STRUCTURE, EXPRESSION AND EVOLUTION OF THE 16 KILODALTON

HEAT SHOCK PROTEIN GENE FAMILY OF C. elegans

By

ROLAND HANS RUSSNAK

B.Sc., The University of Calgary, 1979

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Biochemistry

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October 1986

© Roland H. Russnak

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of ___Biochemistry___

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date ___Feb. 2/87___

Abstract

Sequences coding for three related 16 kd heat shock proteins (hsps) of the nematode Caenorhabditis elegans were isolated and characterized. The extensive accumulation of hsp16 mRNA during heat stress facilitated the identification of two cDNAs, CEHS48 and CEHS41, which encoded hsp16 variants. These plasmids were selected by their ability to hybridize to mRNA which directed the synthesis of hsp16 in vitro, and were further characterized by sequence analysis.

Two-dimensional gel electrophoresis of hsp16 synthesized in vitro from mRNA selected by hybridization to either of the cDNAs under conditions of low stringency revealed the existence of at least five electrophoretic variants with significantly different isoelectric points.

The above cDNAs were used as specific probes to isolate recombinant bacteriophage containing C. elegans genomic DNA. Overlapping phage clones were used to define a region of approximately 30 kilobases. The genes coding for hsp16-48, previously identified by cDNA cloning, and for another 16 kd hsp designated hsp16-1 were characterized by DNA sequencing. These two genes were arranged in a head-to-head orientation. Both the coding and flanking regions of these genes were located within a 1.9 kb region which was duplicated exactly to form a perfect 3.8 kb inverted repeat structure. This structure ended in unusual G + C-rich sequences 24 bp in length. The identity of the two arms of the inverted repeat at the nucleotide sequence level implied that the duplication event may have occurred relatively recently in evolution. Alternatively, gene conversion between the two modules could have maintained homology between the two gene pairs.

Comparison of the hsp16-48 gene with its corresponding cDNA revealed the presence of a single, short intron. An intron of comparable length and in an analogous position was also found in the hsp16-1 gene. The introns separated variable and conserved regions within the amino acid sequences of the encoded heat shock proteins. A domain of approximatey 80 amino acids is contained within the conserved second exon and is homologous to a similar region in the small hsps of Drosophila, Xenopus, soybean and man as well as the α-crystallin protein of the vertebrate lens.

Each hsp16 gene contained a TATA box upstream of the start of transcription. Promoter sequences, which have been shown to be required for heat inducibility in various systems, were located upstream of either TATA box.

Northern blot analysis showed that the hsp16-48 and hsp16-1 genes are expressed at levels approximately 20 - 40 fold lower than two closely related genes, hsp16-41 and hsp16-2, upon temperature elevation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Some original material in Sp.
Collection — reduced in thesis.
Pages. 86, 91, 95, 96, 100, 106.

ABBREVIATIONS

A	adenine
ATP	adenosine-5'-triphosphate
BUdR	bromouridineodeoxyribonucleoside
bp	base pairs
C	cytosine
cDNA	complementary DNA
cpm	counts per minute
DNA	deoxyribonucleic acid
dATP	deoxyadenosine-5'-triphosphate
dCTP	deoxycytidine-5'-triphosphate
dGTP	deoxyguanosine-5'-triphosphate
dTTP	deoxythymidine-5'-triphosphate
ddTTP	dideoxythymidine-5'-triphosphate
EDTA	ethylenediamine tetraacetic acid
G	guanine
Gu-HCl	guanidinium hydrochloride
Hepes	N-2-hydroxyethylpiperazine-N-2-ethanesulfonic acid
hsg	heat shock granule
hsp	heat shock protein
hsmRNA	heat shock messenger RNA
HSTF	heat shock transcription factor
HSE, HIP	heat shcok promoter element
kb	kilobases
kd	kilodaltons

mRNA	messenger RNA
N	adenine, cytosine, guanine or thymine
PIPES	piperazine-$\underline{N}$,$\underline{N}$'-bis(2-ethanesulfonic acid)
pI	isoelectric point
polyA$^+$	polyadenylated
polyA$^-$	non-polyadenylated
RNA	ribonucleic acid
RNP	ribonucleoprotein
rpm	revolutions per minute
SDS	sodium dodecyl sulfate
shsps	small heat shock protein
T	thymine
tRNA	transfer RNA
Tris	tris (hydroxymethyl) aminomethane
ts	temperature sensitive

ACKNOWLEDGEMENTS

I. INTRODUCTION

1.1 Response to Thermal Stress

Most cells respond to severe metabolic stress by dramatically altering their pattern of protein synthesis. All organisms studied, including bacteria, plants, fungi, protozoa and animals, both poikilothermic and homeothermic, react in this way to abrupt temperature increases, i.e. heat shock. Altered patterns of protein synthesis during thermal stress can be a result of the regulation of mRNA synthesis or mRNA translation, the usual situation being a coordination of both. In all cases, a well defined set of heat shock proteins (hsps) can be identified.

Historically, the heat shock response was first observed with the light microscope at the level of polytene chromosome puffing in larval salivary glands of dipterans (Ritossa, 1962, Ritossa, 1963). In Drosophila melanogaster, there are nine heat inducible puffs. In 1974 the induction of a set of proteins upon heat shock was reported by Tissières et al. Since then it has become clear that the puffs are sites of RNA transcription and that most of these new RNAs are translated into heat shock proteins. The seven major heat induced proteins of D. melanogaster are designated hsp83, hsp70, hsp68, hsp27, hsp26, hsp23 and hsp22 according to their apparent molecular weights on SDS polyacrylamide gels.

Hsp83 is encoded by a single gene which is located at cytological locus 63B (Holmgren et al., 1979). The hsp83 transcript is characterized by a 1.2 kb intron in the 5' noncoding region (Holmgren et al., 1981; Hackett and Lis, 1983). Sequence analysis of a yeast hsp90 gene (Farrelly and Finkelstein, 1984) and a chicken hsp90 gene (Catelli et al., 1985a) has

demonstrated that the predicted proteins are related to hsp83 of Drosophila (Hackett and Lis, 1983). A polyclonal antibody prepared against chicken hsp90 reacts with proteins of similar mobility in heat shocked human, rodent and frog cells (Kelly and Schlesinger, 1982). Although the function of Drosophila hsp83 is unknown, hsp90 is associated with non-transformed 8S steroid receptors in chick oviduct (Catelli et al., 1985b) as well as pp60^{src} in avian cells infected with Rous sarcoma virus (Oppermann et al., 1981). Drosophila hsp83 and its related proteins are detectable under nonstress conditions, their levels increasing significantly during heat shock.

In D. melanogaster, the hsp70 gene family consists of both heat inducible and constitutively expressed genes. At least five highly homologous heat inducible genes are distributed over loci 87A and 87C (Schedl et al., 1978; Livak et al., 1978; Moran et al., 1979; Artavanis-Tsakonas et al., 1979; Craig et al., 1979; Goldschmidt-Clermont, 1980; Ingolia et al., 1980). Hsp68 is also heat inducible, and shows 75% homology to hsp70 and therefore can be considered part of the same family. It is encoded by a single gene at locus 95D (Holmgren et al., 1979). Hsp70 related proteins are also detectable in normal cells. They are the products of three constitutively expressed genes (heat shock cognate genes) referred to as hsc1, hsc2 and hsc4 which map cytologically to 70C, 87D and 88E, respectively (Ingolia and Craig, 1982a; Craig et al., 1983; Palter et al., 1986). The hsc1 and hsc2 genes contain introns while the heat inducible genes are uninterrupted.

A family of inducible and constitutively expressed genes related to those in Drosophila are found in yeast (Ingolia et al., 1982; Craig and

Jacobsen, 1984; Craig and Jabobsen, 1985; Craig et al., 1985),

Caenorhabditis elegans (Snutch and Baillie, 1984) and mouse (Lowe and Moran,

1984; Lowe and Moran, 1986). Heat inducible hsp70 sequences have been

characterized in rainbow trout (Kothary et al., 1984), Xenopus (Bienz,

1984a), Dictyostelium (Rosen et al., 1985) and man (Voellmy et al., 1985; Wu

et al., 1985; Hunt and Morimoto, 1985). A related cognate gene has been

characterized in rat (O'Malley et al., 1985). In E. coli, protein DnaK,

which is 48% homologous to Drosophila hsp70 (Bardwell and Craig, 1984), is

an extremely abundant protein at normal growth temperature, but its level

increases during heat treatment (Neidhardt et al., 1984). It is encoded by

a single gene (Bardwell and Craig, 1984). At the protein level, a heat

inducible protein migrating with an apparent molecular weight of 70,000 has

been identified in every organism tested, making it one of the most highly

conserved proteins known. One of the constitutively expressed hsp70 species

in mammalian cells has been identified as a clathrin uncoating enzyme with

ATPase activity (Ungewickell, 1985; Chappell et al., 1986).

The discussion which follows will be confined to the structure and

regulation of the genes within the heat shock puff at locus 67B of

Drosophila and to related genes in other organisms. This locus is linked to

the expression of hsp27, hsp26, hsp23 and hsp22.

1.1.1 Locus 67B: Puffing Activity and Protein Synthesis

If Drosophila larvae are subjected to a temperature of 37°C, 25°C

being the normal culture temperature, puffing at locus 67B occurs within one

minute. Puffs continue to grow for 30 - 40 minutes before regressing. If

protein synthesis is blocked by incubating excised salivary glands in medium

containing cycloheximide, the induced puffs fail to regress unless the temperature is returned to normal. Also, the severity of the temperature increase is reflected in the maximum size of puff observed in a given time.

This cytological analysis led to two other important observations. Firstly, most of the puffs existing prior to heat shock regress upon temperature elevation. Upon prolonged exposure to high temperature, they return to normal following the regression of the heat shock puffs. Secondly, locus 67B can be induced to puff by a variety of other conditions including exposure to dinactin (Rensing, 1973), dinitrophenol (Ritossa, 1963; Ellgard and Rensing, 1972, 1973), hydrogen peroxide (Compton and McCarthy, 1978) and valinomycin (Rensing, 1973) or by recovery from anoxia (Ashburner, 1970; Zhimulev and Grafodatskaya, 1974). Thus, puffing activity first defined the parameters and kinetics of gene activity at locus 67B.

Exposure to high temperature also results in the rapid synthesis of approximately seven hsps whereas the rate of synthesis of most cellular proteins normally made at $25^{o}C$ is strongly reduced (Tissières et al., 1974). This phenomenon is observed in excised salivary glands as well as tissue from brain, malpighian tubules and wing imaginal disks. Various Drosophila melanogaster tissue culture lines show similar protein metabolism (McKenzie et al., 1975; Spradling et al., 1975; Moran et al., 1978; Mirault et al., 1978).

Upon heat shock, polysomes from cultured cells disaggregate and newly synthesized heat shock mRNA (hsmRNA) sediments as two new polysomal peaks in sucrose density gradients (McKenzie et al., 1975). The 12S RNA fraction contains hsmRNA that has been shown to direct the synthesis of 4 proteins with molecular weights of 22,000, 23,000, 26,000 and 27,000 in a rabbit

reticulocyte lysate (Mirault et al., 1978; Moran et al., 1978) or in an ascites cell-free extract (McKenzie and Meselson, 1977). These proteins correspond to 4 of the 7 hsps identified in vivo, the other three having molecular weights of 70,000, 68,000 and 83,000. Their corresponding transcripts are found in a 20S RNA polysomal fraction.

Pulse labelled RNA extracted from the 12S RNA peak hybridizes in situ to chromosomal site 67B (Spradling et al., 1977; McKenzie and Meselson, 1977) and thus a direct link can be made between puffing activity and hsmRNA synthesis at locus 67B and the appearance of a specific subset of hsps upon thermal stress.

Pulse labelling with [^{35}S]methionine for 10 minutes at the start of a heat shock is sufficient to detect the initiation of synthesis of all seven hsps. The maximum rate of synthesis is reached between 90 and 120 minutes with a subsequent decline to about 50% after 6 - 8 hours at 37°C (Mirault et al., 1978). In contrast, when the cells are returned to 25°C after a heat shock at 37°C for 1 hour, and the rate of synthesis of hsps is examined by pulse labelling at various times thereafter, there is a gradual decrease of these proteins until, by 8 hours, no hsp synthesis is detected and normal protein synthesis has recovered completely. Mirault et al. (1978) also demonstrated that hsps, pulse labelled during a 1 hour heat shock, are stable for up to one day at 25°C.

1.1.2 Locus 67B: Gene Organization

The cloning of sequences coding for hsp22, 23, 26 and 27 was greatly
facilitated by the fact that their transcripts were highly enriched in the
12S RNA polysomal peak after heat shock. Thus, cDNAs made to 12S RNA, and
coding for hsp23 and hsp26, were identified by translation arrest (Voellmy
et al., 1981) and hybridization selection (Wadsworth et al., 1980)
experiments. These cDNAs do not cross-hybridize under normal stringency but
do hybridize to locus 67B using in situ hybridization. Drosophila
melanogaster genomic DNA clones were subsequently identified that contained
coding regions for all 4 proteins clustered within 11 kb (Craig and
McCarthy, 1980; Corces et al., 1980; Voellmy et al., 1981). The
orientations of the genes were determined using R-loop mapping; the polyA
ends of the mRNA were identified by their hybridization to poly(BUdR)-tailed
plasmid pBR345 (Voellmy et al., 1981), or by the hybridization of
end-labelled DNA fragments derived from the clones to hsmRNA (Craig and
McCarthy, 1980; Corces et al., 1980). The lack of intervening sequences was
suggested at this point by R-loop mapping data (Voellmy et al., 1981) and by
the fact that the regions homologous to hsmRNA were mapped on the basis of
their resistance to S1 nuclease after hybridization (Corces et al., 1980).
These genomic clones selected messages which synthesized the 4 hsps in vitro
(Craig and McCarthy, 1980; Corces et al., 1980). Furthermore, the hsp27, 26
and 23 genes showed partial homology under reduced stringency (Corces
et al., 1980).

