GENETIC AND MOLECULAR BASIS OF HEAVY METAL TOLERANCE AND THE HEAT SHOCK RESPONSE IN THE MEDITERRANEAN FRUIT FLY: *Ceratitis capitata*

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOMEDICAL SCIENCES
(Genetics)

MAY 1995

By

Sujinda Thanaphum

Dissertation Committee:

David S. Haymer, Chairperson
Terrence W. Lyttle
W. Dorsey Stuart
Donald O. McInnis
H. Gert de Couet

Acknowledgements

The author expresses his appreciation to Dr. David S. Haymer for his advice and to other committee members, Dr. Terrence W. Lyttle, Dr. W. Dorsey Stuart, Dr. Donald O. McInnis, and Dr. H. Gert de Couet.

Abstract

The Mediterranean fruit fly (Medfly), *Ceratitis capitata*, is a major pest of agricultural commodities world wide. Genetic and molecular studies of heavy metal tolerance and the heat shock response in the Medfly can provide a better understanding of how eukaryotes respond to environmental stresses. Potential uses of heavy metal tolerance and a heat shock gene in biological control programs have also been evaluated.

Three heavy metal tolerant strains were established. Genetic analyses of test crosses between these strains and a nonselected strain suggested different modes of inheritance underlying the heavy metal tolerance trait. An attempt to generate a Y-autosome translocation involving this trait appeared to not be feasible.

PCR products were amplified from the Medfly genome based on conserved domains of a metallothionein gene. Genomic DNA sequences homologous to these PCR products demonstrated coding potential metallothionein proteins, although transcripts homologous to these sequences were not detected in poly(A) selected RNA. This coding potential included a long open reading frame with a potentially cysteine rich region and alignment to mammalian metallothionein proteins.

A heat shock like gene (Cerhsc1) was cloned from the Medfly using PCR. This PCR used heavy metal induced cDNA as a template and a PCR primer designed from a conserved domain of metallothionein proteins. The Cerhsc1 gene shows some

structural characteristics of a heat inducible gene including the presence of several heat shock elements and the absence of an intron. The Cerhsc1 gene also exhibits a heat shock response resulting in a 10 to 20% increase in the level of transcription.

Developmentally, expression of the Cerhsc1 gene is very abundant in the adult stage, less so in the pupal stage and not detectable in the embryo stages. Overall, the Cerhsc1 gene exhibits structural and functional similarities to both heat inducible and non-heat inducible genes from a wide range of organisms.

Table of Contents

List of Tables

List of Figures

ATP: Adenosine-5'-triphosphate

Ceractin: an actin gene of *Ceratitis capitata*

C-terminal: carboxy terminal

cDNA: complementary DNA

Cd: cadmium

$CdCl_2$: cadmium chloride

Cu: copper

$CuSO_4$: copper sulphate

CHISM: chloroform and isoamyl alcohol in the ratio of 24:1

CRR: cysteine rich region

DTT: Dithiothreitol

DNA: Deoxyribonucleic acid

$Dp(Mtn^1)$: Duplicated Mtn^1 allele

Drohsc: heat shock cognate of *Drosophila melanogaster*

dATP: deoxyadenosine triphosphate

dCTP: deoxycytidine triphosphate

dCTP[α-^{32}P]: deoxycytidine5'-triphosphate, [α-^{32}P]

dGTP: deoxyguanosine triphosphate

dTTP: deoxythymidine triphosphate

dNTP: deoxynucleotide triphosphate

ddH_2O: double distilled water

ddATP: dideoxyadenosine triphosphate

ddCTP: dideoxycytidine triphosphate

ddGTP: dideoxyguanosine triphosphate

ddTTP: dideoxythymidine triphosphate

EDTA: Ethylenediaminetetraacetic acid

GCG: Genetics computer group

HCl: hydrochloric acid

HCN: hydrocyanic

HSP70: heat shock 70 kilodalton protein

Hsp70: heat shock 70 kilodalton gene

HSC: heat shock cognate protein

Hsc: heat shock cognate gene

HSF: heat shock factor

Hse: heat shock element

LB: Luria-Bertani bacterial medium

LiCl: lithium chloride

MK: Menkes disease

MOPS: morpholinopropanesulfonic acid

mRNA: messenger RNA

MT: metallothionein

Mtn^1: a metallothionein allele from *D. melanogaster*

Mtn^3: a metallothionein allele from *D. melanogaster*

N-terminal: amino terminal

NAD-GDH: nicotinamide-adenine dinucleotide-glyceral phosphate-dehydrogenase

NBT: nitro-blue-tetrazolium salt

NZCYM: bacterial medium

ORF: open reading frame

PCR: polymerase chain reaction

poly(A) RNA: polyadenylated RNA

RNA: ribonucleic acid

^{35}S-dATP: deoxyadenosine 5'-α-thiotriphosphate [^{35}S]

SDS: sodium dodecyl sulfate

Sod: superoxide dismutaste

SSC: sodium chloride and citric acid buffer

STET: sucrose, tris, EDTA, Triton X-100 buffer

TE: Tris-HCl and EDTA buffer

TEL: tetraethyl lead

Tris: Tris-hydroxymethyl-aminomethane

WND: Wilson's diesease

Zn: zinc

$ZnCl_2$: zinc chloride

Measurement

bp: base pair

^oC: degree celsius

Ci: curie

cm: centimeter

cpm: count per minute

gm: gram

Gy: Gray

kb: kilobase

kDa: kilodalton

mM: millimolar

M: molar

ng: nanogram

pM: picomolar

rpm: round per minute

uCi: microcurie

ul: microlitre

v: volt

Amino acids

Three letter code	One letter code	amino acid residue
Ala	A	Alanine
Arg	R	Arginine
Asn	N	Asparagine
Asp	D	Aspartic acid
Cys	C	Cysteine
Gln	Q	Glutamine
Glu	E	Glutamic acid
Gly	G	Glycine
His	H	Histidine
Ile	I	Isoleucine
Leu	L	Leucine
Lys	K	Lysine
Met	M	Methionein
Phe	F	Phenylalanine
Pro	P	Proline
Ser	S	Serine
Thr	T	Threonine
Trp	W	Tryptophan
Tyr	Y	Tyrosine
Val	V	Valine
	X	any amino acid residues

Bases in nucleic acids

A: adenine

C: cytosine

G: guanine

T: thymine

U: uracil

N: A or C or G or T

Y: C or T

Chapter 1

Introduction

The interrelationship between living beings and the environment is a fundamental aspect of life on Earth. As the planet Earth has changed over billions of years, the biosphere has changed as well. As the environment changes, species need to adapt and evolve in order to survive and reproduce.

Living cells exhibit both cellular homeostasis and the ability to react to external disturbances in physiological ways. Disturbance of the normal functioning of a biological system is defined as stress (Hoffmann and Parsons 1993). Frequently, the cellular response to stress involves the induction of specific gene expression. The induction of stress responsive genes generally produces abundant protein products which balance the physiological factors necessary for survival and growth under stressful conditions.

Among several type of stress responsive genes, metallothionein genes and heat shock genes are well-characterized and ubiquitious systems in most organisms. Heat shock genes are usually expressed in response to several kinds of stresses (heavy metals, anoxia, cold shock, amino acid analog, and etc.), although heat stress appears to be a major inductive agent (Parsell and Lindquist 1993). By comparison, the metallothionein genes appear to be specifically induced in response to high levels of particular heavy metals (Hamer 1986).

This work studies the responses to high levels of heavy metals in the Mediterranean fruit fly, *Ceratitis capitata* (Medfly). As part of this study, a heat shock like gene was isolated and characterized from the genome of the Medfly. Therefore, the molecular studies of stress response are examined both in terms of the heat shock response and the response to heavy metal.

1.1 Biochemical roles of heavy metal

Heavy metals are known to be essential for life (Klaassen and Suzuki 1991). Heavy metals are defined as those with an atomic weight larger than sodium (Hopkin 1989). However this definition of a heavy metal does not categorize nor specify biochemical properties.

Although heavy metals are usually found in trace amounts, they are often involved in crucial biological pathways. For example, ribozymes, which are thought to be one of the earliest functional molecules, usually require a metal ion such as zinc, magnesium, calcium and strontium for efficient chemical catalysis and for the stabilization of the ribozyme structure itself (Pyle 1993). Transcriptional regulatory factors such as the zinc-finger domain use heavy metals in the formation of specific DNA binding structures. In the case of the zinc-finger domain, zinc atoms and cysteine residues are coordinated in such a way that a finger-like DNA binding domain is formed (Klug and Rhodes 1987). Another example is that cobalt is a component of vitamin B12 which has a role in nucleoprotein synthesis (Hopkin 1989).

In cells, there are specific proteins associated with metals. These proteins are collectively called metalloproteins. Metalloproteins usually contain of metal ligands and/or metal clusters which appear to be necessary to maintain protein structure and function. Among these metalloproteins, some have roles in the control of metal-responsive gene expression (Karlin 1993, O'Halloran 1993). Example of metalloproteins are carboxypeptidase and phosphatase (Karlin 1993).

Although these heavy metals play crucial roles in molecular recognitions, it is also necessary to ensure optimal levels of the heavy metals within cells. When the level of any essential heavy metal is imbalanced, the effect can become toxic. In part this is because heavy metals can also participate in biochemical reactions and protein structures in which they are not normally involved.

The basis of heavy metal toxicity can be demonstrated using the example the of zinc element, an essential component of carbonic anhydrase. Carbonic anhydrase is an enzyme which catalyses the formation of calcium carbonate. If zinc is replaced by other heavy metal elements the activity of the enzyme is greatly reduced. For example, the activity of the carbonic anhydrase is reduced to 1% when zinc is substituted by copper (Hopkin 1989).

Organisms have various systems to maintain homeostasis of heavy metals and to counteract high levels of heavy metals found in the environment. Uptake systems are a primary process by which heavy metals enter cells. Delivery systems bring heavy metals from an uptake site (for example a digestive tract) to cells.

Storage systems usually keep excess heavy metals within appropriate molecules, organelles or tissues for future use. Detoxification and excretion systems dispose of extraneous heavy metals which would otherwise become toxic (Hopkin 1989).

High levels of heavy metals in a terrestrial environment can pose several problems for organisms because they persist longer than toxic organic compounds that can be broken down into less toxic molecules. Because of this, in certain environments, there may be strong selection in favour of individuals who can tolerate higher levels of heavy metals. Previous work has shown that there is genetic variation for tolerance to heavy metals (Maroni *et al.* 1987).

A major portion of this dissertation describes efforts to study the genetic basis of heavy metal tolerance in the Mediterranean fruit fly, *Ceratitis capitata* , and to isolate and characterize genes that may be directly associated with heavy metal tolerance.

1.2 Heavy metal tolerance

Metal tolerance describes the combined abilities of an organism to prevent, minimize, or repair the adverse effects of metals that have entered the body. These abilities are usually associated with major traits controlling the uptake, storage, and excretion of metals. Heavy metal tolerence traits are quite diverse in nature because of different feeding strategies, rates of food comsumption and assimilation, and the concentrations and chemical forms of metals in food. The anatomical physiology of a

digestive system and life history traits of an organism can also influence metal tolerence strategies (Hopkin 1989).

In organisms, heavy metal tolerance can also be associated with other traits. For instance in the collembolan insect, *Orchesella cincta*, cadmium tolerance is associated with traits such as cadmium excretion efficiency, molting frequency and relative growth rate (Posthuma *et al.* 1993). Therefore, the study of a metal tolerance trait may encompass other traits as well. This type of relationship often implies a higher level of complexity for molecular characterization and genetic analysis. However, examples where differences in heavy metal tolerance are clearly due to a single gene are known as well. For example, in *Drosophila melanogaster* there is a strong correlation between genetic variation at a metallothionein locus and levels of heavy metal tolerance (Maroni *et al.* 1987). Genetic research on heavy metal tolerance of this type, where genetic polymorphisms can be closely linked with a pattern of heavy metal tolerance, may allow researchers to analyse the underlying mode of inheritance of heavy metal tolerance in other organisms as well.

1.3 Mode of inheritance of heavy metal tolerance

Genetic studies of heavy metal tolerance trait have begun, as described above, with an observation that there are differences in the levels of metal tolerance among strains within a species. Primary observations of differential heavy metal tolerance have been made when certain organisms, especially insects, manifested

resistance to agents such as DDT and heavy metal-based insecticides (Melander 1914).

Resistance to insecticides is a ubiquitious problem. For example, insects have developed resistance to DDT in diverse regions of the world. In laboratory experiments, highly resistant strains can be selected by exposing insects to sublethal doses of insecticides for several generations. Test crosses between the highly resistant strains and non-resistant strains can reveal modes of inheritance of the resistance traits. As early as 1914, it was noticed that the use of hydrocyanic gas to control the red scale insect, *Aonidiella airantic* required increasing doses from year to year. Later on, it was shown that there were genetically distinct strains of the red scale insects with regard to their relative resistance to hydrocyanic gas (Dickson 1940). A test cross between a hydrocyanic (HCN) gas resistant strain and a non-resistant stain gave rise to F1 progeny with intermediate levels of hydrocyanic gas resistance. In the F2 progeny, the segregation pattern suggested that the genetic basis for the hydrocyanic gas resistance trait involved a single incompletely dominant gene (Dickson 1940).

Magnusson and Ramel (1986) studied genetic variation in the susceptibility to mercury and other heavy metal compounds in *D. melanogaster*. They found that the pattern of heavy metal tolerance was very different when different heavy metal compounds were applied. This suggests that the mechanism of genetic control of the resistance to heavy metal compounds is

specific for each compound. In a test cross between a highly resistant strain and a strain sensitive to methylmercury, they found that progeny from both reciprocal crosses showed levels of larval hatching close to the levels of the highly resistant parental strain. This result indicates dominance for the methylmercury resistance trait although in addition, there may be a polygenic basis to this resistance because it was found that the variation in the methylmercury tolerance among 12 strains of *D. melanogaster* was very high (Magnussan and Ramel 1986).

Studies of resistance to tetraethyl lead (TEL) suggest that the TEL tolerance trait is under additive genetic control. In an experiment involving TEL tolerance, Nassar (1979) selected *D. melanogaster* for 25 generations using a TEL treated diet. This selection give rise to an increase in fecundity, hatchability, and larva-to-adult viability (when raised with untreated diet) which may be directly associated with the TEL tolerance trait.

1.4 Gene families associated with heavy metal tolerance

There are gene familes that are associated with heavy metal tolerance traits. One of these is the Metallothionein gene family. These proteins encoded by these genes function as part of a natural biological defence against excess heavy metals (Karin 1985, Hamer 1986). Changes in the makeup of these metallothionein genes can directly effect levels of heavy metal tolerance (Maroni *et al.* 1987).

Another major gene family associated with heavy metal tolerance is known as the stress sensitive gene family. This gene family is found both in higher and lower eukaryotes. In higher eukaryotes, this stress sensitive gene family is usually referred to as the heat shock 70 (*hsp70*) gene family (Parsell and Lindquist 1993). Genes in the *hsp70* family respond to several kind of stresses such as elevated temperature as well as heavy metals (Lindquist and Craig 1988).

Apart from metallothionein and heat shock gene families there are also genes associated with heavy metal transport. Examples of these transporter genes include the MK (Menkes disease) and WND (Wilson's disease) genes which encode proteins belonging to a P-type ATPase family of cation transporters (Bull and Cox 1994). The MK and WND genes play a role in copper transport in humans. Mutations in these genes will disrupt copper transport and are the basis of Menkes disease and Wilson's disease. The relation of these heavy metal transporter genes to other aspects of heavy metal tolerance are not well understood. However, these genes are known to have amino acid sequences at the amino terminal end which are remarkably similar to sequences in bacterial heavy metal resistance proteins. This conserved region also includes the CXXC amino acid sequence that is known to play a role in heavy metal binding.

1.5 The metallothionein gene family

1.5.1 Metallothionein protein structure and function

Metallothioneins (MTs) are defined as "small cysteine-rich metal binding proteins" (Hamer 1986, Kojima 1991). These MTs bind to major heavy metals eg. zinc (Zn), copper (Cu), and cadmium (Cd) in the cell. Despite the fact that the function of MTs remains somewhat unclear, experimental evidence suggests a biochemical role for the MTs at least in terms of metabolism and detoxification (Bremner 1991, Templeton and Cherion 1991).

A Cd detoxification role has been demonstrated in various cell culture lines (Bremner 1991, Hamer 1986, Mokdad *et al.* 1987). The degree of Cd sensitivity (or Cd resistance) in each cell line corresponds to the level of the MTs, suggesting that MTs play a role in absorbing excess intracellular Cd (Mokdad *et al.* 1987). MTs also play many roles in biological processes involving Cu and Zn (Bremner 1991, Hamer 1986, Karin 1985, Klaassen and Suzuki 1991).

The MT proteins also have high cysteine residue contents. The content and position of cysteine residues are significant features of the MTs (Hamer 1986, Nemer *et al.* 1985). Mammalian MT proteins containing a total of 61 to 62 amino acids have 20 cysteine residues at conserved positions (Hamer 1986, Karin and Richards 1982). From a consensus sequence derived from mammalian MT genes, the cysteine residues usually are present in the form of a "metallothionein sequence motif" such as C-X-C or C-C, where X is any other amino acid residue (Nemer *et*

al. 1985). There is little similarity in terms of the cysteine positions between MT in mammals and other taxa. However, there are conserved sequences of 10 amino acids referred to as the "central segment" located at the center of the MTs in *Neurospora*, yeast, and *Drosophila* as well as in mammalian MT proteins (Nemer *et al.* 1985).

1.5.2 Molecular cloning, gene structure, and genomic representation of metallothionein genes

Most of MT genes were cloned using cDNAs and a plus/minus screening strategy. cDNA clones corresponding to MT genes from the genome of mouse (Durnam *et al.* 1980), human (Karin and Richards 1982), monkey (Schmidt and Hamer 1983), chinese hamster (Griffith *et al.* 1983), sea urchin (Nemer *et al.* 1984), and fruit fly (Lastowski-Derry *et al.* 1985, Mokdad *et al.* 1987) have been isolated.

In such a plus/minus screening experiments, cDNA derived from a heavy metal induced mRNA population is compared with cDNA from an uninduced source. Presumably, the abundance of mRNA corresponding to MT gene expression will be distinctively greater in the induced mRNA pool. MT specific cDNAs can be isolated after discounting all messages that are otherwise common between the two pools.

It may also possible to use heterologous probes in cloning experiments. However, the possibility of success of this approach is not clear because the MT DNA sequences are not highly

conserved. This approach has been used successfully when heterologous probes come from a closely related species. For example, a MT cDNA from *D. melanogaster* was used to screen MT genes from *D. simulans* and *D. ananassae* (Lange *et al.* 1990, Stephan 1994 *et al.*).

The overall structure of MT genes mammals and avians have two introns and three exons at precisely homologous positions (Hamer 1986, Klaassen *et al.* 1991). All of these intron/exon junctions conform to the GT...AG splice/donor consensus (Lewin 1994). Typical putative TATA boxes in 5' untranslated regions and polyadenylation signals in 3' untranslated regions have also been found in all MT genes characterized so far.

In higher organisms, the MT genes are usually found in multiple copies, and they typically show a close linkage relationship. All of the five functional human MT genes as well as some of the pseudogenes are closely linked on chromosome 16 (Schmidt *et al.* 1984).

There appear to be only two metallothionein genes in *D. melanogaster* (Maroni *et al.* 1986, Mokdad *et al.* 1987). These two are the Mtn and Mto genes. They are located on the right arm of the third chromosome at positions 85E10-15 and 92E, respectively.

1.5.3 Regulation of metallothionein gene expression

MT genes are usually considered to be housekeeping genes because they are expressed in various tissues and cell types. However, there is substantial quantitative variation in terms of gene expression during cellular differentiation as well as in different tissues of a mature organism (Hamer 1986, Nemer *et al.* 1984).

Studies have demonstrated that heavy metals can induce MT gene expression at the level of mRNA synthesis (Durnam *et al.* 1980, Durnam *et al.* 1981, Griffith *et al.* 1983, Karin and Richards 1982, Lastowski-Derry *et al.* 1985, Modad *et al.* 1987, Nemer *et al.* 1984, Schmidt *et al.* 1983). *In vitro* transcription of isolated-nuclei derived from a heavy metal induction system show higher levels of MT mRNA compared to an uninduced system. Since there is no differential mRNA degradation and processing of transcribed RNAs in these *in vitro* systems, these experiments suggest that transcriptional initiation is a major control point for MT gene regulation (Hamer 1986).

In higher organisms, a variety of heavy metals induce mammalian MT gene expression (Hamer 1986, Karin 1985, Klaassen and Suzuki 1991). However, it appears that in fungi only Cu is an inducer (Lerch 1980). *D. melanogaster* has both a Cu responsive gene (Mtn) and a Cd responsive gene (Mto) (Maroni *et al.* 1986, Mokdad *et al.* 1987). The optimal concentration of the heavy metals for induction varies from system to system.

Typically, it is just below the level causing cell toxicity (Hamer 1986).

In order to identify cis-acting control elements, the 5' regions of MT genes have been serially deleted and fused to heterologous coding sequences. Gene transfer of these hybrid constructs have been used to verify the DNA sequences necessary for MT gene expression with regard to various types of inducers and levels of transcription.

In the mouse MT I and human MT II genes, less than 60 base pair upstream of the transcriptional initiation site are required for heavy metals to induce MT gene expression. In addition, deletion constructs extending between the transcriptional initiation site up to the position -126 bp upstream have also shown heavy metal responsiveness. This fact indicates the coexistence of two regulatory elements. Within these two elements, there are imperfectly duplicated 12 nucleotide sequences (TGCGCYGGCCC) which might bind to similar molecules of a trans-acting factor (Karin *et al.* 1984, Stuart *et al.* 1984). In *D. melanogaster*, the Mtn gene control region is located from between positions 0 and -373 bp upstream of the transcriptional initation site. This region confers both heavy metal induced expression as well as tissue specific expression (Otto *et al.* 1987).

Trans-regulatory factors are less well understood in MT genes. Their molecular structure and mechanisms of regulation still remain unknown, however their regulatory role has been established. In the mouse system, positive regulation is evident.

This fact was deduced from competitive experiments showing that multiple copies of transfected cis-acting control elements titrate out the endogenous trans-regulatory factors (Carter *et al.* 1984). Afterward, the lower level of the trans-regulatory factor decreased the level of induced transcription but did not elevate the level of basal transcription. These regulatory factors appear to show characteristics of a transcriptional activator rather than a repressor.

1.5.4 Metallothionein gene duplication and amplification

An MT gene dosage effect is observed in invertebrates and lower organisms. Organisms possessing higher copy numbers of MT genes are less sensitive to heavy metals and show higher levels of MT gene expression. Yeast haplotypes have been identified containing 10 or more tandem copies of a 2 kb segment of DNA that includes the copper MT gene coding sequence (Fogel and Welch 1982). This haplotype was found in a heavy metal resistant line while a $cup^{S}1$ allele ,which contains only a single copy of the repetitive unit, was found in the wild type genome. The MT locus of *D. melanogaster* was also found to be duplicated in some individuals. These individuals usually exhibit approximately double amount of transcription as compared to the individuals possessing an unduplicated MT gene (Maroni *et al.* 1987).

Beyond this, the molecular mechanics of the MT gene duplication and amplification are not well understood. It is certainly possible that this might be advantageous in a heavy metal polluted environment (Lange 1990, Maroni *et al.* 1987, Theodore *et al.* 1991). In *D. melanogaster*, restriction fragment analysis of four different MT duplications did not suggest any obvious gene duplication mechanism. Each type is apparently different only in the length of the intervening sequences between the two duplicated MT genes (Maroni *et al.* 1987).

1.5.5 Evolution of metallothionein genes

The coding regions of MT genes are highly homologous among different mammals. The human MT II gene is 10% and 30% divergent from the mouse MT I gene for nonsynonomous and synonomous substitutions, respectively. The 5' and 3' untranslated sequences are typically more diverse, although some homologies are observed. The 5' and 3' untranslated regions of the monkey MT 2 gene and the human MT II gene are greater than 90% homologous (Schmidt and Hamer 1983).

The evolution of MT genes even among mammals is not clearly understood. For example, the degree of the MT sequence similarity does not always agree with mammalian phylogenetic relationships (Griffith *et al.* 1983, Hamer 1986, Schmidt and Hamer 1983). Hypotheses such as gene conversion and concerted evolution have been proposed to explain these results.

Nonetheless, it is difficult to test these hypostheses because the MT gene family is large and complex in these taxa.

Dipteran species appear to have a simpler and smaller MT gene family. There is only a single Mtn locus in *D. melanogaster*, *D. simulans*, *D. yakuba*, *D. tessieri*, *D. sechellia* (Theodor *et al.* 1991), *D. ananassae* and *D. pseudoobscura* (Stephan *et al.* 1994). Allelic variation has been found in *D. melanogaster* (Maroni *et al.* 1986, Maroni *et al.* 1987, Theodore *et al.* 1991). One allele designated Mtn^3 appears to be an ancestral allele of another allele (Mtn^1). A duplication ($DpMtn^1$) of this allele is also common among *Drosophila* species (Theodore *et al.* 1991, Stephan *et al.* 1994).

Since the surveyed duplication allele $Dp(Mtn^1)$ appears to derive from Mtn^1, it has been proposed that there were two evolutionary steps, ie. one step from Mtn^3 to Mtn^1 and a second step from Mtn1 to $Dp(Mtn^1)$ (Theodore *et al.* 1991). This appears to have been accompanied by an overall 5 to 6 fold increase in mRNA accumulation. Mtn^1 mRNA is 65% to 70% more abundant than Mtn^3 mRNA and $Dp(Mtn^1)$ mRNA doubles the amounts of Mtn1 transcripts. This observation suggests directional selection favoring heavy metal tolerance traits in natural populations. Further studies of the genetic basis on MT genes and heavy metal tolerance are essential to understand more about how these heavy metal tolerance traits might have evolved.

1.6 The heat shock 70 kDa gene family

1.6.1 Heat shock 70 protein function

The products of the *hsp70* genes are a large group of proteins with a molecular weight around 70 kDa, actually ranging from 67 to 74 kDa. These are collectively referred to as HSP70 proteins. After treatment with either moderately or extremely high temperature, these are among the most abundant proteins found in a cell, and they are usually referred to as heat inducible *hsp70s*. Another type of HSP70 protein appears throughout normal developmental stages expressed in a constitutive manner. These are the so-called heat shock cognate (HSC) proteins (Ingolia and Craig 1982). The genes producing these HSC proteins are different from the *hsp70* inducible genes and are referred to as the heat shock cognate (*hsc*) genes. The *hsc* gene products are localized to a variety of cellular compartments in addition to the cytoplasm. There is usually also at least one specific heat shock cognate protein in the endoplasmic recticulum (ER), as well as the mitochondrion or chloroplasts.

Since *hsp70* genes appear to exist in all organisms, their function would appear to be of fundamental importance. The HSP70 proteins also demonstrate a general role as a molecular chaperones. Molecular chaperoning refers to a cellular process by which biomolecules acts as a facilitators or mediators of various cellular processes (Ellis and van der Vies 1991). HSP70 is known to be able to recognize and stabilize nascent chains of unfolded proteins, partially unfolded intermediate proteins, transporting

proteins, as well as denatured or abnormal proteins. They also facilitate assembly and disassembly of multimeric structures within several cellular compartments. The HSP70 protein is distinct from other heat shock proteins partly because their function as chaperones appears to be more general. The HSP70 protein appears to recognise simple structural motifs which are shared by many unfolded proteins (Flynn *et al.* 1991, Gething and Sambrook 1992, Landry *et al.* 1992). However, specific roles for each *hsp70* gene can be demonstrated. For example, a mutation in a *hsp70* gene member cannot always be completely rescued by other related *hsp70* genes (Werner-Washburne 1987).

The specific roles that the different HSP70 proteins appear to have in different places in cells have in many cases been demonstrated by experiments in yeast. Here, it has been shown that a disruption of one *hsp70* gene cannot be completely rescued by the other functional *hsp70* genes, even for those which are which are localized in the same cellular compartment (Craig and Jacobsen 1984).

In a (stressful) situation where cellular proteins could be damaged, a substantial amount of HSP70 protein would be needed to ensure that: 1) all forms of unfolded proteins are well protected and 2) the already degraded proteins will not induce further damage by aggregation. Clearly, the heat inducible heat shock proteins of the HSP70 protein family have to be capable of being produced at high levels in order to have protective roles against general protein damage.

1.6.2 Heat shock 70 kDa protein structure

HSP70 is among the most highly conserved proteins known. An HSP70 protein of *Halobacterium marismortui* shows between 44 and 47% amino acid identity with various eukaryotic *hsp70* proteins and between 51 and 58% amino acid identity with its eubacterial and archaebacterial HSP70 proteins (Gupta and Singh 1992).

Boorstein *et al.* (1994) compared 29 different HSP70 protein sequences derived from 24 organisms across various taxa. Comparison of the HSP70 proteins demonstrated a high level of conservation at this level. More than one-sixth of the amino acids from these 29 HSP70 proteins are identical in the aligned region. This is consistent with the 50% identity typically among heat shock amino acid sequences. Most of these identical amino acid residues appear to be clustered as small blocks within the first one third of the sequence. Amino acid residues located at the extreme C-terminal end in various HSP70 proteins appear to be in a more variable region in terms of amino acid sequence content and length (Boorstein *et al.* 1994).

The HSP70 proteins usually share two highly conserved functional and structural domains, one at the amino terminal and one at the carboxy terminal (figure 1). A 44 kDa proteolytic fragment from the amino terminal end of a mammalian cytosolic HSP70 protein is capable of binding to ATP and it possesses ATPase activity (Flynn *et al.* 1991). The three dimentional structure of this ATP binding domain reveals a two-lobed

structure containing a groove in which an ATP molecule can bind (Flaherty *et al.* 1990). The ATP binding domain appears to be the most conserved region in HSP70 proteins in term of amino acid sequence. The three dimensional structure of this ATP binding domain is similar to an ATP binding motif of a number of different proteins. These proteins are actin, hexokinase, and HIV-1 reverse transcriptase (Flaherty *et al.* 1990, Artymiuk *et al.* 1993). The multiple sequence alignment in the ATP binding motif of the actin, hexokinase and HSP70 proteins also showed a set of common conserved amino acid residues distributed in five sequence motifs (Bork *et al.* 1992). These similarities suggest that the actin, hexokinase, and the HSP70 proteins belong to a protein superfamily.

The C-terminal half of the HSP70 protein is a less conserved domain (figure 1). The C-terminal region is hypothesized to function as a general protein binding domain. A consensus sequence drawn from an alignment of 33 different HSP70 proteins suggested a possible secondary structure with beta four alpha motifs repeated twice. These two beta four alpha motif repeats are usually followed by two long (20-25 residues) alpha-helices and a terminus rich in proline and glycine residues. This beta four alpha helix structure is also characteristic of the secondary structure of a peptide binding domain of human leucocyte antigen (HLA) class I protein (Rippmann *et al.* 1991).

Although all investigated HSP70 proteins share these two domains, the overall amino acid sequence similarity appears to be

higher between HSP70 proteins that exhibit the same subcellular localization. For example, the amino acid sequence of a cytoplasmic HSP70 protein of *D. melanogaster* is more similar to a cytoplasmic HSP70 of humans than to an endoplasmic recticulum localized HSP70 of *D. melanogaster* . The amino acid sequence of this cytoplasmic HSP70 of *D. melanogaster* shows 82% identity to the cytoplasmic HSP70 in human while showing only 61% identity to the endoplasmic reticulum localized HSP70 of *D. melanogaster* (Perkin *et al.* 1990, Rubin *et al.* 1993).

In addition to the overall amino acid sequence similarity among organelle specific HSP70 proteins, endoplasmic reticulum localized HSP70 proteins usually have a leader sequence consisting of a large number of hydrophobic amino acid residues including A, V, L, I, P, W, F, and M at the amino terminal end (figure 1). At the carboxy terminal end conserved tetrapeptide sequences KDEL or HDEL are also always found (figure 1). The protein produced by the *hsc3* gene of *D. melanogaster* is a prime example of this class of *hsp70* like genes (Rubin *et al.* 1993). This characteristic of having conserved sequences at the termini is common in several different types of proteins that are associated with luminal endoplasmic reticulum.

Mitochondrion localized HSP70 proteins, such as *hsc5* in *D. melanogaster*, also usually contain leader sequences consisting of basic amino acids (such as K, R, and H) spaced 3 to 5 residues apart, as well as S residues interspersed though out the length of the leader sequence. This type of leader is again a characteristic

shared among numerous mitochondrion associated proteins (Rubin *et al.* 1993).

The HSP70 proteins localized to the cytoplasm do not appear to contain similar leader sequences. However, it is characteristic for them to contain the amino acid sequences like GP(T/K)(V/I) EEVD at the end of the carboxy terminal region (figure 1) (Boorstein *et al.* 1994).

1.6.3 Molecular cloning of *hsp70* genes

An *hsp70* gene was first isolated from *D. melanogaster* (Lis *et al.* 1978) using labeled cytoplasmic poly(A)+ RNA extracted from heat shocked cells to probe a genomic library. In order to diminish possibilities that genomic clones carrying genes other than the *hsp70* were inadvertantly isolated, all of the isolated genomic clones were probed with similar probe that derived from a non heat shock cells. The *hsp70* gene was subsequently used as a probe to isolate the other *hsp70* gene family members from the *D. melanogaster* genome.

The *D. melanogaster hsp70* gene has also been used as a heterologous probe in the isolation of homologous genes from many other species, including diverse taxa such as yeast (Ingolia *et al.* 1982) and humans (Dworniczak and Mirault 1987). If the isolated heat shock genes were heat inducible, they were called heat inducible *hsp70*. Likewise, if the isolated heat shock genes exhibited no obvious heat inducibility but constitutively

expressed at high levels at normal conditions, they were called heat shock cognate (*hsc*) genes.

Probes containing the protein coding regions of these *hsp70* genes were subsequently used to isolate other members of the *hsp70* gene family within these organisms. Because of this, it should not be surprising that there is some sequence similarity among almost all of the *hsp70* genes that have been isolated.

