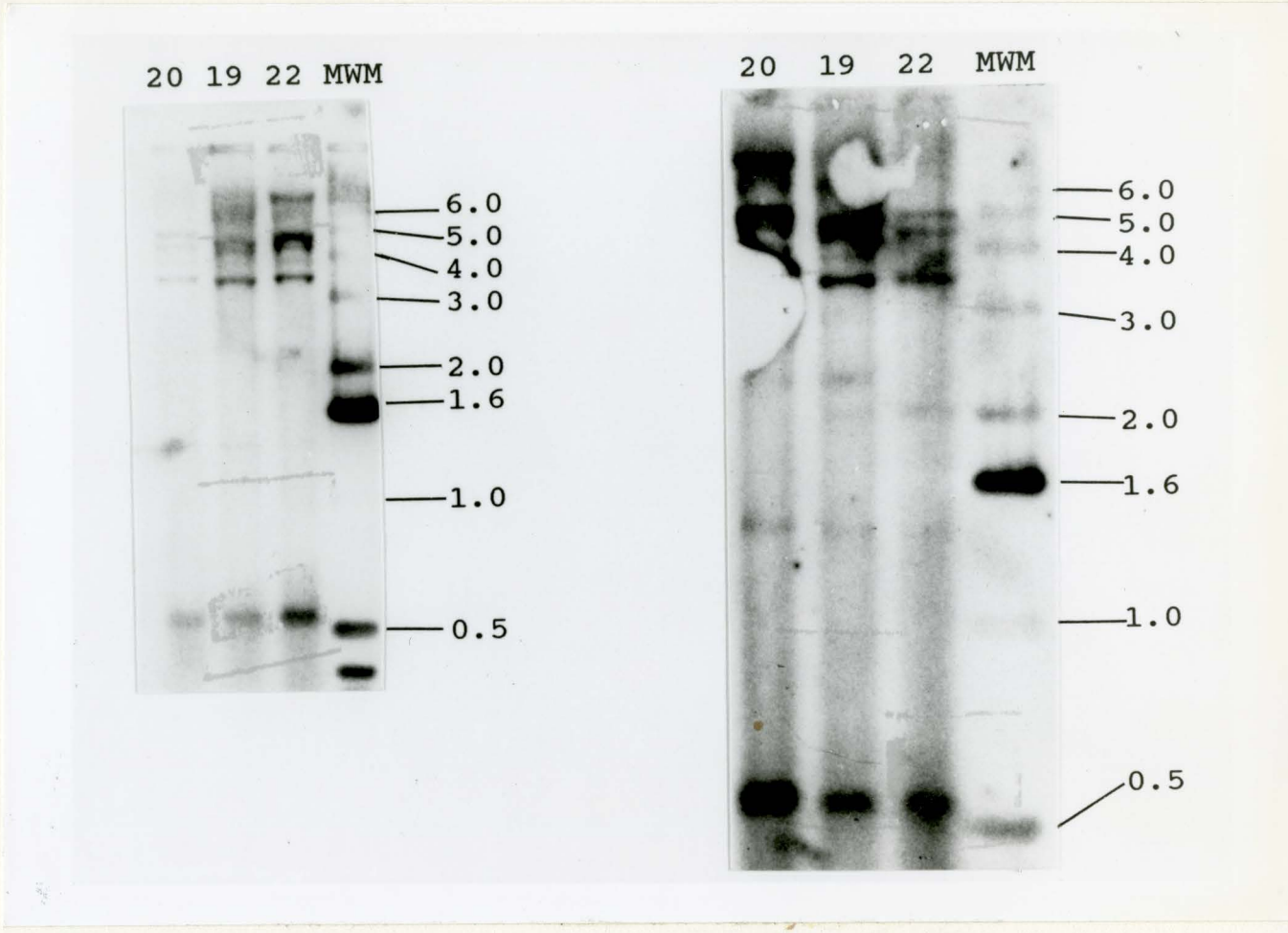
In summary, the human MT-II_A promoter region was isolated and sequenced for 12 individuals and found to be highly conserved. However, the southern analysis did not provide the expected results to verify the polymorphism identified. Ideally, such data is obtained to prove that the polymorphism exists definitively. The polymerase chain reaction was attempted as well, to verify the identified polymorphism, as well as to establish the frequency of change in the population. However, the expected results were not obtained. It is possible that this was due to genomic-primer cross hybridization and difficulty establishing proper experimental conditions. Utilization of this technique could also have allowed for further analysis of the degree of nucleotide promoter variation in the gene, since many more individuals could have been sequenced using it.