The first hint that these genes were developmentally regulated came
from workers who identified a genomic clone which hybridized more strongly
to pupa-specific cDNA than to embryo-specific cDNA (Sirotkin and Davidson,

1982). This clone hybridized in situ to locus 67B and overlapped the clone
identified by Corces et al. (1980). Five regions homologous to pupa RNA
were identified and the orientations of four of them were determined by
R-loop mapping using poly(BUdR). Two of these genes code for hsp26 and
hsp23 while the other three have been named genes 1, 2 and 3 (Southgate
et al., 1985). Genes 1, 2 and 3 are heat inducible although their
corresponding messages do not accumulate to the same high levels as those of
hsp23 and 26 (Ayme and Tissières, 1985). Gene 1 cross-hybridizes to the
hsp27 gene while genes 2 and 3 share a weak homology to the hsp22 gene (Ayme
and Tissières, 1985). These seven genes, the organization of which is shown
in Figure 1, will be referred to collectively as the small heat shock genes,
and their products as the small hsps (shsps) of Drosophila.

1.1.3 Locus 67B: Protein Coding Regions

The protein coding regions of the hsp22, 23, 26 and 27 genes have been
sequenced by two groups (Ingolia and Craig, 1982b; Southgate et al., 1983).
Both sequence determinations indicate the same start and stop codons for the
four uninterrupted open reading frames. However, there are some differences
(96% homology at the amino acid level) which are probably not due to
sequencing errors since some changes demonstrably either create or destroy
18 restriction sites (Southgate et al., 1983). Since both clones were
derived from strain Oregon R DNA, the differences must be due to
polymorphism in the fly population. The protein coding region of gene 1 has
been sequenced also (Ayme and Tissières, 1985).

The derived molecular weights for the unmodified proteins are 19,705,
20,603, 22,9976, 23,620 and 26,560 which correspond to hsp22, hsp23, hsp26,
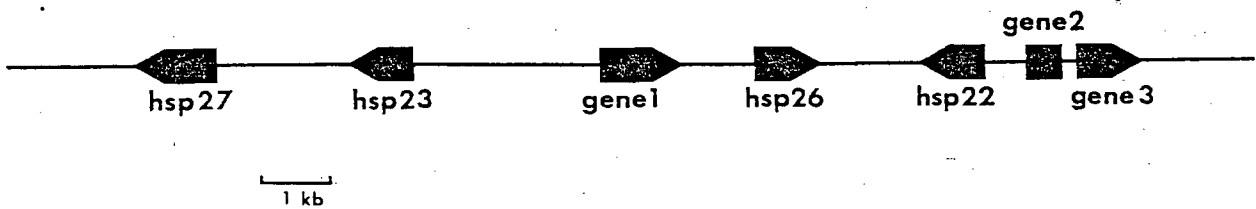
Figure 1.   Gene organization at locus 67B of <u>Drosophila</u>.

The direction of transcription is indicated by the arrows except for gene 2 for which the orientation has not yet been determined.   Genes 1, 2 and 3 are referred to as genes 1, 4 and 5 by Sirotkin and Davidson (1982).

hsp27 and the gene 1 product respectively. Although genes 1, 2 and 3 are heat inducible as measured by transient changes in corresponding transcript levels, protein products have not been detected for these genes. However, in two-dimensional gel electrophoretic separations of both in vivo (Mirault et al., 1978; Buzin and Peterson, 1982) and in vitro (Buzin and Peterson, 1982) translated products, hsp27 consists of multiple spots. Similarly, in heat shocked salivary glands from third instar larvae, hsp22 runs as 2 spots on two-dimensional gels (Buzin and Peterson, 1982).

A schematic illustration of potential functional domains within the Drosophila shsps is shown in Figure 2. Hsp22, 23, 26 and 27 are homologous over a stretch of 108 amino acid residues in which the same amino acid is used in all four proteins at 35% of the positions. The same amino acid is used by three of the four proteins in 71% of the positions. Among each other, hsp 27, 26 and 23 have nearly twice as many similarities within this region as hsp22 does with respect to the other three, thus confirming earlier hybridization selection experiments (Corces et al., 1980). The size differences between these four proteins are accounted for by two heterologous regions, variable in length, on either side of the conserved domain. Surprisingly, the first 83 amino acids of the conserved stretch of 108 amino acids shows a high degree of homology with amino acids 70 - 152 of the bovine α-crystallin B_2 chain (van der Ouderaa et al., 1973), the α-crystallins comprising one of the most abundant protein classes in the vertebrate lens. It is this 83 amino acid region which links the gene 1 product to the other shsps. The gene 1 product shows the most similarity to hsp27 and, interestingly, lacks the conserved 25 amino acid stretch on the carboxy-proximal side which is common to the other four proteins.
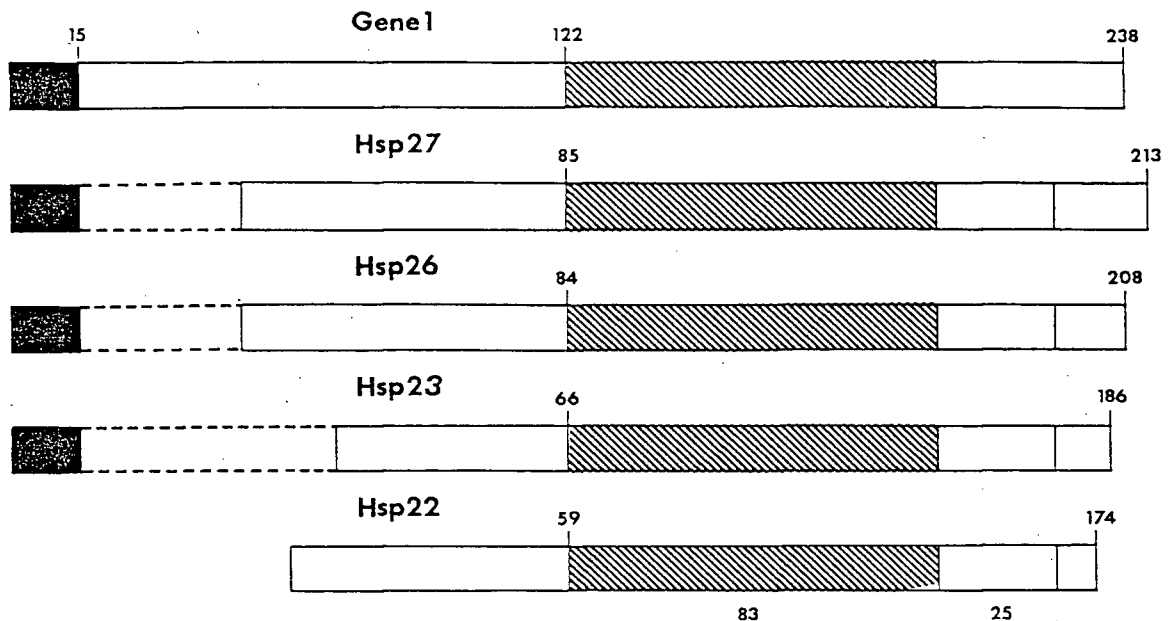
Figure 2. A schematic diagram of the various amino acid homologies between

the shsps of Drosophila melanogaster.

The predicted amino acid sequences are based on the nucleotide sequence

of their coding regions (Ingolia and Craig, 1982b; Southgate et al., 1983;

Ayme and Tissières, 1985).   The 83 amino acid domain which is common to all

shsps and to the vertebrate α-crystallins is shown by hatched boxes.

Other homologous regions are defined by the amino-terminal 15 amino acid

residues and the 25 amino acids located carboxy-proximal to the major

domain.   Also shown, for each protein, is the amino acid numbering which

corresponds to the regions discussed above.

A second region of weaker homology exists within the first 15 predominantly hydrophobic amino acid residues of hsp23, 26, 27 and the gene 1 product. The significance of this hydrophobicity in the N-terminal portion of 4 of the 5 proteins remains unknown; it does not constitute a signal peptide, since a comparison of in vivo and in vitro translation products suggests that there is no post-translational processing (Mirault et al., 1978).

The most striking feature of hydropathy profiles is a prominent hydrophilic peak spread over 16 amino acids in the middle of the 83 amino acid conserved domain (Southgate et al., 1983).

1.1.4 Locus 67B: Activity During Development

Although all seven genes at locus 67B respond to an environmental stimulus such as heat shock, it is now well established that they are under complex and non-coordinate control during fly development.

The hsp22, 23, 26 and 27 genes are normally transcribed from the third instar larval to the mid-pupal stages (Sirotkin and Davidson, 1982; Cheney and Shearn, 1983; Mason et al., 1984; Ayme and Tissières, 1985); however, their levels of expression vary. The hsp23 transcript is the most abundant by 2 - 5 fold as compared with hsp26 and hsp27 mRNAs, while hsp22 mRNA is barely detectable (Ayme and Tissières, 1985). Genes 1 and 3, but not gene 2, are also expressed during the same period (Sirotkin and Davidson, 1982; Ayme and Tissières, 1985), at levels equal to those of hsp23 and hsp26/27, respectively (Ayme and Tissières, 1985). This period of development in Drosophila is characterized by high ecdysterone levels (Hodgetts et al., 1977).

The effects of ecdysterone on hsp synthesis can be mimicked in vitro. Incubation of Drosophila tissue culture (S3) cells with ecdysterone induces the synthesis of hsp22, 23, 26 and 27 to levels consistent with their relative transcript accumulation in vivo (Ireland and Berger, 1982). In isolated imaginal disks from late third instar larvae, hsp23, 26 and 27 mRNA levels are further induced when disks are incubated in ecdysterone (Ireland et al., 1982). Transcripts corresponding to genes 1 and 3 are only weakly induced in either hormone treated S3 cells or imaginal disks (Ireland et al., 1982). In primary cell cultures prepared from Drosophila gastrulation stage embryos, ecdysterone and known teratogenic drugs such as coumarin and diphenylhydantoin inhibit muscle and neuron development in vitro. Under these conditions, only hsp22 and hsp23 are induced (Buzin and Bournias-Vardiabasis, 1984).

A subset of the genes found at locus 67B are also expressed normally during embryogenesis. Messages corresponding to hsp27 and hsp26, as well as hsp23, can be detected in the first 4 hours of embryonic development (Zimmerman et al., 1983; Ayme and Tissières, 1985) while gene 2 mRNA is present in 3 - 24 hour embryos (Sirotkin and Davidson, 1982; Ayme and Tissières, 1985). Early hsp27 and hsp26 mRNAs accumulate in ovarian nurse cells within the egg chamber and are passed into the oocyte (Zimmerman et al., 1983). Thus, these messages are also detected in 3 - 5 day old females (Zimmerman et al., 1983; Ayme and Tissières, 1985). Hsp23 (Mason et al., 1984; Ayme and Tissières, 1985) and gene 1 (Ayme and Tissières, 1985) transcripts are detectable in freshly eclosed male and female adult flies.

If a hyperthermic shock is applied to animals from the larval to adult stages, transcription of the hsp22, 23, 26 and 27 genes increases significantly over their basal, developmentally controlled levels, the transcripts accumulating to approximately the same level at all 6 stages tested (Ayme and Tissières, 1985). The hsp23 gene, which gives the higher level of developmentally regulated transcripts, is comparatively less highly induced during heat shock than the hsp22, 26 and 27 genes.

In contrast, genes 1 and 3 are heat shock induced to 10-fold higher levels in white pre-pupae and middle pupae than in third instar larvae, late pupae or adults. Gene 2 becomes moderately transcribed after heat shock in all larval to adult stages.

In pre-blastoderm embryos (0 - 3 hours), a period in which there is no nuclear RNA synthesis, the genes coding for hsp22, 23, 26 and 27 are unresponsive to heat shock (Dura, 1981; Zimmerman et al., 1983).

These complex patterns of transcriptional activity demonstrate that the seven genes clustered within 15 kb of DNA can each be regulated independently of its neighbors under non-stress conditions during development. Even under heat shock, there is a variable stage-specific accumulation of transcripts coding for the putative products of genes 1, 2 and 3. The specific induction of certain genes such as the expression of gene 2 only in 3 - 24 hour embryos, for example, suggests that their products may serve a vital function in normal development.

1.2 Conserved shsp Function in Eucaryotes

1.2.1 Higher Plants: Soybean

Upon temperature elevation to $35^{\circ}C - 43^{\circ}C$ from $28^{\circ}C$, soybean (Glycine max) seedlings (Key et al., 1981; Altschuler and Mascarenhas, 1982) and tissue culture cells (Barnett et al., 1980) respond by shutting down existing protein synthesis with the concomitant induction of heat shock protein synthesis in a situation analogous to that in Drosophila. Hsp synthesis is characterized by major protein products with molecular weights of 15,000 - 18,000 which are resolved into at least 10 components on two-dimensional gels (Key et al., 1981). The pattern of protein synthesis is common to both root and hypocotyl tissue as well as to both in vivo and in vitro translated products.

Based on hybridization selection experiments, cDNAs made to polyA^{+} mRNA from heat shocked hypocotyl tissue fall into 2 classes (Schöffl and Key, 1981). Class I cDNAs hybridize to mRNAs 800 - 900 nucleotides in length and which code for 20 hsps in the 15 - 18 kd size range when translated in vitro using a wheat germ lysate. A class II cDNA selects a message that codes for a single 18 kd protein on two-dimensional gels.

The characterization of several genomic DNA clones containing sequences coding for the shsp (hsp17) family of soybean has been undertaken by two groups. These genes and their corresponding protein products are listed in Table I.

Table I. Soybean Hsp17 Gene Family

		Predicted Protein		
Gene	cDNA Class	Molecular Weight	Amino Acid Residues	Reference
17.5-M	I	17,544	153	Nagao et al., 1985
17.6-L	I	17,570	154	Nagao et al., 1985
17.5-E	II	17,533	154	Czarnecka et al., 1985
6871	I	17,345	153	Schöffl et al., 1984
6834	–	–	–	Schöffl et al., 1984

The 6834 gene is incomplete and is linked to the 6871 gene, separated by approximately 4 kb. The linkage of the other genes has not yet been determined.

None of the four complete genes contain introns and at the amino acid level they show an overall homology of 90%. The proteins contain a region in their carboxy-terminal half which shows extensive homology to the conserved domain of the proteins encoded by locus 67B in Drosophila. Due to the apparent conservation of a functional domain it is not surprising that hydropathy profiles of the soybean shsps display a prominent hydrophilic peak in an analogous position to that found in Drosophila (Czarnecka et al., 1985).

The shsps of soybean can also be induced with arsenite and cadmium (Czarnecka et al., 1984).