There are only a few heat shock like genes that have been isolated by a method other than DNA probing and hybridization using the *hsp70* sequence of *D. melanogaster*. One example of this is a constitutively expressed *hsc* gene in a fungus-like organism, *Achlya klebsiana*. This *hsc* gene was isolated using an antibody screening of cDNA expression library. The antibody used here was a polyclonal antibody raised against a purified subunit enzyme NAD^+-specific glutamate dehydrogenase (NAD-GDH) (Yang and LeJohn 1994). Both the NAD-GDH gene and the constitutive *hsc* gene are inducible by L-glutamine. Under conditions of nutritional stress, L-glutamine augments the expression of this normally constitutive *hsc* gene (LeJohn *et al.* 1994). The HSC protein sequence of this gene is 69.4% similar to a human HSP70 protein. At the nucleic acid level, it is 67.3% similar to the human *hsp70* gene.

A *hsp70* like gene that has been isolated and characterized in the work described here is another example. This *hsp70* like gene is also one of the few heat shock genes that have been isolated on a basis other than DNA probing and hybridization

using previously isolated *hsp70* genes. Different molecular cloning approaches should reveal different characteristics of the *hsp* and *hsc* gene families.

1.6.4 Genomic representation of heat shock genes

There are a few model organisms in which the heat shock gene family has been extensively studied. A heat shock gene family usually includes heat inducible *hsp* genes and constitutively expressed *hsc* genes.

In *D. melanogaster*, the *hsp70* gene family includes a total of five heat inducible genes, one *hsp68* gene, and at least 7 *hsc* genes (*hsc1* to *hsc7*). The heat inducible *hsp70* genes are found at two polytene chromosome locations, 87A7 and 87C1. At the 87A7 locus, there are two divergently transcribed *hsp70* genes separated by a 1.6 kb spacer. At the 87C1 locus, a proximal gene (*hsp70c1p*) is separated from two distal genes by 40 kb. (Maroni 1993). The *hsc1* to *hsc6* genes were mapped cytologically at the 70C, 87D, 10E, 88E, 50E, and 5C regions, respectively (Lindquist and Craig 1988). The heat inducible *hsp68* gene maps to the 95D region.

The yeast, *Saccharomyces cerevisiae*, has four *hsp70* subfamilies. Subfamily SSA consists of genes designated SSA1 to SSA4. These genes are both heat inducible and cytoplasmically localized in terms of expression. Two genes in the subfamily SSB are also cytoplasmically localized, but these genes are constitutively expressed. The SSC1 gene represents another

hsp70 subfamily which is localized in the mitochondrion. The so called SSD1 gene (also known as KAR2) is localized to the endoplasmic recticulum (Boornstein *et al.* 1994). The genomic distribution of some of these *hsp70* genes is dispersed. The SSA1 gene is located on the left arm of chromosome 1 (Lindquist and Craig 1988); the SSC1 gene is on the right arm of chromosome X; the KAR2 gene is on the left arm of chromosome X; the SSA3 gene is located on the left arm of chromosome II and the SSB2 gene was mapped to the left arm of chromosome 14 (Boorstein *et al.* 1994).

In the human genome, there are at least ten *hsp70* related genes, although the number of these which are functional is not clear (Mues *et al.* 1986). *Hsp70* related genes have been shown to be located on chromosomes 6, 14, and 21 (Goate *et al.* 1987, Harrison *et al.* 1987). Dworniczak and Mirault (1987) screened a human genomic DNA library with an *hsp70* probe from *D. melanogaster*. After sequencing a number of the human *hsp70* related genes isolated, they found that some were processed pseudogenes. There is also a heat inducible human *hsp70* gene that has recently been characterized (Hunt 1985). A human *hsc* gene was also discovered. This human *hsc* gene had several long intervening sequences, therefore the whole *hsc* gene was at least 5 kb long (Dworniczak and Mirault 1987).

In maize, there are a number of heat inducible *hsp70* and *hsc* genes which have been identified by DNA hybridization to a *D. melanogaster hsp70* sequence. Genomic locations of these maize

hsp70 genes have not been studied to the same extent as in other model organisms. However, a striking feature of these plant *hsp70* genes is that they exhibit characteristics of both heat-inducible and heat shock cognate genes. Two of the maize *hsp70* genes show 40 to 60 fold increases in transcription levels at high temperatures. However these genes also have an intron in a position identical to the *hsc1* gene in *D. melanogaster* and to a human *hsc* gene. Heat inducible genes in animals do not have introns in this position. One of the maize HSP70 protein sequence is also more similar to the amino acid sequence of the *hsc1* gene of *D. melanogaster* (75% homology) than it is to the amino acid sequences of the heat inducible *hsp70* gene of *D. melanogaster* (68% homology) (Rochester *et al.* 1986).

There are also *hsp70* like genes (the so called dnak gene) in prokaryotes such as *Escherichia coli.* The dnak gene encodes a protein that has 50% identity at amino acid level to a *hsp70* gene of *D. melanogaster* (Bardwell and Craig 1984). The dnak gene exhibits a baseline level of expression but is also heat inducible. For a number of years, it was believed that the dnak gene was the only *hsp70* like gene in the prokaryotic genome. It has recently been learned that there are two more related genes, *hsc66* and *Mre*, within the prokaryotic genome (Seaton and Vickery 1993, Gupta and Singh 1992). These two bacterial *hsp70* genes show somewhat lower levels of amino acid sequence identity when compared to the dnak protein. The lower level of the protein sequence identity may explain the delay in identification of these

related genes. This may also be another example of how the study of *hsp70* genomic representation may be constrained if it is only done on the basis of DNA hybridization and nucleic acid sequence similarity. More recently, a number of other distantly related members of the *hsp70* gene family have been identified without relying on *hsp70* DNA homology. Perhaps not surprisingly, their amino acid sequences usually reveal low (but significant) identity compared to the previously isolated *hsp70* genes (Folz *et al.* 1993).

1.6.5 *Hsp70* gene structure

Many of the *hsp70* coding regions are similar in length. The lengths of the *hsp70* coding regions usually appear to be no larger than 650 amino acids (1950 nucleotides) long (Boorstein *et al.* 1994). In higher eukaryotes, the coding regions of some *hsp70* genes are interrupted by introns. This is especially true for the constitutively expressed *hsc* genes. The introns appear to be absent in the heat inducible *hsp70* genes. It was proposed that the absent of intron would make heat induction more effective because there is no need to excise introns.

The appearance of introns within the *hsp70* multigene family appears to be consistent with a theory that an explosion of exon-shuffling occured at the time of metazoan radiation (Gilbert 1987, Patthy 1994). This theory states that exon-shuffling is a relatively late evolutionary event. This is thought to have occured at a time when novel proteins were needed to maintain cell-to-

cell communication. The presence of introns in the heat shock 70 kDa multigene family appears to be restricted to higher eukaryotes. This suggests that introns invaded the heat shock multigene family after the split from lower eukaryotes. For example, an intron position corresponding to (amino acid position) codon 65 of *hsc1* of *D. melanogaster* is found to be conserved in a *hsc* gene in both the human and maize genomes (Ingolia and Craig 1982, Dworniczak and Mirault 1987, Rochester *et al.* 1986). Introns are also found in *hsc* genes of *D. melanogaster* as well as in many other higher eukaryotes. For examples, the *hsc4* gene of *D. melanogaster* has one intron in the 5' untranslated region (Perkin *et al.* 1990). The *hsc3* and *hsc5* have one and three introns, respectively (Rubin *et al.* 1993). In contrast, there is no evidence of an intron in the *hsp70* multigene family of lower eukaryotes. For example, none of at least five *hsp70* genes of yeast representing all four subfamilies SSA, SSB, SSC, and SSD have any introns (Ingolia *et al.*1982, Craig *et al.* 1989, Rose *et al.* 1989, Boornstein *et al.* 1994).

Several of the 5' regulatory regions of the *hsp70* genes have been investigated to determine whether or not there are common molecular features underlying the regulation of heat shock gene expression. Common elements referred to as heat shock elements (*hse*) have often been found within 150 bases upstream of TATA boxes for all investigated *hsp70* genes (Pelham 1982) The *hse* consensus sequence is CTNGAANNTTCNAG (figure 2) (Pelham 1982, Pelham and Bienz 1982). These studies showed that this

consensus sequence consists of alternating repeats of a 5 base unit NGAAN and its reverse complement NTTCN (Pelham 1982, Xiao and Lis 1988, Perisic *et al.* 1989, Xiao *et al.* 1991). This *hse* consensus sequence was deduced primarily from a number of comparative studies of different *hsp70* genes which included both heat inducible *hsp70* genes as well as the constitutively expressed *hsc* genes (Ingolia *et al.* 1980, Holmgren *et al.* 1981, Pelham 1982, Hoffman and Corces 1980). In a related investigation, *hsp70* promotor regions were serially deleted and rejoined either to their own coding regions or to marker genes for expression in heterologous systems. This experiment demonstrated that a segment of DNA which included at least two copies of *hse* are necessary, but not sufficient, for a normal level of heat inducible *hsp70* transcription (Pelham 1982). The segment of DNA containing at least two copies of *hse* together is called a regulatory unit. The regulatory unit works in both orientations. In addition, the distance between the heat shock regulatory unit and the TATA box was not critical. Finally, *hsp70* genes with 3 to 5 copies of *hse* are usually more heat inducible than *hsp70* genes with only two copies of *hse* (Amin *et al.* 1987). These findings suggest a positive mode of *hsp70* gene regulation.

However, there is also evidence for a negative control model. Corces *et al.* (1984) showed that a deletion of an *hse* decreased only the high level of transcription but not the heat inducibility. In addition, there are two sequences flanking the TATA box that when deleted cause the heat shock promotor to

lose its inducibility. Hence, there might be at least two different regulatory elements, specifically one (*hse*) is that controls the high level of gene expression and another that confers inducibility (Corces *et al.* 1984). The idea that the *hse* regulatory unit is important for high levels of gene expression is also supported by the fact that there are similar regions in *hsc* promotors which show high levels of expression but no obvious heat induction capability (Craig *et al.* 1983).

1.6.6 *Hsp70* gene regulation

In higher eukaryotes, transcription of the *hsp70* genes appear to be activated by a protein-DNA interaction involving a protein known as the heat shock factor (HSF) and the *hses*(figure 2) (Xiao *et al.* 1991). HSF seems to exist within cells even before heat induction. However after heat induction, the HSF appears to acquire a higher DNA binding affinity probably through auto-oligomerization (Clos *et al.* 1990, Westwood *et al.* 1991).

Another *hsp70* related transcriptional factor is called GAGA. This GAGA factor was found to bind to a segment of alternating CT or GA sequences in the 5' regulatory region of *hsp70* genes (figure 2). This GAGA protein is a constitutively expressed transcriptional factor that can bind to GA/CT rich sites present in many *Drosophila* genes (Biggin and Tjian 1988). It was demonstrated *in vitro* that the introduction of the GAGA protein with ATP altered chromatin structure of *hsp70* promotors during and after nucleosome assembly (Tsukiyama *et al.* 1994). This

chromatin alteration could be indicative of an accessible promotor region poised to respond to a formation of a general transcriptional activation complex as well as to the HSF protein.

It has been shown (in *Drosophila*) that at a normal temperature RNA polymerase II binds to a region representing the transcriptional initiation site of the *hsp70* genes. Transcription is initiated, but pauses after synthesizing a short transcript approximately 25 nucleotides in length. It is only after the binding of the HSF that transcription continues and enlongation proceeds at a much elevated rate. However, even in extreme heat shock conditions, the presence of transiently paused polymerases can still be detected (O'Brien and Lis 1991, Rougvie and Lis 1988, Rasmussen and Lis 1993). This suggestes that transcriptional elongation is a rate limiting step in *hsp70* gene induction.

In *D. melanogaster*, *hsp70* mRNA was found to be very stable and efficiently translated at 36°C, however, at 25°C, it had a half life of only 15 to 30 minutes. The short half life was found to be associated with an AU rich sequence in the 3' untranslated region of the *hsp70* mRNAs (Petersen and Lindquist 1988). The temperature-dependent half life of the *hsp70* mRNA could be a mechanism to further regulate the *hsp70* expression.

At temperatures above 35°C, the translational apparatus of *D. melanogaster* shifts from production of normal proteins toward HSP70 proteins. This shift toward the translation of *hsp70* mRNA was found to be associated with information embedded within the

5' untranslated leader sequence of the *hsp70* mRNAs (figure 2) (Kelmenz *et al.* 1985).

1.6.7 Previous work on heat shock proteins in the Medfly

In the Medfly, previous studies of heat shock gene expression have been at the level of protein synthesis (Stephanou *et al.* 1983, Stephanou 1987, Jang 1992). Salivary glands from the third instar larvae of the Medfly appear to synthesize a set of at least eight specific polypeptides in response to heat shock at 37°C for 30 minutes. Here again, overall protein synthesis is reduced. The most predominant heat shock protein produced has a molecular weight of 69 kDa which might correspond to the HSP70 proteins of *D. melanogaster*. The other heat shock proteins have been named HSP87, HSP34, HSP20, HSP16, HSP14, HSP13 and HSP12 after their corresponding molecular weights in kDa units. When protein synthesis was studied at a range of temperature from 25°C to 41°C, it revealed that each of the eight heat shock proteins respond differently. The HSP69 protein is predominantly synthesized between 37°C and 41°C, while the HSP87 protein is synthesized between 33°C and 35°C (Stephanou *et al.* 1983). The HSP69 protein appears to be optimally produced in response to a heat shock between 30 to 60 minutes in length. A gradual reduction of the HSP69 protein synthesis was noticed for a heat shock period of 30 minutes to 150 minutes. Synthesis usually

ceases approximately 180 minutes after the beginning of the heat induction.

There is a strong positive correlation between the ability of the Medfly larvae to survive a heat shock and the cellular level of the HSP69 protein synthesis. It has also been shown that when Medfly larvae are mildly preheated at 35°C for 30 minutes to one hour, they demonstrate higher levels of HSP69 protein synthesis and higher levels of larval survival compared to untreated larvae in response to a subsequent high temperature heat shock at 45°C for 40 minutes (Stephanou *et al.* 1983).

The control of HSP69 protein synthesis in the Medfly appears to operate at both transcriptional and translational levels. At the transcriptional level, cultured cell lines pretreated with actinomycin D did not show HSP69 protein synthesis after a heat shock at 37°C for 30 minutes to one hour (Jang 1992). Since actinomycin D suppresses mRNA production, this infers that there is not a high level of mRNA of the HSP69 protein available for selective translation prior to a heat shock treatment.

However, there is also some evidence for translational control of the HSP69 protein in the Medfly. This is because there appeared to be equal amounts of transcripts of the HSP69 protein from two systems that show different levels of the HSP69 protein synthesis (Stephanou 1987).

Optimum temperatures for HSP69 protein synthesis in the Medfly appear to be between 37°C and 41°C. This compares to between 35°C to 37°C for *D. melanogaster*. The difference in the

temperature for heat shock induction between the Medfly and *D. melanogaster* may reflect differences in the optimal environmental temperatures for the two species (Stephanou 1987).

The temperature range for the heat shock induction in the Medfly appears to be similar to many other insect species. For example, the optimal heat shock response temperature for *Manduca sexta* (tobacco hornworm) is 42°C (Fittinghalt and Riddiford 1988), for *Locusta migratoria* (locusts) it is 45°C (Walker *et al.* 1986), and for *Sarcophaga crassipalsii* (the flesh fly) it is between 43°C to 45°C (Joplin and Denlinger 1990). Studies of heat shock gene structure and expression in these different insects may provide new insights into this complex gene family.

1.7 Background on the Medfly

The Medfly appears to have originated in Africa. Within the last few hundred years it has spread to a number of countries around the world. This insect is polyphagous and multivotine (Fletcher 1989).

The Medfly is one of the most destructive pests around the world. This agriculture pest appears to be critical throughout the Mediterranean region, Central and South America, South Africa, South-western Australia, Hawaii and so on. The females insert their eggs into various kinds of fruit and also induce pathogenic infections. Therefore, many control programs have been initiated (Saul 1986).

The Sterile Insect Release Method (SIRM), which is part of a desirable integrated pest management (IPM) program, has been developed to control the Medfly (La Chance 1979). However, one disadvantage in this practice at present is the release of females. The ability to employ techniques such as genetic engineering and germ line transformation on agriculture pest species such as the Medfly should be of great value in increasing the efficiency of SIRM programs (Robinson 1989).

A possibility of using molecular genetics to improve the SIRM programs relies on basic knowledge of the molecular biology of tephritid species. To date, only a few genes from Medfly have been characterized. Example of these genes are an actin gene (Haymer *et al.* 1990), chorion genes (Konsolaki *et al.* 1990), vitellogenin genes (Rina and Savakis), a Cu, Zn superoxide dismutase gene (Kwaitowski *et al.* 1992), and an alcohol dehydrogenase gene (Gasperi *et al.* 1994). Molecular genetic studies of these Medfly genes have provided valuable basic knowledge even though none of these genes serve directly in any SIRM program. This work describes efforts to isolate and characterize genes such as metallothionein and heat shock genes that may be of more direct value in such programs. These genes have been used in germline transformation schemes in other organisms, and these genes appear to work best when they are endogenous or derived from very closely related species.

Figure 1. **Molecular features of heat shock 70 kDa proteins**

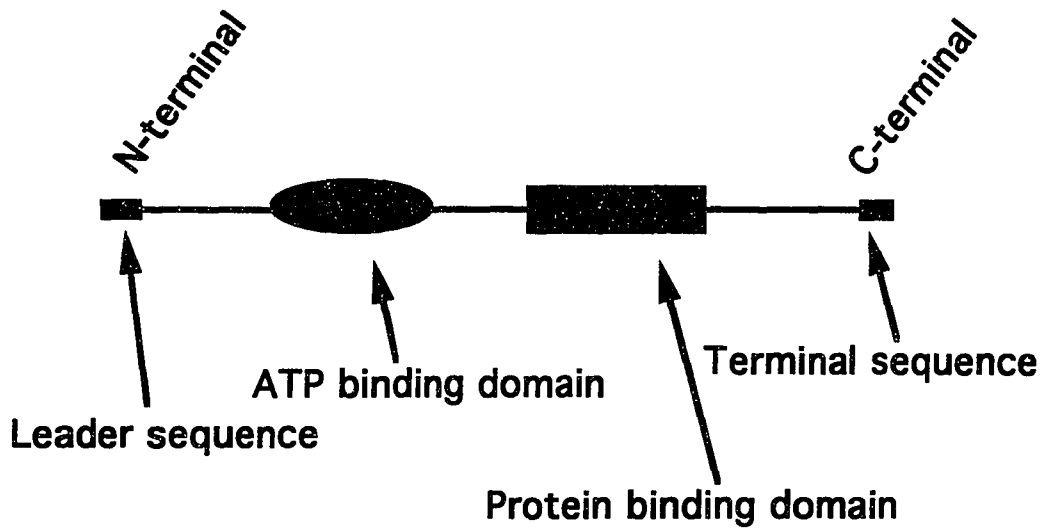
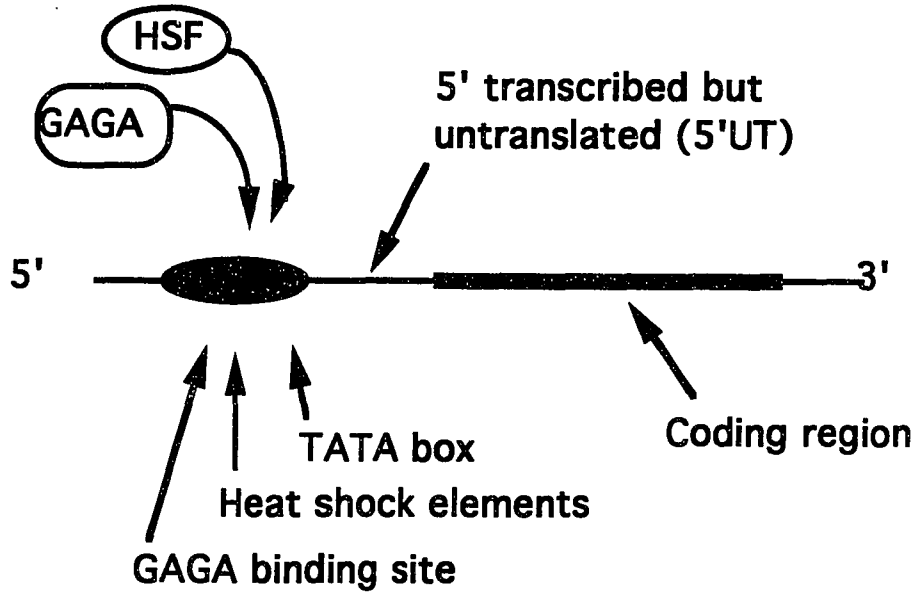


Figure 2. **Molecular features of heat shock 70 kDa genes**



Heat shock element consensus sequence is CTNGAANNTTCNAG.

GAGA binding sites are GA/CT rich sites.

5' UT region contains information necessary for translation at high temperature.

HSF (heat shock factor) is a transcriptional factor that binds to heat shock elements during heat shock.

GAGA is a constitutive transcriptional factor that alter chromatin structure of heat shock 70 promotors.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Medfly strains

The hilab strain originated from Kula, Maui (Hawaii). It has been maintained at the USDA/ARS laboratory, Honolulu, Hawaii for approximately 39 years. All genomic DNA and RNA extractions used in these experiments came from this strain.

The A4A, A5A3, and 18A2A strains are heavy metal tolerant strains. These strains were isolated from the Hilab strain by Dr. David S. Haymer.

2.1.2 Reagents and systems

2.1.2.1 Enzymes

Restriction endonucleases were obtained from either Boehringer Mannheim Inc. or Promega Inc. Deoxyribonuclease I and Ribonuclease were from Sigma Inc.. Proteinase K and T4 DNA ligase were from Boehringer Mannheim Inc. Lysozyme was from Sigma Inc.. These enzymes were stored and used as described according to manufacturer's recommendations or as in Sambrook, Fritsch and Maniatis (1989).

Other enzymes came with system kits which will be described later.

2.1.2.2 Chemical reagents

Chemical reagents were usually purchased from major chemical companies. These companies are Sigma Inc., Gibco BRL, Promega, Boehringer Mannheim, BioRad Inc., International Biotechnologies Inc. (IBI) , Research Organic Inc., Phamacia Biotech Inc., Aldrich, United states Biochemical (USB), Stratagene, and etc. In all cases, molecular biology grade reagents were used if available.

2.1.2.3 System products

The random primed DNA labeling kit was purchased from Boehringer Mannheim, Inc. This random primed DNA labeling kit consists of reagents used for ^{32}P-DNA labeling reactions. This type of labeling reaction was used for making DNA probes for RNA slot blots, northern blots and plus/minus cDNA library screening experiments.

The Genius™ nonradioactive DNA-labeling and detection kit was also purchased from Boehringer Mannheim Inc. This labeling and detection kit was used to label DNA probes for Southern blots, genomic DNA and cDNA library screening experiments.

The Geneclean™ kit was purchased from Bio101, Inc. It was used as a general procedure to purify DNA before enzymatic manipulations.

The GeneAmp PCR kit was purchased from Perkin Elmer Cetus or from Boehringer Mannheim, Inc. AmpliTaq^{R} DNA

Polymerase was provided with this kit and also with reagents necessary for general PCR reactions.

The PolyATtract^{R} mRNA Isolation System was obtained from Promega. This PolyATtract^{R} mRNA Isolation System uses streptavidin-paramagnetic particles to magnetically isolated poly(A) RNA out of total RNA samples. These isolated poly(A) RNAs were subsequently used for northern blots and as RNA templates in cDNA synthesis.

The RiboClone cDNA Synthesis System was also purchased from Promega. It is designed for the complete synthesis of double stranded cDNA using poly(A)+ RNA. An oligo(dT) primer and AMV reverse transcriptase were employed for first strand cDNA synthesis. Second strand synthesis uses RNase H to create nicks and gaps in the hybridized mRNA strand, which provide priming sites for DNA synthesis by DNA polymerase I. Double stranded cDNA was used as PCR templates and for cDNA library construction.

The Riboclone^{R} EcoRI Adaptor Ligation System was from Promega. This EcoRI adaptor is a molecular adaptor to ligate cDNA to vector arms in cDNA library construction.

The Protoclone^{R} λ gt 10 system was purchased from Promega. It is a vector system used for the construction of cDNA libraries. The cloning capacity of this vector is 0 to 7.6 kb. This vector contains a single EcoRI site within the phage repressor gene. The bacterial host strains C600 and C600*Hfl* are provided with the Protoclone λ gt 10 system.

The Packagene^{R} Lambda DNA Packaging System was from Promega. It is an extract containing the components and buffers needed to perform the packaging reaction. This packaging reaction is a process by which recombinant phage DNA molecules are packed into phage to produce a genomic library.

The pCR-Script^{TM} SK(+) cloning kit was purchased from Stratagene. This pCR-Script^{TM} SK(+) cloning kit is designed for cloning of PCR products at high efficiency. The pCR-Script cloning vector is based on the pBluescript II SK(+) vector.

Subcloning efficiency^{TM} E. coli DH5a competent cells were purchased from BRL. These competent cells were used in subcloning procedures.

The Erase-a-Base^{R} system was purchased from Promega. This Erase-a-Base^{R} system is used to produce unidirectional nested deletions in a DNA insert. Reagents used are Exonuclease III, S1 nuclease, and DNA polymerase I.

The Sequenase version 2.0 kit was purchased from United States Biochemical. This Sequenase version 2.0 kit contains all chemical reagents and standard protocols needed for dideoxy chain-termination DNA sequencing procedures.

2.1.3 Genomic DNA libraries

The λ gt 10 genomic libraries are EcoRI restriction fragment genomic library constructed from Medfly genomic DNA. This λ gt 10 genomic library was constructed by a complete digestion of genomic DNA with EcoRI, insertion in bacteriophage λ gt 10 arms

followed by a packaging reaction. This vector has a cloning capacity of 0-7.6 kb.

The EMBL 4 genomic library was constructed by a partial digestion of genomic DNA from the Medfly with Mbo I. Insert fragments and the vector arms were ligated together at the Bam HI site. The library has a cloning capacity of 7-20 kb and uses the *E. coli* strain K802 as a host. This library was used after being amplified.

2.1.4 cDNA libraries

Two cDNA libraries were used in the library screening experiments. One of these cDNA libraries was derived from the late pupal stage (20 days after egg laying). The other cDNA library was derived from the embryo stage. These two cDNA libraries use λ gt 10 as vector (Haymer *et al.* 1990).

2.1.5 DNA clones

The pmed21 clone is a 4.7 kb EcoRI fragment isolated from *C. capitata*. It contains the complete coding region of the CcA1 actin gene. It was subcloned in pUC9 (Haymer *et al.* 1990). This pmed 21 was used as a control in Southern blots, RNA slot blots, and poly(A) selected RNA northern blots.

pDm131 is an EcoRI genomic DNA fragment containing a metallothionein gene (Mtn 1) from *D. melanogaster* (Maroni *et al.* 1986). A Hpa II fragment within this pDm131 fragment was used

as a probe to identified homologous restriction fragments within the Medfly genome by a Southern blot experiments.

cDm51 is a cDNA clone that contains a transcribed sequence of the Mtn 1 gene. It was also used as a probe to identified homologous sequence in the Medfly genome (Maroni *et al.* 1986).

2.1.6 PCR Primers

Primer I	5'-GCA ATC ACA NCC GCA ATT ACA-3' G GGT G G	(21 mer)
Primer II	5'-TGC CCA TGC GGN ACN AGT-3' T TG G	(18 mer)
Primer III	5'-CCT GGT GCT TGC CGN TAC CCN-3' T GT G	(21 mer)
Primer IV	5'-CAG GAC NGC ACA NCG ACA AGC-3' G GT G	(21 mer)
Primer V	5'-TTA CCA ACG CCG TCA TCA CTG-3'	(21 mer)
Primer VI	5'-CCT CTT TCT CAG CCA ACT GGT TAG-3'	(24 mer)

Primers I and II were designed as described in figure 10. Primers III and IV were designed as described in figure 16. Primer V was designed to bind to a genomic DNA region corresponding to nucleotide positions 1589 to 1609 on the 3,286 base pair fragment of the Cerhsc1 gene. Primer V binding site is 314 nucleotides upstream of a junction site between the 6·5 and 6-9 subclones. Primer VI was designed to bind to a genomic DNA region corresponding to nucleotide position 2,944 to 2,921 on the 3,286 base pair fragment of the Cerhsc1 gene. Primer VI binding

site is 126 nucleotides downstream of the junction site between the 6-5 and 6-9 subclones.

2.2 Methods

2.2.1 Heavy metal treatments

Heavy metals ($CuSO_4$ or $CdCl_2$ or $ZnCl_2$) were applied to the Medfly larvae by mixing solutions of any of these heavy metals in larvae food as follows. In general a unit of the Medfly larvae food consists of 50 gm of bran, 25 gm of yeast (torular), 25 gm of sugar, 2 gm of Vanderzant vitamin fortification insects (USB), 2 ml of mold inhibitor solution, 125 ml concentrated HCl per 8 lites of water). To make heavy metal larvae food, the HCl water was mixed with an appropriate volume of a heavy metal stock solution to meet a designated final concentration.

In order to treat the Medfly adults a heavy metal solution was provided as drinking water for newly emerged adults for 3 days.

2.2.2 Extraction of high molecular weight genomic DNA

Genomic DNA from *C. capitata, Bactrocera dorsalis, Bactrocera cucurbitae*, and *D. melanogaster* was prepared by the Lifton procedure (Haymer and McInnis 1994). Approximately 0.3 gm flies (fresh or from -70°C) were homogenized in a 5 ml grinding buffer (0.2 M Sucrose, 50 mM EDTA, 100 mM Tris pH 9.0, 0.5% SDS). The homogenate was filtered through a sterile polyester

fiberfil using a 10 cc syringe. This homogenate was treated with 50 ul of proteinase K (final concentration of 0.2 mg/ml) at 65°C for one hour. Then, 750 ul of 8M potassium acetate was added before an incubation on ice for at least one hour. This mixture was centrifuged for 15 minutes at 10,000 rpm at 4°C (Sorvall, type SS-34 rotor) and the supernatant was transferred to a new test tube. The supernatant was mixed with 2 volumes of 95% ethanol and centrifuged again as before. A DNA pellet derived from this centrifugation was resuspended in 500 ul of TE, pH 8.0. The DNA solution was treated with 2.5 ul of 10 mg/ml RNase A at room temperature for 30 minutes. This solution was then extracted using phenol, phenol/CHISM, and CHISM. Finally, the extracted DNA was precipitated with 0.1 volume of 3M sodium acetate and 2 volumes of absolute ethanol at -20°C for at least 2 hours. This precipitated DNA was pelleted by microcentrifugation using at least 10,000 rpm for 15 minutes. The DNA pellet was washed with 70% ethanol, centrifuged and resuspend in 100 to 200 ul of TE pH 8.0. Quality and quantity of the extracted DNA was determined either by optical density readings or agarose gel electrophoresis.

2.2.3 Phage DNA preparation

For phage DNA isolation, phage clones were propagated and harvested from confluent plates as lysates (Maniatis *et al.* 1982). To extract phage DNA, lysates from two confluent plates of each clone were combined. The combined lysates were adjusted with SM to a final volume of 10 ml. Chloroform (166 ul) was added to

lyse any bacterial cells. The tube was spun at 5,500 rpm (Sorvall SS 34 rotor) for 10 minutes at 4°C. The supernatant was transferred to a new tube. DNase and RNase were added to 50 ng/ul final concentration and incubated at 37°C for one hour. After the incubation, 2 ml of 70°C prewarmed lysis buffer (0.5M Tris-HCl, 0.25M EDTA, 2.5% SDS) was added and incubated at 70°C for 35 minutes. Next, 2.5 ml of 8 M potassium acetated was added and the mixture was kept on ice for 30 minutes. The mixture was spun at 10,000 X g for 20 minutes. Isopropanol (0.6 volume) was added and the mixture was left for 10 minutes at room temperature. To pellet the DNA, the mixture was spun at 10,000 X g, 20°C for 10 minutes. The pellet was drained well and resuspended in 900 ul of 3M sodium acetate pH 6.0 and transferred to a microfuge tube. Isopropanol (0.6 volume) was added to precipitate the DNA at room temperature for 10 minutes. The DNA pellet was centrifuged at 12,000 X g for 10 minutes and resuspended in 500 ul of 0.3M sodium acetate. This mixture was extracted with one volume of phenol. The mixture was also extracted with CHISM twice. The DNA was precipitated by adding 0.6 volume of isopropanol at room temperature, incubation for 10 minutes and spinning at 12,000 X g for 10 minutes at room temperature. The DNA pellet was washed by adding 70% ethanol, centrifuged, air-dried and resuspended in 50 ul of TE pH 8.0. This procedure yields approximately 50 ug of purified phage DNA.

2.2.4 Plasmid DNA preparation

Plasmid DNA derived from the alkaline plasmid DNA preparation is usually used for further enzymatic manipulations such as restriction digestion and exonuclease deletion for DNA sequencing. These bacterial colonies were grown in 5 ml LB broth with an appropriate selective antibiotic such as amplicillin (100 ug/ml) overnight. 1 ml of this culture was centrifuged at 10,000 X g for 15 seconds. The pellet was frozen in a dry-ice-ethanol bath for approximately 30 minutes. The frozen pellet was then resuspended with 10 ul of solution I (50mM glucose, 10mM EDTA, 25mM tris pH 8.0, lysozyme to a final concentration of 10 mg/ml). After the bacterial cells were lysed, 20 ul of solution II (0.2N sodium hydroxide, 1% SDS) was added and incubated on ice for 10 minutes. Later, solution III (3M potassium acetate pH 4.8) was added and mixed.

This mixture was centrifuged at 12,000 X g for 15 minutes. The supernatant containing the DNA was transferred to a new microfuge tube and mixed with 24 ul isopropanol. The tube was kept at -20°C for at least 2 hours to precipitate the plasmid DNA. After spining 15 minutes at 4°C (10,000 X g), the DNA pellet was air dried, resuspended in 20 ul of TE pH 8.0 and treated with 2 ul of RNase A (10 mg/ml) at 37°C for 30 minutes before use.

2.2.5 Rapid boiling plasmid DNA preparation

The rapid boiling method is an alternative plasmid DNA preparation procedure because it is relatively easier to employ

than the alkaline plasmid preparation. However the quality of the plasmid DNA is only suitable for an immediate use and when only small amounts of plasmid DNA are required. This method has been used mostly in subclone selection and labeling reactions for DNA sequencing.