Figure 10: Southern Analysis



Southern analysis, done according to section 2.2.2. Genomic DNA, purified from peripheral lymphocytes from donors 20, 22, 23, and 24 were digested with SmaI and run on a 1.5 % 1 X TBE agarose gel. This was transferred to Zeta-probe nylon filter and probed with the human MT-II_A promoter region derived from the plasmid pHS1b. The 1 kbp molecular weight marker was run concurrently as shown.

Figure 11: Southern Analysis



A.) and B.) show two different southern gels in which genomic DNA which was purified from the peripheral lymphocytes of donors 19, 20, and 22. The DNA was digested with SmaI in both cases, and run on 1.0 % 1 X TBE agarose gels. The gels were transferred to Zeta-probe nylon filters and probed with the human MT-II_A promoter fragment derived from the pHS1b plasmid. The 1 kbp molecular weight marker was run concurrently. Southern gel analysis is described in section 2.2.2.

4. DISCUSSION

4.1 Polymorphism in the MT-IIA Promoter region:

In this study, the MT-II_A promoter regions, that were cloned and sequenced from ten different people were found to be highly conserved in the area analyzed. Only one base was found to be polymorphic. This was found for all clones identified for one individual. Another individual appeared to be heterozygous for the polymorphism since one clone identified contained the polymorphism and another contained the wild-type sequence. From these results, a rate of polymorphism was determined to be 0.14-0.19 changes/ 100 bp, or 0.14-0.19 % (Table 3). The higher value includes the 518 bp area which was sequenced for all subclones analyzed. The lower value includes data from seven people who were sequenced in both strands in 713 bp , indicating the rate of change if all the clones had been sequenced to the maximum obtained in these sequences. The polymorphic site was found in the 5' promoter region, 35 bp 3' from the GRE sequence, and 57 bp 5' from MRE4, giving a rate of polymorphism of 0.23-0.37 changes/ 100 bp, or 0.23-0.37 % polymorphism, for that specific region. The other regions in which sequence was obtained included the transcribed, untranslated region, as well as exon 1 and the first 150-195 bp of intron 1. These areas showed no polymorphism for the people analyzed. Although the sequence for intron 1 appeared definitive, it must be regarded cautiously, since only one strand was determined. However, it does appear that the rate of polymorphism in the MT-

II_A promoter region is small in respect to previous nucleotide polymorphism studies.

4.2 Polymorphism Studies in other Genes and comparison with MT-IIA:

Other polymorphism studies have been done previously, but few have looked at the rates of nucleotide polymorphism. D'Urso et al (1988) identified a polymorphic silent mutation in the human glucose-6-phosphate dehydrogenase gene by analyzing the X chromosome of 12-57 individuals with 18 enzymes and 14 sequencing probes. However this study was an RFLP analysis, so any mutations outside of sequences not recognized by the restriction enzymes utilized would not have been identified and only those areas seen to result in fragment variation were analyzed. Thus, variation in the promoter region as well as other areas could be present, but not identified. As well, it would be possible to see insertion/ deletion changes without a new restriction site being present. Other studies have been done, but these have either used RFLP analysis initially to find changes to sequence (Georges et al, 1987), or have selected for spontaneous mutations in tissue culture which would select for functional amino acid changes resulting in changes in gene function, and select against changes in the silent intron or promoter regions. As well, these would not identify the presence of third position codon changes which would not result in amino acid changes. Effects in these promoter regions would not be obvious since seeing the change would rely on

expression of the gene. Analysis of genes like glucose-6-phosphate dehydrogenase have indicated many changes and high rates of polymorphisms, but these have emphasized the analysis of the coding regions, and have involved isolation of phenotypically different genes (Beutler and Yoshida, 1988).