Homologous shsp genes are found in other higher plants since hsp17 class I cDNAs cross-hydridize to other plant species such as pea, millet, maize, and sunflower (Key et al., 1983). At the protein level, major heat inducible products with similar molecular weights occur in a tobacco (Nicotiana tobacum) tissue culture cell line (Barnett et al., 1980), in field grown cotton (Gossyupium hirsutum) (Burke et al., 1985), and in suspension cultures of tomato (Lycopersicon peruvianum) (Scharf and Nover, 1982).

In maize (Zea mays) seedlings, 4 - 8 hsps of 18 kd are seen on two-dimensional gels (Basczczynski et al., 1982). These proteins are synthesized in a great variety of maize tissues after heat shock (Cooper et al., 1984; Baszczynski et al., 1985). They are, however, not observed in the germinating pollen grains of the maize plant (Cooper et al., 1984), a stage in which RNA synthesis takes place under non-stress conditions. This non-responsiveness is therefore analogous to that seen with hsp30 in early Xenopus development (section 1.2.3). The absence of hsp synthesis in germinating pollen grains has also been observed in another plant, Tradescantia paludosa (Xiao and Mascarenhas, 1985).

1.2.2 Avian and Mammalian Systems

A 25 kd protein is inducible in chick embryo fibroblast (CEF) cells during a 41°C - 45°C incubation (Kelly and Schlesinger, 1978). The synthesis of this hsp can also be induced with a wide variety of agents which are summarized in Table II.

Table II. Hsp25 Induction in CEF Cells

Inducer	Reference
1. Amino Acid Analogues	
canavanine	Kelly and Schlesinger, 1978; Hightower, 1980
p-fluorophenylalanine	Hightower, 1980
hydroxynorvaline	Kelly and Schlesinger, 1978
o-methyl threonine	Kelly and Schlesinger, 1978
2. Copper Binding Ligands	
kethoxal bis(thiosemicarbazone)	Levinson et al., 1978a
disulfiram	Levinson et al., 1978b
3. Chelating Agents	
o-phenanthroline	Levinson et al., 1979
8-hydroxyquinoline	Levinson et al., 1979
4. Viral Infection	
Herpes simplex virus	Notarianni and Preston, 1982
Newcastle disease virus	Collins et al., 1980
5. Metals	
copper, cadmium, zinc, mercury	Levinson et al., 1980

Table II. Hsp25 Induction in CEF Cells (cont'd)

Inducer	Reference
6. Others	
arsenite	Levinson et al., 1980
puromycin	Hightower, 1980

Both in vivo and in vitro synthesized chicken hsp25 appears as two species on 2-D gels (Johnston et al., 1980; Wang et al., 1981). Minor amounts of hsp25 have also been observed in uninduced cells (Johnston et al., 1980; Wang et al., 1981).

Using clones coding for Drosophila hsp27, 26 and 23, no hybridization was found to chicken mRNA (White and Hightower, 1984). Similarly, antibodies against hsp25 (Kelly et al., 1980) fail to react with cell extracts from other organisms including yeast, slime mold (Dictyostelium), maize seedling roots, Caenorhabditis elegans, Drosophila melanogaster, Xenopus kidney cells, mouse L929 cells and human WI38 cells (Kelly and Schlesinger, 1982). However, anti-hsp25 cross-reacts with a protein in extracts prepared from 11 day old embryonic chicken lens and thus can be considered analogous to the shsps of Drosophila, Xenopus and soybean (Schlesinger, 1985).

Anti-hsp25 cross-reacts with an avian muscle protein of subunit molecular weight 22,000 which is unusually abundant in chicken embryonic heart as well as in skeletal muscle (Schlesinger, 1985). This antibody also

shows that small amounts of hsp25 are present in unstressed cells, confirming other observations.

In developing chicken embryos, hsp25 displays variable levels of induction in different tissues after a heat treatment. Hsp25 is much more abundant in heart and liver tissue than in brain and lung (Voellmy and Bromley, 1982). Similarly, other workers failed to detect the induction of hsp25 synthesis in brain isolated from 9 day old embryos under conditions in which it was easily seen in limb and breast tissue (Atkinson et al., 1985).

A 25 kd protein is also one of the major inducible proteins in primary cultures of japanese quail breast tissue (myoblasts) when these are subjected to a temperature of $43^{\circ} - 46^{\circ}$C (Atkinson, 1981). Interestingly it loses its heat inducibility by 100 – 120 hours of development when the cells have differentiated into myotubes, but its ability to be induced by arsenite, copper or zinc remains (Atkinson et al., 1983). Arsenite also induces hsp25 in chicken embryonic myotubes (Wang et al., 1981). Furthermore, quail blood cells (white and red) do not synthesize hsp25 when treated with copper or zinc (Atkinson et al., 1983) but have the ability to produce hsp25 under heat shock conditions. These results suggest that different inducers of the heat shock response may act through different mechanisms.

Quail hsp25 runs as a single band with a pI of 5.4 (Atkinson, 1981) and appears to be analogous to its chicken counterpart since it reacts with anti-chicken hsp25 (Atkinson et al., 1983).

The induction of a protein with an approximate molecular weight of 25,000 has been observed in a variety of mammalian cell lines which are reviewed in Table III.

Table III. Hsp25 Induction in Mammalian Cells

Cell Type	Inducer	Reference
human HeLa	heat	Hickey and Weber, 1982
human lymphocytes	arsenite/ethanol	Rodenhiser et al., 1985
rat embryo fibroblasts	heat	Welch and Feramisco, 1985
rat embryo fibroblasts	arsenite/proline analogue	Welch, 1985
rat myoblasts	heat/arsenite/arsenate	Kim et al., 1983
rat primary fibroblasts	arsenate	Kim et al., 1983
rat primary hepatoma	arsenate	Kim et al., 1983
rat pituitary tumor	arsenate	Kim et al., 1983
rat hepatoma MH-7777	heat	Lamarche et al., 1985
mouse embryonal carcinoma	arsenite	Bensaude and Morange, 1983
mouse primary fibroblasts	arsenite	Bensaude and Morange, 1983
mouse lymphocytes	arsenite/ethanol	Rodenhiser et al., 1985
mouse myeloma	heat	Hickey and Weber, 1982
mouse 3T3 fibroblasts	heat	Hickey and Weber, 1982

Table III. Hsp25 Induction in Mammalian Cells (cont'd)

Cell Type	Inducer	Reference
chinese hamster ovary	heat	Bouche et al., 1979
chinese hamster fibroblast	heat	Li and Werb, 1982
hamster primary fibroblast	heat	Atkinson and Pollock, 1982
rabbit lymphocyte	arsenite/ethanol	Rodenhiser et al., 1985

The human variant does not incorporate [^{35}S]methionine but can be labelled using [^{3}H]leucine (Hickey and Weber, 1982). Thus the induction of a shsp could not be observed in human foreskin cells (Levinson et al., 1980), in a human fibrosarcoma (HT1080) cell line (Slater et al., 1981) or in human lymphocytes (Atkinson and Dean, 1985) when [^{35}S]methionine was used in the protein labelling procedure. In human epidermoid carcinoma (KB) cells, hsp25 is not detected in heat shocked, [^{14}C]leucine treated cells (Atkinson and Pollock, 1982) and may be an example of non-responsiveness in certain transformed states.

Human hsp25 runs as two components on two-dimensional gels, the more basic one being present in cells grown at 37°C (Hickey and Weber, 1982).

In non-human mammalian cells, hsp25 can be detected using methionine as the radioactive amino acid. The only reported failure to detect an

$[^{35}S]$methionyl hsp25 under conditions which readily induced the other hsps as well as hsp25 in CEF cells, was for mouse L cells and baby hamster kidney cells (Kelly and Schlesinger, 1978). This cannot be a species specific phenomenon since a $[^{35}S]$methionyl hsp25 is induced in chinese hamster ovary and chinese hamster fibroblast cells as well as a mouse primary fibroblast cell culture (see Table III for references).

As in HeLa cells and chicken cells, other mammalian shsps appear as multiple components on two-dimensional gels. Three variants of hsp25 have been observed in heat shocked primary cultures of hamster fibroblasts (Atkinson and Pollock, 1982). The detailed analysis of hsp25 isoforms in various rat cell types (Kim et al., 1983; Welch, 1985) emphasizes the complexity of shsp regulation. In rat myoblast cells, 4 heat inducible proteins in the 25 kd - 30 kd range can be distinguished on two-dimensional gels (Kim et al., 1983). Of these, 2 are not labelled with methionine while arsenite or arsenate treatment induces an additional 2 proteins of comparable size. The same authors have also shown that arsenate induces a 30 kd protein in 13 day old chick embryo muscle in addition to hsp25. Furthermore, the induction of these proteins varied in the different rat tissue cell lines that were used. Welch (1985) has reported the existence of four hsp25 isoforms in rat embryo fibroblast cells, none of which are labelled with methionine. The induction of these four proteins, which were shown to be related, differed depending upon the agent used which included heat, arsenite and amino acid analogues.

In rat myoblast cells, two hsp25 proteins are phosphorylated at seryl residues (Kim et al., 1984). Similarly, it has been shown that three hsp25 isoforms in rat embryo fibroblast cells are phosphoproteins in both normal

and stressed cells (Welch, 1985). The phosphorylation of two of these proteins increases when the cells are treated with a phorbol diester (phorbol-12-myristate-13 acetate), a calcium ionophore, A23187, or if quiescent cells are given fresh serum. None of these treatments induces a stress response.

The characterization of two genes related to the 4 member gene family encoding human hsp25 has recently been reported (Hickey et al., 1986). One gene, designated HS11, is a member of a cluster of three genes linked within a 14 - 18 kb region of the genome. S1 nuclease protection experiments confirm that this gene is expressed at low levels in control cells but is induced 20-fold during heat shock. The open reading frame, which is interrupted by two intervening sequences of 723 and 120 bp, encodes a polypeptide of 22,300 deduced molecular weight. The second isolated gene, designated HS8, appears to be a processed pseudogene lacking promoter elements and is unlinked to the other members of the hsp25 gene family. The deduced amino acid sequence of hsp25 shows homology to the vertebrate α-crystallins and to the shsps of Drosophila, soybean and Xenopus and thus is indicative of a carefully conserved essential physiological role.

1.2.3 Xenopus

In Xenopus laevis somatic cells or cultured fibroblasts, a 30 kd protein is induced at temperatures between $32^{o}C$ and $37^{o}C$ (Bienz, 1982). In vitro translation experiments demonstrate that its corresponding mRNA is newly synthesized during heat shock and thus is under transcriptional control, as are the genes at locus 67B in Drosophila.

Sequences encoding hsp30 have been isolated from cDNAs made to heat shocked Xenopus polyA RNA from an epithelial kidney fibroblast cell line (Bienz, 1984a). DNA sequence data reveals a 45% homology to the 83 amino acid domain of the Drosophila shsps. No homology is detectable in the preceding 80 amino acid residues or in the carboxy-terminal 40 amino acid residues. Clones from a Xenopus laevis blood cell genomic DNA library have been isolated and two hsp30 coding regions have been identified as hsp30A and hsp30B (Bienz, 1984b). Neither gene is completely identical to the sequenced cDNA. Based on this, along with unpublished genomic Southern blot results, Bienz (1984b) suggests that there may be a multi-gene family of 5 - 10 members coding for hsp30. This is supported by the observation that 5 heat inducible 30 kd proteins can be separated on two-dimensional gels (Guedon et al., 1985).

The hsp30 cDNA has been used to quantify corresponding mRNA levels during Xenopus development and in various somatic tissues (Bienz, 1984a). Whereas hsp70 mRNA can be heat induced at the blastula stage, hsp30 is not heat inducible until the swimming tadpole stage. This is in agreement with the observation that hsp30 could not be induced in Xenopus oocytes (Bienz and Gurdon, 1982; Bienz, 1982) but in conflict with Guedon et al. (1985) who observed that it could.

Hsp30 also shows some degree of tissue specificity, its messages accumulating to levels 10-fold higher in the kidney and gut as compared to lung or liver after heat shock.

Although hsp30 genes may be repressed during early development in Xenopus, becoming responsive to heat shock in later stages, there is no report of these genes being transcribed under non-stress conditions.

1.2.4 Other Organisms

Based on similarities in molecular weight, many hsps reported in other systems may be related to the shsps already discussed. A hsp profile analogous to Drosophila melanogaster is found in another dipteran, Chironomus tentans (Tanguay and Vincent, 1981). In the skipper butterfly (Calpodes ethlius), a lepidopteran, shsps of 26 kd and 22 kd resolve into multiple components on two-dimensional gels (Dean and Atkinson, 1983). A 28 kd protein is induced with cadmium and zinc in chinook salmon embryo cells (Heikkila et al., 1982) while a pair of proteins approximately 30 kd in size are induced with heat and arsenite in rainbow trout fibroblast cells (Kothary and Candido, 1982). Heikkila et al. (1982) have observed that the metal induced protein is not induced by a temperature increase to 24$^{\text{O}}$C but this may have been below the threshold of induction since the hsps of trout are induced at 27$^{\text{O}}$C (Kothary and Candido, 1982).

Heat shock induces the synthesis of approximately 10 proteins in the molecular weight range of 23,000 to 32,000 in the slime mold Dictyostelium discoideum (Loomis and Wheeler, 1982) and in the protozoan Tetrahymena pyriformis (Guttman et al., 1980). In Tetrahymena, these shsps are also induced by deciliation or by a release from anoxia.

In Volvox cultures, approximately 6 heat inducible proteins between 18 kd and 26 kd have been observed (Kirk and Kirk, 1985). Interestingly, these proteins show different patterns when synthesized in vitro or in vivo, suggesting perhaps that post-translational modification is involved.

Fungal systems also synthesize shsps under hyperthermic conditions. A 23 kd hsp which runs as a single basic band on two-dimensional gels has been identified in Neurospora crassa (Kapoor, 1983). In yeast (Saccharomyces

cerevisiae), a 26 kd protein is synthesized during heat shock. It is also induced early in sporulation under normal growth conditions, which is reminiscent of the developmental control of the genes at locus 67B in Drosophila (Kurtz et al., 1986).

1.3 Transcriptional Control and Induction Mechanisms

The expression of shsps alters dramatically upon heat shock. The complexity of the response becomes evident when the same proteins are found to be induced by a wide variety of agents and is compounded by phenomena such as differential synthesis during certain developmental stages, non-responsiveness to stress during gametogenesis and embryogenesis, tissue specific regulation and the differential response to different stressors within the same cell type.