Bacterial were grown in 2 ml of media overnight. This culture was microcentrifuged for 30 seconds. The pellet was resuspended in 300 ul STET buffer (8% sucrose, 50mM Tris pH8.0, 50mM EDTA, 5% Triton X-100, filter steriled), mixed with 15 ul of lysozyme (20 mg/ml in 0.001M Tris, pH 8.0), incubated on ice for 10 minutes. The mixture was boiled for 45 seconds and then centrifuged at room temperature for 15 minutes to pellet the cellular debris and chromosomal DNA. 210 ul of supernatant was transferred to a new tube, mixed with 230 ul of isopropanol, incubated at room temperature for 10 minutes, microcentrifuged at 4°C for 15 minutes, washed with 70% ethanol twice, and spun again. The plasmid DNA pellet was dried under vacuum and resuspended in 50 ul of ddH_2O.

2.2.6 Extraction of total RNA from the whole fly

The total RNA extraction was carried out with special precautions and sterile procedures due to fragile nature of the RNA molecule and the presence of RNAses (Sambrook, Fritsch, and Maniatis 1989).

One gram of Medfly material was mixed with 0.5 ml extraction buffer (3M LiCl, 6M urea, 10mM Sodium acetate pH 5.2,

0.2 mg/ml Heparin, 0.1% SDS) in a tube, homogenized on dry ice, and kept at -20°C overnight. The homogenate was microcentrifuged at 10,000 X g for 15 minutes at 4°C. The homogenate was kept on ice. The homogenate was mixed with another volume of extraction buffer and resuspended with 0.5 ml precooled wash buffer (4M LiCl, 8M urea). The mixture was microcentrifuged at 10,000 X g for 15 minutes at 4°C. The pellet was kept on ice, mixed with 300 ul dissolving buffer (0.1M sodium acetate pH 5.2, 0.1% SDS) and gently agitated for 5 minutes at room temperature. The mixture was then phenol extracted using 1 volume of TE pH 8.0 saturated phenol and 1 volume of CHISM with gentle agitation for 20 minutes at room temperature. The mixture was microcentrifuged at 10,000 X g for 5 minutes to ensure phase separation. The bottom organic phase was discarded. The mixture was reextracted with an equal volume of CHISM and microcentrifuged at 10,000 X g for 5 minutes. The top aqueous phase was transferred to a new tube, precipitated with 1/20 volume of 3M sodium acetate pH 5.2 and 2.5 volumes of absolute ethanol, and kept at -20°C overnight. The precipitated RNA was microcentrifuged at 10,000 X g for 15 minutes at 4°C. The pellet was washed with 70% ethanol twice, dried under vacuum, and resuspended in 50 to 100 ul of ddH_2O. The quality and quantity of the total RNA was monitored by standard optical density reading (Sambrook, Fritsch, and Maniatis 1989). The RNA solution was stored at -70°C until use.

2.2.7 Poly(A) RNA selection

The poly(A) tract mRNA isolation system uses a biotinylated oligo(dT) primer to hybridize in solution to the 3' poly(A) region present in most mature eukaryotic mRNA species. The hybrids are isolated using streptavidin associated paramagnetic particles and a magnetic separation stand. The mRNA is eluted from the paramagnetic complex by an addition of ribonuclease-free deionized water. The mRNA isolated was used for northern blot analyses and as a template for cDNA synthesis.

2.2.8 cDNA synthesis and cDNA library construction

Approximately 2 ug of total mRNA templates was hybridized with a primer-adaptor supplied by the the Riboclone^R cDNA synthesis system. This primer-adaptor consists of oligo$(dT)_{15}$ adjacent to a unique restriction site (Xba I). AMV reverse transcriptase is used as described in a technical manual of the Riboclone^R cDNA synthesis system to synthesize the first strand of the cDNA. The first strand cDNA was used as a template for a reaction with DNA polymerase I and RNase H to synthesize the second strand of the cDNA. Ends of the newly synthesized cDNA molecules were made flush using T4 DNA polymerase. The cDNA was first ligated to EcoRI adaptors at both ends and then ligates to the λ gt 10 arms having EcoRI sites in order to construct a cDNA library.

To achieve the optimal ligation conditions, the molar ratio of λ gt 10 arms and cDNA insert was varied as recommended in the

technical bulletin. Each of recombinant λ gt 10 DNA constructs was packaged using the Packagene^{R} Lambda DNA Packaging System and plated out on a bacterial lawn. Approximately 3 to 15 thousand plaques were formed from each of the cDNA library constructions. The 3 to 15 thousand plaques reflected that the cDNA library construction was successful. Nonetheless, the titer number is at the low end for a typical range of titers. These cDNA libraries were used for a plus/minus library screening.

2.2.9 Southern blots

Approximately 3 to 5 ug of genomic DNA or 0.5 to 1.0 ug of plasmid DNA or 1 ug of phage DNA was loaded on to a 0.7% agarose gel containing ethidium bromide. The gel was run submarine at approximately 1.5 v/cm for 12 to 16 hours in TBE buffer (0.089M Tris-borate, 0.089M boric acid, 0.002M EDTA). After the electrophoresis had been completed, the DNA within the gel was visualized and photographed on a UV light transilluminator to record DNA mobility.

To localize a particular DNA fragment within the gel, the DNA was transferred onto a nylon membrane (Micron Separation Inc.) by the Southern blot transfer technique (Sambrook, Fritsch, and Maniatis 1989). For this procedure, the gel was moved to a pyrex dish and submerged in a denaturation solution (0.5M sodium hydroxide, 1.5M sodium chloride) with gentle shaking for one hour. The denaturation solution was drained off and replaced

with neutralization solution (1M Tris pH 8.0, 1.5M sodium chloride) for another one hour on the shaker.

To prepare the blot transfer, a 25 cm X 30 cm X 5 cm sponge was saturated with 4X SSC. Two 25 cm X 30 cm filter papers were wetted with 2X SSC and placed on the sponge. The neutralized gel was placed on the filter papers.

A gel-sized nylon membrane was wetted completely with 2X SSC and placed on the top of the gel. All air bubbles that were trapped between the gel and the filter were carefully removed. The area of the filter paper not covered by the gel was covered with parafilm strips. Two more filter papers were wetted with 2X SSC and placed on the membrane.

A stack of paper towels (at least 10 cm high) was placed on the top of the blot. A 15 cm X 15 cm plexiglass plates was placed on the stack before a 500 gm weight was applied at the top. This set up was left for 12 to 24 hours to allow DNA to be transferred. After the transfer, the nylon membrane was removed from the set up and rinsed with 2X SSC. The rinsed membrane was air dried between blotting papers for 30 minutes. The nylon membrane was then baked at 80°C under vacuum for 45 minutes.

All Southern blots were probed with nonradioactively labeled DNA probes. The blotted membrane was inserted into a sealable plastic bag and 10 to 20 ml of prehybridization solution (2X SSC, 5% blocking reagent, 0.5% N-lauroylsarcosine, 0.1% sodium chloride, 0.1% SDS, 50% deionized formamide (high stringency) or 43% deionized formamide (low stringency) or 30% deionized

formamide (very low or lowest stringency)). This was incubated at 42°C for 2 to 24 hours. This solution was replaced by a solution containing a probe and fresh hybridization solution (as above). The solution containing the DNA probe was denatured at 95°C for 10 minutes before use. The hybridization was carried out at 42 °C for 18 to 24 hours.

The hybridized filter was then separated from the hybridization solution and washed. The first and second washes were either at 65°C (high stringency) or 55°C (low stringency) or 37°C (very low stringency or lowest stringency) using a solution of 0.1% SDS and 2X SSC for 15 minutes. The third and fouth washes used a solution of 0.1% SDS and 0.5X SSC for high stringency or 1X SSC for low stringency or 2X SSC for very low or lowest stringency for 15 minutes each.

Probe detection by a color reaction consisted of washing the filter with buffer 1 (0.1M Tris-HCl pH 7.5, 0.15M sodium chloride) for 1 minute, incubating in buffer 2 (0.5% blocking reagent in buffer 1) for 30 minutes and washing with buffer 1 for 1 minute. The filter was the inserted in to a bag, and a dilute<Dig>Ap-conjugate (150 unit/ml) was added. After gently agitating for 30 minutes at room temperature, the filter was washed again twice with buffer 1 for 15 minutes, and then with buffer 3 (0.1M Tris-HCl pH 9.5, 0.1M sodium chloride, 0.05M magnesium chloride) for two minutes. The filter was exposed to a color solution (45 ul of NBT and 35 ul of X-phosphate in 10 ml of buffer 3) and allowed to develop the dark. Up to 48 hours could be required for the color

reaction to be completed. The color image of the Southern blot was fixed using with solution 4 (10mM Tris-HCl pH 8.0, 1mM EDTA) for 5 minutes when positive signals were detected. The color image filter was kept dry and photographed.

2.2.10 Northern blots

The gel electrophoresis apparatus used for northern blots was treated with 0.1M sodium hydroxide for 30 minutes to eliminate RNase contamination. Likewise, all chemicals and glassware used in for northern blot were handle using special precautions described earlier.

The gel was prepared with 1.4% agarose in RNA gel buffer pH 7.0 (0.25M MOPS, 0.005M sodium acetate, 0.001M EDTA) with 2.2M formaldehyde. Approximately 1 ug of poly(A) selected RNA was dissolved in 20 ul of loading buffer (1X electrophoresis buffer, 50% formamide, 1.8M formaldehyde), denatured at 70°C for 10 minutes, chilled on ice for 3 minutes, mixed with 2 ul of bromophenol blue solution, and loaded on to the gel for electrophoresis. The RNA gel was electrophoresed at 20 to 25 volts for 18 to 20 hours using a buffer circulation system.

The RNA gel was soaked in 20X SSC for 40 minutes and then sandwich blotted onto a wetted nylon membrane in 10X SSC at 4 °C for overnight. The sandwich blot set up was as described in Sambrook, Fritsch, and Maniatis (1989).

The nylon membrane was removed from the sandwich blot dried and baked at 80°C for 45 minutes. This membrane was then

prehybridized in 15 ml of prehybridization solution (50% formamide for high stringency or 43% formamide for low stringency, 5X SSC, 10mM sodium phosphate pH 6.5, 250 ug/ml sheared salmon sperm DNA, 1X Denhardt's solution.) at 42°C for 12 to 24 hours.

The prehybridized nylon membrane was hybridized with a heat denatured ^{32}P-labeled probe in hybridization solution at 42°C for 12 to 24 hours. The hybridization solution was identical to the prehybridization solution except that it contained approximately 2 X 10^7 count per minute (cpm) of labeled DNA probe.

The hybridized membrane was washed as follows. The first wash was in 20 ml of wash solution (50% formamide (high stringency) or 43% formamide (low stringency), 2X SSC, 0.1% SDS) for 15 minutes at room temperature. The second and third washes were in 500 ml of another wash solution (2X SSC, 0.1% SDS) for 15 minutes at 50°C (high stringency) or 40°C (low stringency). The fourth and fifth washes were in 500 ml of wash solution (0.2X SSC, 0.1% SDS) for 15 minutes at 50°C (high stringency) or 40°C (low stringency). The washed nylon membrane was air dried and autoradiographed using Fuji Rx X ray film.

2.2.11 RNA slot blots

A slot blot manifold apparatus (Schleicher & Schuell) was submerged in 0.1N sodium hydroxide for 30 minutes and rinsed with ddH_2O. Each total RNA sample (5 ug for adult total RNA and

10 ug for larvae total RNA) in 10 ul ddH_2O was mixed with 20 ul 100% formamide, 7 ul of 37% formaldehyde, 2 ul of 20X SSC, denatured at 68°C for 15 minutes, chilled on ice for 5 minutes, and mixed with another 2 volumes of 20X SSC. The manifold was assembled as described by the manufacture with a 20X SSC soaked nylon membrane. Each slot was washed with 500 ul of 10X SSC using gentle air suction before loding RNA samples. Each loaded slot was rinsed twice with 10X SSC using continuous air suction. The nylon membrane was then air dried for 30 minutes before baking under vacuum at 80°C for 45 minutes. The prehybridization and hybridization procedures were identical to the previously described protocol for nothern blots (high stringency).

2.2.12 DNA labeling

2.2.12.1 Radioactive DNA labeling

Linear DNA fragments were radioactively labeled by a "random primed DNA labeling" approach. The labeling reactions were conducted in a 20 ul volume consisting of 1 ul DNA solution (100 ng/ul TE pH 8.0), 3 ul dATP dGTP dTTP mixture (1:1:1, 5uM each), 5 ul (50 uCi) dCTP [α-^{32}P] 3000 Ci/mM, 2 ul hexanucleotide mixture, 1 ul Klenow enzyme, and 8 ul ddH_2O.

A DNA solution was denatured by heating for 10 minutes at 95°C after which it was quickly cooled on ice. The other components were added and the reaction was incubated at 37°C

for 30 minutes. The labeling reaction was stopped by adding 2 ul 0.2M EDTA.

Separation of unincorporated nucleotides was accomplished by bringing the 20 ul volume to 50 ul with water, adding 1 ul 0.1M spermine solution, vortexing and incubating on ice for 15 minutes. After 15 minutes of centrifugation at 12,000 X g and 4°C, the DNA pellet was resuspended in 50 to 100 ul TE pH 8.0. The percent incorporation of the reaction was measured by scintillation count of the radioactivity found in both the incorporated and unincorporated fractions. Percent incorporations were usually between 20 to 70% corresponding to specific activities of 10^8 cpm/ug.

2.2.12.2 Nonradioactive DNA labeling

Linear DNA templates were labeled with Dig-dUTP provided by the Genius™ nonradioactive DNA-labeling and detection kit. Approximately 1 ug DNA template was used in a volume of 15 ul in ddH_2O, denatured at 95°C for 10 minutes, and chilled on ice. The denatured DNA solution was mixed with 2 ul hexanucleotide mixture, 2 ul Dig-DNA mix (dATP, dCTP, dGTP, dTTP, Dig-dUTP) and 1 ul Klenow enzyme, and incubated at 37°C for 1 hour. The labeling reaction was stopped by adding 2 ul 0.2M EDTA. This labeled DNA was precipitated by adding 2ul LiCl (4M), 60 ul prechilled (-20°C) ethanol, and incubating at -20°C for 2 hours. This precipitated DNA was microcentrifuged for 15 minutes at 4°C, washed with 70% ethanol, microcentrifuged again for 15 minutes

at 4°C, air dired in vacuum, and resuspended in 50 ul TE pH 8.0. This labeled DNA was stored at -20°C.

2.2.13 Genomic DNA library and cDNA library screening

For genomic DNA or cDNA library screening experiments, approximately 20,000 plaques from the corresponding libraries were plated out onto two 150 mm diameter NZCYM agar plates as described in Sambrook, Fritsch, and Maniatis (1989).

The plaques were transferred by laying a nylon membrane on top of the plate for 1 minutes to adsorb phage particles. A duplicate nylon membrane was laid in a similar manner except that the time was prolonged for the second transfer up to 5 minutes.

The membranes were air dried for 5 minutes, floated on a denaturation solution (0.5M sodium hydroxide, 1.5M sodium chloride) for 30 seconds, immersed into the same solution for 30 seconds, transferred to a neutralization solution (1M Tris-HCl pH 8.0, 1.5M sodium chloride) for 5 minutes, rinsed in 2X SSC, air dried for 30 minutes, and baked in a 80°C vacuum chamber for 45 minutes.

The nylon membranes were then treated in the same manner described for the Southern blot hybridization and color reaction.

2.2.14 Plus/minus cDNA library screening

Duplicate membranes were prepared as described from a heavy metal induced cDNA library. One of these nylon membranes was screened with a ^{32}P-labeled heavy metal induced cDNA probe (high stringency). The other membrane was screened with a ^{32}P-labeled non induced cDNA probe (high stringency). cDNA clones that exhibited stronger hybridization signal to the ^{32}P-labeled heavy metal induced cDNA probe than to the ^{32}P-labeled non induced cDNA probe were isolated. These isolated cDNA clones were replated with low plaque density and rescreened to confirm the primary results.

The heavy metal induced cDNA was prepared from using newly emerged Medfly adults that had been given a 20mM solution of $CuSO_4$ in their drinking water solution for 3 days before being scarified. The noninduced flies were treated with water only.

The cDNA library lift procedure was similar to the previously described genomic library screening except that a radioactive probe was used.

2.2.15 Polymerase chain reaction (PCR)

All PCR reactions were conducted in a total volume of 100 ul. The GeneAmp PCR reagent kit supplied all reagents used except primers and DNA templates. These PCR reactions consist of 200uM of dNTPs, 10 ul of 10X PCR buffer, magnesium chloride, 2.5 units of AmpliTaq^{R} DNA polymerase, 10 ng of DNA templates or cDNA

templates, and 200pM of degenerate primers or 50pM of specific primers. The PCR reaction mixture were covered with 50 ul of mineral oil in a 500 ul microtube before being placed in a PCR temperature cycling device (Coy TempCycle II).

Primers I, II, III, and IV are degenerate primers. These primers were dissolved in filtered ddH_2O to a final concentration of 200 pM/ul. Primers V and VI are specific primers. These primer V and VI were dissolved in filtered ddH_2O to a final concentration of 50 pM/ul. The oligo(dT) (15-mer) primer was also prepared to be 200 pM/ul. All genomic DNA templates and cDNA templates were also diluted to a final concentration of 10 ng/ul with filtered ddH_2O.

Primers I and II were designed from a coding region of a metallothionein gene from *D. melanogaster*. Details of the primers are in figure 10. Primer annealing temperature was varied among different experiments in order to search for an optimal conditions for specific binding to the genomic DNA templates. The profiles for the PCR reactions was a cycle of 94°C for 2 minutes (1 cycle), followed by 40 cycles of 94°C for 1 minute, 48°C or 55°C for 1 minute, and 72°C for 2 minutes. Genomic DNA templates from *D. melanogaster* was used as a positive control.

Primer IV was used with heavy metal induced cDNA templates to amplified coding region of genes associated with heavy metal induction. The PCR profile was a cycle of 94°C (1 cycle) followed by 40 cycles of 94°C for 1 minute, 58°C for 1 minute, and 72°C for 2 minutes.

A combination of primers II and IV were used with heavy metal induced cDNA templates in a PCR reaction. Three specific PCR products were amplifled from this PCR reaction. The PCR profile used in this case was a cycle of 94°C for 2 minutes (1 cycle) followed by 40 cycles of 94°C for 1 minute, 53°C for 1 minute, and 72°C for 2 minutes.

Many other combinations of primers I, II, III, IV and oligo(dT) were also used with heavy metal induced cDNA templates in several PCR reactions. Temparature profiles used in these cases were similar to that was described above, however, the annealing temperature were varied as 53°C or 55°C or 58°C or 60°C or 65°C, in order to search for an optimal annealing temperature. No specific PCR products was recovered from these PCR reactions.

In order to confirm organization of the 6-5, 6-9, and 6-1 EcoRI fragments in the Medfly genome (figure 17), primers V and VI were designed using a computer program to searched the 3,286 base pair sequence of the Cerhsc1 gene. These primers were designed to amplify a 1,274 base pair fragment corresponding to a portion of the Cerhsc1 gene from total genomic DNA templates. The cycling profile for these primers was one cycle of 94°C for 2 minutes followed by 30 cycles of 94°C for 1 minutes, 55°C for 1 minutes, 72°C for 2 minutes. A 1274 base pair fragment was amplified, cloned and sequenced from both ends to determine nucleotide sequences around the junction regions of the 6-5, 6-9, and 6-1 subclones.

2.2.16 Nested deletion

The Erase-a-Base^R system provides a way to achieve a nested set of deletions in DNA. Details procedures to generated this nested set of deleted DNA is provided by the technical manual of the Erase-a-Base^R system.

The Erase-a-Base^R system is designed for the rapid construction of vector DNA containing progressive unidirectional deletions of any inserted DNA. The system employs exonuclease III to specifically digest inserted DNA from a 5' protruding or blunt end restriction site. The other end of the same inserted DNA, at which the primer binding site is located, is protected from the exonuclease deletion by a 3' overhang restriction site or by a α-phosphorothioate filled end.

Plasmids containing inserted DNA fragments of interest were prepared by alkaline plasmid DNA preparation procedure. Approximately 3 to 5 ug of the plasmid DNA was restriction digested to generated desirable 5' protruding or blunt ends ready for the exonuclease deletion and a protected end.

The digested plasmid DNA was precipitated, resuspended in 60 ul of exonuclease III 1X buffer, and warmed at 37^{o}C in a waterbath. This warmed DNA solution was mixed rapidly with 300 to 500 units of exonuclease III, and incubated at 37^{o}C. Since exonuclease III digestion proceeds about 450 bases/minute at 37^{o}C, the reaction mix was proportionally aliquoted at intervals of 20 to 30 seconds to generate a set of serially deleted plasmid

DNAs. Each aliquoted sample was mixed immediately with S1 nuclease in a microtube on ice to stop the exonuclease digestion. The samples were then heated at 70°C for 10 minutes to inactivate the S1 nuclease. 2ul aliquot from each sample were loaded to a minigel in order to examine the extent to which the plasmid DNAs were deleted.

The serially deleted DNA samples were filled in with dNTPs by Klenow enzyme. These DNA samples were ligated, and transformed into *E. coli* DH5α competent cells. Transformants were selected for plasmids containing nested sets of deletions in the target DNA.

2.2.17 DNA sequencing

Plasmids containing inserted DNA of interest was prepared by the rapid boiling method. 3 to 5 ug aliquots of the plasmid DNA were resuspended in 45 ul TE 8.0. This plasmid DNA solution was mixed with 180 ul of a denaturation solution (0.2M sodium hydroxide, 0.2mM EDTA pH 8.0), and incubated at 37°C for 30 minutes. The denatured DNA was neutralized and precipitated with a 0.1 volume of 2M ammonium acetate and 2 volumes of ethanol at -20°C for one hour. The precipitated DNA was microcentrifuged at 4°C for 15 minutes. The DNA pellet was washed with 70% ethanol, dried under vacuum, and resuspended in 7 ul ddH_2O.

Labeling of the DNA templates was carried out using reagents and a standard protocol provided by the Sequenase^R 2.O

kit. Since all DNA templates were in the pUC19 or the pCR-Script vectors, the primers used in the sequencing reaction were the forward or reverse primers which bind to immediate regions flanking a polylinker region in the plasmid (Boehringer Mannheim). DNA labeling procedures began with annealing the primer to the DNA templates in a mixture of 7 ul DNA in ddH_2O, 2 ul reaction buffer, 1 ul primer (0.5 pM/ul). The annealing mixture was incubated at 65°C for 2 minutes, cooled down slowly to 30°C in about 30 minutes, and set on ice.

Four small microtubes were marked A, C, G, T and added 2.5 ul of each dideoxynucleotide termination buffer (ddATP, ddCTP, ddGTP, ddTTP), respectively, to each marker tube. These tubes were warmed at 37°C for at least 1 minute. The primer annealed DNA template sample was labeled by mixing with a 5.5 ul reaction mix stock (1.6 ul ddH_2O, 1.0 ul DTT, 0.4 ul 5X labeling mixture, 0.5 ul ^{35}S-dATP, and 2 ul diluted sequenase 2.0) and incubated at room temperature for 2 minutes. This labeling reaction was specifically terminated by adding a 3.5 ul volume from this labeled DNA mixture to each of the previously aliquoted dideoxynucleotide termination buffers. These dideoxynucleotide termination reactions were incubated at 37°C for 5 minutes and stopped by adding 4 ul of stop solution. These labeled DNA samples were stored at -20°C before polyacrylamide gel electrophoresis.

The buffer gradient polyacrylamide sequencing gel (6%) was a gel of choice for this DNA sequencing work. Buffer gradient polyacrylamide gels were prepared according to Sambrook, Fritsch,

and Maniatis (1989). Gels were run using the BRL model S2. Gels were prerun for 30 minutes at 60 watts before loading the sample. After the prerun, each well of the gel was washed with 0.5X TBE buffer solution. The previously labeled DNA samples were heated at 90°C for 2 minutes and chilled on ice. 2 to 3 ul of each sample were loaded and the gel was run at 60 watt for 2 to 3 hours. Gels were removed from the sequencing apparatus, fixed with a solution (15% methanol, 5% acetic acid) for 15 minutes, dried on a gel dryer, then autoradiographed with an X-ray film (Fuji-RX). The X-ray image of this sequencing gel was read manually and entered in to a computer using the DNA Inspector program (Textco, Inc) or other computer programs for DNA sequence analysis.

2.2.18 Database search comparisons

Database search comparisons were carried out using a computer to search through the nucleic or protein sequence databases for similarity to a query sequence. The program used in all database search comparisons is from the Genetics Computer Group, Inc. (GCG). The nucleic acid sequence database was GenEMBL which consists of GenBank and those sequences in EMBL that are not represented in GenBank. This GenEMBL database contains at least 106,953 sequences representing many taxa.

The database search comparison at the nucleotide level was done by using a subprogram "FASTA" within the GCG program. FASTA uses the method of Pearson and Lipman (1988) to search for similarities between one sequence and any group of sequences.

FASTA begins to search for similar sequences by registering sequences that have the largest number of short perfect match (words). The word size for the nucleic acid sequences database search comparison is 6 nucleotides for this work. Output files of the FASTA database search programs are list of sequences which are ranked based on degree of similarity to the query sequence. These sequences are also individually aligned to the query sequence together with a %identity value.

At the amino acid level, the subprogram "TFASTA" was used for the database search comparisons. TFASTA also does a Pearson and Lipmann search for similarity between a query peptide sequence and any group of nucleotide sequences. However, the TFASTA translates the nucleotide sequences of database in all six reading frames before performing the comparison. The word size for TFASTA is 2 amino acid residues for this work. Output files of the TFASTA database search programs are a list of sequences which are ranked based on degree of similarity to the query sequence. These sequences are also individually aligned to the query sequence together with %identity value.

2.2.19 Multiple sequence alignments

Multiple sequence alignments of nucleic acid and amino acid sequences were carried out using a subprogram "PILEUP" within the GCG program. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. This progressive, pairwise alignment is simplified

from a method of Feng and Doolitle (1987). The multiple alignment procedure begins the pairwise alignment of the two most similar sequences and generates a cluster of two aligned sequences. This cluster sequence is used to aligned to the next most related sequence or another cluster of sequences. In order to calculate similarity between a pair of sequences, PILEUP has gap weight equal to 3.00 and gap length weight equal to 0.10 for this work. The output file of PILEUP is a multiple sequence alignment of a group of sequences in an input file (figures 24 and 25). PILEUP also plots a dendrogram showing the clustering relationships used to determine the order of the pairwise alignment, allowing generation of overall multiple sequence alignment (figures 22 and 23).

The multiple sequence alignment data was used to generate a graphic presentation of average similarity across the entire alignment (figures 26 and 27). This was done by a subprogram "PLOTSIMILARITY" within the GCG program. PLOTSIMILARITY calculates and plots the average similarity among all members of a group of aligned sequences at each position in the alignment using a window size of 10.

Chapter 3

Genetic basis of heavy metal tolerance in the Medfly

3.1 Base line levels of heavy metal tolerance

Studies of heavy metal tolerance in the Medfly are in part related to an effort intended to document the feasibility of using heavy metal tolerance as a selectable genetic marker. This genetic marker might serve in germ line transformation systems as well as in genetic sexing strains. The possibility that heavy metal tolerance may provide a means to select individuals (transformants or chromosomal rearrangements) relies on experimental results from work with *D. melanogaster*. These experiments showed that a complete metallothionein gene duplication resulted in a higher levels of heavy metal tolerance (Maroni *et al.* 1987, Lange *et al.* 1990). Theoretically for any species, the numbers of copies of metallothionein gene could be directly related to the level of heavy metal tolerance in an individual.

In addition, differential heavy metal tolerance among natural populations of *D. melanogaster* has been observed (Lange *et al.* 1990). This suggests that there might be genetic variation underlying the heavy metal tolerance trait in the Medfly as well. One way to test this is to determine the base line levels of heavy metal tolerance in different strains over several generations.

Haymer (1990) had embarked on studies of base line levels of heavy metal tolerance in the Medfly. These studies were

carried out using various concentration of zinc chloride and cadmium chloride solution in the Medfly larval diet. Analyses of the number of pupae counted on the fourteenth day after an egging was a method of choice to measure heavy metal tolerance in the Medfly. Base line levels of heavy metal tolerance were established using a range of heavy metal doses compared to a number of pupae recovered from a normal diet.

Three different Medfly strains selected that demonstrated higher levels of heavy metal tolerance compared to the wild type Hilab strain from which they were isolated. These strains were designated A4A, A5A3, and 18A2A.

A base line level of zinc tolerance is shown in figure 3. The number of pupae recovered at a 0mM concentration was used as a 100% reference for each of the Medfly strains. At a 25mM zinc chloride concentration, no obvious differences are observed in the percent of pupal recovery among the Hilab strain and the selected strains. The differences of the levels of tolerance become more obvious in a 50mM zinc diet. It was demonstrated that the A4A, A5A3, and 18A2A strains showed approximately 75% pupal recovery while the Hilab strain yielded only 25% pupae. A 75mM zinc diet seemed to suppress nearly all pupal recovery.

Copper sulphate doses for the base line levels of tolerance studies were also studied (figure 4). A 2.5mM copper diet did not rusult in an obvious differences in the pupal recovery among the Medfly strains, although the overall percent of pupal recovery appeared to be lower that seen for the 0mM diet. Data from a

7.5mM copper diet suggested that the A4A strain was the only strain exhibiting any copper tolerance. This strain yielded approximately 65% pupal recovery while the other Hilab, A5A3, and 18A2A strains yielded less than 20% pupal recovery. The percent of pupal recovery was very low approaching 0% in a 10mM copper diet, suggesting this was probably a lethal dose for all of the Medfly strains.

Cadmium chloride is toxic at lower levels when compared to both the copper and zinc toxic level. Figure 5 shows that pupal recovery in all of the Medfly strains were totally suppressed in a cadmium chloride diet with a concentration as low as 0.75mM. In a 0.25mM and a 0.50mM cadmium diets, there were no clear differences in the levels of cadmium tolerance among the Hilab, A4A, A5A3, and 18A2A strains (figure 5).

Heavy metal tolerance levels appear to vary by the type of heavy metal. The A4A, A5A3, and 18A2A strains exhibited some zinc tolerance (figure 3). However, these same strains did not appear to be either cadmium or copper tolerant, with the possible exception that the A4A strain may exhibit some copper tolerance (figure 4).

The zinc tolerance trait exhibited by the strains A4A, A5A3, and 18A2A appeared to be heritable because the trait has been transmitted for generations. Hence it may be possible to identify an underlying genetic basis for this phenomenon. Reciprocal crosses were carried out using the Hilab and each of the A4A,

A5A3, and 18A2A strains to determine a mode of inheritance of the zinc tolerance trait.

3.2 Reciprocal crosses for testing modes of inheritance of the heavy metal tolerance trait

Comparative toxicity tests involving crosses between the A4A and Hilab strains were done in order to assess any genetic basis for this resistance phenotype. As shown for a 50mM zinc diet (figure 6), the data are consistent with a dominant mode of inheritance for the zinc tolerance trait. This is because the percent of pupal recovery for the crosses were as high as the zinc tolerance exhibited by A4A strain alone. There are differences in the percent of the pupal recovery between the progeny of the cross ♂♂ A4A X ♀♀ Hilab and the progeny of the reciprocal cross ♂♂ Hilab X ♀♀ A4A (approximately 50% difference in pupal recovery) although both demonstrated zinc tolerance increased over the level exhibited by the Hilab strain. Such reciprocal crosses differences suggest that sex chromosomes of the Medfly might carry genes associated with zinc tolerance. However, more detailed studies are needed to test this hypothesis.

In reciprocal crosses between the ♂♂ A5A3 and ♀♀ Hilab strains, the pupal recovery in a 50mM zinc diet also suggested a dominance effect for zinc tolerance trait (figure 7). However, this may not be the same zinc tolerance trait exhibited by the A4A strain. This is because the crosses involving the ♂♂ A5A3 and ♀♀ Hilab strains produced different results compared to what had

been seen with the A4A strain. Specifically, these results differed according to the direction of the crosses (in terms of male versus female). In contrast to the A4A test crosses, the progeny that show higher levels of zinc tolerance (approximately 35% higher in pupal recovery) is from crosses ♂♂ A5A3 X ♀♀ Hilab. This suggests that there are different genes associated with zinc tolerance located on different chromosomes compared to what appeared in strain A4A. However, further genetic studies will be needed to verify this suggestion.

Zinc tolerance in the 18A2A strain also gives some indications of dominance. Data from a 50mM zinc diet (figure 8) show that progeny of the ♂♂ 18A2A X ♀♀ Hilab and ♂♂ Hilab X ♀♀ 18A2A reciprocal crosses tolerated levels of zinc in the diet similar to those of the 18A2A parent. However, again the reciprocal crosses involving this strain showed a different pattern of zinc tolerance. This suggests the possibility of a variable genetic basis underlying the zinc tolerance trait in the Medfly strain. This may be further elucidated by conducting reciprocal crosses among the A4A, A5A3, and 18A2A zinc strains. This could demonstrate the extent to which a single gene or multiple genes are involved in the zinc tolerance phenotype.

Another test may involve crosses among other wild type strains showing similar base line levels of zinc tolerance. These crosses would provide opportunities to investigate effects due to heterosis. This heterosis effect may increase the vigour of F1 progeny over the parent due to heterozygote advantage (Wright

and Dobzhansky 1946). The result would be that the progeny of crosses for two similar wild type strains may tolerate higher levels of the zinc compared to both of the parents due to heterosis, not allelic differences in heavy tolerance genes.

Molecular genetic studies of metallothionein loci may be another approach to reveal genetic differences among the zinc tolerance strains. For example, genetic polymorphisms may be found at a metallothionein locus, and these polymorphisms could be related to the zinc tolerance exhibited by these strains consistent with what has been shown in *D. melanogaster* (see chapter 1).

3.3 Zinc tolerance based genetic sexing strains in the Medfly

One possible application of this research is to use heavy metal tolerant strains to achieve genetic sexing. A genetic sexing strain would be beneficial to sterile insect releasing programs where "males only" capability can be desirable (La chance 1979). In the Medfly, males are the heterogametic sex and they have a sex determining factor on the Y chromosome. Thus, Y chromosome translocation events had made genetic sexing strains possible (Robinson 1984, Kerremans *et al.* 1991). If a chromosomal fragment bearing a zinc tolerance alleles is translocated to a Y chromosome of the Medfly, any male carrying this Y translocated chromosome may tolerate zinc in the diet better than normal females.