The study by Kreitman (1983) presented in the Introduction is the only study which has analyzed polymorphism in a gene at the nucleotide level without selecting for changes first. Eleven cloned ADH genes from five natural populations were sequenced and 43 polymorphisms were identified, only one of which resulted in an amino acid change. Each gene region was analyzed separately (see Table 1) and it was found that the highest ratios of changes were found in the non-coding regions introns 1-3, translated parts of exon 4, and the 5' flanking region. The 5' flanking region had the second highest rate, but the polymorphic sites were found in position -1, -2, and -3. No other changes were found in the remaining 5' flank, but only 60 bp were actually analyzed. The exon regions 1-3, 3' untranslated region, 3' flank, and 5' flanking region without positions -1, -2, -3 were the most highly conserved, with rates of polymorphism of 0-1 %, 1.1 %, 0.6 %, and 0 %, respectively.

In the 1991 study, the 5' and 3' regions were extended. The other regions were found to have essentially the same results and therefore only the 1983 data was presented in Table 1 for these areas. Because more 5' data was given, the complete 660 bp of distal adult ADH promoter could be

compared to the proximal ADH and MT promoter regions. This data are presented in Tables 1 and 3 in sections 1.8 and 3.4 respectively. The comparison of the promoter regions is found in section 3.5. Statistical analysis was done using the chi-squared distribution, the results and explanation of which are found in Table 4. Comparison of the three promoter sequences together showed that they were statistically different in their respective rates of polymorphism, as did comparisons of the two ADH promoters on their own, and the ADH distal adult promoter with the MT promoter region analyzed. However, comparison of the ADH proximal larval promoter segment with the MT sequence revealed that the two regions were not statistically different. Results did not depend on whether or not the 430 bp maximum or 270 bp minimum values were used. Comparison of the first 225 bp of the three promoters was found not to be statistically different in respect to their rates of polymorphism, which was not the case for the comparison of the three for their full sequences, or the first 270 bp of each.

From these comparisons, it would appear that the three promoters are not all conserved to the same degree. the distal adult promoter was the least with 3.0 % polymorphism, but the precise regions required for the function of this promoter have not been mapped. It may be that the appropriate regions are not being compared. The important aspects of the promoter may be highly conserved with areas of changes interspersed between them. As well, the important aspects may be located in the first 225 bp with the remainder of

the 660 bp region not being functionally important. Because of this, the comparison of this area to the other promoter segments may not be appropriate because of the lack of knowledge about the promoter. This is also the case for the 385 bp proximal larval promoter segment, since sequence greater than 2 kbp upstream has been implicated in the function of the promoter. Although the 385 bp is relatively conserved with a % polymorphism of 0.5, the upstream portion has not been analyzed, so the comparison is not really complete. Thus, the relevance of these areas to promoter nucleotide polymorphism analysis is questionable. As well, although the MT promoter region sequenced was found to be highly conserved with a % polymorphism of 0.23 to 0.37, the complete promoter region was not sequenced. The extent of sequencing covered from the +1 site to the GRE, but the promoter extends as far as 500 bp upstream from this, where the IRE element is found. Although the region analyzed contained most of the regulatory elements and was found to be conserved, the remaining areas could also contain polymorphic sites which could influence promoter function. One might expect more change in this more 5' sequence since it apparently has less regulatory function associated with it. Despite the fact that all three promoter regions were not all highly conserved, it was found that the first 225 bp of each were, with % polymorphisms of 0.4 % for the MT region, and 0.9 and 1.3 % for the ADH promoters. These were not statistically different, indicating that this area of all three genes may require a higher level of

conservation between individuals and species to maintain proper transcription. This makes sense, since it contains the elements needed for appropriate mRNA transcriptional initiation and polymerase interaction. Otherwise it is difficult to make firm conclusions from the comparisons. The apparent differences in the promoter organizations and differential régulation make a meaningful comparison difficult. As well, the variable extent of sequencing and lack of knowledge of the ADH promoter analysis are complicating factors also.