Despite this complexity, some important generalizations can be made. With the exception of early hsp27 and hsp26 expression in Drosophila embryos due to stored mRNA in the oocyte (Zimmerman et al., 1983), the appearance of shsps from plants to man is due to new transcriptional activity. Although it has been suggested that the induction of hsp25 in canavanine-treated CEF cells is due to the translation of existing message (White and Hightower, 1984) this is in direct conflict to inhibitor studies in which hsp25 synthesis under the same conditions is blocked by actinomycin D (Kelly and Schlesinger, 1978). Additional support for the above conclusion comes from the observation that in CEF cells, hsp25 synthesis can occur in virus infections only when the virus is defective in terminating host RNA synthesis (Collins et al., 1980; Notarianni and Preston, 1982).

1.3.1 Gene Promoter Function

The other generalization that can be made comes from the available
sequence analysis of gene promoter regions. Without exception, all heat
inducible genes including gene 1 and the genes coding for hsp22, 23, 26 and
27 in Drosophila, the genes coding for four hsp17 variants in soybean, the
human HS11 gene and the gene coding for hsp30A in Xenopus contain heat shock
elements (HSEs) upstream from the TATA motif. These elements resemble the
14 bp palindromic consensus sequence which has been shown to confer heat
inducibility upon genes (Bienz and Pelham, 1982; Pelham, 1982; Pelham and
Bienz, 1982; Mirault et al., 1982). It should be mentioned that in the
hsp30B gene of Xenopus, which may be a pseudogene, the proximal HSE has
diverged significantly compared to the hsp30A gene (Bienz, 1984b).

It is becoming more evident that non-HSE promoter function may be
responsible for transcriptional activation in the case of some inducers
other than heat shock. An understanding of multiple promoter function has
come from the detailed analysis of promoter deletions reintroduced,
preferably, into a homologous environment.

A size variant of hsp27 (hsp18.5) has been introduced into the
Drosophila genome using P-element mediated transformation (Hoffman and
Corces, 1984) where it displays heat inducibility and temporal regulation
through development identical to endogenous hsp27. Deletion analysis of
upstream sequences has demonstrated that heat inducibility and ecdysterone
responsiveness functions are separate (Hoffman and Corces, 1986). Deletions
that remove 80% of its heat inducibility (-2.1 kb to -1.1 kb) have no effect
on hsp18.5 mRNA levels in pre-pupal stages. A construct which retains only
124 bp 5' to the start of transcription completely abolishes heat

activation yet retains 30% of its hormone responsiveness. This region

contains homologies to sequences shown to be important for ecdysterone

induced expression of the Drosophila melanogaster glue genes (Muskavitch and

Hogness, 1982; McGinnis et al., 1983; Bourouis and Richards, 1985). Full

heat inducibility of the hsp27 gene may be due to the concerted effect of

multiple HSEs, 4 clustered within 125 bp approximately 300 bp upstream of

the mRNA initiation site, and 2 located 30 bp apart at a distance of greater

than 1 kb from the transcription start and located, in fact, within the

adjacent hsp23 coding region (see Figure 1). By assaying various hsp27

promoter constructs in Drosophila tissue culture cells, Riddihough and

Pelham (1986) have also found that ecdysterone and heat inducibility

elements are distinct. These authors, however, separate the functional

elements into two clusters, each covering approximately 100 bp, some 300 and

500 bp upstream of the cap site. The region between -579 and -455 is

necessary and sufficient for ecdysterone induction but is not required for

heat induction. Full heat inducibility is accomplished by the multiple HSEs

at approximately -300.

The heat inducibility and developmental regulatory functions have also

been shown to be distinct at the promoter level for the hsp26 gene (Cohen

and Meselson, 1985). Upstream regions of the hsp26 gene were fused to

bacteriophage lambda DNA and introduced into Drosophila using the P-element

system. Sequences required for the full heat shock response are located

within 341 bp upstream of the 5' end of the mRNA and contain 6 HSEs.

Sequences approximately 200 bp further upstream are responsible for the

normal ovarian and pupal responses.

Using P-element mediated transformation, Klemenz and Gehring (1986) have introduced promoter deletions of the hsp22 gene into a fly strain which synthesizes an electrophoretic variant of hsp22. The hsp22 gene contains two closely spaced HSEs which are just upstream of the TATA box while a third distal element is located a further 100 bp upstream. Although deletions retaining all three HSEs maintain full heat inducible function and full developmental expression, the authors suggest that the sequence requirements for the two modes of gene activation might not be completely congruent. A deletion containing both of the proximal HSEs still confers approximately 20% of the heat inducibility but expression during early pupal stages is undetectable. The use of different promoter sequences for hormone and heat induction has also been shown for hsp23 using hsp23/beta-galactosidase hybrid genes in Drosophila tissue culture cells (Mestril et al., 1986).

The soybean 17.5-E gene has been introduced into primary sunflower tumors with the T-DNA region of Agrobacterium tumefaciens (Gurley et al., 1986). A construct containing 3.2 kb of upstream sequences is strongly induced by heat shock and arsenite and to a lesser degree by cadmium. Once again, additional analysis will be required to ascertain whether the arsenite and/or cadmium responsive functions can be separated from the heat inducible ones.

Cadmium and other heavy metal ions may exert their influence on heat shock gene induction through sequences similar to the metal ion response element (MRE) which has been shown by deletion analysis to be essential for induction of the human metallothionein-IIA gene by cadmium (Karin et al., 1984). Significant homologies to this region have been reported in the

soybean 17.5-E gene as two repeats between the TATA motif and the transcription start (Czarnecka et al., 1985).

The fact that arsenite induces the expression of proteins which are not heat inducible in rat myoblast cells (Kim et al., 1983), in chick embryo muscle tissue (Kim et al., 1983) and in rainbow trout fibroblast cells (Kothary and Candido, 1982) suggests that arsenite, as well, may be inducing gene activity through sequences other than HSEs. Furthermore, the hsp25 gene is induced by arsenite, copper and zinc in quail myotubes which have lost their thermal sensitivity (Atkinson et al., 1983).

1.3.2 Heat Shock Regulons

The number and variety of ways in which shsps can be induced strongly suggests that a common mechanism may be involved for many of them. Consistent with this model would be the activation of a factor during cellular stress or damage which initiates transcription of a battery of genes (regulon) through a common sequence element.

A heat shock transcription factor (HSTF) has been partially purified from nuclear extracts of Drosophila tissue culture cells which is specifically required for transcription of a Drosophila hsp70 gene in vitro (Parker and Topol, 1984). DNAseI footprinting data demonstrates that HSTF has a high affinity for HSEs in vitro (Parker and Topol, 1984; Topol et al., 1985). Cooperative binding occurs at a pair of adjacent HSEs located in the Drosophila hsp70 gene (Topol et al., 1985). Both are required for maximal heat inducibility in vivo (Dudler and Travers, 1984; Simon et al., 1985) and in Drosophila melanogaster tissue culture cells (Amin et al., 1985). This region is also protected by HSTF from exoIII nuclease digestion in nuclei

purified from heat shocked <u>Drosophila</u> cells but not in nuclei isolated from cells grown at normal temperature (Wu, 1984).

The procaryotic counterpart to HSTF is a 32 kd sigma factor (sigma 32) that stimulates transcription initiation from heat shock promoters (Grossman <u>et</u> <u>al</u>., 1984; Landick <u>et</u> <u>al</u>., 1984; Yura <u>et</u> <u>al</u>., 1984). Interestingly, a consensus sequence derived from six of the known heat shock genes of <u>Escherichia</u> <u>coli</u> consists of a -35 element which contains one half of the eucaryotic HSE concensus (Cowing <u>et</u> <u>al</u>., 1985).

Mutations within the gene coding for sigma 32 (<u>rpoH</u> gene) have shown that the heat shock genes are under the positive control of sigma 32. A nonsense mutation in a strain (<u>rpoH165</u>) carrying a tRNA that suppresses amber mutations at 28°C but not 42°C, eliminates the synthesis of hsps after a shift to the high temperature and, in fact, the cells die after a brief period at 42°C (Neidhart and Van Bogelen, 1981; Yamamori and Yura, 1982).

Significant amounts of HSTF and sigma 32 are, however, found in normally growing cells (Beckman and Cooper, 1973; Parker and Topol, 1984). Whether or not they are required under normal conditions is not yet known. If the <u>rpoH</u> gene carrying a nonsense mutation is placed in a non-suppressing environment, the one hsp analyzed (groE) is still synthesized at 30°C although no induction is observed at 42°C (Yura <u>et</u> <u>al</u>., 1984). This suggests that sigma 32 is dispensable at low temperatures.

HSTF is probably activated during heat shock since HSTF derived from heat shocked cells is more active in transcription assays (Parker and Topol, 1984).

An indication that sigma 32 may be found in 2 different forms comes from the observation that sigma 32 from cell lysates migrates as 2 spots on two-dimensional gels, but only as 1 spot when copurified with RNA polymerase (Grossman et al., 1984).

1.3.3 Activation of Transcription Factors

The evidence already presented suggests that an activated transcription factor interacts with genes within the heat shock regulon to initiate coordinate expression.

Recently, several small molecules have been identified in Salmonella typhimurium and E. coli that accumulate under heat shock or ethanol stress. The intracellular concentration of five adenylated nucleotides (AppppA, AppppG, ApppG, ApppA, ApppGpp) increases 5 - 10 fold within 5 minutes after insult (Lee et al., 1983). The observation that these nucleotides still accumulate in the absence of hsp synthesis in the rpoH165 mutant led the authors to suggest that they were alarmones signalling the onset of oxidative stress and triggering the heat shock response. The accumulation of only ApppA occurs in heat shocked yeast cells (Denisenko, 1984).

Adenylated nucleotides are believed to be synthesized in vivo by a side reaction of aminoacyl-tRNA synthetases (Rapaport et al., 1975). Curiously, one of the hsps in E. coli is an inducible species of lysyl-tRNA synthetase encoded by the lysU gene (Van Bogelen et al., 1983).

Many observations, however, conflict with this model. The kinetics of the accumulation of these alarmones is not consistent with the fact that in bacteria, transcription after heat shock is observed within one minute and

has reached a maximum level and begun to decrease within 5 minutes of the temperature shift. Injection of Xenopus oocyte nuclei with either ApppA or ApppA does not induce hsp synthesis (Guedon et al., 1985). Also, in Xenopus oocytes the accumulation of ApppA to 10 times its basal level occurs only under severe heat conditions such as 45°C whereas hsp synthesis readily proceeds at 33°C (Guedon et al., 1985). Finally, treatment of Drosophila tissue culture cells with cadmium chloride under conditions which induce the heat shock response (Courgeon et al., 1984) does not significantly affect the cellular pool of ApppN and ApppN nucleotides (Brevet et al., 1985).

There is mounting evidence that the heat shock response is intimately associated with protein degradation systems. This should not be surprising since most of the known inducers must indeed lead to protein damage.

In some flightless Drosophila mutants, molecular lesions in actin III isoforms leads to the disruption of myofibrils in indirect flight muscle, the only tissue in which they are expressed. This is associated with constitutive hsp synthesis only within the same cell type (Karlik et al., 1984; Hiromi and Hotta, 1985). P-element mediated transformation with these mutant actin genes also results in constitutive expression of the heat shock genes within the indirect flight muscle (Hiromi et al., 1986).

A mouse cell line called ts85 contains a thermolabile ubiquitin-activating enzyme (E-1) (Finley et al., 1984). At the non-permissive temperature of 39°C, short-lived proteins and abnormal proteins are not efficiently degraded since they do not become ubiquitinated (Ciechanover et al., 1984). At this temperature, which is well below the normal threshold of hsp induction, ts85 cells synthesize hsps at high levels (Ciechanover et al., 1984).

In E. coli, the synthesis of large amounts of aberrant polypeptides due to canavanine incorporation, the production of truncated proteins in the presence of puromycin, the induction of translational errors with streptomycin, or the synthesis of a cloned foreign protein (human tissue plasminogen activator) leads to the induction of the heat shock response (Goff and Goldberg, 1985). This phenomenon is not seen in the rpoH165 mutant in which the sigma factor specific to the heat shock regulon is defective.

The induction of the heat shock response by amino acid analogues is, presumably, a direct result of their incorporation into newly synthesized proteins, rendering them nonfunctional. Homoarginine, which is not incorporated into proteins (Nazario and Evans, 1974), does not induce a heat shock response in CEF cells (Hightower, 1980).

The most direct evidence for a link between protein degradation and the heat shock response comes from the work of Ananthan et al. (1986). A Drosophila hsp70/beta-galactosidase hybrid gene is activated in Xenopus oocytes when co-injected with bovine beta-lactoglobulin or bovine serum albumin that has been denatured by reductive carboxymethylation. The native proteins have no effect. Native hemoglobin also has no effect, but globin monomers which result from extraction of the heme groups do. Furthermore, the analysis of promoter deletions demonstrates that the proximal HSE is required for both heat shock induction and for activation by denatured protein.

All of these results suggest that a depletion of activated ubiquitin pools in eucaryotes and the overburdening of protein degradation systems in procaryotes leads to the activation of the heat shock response.

Interestingly, E. coli protease La is a hsp under the regulation of sigma 32 (Goff et al., 1984; Phillips et al., 1984; Baker et al., 1984). Similarly, it has been shown that ubiquitin mRNA and ubiquitin synthesis increase 5-fold after heat shock in chicken embryo fibroblasts (Bond and Schlesinger, 1985).

In eucaryotes, ubiquitination of damaged proteins could deplete activated ubiquitin pools, mimicking the lesion found in ts85 cells. Under these conditions, all available ubiquitin stores would be mobilized to degradation pathways. This is consistent with the observation that ubiquitinated histone H2A is depleted in ts85 cells at the non-permissive temperature (Matsumoto et al., 1983; Finley et al., 1984) or in Drosophila cells under heat shock (Glover, 1982a). Thus it is possible that during heat shock or metabolic stress, HSTF may be activated by the conversion of a ubiquitinated form to a non-ubiquitinated form although there is no direct evidence to support this as yet. Alternatively, HSTF and sigma 32 may have a high turnover in control cells, being susceptible to protein degradation pathways. Under metabolic stress, general cellular protein damage would divert the pathway and HSTF would be stabilized. Similarly, a defective protein degradation system such as in ts85 cells at $39^{o}C$, would stabilize the transcription factor and lead to hsp synthesis.