To induce translocations, Medfly males were irradiated one day after an emergence with 50 Gy of gamma radiation (Busch-Peterson and Southern 1987, Kerremans *et al.* 1991). These irradiated males were from the A4A, A5A3, and 18A2A strains. A mating scheme was carried out to recover any male line that may be the founder of a male specific zinc tolerant sexing strain (figure 9). In the mating scheme, individual irradiated males were mated with three females from the Hilab strain. F1 progeny were raised in a 50mM zinc larvae diet in order to facilitate a selection for zinc tolerance phenotype. Each of the F1 surviving males from this selection were subsequently back crossed to three females from the Hilab strain in order to identify descendants that inherited the zinc tolerance phenotype through males only as a result of Y linkage.

About 130 irradiated males total from the A4A, A5A3, and 18A2A zinc tolerance strains were tested using the previously described mating scheme (figure 9). In preliminary observations, male descendants of 11 irradiated males (out of 130 tested) demonstrated a sex ratio deviation in favour of males in their corresponding F2 broods when these F2 larvae were raised with zinc diet. However, these lines did not demonstrate a consistent sex ratio in favour of males when again raised in a 50mM diet (data not shown).

Although only 130 irradiated males were tested, this number seemed reasonable based on previous work involving qualitative phenotypes such as dark pupae and apricot eye (Kerremans *et al.*

1991) and translocations in the Medfly. These qualitative traits are easier to score than quantitative phenotypes such as a zinc tolerance trait. Such quantitatively based traits can be confounded in a numbers of ways in terms of pupal recovery. For example, changes in traits related to larvae growth rate and to larvae viability might lower or elevate the pupal recovery number on any number of days used as a standard scoring time point.

The inability to generate a translocation based sexing strain here could also have resulted from the fact that the mode of inheritance of zinc tolerance is additive or due to polygenic interactions. This would make a single Y chromosome translocation event an ineffective means to generate a sexing strain. Thus further studies on the genetic basis for heavy metal tolerance will be essential to determine the feasibility of constructing of heavy metal tolerant genetic sexing strains.

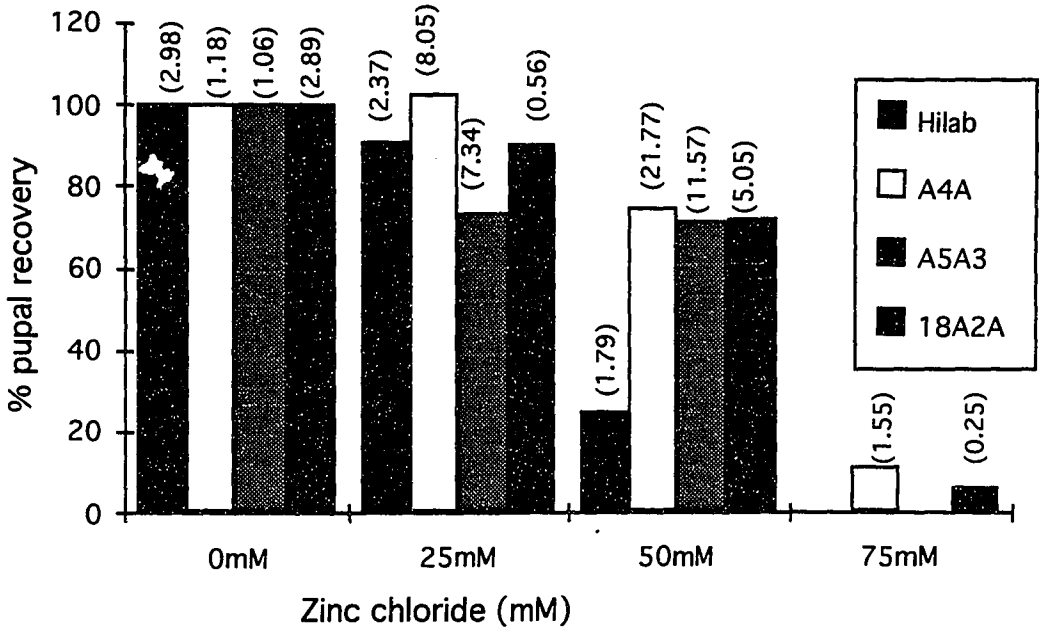


Figure 3. **Base line levels of zinc tolerance.**
Standard errors for the percent of pupal recovery are designated in a parenthesis on the top of each bar.

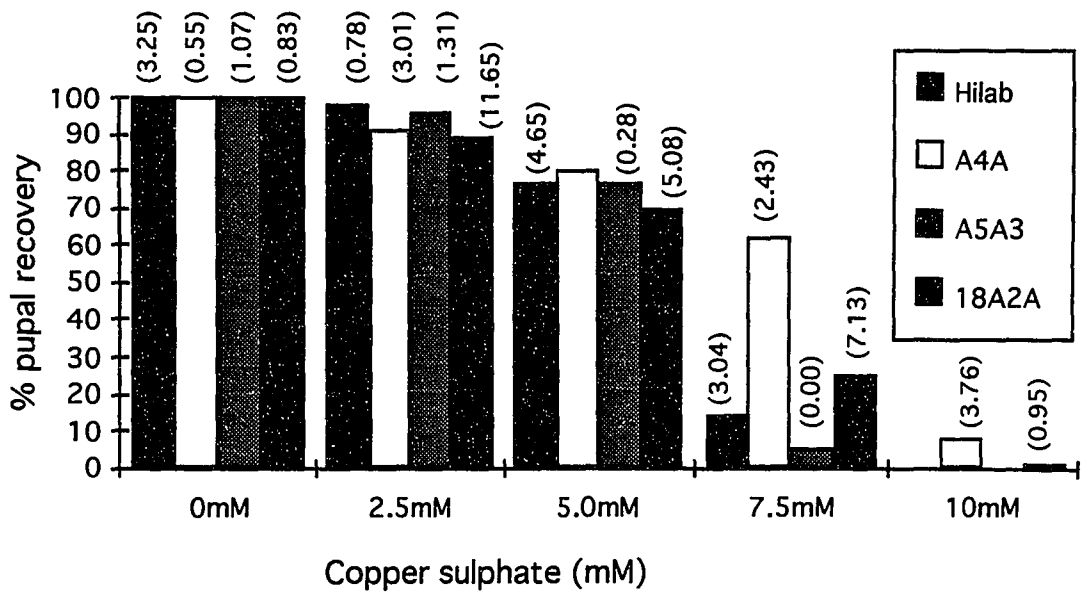


Figure 4. **Base line levels of copper tolerance.**
Standard errors for the percent of pupal recovery are designated in a parenthesis on the top of each bar.

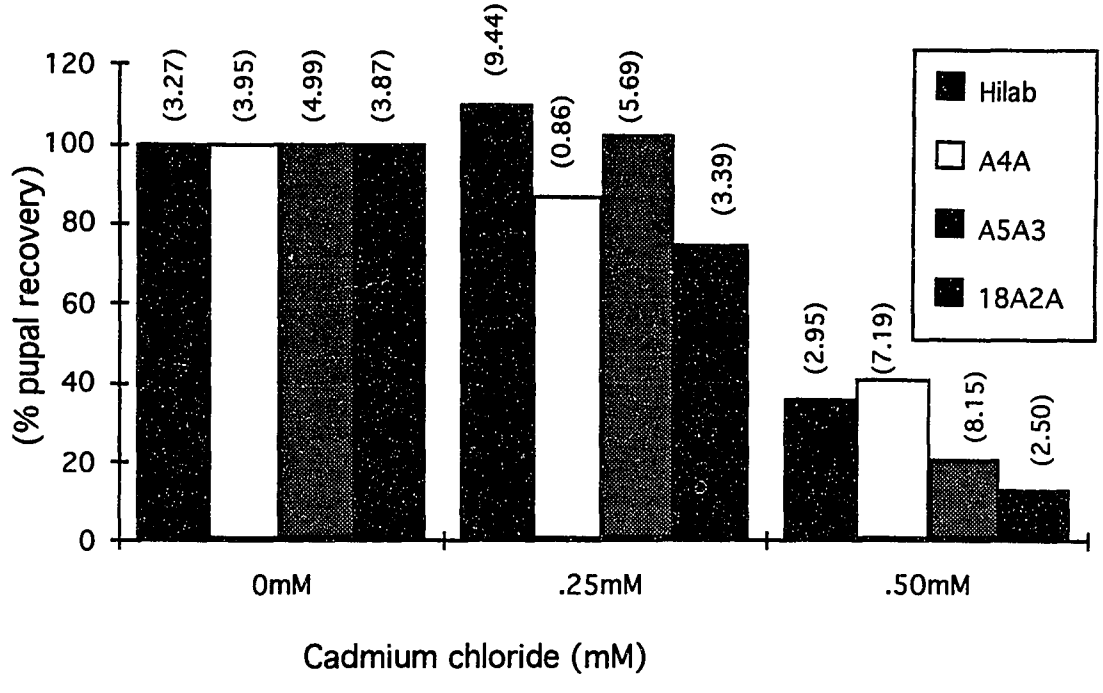


Figure 5. **Base line levels of cadmium tolerance.** Standard errors for the percent of pupal recovery are designated in a parenthesis on the top of each bar.

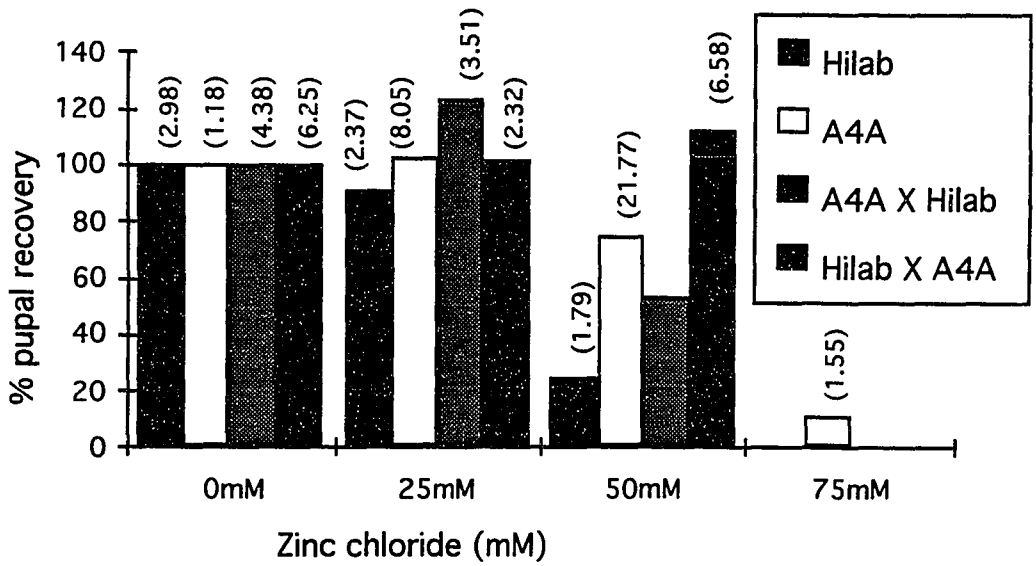


Figure 6. **Base line levels of zinc tolerance of A4A, Hilab strains, and reciprocal crosses.**
Standard errors for the percent of pupal recovery are designated in a parenthesis on the top of each bar.

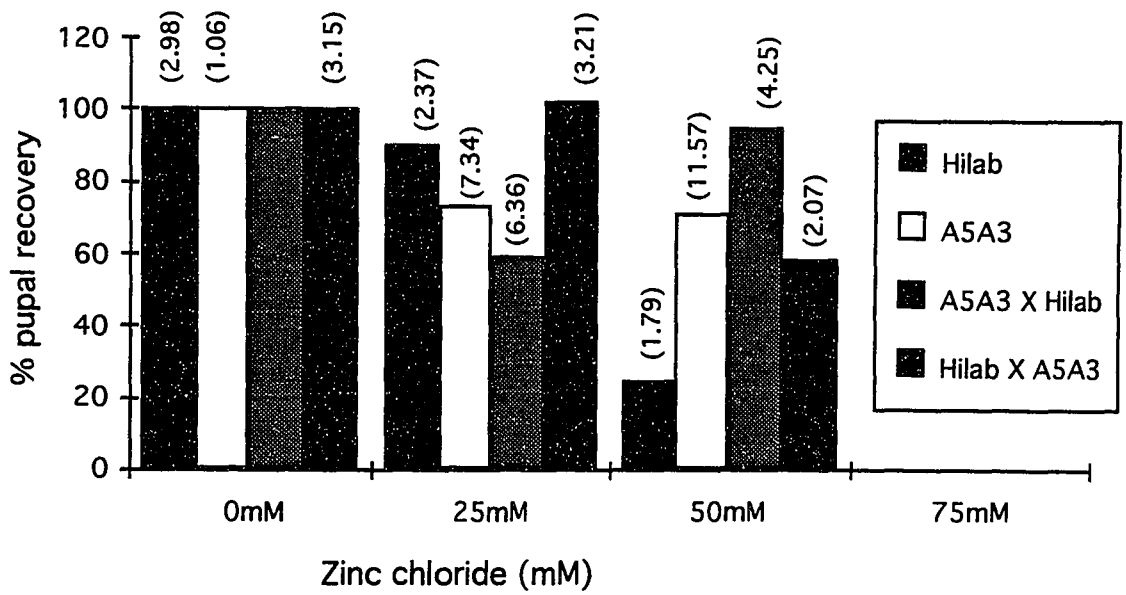


Figure 7. **Base line levels of zinc tolerance of A5A3, Hilab strains, and reciprocal crosses.**
Standard errors for the percent of pupal recovery are designated in a parenthesis on the top of each bar.

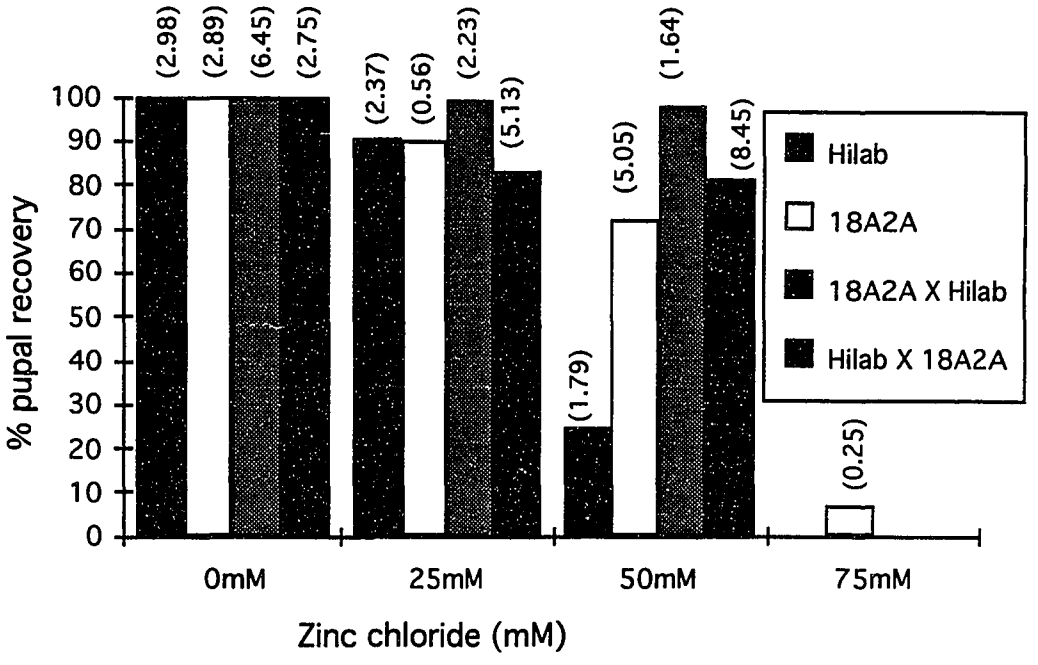


Figure 8. **Base line levels of zinc tolerance of 18A2A, Hilab strains, and reciprocal crosses.** Standard errors for the percent of pupal recovery are designated in a parenthesis on the top of each bar.

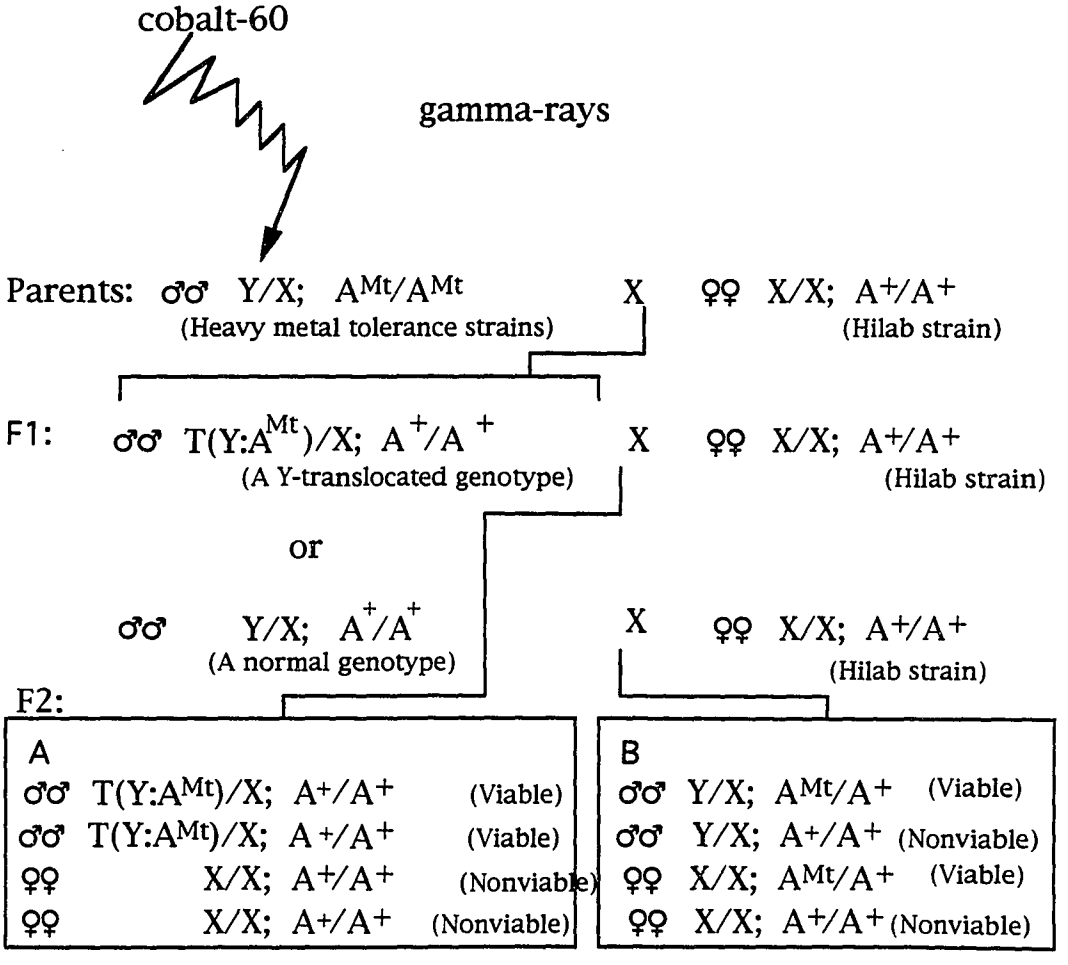


Figure 9. **Mating scheme for selecting zinc tolerant sexing strains.**

Box A demonstrates genotypes of a brood from which a sex ratio deviation in favour of the male gender is expected in 55mM zinc diet.

Box B demonstrates genotypes of a brood from which a sex ratio deviation in favour of the male gender is not expected in 55mM zinc diet.

Chapter 4

Genes associated with heavy metal tolerance trait

A PCR based approach was a method of choice in attempting the molecular cloning of metallothionein gene. Although the use of heterologous probes and plus/minus library screening have been successful approaches in other circumstances (Durnam *et al.* 1980, Karin and Richards 1982, Schmidt and Hamer 1983, Griffith *et al.* 1983, Nemer *et al.* 1984, Lastowski-Berry *et al.* 1985, Mokdad *et al.* 1987), they proved to be not feasible for the Medfly. Both a metallothionein cDNA probe (cDm51) and a metallothionein genomic DNA probe (pDm131) from *D. melanogaster* (Maroni *et al.* 1986) did not cross hybridize to the Medfly total genomic DNA at low stringency conditions (data not shown). Plus/minus cDNA library screening was also carried out. However, it was not successful due to a questionable quality of cDNA libraries and a lack of knowledge about the proper condition needed to induce metallothionein gene expression (data not shown).

A pair of 256 fold degenerate primers (so called primer I and primer II) were designed based on conserved cysteine rich metallothionein domains of *D. melanogaster* and based on the knowledge of codon usage of a Medfly actin gene (He and Haymer 1992) (figure 10). Primer I was designed based on a region coding for an amino acid sequence C N C G S D C (figure 10). This amino acid sequence is found in the second exon of the Mtn 1 gene (Maroni *et al.* 1986). The cysteine residues in this amino acid

sequence are also conserved in metallothioneins from various taxa (Nemer *et al.* 1985). The primer II sequence was designed from an amino acid sequence C P C G S G located in the first exon of the Mtn 1 gene (figure 10). Hence the use of the primer I and the primer II was expected to amplified a part of any metallothionein gene which was similar to the Mtn gene.

4.1 Characterization of PCR products

PCR reactions were conducted using primer I, primer II, and Medfly genomic DNA templates. Temperature profiles for PCR reactions were set to achieve specific PCR products. Primer annealing temperature was varied among different experiments in order to search for an optimal condition for specific binding to the genomic DNA templates. Trial annealing temperature were 48°C or 55°C. Genomic DNA templates from *D. melanogaster* were used as a positive control. This was because the primer I and the primer II should at least bind to the Mtn 1 gene.

PCR products were electrophoresed, Southern blotted, and probed with the PCR products themselves (figure 11). A few major bands of the PCR products are seen with molecular weights below a 0.58 kb DNA marker. The sizes of these PCR products appeared to be consistent with what might expected for a Mtn gene (Maroni *et al.* 1986). Specificity of the PCR products appeared to not differ between PCR reactions using 48°C and 55°C annealing temperatures. This can be seen in figure 11.

Since it was possible that the PCR products may contain sequences from metallothionein genes, they were used as a probe to screen a Medfly EcoRI restriction fragment genomic DNA library. Three genomic clones were isolated from a genomic DNA library screening experiment. These three genomic clones were named Mt1, Mt2, and Mt3, respectively. The Mt1, Mt2, and Mt3 genomic clones were subcloned in to plasmid pUC 9 for further molecular characterization.

4.2 Characterization of the Mt1, Mt2, and Mt3 subclones

The Mt1, Mt2, and Mt 3 subclones were restriction digested with EcoRI. The subclones and EcoRI restriction digested genomic DNA from the Medfly were Southern blotted and probed with the previously described Medfly PCR probe. Figure 12 shows that the Mt1, Mt2, and Mt3 DNA fragments might contain sequences similar to the previously characterized PCR products because they cross hybridized well to the Medfly PCR probe.

A Southern blot hybridization between the Medfly PCR probed and an EcoRI restriction digest genomic DNA of the Medfly produced multiple bands of hybridization (figure 12). This suggests that the PCR products contained repetitive DNA, and that these may be dispersed in the Medfly genome. Hybridization of the Mt1 or Mt2 or Mt3 probes to similar genomic Southern blots also showed multiple bands signals (data not shown). This suggested that there were also repetitive sequences within the Mt1, Mt2, and Mt3 subclones.

The Mt1, Mt2, and Mt3 DNA fragments were individually used as templates together with the primer I and the primer II in PCR reactions. The Mt1 and the Mt2 DNA templates produced a PCR product similar in size to the PCR product derived from the Medfly genomic DNA templates (data not shown). This suggested that the Mt1 and Mt2 subclones contain DNA sequences similar to the PCR product which derived from the genomic DNA templates.

In order to evaluate the possibility that any of the Mt1, Mt2, and Mt3 fragments may contain a coding region of a metallothionein gene, a search for a long open reading frame (which may code for cysteine rich polypeptide) was carried out by analyses of DNA sequences within the Mt1, Mt2, and Mt3 fragments.

A 195 base pair open reading frame (ORF) was discovered in the Mt1 fragment. This 195 base pair ORF possessed a deduced cysteine rich amino acid sequence (figure 13). It contained 10 cysteine residues out of a total of 65 amino acid residues (15%). This was an unusally high cysteine composition for typical proteins. For example, an actin protein of the Medfly had only 1.59% cysteine residues (He and Haymer 1992). A heat shock cognate 1 gene had only 0.76% cysteine residues (table 2). Thus, this 195 base pair ORF might be a coding region for a cysteine rich protein such as a metallothionein. A database search comparison demonstrated that deduced amino acid sequence of the 195 nucleotides ORF was similar to a number of mammalian metallothioneins (figure 16). Therefore, there was a potential that

the 195 nucleotides ORF could be used as a molecular tool for the isolation and characterization of a metallothionein gene.

A 382 base pair Pst I- Xho I fragment was taken out from the Mt1 subclone (figure 13a). This Pst I- Xho I fragment was named a Cysteine rich region (CRR) because it contained the previously described 195 base pair ORF sequence flanked by a 90 base pair and a 97 base pair DNA fragments (figure 13a). The CRR fragment was further characterized in a variety of ways in order to evaluate its potential as a metallothionein gene.

4.3 Characterization of cysteine rich regions (CRRs)

4.3.1 Restriction map and Southern blot analyses

The CRR sequence appeared to be in all of the Mt1, Mt2, and Mt3 subclones because the CRR fragment hybridized to them on a Southern blots under high stringency condition (figures 14 and 15). Restriction maps and Southern blot analyses of the Mt1, Mt2, and Mt3 subclones demonstrated that flanking regions of the CRR fragments of the Mt1 and Mt3 subclones could hybridized to flanking regions of the CRR fragment of the Mt2 subclone. A Mt1 probe hybridized to a Pst I-Xba I fragment of the Mt2 while the CRR probe did not. Likewise, a Mt3 probe hybridized to a Xho I-Cla I fragment of the Mt2 (figure15). Similarities among the Mt1, Mt2, and Mt3 subclones suggested that their CRR fragments as well as the other restriction fragments may share coding potential for cysteine rich proteins.

4.3.2 RNA slot blot analyses

In order to test if any of the CRR containing Mt1, Mt2, and Mt3 fragments coded for a metallothionein gene, the CRR, Mt1, Mt2, and Mt3 fragments were individually used as ^{32}P labeled probes to a number of total RNA slot blots. Total RNA used in these RNA slot blot experiments came from heavy metal induction systems. A pmed 21 actin gene probe was used in the RNA slot blot experiment to standardize the densitometer readings of the results.

In general, induction of metallothionein gene expression usually requires sublethal doses of heavy metal (Hammer 1986). These sublethal doses of heavy metal could be deduced from the studies of baseline levels of heavy metal tolerance described in chapter 3. Since the Medfly adults ingested different type of diet, the same heavy metal solutions used in preparation of the heavy metal larvae diets were given to the Medfly adults as liquid for three days. These adults were subsequently used as sources of total RNA.

Densitometer readings of RNA slot blots demonstrated that all of the CRR, Mt1, Mt2, and Mt3 probes yield higher intensity of hybridization signals in heavy metal induced total RNA samples (data not shown). Although this suggested that the CRR, Mt1, Mt2, and Mt3 fragments might code for heavy metal induced transcripts, RNA northern blot experiments using poly(A) RNA selected from the same total RNA meterials could not confirm this suggestion.

4.3.3 Poly(A) selected RNA northern blot analyses

Poly(A) selected RNA northern blots did not show any hybridization signals in all samples (data not shown). This suggested that the hybridization signals observed in the total RNA slot blot experiments were not derived from mRNA transcripts. This was supported from a fact that the total RNA samples and poly(A) minus RNA samples on the same northern blots demonstrated a high molecular weight hybridization signal similar in size with a signal observed in a total genomic DNA control lane (data not shown). This suggested that it could be the contamination of genomic DNA in the total RNA samples that hybridized to the CRR, Mt1, Mt2, and Mt3 probes.
This contaminated genomic DNA could strongly hybridized to the Mt1, Mt2, and Mt3 probes because they contained highly repetitive DNA. Hence, the RNA slot blot signals might not be evidence for heavy metal induction.

The analyses of the total RNA slot blot experiment and the poly(A) selected RNA northern blot experiment inferred that the sequences in the CRR, Mt1, Mt2, and Mt3 subclones did not code for any heavy metal induced transcripts. Nonetheless, it was possible that the transcripts exist at such a low level that they appear undetectable.

4.3.4 cDNA library screening for the CRR transcripts

The other attempt to detect CRR transcripts was to use the CCR as a probe to screen cDNA libraries. An amplified embryo cDNA library and an amplified pupal cDNA library were screened with the CRR probe. There was no hybridization signal to the CRR probe when twenty thousand plaques from the embryo cDNA library and the pupal cDNA library were screened.

Although an attempt to identify homologous clones in the cDNA libraries with this probe was not successful, this does not infer that cDNA library screening is not a useful approach. This amplified cDNA library may not be representative. A large portion of cDNA clones might be lost in the library amplification processes. Moreover, the embryo stage and the pupal stage might have such low levels of the transcript that twenty thousand plaques from the embryo and the pupal derived cDNA libraries was not a sufficiently large number. Screening of a cDNA library derived from a heavy metal induced larvae could lead to better success because this cDNA library might be rich in heavy metal induce transcripts. An attempt to construct this type of library was carried out. However, the titer of this newly constructed cDNA library was too low to be used.

4.3.5 Amplification from heavy metal induced cDNA templates

Another method to detect transcription of the CRR fragment was to employ a PCR approach. If there were heavy metal induced

transcripts from the CRR fragment and/or the metallothionein genes, it should be possible to amplify these transcripts from a pool of cDNA templates derived from a heavy metal induced system.

A new pair of degenerate PCR primers (so called primer III and primer IV) had nucleotide sequences deduced from two specific regions within the CRR nucleotide sequence (figure16). These two specific regions represented regions where blocks of conserved amino acid residues of the CRR fragment aligned to a number mammalian metallothioneins (figure16).

In order to synthesize heavy metal induced cDNA template, newly emerged Medfly adults had been treated with a 20mM $CuSO_4$ drinking water solution for 3 days before they were sacrificed for a total RNA extraction. The total RNA was subsequently selected for poly(A) RNA. The poly(A) selected RNA was then used as a template for an oligo-dT primed cDNA synthesis.

Primers I, II, III, and IV each were also used with an oligo-dT (15-mer) primer to amplify a pool of heavy metal induced cDNA templates. The Mt1 fragment was also used as a positive control template to assure optimal PCR reactions. Primer III and IV had been designed based on nucleotide sequences within the Mt1 fragment.

The use of the oligo-dT primer with any of the primers I, II, III, and IV did not appeared to amplified any PCR product.

However, specific PCR products were achieved when primer IV was used alone or primers II and IV were used together.

An approximately 500 base pair specific PCR product was amplified using heavy metal induced cDNA templates and primer IV alone. This 500 base pair PCR product was subsequently cloned and name P4CRF1. Another three specific PCR products, approximately 300 to 500 nucleotides in length, were produced from PCR reactions using primer II and primer IV on the heavy metal induced cDNA templates. Likewise, the three specific PCR products were subsequently cloned and named P2:4CRF1, P2:4CRF2, and P2:4CRF3, respectively.

4.4 Identification of coding potential of the P4CRF1, P2:4CRF1, P2:4CRF2, and P2:4CRF3 clones

The P4CRF1, P2:4CRF1, P2:4CRF2, and P2:4CRF3 fragments were used to obtain approximately 150 base pair of sequence information from both of their termini. This was to generate nucleotide sequence and deduced amino acid sequences for a database search. A search for similar sequences within the entire EMBL sequence database was performed using any sequences from the P4CRF1, P2:4CRF1, P2:4CRF2, and P2:4CRF3 clones as a query. The computer listed a rank of similar sequences based on their similarity scores to the query sequence.

This database search comparison using the P4CRF1 sequence information gave rise to a list of *hsp70* multigene family sequences. The heat shock 70 kDa sequences appeared in the

database search comparison at both the nucleotide level and the amino acid level (data not shown). The heat shock 70 kDa sequences in the lists were from a number species across taxa. This included a top ranked *hsc4* from *D. melanogaster* (Drohsc4). A deduced amino acid sequence from the P4CRF1 clone perfectly matched to an amino acid sequence from Drohsc4 for a length of 22 amino acids. This suggested that the P4CRF1 fragment was probably from a coding region of a heat shock 70 kDa like gene.

Database search comparisons for the P2:4CRF1 sequence gave rise to a list of ranked sequences of vacuolar ATPase B subunit genes from various taxa. The P2:4CRF1 sequence showed 97% indentity for a length of 36 amino acids to a vacuolar ATPase subunit gene from the *D. melanogaster.* Hence the P2:4CRF1 may be part of a vacuolar ATPase subunit gene.

Likewise, database search comparisons for a sequence from the P2:4CRF2 clone demonstrated that it might be a coding region for a gene similar to a flightin gene in *D. melanogaster.* This flightin sequence was ranked at the top of the list of sequences at both the nucleotide level and the amino acid level of the database search comparisons. A deduced amino acid sequence from the P2:4CRF2 clone showed 76% identity for the length of 30 amino acids to the flghtin gene.

Database search comparisons for sequences from the P2:4CRF3 clone gave rise to a list of sequences that appeared to be unrelated at both the nucleotide level and the amino acid level. The P2:4CRF3 clone might be coding for a type of protein that was

not conserved at either the amino acid level or the nucleotide level. Alternatively, the P2:4CRF3 clone might code for a conserved protein domain that was shared by a number of different type of proteins. A longer sequence information from the P2:4CRF3 clone might be more informative for the database search comparisons because it would allow the search to be more specific.

The P4CRF1, P2:4CRF1, P2:4CRF2, and P2:4CRF3 clones did not appear to have coding potential as metallothionein genes. They did not produce any cysteine rich regions nor did they have metallothionein sequences appearing in the lists of similar sequences derived from database search comparisons.

Characterization of nucleotide sequences at terminal regions of the P4CRF1, P2:4CRF1, P2:4CRF2, and P2:4CRF3 clones demonstrated that there were a few short open reading frames with a number of deduced cystein residues (data not shown). This suggested that the PCR primers I, II, III, and IV did actually bind to the TGC and/or TGT rich regions.