4.3 Consideration of the relevance of these comparisons:

The issue of whether or not the populations are comparable is also important. The ADH study used genes from 11 strains from 5 different, geographically isolated populations. The current study used 10 individuals, but the population cannot be considered to be an isolated or single population, since no attempt was made to control for this. Thus, people from a variety of backgrounds were included, making the analysis more of a comparison of different populations as well, but not isolated populations. Data was not collected for such a consideration. The ADH study compares strains from different areas, so one might expect higher degrees of polymorphism in this comparison. With an isolated population one might expect a greater degree of inbreeding with a development towards a greater degree of homozygosity, in respect to polymorphic or recessive alleles. With an

increase in migration from other populations, an increase in heterozygosity is expected as different alleles, or polymorphisms are brought into the gene pool. Thus, if one compares isolated populations, a higher rate of polymorphism would not be unexpected because of this increase in homozygosity in each population as differences between them are emphasized. Because the current study is essentially an analysis within an unisolated population, and likely contains people from many different populations, it is difficult to compare. One might expect that because of this a lesser degree of homozygosity would be expected. However, it is important to keep in mind that these are very small isolated areas of DNA that have been analyzed. At best, 430 bp of MT-II_A sequence has been studied, and it indicates that a high degree of conservation. It is impossible to extrapolate this to other genes or populations, since one must account for the importance of the genes function and the selective pressures for a gene to change or not change in a given population in a given environment.

In summary, it is difficult to really compare these populations and to predict what one would expect in terms of outcome, since the selection of subjects or the MT study did not involve maintaining a specific subset of the population which might have been referred to as a more isolated group. Having done this might have made this more meaningful, but again, the region in question is very small and the number of people studied is relatively small. Thus, the promoter regions are different in degree of polymorphism

and it is difficult to say if this was expected or not. However, the first 225 bp were conserved to the same extent, which is likely significant. The fact that the two ADH promoter regions were not conserved to the same extent may indicate the existence of different selective pressures on each region, with one being able to accept more change with maintanence of function while the other remains relatively unchanged. Again, the factors above make this difficult to analyze properly.

4.4 Consideration of the MT-IIA promoter analysis in respect to the MT mRNA induction studies:

This analysis shows that the MT-II_A promoter region is highly conserved, indicating the gene likely has an important function for which its transcriptional ability remains well preserved. This is not necessarily unexpected since MT proteins have been found distantly in very primitive organisms, including yeast. It also has several potentially important functions, which could put selective pressure on the gene and its promoter region for sequence conservation to preserve function. Because of this, it appears that the gene likely has an important role now, as well as throughout evolution.

The degree of difference between the different individuals in the MT mRNA induction ratios between Cd treated and untreated cells was not likely a function of nucleotide variation within the MT genes of these people. If this

were the case, transcription would be the same for each. Thus, the variation seen between different people for MT mRNA Cd induction ratios is likely a function of another variable, or combination of variables. This could include differences in the expression of cellular transcriptional factors, differences in mRNA stability or degradation, variation in intestinal or cellular absorption of Cd, or differences in the exposure to other MT inducing agents that may affect an individuals ability to respond to Cd. The variation could be due to differences in the ability of different factors to interact with each other and with the MRE/BLE sequences, which would not appear to be a consequence of the sequences themselves. In order to draw legitimate conclusions from this kind of analysis in respect to MT mRNA induction variability, it would be necessary to establish a larger number of individuals. As well, this group would have to be representative of the spectrum of induction ranges seen in the previous study, including those with high (21-30 and 31-40 fold), medium (11-20 fold), and low (0-10 fold) Cd mRNA induction over basal mRNA expression ratios. In the previous analysis, most individuals were in the medium induction ratio group (21 people), with only a few in the high (5 at 21-30 fold, and 3 at 31-40 fold) and low (10 people) induction ratio groups. More people with equal representation in each level would be needed to achieve a statistically significant result. The current study population was derived largely from the medium induction ratio group since it was not possible to obtain a good distribution of subjects from those currently available from the previous study.

Thus, more people would need to be tested for purified peripheral lymphocyte MT induction ratios in response to cadmium in vitro, and then their MT genes would need to be cloned and sequenced in at least duplicate. This would be a great deal of work, which would not be warranted from the results obtained here. As a result, definitive conclusions in respect to this aspect of the study cannot be made.