1.3.4 Differential HSmRNA Stability

Although the heat shock response is characterized by new transcriptional activity, variability in accumulated mRNA levels due to differences in stability may result in altered patterns of hsp synthesis. The synthesis of shsps in heat shocked versus ecdysterone treated

Drosophila _melanogaster_ tissue culture cells is one example of this. The ratio of hsp23 to hsp22 varies from 3 during heat shock to 20 during ecdysterone treatment (Vitek and Burger, 1984). Under heat shock conditions, the mRNA levels for hsp23, 22, 26 and 27 increase rapidly for 30 minutes and plateau at similar levels. With hormone treatment, however, hsp23 mRNA increases steadily for 24 hours, reaching a level approximately 50% that of heat shock. Hsp22 mRNA, on the other hand, increases during the first 4 hours and reaches a plateau so that by 24 hours, the ratio of hsp23 mRNA to hsp22 mRNA is approximately 6. Under the same conditions hsp26 and hsp27 transcripts rise to intermediate levels. Vitek and Burger (1984) further showed by pulse-chase experiments that the variability in steady-state mRNA levels is due to differences in mRNA stability. Interestingly, hsmRNAs are more stable during a chase at 35°C as compared to 25°C.

1.4 Translational Control

The changing patterns of protein synthesis during heat shock usually include the repression of existing protein synthesis as well as the appearance of heat shock proteins. The degree to which this occurs depends on the cell type in question and the nature or severity of the stress. For example, there is no cessation of translation of existing messages upon heat shock in _Xenopus_ somatic cells as there is in _Xenopus_ oocytes (Bienz, 1982). Storage proteins in the soybean embryo (seed) continue to be synthesized at high temperatures along the hsps (Altschuler and Mascarenhas, 1982). In _Drosophila_ cells, heat shock, arsenite, and canavanine treatments

vary in their ability to shut down normal protein synthesis even though hsps
are induced in all three situations (Olsen et al., 1983).

These changes in protein synthesis might occur via changes in the
specificity of the translational machinery so that a subset of mRNAs is
translated more efficiently than another, or by changes in the mRNA pools
that are available to the ribosomes. Both of these mechanisms are used, the
latter situation being utilized in yeast.

1.4.1 HSmRNA Selectivity

In most organisms pre-heat shock mRNAs (normal mRNAs) are stable under
heat shock conditions since they are efficiently translated in a variety of
in vitro systems but not in vivo. In Drosophila cells (Storti et al., 1980)
and HeLa cells (McCormick and Penman, 1969; Hickey and Weber, 1982) normal
protein synthesis resumes in the presence of actinomycin D upon return to
normal temperature suggesting that translation takes place on existing
messages which have been inactive during stress. Efficient translation of
normal mRNAs in vitro further suggests that they do not become modified in
any way even though they are preferentially ignored by the translational
machinery in vivo.

Polysome breakdown is an immediate response to heat shock. This has
been shown for Drosophila (McKenzie et al., 1975) and soybean (Key et al.,
1981). In Drosophila, the percentage of ribosomes in monosome form compared
to polysome form increases from approximately 20% to 50% during heat shock
(Lindquist, 1980; Ballinger and Pardue, 1983). These changes are not due to
a flood of newly synthesized hsmRNA since they occur in actinomycin D
treated cells (Lindquist, 1980). As discussed earlier, a bimodal

distribution of polysomes is found in heat shocked <u>Drosophila</u> cells, which reflects the commitment of these cells to the translation of the two major size classes of hsmRNA. <u>In vitro</u> translation of polysomal mRNA from heat shocked <u>Drosophila</u> cells (Kruger and Benecke, 1981; Ballinger and Pardue, 1983) and azetidine (proline analogue) treated HeLa cells (Thomas and Mathews, 1982) demonstrates, however, that a significant proportion of normal messages are still associated with polysomes even though they are not being translated. Using specific hybridization probes, it was shown that alpha-tubulin, beta-tubulin and actin fall into this category (Kruger and Benecke, 1981).

Ballinger and Pardue (1983) have estimated that there is a 15 – 30 fold decrease in the initiation/elongation rates of ribosomes on normal mRNAs in heat shocked <u>Drosophila</u> cells. In HeLa cells, hsp translation is more sensitive to cycloheximide suggesting that hsp mRNAs initiate translation more efficiently than most normal mRNAs (Hickey and Weber, 1982).

The preferential translation of hsmRNAs can be reproduced <u>in vitro</u> using cell lysates prepared from heat shocked <u>Drosophila</u> tissue culture cells (Storti <u>et</u> <u>al</u>., 1980; Kruger and Benecke, 1981). These lysates have an optimum temperature for protein synthesis of 28°C suggesting that a stable change takes place during heat shock at 36°C. A lysate prepared from cells grown at 25°C translates both normal mRNAs and hsmRNAs.

Fractionation of lysates and subsequent supplementation experiments demonstrated that the ribosomal fraction (222,600 x g pellet) of control lysates could rescue normal mRNA translation in heat shock lysates (Scott and Pardue, 1981). The ability to rescue normal translation was diminished by a 0.5 M KCl wash. None of the heat shock lysate fractions could cause

the 25°C lysate to change its specificity suggesting that there is a

negative control or inactivation of a component in the heat shock lysate.

A candidate for this type of modification is the dephosphorylation of

ribosomal protein S6 during heat shock which has been shown to occur in

Drosophila (Glover, 1982b), in suspension cultures of tomato (Scharf and

Nover, 1982), in primary cultures derived from human skin fibroblasts and

meningiomas (Richter, 1983), as well as in human HeLa and baby hamster

kidney cells (Kennedy et al., 1984). The observations made by Olsen et al.

(1983) are inconsistent with this simple correlation: in their studies, the

rephosphorylation of S6 in Drosophila cells was found not to occur until

recovery had proceeded for 8 hours, well after normal protein synthesis had

resumed. Also, arsenite and canavanine treatment of Drosophila cells

induced hsp synthesis but dephosphorylation of S6 was not observed.

Recently, it was reported that the ribosomal supernatant from 25°C

Drosophila cell lysates could rescue the translation of normal mRNAs in heat

shock lysates (Sanders et al., 1986). The fact that these authors used

higher ionic strength buffers in their ribosomal purifications is consistent

with the initial observation made by Scott and Pardue (1981) that rescue was

reduced with salt washed ribosomes. Sanders et al. (1986) also showed that

a reconstituted system containing heat shocked ribosomes and a control

supernatant had the ability to synthesize normal proteins.

The change in the specificity of translation during heat shock may

therefore depend on soluble factors such as translation initiation factors.

Detailed immunoblot analysis of HeLa cell lysates demonstrates that a

variety of modifications occur in various initiation factors upon heat shock

(Duncan and Hershey, 1984). Also, phosphorylation of eIF-2 alpha occurs in reticulocyte lysates heated to 42°C (Ernst et al., 1982).

In Drosophila cells, some normal messages escape the translational selectivity during heat shock. Histone H2B synthesis increases 3 - 4 fold during heat shock (Sanders, 1981). Although transcription of the histone gene is relatively unaffected, a greater abundance of histone H2B specific mRNA is found in polysomes isolated from heat shocked cells as compared to control cells (Farrell-Towt and Sanders, 1984). Also, in Drosophila Schneider 2 cells infected with the double stranded DNA virus HPS-1, HPS-1 specific proteins continue to be synthesized during heat shock (Scott et al., 1980).

The sequences required for high temperature translation are now under investigation. Drosophila hsp70 genes with deleted leader sequences or fusions between hsp70 promoter sequences and an alcohol dehydrogenase (AdH) gene containing its own leader and coding regions have been introduced into Drosophila tissue culture cells (McGarry and Lindquist, 1985) and flies (Klemenz et al., 1985). In both cases, transcripts accumulate to high levels during heat shock but they are not translated until the cells are returned to normal temperature. These results suggest that the 5' untranslated leader sequences of hsp mRNAs are required for preferential translation at elevated temperature. Initial observations (DiNocera and Dawid, 1983; Lawson et al., 1984) suggested that only the terminal sequences of heat shock mRNA leaders fulfilled this function.

The fusion of only the first 95 bp of the hsp70 250 bp leader sequence to the wild type AdH leader is sufficient to direct translation at high temperature (Klemenz et al., 1985). Deletions of the same leader between +3

and +26 or between +14 and +114 have no effect on wild type translation (McGarry and Lindquist, 1985). Surprisingly, a tandem duplication between -29 and +2 which results in an additional 39 bp attached to the 5' end of a perfectly normal 250 bp leader, destroyed the translational selectivity of that mRNA upon heat shock (McGarry and Lindquist, 1985).

Similar experiments demonstrate that only approximately the first 30 bp of the hsp22 mRNA 250 bp leader sequence are required for its translation during stress (Hultmark et al., 1986).

1.4.2 High mRNA Turnover

In yeast (Saccharomyces cerevisiae), the pattern of protein synthesis alters rapidly upon a shift from $23^{o}C$ to $36^{o}C$ even though both temperatures are considered to be within its normal growth range (McAlister and Finkelstein, 1980a; Lindquist, 1981). In this organism, however, the in vitro translation pattern for RNA isolated from heat shocked cells corresponds to that seen in vivo (McAlister and Finkelstein, 1980a; Lindquist, 1981). This suggest that the reduced synthesis of particular proteins during heat shock correlates with the degradation of their mRNAs. Hybridization of specific gene probes to mRNA supports this since ura-3 mRNA levels decrease 5 fold within 1 hour while lev-2 mRNA levels decrease to 5% that of normal during the same time period (Lindquist, 1981).

Both the appearance of hsmRNAs and the rapid disappearance of normal mRNAs are dependent upon new transcription. If transcription is inhibited by the zinc-chelating antibiotic lomofungin or if tsrna1 strains are used in which RNA transport/processing is defective at $36^{o}C$, then repression of protein synthesis is greatly reduced (Miller et al., 1979; McAlister and

Finkelstein, 1980a). This phenomenon, however, does not require protein

synthesis since cycloheximide treatment or the use of strains which have a

temperature sensitive defect in translation initiation (ts187) have no

effect on changing mRNA pools during heat shock as assayed by in vitro

translation (McAlister and Finkelstein, 1980a).

1.5 Recovery and Autoregulation

The effects of heat shock and other stress conditions are always

reversible upon return to normal temperature, or to a lesser and more

variable degree upon prolonged exposure to stress depending on the

adaptibility of the cell type.

In most organisms such as Drosophila, mammals, and higher plants, the

return to normal protein synthesis is probably simply a matter of the

resumption of translation of existing messages, since recovery is

insensitive to inhibitors of transcription. Yeast, on the other hand,

requires new mRNA synthesis in order to recover the pre-existing protein

synthesis pattern.

The situation in avian systems, specifically in chick embryo

fibroblasts, presents an enigma. Although in vitro translation results

indicate that normal mRNAs are present in heat shocked cells (Kelly et al.,

1980; Johnston et al., 1980; Voellmy and Bromley, 1982), recovery is blocked

by actinomycin D (Hightower, 1980; Schlesinger et al., 1982). If

cycloheximide is present during an 8 hour recovery period and removed prior

to labelling of newly synthesized protein, the heat shock profile is seen,

suggesting that there is also a need for new protein synthesis in order for

recovery to proceed (Hightower, 1980; Schlessinger et al., 1982).

In <u>Drosophila</u>, all normal mRNAs resume translation at the same rate in recovering cells; on the other hand, the repression of hsp synthesis is a-synchronous and occurs in a reproducible order (DiDomenico <u>et al</u>., 1982a). Although recovery times can vary depending on the severity of the shock, the repression of hsp70 always occurs first, hsp83 last and the shsps in between. The repression of hsp70 is correlated with the recovery of normal protein synthesis. Repression of hsp synthesis during recovery is a result of hsmRNA degradation in <u>Drosophila</u> (Lindquist, 1980; DiDomenico <u>et al</u>., 1982a) and in HeLa cells (Hickey and Weber, 1982).

The autoregulation of the heat shock response has been demonstrated in canavanine treated <u>Drosophila</u> cells (DiDomenico <u>et al</u>., 1982b). In these cells, there is no recovery of normal protein synthesis at 25°C after a 1 hour heat shock and hsp synthesis continues unabated. Also, if cycloheximide is added to normally growing cells immediately before heat shock, synthesis of hsmRNAs continues for at least 4 hours upon recovery at 25°C where they remain stable (DiDomenico <u>et al</u>., 1982b). The same authors also demonstrated that when the rate of hsp synthesis is limited by decreasing the concentration of hsmRNA with actinomycin D, both the repression of hsp synthesis and the restoration of normal synthesis are delayed, apparently until a specific amount of functional hsp has accumulated.

If <u>Drosophila</u> cells are maintained at high temperature for prolonged periods, repression of hsp synthesis and the onset of normal protein synthesis initiates upon the accumulation of hsps as before; however, the repression of hsp synthesis is prolonged due to increased hsmRNA stabilities at high temperature (DiDomenico <u>et al</u>., 1982b).

The kinetics of hsp70 repression and its close correlation with the recovery of normal protein synthesis implicates it as the protein directly involved in the feedback control mechanism which seems to be operable in the heat shock response. In E. coli, this hypothesis is substantiated by the following observations. The dnaK gene of E. coli is a heat shock gene whose product is approximately 50% homologous to Drosophila hsp70 (Bardwell and Craig, 1984). Temperature sensitive dnaK mutants such as dnaK756 fail to turn off the transient heat shock response at 43°C while bacteria that overproduce dnaK protein at all temperatures undergo a drastically reduced heat shock response (Tilly et al., 1983). The role played by the shsps in autoregulation, if any, remains to be elucidated.

An analysis of hsmRNA sequences which are required for their own destabilization during recovery has begun. Drosophila strain Df(3R) kar^{D2} contains an X-ray induced 3' deletion mutation of hsp70 and synthesizes a truncated protein (hsp40) during heat shock. In these flies or in tissue culture cells transfected with the truncated gene, hsp40 mRNA persists during recovery whereas endogenous hsp70 mRNA decreases rapidly (Simcox et al., 1985). These data suggest that the 3' sequences of hsp70 mRNA are involved in the active destabilization of the hsp70 mRNA after release from heat shock.