Although a metallothionein gene was not cloned by employing the previously described PCR approach, a stress associated heat shock like gene had been cloned. The P4CRF1 clone was selected for further molecular characterization because it was the only coding sequence that appeared to be associate with a stress response gene. It was known that heat shock genes could also respond to heavy metal induction in a number of systems. Hence, the P4CRF1 clone may be used as a probe for the isolation

and characterization of a stress response associated heat shock gene.

4.5 Speculation on the nature of the 195 base pair CRR open reading frame

The 195 base pair CRR open reading frame is unlikely to be a random nucleotide sequence representing a noncoding region. This is because the probability that a randomly generated nucleotide sequence gives rise to a 195 nucleotide open reading frame would be $(61/64)^{n/3} = 0.044$ (4.4%) where n is the length of nucleotide (Ohno 1984). However, it is unlikely that the 195 base pair fragment represents a coding region of a functional gene because there was no evidence of any transcription.

The presence of any internally repetitious coding sequence could lead to a higher likelihood of a long open reading frame because the internal repetitions do not have stop codons (Ohno 1984). But the 195 base pair sequence here does not show obvious characteristics of an internally repetitious sequence. This is suggested by a self homology search and a directed repetitive sequence search. However it is possible that the 195 base pair sequence was once an internally repetitious sequence and it has been changed to an extent that its repeating unit was no longer recognizable. The possibility of once having an internally repeated sequence was also supported by a fact that the complementary strand of this 195 base pair sequence is also an approximately 180 base pair open reading frame. This fact is consistent with the

rationale that if one sequence has internal repetitions, its complementary sequence should be internally repetitious too. Additional evidence for the internally repetitive nature of this 195 base pair sequence is that the CRR fragment appears to contain and/or is surrounded by repetitive DNA. When the flanking regions of the 195 base pair sequence were used as probes, they showed a repetitive pattern of hybridization to the Medfly genomic DNA. It is known that repetitive DNA islands are conducive to generate further repetitive DNA due to intrastrand slipped-mispair during DNA replication (Levinson and Gutman 1987).

An alternative possibililty is that the 195 base pair could be a metallothionein pseudogene. The translated amino acid sequence of this 195 base pair fragment shows a number of conserved amino acid residues similar to mammalian metallothionein proteins (figure 16). Moreover, 7 out of 11 of these amino acid residues are cysteine residues representing a characteristic of metallothionein protein.

Figure 10. **Design of primer I and primer II.**

An amino acid sequence of a metallothionein (Mtn 1) gene from *D. melanogaster* is shown at the top. Two conserved metallothionein domains were designated in the open boxes I and II (Nemer *et al.* 1985). The sequence of the primer I was designed from the underlined amino acid sequence in the box I. Likewise, the sequence of the primer II was designed from the underlined amino acid sequence in box II.

Primer I (21 mer) 5'-GCA ATC ACA NCC GCA ATT ACA-3'
G GGT G G

Primer II (18 mer) 5'- TGC CCA TGC GGN ACN AGT-3'
T TG G

Figure 11. **A Southern blot of Medfly PCR products.**

All of the PCR products were amplified using primer I and primer II. The first two PCR samples (left panel) came from the Medfly genomic DNA templates. Likewise, the last two PCR samples (right panel) came from genomic DNA templates of *D. melanogaster*. Annealing temperatures (^{o}C) of all PCR reactions for the PCR samples are designated in parentheses. The PCR samples were self-probed by a labeled Medfly PCR product (48 degree) in the left panel. Likewise, the PCR samples of the right panel were self-probed by a labeled *D. melanogaster* (48 degree) at a high stringency.

Medfly PCR product
(48 degree)
Medfly PCR product
(55degree)
D.m. PCR product
(48 degree)
D.m. PCR product
(55 degree)
-23.1 kb.
-9.46 kb.
-6.75 kb.
-4.62 kb.
-2.26 kb.
-1.98 kb.
-0.58 kb.

Figure 12. **A Southern blot of the M1, Mt2, Mt3, and genomic DNA from the Medfly probed with the Medfly PCR product.**

The Mt1, Mt2, Mt3 subclones and the genomic DNA of the Medfly were restriction digested with EcoRI, Southern blotted and probed with a Medfly PCR probe (described in the figure 11) under high stringency conditions.

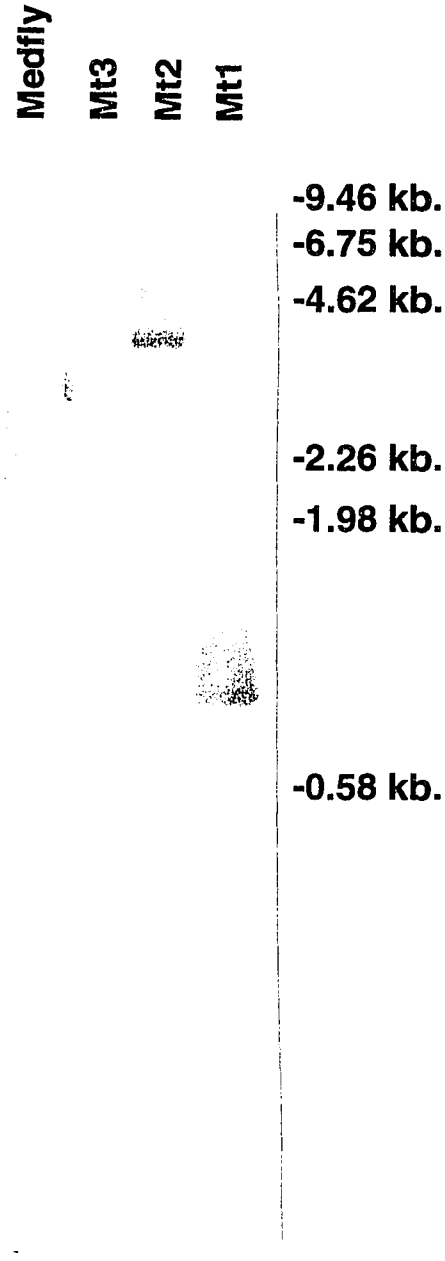

Medfly
Mt3
Mt2
Mt1
-9.46 kb.
-6.75 kb.
-4.62 kb.
-2.26 kb.
-1.98 kb.
-0.58 kb.

Figure 13. **A. A nucleotide sequence of an Xho I- Pst I fragment containing a cysteine rich region (CRR).**

B. A deduced amino acid sequence of a 195 base pair open reading frame within the CRR fragment.

A 195 base pair open reading frame sequence is underlined. Xho I and Pst I recognition sequences are designated with open boxes.

A.

Xho I

CTCTCGAGGC AGCTCAAGCT CGTTGTCTGC TATGGGTGGT

GGTATTTCTC CGAGTACGCC ATTCACTGGA AGGGAGTCCT

TGCCGGTGTT GACTGACTCC TGGTGCTTGT CGGTACGCCG

TCCTGTCCAG CGGTTGTTGT ATGTGTTGCT GGATTGTGTC

CTTTAAGTCT AGTACTCCCT TTCTGATGTG CCTAGGTGGA

GGTTCAGCCT CTATTAGGTG CCTACTGGCG TGCTCTTGCG

GTAGCATCCC AATAGGAACT GCTTGCACAG CAAAGCATTA

TGCTCCTTAA CTCGTAGCAT AGATGCTTCC CTGTGCAAGT

GATGCTCCGG CGACATCAGG AGGTTTCCTG TCGCGGTCCT

AGGGCGGTGT TCTGACAACC TGCAGCTTCT TCATCTGCGT

Pst I

B.

LTPGACRYAVLSSGCCMCCWIVSFKSSTPFLMCLGGGSASIRCLLACSCG

SIPIGTACTAKHYAP

Figure 14. **A Southern blot of the Mt1, Mt2, and Mt3 subclones probed with the CRR fragment.**

The Mt1, Mt2, and Mt3 subclones were restriction digested with EcoRI, Southern blotted and probed with the CRR fragment under high stringency conditions.

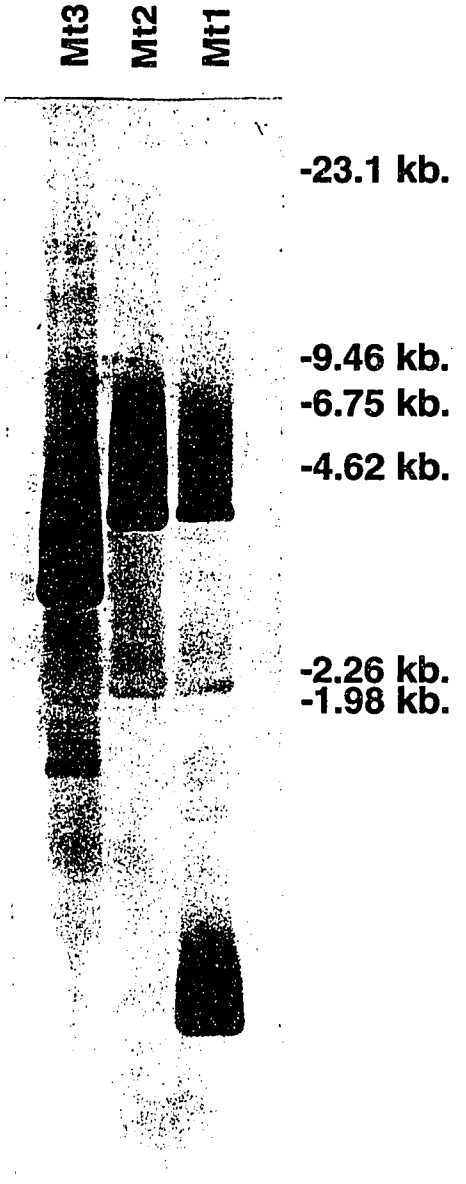

Mt3
Mt2
Mt1
-23.1 kb.
-9.46 kb.
-6.75 kb.
-4.62 kb.
-2.26 kb.
-1.98 kb.

Figure 15. **Restriction maps of the Mt1, Mt2, and Mt3 fragments.**

A thick line within each of the Mt1, Mt2, and Mt3 restriction maps represents a restriction fragment on which a CRR like sequence is located.

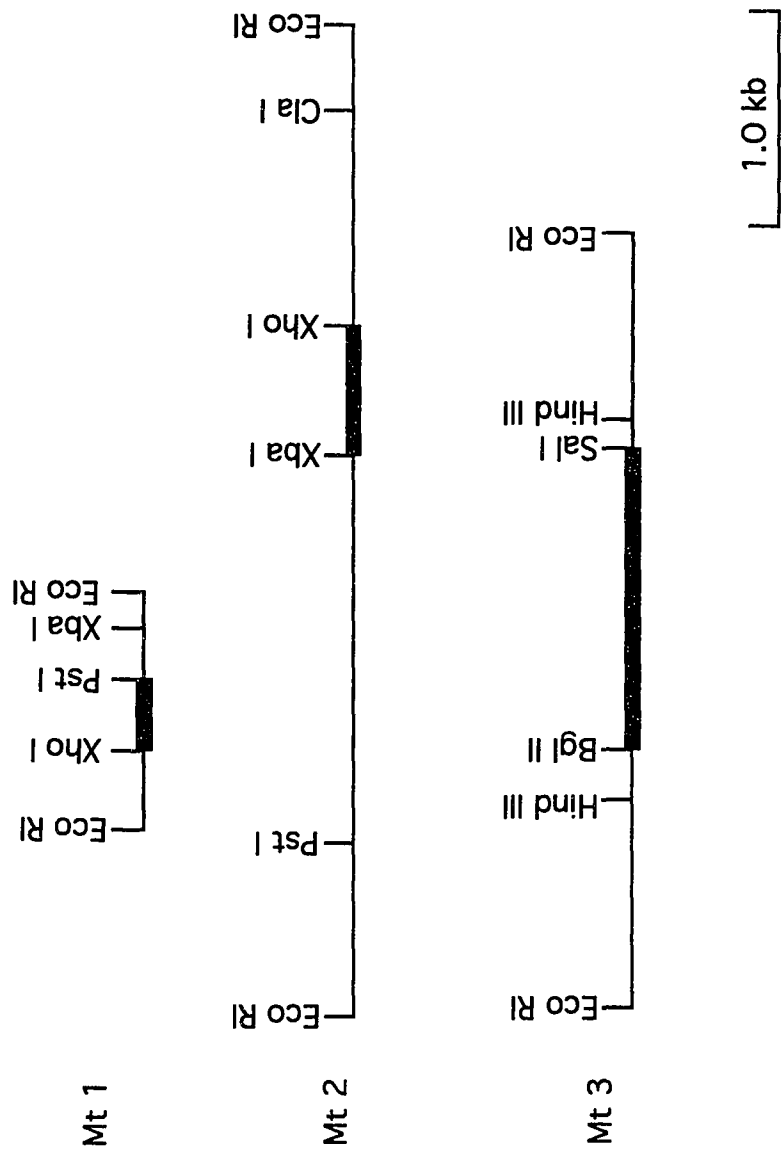
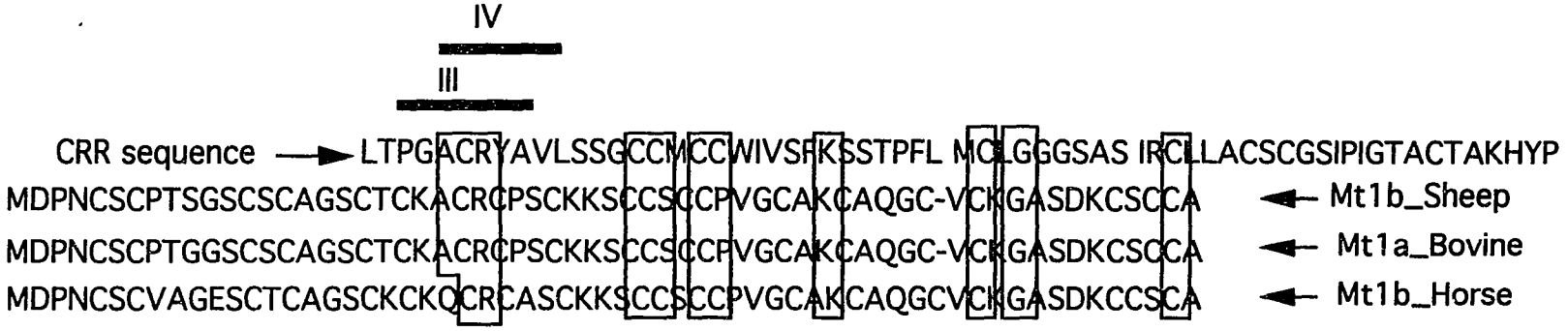

Mt 1
Eco RI
Xho I
Pst I
Xba I
Eco RI
Mt 2
Eco RI
Pst I
Xho I
Xba I
Cla I
Eco RI
Mt 3
Eco RI
Hind III
Bgl II
Sal I
Hind III
Eco RI
1.0 kb

Figure 16. **Alignment of deduced amino acid sequences from the CRR fragment and mammalians metallothionein proteins.**

A number of conserved amino acid residues of a sheep metallothionein (Mt1b), a bovine metallothionein (Mt1a), a horse metallothionein (Mt1b) and the CRR fragment are designated in open boxes.

The oligonucleotide sequence for primer III was deduced from a block of conserved amino acid residues. This block of conserved amino acid is indicated by a thick bar (designated III). Likewise, the oligonucleotide sequence of primer IV was deduced from a block of conserved amino acid residues.



Primer III (21 mer) 5'-CCT GGT GCT TGC CGN TAC CCN-3'
T GT G

Primer IV (21 mer) 5'-CAG GAC NGC ACA NCG ACA AGC-3'
G GT G

Chapter 5

Characterization of a heat shock like gene

The PCR product (P4CRF1-1) carrying 400 bases of a potentially heat shock like coding sequence was used to probe a Southern blot containing restriction digested Medfly genomic DNA. An 887 base pair EcoRI fragment appeared to be the only genomic DNA fragment hybridizing to the P4CRF1-1 probe. This suggested that the P4CRF1-1 probe would be specific enough to isolate this putative heat shock like gene from the genome of the Medfly. Four positively hybridizing genomic clones were subsequently isolated from a screen of a total of 20,000 plaques of a Medfly EMBL4 genomic library. DNA from these genomic clones were restriction digested with EcoRI, Southern blotted, and reprobed with the P4CRF1-1 probe. It was verified that each of these isolated genomic clones contained the 887 base pair EcoRI fragment seen previously (data not shown).

The four isolated phage clones appeared to contain two types of insert fragments. One pair of phage clones had identical insert fragments. The other two phage clones also contained identical inserts, but only the 887 base pair fragment was common to all cases. The 887 base pair EcoRI fragments from these two different types of genomic clones were subcloned and partially sequenced. It was found that for at least 169 nucleotides, these two EcoRI fragments were identical in terms of nucleotide sequence. This suggested that the two 887 base pair EcoRI fragments were either

derived from a single genomic location or from very recently duplication events representing different loci.

5.1 Genomic organization of a heat shock like gene

The previously described 887 base pair EcoRI fragment (designated 6-9) appears to be a portion of a heat shock like gene because it has nucleotide sequences very similar to a heat shock cognate 4 of *D. melanogaster* (Drohsc4). Nucleotide sequence aligment between the sequences of the 6-9 fragment and the Drohsc4 gene suggests that the flanking regions of the 6-9 fragment should contain coding sequence as well. To subclone flanking regions of the 6-9 fragment, one of the isolated phage clones (P4CRF3) was restriction digested with EcoRI and shot gun subcloned. Six different EcoRI subcloned fragments were then partially sequenced from both termini in order to identify subclone fragments which may be flanking the 6-9 fragment and may contain a heat shock coding region. Three EcoRI subcloned fragments appear to have nucleotide sequences aligning with the coding region of the Drohsc4 gene. These three fragments are the 887 bases EcoRI subclone (6-9) itself and two EcoRI flanking fragments designated 6-5 and 6-1 (figure 17).

In order to document whether the 6-5, 6-9, and 6-1 fragments were contiguous in the genome, a specific pair of PCR primers were designed based on nucleotide sequences from both the subclones 6-5 and 6-1. When these PCR primers were used in a PCR reaction with the Medfly total genomic DNA templates, a

specific PCR product spanning these fragments was amplified. This PCR product was then partially sequenced from both ends. Over 400 bases of these nucleotide sequences were perfectly matched to connected regions of the 6-5, 6-9, and 6-1 fragments. Moreover, EcoRI sites were found at each the connecting points among the fragments (figure 17).

The entire heat shock like nucleotide sequence was obtained from a 3,286 base region DNA which comprised the entire 6-5 and 6-9 EcoRI fragments as well as a small portion of the 6-1 EcoRI fragment (figure 18). This heat shock like gene from the Medfly genome was named heat shock cognate1 (Cerhsc1) because it was the first characterized heat shock like gene from the Medfly genome.

5.2 Gene structure

The 3.28 kb DNA fragment containing the Cerhsc1 gene includes an open reading frame (ORF) 1,962 nucleotides in length. This ORF begins with an ATG codon and ends with a stop codon (TAA) (figure 18). This ORF shares 96.2% identity at the amino acid level and 81.8% identity at the nucleotide level to the Drohsc4 gene. This suggests that this is likely to be an ORF of a heat shock like gene. A putative promotor region was located approximately 650 bases upstream of the start codon ATG. Within this region, a number of heat shock elememts, CAAT boxes, and a few putative TATA boxes were identified. Also, a number of putative polyadenylation sites located 150 bases downstream of the TAA

stop codon were identified. This coding region appeared to be intronless because the deduced amino acid sequence alignment to the Drohsc4 gene did not reveal any coding interruption.

The Cerhsc1 gene could be predicted to produce a protein 653 amino acids long having a predicted molecular weight of 71.1 kDa. This molecular weight of 71.1 kDa is consistent with the idea that the Cerhsc1 gene belongs to the *hsp70* multigene family. The %GC content of the coding region of the Cerhsc1 gene is 47.14%. This %GC content is in a typical range for a number of genes in the Medfly genome. For example, an actin gene had 52 %GC (He and Haymer 1992) and an Cu, Zn superoxide dismutase (Sod) gene had 47 %GC (Kwaitowski *et al.* 1992). In *D. melanogaster,* homologous genes appear to show higher %GC content. The *hsc1* gene (Drohsc1), the *hsc4* gene (Drohsc4) , the Cu, Zn superoxide dismutase gene (DroSod), and an actin gene from a genome of *D. melanogaster* have 57%, 59%, 60%, and 61% GC contents, respectively. The %GC content differences between *D. melanogaster* genes and the Medfly genes reflect codon usage differences in these species (Kwiatowske *et al.* 1992, He and Haymer 1992). The %GC content in the immediate 5' noncoding region of the Cerhsc1 gene was as low as 27%. This low %GC content was consistent with a similar noncoding region of the Medfly Sod gene (Kwiatowski *et al.* 1992).

The codon usage of the Cerhsc1 gene is shown in comparison to the codon usage of the Medfly actin gene, the Drohsc4 gene and the Drohsc1 gene in table 1. The overall pattern of codon usage for

the Cerhsc1 gene appeared to be more similar to the Medfly actin gene than to the Drohsc1 gene and to the Drohsc4 gene. This suggests that, the codon usage may be more similar among different genes from the same genome than among homologus genes from different genomes. However, this suggestion need to be confirmed by characterization a number of genes. The codon usage patterns for the Drohsc4 gene and the Drohsc1 gene appear to be very similar (table 1), probably reflecting the fact that they are members of a gene family. However, there are some examples of codon usage that seem to be different in the case of the Drohsc1 gene. With arginine, for example, only two out of six possible codons were used in the Drohsc4 gene while all of the six possible arginine codons were used in the Drohsc1 gene.

Amino acid composition for the Cerhsc1 gene reveals a different type of comparison. The overall amino acid composition for the Cerhsc1 protein was more similar to the Drohsc4 protein and the Drohsc1 protein than to the Medfly actin gene. The Cerhsc1 had 8.57% alanine, 8.42% glycine, and 8.72% threonein. These alanine, glycine, and threonein levels also appear in the Drohsc1 and Drohsc4 proteins (table2). In the case of arginine, there was 6.7% arginine in the Drohsc1 protein but only 4.44% and 4.59% arginine in the Cerhsc1 protein and the Drohsc4 proteins, respectively (table 2). These types of similarities in amino acid composition among heat shock cognate genes may represent functional constraints to conserve protein sequences.

The Cerhsc1 gene was low in cysteine composition (0.76%) This low cysteine composition distinguishes the Cerhsc1 gene from metallothionein genes. Moreover, there is no cysteine rich region on the Cerhsc1 protein. These five cysteine residues found are dispersed along the length of the Cerhsc1 protein (figure 19).

The whole Cerhsc1 coding region was compared with 18 different heat shock 70 kDa genes from 10 species across a number of taxa (table 3). At the amino acid level, the Cerhsc1 gene is most similar to the Drohsc4 gene (96% identity). The Cerhsc1 protein sequence can also be grouped with a number of cytoplasmic localized *hsc* genes from various species (figure 22). This grouping was based on the degree of amino acid sequence similarity among all aligned sequences. The Cerhsc1 amino acid sequence showed 82.2% identity to the Drohsc1a, 82.7% identity to the Humhspa2a, and 72.5% identity to the Drohsp7a2. In addition, the Cerhsc1 protein sequence was also grouped with the *hsc* Celhsp70 and Scssa1 proteins from the genome of *C. elegan* and *S. cerevisae*. Hence, it appears that the Cerhsc1 gene is likely to be a *hsc* gene which has a gene product localized to the cytoplasm.

Taking all of the amino acid sequences comparisons together, these results are consistent with a previous observation that HSP70 proteins localized in the same subcellular compartment and expressed under similar heat shock conditions seem to be more similar in their overall amino acid sequences (Boornstein *et al.* 1994). Hence, there appears to be a strong relationship between structure and function within the *hsp70* multigene family.

In figure 26, average similarity scores along the whole length of compared HSP70 protein sequences suggest that the first 50 amino acids and the last 60 to 70 amino acids are the most variable regions. This indicates that the end of each HSP70 proteins can be crucial regions for distinguishing one HSP70 member from one another. Amino acid sequence alignments also show that the mitochondrion localized Drohsc5 has a 50 amino acid leader sequence (figure 24). This leader sequence possesses amino acid residues similar to a number of other general mitochondrion associated proteins (figure 24). The leader sequence consists of basic amino acids (such as K, R, and, H) spaced 3 to 5 residues apart, as well as S residues interspersed though out the length of the leader sequence. Likewise, all of the endoplasmic recticulum localized HSP70 proteins (Drohsc3a, kar2, and Tobacco) showed long leader sequences with amino acid residues that are common among most if not all endoplasmic recticulum associated proteins. The long leader sequence consist of a large number of hydrophobic amino acid residues including A, V, L, I, P, W, F, and M. All of the cytoplasmic localized HSP70 proteins, including the Cerhsc1 protein, appear to have no such leader sequences, This can be seen from the fact that these sequence align beginning with a conserved region around amino acid position 51 (figure 24).

Amino acid sequences at the C-terminal end of the Cerhsc1 protein share a conserved terminal sequence with other cytoplasmic localized HSP70 proteins (figure 25). These amino acid

sequences consist of (T/K) (I/V) E E V D residues shown as sequence alignments from the C-terminal end.

The nucleotide sequence of the Cerhsc1 gene was also compared to the same set of the *hsp70* genes (table 3). Multiple nucleotide sequence alignments were made using the entire coding region of all the genes analysed. The Cerhsc1 nucleotide sequence shows 81% identity to the Drohsc4 gene, 70.3% identity to the Drohsc1 gene, 68.1% identity to the Drohsp2a and 71.2% identity to the Humhsp70 gene. The overall % identity at nucleotide level is lower than the overall % identity at the amino acid level. This is not surprising because of selection pressure at the amino acid sequence level and the influence of codon usage in the different genomes. The overall Cerhsc1 nucleotide sequence is most similar to the Drohsc4 gene. The Cerhsc1 and Drohsc4 nucleotide sequences overall are most similar to other cytoplasmic *hsc* genes such as Drohsc1 gene and Humhspa2a gene. In figure 23, a dendrogram was plotted based on average degree of nucleotide sequence similarity along the entire length of coding regions of all analysed *hsp70* genes.

In the dendrogram based on nucleotide sequences a new group of sequences appeared (figure 23) compared to dendrogram based on amino acid sequences (figure 22). This new group is comprised of sequences from the *hsp70* genes of lower eukaryotic organisms including Ddhsc70, Scssa1, Scssb1, and kar2 from the genomes of *Dictyostelium discoideum* and *S. cerevisiae*. The fact that *hsp70* genes with functions in different cellular compartments

could be more similar to one another compared to *hsp70* genes from the same cellular compartments suggests that the evolution of the *hsp70* multigene family in lower eukaryotic organisms might be different from what it is in higher eukaryotes.

An average similarity score plot for the *hsp70* nucleotide sequences is presented in figure 27. Here it is also clear that the 5' and 3' ends of the nucleotide sequence alignments are the most variable regions in this multigene family. This is consistent with the amino acid sequence similarity plot shown in figure 26.

5.3 The 5' upstream region of the Cerhsc1 gene

A 280 base pair region located between 650 to 930 nucleotides upstream of an ATG start codon of the Cerhsc1 gene appears to contain a promotor (figure 18). Three putative TATA boxes can be found at 656, 666, and 854 bases upstream from this start codon. Five putative *hses* are also located in this promotor region (figure 20). These promotor elements share a consensus with the sequence NGAAN alternating with NTTCN. This are characteristic of so-called *hse* often seen in the regulatory of such genes (Pelham 1985, Bienz and Pelham 1987, Xiao and Lis 1988, Perisic *et al.* 1989, Xiao *et al.* 1991). Among these *hses* are two "CAAT box" consensus sequences. Overall, this Cerhsc1 gene has a promotor region similar to other known *hsp70* genes. However, the number of *hse* in this promotor region is larger than what is typical for a *hsc* gene. In the promotor regions of the Drohsc4 gene

and a human *hsc* gene (Perkin *et al.* 1990, Dworniczak and Mirault 1987), for example, only two *hses* were found.

The larger number of *hses* are usually more characteristic of an inducible *hsp70* gene. Typically four to five *hses* are found in inducible *hsp70* genes in the *D. melanogaster* genome (Maroni 1993). In general, *hsc* genes were believed to be nonresponsive to heat induction because they did not have enough number of the *hse* to bind heat shock factors (HSF). The *hses* work cooperatively to bind HSF (Amin *et al.* 1986). Since the promotor region of the Cerhsc1 has at least five *hses*, it suggests that either this Cerhsc1 gene might function as a heat inducible *hsp70* gene or that other factors are involved in inducibility or noninducibility. It is known that a heat inducible *hsp70* gene from *D. melanogaster* has a different regulatory element responsible for heat induction (Corces *et al.* 1984). If the Cerhsc1 gene are not heat inducible, it could imply that the high number of *hse* is not sufficient for the heat induction. Therefore there might be another regulatory element required for heat inducibility. If the Cerhsc1 gene is heat inducible, it suggests that these *hses* are regulatory unit conferring heat induction. This would be a similar case to a heat shock cognate in maize. In maize that a *hsp70* gene with a highly conserved nucleotide sequence and intron position compared to the Drohsc1 gene is heat responsive, similar to the Drohsp70 genes (Rochester *et al.* 1986).

5.4 The 3' untranslated region of the Cerhsc1 gene

At least 3 putative polyadenylated sites (AATAAA) can be identified within a 150 base pair region immediately downstream of the stop codon TAA (figure 21). Although these polyadenylation sites appeared to be very close to the stop codon, it is possible that these are real. It is not uncommom for such polyadenylation sites to be close to a stop codon in *hsp70* genes. The *Drosophila hsp70* genes and the human *hsp70* genes also have polyadenylation sites located within 200 bases downstream of stop codons (Dworniczack and Mirault 1987, Maroni 1993). Further studies on 3' untranslated sequences from cDNA would be necessary to identify the exact point where polyadenylation occurs.

5.5 Representation of the *hsp70* gene family in the Medfly genome

A study of the *hsp70* gene family representation in the genome of the Medfly was carried out by using a 5' highly conserved portion of the Cerhsc1 gene as a probe for a genomic Southern blot. This 5' Cerhsc1 probe extended from amino acid position 34 to 224 and it includes a region that is normally highly conserved. This is about one third of the entire coding region. This region contains the conserved ATP binding domain for all known *hsp70* genes (Parsell & Lindquist 1993).

The Southern blot analysis shown in figures 28 and 29 include restriction digested genomic DNA from the Medfly

(*C. capitata*), the Oriental fruit fly (*Bactrocera dorsalis*), the Melon fly (*Bactrocera cucurbitae*) and *D. melanogaster* using the previously described Cerhsc1 probe. The genomic DNA digested with EcoRI shows different numbers of bands in the Medfly, the Oriental fruit fly, the Melon fly and *D. melanogaster* genomes (figure 29).

In the Medfly genome, there may be more than one *hsp70* gene. The 2.0 kb EcoRI fragment presumably represents the coding region of the Cerhsc1 gene (figure 29). The other two EcoRI fragments (2.2 kb and 2.5 kb) may represent different *hsp70* genes. This is suggested from a fact that it is only the 2.0 kb EcoRI fragment that correspond in size to the 6-5 EcoRI fragment of the Cerhsc1 gene (figure 17). Likewise, there may be a number of *hsp70* genes represented in the genomes of the Oriental fruit fly and the Melon fly.

The genomic Southern blot also suggests that the high degree of nucleotide sequence similarity among members in the *hsp70* multigene family may be restricted to related species. In *D. melanogaster*, there are 13 members of the *hsp70* gene family, but clearly not all of these were detected with this probe. This could be due to the fact that *D. melanogaster* has a very different codon usage pattern compared to tephritids such as the Medfly, the Melon fly, and the Oriental fruit fly (He and Haymer 1992).

The pattern of cross hybridization of the 5' Cerhsc1 probe in the *D.melanogaster* genome is strikingly different from the pattern

obtained using a Medfly actin probe with the *D. melanogaster* (data not shown). The Medfly actin probe recognised several members of its own multigene family in *D. melanogaster.* Although both the actin genes and the *hsp70* genes are considered to be highly conserved multigene families, the *hsp70* multigene family might be more complex in evolution. The *hsp70* genes are more diverse both in terms of the patterns of gene expression and in subcellular localization (Boorstein *et al.* 1994). The restricted cross hybridizing ability of the Cerhsc1 probe may also reflect other evolutionary phenomena in the *hsp70* gene family. Phenomena such as gene conversion, especially at conserved regions of the *hsp70* genes, might homogenize the nucleotide sequences among the *hsp70* genes.

Specific region known as signature boxes, two blocks of the highly conserved amino acid sequences "I D L G T T Y S" and "D L G G G T F D", are also found within most of the HSP70 proteins (LeJohn *et al.* 1994). The signature box "I D L G T T Y S" is usually located at the very begining of the amino terminal region of the HSP70 kDa proteins. In the case of the Cerhsc1 gene, the signature box "I D L G T T Y S" is found at amino acid position 9 (figure 19). The other signature box "D L G G G T F D" is usually located at approximately amino acid position 200. This signature box is located at amino acid position 198 of the Cerhsc1 gene (figure 19).

5.6 Expression of the Cerhsc1 gene

Expression of the Cerhsc1 gene was studied at the level of transcription. A number of different poly(A) selected mRNAs were prepared from different developmental stages of the Medfly including early embryos (3-4 hours), late embryos (30 hours), third instar larvae, and one day old adults. The effect of heavy metal exposure on the Cerhsc1 gene was studied by raising the Medfly larvae in a 5.0mM $CuSO_4$ diet. Poly(A) selected mRNA was also extracted from these copper treated larvae at the third instar. The heat response of the Cerhsc1 gene was also investigated using poly(A) selected mRNA from third instar larvae which had been heat shocked at 37°C for one hour. The heat shock derived poly(A) selected mRNAs were compared to poly(A) selected mRNAs derived from the third instar larvae raised at a normal temperature (22°C).