As well, it should be noted that this is a very small sample size. In order to accurately estimate the variation within this region, one would ideally want to sequence many more MT gene promoter regions from other individuals. However, it does give some indication of nucleotide variation, and suggests that promoter variation is not responsible for mRNA induction variability as discussed above. In respect to the base change identified at position 557 for two of the ten people studied, analysis of the MT-II_A gene in many more people would be required in order to get an indication of its frequency in the population. This could be carried out using the polymerase chain reaction to generated DNA from genomic DNA samples and analyzed for the presence or absence of a SmaI site at 557. This was attempted unsuccessfully. As well, the implications if any, of this base change located between the GRE and MRE4 consensus sequences is not known, and would require further analysis. This could be done in an in vitro transcription system or by the use of an in vitro transfection system, to study the variability of the activity of each in a relatively controlled and isolated situation. Considering

that only one change has been found, it is unlikely that such a study would be justified, unless changes were found in the regulatory elements with the sequencing of more genes from more people.

4.5 Discussion of the MT-IIA sequence changes noted in the previous literature:

This study also has implications in respect to the accuracy of published sequences. Although the differences seen in the MT-II_A promoter sequences between the 1982, 1984, and 1987 publications could have been due to variability in the individuals for which the genes were cloned, it is more likely that at least some of these errors were due to misreading of the sequences and later correction, or errors in transposition of the sequence into the published or database sequences. These kinds of errors can have different implications. In this study, it is possible that an error in the original sequence resulted in the use of a mismatched sequencing probe, with the end result of poor sequencing results from that area. As well, comparison studies of different genes could also be affected. The initially published TRE consensus sequence was later revised and changed considerably. This sequence falls within the MT-II_A sequence which was found to have many errors in the 1982 and 1984 publications. The incidence in which such errors occur is not known, but from this analysis it could be postulated that there are likely many errors in published DNA sequences. It is also possible that this is specific to

the laboratory that published the earlier results, meaning that this kind of error is not very common and was not representative of the sequencing database in general. The identification of these discrepancies is significant because of the potential effects of errors in published genomic sequences, especially considering the current efforts to sequence the human genome. Errors in sequence could result in errors in the synthesis of probes used for sequencing as suggested above, as well as for hybridization for RFLP analysis. It could also create errors in database searches and sequence comparisons. These sequences are depended upon for many experiments, and errors in them can result in significant losses of time and resources.

4.6 Summary:

In conclusion, the MT-II_A promoter was analyzed from 10 people and was found to be highly conserved, with only one polymorphism found outside of the important regulatory elements. Although it would appear that polymorphism likely has little role in the variability of MT Cd-induced expression over basal level expression between different people from this study, further work would be needed to prove this, as described above. Ideally, it would be nice to sequence many more promoter regions to obtain a better estimation of the degree of conservation present. As well, the sequence for the entire promoter was not obtained. Thus, the study would be more meaningful if the entire gene was sequenced for each clone obtained for the

individuals studied. Current efforts continue in the laboratory to utilize PCR techniques to establish the sequence for more people, as well as the frequency of the polymorphism at position 557 between the GRE and MRE4 sequences, in the population. Although this was tried before, these techniques have become more refined and may allow for the progression of the study. The study also indicated that the actual sequence of the promoter region obtained has differences from the published sequences, and this sequence has been changed in the published database in the past without note. The importance of this is not known, but it may indicate that more error exists in published sequence than would be expected.

Thus, although efforts continue to establish the frequency of the identified polymorphism, and one would ideally want to study the full gene sequence in respect to its degree of polymorphism, it appears that the MT-II_A promoter is highly conserved, with only one change found in the region analyzed, and that the variability in Cd-induced MT-II_A mRNA transcripts over basal level expression is likely due to factors other than nucleotide polymorphism.

5. BIBLIOGRAPHY:

Andersen, R.D., Taplitz, S.J., Wong, S., Bristol, G., Larkin, B., and Herschman, H.R. (1987) Mol. Cell. Biol. 7: 3574-3581.

Andersen, R.D. and Weser, U. (1978) Biochem. J. 175: 841-852.

Angel, P., Baumann, I., Stein, B., Delius, H., Rahmsdorf, J.H., and Herrlich, P. (1987A) Mol. Cell. Biol. 7: 2256-2266.

Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R.J., Rahmsdorf, H.J., Jonat, C., Herrlich, P. and Karin, M. (1987B) Cell 49: 729-739.

Beato, M. (1989) Cell 56: 335-344.