1.6 Function of shsps

Sequence information for the shsp genes of Drosophila, soybean, Xenopus and man indicates that the proteins share a domain of approximately 80 amino acids with α-crystallin chains of vertebrates. Also, anti-chicken hsp25 reacts with a protein derived from 11 day old embryonic chicken lens. The

significance of this homology is not clear. Vertebrate α-crystallins are the major protein components of the vertebrate eye lens, forming large aggregates with an average molecular weight of 800,000 (Bloemendal et al., 1971). The aggregation properties of alpha-crystallins are interesting in light of the observations made for shsps during cellular localization studies (see below).

Ingolia and Craig (1982b) postulated that the region of homology between α-crystallin and the shsps of Drosophila represented a domain that promoted aggregation. Wistow (1985) has suggested that the domain represents a thermodynamically stable structure which pre-existed the lens and was borrowed from ancestral heat shock genes to build a protein capable of surviving for years without turnover in the enucleated, avascular lens (Wannemacher and Spector, 1968).

1.6.1 Intracellular Localization

Early autoradiographic studies on Drosophila salivary glands (Mitchell and Lipps, 1975; Velazquez et al., 1980) and tissue culture cells (Velazquez et al., 1980; Arrigo et al., 1980) showed that newly synthesized hsps were rapidly transported to the nucleus during heat shock. Subsequent cellular fractionation studies with Drosophila Kc cells confirmed that hsp22, 23, 26 and 27 were associated with nuclear fractions during heat shock and were translocated to the cytoplasm during recovery (Arrigo et al., 1980; Tanguay and Vincent, 1982). Immunofluorescence studies using an antibody specific for hsp23 revealed a similar pattern (Arrigo and Ahmad-Zadeh, 1981). The association of the shsps with the nuclear pellet is resistant to extensive

nuclease digestion and to 2.0 M salt treatment (Sinibaldi and Morris, 1981; Levinger and Varshavsky, 1981).

A nuclear function for the shsps is now in doubt, however, due to a better understanding of the changes which occur within the cellular intermediate filament (10 nm) network during heat shock. It was observed that during heat shock in Drosophila, a 46 kd and a 40 kd protein, normally found in microsomal fractions, became enriched in nuclei (Faulkner and Biessman, 1980; Tanguay and Vincent, 1982). An antibody made to the 46 kd protein cross-hydridized to the 40 kd protein as well as to proteins with molecular weights of 55,000 and 52,000 which have been identified as vimentin and desmin, respectively, in baby hamster kidney cells (Faulkner et al., 1981). These proteins are major components of the vertebrate intermediate filament cytoskeletal system.

Immunofluorescence studies, using the antibody made against the Drosophila 46 kd intermediate filament protein, showed that, upon heat shock, the protein was localized in the peripheral region of the nucleus (Faulkner et al., 1981). Immunoelectron microscopy further characterized the intermediate filament structure of Drosophila and confirmed that it collapses upon the nucleus after heat shock (Walter and Biessman, 1984). Furthermore, antibodies made against hsp23 and hsp26 behave like antibodies made to the 46 kd vimentin-like protein in their subcellular localization, (Leicht et al., 1986), suggesting that the shsps are associated with the intermediate filament network. Thus the apparent translocation of shsps into the nucleus upon heat shock may be artifactual due to the collapse of the cytoskeletal structure, throwing into doubt the validity of earlier fractionation results.

Support for this explanation comes from the observation that the shsps are found entirely in the cytoplasmic fraction of ecdysterone treated Drosophila larval imaginal disks (Ireland et al., 1982). Interestingly, the crystallins show immunological cross-reactivity with keratins which are also members of the intermediate filament protein family (Kodama and Eguchi, 1983).

In Drosophila, hsp22, 23, 26 and 27 become associated with cytoplasmic RNP particles which sediment at approximately 20S in sucrose density gradients (Arrigo et al., 1985; Shuldt and Kloetzel, 1985). The shsps are detected in control cells at levels less than 10% of the amount which accumulates after 6 hours of recovery from heat shock. These RNP particles are characterized by the presence of 16 – 20 proteins in the 20 – 30 kd range on two-dimensional gels, contain a set of small RNA species between 50 – 200 nucleotides and have a buoyant density of 1.365 – 1.380 g/cm^3 in CsCl after UV crosslinking (Arrigo et al., 1985; Shuldt and Kloetzel, 1985). These particles are very similar to the prosome which has been characterized in duck erythroblasts and mouse erythropoietic cells (Schmid et al., 1984). Under the electron microscope they appear as ring shaped particles, 12 nm in diameter (Arrigo et al., 1985; Shuldt and Kloetzel, 1985).

A 74 nucleotide long RNA associating with the Drosophila prosome has been sequenced (Arrigo et al., 1985) and shown to be homologous to mammalian U6 small nuclear RNA (Ohshima et al., 1981; Harada et al., 1980; Epstein et al., 1980).

In suspension cultures of tomato, the shsps form cytoplasmic heat shock granules (hsg) which are resistant to RNase, 0.5 M KCl, EDTA, detergent and

sonication (Nover et al., 1983). These are different from Drosophila prosomes in many respects. First of all, cytoplasmic granules are not seen in the cytoplasm of Drosophila cells. In tomatoes, hsgs undergo a massive and rapid accumulation upon heat shock and disappear slowly during recovery. In contrast, the prosome is found in equal amounts in both control and heat shocked cells, the amount of associated shsps increasing substantially after heat shock.

The function of hsgs in plants and of prosomes in animal cells is not known.

1.6.2 Thermotolerance: Role of shsps

Thermotolerance was first used to describe the phenomenon in which an initial mild, non-lethal heat treatment causes cells to be resistant to a brief shift to higher temperatures which are normally lethal. In Drosophila, a mild heat treatment at 35°C for 50 minutes protects all developmental stages and tissue culture cells from cell death at 40.5^{o}C for 20 minutes (Mitchell et al., 1979). In Chinese hamster fibroblast cells, a short 46°C treatment followed by a recovery period of 4 - 6 hours or treatment at 41°C for several hours protected the cells from death at 45°C for 45 minutes (Li and Werb, 1982). In yeast, a shift to 36°C for 90 minutes results in a transient protection from death at 52°C (McAlister and Finkelstein, 1980b). In all of the pre-treatments described, it was demonstrated that hsp synthesis occurs. Furthermore, the persistence of thermotolerance for up to 36 hours after heat shock correlates well with the persistence of heat shock proteins in hamster cells (Li and Werb, 1982).

Many inducers of the heat shock response also induce thermotolerance. Prior exposure of hamster cells to arsenite and ethanol or a release from anoxia results in an acquired tolerance to a subsequent lethal heat challenge (Li and Werb, 1982). If amino acid analogues such as canavanine or azetidine are used to induce hsp synthesis in hamster cells, then thermotolerance is not acquired (Li and Lazlo, 1984), presumably because hsps incorporating analogues have altered function. Ethanol treated yeast cells, in which hsp synthesis is less than in heat shocked cells, are correspondingly less thermotolerant (Plesset et al., 1982). Exposure of yeast cells to ionizing radiation (gamma-rays) also induces thermal resistance (Mitchel and Morrison, 1982) although it is not known whether there is any heat shock protein synthesis under these conditions.

These findings show a strong correlation between the expression of the hsps and the development of thermotolerance but they do not prove that these events are functionally related. The requirement for heat shock induced transcription in the acquisition of thermotolerance has been demonstrated in ts mutants of yeast in which a defect in RNA transport/processing results in no increase of thermotolerance after heat shock. Thermotolerance is acquired, however, if the cells are allowed to recover following a heat shock (McAlister and Finkelstein, 1980b).

Protein synthesis inhibitor studies have also been used to address this question. The elimination of thermotolerance with cycloheximide treatment prior to a pre-treatment has been demonstrated in yeast by several workers (McAlister and Finkelstein, 1980b; Mitchel and Morrison, 1982; Craig and Jacobsen, 1984) but not all (Hall, 1983). Hall (1983) also showed that yeast cells treated with a phenylalanine analog did not become

thermotolerant in agreement with experiments on analog treated hamster cells (Li and Lazlo, 1984). However, if these cells were given a heat treatment at 37°C, a condition in which hsps should still be non-functional, thermotolerance was acquired. Furthermore, whereas cycloheximide has been shown to block the thermotolerant state in Dictyostelium (Loomis and Wheeler, 1982), treatment of rat embryonic fibroblast (Rat-1) cells with cycloheximide for 6 hours at 37°C after a 20 minute interval at 45°C inhibits protein synthesis, including hsp synthesis, but has no effect on subsequent survival at 45°C (Widelitz et al., 1986). The reasons for these discrepancies are not known.

The strongest evidence for a function of hsps in thermotolerance is found in the rpoH165 mutant of E. coli in which hsps are not synthesized. These cells also fail to acquire resistance to a 55°C challenge after a 42°C pre-treatment (Yamamori and Yura, 1982).

Some observations suggest that the shsps may play a role in thermotolerance. In Drosophila cells, tolerance is acquired by continued exposure to ecdysterone in which only the shsps are induced (Berger and Woodward, 1983). Also, Drosophila pupae (a developmental stage in which the shsps are specifically synthesized) appear to be more resistant to heat induced death (Mitchell et al., 1979; Berger and Woodward, 1983).

A mutant cell line of Dictyostelium discoideum called HL122 has been identified due to its defect in acquiring thermotolerance (Loomis and Wheeler, 1982). These cells do not synthesize any of the shsps. However, even though hsp70 is present in heat shocked HL122 cells, it accumulates to levels well below those found in wild type cells and may affect the ability of these cells to develop thermotolerance.

The acquisition of thermotolerance in heat shocked Xenopus laevis embryos argues against a role for hsp30 in this phenomenon (Heikkila et al., 1985). Thermotolerance is developed by the late blastula and early gastrula stages in which the hsp30 genes are not responsive to heat shock (Bienz, 1984a). Also, in yeast strains containing an inactive hsp26 gene, thermotolerance is, nevertheless, developed in log phase or in stationary phase cells, in mature or germinating spores, and during spore development (Petko and Lindquist, 1986).

1.6.3   Inhibition of the Heat Shock Response

Both deuterium oxide and polyhydroxyl alcohols such as glycerol block the induction of hsp synthesis by heat or arsenite in chick embryo fibroblasts (Hightower et al., 1985). Both of these compounds are known to stabilize macromolecules and to protect cells from thermal killing. They may protect stress-sensitive proteins from irreversible denaturation, possibly by direct interaction but more likely by generalized solvent effects. It has been suggested that hsps, which are resistant to denaturation by heat or ethanol, can non-specifically stabilize other proteins that are highly susceptible to inactivation (Minton et al., 1982). This is consistent with the idea that hsps play a role in protecting cells against stress and in the acquisition of thermotolerance. It also supports the model whereby protein damage may trigger the heat shock response, as discussed above.

1.7   The Biology of Caenorhabditis elegans

Caenorhabditis elegans, a member of the family Rhabditidae, is a microbivorous, free-living soil nematode. Adult nematodes are found as males or self-fertilizing hermaphrodites, the latter being slightly larger and reaching a length of 1 mm. The life cycle of C. elegans is rapid and takes about 3.5 days at $20^{\circ}C$. After fertilization, the eggs begin to cleave within the hermaphrodite and are laid at about the 30-cell stage (mid-gastrula). Each hermaphrodite produces 200 - 300 progeny. After embryogenesis, a juvenile containing about 550 cells hatches from the egg case and develops through four larval stages, L1-L4, before reaching the adult stage. Many cell divisions occur during the larval period, the somatic cell number increasing to about 1000 in mature adults. The first larval stage contains only two germ line cells while the adults contain 1000 - 2000 germ line nuclei. L2 stage larvae can transform into a resistant stage known as dauer larvae under adverse environmental conditions. In this state, the nematode can tolerate starvation for many months, resuming normal development when nutrients become available.

Since C. elegans is transparent, development of the living organism can be observed with a light microscope using Nomarski optics. In this manner, all of the cell divisions, deaths, and migrations that produce a mature adult from a single egg have been determined (Sulston et al., 1983). This achievement is a first for an organism of this degree of complexity, and has been facilitated also by the fact that nematode development is invariant. This means that the fate and position of a given precursor cell is the same in all individuals. The anatomy and wiring of the complete nervous system, which is composed of only 302 neurons, has also been determined.

C. elegans is ideally suited to genetic analysis. Thousands of
mutations have been mapped to about 500 genes. Many include genes which
affect cell fates during development, and others which affect neuromuscular
function.

C. elegans is a diploid organism, containing five autosomal chromosomes
and one sex (x) chromosome. Hermaphrodites contain two sex chromosomes
while males carry only one. The haploid DNA content is 8×10^7 bp per
genome (Sulston and Brenner, 1974), which is only 20 times that of E. coli.
Repetitive DNA makes up only 17% of the total DNA content which has an
overall low G/C base composition of 36% (Sulston and Brenner, 1974). The
simplicity of the nematode genome has sparked a cooperative effort to map
the entire C. elegans genome. Presently, about 65% of the genome has been
categorized into about 900 segments, many greater than 200 kb and one
greater than 600 kb.

1.8 The Present Study

By the end of 1980, the heat shock response had been well characterized
in Drosophila (see review by Ashburner and Bonner, 1979), but it was
becoming evident that the phenomenon was conserved in a diverse group of
organisms including Escherichia coli (Lemaux et al., 1978; Yamamori et al.,
1978), protozoans such as Naegleria gruberi (Walsh, 1980) and Tetrahymena
pyriformis (Fink and Zeuthen, 1980; Guttman et al., 1980), slime molds
including Dictyostelium discoideum (Loomis and Wheeler, 1980) and
Polysphondylium pallidum (Francis and Lin, 1980), yeast (McAlister and
Finkelstein, 1980a; 1980b), sea urchin (Guidice et al., 1980), another
dipteran Chironomus tentans (Vincent and Tanguay, 1979), plants including

tobacco and soybean (Barnett et al., 1980), chinese hamster ovary cells (Bouche et al., 1979) and chick embryo fibroblasts (see Table II, section 1.2.2).

The inducibility of the heat shock genes to high levels, in a reversible fashion, made them attractive models for the study of gene activation. The late seventies brought about intense activity directed towards the cloning and characterization of the major heat inducible genes of Drosophila and at the end of 1980, Ingolia et al. (1980) reported the DNA sequence of one complete hsp70 gene along with the comparison of three hsp70 5' flanking sequences.