Northern blots representing these previously described poly(A) selected mRNA samples were prepared (figures 30, 31, and 32) and probed with different probes derived from various regions of the Cerhsc1 gene. An N-terminal probe consisted of a 771 base pair Acc I - Eco RI DNA fragment within the 3.28 kb Cerhsc1 DNA fragment. This probe begins approximately 40 bases upstream of the ATG start codon. This region also contains a highly conserved ATP binding domain (Flynn *et al.* 1991). The C-terminal probe was a 887 base pair EcoRI fragment derived from the 3.28 kb Cerhsc1 DNA (figure18). This C-terminal fragment

contains another conserved region which has been hypothesized to be a protein binding domain (Reppmann *et al.* 1991).

A medfly actin probe pmed21 (Haymer *et al.* 1990) was also used as a control to assure standard quality and quantity of all poly(A) selected mRNAs used in northern blots. Figure 32 shows that the pmed 21 probe hybridized well to messages 1.8 kb size as expected. However, trascription of the actin gene in the heat shocked larvae was absent. Lower levels of actin transcription were also evident in the heavy metal treated larvae. This is consistent with knowledge that heat shock and heavy metal exposure may exert cellular stress, and that under such stress even abundantly expressed genes such as actin are suppressed.

The Cerhsc1 gene appeared to detect messages approximately 2.5 kb in size. This size approximation was based on an interpolation from a semilog plot using the mobility of RNA markers of known size. Transcripts 2.5 kb in size are consistent with sizes of *hsp70* transcripts in general. Transcripts of this size are also in agreement with a 650 base pair transcribed but untranslated region including the promotor and the coding region of the Cerhsc1 gene. If this whole 650 base pair region is transcribed, the transcript size would be equal to 650 nucleotides plus 1,962 nucleotides of coding sequence, totaling approximately 2.5 kb. This also suggests that the Cerhsc1 gene does not have an introns interrupting in the 5' untranslated region, as is the case for the Drohsc4 gene (Perkins *et al.* 1990).

To verify that there is not even a small intron at the 5' untranslated region of the Cerhsc1 gene, cDNA sequence information would be necessary. The 2.5 kb size approximation here can not rule out a likelihood that there might be a small intron. Nonetheless, the possibility of not having any intron in the Cerhsc1 gene is also suggested by the fact that there are no consensus exon-intron junction sequences found in this region.

Expression of the Cerhsc1 gene appears to be most abundant in adult stage. Densitometer readings indicate that the Cerhsc1 mRNA is at least twice as abundant in the adult stage compared to the third instar larvae stage. Transcription of the Cerhsc1 gene was not detectable in both the early and late embryo stages.

The developmental expression of the Cerhsc1 gene appears to be different from that of the Drohsc4 gene. In *D. melanogaster*, the Drohsc4 transcriptional level appears to be equally abundant in all developmental stages (Craig *et al.* 1983). This includes the embryo, larvae, and adult stages.

Although the Cerhsc1 gene is most similar to the Drohsc4 gene in terms of nucleotide sequence and amino acid sequence, the pattern of gene expression is more similar to the Drohsc1 gene and the Drohsc2 gene. Transcripts from these gene are most abundant at the adult stage and are not detectable or exist at very lower level in the larvae stage and embryo stage (Craig *et al.* 1983). This suggests that there are differences in the structural and functional relationships of the Cerhsc1 and the Drohsc4 genes.

5.7 Heat shock responses of the Cerhsc1 gene

The Cerhsc1 gene appears to respond to heat induction producing a 10 to 20% increase in the level of transcription. This is based on densitometry readings from hybridization signals in a northern blot carried out using high stringency conditions (figure 30). The heat shock responsiveness of the Cerhsc1 gene may be consistent with the high number of *hses* found in its promotor region.

The presence of heavy metals appeared to suppress the expression of the Cerhsc1 gene (figure 30) as well as the expression of the actin gene (figure 32). Although it is known that certain *hsp70* gene respond to stresses generated by heavy metal exposure, it appears here that the Cerhsc1 gene did not respond. In the Medfly, it may be that only the metallothionein genes play a major role in response to heavy metals. Clearly, there might be more than one type of stress responsive multigene familiy in the Medfly genome. Further studies on different multigene families such as metallothioneins would enhance the understanding of stress responsiveness in the Medfly genome.

5.8 Expression of the other heat shock genes in the Medfly genome

Figure 31 shows that there also appear to be a number of smaller transcripts (<2.5 kb) which responded to heat induction. The heat induction of this group of small transcripts were elevated approximately 57% above the basal level of transcription. The

northern blot in figure 31 was done at a lower stringency condition compared to the northern blot in figure 30. This lower stringency condition could allow the N-terminal and C-terminal probes to cross hybridize to transcripts from other genes in the heat shock multigene family.

The smaller transcripts appeared to be detected only by the C-terminal probe. This can be seen by comparing the results of the similar lanes probed with the N-terminal region (figure 31). This additional hybridization using the C-terminal probe suggests that this region is more conserved than the N-terminal region among the heat shock genes in the Medfly genome. It also suggests that these other heat shock proteins and the Cerhsc1 protein may have similar protein binding domains (Reppmann *et al.* 1991). The heat shock genes that share similar protein binding domains may have overlapping functions in response to heat induction. Hence, it is possible that the heat shock response involves a number of heat shock genes. The C-terminal probe could be a molecular tool to isolate and characterize these other heat shock genes.

Previous work on the heat shock response at the protein level in the Medfly also demonstrated existence of small heat shock proteins (Stephanou *et al.* 1983). In that work, the small heat shock proteins were collectively referred to as "minor heat shock proteins" The minor heat shock proteins were 12 kDa, 13 kDa, 14 kDa, and 34 kDa in size. These minor heat shock proteins may represent translational products of the small heat inducible transcripts observed in this work.

Although some heat shock response in the Medfly can be attributed to the Cerhsc1 gene, it may also be associated with a minor heat shock gene family as well. To fully understand the heat shock response, it will be essential to characterize these minor heat shock genes as well as other *hsp70* genes. The N-terminal probe and the C-terminal probe from the Cerhsc1 gene offers opportunities to isolate and characterize these other heat shock genes and to better understand how the Medfly genome interacts with environmental stresses.

Table 1. Codon usage table of genes from the Medfly and D. melanogaster

Amino acid	Codon	Ceractin	Cerhsc1	Drohsc4	Drohsc1
Ala	GCG	2/0.07	3/0.05	5/0.09	17/0.30
Ala	GCA	4/0.13	4/0.07	2/0.03	9/0.16
Ala	GCT	12/0.40	32/0.57	18/0.31	8/0.14
Ala	GCC	12/0.40	17/0.30	33/0.57	23/0.40
Arg	AGG	0/0.00	0/0.00	0/0.00	5/0.12
Arg	AGA	0/0.00	0/0.00	0/0.00	1/0.02
Arg	CGG	0/0.00	0/0.00	0/0.00	9/0.21
Arg	CGA	1/0.05	0/0.00	0/0.00	7/0.16
Arg	CGT	12/0.63	22/0.76	18/0.60	7/0.16
Arg	CGC	6/0.31	7/0.24	12/0.40	14/0.33
Asn	AAT	4/0.44	14/0.42	3/0.09	14/0.45
Asn	AAC	5/0.56	19/0.58	29/0.91	17/0.55
Asp	GAT	9/0.41	27/0.61	21/0.44	23/0.53
Asp	GAC	13/0.59	17/0.39	27/0.56	20/0.47
Cys	TGT	2/0.33	2/0.40	0/0.00	3/0.75
Cys	TGC	4/0.67	3/0.60	5/1.00	1/0.25
Gln	CAG	8/0.62	10/0.45	23/1.00	28/0.85
Gln	CAA	5/0.39	12/0.55	0/0.00	5/0.15
Glu	GAG	17/0.61	22/0.44	45/0.96	37/0.77
Glu	GAA	11/0.39	28/0.56	2/0.04	11/0.23
Gly	GGG	1/0.03	1/0.02	0/0.00	0/0.00
Gly	GGA	0/0.00	2/0.04	7/0.13	9/0.18
Gly	GGT	18/0.62	42/0.76	24/0.45	12/0.24
Gly	GGC	10/0.35	10/0.18	22/0.42	29/0.58
His	CAT	3/0.38	2/0.25	2/0.25	7/0.78
His	CAC	5/0.63	6/0.75	6/0.75	2/0.22
Ile	ATA	2/0.07	0/0.00	0/0.00	4/0.10
Ile	ATT	12/0.42	23/0.51	9/0.21	10/0.26
Ile	ATC	14/0.50	22/0.49	34/0.79	25/0.64

Amino acid	Codon	Cer actin	Cerhsc4	Drohsc4	Drohsc1
Leu	TTG	13/0.46	34/0.71	4/0.08	5/0.10
Leu	TTA	1/0.035	1/0.02	0/0.00	1/0.02
Leu	CTG	6/0.21	4/0.08	37/0.74	27/0.56
Leu	CTA	2/0.07	5/0.10	1/0.02	1/0.02
Leu	CTT	1/0.04	2/0.04	1/0.02	5/0.10
Leu	CTC	5/0.17	2/0.04	7/0.14	9/0.19
Lys	AAG	15/0.79	35/0.69	50/0.98	27/0.82
Lys	AAA	4/0.21	16/0.31	1/0.02	6/0.18
Met	ATG	15/1.00	12/1.00	12/1.00	11/1.00
Phe	TTT	3/0.25	3/0.13	1/0.04	9/0.33
Phe	TTC	9/0.75	21/0.88	23/0.96	18/0.67
Pro	CCG	1/0.05	0/0.00	4/0.16	4/0.15
Pro	CCA	8/0.42	14/0.56	3/0.12	11/0.42
Pro	CCT	4/0.21	7/0.28	2/0.08	2/0.08
Pro	CCC	6/0.32	4/0.16	16/0.64	9/0.35
Ser	AGT	0/0.00	3/0.08	0/0.00	2/0.06
Ser	AGC	1/0.04	2/0.06	4/0.13	8/0.24
Ser	TCG	9/0.34	4/0.11	13/0.42	10/0.29
Ser	TCA	4/0.15	13/0.36	0/0.00	1/0.03
Ser	TCT	3/0.11	9/0.25	6/0.19	1/0.03
Ser	TCC	8/0.31	5/0.14	8/0.26	12/0.35
Thr	ACG	0/0.00	1/0.02	4/0.07	14/0.29
Thr	ACA	4/0.17	6/0.11	1/0.02	5/0.10
Thr	ACT	4/0.17	15/0.28	6/0.11	6/0.12
Thr	ACC	15/0.65	32/0.59	44/0.80	24/0.49
Trp	TGG	4/1.00	2/1.00	2/1.00	2/1.00
Tyr	TAT	9/0.56	4/0.29	2/0.14	6/0.46
Tyr	TAC	7/0.44	10/0.71	12/0.86	7/0.54
Val	GTG	13/0.56	7/0.17	26/0.65	25/0.61
Val	GTA	3/0.13	13/0.32	0/0.00	3/0.07
Val	GTT	3/0.13	11/0.28	3/0.08	3/0.07
Val	GTC	4/0.17	9/0.22	11/0.28	10/0.24
End	TGA	0/0.00	0/0.00	0/0.00	0/0.00

Table 2.
Amino acid composition (%) of the Medfly and D. melanogaster proteins

Amino acid	Ceractin	Cerhsc1	Drohsc4	Drohsc1
Ala	6.38	8.57	8.89	8.89
Arg	5.05	4.44	4.59	6.69
Asn	2.39	5.05	4.89	4.83
Asp	5.85	6.58	7.36	6.69
Cys	1.59	0.76	0.76	0.62
Gln	3.45	3.37	3.52	5.14
Glu	7.44	7.65	7.19	7.49
Gly	7.71	8.42	8.12	7.69
His	2.12	1.22	1.22	1.39
Ile	7.44	6.89	6.59	6.01
Leu	7.44	7.35	7.69	7.47
Lys	5.05	7.96	7.82	5.14
Met	3.99	1.83	1.84	1.71
Phe	3.19	3.67	3.68	4.19
Pro	5.05	3.83	3.83	4.05
Ser	6.64	5.51	4.75	5.29
Thr	6.11	8.72	8.43	7.46
Trp	1.06	0.29	0.29	0.31
Tyr	4.25	2.14	2.14	2.02
Val	6.11	6.12	6.13	6.38

Table 3.
Lists of heat shock 70 kDa genes and proteins analyzed in comparison to the Cerhsc1 gene.

HSP70 sequence	Organism	Accession #	Reference
Drohsc5a	*D. melanogaster*	L01502	Rubin et al. 1993
Ecolidnak	*E. coli*	P04475	Schmid et al. 1992
Cyanobac	*Synechocystis sp.*	M57518	Chitnis and Nelson 1991
Msqhsp70a1	*Anopheles albimanus*	M96662	Bebedict et al. 1993
Msqhsp70a2	*Anopheles albimanus*	M96662	Bebedict et al. 1993
Drohsp7a2	*D. melanogaster*	J01103	Ingolia et al. 1980
Drohsp7d1	*D. melanogaster*	J01104	Ingolia et al. 1980
Cerhsc1	*C. capitata*	-	-
Drohsc4a	*D. melanogaster*	L01500	Perkin et al. 1990
Humhspa2a	*Homo sapiens*	L26336	Yu et al. 1994
Celhsp70	*Caenorhabditis elegans*	M18540	Snutch et al. 1988
Ddhsc70	*Dictyostelium discoideum*	X75263	Haus et al. 1993
Scssa1	*S. cerevisiae*	X12926	Slater and Craig 1989
Maize	*Zea mays*	X03714, X03697	Rochester et al. 1986
Drohsc3a	*D. melanogaster*	L01498	Rubin et al. 1993
Tobacco	*Nicotiana tabacum*	X60057	Denecke et al. 1991
Kar2	*S. cerevisiae*	M25064	Rose et al. 1989
Scssb1	*S. cerevisiae*	X13713	Slater and Craig 1989

Figure 17. **Genome organization of cloned fragments containing the Cerhsc1 gene.**

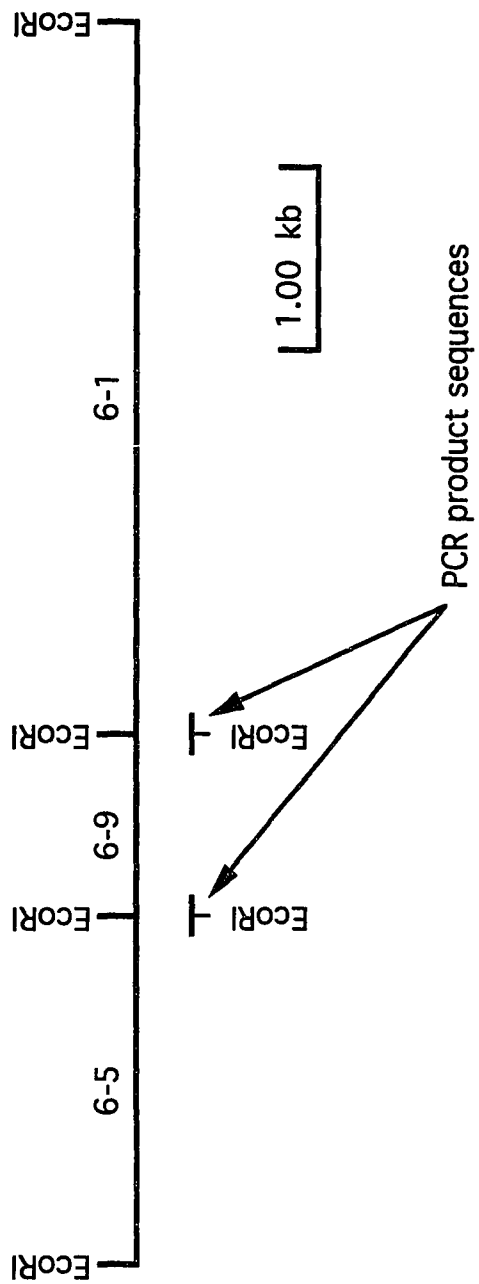

EcoRI
6-1
1.00 kb
PCR product sequences
EcoRI
EcoRI
6-9
EcoRI
EcoRI
6-5
EcoRI

Figure 18. **Schematic representation of a gene structure of the heat shock cognate1 gene (Cerhsc1) and corresponding restriction map.**

An oval shade designates a promotor region. A thick solid line designates a coding region, 1962 nucleotides long.

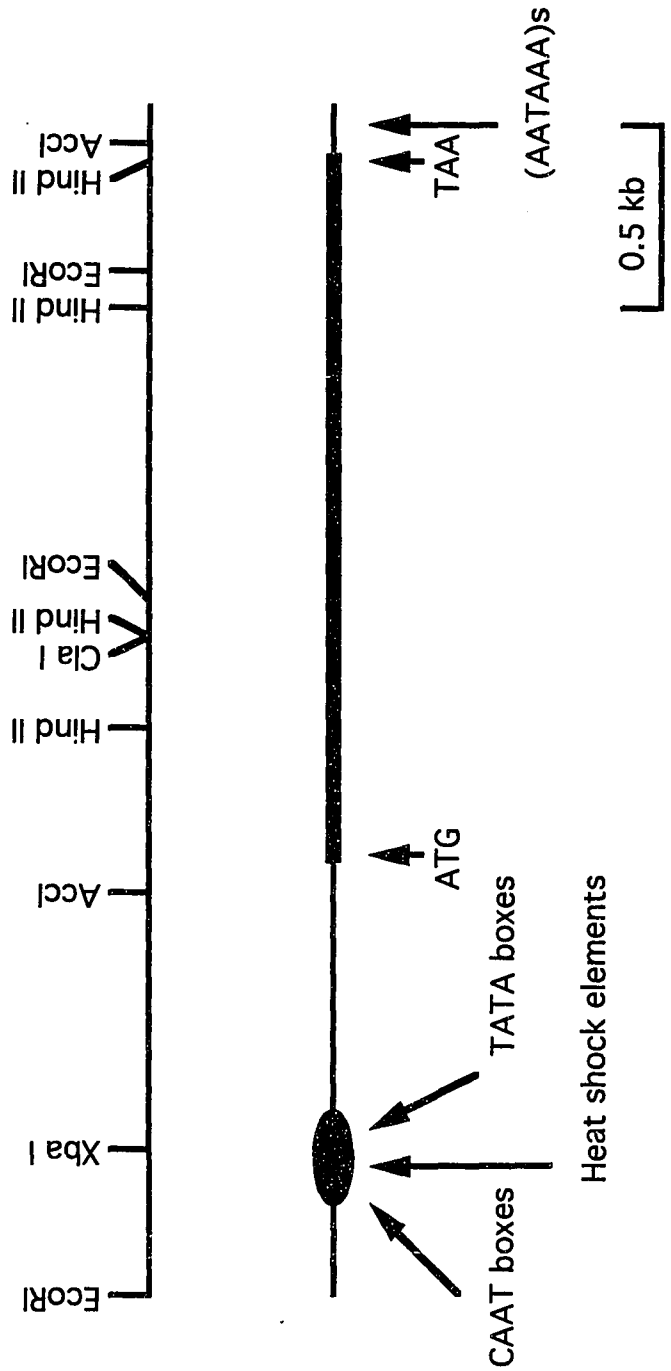

EcoRI
Xba I
AccI
Hind II
Cla I
Hind II
EcoRI
Hind II
EcoRI
Hind II
AccI
CAAT boxes
TATA boxes
Heat shock elements
ATG
TAA
(AATAAA)s
0.5 kb

Figure 19. **Nucleotide sequence and deduced amino acid sequence of the coding region of the Cerhsc1 gene.**

```
      M  S  K  A  P  A  V  G  I  D  L  G  T  T  Y  S  C  V  G  V
1     ATGTCTAAAGCTCCAGCTGTTGGTATTGATTTGGGTACTACCTACTCCTGTGTGGGTGTA     60

      F  Q  H  G  K  V  E  I  I  A  N  D  Q  G  N  R  T  T  P  S
61    TTCCAGCACGGCAAAGTTGAGATCATTGCCAATGACCAGGGTAATCGCACTACCCCATCA    120

      Y  V  A  F  T  E  T  E  R  L  I  G  D  A  A  K  N  Q  V  A
121   TATGTTGCCTTCACCGAGACAGAGCGTCTGATCGGCGATGCTGCCAAGAATCAGGTTGCC    180

      M  N  P  T  N  T  I  F  D  A  K  R  L  I  G  R  K  F  D  D
181   ATGAACCCTACAAACACCATTTTCGATGCCAAGCGTTTGATTGGTCGTAAATTCGATGAT    240

      A  N  V  Q  S  D  M  K  H  W  P  F  E  V  V  S  A  D  G  K
241   GCTAATGTACAATCTGATATGAAGCATTGGCCATTCGAAGTGGTTAGTGCTGATGGCAAA    300

      P  K  I  S  V  S  Y  K  D  E  K  K  T  F  F  P  E  E  I  S
301   CCAAAGATCAGTGTCTCCTACAAAGATGAAAAGAAAACTTTCTTCCCAGAGGAGATCTCA    360

      S  M  V  L  T  K  M  K  E  T  A  E  A  Y  L  G  K  T  V  T
361   TCTATGGTGTTGACTAAGATGAAGGAAACCGCTGAAGCCTACTTGGGCAAGACTGTTACC    420

      N  A  V  I  T  V  P  A  Y  F  N  D  S  Q  R  Q  A  T  K  D
421   AACGCCGTCATCACTGTACCCGCATACTTCAACGATTCTCAACGTCAAGCTACCAAGGAT    480

      A  G  T  I  A  G  L  N  V  L  R  I  I  N  E  P  T  A  A  A
481   GCCGGTACCATTGCTGGTTTAAATGTATTGCGTATTATTAACGAGCCTACCGCTGCTGCT    540

      I  A  Y  G  L  D  K  K  A  V  G  E  R  N  V  L  I  F  D  L
541   ATTGCGTACGGTTTGGACAAGAAAGCTGTTGGTGAACGCAATGTATTGATCTTCGATTTG    600

      G  G  G  T  F  D  V  S  I  L  S  I  D  D  G  I  F  E  V  K
601   GGTGGTGGTACTTTTGATGTGTCCATTCTATCAATCGATGATGGTATCTTTGAGGTCAAG    660

      S  T  A  G  D  T  H  L  G  G  E  D  F  D  N  R  L  V  T  H
661   TCAACCGCTGGTGACACTCACTTGGGTGGTGAAGATTTCGACAATCGTCTAGTCACCCAC    720

      F  V  Q  E  F  K  R  K  H  K  K  D  L  T  T  N  K  R  A  L
721   TTCGTTCAGGAATTCAAGCGCAAGCACAAGAAGGACCTCACCACCAACAAGCGTGCCCTA    780

      R  G  L  R  T  A  C  E  R  A  K  R  T  L  S  S  S  T  Q  A
781   CGTGGTTTGCGCACAGCTTGTGAACGCGCGAAGCGTACGCTATCTTCATCCACTCAGGCC    840

      S  I  E  I  D  S  L  F  E  G  V  D  F  Y  T  S  I  T  R  A
841   AGCATTGAAATTGACTCATTGTTCGAAGGTGTCGATTTCTACACCTCAATCACACGTGCC    900

      R  F  E  E  L  N  A  D  L  F  R  S  T  M  D  P  V  E  K  A
901   CGTTTCGAGGAGTTGAACGCTGACTTGTTCCGTAGCACCATGGACCCTGTGGAGAAGGCA    960

      I  R  D  A  K  L  D  K  S  A  I  H  D  I  V  L  V  G  G  S
961   ATTCGTGATGCTAAATTGGACAAATCTGCTATTCACGATATTGTGTTGGTTGGTGGTTCA   1020

      T  R  I  P  K  V  Q  R  L  L  Q  D  L  F  N  G  K  E  L  N
1021  ACCCGTATCCCTAAGGTACAACGCTTGTTGCAGGATTTGTTCAACGGCAAAGAGCTGAAC   1080

      K  S  I  N  P  D  E  A  V  A  Y  G  A  A  V  Q  A  A  I  L
1081  AAATCAATCAATCCCGATGAAGCTGTAGCTTACGGTGCTGCCGTCCAAGCTGCTATCTTG   1140

      H  G  D  K  S  Q  E  V  Q  D  L  L  L  L  D  V  T  P  L  S
1141  CATGGTGATAAGTCGCAGGAAGTACAAGACTTGTTGCTTCTCGATGTTACTCCCCTGTCA   1200
```

```
     L  G  I  E  T  A  G  G  V  M  S  V  L  I  K  R  N  T  T  I
1201 TTGGGTATTGAAACCGCTGGCGGCGTGATGAGTGTCTTGATCAAGCGTAACACAACCATT 1260

     P  T  K  Q  T  Q  T  F  T  T  Y  S  D  N  Q  P  G  V  L  I
1261 CCAACCAAGCAGACTCAGACATTCACTACCTACTCAGATAATCAACCTGGTGTATTGATT 1320

     Q  V  Y  E  G  E  R  A  M  T  K  D  N  N  L  L  G  K  F  E
1321 CAAGTATATGAAGGTGAACGTGCAATGACCAAGGACAACAATCTGCTTGGAAAATTCGAG 1380

     L  S  G  I  P  P  A  P  R  G  V  P  Q  I  E  V  T  F  D  I
1381 TTGTCTGGTATTCCACCAGCACCACGTGGTGTACCACAAATTGAAGTTACCTTCGACATT 1440

     D  A  N  G  I  L  N  V  T  A  L  E  R  S  T  N  K  E  N  K
1441 GATGCCAACGGTATCTTGAACGTAACCGCTTTGGAACGTTCCACCAACAAGGAAAACAAG 1500

     I  T  I  T  N  D  K  G  R  L  S  K  E  D  I  E  R  M  V  N
1501 ATTACCATCACTAACGACAAGGGTCGTCTATCGAAGGAAGACATCGAACGTATGGTCAAC 1560

     E  A  E  K  Y  R  S  E  D  E  K  Q  K  E  T  I  A  A  K  N
1561 GAAGCCGAGAAATACCGTTCGGAAGATGAGAAGCAAAAGGAGACCATTGCTGCCAAGAAT 1620

     S  L  E  S  Y  C  F  N  M  K  A  T  L  D  E  E  N  M  K  T
1621 TCGTTGGAATCATACTGCTTTAACATGAAGGCCACTTTGGATGAGGAGAATATGAAGACC 1680

     K  I  S  E  S  D  R  T  T  I  L  D  K  C  N  E  T  I  K  W
1681 AAGATCTCTGAGTCAGACCGCACCACCATTTTGGACAAGTGCAATGAAACCATCAAATGG 1740

     L  D  A  N  Q  L  A  E  K  E  E  Y  E  H  R  Q  K  E  L  E
1741 TTGGATGCTAACCAGTTGGCTGAGAAAGAGGAATATGAGCACCGTCAAAAGGAATTGGAA 1800

     S  V  C  N  P  I  I  T  K  L  Y  Q  G  A  G  G  A  P  G  G
1801 TCTGTATGCAATCCAATCATCACTAAATTGTATCAAGGTGCTGGTGGTGCTCCAGGTGGT 1860

     M  P  G  G  I  P  G  G  F  P  G  A  G  G  A  P  G  A  G  G
1861 ATGCCCGGTGGCATCCCTGGTGGTTTCCCAGGAGCTGGTGGGGCCCCTGGTGCGGGTGGT 1920

     A  G  T  G  A  G  P  T  I  E  E  V  D  *
1921 GCTGGCACCGGTGCTGGCCCAACCATCGAAGAGGTCGACTAA 1962
```

Figure 20. **Nucleotide sequence at the promotor and 5' untranslated region of the Cerhsc1 gene.**

Five putative heat shock elements are designated in rectangular boxes. Three putative TATA boxes are designated in ovals. Two putative CAAT boxes are underlined. An ATG start codon is also underlined at nucleotide position 1,174.

First nt.

+1	GAATTCCCCG	GATCAGTTCA	TTTAGAAAGT	GGGGTTTTGT	TAGCTATTTC
	GCAATGTACA	CTTTAATCCT	TTAATAAAAT	AGTTTCTTAA	TTAGGAATAT
+101	GTGAGAAAAT	GAGCTTCGGT	GTTGCCTCGT	GCAGTGGTCT	GATATACATA
	TGTATATATA	TTCTCCATAT	TTTTACTTTG	TTAGCATATT	GTTTATTTTT
+201	AGCACCGTTT	CAAACTTAAT	GCTTACACAT	GTGTGTATGT	GAATAACGCA
	AGTTCATGGC	GTGAACGGTG	TTACCTAGAA	GATTCCAACT	CGGGTATAGC
+301	CTACACACAG	ATCAGGAATT	GCACTGTGTA	TTGAAAGTCC	AACGTTGCGT
	GGTTTTCTTC	CATTTCAATA	GGCTATATAA	ATGTGAAATG	ATCTTATCTT
+401	TATGATTTCT	AGAAAACAAT	AGATTTGTGA	AATTTAGAAA	TTCAGCATTG
	GTTTTCTTTT	CAACTTCCTA	ATTAGGTGAA	TTTATGTTTG	GCGATAACAT
+501	GAAGTTATAA	TAAGTCTATA	TTAATTTCTC	TCCCAATGAA	ATGAAATCGT
	GTTAAAGAAC	TCATTTTGGA	AATCTAAATG	CATATTTGCT	CCCGCTTTTT
+601	GAACGTATGT	ACGTCCGTTT	CGGAAGAACG	TACAGTACGT	CAAGTATAAT
	AAAAAGTGTT	CTAAAACTAT	TCTTATCATT	TTAATGAAAT	ACGAATTTGA
+701	TTTGCAGTTT	ATAATAAACC	GCTCGCACAT	TACTTTTAAA	GATGATTGCC
	CTGCTAAAGT	ATATCAGTTT	TTGGTTCGTT	CGTTTTATAT	TTTAAACTGG
+801	TTTTATTTTT	AATCAATATA	TGGTATTAAC	CCAAGTCATG	ACTCATTTAT
	TCTTATGAGA	AGTGACGAGT	CTCTGAAATA	TCTTACTTCA	TGTTTTACCT
+901	TATTAGTTAT	TGGTCGAAGA	ATGTATAGTG	ATATACTCGT	AAATTAGTTA
	TTTTTACAAT	TTATAACAGT	TCTTTATGCG	TTGTTGAGTT	TTCTTTTATA
+1001	TTATGAATTT	GATTTTTTTT	CTTTAATTAT	TTGGTATTTT	TCCAAAATCT
	TATAACGAAT	AGAAATTTTA	TCATTCTATC	ATTTTTCCCT	CTTCCTCTTG
+1101	CACCATACAA	TAAAATACAA	GTAACATGAA	TGTAGACACA	TCTATAATAA
	AATTATTTAT	TCTCTAATTT	CAGATGTCTA	AAGCTCCAGC	TGTTGGTATT

Figure 21. **Nucleotide sequence at the 3' end of the Cerhsc1 gene.**

A TAA stop codon is underlined at nucleotide position 3,133. Three putative polyadenylation sites are designated in ovals.