Becker, P.B., Gloss, B., Schmid, U.S., and Schutz, S. (1986) Nature 324: 686-688.

Bell, J.U., Lawman, M.J.P., Lopez, J.M., DesJardins, L.E., and Applewhite, L.A.D. (1987) Experientia Supplementum 52: 581-586.

Beutler, E. and Yoshida, A. (1988) Medicine 67: 311-334.

Brady, F.O., Helvig, B.S., Funk, A.E., and Garrett, S.H. (1987) Experientia Supplementum 52: 555-563.

Bremner, I. (1987) Experientia Supplementum 52: 81-107.

Bremner, I., Mehra, R.K., and Sato, M. (1987) Experientia Supplementum 52: 507-517.

Brinster, R.L., Chen, H.Y., Warren, R., Sarthy, A., and Palmiter, R.D. (1982) Nature 296: 39-42.

Carter, A.D., Felber, B.K., Walling, M., Jubier, M., Schmidt, C.J., and Hamer,

D.H. (1984) Proc. Natl. Acad. Sci. USA 81: 7392-7396.

Chandler, V.L., Maler, B.A., and Yamamoto, K.R. (1983) Cell 33: 489-499.

Compere, S.J. and Palmiter, R.D. (1981) Cell 25: 233-240.

Cousins, R.J. (1985) Physiol. Review 65: 238.

Cousins, R.J. and Coppen, D.E. (1987) Experientia Supplementum 52: 545-553.

Cousins, R.J., Dunn, M.A., Leinart, A.S., Yedinak, K.C., and DiSilvestro, R.A. (1986) Am. J. Physiol. 251: E688-E694.

Cousins, R.J. and Leinart, A.S. (1988) FASEB J. 2: 2884-2890.

Culotta, V.C. and Hamer, D.H. (1989) Mol. Cell. Biol. 9: 1376-1380.

Danielsen, M., Hinck, L., and Ringold, G.M. (1989) Cell 57: 1131-1138.

DeFranco, D., Morris, S.M., Leonard, C.M. and Gluecksohn-Waelsch, S. (1988) Proc. Natl. Acad. Sci. USA 85: 1161-1164.

Dunn, M.A., Blalock, T.L., and Cousins, R.J. (1987) Pro. Soc. Exp. Bio. Med. 185: 107-119.

Durnam, D.M., Hoffman, J.S., Quaife, C.J., Benditt, E.P., Chen, H.Y., Brinster, R.L. and Palmiter, R.D. (1984) Proc. Natl. Acad. Sci. USA 81: 1053-1056.

Durnam, D.M. and Palmiter, R.D. (1981) J. Biol. Chem. 256: 5712-5716.

D'Urso, M., Luzzatto, L., Perroni, L., Ciccodicola, A., Gentile, G., Peluso, I., Persico, M. G., Pizzella, T., Toniolo, D., and Vulliamy, T.J. (1988) Am. J. Hum. Genet. 42: 735-741.

Dyan, W.S. and Tijan, R. (1983) Cell 35: 79-87.

Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132: 6-13.

Friedman, R.L. and Stark, G.R. (1985) Nature 314: 637-639.

Foster, R., Jahroudi, N., Varshney, U. and Gedamu, L. (1988) J. Biol. Chem. 263: 11528-11535.

Furst, P., Hu, S., Hackett, R. and Hamer, D. (1988) Cell 55: 705-717.

Galas, D. and Schmitz, A. (1978) Nuc. Acid Res. 5: 3157-3170.

Garg. L.C., Dixit, A., Webb, M.L., and Jacob, S.T. (1989) J. Biol. Chem. 264: 2134-2138.

Georges, M., Lequarre, A.S., Hanset, R., and Vassart, G. (1987) Animal Genetics 18: 41-50.

Glanville, N., Durnam, D.M. and Palmiter, R.D. (1981) Nature 292: 267-269.

Hager, L.J. and Palmiter, R.D. (1981) Nature 291: 340-342.

Hamer, D.H. (1986) Ann. Rev. Biochem. 55: 913-951.

Hanke, T., Tyers, M. and Harley, C.B. (1988) FEBS Letters 241: 159-163.

Harley, C.B., Menon, C.R., Rachubinski, R., and Nieboer, E. (1989) In press.