It was at this time that the present investigation was undertaken to characterize sequences coding for the shsps of the nematode Caenorhabditis elegans. Until then there had been no sequence information available regarding the shsp genes of any organism including Drosophila. In the next five years it would become clear that the shsps from a variety of systems were related and highly conserved throughout evolution. During the same period, other workers were focusing their attention on the hsp70 gene family of Caenorhabditis elegans (Snutch and Bailie, 1983; Snutch and Baillie, 1984).

II.   EXPERIMENTAL PROCEDURES

2.1   Maintenance of Nematodes

Caenorhabditis elegans strains Bristol (N2) and Bergerac (BO) were

maintained on NG plates (0.3% NaCl, 0.25% bactotryptone, 5.0 µg/ml

cholesterol, 1.0 mM $CaCl_2$, 1.0 mM $MgSO_4$ and 25 mM KH_2PO_4 pH 6.0)

containing E. coli OP50 as described by Brenner (1974).   OP50 is a uracil

requiring mutant which prevents the overgrowth of the bacterial lawn.   The

Bergerac strain was maintained at $17^{o}C$ while the Bristol strain was kept

at ambient room temperature or at $17^{o}C$.   Synchronous nematode populations

were started from isolated eggs by dissolving gravid adults in 2% sodium

hypochlorite, 0.05M NaOH for 10 minutes (Emmons et al., 1979).

2.2   Analysis of Heat Shock Proteins

2.2.1   [^{35}S]sulfate Labelling of E. coli

E. coli K12 was grown to stationary phase in 100 ml of minimal medium

containing 5 mCi of [^{35}S]sulfate (New England Nuclear) according to the

procedure of Bretscher and Smith (1972).

2.2.2   In vivo Labelling of C. elegans (Bristol) Proteins and Induction of
Heat Shock Polypeptides

In a typical experiment, a plate of synchronous adult nematodes growing

at room temperature ($22^{o}C$) was transferred to a $35^{o}C$ incubator.   After 1

hour, the nematodes were washed off with $35^{o}C$ distilled water and allowed

to settle by gravity at $35^{o}C$.   The nematodes were then washed twice at

35°C and transferred to prewarmed NG plates onto which approximately 25

μCi of [^{35}S]-labelled E. coli K12 paste had been spread. Following a

labelling period ranging from 30 minutes to 5 hours at 35°C, the nematodes

were washed off as before and transferred to NG plates containing unlabelled

bacteria (either E. coli K12 or E. coli OP50) for at least 30 minutes at the

elevated temperature. This was carried out to remove any labelled bacteria

remaining in the gut. At this time or after a recovery period at room

temperature, the nematodes were washed thoroughly at room temperature and

resuspended in 25 - 50 μl of three times concentrated Laemmli sample

buffer (Laemmli, 1970). Proteins were then solubilized by 2 - 3 rapid

freeze-thaw cycles followed by boiling for 5 minutes and were analyzed using

SDS-polyacrylamide gel electrophoresis and autoradiography.

2.2.3 Analysis of in vitro Labelled Proteins

Total RNA, polyadenylated RNA (polyA$^+$RNA) or hybridization selected

RNA was translated in a rabbit reticulocyte system (New England Nuclear) as

described (Pelham and Jackson, 1976) using [^{35}S]methionine. Translation

products were fractionated on SDS-polyacrylamide gels and analyzed by

autoradiography. For two-dimensional gel electrophoresis, the 25 μl

translation reaction was precipitated in 10 volumes of acetone at -20°C

overnight. The pellet was washed with ethanol, dried and resuspended in

O'Farrell loading buffer A (O'Farrell, 1975).

2.2.4 Polyacrylamide Gel Electrophoresis of Proteins

SDS slab gels (0.08 cm X 7.5 cm X 10 cm) containing 15% polyacrylamide

with a 4.5% stacking gel were prepared using the discontinuous buffer system

of Laemmli (1970). Two-dimensional polyacrylamide gel electrophoresis was carried out as described by O'Farrell (1975). After destaining, the gels were dried and autoradiography was carried out using Kodak X-Omat AR film.

2.3 RNA Analysis

2.3.1 Isolation of RNA from Bristol N2 Nematodes

Control nematodes from which RNA was to be isolated were washed off of NG plates with sterile 0.14 M NaCl at 4°C. The nematodes were then washed twice with cold 0.14 M NaCl after allowing them to settle on ice by gravity and removing the supernatant by aspiration. Finally they were centrifuged at 12,000 g for 5 minutes at 4°C and the pellet was stored at -70°C. Nematodes were heat shocked by incubating NG plates at 35°C for 2 - 4 hours before carrying out the washing procedure described above. One hundred plates typically yielded 1.5 - 2.0 gram of nematodes.

The RNA isolation procedure was based on that of Chirgwin et al. (1979) with the following modifications. Nematodes were passed twice through a pre-cooled French press at 8000 psi in ten volumes of 6 M guanidinium hydrochloride (Gu-HCl), 0.2 M sodium acetate pH 5.0, 0.1 M ß-mercaptoethanol at 4°C. After centrifugation at 12,000 g for 10 minutes at 4°C, 0.5 volume of 95% ethanol was added to the supernatant and the RNA was precipitated overnight at -20°C. The subsequent purification steps were as described (Chirgwin et al., 1979). Typical yields were 10 mg per gram of starting material. Control and heat shock polyA^{+}RNA were prepared from total RNA by two passages through an oligo-dT cellulose column (Collaborative Research Inc., Type 2) using the procedure of Aviv and Leder (1972) except that the final polyA^{+}RNA fractions were eluted with

sterile distilled water. All glassware and solutions were treated with 0.1%

diethylpyrocarbonate (DEP) and baked or autoclaved, respectively, before

use.

2.3.2 Electrophoresis of RNA and Northern Transfers

RNA was denatured with glyoxal according to the procedure of McMaster

and Carmichael (1977) except that dimethylsulfoxide was omitted. After

fractionation on agarose gels in 10 mM NaH_2PO_4 pH 7.0 with buffer

recirculation, the RNA was transferred directly to nitrocellulose

(Schleicher and Schuell) in 20 X SSPE (Thomas, 1980). 1 X SSPE is 0.1 mM

EDTA, 10 mM NaH_2PO_4 pH 7.0 and 0.18 M NaCl. Alternatively, RNA was

denatured with formaldehyde and electrophoresed through formaldehyde agarose

gels (Maniatis et al., 1982). Formaldehyde gels were processed after

electrophoresis as described (Maniatis et al., 1982) and the RNA was

transferred to nitrocellulose as above. After transfer, the filters were

air dried and baked at $80^{\circ}C$ for 2 hours.

2.3.3 S1 Nuclease Mapping

S1 nuclease protection analysis was carried out essentially as

described by Berk and Sharp (1977). For mapping the hsp16-48 mRNA, 10 to 20

ng of a TaqI-RsaI fragment (map coordinates +44 to -191), 5'-end labelled

with polynucleotide kinase and $[\gamma-^{32}P]ATP$ at the TaqI site, was

hybridized with 200 ng of heat shock polyA$^+$RNA for 20 hours at $47^{\circ}C$ in

10 μl of hybridization buffer (50% formamide, 0.4 M NaCl, 40 mM PIPES, 1.0

mM EDTA). For the hsp16-1 mRNA, 10 to 20 ng of an AluI-XbaI fragment (+28

to -85), 5'-end labelled as above at the AluI end, was hybridized to 500 ng

of the same RNA for 12 hours at 50°C in 10 µl of hybridization buffer

without formamide. In each case, hybridization was carried out by first

heating the mixture at 85°C for 15 min and then quickly submerging the

tube in a water bath at the appropriate temperature. Hybridization was

terminated by the addition of 300 µl of ice-cold S1 nuclease mixture

containing 0.28 M NaCl, 50 mM sodium acetate pH 4.6, 4.5 mM ZnSO$_4$, and 200

U of S1 nuclease (Boehringer-Mannheim). S1 nuclease digestions were carried

out at 37°C for 30 min and terminated with 50 µl of 4.0 M ammonium

acetate-100 mM EDTA. The protected fragments were precipitated with

isopropanol with 20 µg of E. coli tRNA as carrier and analyzed on 8%

acrylamide gels containing 8 M urea (see DNA Sequencing).

2.4 Identification of cDNAs

2.4.1 Screening of a cDNA Library

A cDNA library made with polyA$^+$mRNA purified from heat shocked

Bristol nematodes was kindly provided by Don Jones. The cDNA had been

inserted into the PstI site of pBR322 using G-C tailing. Annealed DNA was

used to transform E. coli RR1 to create the library which was used in this

study. The cDNA library was screened by the colony hybridization method of

Grunstein and Hogness (1975), using as a probe [^{125}I]-labelled heat shock

mRNA which had been size fractionated (see section 2.10.1 below).

Ampicillin sensitive, tetracycline resistant colonies were individually

transferred from a master plate to a nitrocellulose filter (Schleicher and

Schuell) and allowed to grow overnight at 37°C on LB (1.0% bactotryptone,

0.5% yeast extract, 1.0% NaCl pH 7.5) plates containing tetracycline

(15 µg/ml). In some cases, the filter was transferred to an LB plate
containing 170 µg/ml chloramphenicol after the colonies had grown to a
diameter of approximately 1.0 mm and the plasmids were allowed to amplify
for 12 hours at 37°C. This alternate procedure, however, did not
significantly increase the hybridization signal of positive recombinants.
Cell lysis was carried out with SDS and NaOH as described by Maniatis et al.
(1982). The filters were air dried and baked at 80°C for 2 hours prior to
hybridization. Putative positives were purified from the master plate and
the plasmids were analyzed further.

2.4.2 Hybridization Selection Analysis of cDNAs

10 µg of plasmid DNA was linearized with BamHI, phenol-extracted and
precipitated with 2 volumes of 95% ethanol. The DNA was then applied to
nitro-cellulose filters (Schleicher and Schuell, B-6) using a Millipore
sintered glass filtration unit as described by Young et al. (1980).
Hybridization, washing and elution were carried out using the protocol of
Tilghman et al. (1978). Typically, the filters carrying plasmid DNA were
incubated with 10 µg of heat shock polyA^{+}RNA in 400 µl of
hybridization buffer for 20 hours at 43°C. Non-hybridized RNA in the
hybridization buffer was precipitated with 2 volumes of 95% ethanol and
1 µg was used for in vivo translation. Hybridized RNA was eluted in 90%
formamide, 10 mM Tris-HCl pH 7.5 and 1.0 mM EDTA at 45°C for 1 hour,
precipitated after the addition of 10 µg of carrier E. coli tRNA, and
translated.

2.5 General Methods for Plasmid Analysis

2.5.1 Bacterial Strains

E. coli RR1 was used to propagate pBR322 and pBR325-derived recombinants including the cDNA library and was grown in LB medium. E. coli strains JM101 and JM103 were used for transformation with M13 and pUC-derived plasmids and were grown in YT medium (0.8% bactotryptone, 0.5% yeast extract, 0.5% NaCl pH 7.0).

2.5.2 Transformations

Bacterial cells were made competent for transformation using 50 mM $CaCl_2$ as described by Messing (1983). Competent cells were transformed with either purified plasmid DNA, ligation mixes or single-stranded M13 DNA. Ligations were usually carried out with 10 ng of vector DNA and 20 - 60 ng of insert DNA in 20 µl of 50 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1.0 mM spermidine, 1.0 mM ATP and 100 µg/ml bovine serum albumin. Ligations were done at $4^{\circ}C$ or $15^{\circ}C$ for at least 12 hours using 1.0 - 10 Weiss units of T4 DNA ligase.

DNA inserted into the PstI site of pBR322 rendered the transformed colonies tetracycline-resistant and ampicillin-sensitive. pBR325 derivatives containing inserts in the EcoRI site were screened by their resistance to ampicillin and their sensitivity to chloramphenicol. The concentrations of antibiotics used are taken from Maniatis et al. (1982). Plaques infected with recombinant M13 phage were assayed for their inability to cleave 5-bromo4-chloro3-indolylgalactoside (X-gal) as described by Messing (1983). The same color assay was used to screen bacterial colonies containing pUC plasmids.

2.5.3 Purification of Plasmid DNA

For the large-scale isolation of plasmid DNA, 500 ml cultures were amplified in the appropriate medium with chloramphenicol according to Maniatis et al. (1982). pBR325 and its derivatives were amplified with spectinomycin at a concentation of 300 μg/ml. For the large-scale isolation of M13 replicative form DNA, 5.0 ml of an exponentially growing JM101 or JM103 culture and 100 - 200 μl of a M13 infectious phage supernatant (see sequencing section) were added to 500 ml of YT and incubated at 37°C for 5 - 7 hours.

Plasmids were purified using the alkali lysis procedure of Birnboim and Doly (1979) as described by Maniatis et al. (1982). The final DNA pellet was dissolved in 8 ml of TE and was extracted once with an equal volume of phenol/chloroform (1:1) and once with the same volume of chloroform. RNaseA was added to a final concentration of 25 μg/ml and digestion was allowed to proceed at room temperature for one hour. The DNA was banded in two 4 ml cesium chloride-ethidium bromide density gradients (Maniatis et al., 1982) in a Beckman VTi65 rotor at 60,000 rpm (12 - 16 hours, 20°C). After collecting the plasmid DNA band from the gradient, the ethidium bromide was removed by several extractions with water-saturated 1-butanol. The final aqueous phase was diluted 3-fold with water prior to the addition of 2 volumes of ethanol. Precipitation was allowed to proceed at -20°C overnight and the DNA was pelleted by centrifugation at 12,000 g for 20 minutes. The DNA was washed with 95% ethanol, dried and resuspended in TE. Yields ranged from 100 μg to 1.0 mg per 500 ml culture depending on the nature of the plasmid.

Rapid small-scale plasmid isolations were carried out on 1.5 ml of overnight cultures as described by Maniatis et al. (1982) using the alkali lysis procedure of Birnboim and Doly (1979). M13 replicative forms were purified in the same manner from cultures infected with M13 phage.