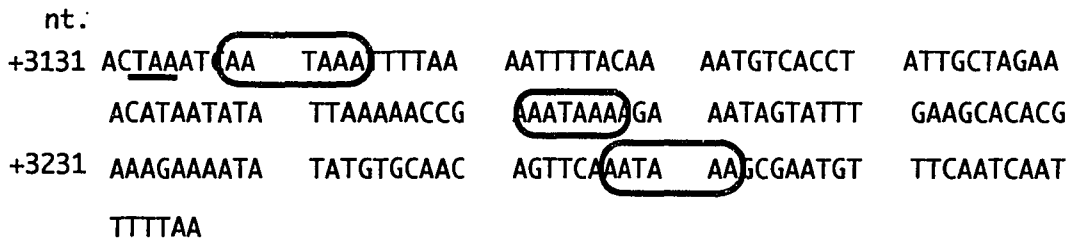

nt.
+3131 ACTAAATCAA TAAATTTTAA AATTTTACAA AATGTCACCT ATTGCTAGAA
ACATAATATA TTAAAAACCG AAATAAAAGA AATAGTATTT GAAGCACACG
+3231 AAAGAAAATA TATGTGCAAC AGTTCAAATA AAGCGAATGT TTCAATCAAT
TTTTAA

Figure 22. **Dendrogram based on amino acids sequence similarity among HSP70 proteins listed.**

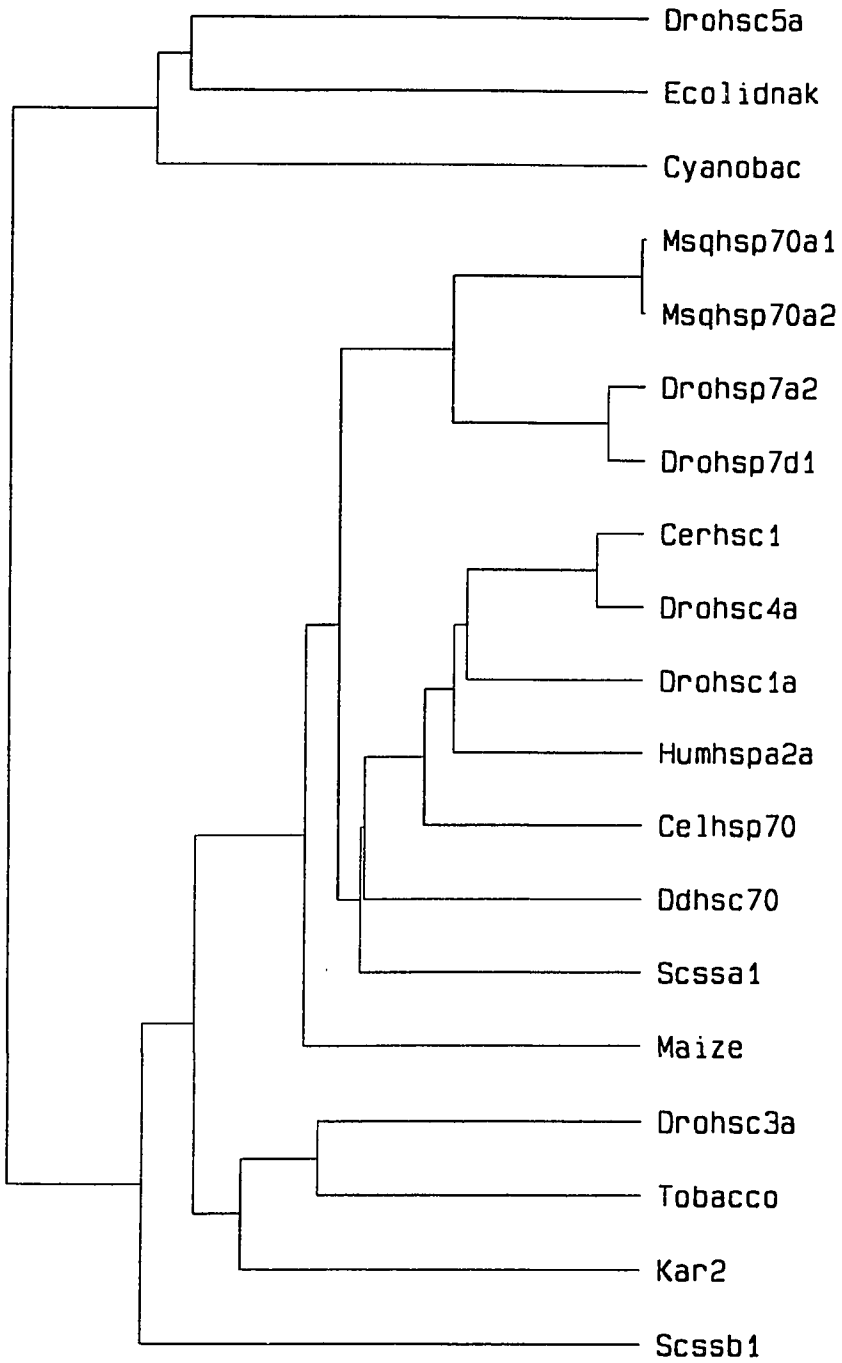

Drohsc5a
Ecolidnak
Cyanobac
Msqhsp70a1
Msqhsp70a2
Drohsp7a2
Drohsp7d1
Cerhsc1
Drohsc4a
Drohsc1a
Humhspa2a
Celhsp70
Ddhsc70
Scssa1
Maize
Drohsc3a
Tobacco
Kar2
Scssb1

Figure 23. **Dendrogram based on nucleotide sequence similarity from the entire coding regions of the *hsp70* genes.**

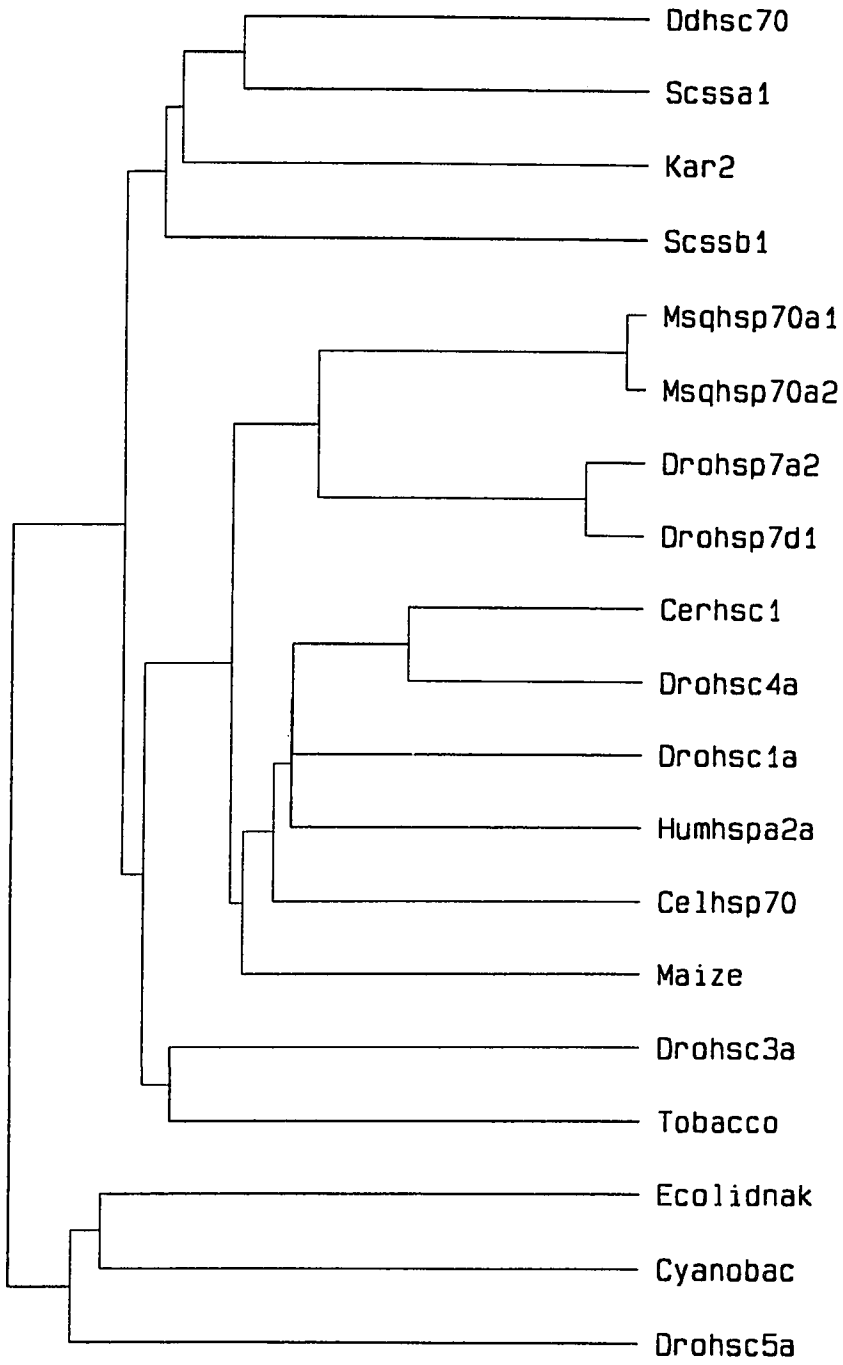

Ddhsc70
Scssa1
Kar2
Scssb1
Msqhsp70a1
Msqhsp70a2
Drohsp7a2
Drohsp7d1
Cerhsc1
Drohsc4a
Drohsc1a
Humhspa2a
Celhsp70
Maize
Drohsc3a
Tobacco
Ecolidnak
Cyanobac
Drohsc5a

Figure 24. **Multiple amino acid sequence alignments of HSP70 proteins at the N-terminal end.**

```
             1                                                  50
Drohsc5a     MLRVPKFLPR LARQAGVVPS HMSGASSMFR NLPGASNGIS SQLRYKSGEV
Ecolidnak    .......... .......... .......... .......... ..........
Cyanobac     .......... .......... .......... .......... ..........
Msqhsp70a1   .......... .......... .......... .......... ..........
Msqhsp70a2   .......... .......... .......... .......... ..........
Drohsp7a2    .......... .......... .......... .......... ..........
Drohsp7d1    .......... .......... .......... .......... ..........
Cerhsc1      .......... .......... .......... .......... ........MS
Drohsc4a     .......... .......... .......... .......... ........MS
Drohsc1a     .......... .......... .......... .......... ........MP
Humhspa2a    .......... .......... .......... .......... .......MSA
Celhsp70     .......... .......... .......... .......... ........MS
Ddhsc70      .......... .......... .......... .......... ..........
Scssa1       .......... .......... .......... .......... ..........
Maize        .......... .......... .......... .......... ......MAKG
Drohsc3a     .......... MKLCILLAVV AFVGLS.... .........L GEEKKEKDKE
Tobacco      ..MAGGAWNR RTSLIVFGIV LFGCLF.... .........A FSIATEEATK
Kar2         ..MFFNRLSA GKLLVPLSVV LYALFVVILP LQNSFHSSNV LVRGADDVEN
Scssb1       .......... .......... .......... .......... .....MAEGV

             51                                                 100
Drohsc5a     KGAVIGIDLG TTNSCLAVME GKQAKVIENA EGARTTPSHV AFTKDGERLV
Ecolidnak    .GKIIGIDLG TTNSCVAIMD GTTPRVLENA EGDRTTPSII AYTQDGETLV
Cyanobac     MGKVVGIDLG TTNSCVAVME GGKPTVIANA EGFRTTPSVV GYAKNGDRLV
Msqhsp70a1   MPSAIGIDLG TTYSCVGVFQ HGKVEIIAND QGNRTTPSYV AFS.DTERLI
Msqhsp70a2   MPSAIGIDLG TTYSCVGVFQ HGKVEIIAND QGNRTTPSYV AFS.DTERLI
Drohsp7a2    .MPAIGIDLG TTYSCVGVYQ HGKVEIIAND QGNRTTPSYV AFT.DSERLI
Drohsp7d1    .MPAIGIDLG TTYSCVGVYQ HGKVEINAYD QGNRTTPSYV AFT.DSERLN
Cerhsc1      KAPAVGIDLG TTYSCVGVFQ HGKVEIIAND QGNRTTPSYV AFT.ETERLI
Drohsc4a     KAPAVGIDLG TTYSCVGVFQ HGKVEIIAND QGNRTTPSYV AFT.DTERLI
Drohsc1a     KLPAVGIDLG TTYSCVGVFQ HGKVEIIAND QGNRTTPSYV AFT.ESERLI
Humhspa2a    RGPAIGIDLG TTYSCVGVFQ HGKVEIIAND QGNRTTPSYV AFT.DTERLI
Celhsp70     KHNAVGIDLG TTYSCVGVFM HGKVEIIAND QGNRTTPSYV AFT.DTERLI
Ddhsc70      .MSSIGIDLG TTYSCVGVWQ NDRVEIIAND QGNRTTPSYV AFT.DTERLI
Scssa1       MSKAVGIDLG TTYSCVAHFA NDRVDIIAND QGNRTTPSFV AFT.DTERLI
Maize        EGPAIGIDLE TTYSCVGVWQ HDRVEIIAND XGNRTTPSYV AFT.DTERLI
Drohsc3a     LGTVIGIDLG TTYSCVGVYK NGRVEIIAND QGNRITPSYV AFTADGERLI
Tobacco      LGTVIGIDLG TTYSCVGVYK NGHVEIIAND QGNRITPSWV AFT.DGERLI
Kar2         YGTVIGIDLG TTYSCVAVMK NGKTEILANE QGNRITPSYV AFT.DDERLI
Scssb1       FQGAIGIDLG TTYSCVATYE .SSVEIIANE QGNRVTPSFV AFTPE.ERLI
```

Figure 25. **Multiple amino acid sequence alignments of HSP70 proteins at the C-terminal end.**

```
            651                                                    700
  Drohsc5a  TADLEEVRKA TSSLQQSSLK LFELAYKKMS AE.....RET NAGAGSSDSS
 Ecolidnak  KAAIEAKMQE ...LAQVSQK LMEIAQQQHA QQ.....QTA GADASANNAK
  Cyanobac  DAKIQTVMPE ...LQQVLYS IGSNMYQQAG AEAGVGAPGA GPEAGTSSGG
Msqhsp70a1  MAEKEEYEHQ MQELSRVCSP IMTKLHQQAA G...GP.... ....QPTSCG
Msqhsp70a2  MAEKEEYEHQ MQELSRVCSP IMTKLHQQAA G...GP.... ....QPTSCG
 Drohsp7a2  TAEKEEFDHK LEELTRHCSP IMTKMHQQGA GAGAGG.... ....PGANCG
 Drohsp7d1  TAEKEEFDHK MEELTRHCSP IMTKMHQQGA GA.AGG.... ....PGANCG
   Cerhsc1  LAEKEEYEHR QKELESVCNP IITKLYQGAG GAPGGMPGGI PGGFPGAGGA
  Drohsc4a  LADKEEYEHR QKELEGVCNP IITKLYQGAG FPPGGMPGG. PGGMPGAAGA
  Drohsc1a  LAERQEFEHK QQELERICSP IITRLYQGAG MAPPPTAGG. ..........
 Humhspa2a  MAEKDEYEHK QKELERVCNP IISKLYQG.. .......... .....GPGGG
  Celhsp70  TAEKEEFESQ QKDLEGLAKP DLSKLYQSAG GAPP...... ........GA
   Ddhsc70  TAEKDEYEDK MKALEAVVNP IMSKLYQEGG MPQG...... .....GGMPG
    Scssa1  TASKEEFDDK LKELQDIANP IMSKLYQ.AG GAPGGAAGGA PGGFPG....
     Maize  LAEVEEFEDK MKELEGICNP IIAKMYXGEG AGMGA..... .....AAGMD
  Drohsc3a  DADPEEYKKQ KKDLEAIVQP VIAKLYQGAG GAPPPE.GGD DADLKDEL*.
   Tobacco  SAEKEDYEEK LKEVEAVCNP IITAVYQKSG GAPGGESGAS EDDDHDEL*.
      Kar2  TAIAEDFDEK FESLSKVAYP ITSKLYGGAD GSGAADYDDE DEDDDGDYFE
    Scssb1  SA..DELRKA EVGLKRVVTK AMSSR*.... .......... ..........

            701              719
  Drohsc5a  SSSDTSASQE GREELN*..
 Ecolidnak  DDDVVDAEFE EVKDKK...
  Cyanobac  GDDVIDAEFS EPEK*....
Msqhsp70a1  QQAGGFGGRT GPTVEEVD*
Msqhsp70a2  QQAGGFGGRT GPTVEEVD*
 Drohsp7a2  QQAGGFGGYS GRTVEEVD*
 Drohsp7d1  QQAGGFGGYS GPTVEEVD*
   Cerhsc1  PGAGGAGTGA GPTIEEVD.
  Drohsc4a  AGAAGAG.GA GPTIEEVD*
  Drohsc1a  .SNPGATGGS GPTIEEVD*
 Humhspa2a  SGGGGSGASG GPTIEEVD*
  Celhsp70  APGGAAGGAG GPTIEEVD*
   Ddhsc70  GMSNDSPKSS NNKVDELD*
    Scssa1  .GAPPAPEAE GPTVEEVD*
     Maize  EDAPSGGSGA GPKIEEVD.
  Drohsc3a  .......... .........
   Tobacco  .......... .........
      Kar2  HDEL*..... .........
    Scssb1  .......... .........
```

Figure 26. **Average amino acid similarity scores plotted along the length of aligned amino acid sequences for HSP70 proteins.**

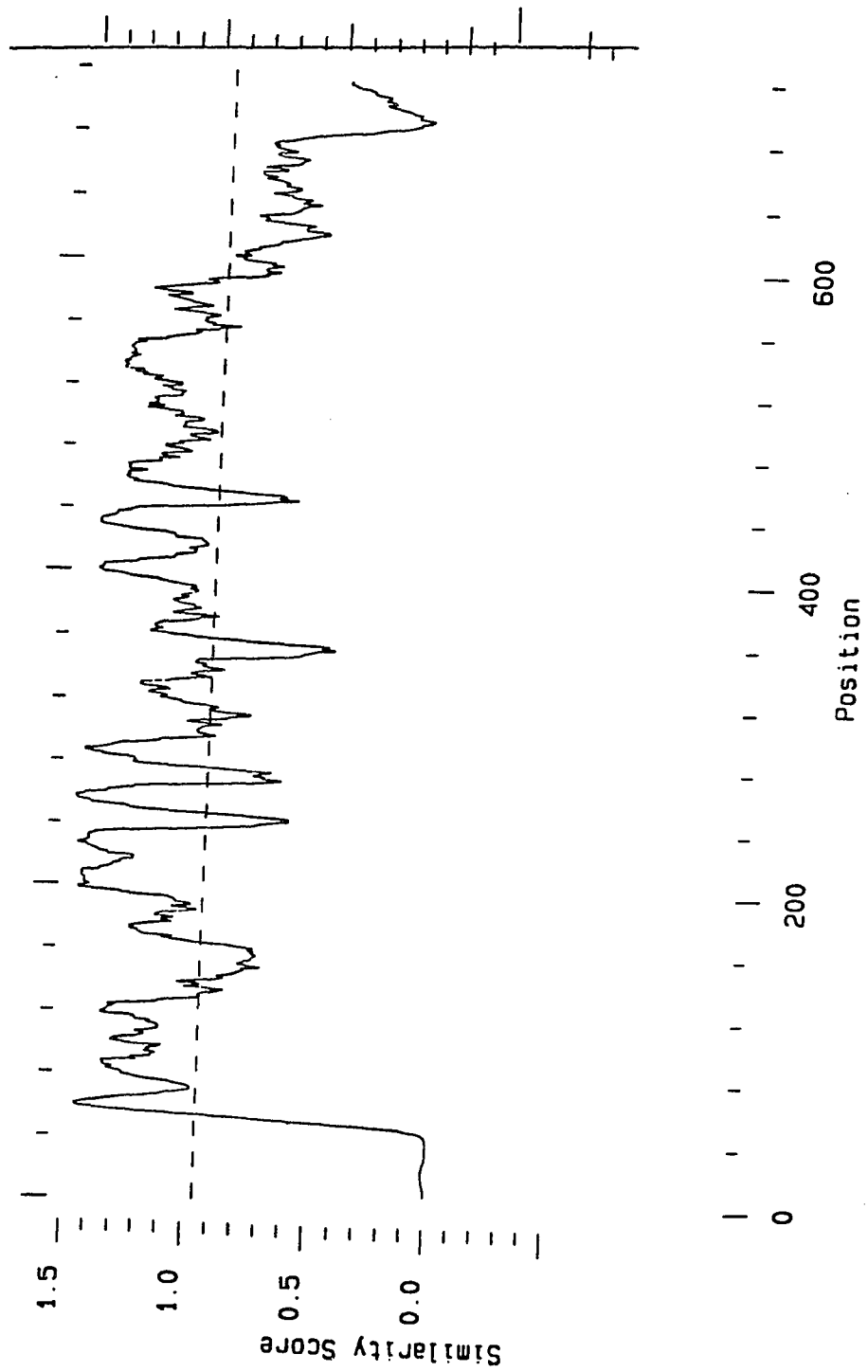

Similarity Score
1.5
1.0
0.5
0.0
0
200
400
600
Position

Figure 27. **Nucleotide sequence similarity scores plotted along the length of aligned *hsp70* coding sequences.**

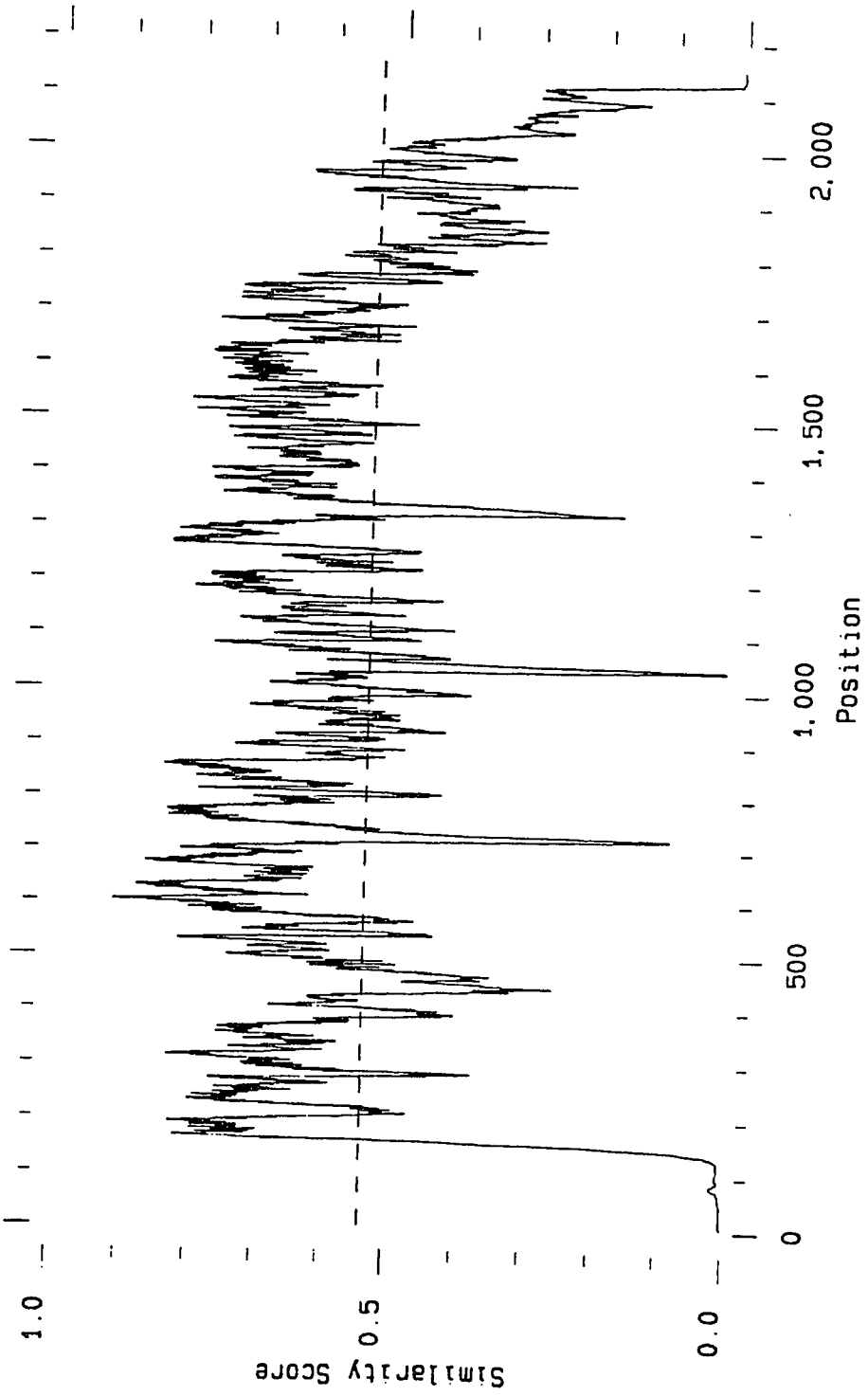

Similarity Score
1.0
0.5
0.0
0
500
1,000
1,500
2,000
Position

Figure 28. **Genomic representation of a *hsp70* gene family in the Medfly genome.**

This Southern blot was probed with a 5' coding region of the Cerhsc1 gene under high stringency conditions.

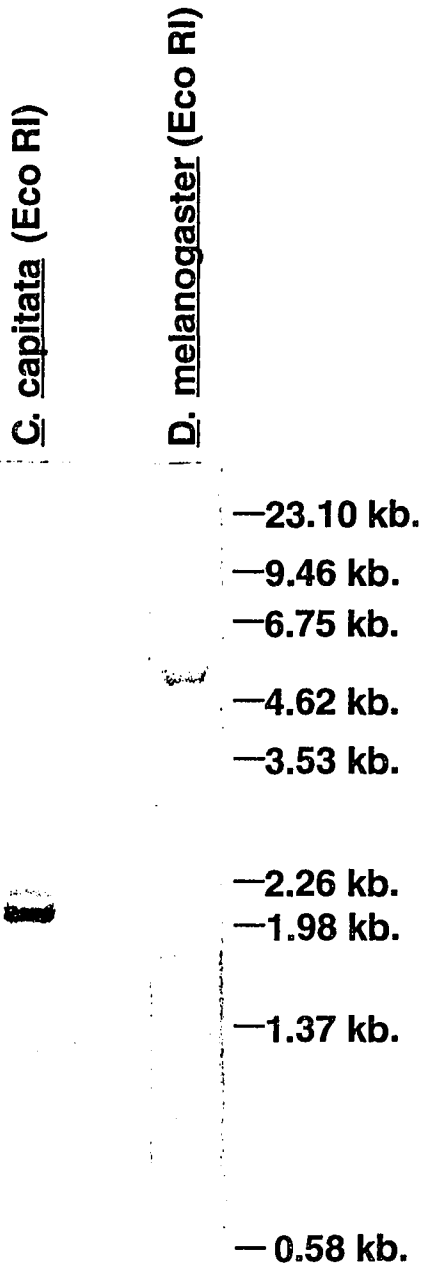

C. capitata (Eco RI)
D. melanogaster (Eco RI)
—23.10 kb.
—9.46 kb.
—6.75 kb.
—4.62 kb.
—3.53 kb.
—2.26 kb.
—1.98 kb.
—1.37 kb.
— 0.58 kb.

Figure 29. **Genomic representation of a heat shock 70 kDa gene family in the genomes of tephritid insects.**

This Southern bot was probed with a 5' coding region of the Cerhsc1 gene using high stringency conditions.

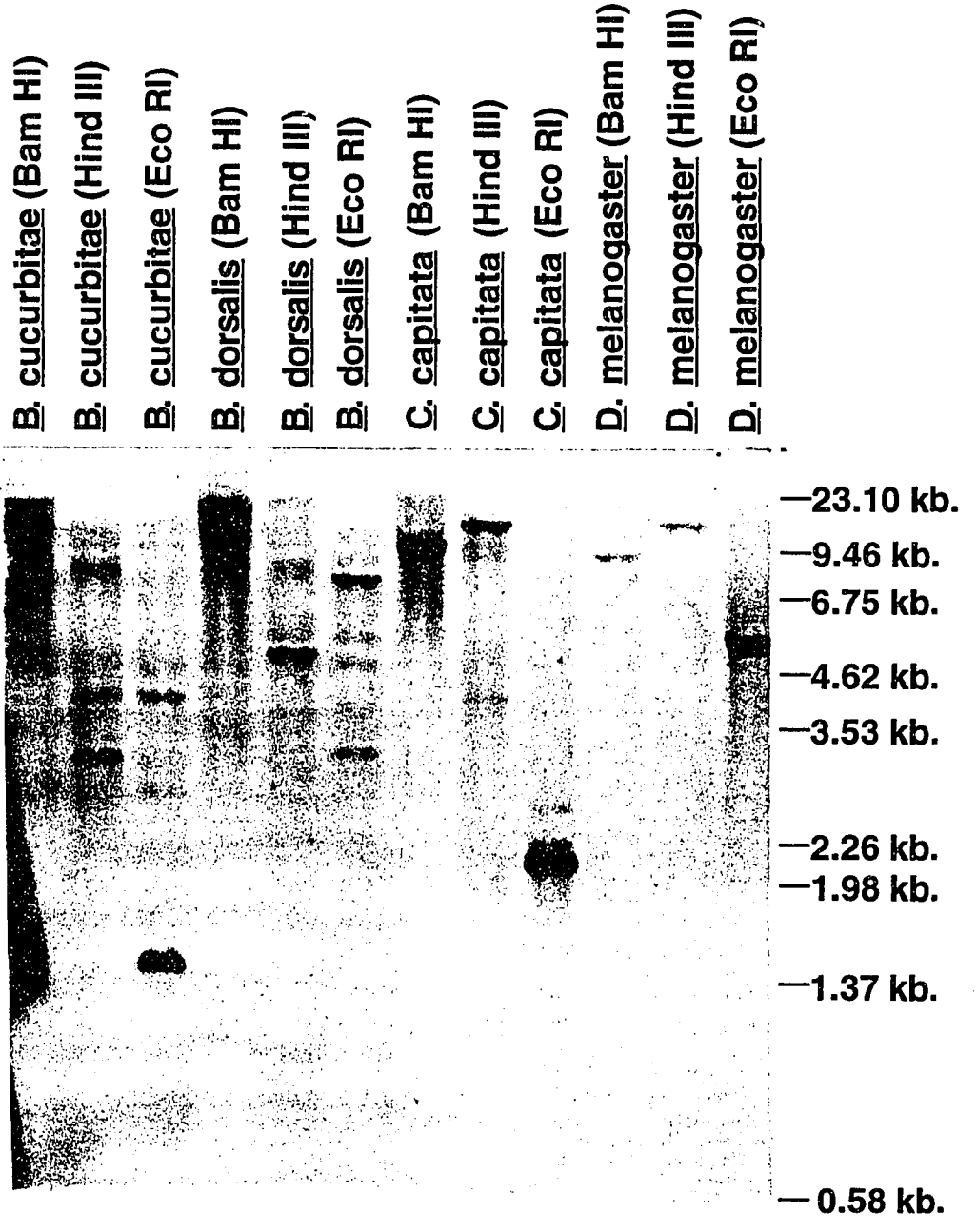

B. cucurbitae (Bam HI)
B. cucurbitae (Hind III)
B. cucurbitae (Eco RI)
B. dorsalis (Bam HI)
B. dorsalis (Hind III)
B. dorsalis (Eco RI)
C. capitata (Bam HI)
C. capitata (Hind III)
C. capitata (Eco RI)
D. melanogaster (Bam HI)
D. melanogaster (Hind III)
D. melanogaster (Eco RI)
—23.10 kb.
—9.46 kb.
—6.75 kb.
—4.62 kb.
—3.53 kb.
—2.26 kb.
—1.98 kb.
—1.37 kb.
— 0.58 kb.

Figure 30. **Developmental expression and stress responses of the Cerhsc1 gene.**

The first six lanes (left panel) were probed with an N-terminal domain portion of the Cerhsc1 gene. The second four lanes (right panel) were probed with a C-terminal domain portion of the Cerhsc1 gene. Both hybridizations were done under high stringency conditions.

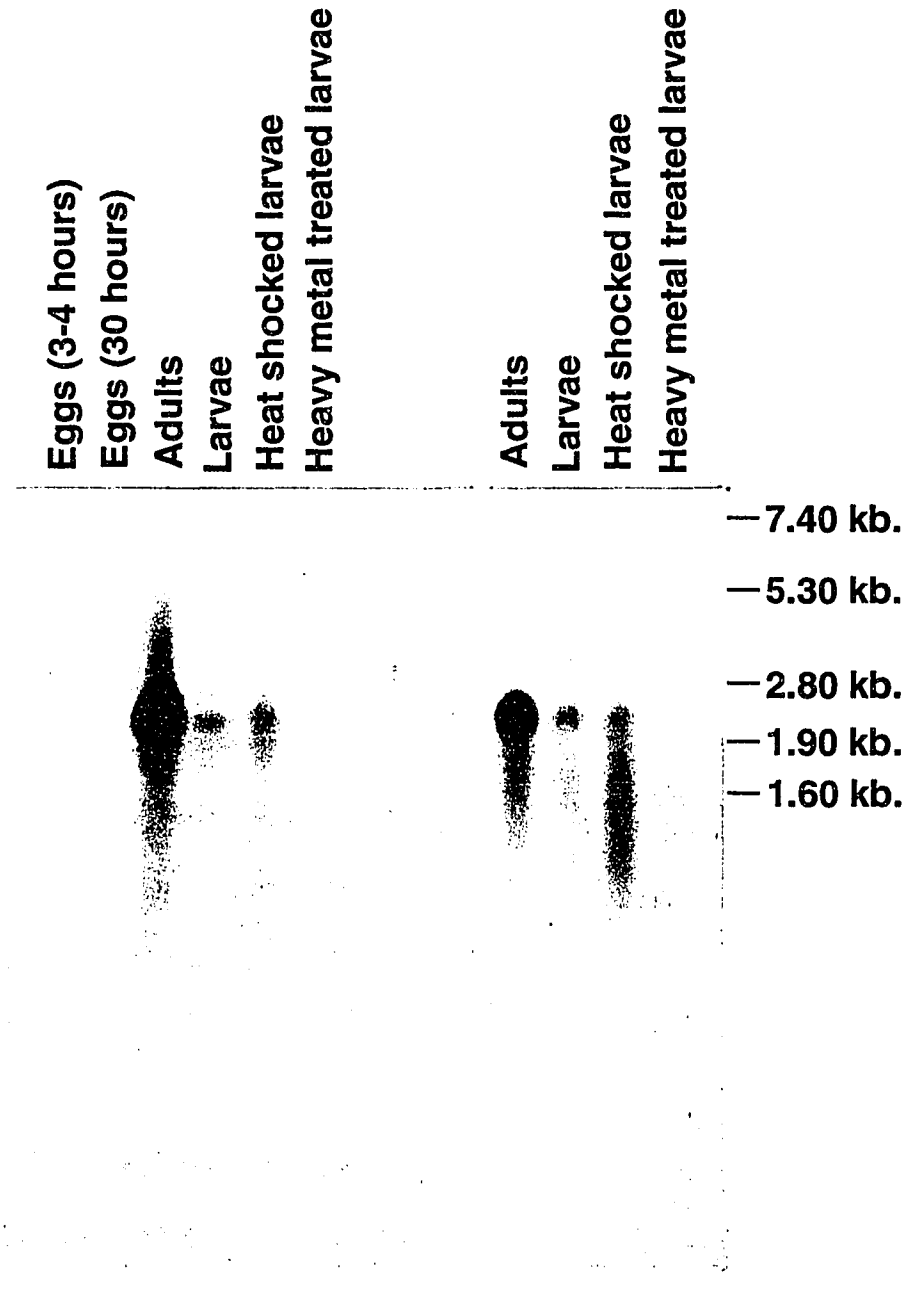

Eggs (3-4 hours)
Eggs (30 hours)
Adults
Larvae
Heat shocked larvae
Heavy metal treated larvae
Adults
Larvae
Heat shocked larvae
Heavy metal treated larvae
7.40 kb.
5.30 kb.
2.80 kb.
1.90 kb.
1.60 kb.

Figure 31. **Developmental expression and stress responses of the Cerhsc1 gene and other potential genes within heat shock gene families.**

The first six lanes (left panel) were probed with an N-terminal domain portion of the Cerhsc1 gene. The second four lanes (right panel) were probed with a C-terminal domain portion of the Cerhsc1 gene. Both hybridizations were done using low stringency conditions.