Haslinger, A. and Karin, M. (1985) Proc. Natl. Acad. Sci. USA 82: 8572-8576.

Heguy, A., West, A., Richards, R.I., and Karin, M. (1986) Mol. Cell. Biol. 6: 2149-2157.

Hildebrandt, C.E. and Cram, L.S. (1979) Proc. Soc. Exptl. Biol. Med. 161: 438-443.

Imbra, R.J. and Karin, M. (1987) Mol. Cell. Biol. 7: 1358-1363.

Israel, A., Kimura,A., Fournier,A., Fellous, M., Kourilsky, P. (1986) Nature 322: 743-746.

Johnson, G.F., Morell, A.G., Stockert, R.J., Sternlieb, I. (1981) Hepatology 1: 243-248.

Kadonaga, J.P., Jones, K.A., Tijan, R. (1986) Trends Biochem. 11: 20-23.

Kagi, J.H.R. and Kojima, Y. (1987) Experentia Supplementum. 52: 25-59.

Kagi, J.H.R., and Nordberg, M. (Editors) (1979) Metallothionein. Basel: Birkhauser-Verlag.

Kagi, J.H.R. and Schaffer, A. (1988) Biochemistry 27: 8509-8515.

Kagi, J.H.R., and Vallee, B.L. (1961) J. Biol. Chem. 236: 2435-2442.

Karin, M. (1985) Cell 41: 9-10.

Karin, M., Andersen, R.D., Slater, E., Smith, K., and Herschman, H.R. (1980) Nature 286: 295-297.

Karin, M., Eddy, R.L., Henry, W.M., Haley, L.L., Byers, M.G., and Shows, T.B. (1984A) Pro. Natl. Acad. Sci. USA 81: 5494-5498.

Karin, M., Haslinger, A., Heguy, A., Dietlin, T., and Cooke, T. (1987A) Mol. Cell. Biol. 7: 606-613.

Karin, M., Haslinger, A., Heguy, A., Dietlin, T. and Imbra, R. (1987B) Experientia Supplementum 52: 401-405.

Karin, M., Haslinger, A., Holtgreve, H., Cathala, G., Slater, E. and Baxter, J.D. (1984B) Cell 36: 371-379.

Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krauter, P., Westphal, H.M., and Beato, M. (1984C) Nature 308: 513-519.

Karin, M. and Richards, R.I. (1982) Nature 299: 797-802.

Karin, M., Slater, E.P., and Herschman, H.R. (1981) J. Cell. Physiol. 106: 63-74.

King, W.J., Greene, G.L. (1984) Nature 307: 745-747.

Koizumi, S., Sone, T., and Kimura, M. (1987) Experienta Supplementum 52: 519-523.

Kreitman, M. (1983) Nature 304: 412-417.

Kreitman, M., and Hudson, R.R. (1991) Genetics 127: 565-582.

Lee, W., Mitchell, P., and Tijan, R. (1987) Cell 49: 741-752.

Mayo, K.E. and Palmiter, R.D. (1981) J. Biol. Chem. 256: 2621-2624.

Maniatis, T., Fritsch, E.F., and Sambrook, J. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press; 1982.

Nebes, V.L., DeFranco, D., and Morris, Jr., S.M. (1988) Biochem. J. 255: 741-743.

Nieboer, E., Hertzman, C., Tom, R.T., Menon, C.R., Jusys, A.A., Crechiolo, J., Harley, C.B., and Muir, D.C.F. (1988) Biological monitoring of Ontario Cd workers: Final report. Report communicated by Dr. Nieboer, Occupational Health Program and Department of Biochemistry, McMaster University, Hamilton, Ontario, Canada.

Ogata, R.G., and Gilbert, W. (1978) Proc. Natl. Acad. Sci. 75: 5851-5854.

Palmiter, R.D. (1987) Experientia Supplementum 52: 63-79.

Peavy, D.L. and Fairchild II, E.J. (1987) Environ. Res. 42: 377-385.

Probst, G.S., Bousquet, N.F., Miya, T.S. (1977) Toxicol. Appl. Pharmacol. 39: 51-60.

Ptashne, M. (1988) Nature 335: 683-689.

Pulleyblank, D.E., Michiluk, M., Daisley, S.L., and Glick, R. (1983) Mol. Biol. Rep. 9: 191-195.

Quaife, C., Hammer, R.E., Mottet, K. and Palmiter, R.D. (1986) Develop. Biol. 118: 549-555.

Remington, R.D., and Schork, M.A. Statistics with applications to the Biological and Health Sciences. Prentice-Hall, Inc., New Jersey; 1985.

Richards, R.I., Heguy, A., and Karin, M. (1984) Cell 37: 263-272.

Riordan, J.R., Jolicoeur-Paquet, L. (1982) J. Biol. Chem. 257: 4639-4645.

Sadhu, C. and Gedamu, L. (1988) J. Biol. Chem. 263: 2679-2684.

Sambrook, J., Fritsch, E.F., and Maniatis, T. Molecular Cloning: A laboratory manual. Cold Spring Harbor Laboratory Press; 1989.

Schaefer, S.W., and Aquadro, C.F. (1987) Genetics 117: 61-73.

Scheidereit, C. et. al. (1983) Nature 304: 749-752.

Schmidt, C.J. and Hamer, D.H. (1986) Proc. Natl. Acad. Sci. USA 83: 3346-3350.

Schmidt, C.J., Hamer, D.H., and McBride, O.W. (1984) Science 224: 1104-1106.

Schmidt, C.J., Jubier, M.F., and Hamer, D.H. (1985) J. Biol. Chem. 260: 7731-7737.

Scholer, H., Haslinger, A., Heguy, A., Holtgreve, H. and Karin, M. (1986) Science 232: 76-80.

Searle, P.F. (1987) Biochem. Soc. Trans. 15: 584-586.

Searle, P.F., Davison, B.L., Stuart, G.W., Wilkie, T.M., Norstedt, G., Palmiter, R.D. (1984) Mol. Cell Biol. 4: 1221-1230.

Searle, P.F., Stuart, G.W., and Palmiter, R.D. (1987) Experientia Supplementum 52: 407-414.

Seguin, C. and Hamer, D.H. (1987) Science 235: 1383-1387.

Seguin, C. and Prevost, J. (1988) Nuc. Acid Res. 16: 10547-10560.

Sendelbach, L.E., and Klaassen, C.D. (1988) Toxicol. Appl. Pharmacol. 92: 95-102.

Shapiro, S.G., Squibb, K.S., Markowitz, L.A., and Cousins, D.J. (1978) Biochem. J. 175: 833-840.

Sone, T., Koizumi, S., and Kimura, M. (1988) Chem.-Biol. Interactions 66: 61-70.

Squibb, K.S., Cousins, R.J. (1977) Biochem. Biophys. Res. Commun. 75: 806-812.

Stuart, G.W., Searle, P.F., Chen, H.Y., Brinster, R.L. and Palmiter, R.D. (1984) Proc. Natl. Acad. Sci. USA 81: 7318-7322.

Szczypka, M.S. and Thiele, D.J. (1989) Mol. Cell. Biol. 9: 421-429.

Thiele, D.J., Walling, M., and Hamer, D.H. (1986) Science 231: 854-856.

Varshney, U. and Gedamu, L. (1984) Gene 31: 135-145.

Varshney, U., Jahroudi, N., Foster, R., and Gedamu, L. (1986) Mol. Cell. Biol. 6: 26-37.

Visentin, D. and Harley, C.B. (1987) Biochemistry 26: 6578-6583.

Vogelstein, B. and Gillespie, D. (1979) Proc. Natl. Acad. Sci. USA 76: 615-619.

Von Fraunhofer, J. A., and Murray, J. J. Statistics in Medical, Dental, and Biological Studies. Tri-med Books Ltd., London; 1976.

Webb, M. (1987) Experientia Supplementum 52: 109-134.

Welshons, W.V., Lieberman, M.E., Gorski, J. (1984) Nature 307: 747-749.

Westin, G. and Schaffner, W. (1988) EMBO J. 7: 3763-3770.

Willman, T. and Beato, M. (1986) Nature 324: 688-691.

Yamamoto, K.R. (1985) Ann. Rev. Genetics 19: 209-252.

Yamamoto, K.R., Alberts, B.M. (1974) J. Biol. Chem. 249: 7076-7086.