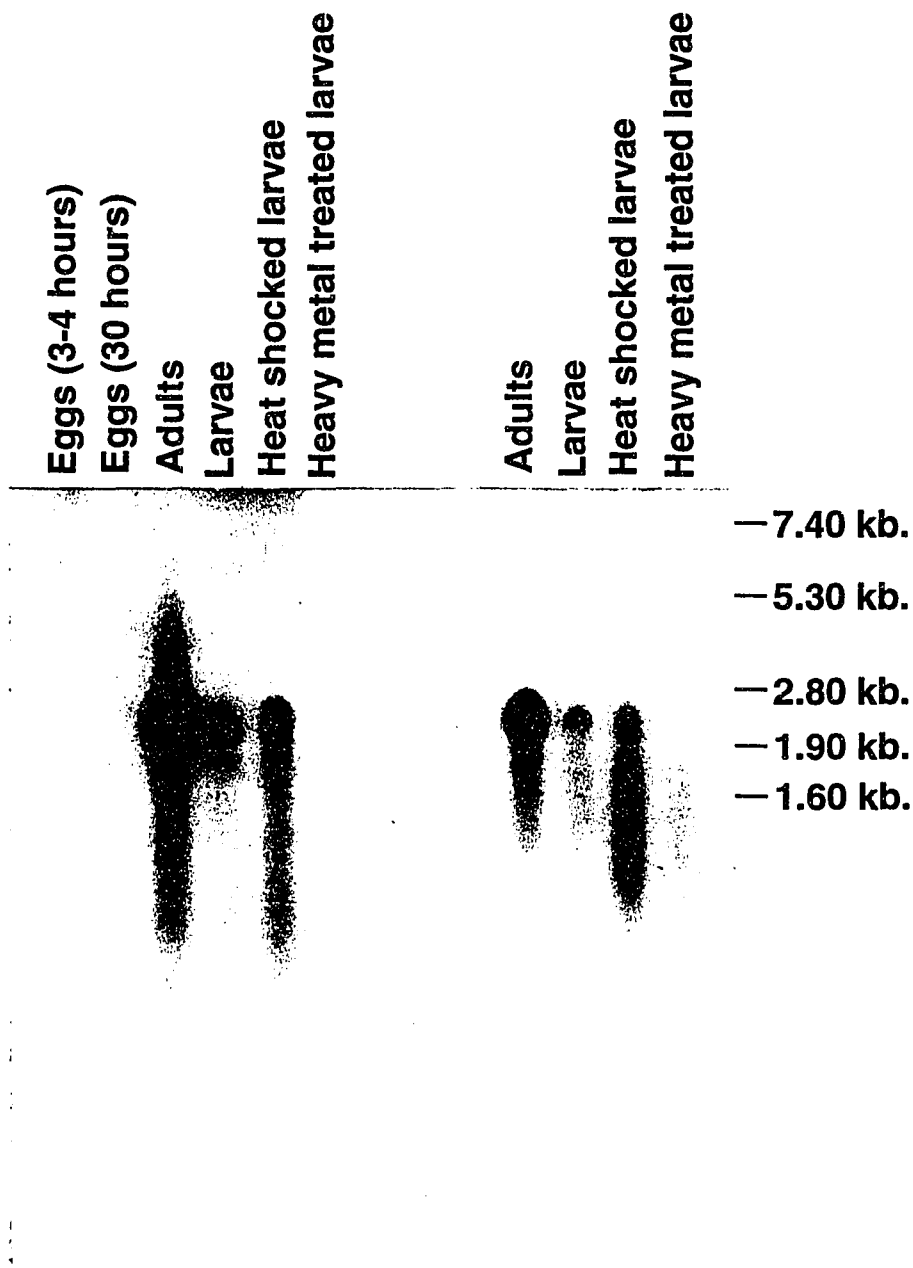
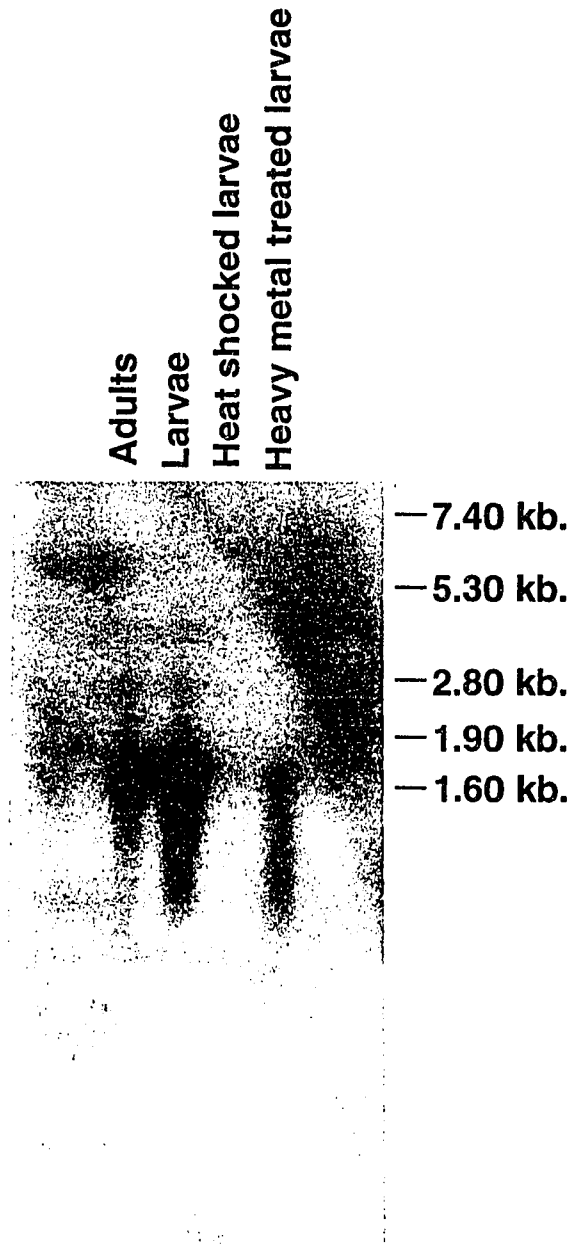

Eggs (3-4 hours)
Eggs (30 hours)
Adults
Larvae
Heat shocked larvae
Heavy metal treated larvae
Adults
Larvae
Heat shocked larvae
Heavy metal treated larvae
—7.40 kb.
—5.30 kb.
—2.80 kb.
—1.90 kb.
—1.60 kb.

Figure 32. **Developmental expression and stress responses of a pmed 21 actin gene.**

This northern blot was probed with the pmed21 actin gene at the same high stringency as the northern blot in figure 30.


Adults
Larvae
Heat shocked larvae
Heavy metal treated larvae
7.40 kb.
5.30 kb.
2.80 kb.
1.90 kb.
1.60 kb.

Chapter 6

Conclusions

Experiments were conducted to study heavy metal tolerance in the Mediterranean fruit fly, (Medfly), *Ceratitis capitata*. A number of strains were established that exhibited some heavy metal tolerance. The A4A, A5A3, and 18A2A strains exhibited some increased ability to tolerate zinc chloride in their larval diet. However, these same strains did not tolerate higher than normal levels of copper sulfate and cadmium chloride. This suggests that genetic basis or mechanism of heavy metal tolerance is specific for each type of heavy metal.

A dominant mode of inheritance could be the underlying genetic basis for the heavy metal (zinc) tolerance trait in these strains. In some cases, F1 progeny of reciprocal crosses between the heavy metal tolerant strains and a nonselected strain exhibited levels of heavy metal tolerance similar to those of the heavy metal tolerant parent. However, the heavy metal tolerance observed could also have a polygenic basis. Each of the heavy metal tolerant strains exhibited somewhat variable patterns of inheritance for this trait. Reciprocal test crosses also suggested that there might be genes associated with the heavy metal tolerance on more than one chromosome in the Medfly. Also, an attempt to generate a Y-autosome translocations involving this trait appears to be not feasible. This analysis was made more complicated due to the fact that the heavy metal tolerance is a quantitative trait which is relatively difficult to score. Further understanding on the genetic

basis of the heavy metal tolerance would help to evaluate potential of this in a translocations based system.

Molecular genetic studies of the heavy metal tolerance were also conducted in an effort to clone a metallothionein gene. PCR products were amplified from the Medfly genome. These PCR products were derived from primers based on conserved domains of a metallothionein gene from *D. melanogaster.* The PCR products were subsequently used to identify homologous sequences in a Medfly genomic library.

The Medfly sequences isolated were similar in terms of nucleotide sequence. Furthermore, they contained a region with potential for coding a metallothionein protein. The coding potential was deduced from the fact that these fragments contained a cysteine rich region (CRR) within their sequences.

The CRR itself appears to possess coding potential for a metallothionein protein because its deduced amino acid sequence can be aligned to a number of mammalian metallothionein proteins. However, transcripts homologous to the CRR were not detected in poly(A) selected RNA.

PCR primers were designed based on conserved amino acid sequences within the CRR. One of these primers (primer IV) amplified a portion of a coding region of a heat shock like gene using heavy metal induced cDNA as a template. This PCR product was used to isolate and characterize an entire gene (Cerhsc1) from a Medfly genomic library.

The Cerhsc1 gene appears to be a *hsc* gene and a member of the *hsp70* gene family. It is most similar to the *hsc4* gene from *D. melanogaster* in terms of the deduced amino acid and nucleotide sequences. A number of DNA sequence and amino acid sequence comparisons of the Cerhsc1 gene suggest that it belongs to a subfamily of *hsp70* genes. These genes produce HSP70 proteins which are localized to the cytoplasm.

The %GC content of the Cerhsc1 coding region is similar to the %GC content of the other Medfly genes. The amino acid composition is similar to that of the *hsp70* genes from *D. melanogaster*. *Hses* are found in the promotor region of the Cerhsc1 gene. This promotor region is connected to the coding region by a potentially long transcribed but untranslated 5' region. The number of *hse* found in this gene are characteristic of a heat inducible gene. Heat inducible genes of this type often lack an intron, and the Cerhsc1 gene also appears to be intronless.

A genomic representation study demonstrates that there may be other members of *hsp70* gene family in the Medfly genome. These *hsp70* genes may be different from the *hsp70* genes of *D. melanogaster* in term of nucleotide sequence and codon usage. There may be several *hsp70* genes in the genome of the Oriental fruit fly and of the Melon fly as well.

Developmental expression of the Cerhsc1 gene is very abundant in the adult stage, less so in the pupal stage and not detectable in the embryo stages. The pattern of the Cerhsc1 gene expression is similar to the *hsc1* and *hsc2*, but not *hsc4*, of *D.*

melanogaster. This indicates that the structural relationship between the Cerhsc1 gene and the Drohsc4 gene does not ensure similarity in their developmental expression.

The heat shock response of the Cerhsc1 gene results in a 10 to 20% increase in the level of transcription. This response demonstrates that the Cerhsc1 gene could function as a heat inducible member as well. This is another different characteristic that distinguishes the Cerhsc1 from the Drohsc4 gene.

The Cerhsc1 gene does not appear to respond to heavy metal induction. On the contrary, its basal level of transcription is suppressed by heavy metal induction. This suggests that there might be another gene family, such as metallothioneins, playing a role in the response to heavy metals.

Under reduced stringency conditions, the Cerhsc 1 probe could detect a number of smaller transcripts which may represent other heat shock genes. These smaller transcripts demonstrated a 57% increase in abundance in response to heat induction. Small heat inducible heat shock proteins have been previously indentified in the Medfly. These other heat shock genes may overlap in function with the Cerhsc1 gene because they seem to share a similar C-terminal protein binding domain. It will be of interest to isolate and characterize these and other heat shock genes in order to have a better understanding in the heat shock response in the Medfly. The Cerhsc1 gene could be used to isolate and characterize these other members of the heat shock gene family.

References

Amin, J., R Mestril, P. Schiller, M. Dreano, and R. Voellmy. 1987. Organization of the *Drosophila melanogaster* hsp70 heat shock regulatory unit. Mol. Cell. Biol. 7(3)1055-1062.

Artymiuk, P.J., H.M. Grindley, K. Kumar, D.W. Rice., and P. Willett. 1993. Three-dimensional structural resemblance between the ribonuclease H and connection domains of HIV reverse transcriptase and the ATPase fold revealed using graph theroretical techniques. Federation of european biochemical societies letters 324(1):15-21.

Bardwell, J.C.A. and E.A. Craig. 1984. Major heat shock gene of *Drosophila* and the *Escherichia coli* heat inducible dnaK gene are homologous. P.N.A.S. 81:848-852.

Benedict, M.Q., A.F. Cockburn, and J.A. Seawright. 1993. The Hsp70 heat-shock gene family of the mosquito *Anophels albimanus*. Insect Mol. Biol. 2:93-102.

Biggin, M. and R. Tjian. 1988. Transcription Factor that activated the ultrabithorax promotor in developmentally staged extracts. Cell 53:699-711.

Boorstein, W.R., T. Ziegelhoffer, and E.A. Craig. 1994. Molecular evolution of the HSP70 multigene family. J. Mol. Evol. 38:1-17.

Bork, P., C. Sander, and A. Valencia. 1992. An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. PNAS. 89(16):7290-7294.

Bremner, I. 1991. Nutritional and physiologic significance of metallothionein. See ref. Riordan, J.F. and B.L. Vallee.

Bull, P.C. and D.W. Cox. 1994. Wilson disease and Menkes disease: new handles on heavy-metal transport. Trend in Genetics 10(7)246-252.

Busch-Petersen, E. and D.I. Southern. 1987. Induced suppression of genetic recombination in females of the Mediterranean fruit fly, *Ceratitis capitata* (Wied.), by translocation heterozygosity. Genetica 72:161-169.

Carter, A.D., B.K. Felber, M. Walling, M.F. Jubier, C.J. Schmidt, and D.H. Hamer. 1984. Duplicated heavy metal control sequences of the mouse metallothionein I gene. P.N.A.S. 81:7392-7396.

Chitnis, P.R. and N. Nelson. 1991. Molecular cloning of the genes encoding two chaperone proteins of the cyanobacterium *Synechocystis sp.* PCC 6803. J. Biol. Chem. 266, 58-65.

Clos, J., J. T. Westwood, P.B. Becker, S. Wilson, K. Lambert, and C. Wu. 1990. Molecular cloning and expression of a hexameric *Drosophila* heat shock factor subject to negative regulation. Cell 63:1085-1097.

Corces, V. and A. Pellicer. 1984. Identification of sequence involved in the transcriptional control of a *Drosophiila* heat shock gene. J. Biol. Chem. 259(23)14812-14817.

Craig E.A., T.D. Ingolia, and L.J. Manseau. 1983. Expression of *Drosophila* heat-shock cognate genes during heat shock and development. Dev. Biol. 99:418-426.

Craig, E.A. and K. Jacobsen. 1984. Mutations of the heat inducible 70 kilodalton genes of yeast confer temperature sensitive growth. Cell 38:841-849.

Craig, E.A., J. Kramer, J. Shilling, M. Werner-Washburne, S. Holmes, J. Kosic-Smithers, and C.M. Nicolet. 1989. SSC1, an essential member of the yeast HSP70 multigene family, encodes a mitochondrial protein. Mol. Cell. Biol. 9(7)3000-3008.

Denecke, J., M.H. Goldman, J. Demolder, J. Seurinck, and J. Botterman. 1991. The tobacco luminal binding protein is encoded by a multigene family. Plant Cell 3:1025-1035.

Dickson, R.C. 1940. Inheritance of resistance to hydrocyanic acid fumigation in the california red scale. Hilgardia 13(9)515-522.

Durnam, D.M., F. Perrin, F. Gannon, and R.D. Palmiter. 1980. Isolation and characterization of the mouse metallothionein-I gene. P.N.A.S. 77(11)6511-6515.

Durnam, D.M. and R.D. Palmiter. 1981. Transcriptional regulation of the mouse metallothionein-I gene by heavy metals. J. Biol. Chem. 256(11):5712-5716.

Dworniczak, B. and M-E Mirault. 1987. Structure and expression of a human gene coding for a 71 kd heat shock 'cognate' protein. Nucleic Acids Research 15(13)5181-5197.

Elis, R.I. and S.M. van der Vies. 1991. Molecular chaperones. Ann. Rev. Biochem. 60:321-47.

Feng, D.F. and R.F. Doolittle. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. 25:351-360.

Fittinghoff, G.M. and L.M. Riddiford 1988. Effect of heat shock on nonheat shock proteins of the insect integument (abstract). J. Cell Biochem. 12D, 268.

Flaherty, K.M., C. DeLuca-Flaherty, and D.B. McKay. 1990. Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. Nature 346:623-628.

Fletcher, B.S. 1989. Life history strategies of tephritid fruit flies. In "World crop pests, fruit flies." eds. A.S. Robinson and G. Hooper. Elsevier press, Amsterdam, 3B:195-208.

Flynn, G.C., J. Pohl, M. Flocco, and J.E. Rothman. 1989. Peptide binding and release by proteins implicated as catalysts of protein assembly. Science 245:385-390.

Flynn, G.C., J. Pohl, M.T. Flocco, and J.E. Rothman. 1991. Peptide-binding specificity of the molecular chaperone BiP. Nature 353:726-730.

Fogel, S. and J.W. Welch. 1982. Tandem gene amplification mediates copper resistance in yeast. P.N.A.S. 79:5342-5346.

Folz, K.R., J.S. Partin, and W.J. Lennarz. 1993. Sea urchin egg receptor for sperm: sequence similarity of binding domain and hsp70. Science 259:1421-1425.

Gasperi, G., D. Kafetzopoulos, A. Christodoulidou, V. Bouriotis, and C. Savakis. 1994. Isolation and partial characterization of two Alcohol dehydrogenase isozymes from the Medfly *Ceratitis capitata.* Insect Biochem. Molec. Biol. 24(1)87-94.

Gething, M.J. and J. Sambrook. 1992. Protein folding in the cell. Nature 355:33-45.

Gilbert, W. 1987. The exon theory of genes. Cold spring harbor symposia on quantitative biology LII:901-905.

Goate, A.M., D.N. Cooper, C. Hall, T.K.C. Leung, E. Solomon, and L. Lim. 1987. Localization of a human heat shock HSP70 gene sequence to chromosome 6 and detection of two other loci by somatic-cell hybrid and restriction fragment length polymorphism analysis. Hum. Genet. 75:123-28.

Griffith, B.B., R.A. Walters, M.D. Enger, C.E. Hildebrand, and J.K. Griffith. 1983. cDNA cloning and nucleotide sequence comparison of chinese hamster metallothionein I and II mRNAs. Nucleic Acid Research 11(3)901-910.

Gupta, R.S. and B. Singh. 1992. Cloning of the HSP70 gene from *Halobacterium marismortui* : relatedness of archaebacteria HSP70 to its eubacterial homologs and a model for the evolution of the HSP70 gene. J. Bacteriol. 174:4594-4605.

Hamer, D.H. 1986. Metallothionein. Ann. Rev. Biochem. 55:913-952.

Harrison, G.S., H.A. Drabkin, F. Kao, J. Hartz, and M. Hart. 1987. Chromosome location of human genes encoding major heat-shock protein HSP70. Somatic Cell Mol. Genet. 13:119-130.

Haus, U., P. Trommler, P.R. Fisher, H. Hartmann, F. Lottspeich, A.A. Noegel, and M. Schileicher. 1993. The heat shock cogante protein from *Dictyostelium* affects actin polymerization through interaction with the actin-binding protein cap32/34. EMBO J. 12:3763-3771.

Haymer, D.S., J.E. Anleitner, M. He, S. Thanaphum., Stephen H. Saul, J. Ivy, K. Houatchens, and L. Arcangeli. 1990. Actin genes in the Mediterranean fruit fly, *Ceratitis capitata.* Genetics : 125:155-160.

He, M. and D.S. Haymer. 1992. A muscle-specific actin gene from the Mediterranean fruit fly, *Ceratitis capitata.* Insect Mol. Biol. 1(1)15-24.

Hoffmann, A.A. and P.A. Parsons. 1993. Evolutionary genetics and environmental stress. Oxford University press, Oxford.

Hoffman, E. and V. Corces. 1986. Sequence involved in temperature and ecdysterone: induced transcription are located in separate regions of a *Drosophila melanogaster* heat-shock gene. Mol. Cell. Biol. 6(2)663-673.

Holmgren, R., V. Corces, R. Morimoto, R blackman, and M. Meselson. 1981. Sequence homologies in the 5' regions of four *Drosophila* heat-shock genes. P.N.A.S. 78(6)3775-3778.

Hopkin, S.P. 1989. Ecophysiology of metals in terrestrial invertebrate. (ed.) Elsevier Applied Sciene, London, UK.

Hunt, C. and R.I. Morimoto. 1985. Conserved features of eukaryotic hsp70 genes revealed by comparison with the nucleotide sequence of human hsp70. P.N.A.S. 82:6455-6459.

Ingolia, T.D., E.A. Craig, and B.J. McCarthy. 1980. Sequence of three copies of the gene for the major *Drosophila* heat shock induced protein and their flanking regions. Cell 21:669-679.

Ingolia, T.D. and E.A. Craig. 1982. *Drosophila* gene related to the major heat shock-induced gene is transcribed at normal temperatures and not induced by heat shock. P.N.A.S. 79:525-529.

Ingolia, T.D., M.R. Slater, and E.A. Craig. 1982. *Saccharomyces cerevisiae* contains a complex multigene family related to the major heat shock-inducible gene of *Drosophila*. Mol. Cell. Biol. 2(11)1388-1398.

Jang, E.B. 1992. Heat shock proteins and thermotolerance in a cultured cell line from the mediterranean fruit fly *Ceratitis capitata*. Arch. Ins. Biochem. and Phys. 19:93-103.

Joplin, K.H. and D.L. Denlinger. 1990. Developmental and tissue specific control of the heat shock induced 70 kDa related proteins in the flesh fly, *Sarcophaga crassipalpis*. J. Insect. Physiol. 36(4)239-249.

Karin, M. and R. I. Richards. 1982. Human metallothionein genes: molecular cloning and sequence analysis of the mRNA. Nucleic Acids Research 10(10)3165-3173.

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

Karin, M. 1985. Metallothioneins: protein in search for functions. Cell 41:9-10.

Karlin, K.D. 1993. Metalloenzymes, structural motif, and inorganic models. Science 261:701-708.

Kerremans, Ph., E. Gencheva, and G. Franz. 1992. Genetic and cytogenetic analysis of Y-autosome translocations in the Mediterranean fruit fly, *Ceratitis capitata*. Genome 35:264-272.

Klaassen, C.D. and K.J. Suzuki. 1991. Metallothionein in biology and medicine. (eds.) CRC press, Boca Raton, pp. 1-414.

Klemenz, R., D. Hultmark, and W.J. Gehring. 1985. Selective translation of heat shock mRNA in *Drosophila melanogaster* depends on sequence information in the leader. EMBO J. 4(8)2053-2060.

Klung, A. and D. Rhodes. 1987. "Zinc fingers": a novel protein motif for nucleic acid recognition. Trends Biochem. Sci. 12:464-469.

Kojima, Y. 1991. Definition and nomenclature of metallothionein. See ref. Riordan, J.F. and B.L. Vallee.

Konsolaki, M., K. Komitopoulou, P.P. Tolias, D.L. King, C. Swimmer, and F.C. Kafatos. 1990. The chorion genes of the medfly, *Ceratitis capitata*, I: Structural and regulatory conservation of the S36 gene relative to two *Drosophila species*. Nucleic Acids Res. 18:1731-1737.

Kwiatowski, J., D. Skarecky, and F.J. Ayala. 1992. Structure and sequence of the Cu, Zn Sod Gene in the Mediterranean fruit fly, *Ceratitis capitata* : intron insertion/deletion and evolution of the gene. Mol. phylo. Evol. 1(1).

La Chance, L.E. 1979. Genetic strategies affecting the success and economy of the sterile insect release method. In"Genetic in relation to insect management." eds. M.A. Hoy and J.J. Mckelvey Jr., The Rockefeller foundation, New York, pp8-18.

Landry, J., H. Lambert, M. Zhou, J.N. Lvoie, and E. Hickey. 1992. Human HSP27 is phosphorylated at serine 78 and 82 by heat shock and mitogen-activated kinases that recognized the same amino acid motif as S6 kinase II. J. Biol. Chem. 267:794-803.

Lange, B.W. 1990. *Drosophila melanogaster* metallothionein genes selection for duplication? Ph.D. thesis, Duke University, North Carolina, pp.1-106.

Lange, B., C.H. Langley, and W. Stephen. 1990. Molecular evolution of *Drosophila* metallothionein genes. Genetics 126:921-932.

Lastowski-Perry, D., E. Otto, and G. Maroni. 1985. Nucleotide sequence and expression of a *Drosophila* metallothionein. J. Biol. Chem. 260(3)1527-1530.

LeJohn, H.B., L.E. Cameron, B. Yang, G. MacBeath, D.S. Barker, and S.A. Williams. 1994. Cloning and analysis of a constitutive heat shock (cognate) protein 70 gene inducible by L-Glutamine. J. Biol. Chem. 269(6)4513-4522.

Lerch, K. 1980. Copper metallothionein, a copper-binding protein from *Neurospora crassa.* Nature 284:368-370.

Levinson, G. and G.A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4(3)203-221.

Lewin, B. 1994. Genes V. Oxford university press, New York, pp1-1272.

Lindquist, S. and E.A. Craig. 1988. The heat-shock protein. Annu. Rev. Genet. 22:631-677.

Lis, J.T., L. Prestidge, and D.S. Hogness. 1978. A novel arrangement tandamly repeated genes at a major heat shock site in *D. melanogaster.* Cell 14:901-919.

Magnusson, J. and C. Ramel. 1986. Genetic variation in the susceptibility to mercury and other metal compounds in *Drosophila melanogaster.* Teratogenesis, Carcinogenesis, and Mutagenesis 6:289-305.

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

Maroni, G., J. Wise, J.E. Young, and E. Otto. 1987. Metallothionein gene duplications and metal tolerance in natural populations of *D. melanogaster.* Genetics 117:739-744.

Maroni, G. 1993. An atlas of *Drosophila* genes: sequences and molecular features. Oxford University Press, New York Oxford.

Melander, A.L. 1914. Can insect become resistant to sprays?. J. Eco. Ento. 7:167-173.

Mokdad, R.A., A. Debec, and M. Wegnez. 1987. Metallothionein genes in *Drosophila melanogaster* constitute a dual system. P.N.A.S. 84:2658-2662.

Mues, G.I., T.Z. Munn, and J.D. Raese. 1986. A human gene family with sequence homology to *Drosopihila melanogaster* Hsp70 heat shock gene. J. Biol. Chem. 26:874-877.

Nassar, R. 1979. Genetics of resistance to tetraethyllead. Aust. J. Biol. Sci. 32:127-132.

Nemer, M., E.C. Travaglini, E. Rondinelli, and J. D'Alonzo. 1984. Developmental regulation, induction, and embryonic tissue specificity of sea urchin metallothionein gene expression. Dev. Biol. 102:471-482.

Nemer, M., D.G. Wilkinson, E.C. Travaglini, E.J. Sternberg, and T.R. Butt. 1985. Sea urchin metallothionein sequence: key to a evolutionary diversity. P.N.A.S. 82:4992-4994.

O'Brien, T., and J.T. Lis. 1991. RNA polymerase II pauses at the 5' end of the transcriptionally induced *Drosophila* hsp70 gene. Mol. Cell. Biol. 11(10)5285-5290.

O'Halloran, T.V. 1993. Transition metals in control of gene expression. Science 261:701-708.

Ohno, S. 1984. Birth of a unique enzyme from an alternative reading frame of the preexited, internally repetitious coding sequence. P.N.A.S. USA 81:2421-2425.

Otto, E., J.M. Allen, J.E. Young, R.D. Palmiter, and G. Maroni. 1987. A DNA segment controling metal-regulated expression of the *Drosophila melanogaster* metallothionein gene Mtn. Mol. Cell. Biol. 7(5)1710-1715.

Parsel, D.A. and S. Lindquist. 1993. The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. Annu. Rev. Genet. 27:437-496.

Patthy, L. 1994. Introns and exons. Current Opinion in Structural Biology 4:383-392.

Pearson, W.R. and D.J. Lipman. 1988. Improved tools for biological sequence analysis. P.N.A.S. 85:2444-2448.

Pelham, H.R.B. 1982. A regulatory upstream promoter element in the *Drosophila* Hsp70 heat-shock gene. Cell 30:517-528.

Pelham, H.R.B. and M. Bienz. 1982. A synthetic heat-shock promotor element confers heat-inducibility on the herpes simplex virus thymidine kinase gene. EMBO J. 1(11)1473-1477.

Perkins, L.A., J.S. Doctor, K. Zhang, L. Stinson, N. Perrimon, and E.A. Craig. 1990. Molecular and developmental characterization of the heat shock cognate 4 gene of *Drosophila melanogaster*. Mol. Cell. Biol. 10:3232-3238.

Perisic, O., H. Xiao, and J.T. Lis. 1989. Stable binding of *Drosophila* heat shock factor to head-to-head and tail-to-tail repeats of a conserved 5 bp. regulatory unit. Cell. 59:797-806.

Peterson, R. and S. Lindquist. 1988. The *Drosophila* hsp70 message is rapidly degraded at normal temperatures and stabilized by heat shock. Gene 72:161-168.

Posthuma, L., R.F. Hogervorst, E.N.G. Joossee, and N.M.Van Straalen. 1993. Genetic variation and covariation for characteristics associated with cadmium tolerance in natural populations of the springtail *Orchesella cincta* (L.). Evolution 47(2)619-631.

Pyle, A.M. 1993. Ribozyme: a distinct class of metalloenzymes. Science 261:709-714.

Rasmussen, E.B. and J.T. Lis. 1993. *In vivo* transcriptional pausing and cap formation on three *Drosophila* heat shock genes. P.N.A.S. 90:7923-7927.

Rina, M. and C. Savakis. 1991. A cluster of vitellogenin genes in the Mediterranean fruit fly *Ceratitis capitata* : sequence and structural conservation in dipteran yolk proteins and their genes. Genetics 127:769-780.

Riordan, J.F. and B.L. Vallee. 1991. Method in enzymology: metallobiochemistry part B metallothionein and related molecules. Academic press, San Diago, California, 205(1-681).

Rippmann, F., W.R. Taylor, J.B. Rothbard, and N.M. Green. 1991. A hypothetical model for the peptide binding domain of hsp70 based on the peptide binding domain of HLA. EMBO J. 10(5)-1053-1059.

Robinson, A.S. 1984. Unexpected segregation ratios from male-liked translocations in the Mediterranean fruit fly, *Ceratitis capitata* (Diptera: Tephritidae). Genetica 62:209-215.

Robinson, A.S. 1989. Genetic sexing methods in the mediterranean fruit fly, *Ceratitis capitata* (Wiedemann). In"World crop pest: fruit flies." eds. A.S. Robinson and G. Hooper, 3B:195-208.

Rochester, D.E., J.A. Winer, and D.M. Shah. 1986. The structure and expression of maize genes encoding the major heat shock protein, hsp70. EMBO J. 5(3)451-458.

Rose, M.D., L.M. Misra, and J.P. Vogel. 1989. Kar2, a karyogamy gene is the yeast homolog of the mammalian BiP/Grp78 gene. Cell 57:1211-1221.

Rougvie, A.E. and J.T. Lis. 1988. The RNA polymerase II molecular at the 5' end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. Cell 54:795-804.

Rubin, M., A.D. Mehta, J. Zhu, S. Shoham, X. Chen, Q.R. Wells, and K.B. Palter. Genomic structure and sequence analysis of *Drosophila melanogaster* HSC70 genes. Gene 128:155-163.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Prsss, New York, Book I to III.

Saul, S.H. 1986. Genetics of the mediterranean fruit fly (*Ceratitis capitata*) (Wiedemann). Agriculture zoology review 1:73-108.

Schmid, D., R. Jaussi, and P. Christen. 1992. From GenEMBL database.

Schmidt, C.J. and D.H. Hamer. 1983. Cloning and sequence analysis of two monkey metallothionein cDNA gene 24:137-146.

Schmidt, C.J., D.H. Hamer, and O.W. McBride. 1984. Chromosomal location of human metallothionein genes: implications for Menkes'disease. Science 224:1104-1106.

Seaton, B.L. and L.E. Vickery. 1993. A gene encoding a new DnaK/ hsp70 homolog in *Escherichia coli.* P.N.A.S. Mannuscript submitted.

Slater, M.R. and E.A. Craig. 1989. The SSA1 and SSA2 genes of the yeast *Saccharomyces cerevisiae.* Nucleic Acids Res. 117: 805-806.

Snutch, T.P., M.F. Heschl, and D.L. Baillie. 1988. The *Caenorhabditis elegans* hsp70 gene family: A molecular genetic characterization. Gene 64:241-255.

Stephan, W., V.S. Rodriguez, B. Zhou, and J. Parsch. 1994. Molecular evolution of the metallothionein gene Mtn in the *melanogaster* species groups: results from *Drosophila ananassae.* Genetics 138:135-143.

Stephanou, S., S.N. Alahiotis, V.J. Marmaris, and C. Christodoulou. 1983. Heat shock response in *Ceratitis capitata.* Comp. Biochem. Physiol. 74(3)425-432.

Stephanou, G. 1987. On the control of heat shock protein synthesis in *Drosophila melanogaster* and *Ceratitis capitata* Insect. Biochem. 17(4) 597-602.

Stuart, G., P. Searle, H. Chen, R. Brinster, and R. Palmiter. 1984. A 12-base-pair DNA motif this is repeated several times in metallothionein gene promotors confers metal regulation to a heterologous gene. P.N.A.S. 81:7318-7322.

Templeton, D.M. and M.G. Cherion. 1991. Toxicological significance of metallothionein. See ref. Riordan, J.F. and B.L. Vallee

Theodore, L., Ann-Shu Ho, and G. Maroni. 1991. Recent evolutionary history of the metallothionein gene Mtn in *Drosophila*. Genet. Res. Camb. 58:203-210.

Tsukiyama, T., P.B. Becker, and C. Wu. 1994. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. Nature 367:525-532.

Walker, V.K., S. Whyard, G.R. Wyatt, and M.R. Kanest. 1986. Heat shock and gene transfer in locusts (abstract) J. Cell. Biochem. 10C, 56.

Werner-Washburne, M., D.E. Stone, and E.A. Craig. 1987. Complex interactions among members of an essential subfamily of hsp70 genes in *Saccharomyces cerevisiiae*. Mol. Cell. Biol. 7:2568-2577.

Westwood, J.T., J. Clos, and C. Wu. 1991. Stress-induced olgomerization and chromosomal relocalization of heat shock factor. Nature 353:822-827.

Wright, S. and Th. Dobzhansky. 1946. Genetics of natural population. XII. Experimental reproduction of some of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura*. Genetics 31:125-56.

Xiao, H. and J.T. Lis. 1988. Germline transformation used to define key features of heat-shock response elements. Science 239:1139-1142.

Xiao, H., O. Perisic, and J.T. Lis. 1991. Cooperative binding of *Drosophila* heat shock factor to arrays of a conserved 5 bp. unit. Cell 64:585-593.

Yang, B. and H.B. LeJohn. 1994. $NADP^+$-specific Glutamate Dehydrogenase. J. Biol. Chem. 269(6)4506-4512.

Yu, C.-E., L.L.C. Bonnycastle, C.R. Hunt, B.J. Trask, K.P. Clancy, J.L.Webe, D. Patterson, and G.D. Schellenberg. 1994. Cloning, sequencing, and mapping of the human chromosome 14 heat shock protein gene (HSPA2). unpublished.

Zakeri, Z.F., D. Wolgemath, and C.R. Hunt. 1988. Identification and analysis of a new member of the mouse Hsp70 gene family and characterization of its unique cellular and developmental pattern of expression in the male germ line. Mol. Cell. Biol. 8:2925-2932.