2.6 Analysis of Bacteriophage

2.6.1 Screening of C. elegans Bristol Genomic DNA Libraries

Two different Bristol (N2) genomic DNA libraries were used in this study. One was a partial EcoRI digest in the lambda Charon4 vector. It was constructed and kindly provided by Terry Snutch of Simon Fraser University. This phage was propagated in E. coli DP50supF or LE392 hosts which were grown in NZYT (1.0% NZ-amine, 0.5% yeast extract, 0.2% casamino acids, 10 mM MgCl$_2$ and 40 µg/ml thymidine pH 7.0). Bacterial strain DP50supF required in addition, diaminopimelic acid at a concentration of 0.1 µg/ml. The other library was a partial MboI digest cloned into the BamHI site of the lambda derivative EMBL4. This library was constructed by Chris Link and kindly provided by Mike Krause both at the University of Colorado, Boulder. This library was used in conjunction with E. coli Q358 or Q359 strains in NZYC media (1.0% NZ-amine, 0.1% yeast extract, 0.5% NaCl, 0.1% casamino acids and 10 mM MgCl$_2$ pH 7.0). Since this library had a low percentage of wild type phage, the Q358 host, in which wild type phage are viable, was preferred since it gave larger plaques and stronger hybridization signals.

Initial screening was carried out on 50,000 - 100,000 plaques, representing approximately 5 - 10 genome equivalents based on the conservative estimate that each recombinant clone contained 10 kb of

C. elegans DNA. Phage were plated at a density of approximately 1,000 – 2,000 plaques per 10 cm petri plate. Aliquots of bacteriophage in a volume of 200 – 300 µl of λ dilution buffer (0.1 M NaCl, 0.01 M Tris-HCl pH 7.5, 0.01 M $MgCl_2$ and 0.02% gelatin) were mixed with an equal volume of an overnight culture of the appropriate bacteria and incubated at $37^{o}C$ for 20 minutes before plating in the appropriate medium containing 0.7% agarose. Both phage libraries were screened as described by Maniatis et al. (1982), based on the procedure of Benton and Davis (1977). Bacteriophage and agarose plugs containing phage clones were maintained in λ dilution buffer at $4^{o}C$ in the presence of chloroform. Recombinant clones of interest were purified by subsequent rounds of replating and rescreening at decreasing densities of phage until an isolate was completely homogeneous.

2.6.2 Establishing High Titer Phage Stocks

Purified phage were plated as described above, at a density of approximately 10,000 plaques per plate. In this case, however, 3 ml of media was added to the phage-host mixture before the addition of 3 ml of top agarose prior to plating. This resulted in a more fluid top layer which could be readily scraped off after the plaques had grown at $37^{o}C$. The plates were rinsed with an additional 2 ml of media. The phage suspension was centrifuged at 12,000 g for 5 minutes to remove the agarose. The supernatant (approximately 3 ml) was stored at $4^{o}C$ in the presence of chloroform. This procedure consistently resulted in a 10-fold increase in phage concentration.

2.6.3   Isolation of Bacteriophage DNA

Phage DNA was routinely purified from 20 ml cultures. Consistent lytic infections were obtained with the following conditions.   200 μl of λ dilution buffer containing $0.5 - 1.0 \times 10^6$ phage was incubated at 37°C for 20 minutes with 100 μl of a stationary phase (overnight) culture of LE392 or Q358 or 200 μl of DP50supF. This mixture was added to 20 ml of the appropriate growth medium in a 125 ml Erlenmeyer flask and incubated at 37°C with vigorous shaking. Lysis usually occurred within 5 - 7 hours, at which time a few ml of chloroform was added to the culture which was left shaking a further 5 - 10 minutes. The contents were transferred to a 30 ml glass (Corex) tube, being careful to leave most of the chloroform behind, and the sample was centrifuged at 12,000 g for 10 minutes. The supernatant was transferred to a clean tube and centrifugation was repeated in order to remove all the bacterial debris. To the supernatant were then added 3 ml of 5.0 M NaCl and 3 g of polyethyleneglycol (molecular weight 8,000 - 15,000). The contents were mixed and left at 4°C for at least 2 hours to precipitate the phage particles. After centrifugation at 12,000 g for 10 minutes, the phage pellet was resuspended in 500 μl of DNase buffer (50 mM Hepes pH 7.5, 5.0 mM $MgCl_2$ and 0.5 mM $CaCl_2$) to which was added 10 μl RNaseA (5 mg/ml) and 5 μl DNaseI (1 mg/ml; Boehringer Mannheim, Grade I). RNaseA (Boehringer Mannheim; analytical grade) was boiled for 5 minutes to inactivate contaminating DNase activity. After incubation at 37°C for 60 minutes, 50 μl of 10 x SET (0.1 M Tris-HCl pH 7.5, 0.2 M EDTA and 5% SDS) was added before digestion with 8 μl of proteinase K (25 mg/ml; Boehringer Mannheim) for 60 minutes at 68°C. This mixture was extracted once with an equal volume of phenol/chloroform (1:1) and once with the same

volume of chloroform. The phases were separated by centrifugation at 15,000 g for 3 minutes. The DNA was precipitated from the aqueous phase with 2 volumes of 95% ethanol at room temperature for 2 minutes, and collected by centrifugation at 15,000 g for 5 minutes. The DNA pellet was washed with 1.0 ml of 70% ethanol and recentrifuged. The final pellet was dried and resuspended in 50 μl of TE (10 mM Tris-HCl pH 7.5, 10 mM EDTA). Typically, 2 - 5 μg of phage DNA was obtained. RNA contamination was usually high, but could be removed by RNaseA digestion during subsequent restriction endonuclease reactions (see below).

2.7 Purification of C. elegans Genomic DNA

Bristol (N2) and Bergerac (BO) nematodes were collected as described above for RNA isolation. DNA preparations were typically done on 0.5 - 1.0 gram of nematodes. Nematode pellets were resuspended in 10 ml of proteinase K buffer (0.1 M Tris-HCl pH 8.5, 0.05 M EDTA, 0.2 M NaCl and 1% SDS) as described by Emmons et al. (1979). Proteinase K was added to a final concentration of 200 μg/ml and the solution was incubated at 65°C for 30 - 60 minutes at which time the solution was clear. This solution was extracted three times with phenol and once with chloroform in a separatory funnel with gentle mixing. Phase separation was carried out in a desk top centrifuge at full speed for 3 minutes. The aqueous phase was chilled and 2 volumes of 95% ethanol at -20°C was gently layered over it. The DNA was precipitated either by winding it upon a glass rod or by rotating the tube at an angle to disturb the interface. The DNA was then washed with chilled 70% ethanol, dried and resuspended in an appropriate volume of TE (1.0 - 2.0 ml). This usually resulted in a DNA solution of approximately 1.0 mg/ml.

RNA was digested with RNaseA in subsequent restriction enzyme reactions (see below). Banding of the DNA in CsCl density gradients was found to be unnecessary.

2.8   General DNA Techniques

2.8.1   Restriction Endonuclease Digestion of DNA

DNA (0.5 - 2.0 µg) was usually digested in a total volume of 15 µl using the buffer system described by Maniatis et al. (1982). Bovine serum albumin (ultrapure grade) was added to a final concentration of 100 µg/ml. In most cases, 1.0 - 5.0 units of restriction enzyme was used for each reaction. Restriction enzymes were purchased from Bethesda Research Laboratories, New England Biolabs, Boehringer Mannheim and Pharmacia. For phage DNA and C. elegans genomic DNA, 5.0 µg of RNaseA was included. Restriction enzyme digestion mixtures were analyzed by electrophoresis in agarose gels as described below after the addition of 0.1 volume of loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol and 25% ficoll).

2.8.2   Electrophoresis of DNA and Southern Transfers

DNA samples or restriction endonuclease reaction mixtures were analyzed on agarose gels which were poured and run in 1 x TBE (89 mM Tris-borate pH 8.3, 89 mM borate and 2.0 mM EDTA) containing 0.5 µg/ml ethidium bromide. DNA bands were visualized under ultraviolet light and photographs were taken with a Polaroid camera using type 57 film. The DNA was transferred to nitrocellulose as described by Southern (1975) except that the acid depurination step was usually omitted. Transfer was carried out in either

20 x SSPE or in 1.0 M ammonium acetate pH 7.0, the latter being more

efficient. The filters were then air dried and baked at 80°C for 2 hours.

2.8.3 Purification of Specific DNA Fragments

DNA fragments were recovered from agarose gels by electroelution into

dialysis tubing using 0.5 x TBE. The DNA was then purified by

chromatography through RPC-5 analog (BRL) columns, NACS PREPAC cartridges

(BRL), or DE-52 columns. In all cases, the DNA was loaded and washed in TE

containing 0.2 M NaCl. The DNA was eluted in 300 µl of TE containing 2.0

M NaCl. After the addition of 0.1 volumes of 3.0 M sodium acetate pH 5.2,

the fragment was precipitated with 2.0 volumes of 95% ethanol and

resuspended in a small volume of sterile distilled water.

2.8.4 End-Labelling of DNA Fragments

Fragment ends containing 5' overhangs generated by restriction enzyme

digestions were labelled in one of two ways. The 3' ends were end-labelled

using the Klenow fragment of E. coli DNA polymerase I and the appropriate

$[\alpha-^{32}P]$deoxynucleoside triphosphates, while the 5' ends were

end-labelled with T4 polynucleotide kinase and $[\gamma-^{32}P]$adenosine

triphosphate (ATP) after dephosphorylation with calf intestinal alkaline

phosphatase (grade I). These reactions were carried out according to

Maniatis et al. (1982). Alternatively, the 3' ends of 3' overhangs were

end-labelled with calf thymus terminal transferase and $[\alpha-^{32}P]$cordycepin

5'-triphosphate as described by Tu and Cohen (1980).

Single-end-labelled fragments were obtained by cleavage with a second

restriction enzyme and were purified on 5% or 8% preparative polyacrylamide

slab gels (0.15 x 15 x 17 cm) containing acrylamide:bisacrylamide at a ratio of 29:1, 1 x TBE, 0.06% ammonium persulphate and 0.03% TEMED (N,N,N',N',-tetramethylethylene diamine). Autoradiography was used to locate the fragment of interest. After electroelution into dialysis tubing as described above, the fragment was precipitated with 0.1 volumes of 3.0 M sodium acetate pH 5.2 and either 2 volumes of 95% ethanol or 1 volume of isopropanol. In addition, either 20 µg of E. coli tRNA or 5 µg of pBR322 was used as carrier during the precipitation step.

2.9 DNA Sequencing

The cDNAs were sequenced using the base modification procedure of Maxam and Gilbert (1980). Maxam and Gilbert reactions were also used to determine mRNA initiation sites in S1 nuclease protection experiments. The sequence of the genomic DNA clones was determined using the dideoxynucleotide chain termination technique of Sanger et al. (1977). A detailed description of methods used, including the subcloning of DNA fragments into M13 vectors, the preparation of single stranded phage DNA and the dideoxynucleotide reactions using the Klenow fragment of E. coli DNA polymerase I is given in Messing (1983). Fragments were cloned into M13mp8, M13mp9 or M13mp11 by using either cohesive-end or blunt-end ligations, the latter after the cohesive ends had been filled in with Escherichia coli DNA polymerase I (Klenow fragment). In most cases, M13 clones were first screened using only ddTTP reactions in order to avoid sequencing redundant clones. Fragments cloned into the RF in the opposite orientation were sometimes screened by their ability to form a figure eight-like structure which migrates slower in agarose gels (Messing, 1983). All sequencing reactions were analyzed on 6%

polyacrylamide slab gels (0.035 x 15 x 35 cm). The gels contained an

acrylamide:bisacrylamide ratio of 19:1 in addition to 8.3 M urea, 1 x TBE,

0.06% ammonium persulphate and 0.03% TEMED. Electrophoresis was in 1 x TBE

at a constant current of 17.5 milliamperes. The gels were dried onto

Whatman filter paper and autoradiographed using Kodak X-Omat RP film.

2.10 Preparation of Hybridization Probes

2.10.1 Preparation of $[^{125}I]$-labelled RNA

Forty µg of heat shock polyA^{+}RNA was fractionated on a denaturing

98% formamide-polyacrylamide slab gel (0.08 x 7.5 x 10 cm) as described by

Maniatis et al. (1975). The gel was cut into 1.0 cm slices and the RNA was

electroeluted in 20 mM Tris-acetate pH 8.0, 0.4 mM EDTA. Ten percent of

each such fraction was assayed in a cell-free translation system after

concentration by ethanol precipitation. The RNA fraction which directed the

translation of hsp16 mRNA was iodinated with $[^{125}I]$iodide (Amersham, 17

mCi/mg NaI) according to the procedure of Commerford (1971) to a specific

activity of 2.5×10^{7} cpm per µg of RNA.

2.10.2 Purification and Labelling of Oligodeoxynucleotides

Two 18mer oligodeoxynucleotides were used as hybridization probes:

I.   CGGGGCCGCGCGCACGCA

II.   CAGGGCCGCGCGCACGCA

The oligodeoxynucleotides were synthesized by Tom Atkinson in the laboratory

of Dr. M. Smith, UBC, with an Applied Biosystems 380A DNA synthesizer.

Oligodeoxynucleotides were purified through 20% sequencing (urea) gels and

isolated by C$_{18}$ SEP-PAK (Millipore) chromatography as described by

Atkinson and Smith (1984). To make hybridization probes, 20 pmoles of oligodeoxynucleotide was labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase as described by Zoller and Smith (1983).

2.10.3 Preparation of Double-Stranded DNA Probes

Purified DNA fragments were nick translated with [α-^{32}P]dCTP and [α-^{32}P]dGTP by the method of Rigby et al. (1977).

M13 templates containing C. elegans genomic DNA fragments were also used to generate probes with high specific activities. An annealing mixture containing 3.0 μl of template (0.5 to 1.0 μg), 2.0 μl of universal primer (P-L Biochemicals; 0.03 A$_{260}$ units per ml), 2.0 μl of 10 x annealing buffer (100 mM Tris-hydrochloride, pH 7.5, 600 mM NaCl, 70 mM MgCl$_2$), was incubated at 65°C for 15 min in a 1.5-ml microfuge tube. After cooling to room temperature (5 to 10 min), 1.0 μl of 20 mM dithiothreitol, 2.0 μl of 0.5 mM dATP, 2.0 μl of 0.5 mM dTTP, 2.5 μl each of [^{32}P]dGTP and [^{32}P]dCTP (25 μCi; 3,000 Ci/mmol), and 0.5 U of E. coli DNA polymerase I (Klenow fragment) were added. The reaction was allowed to proceed for 10 min at room temperature and was then followed by a chase after the addition of 2.0 μl of 0.5 mM dGTP and 2.0 μl of 0.5 mM dCTP for 5 min before termination by heating at 70°C for 10 min. The primer-extended product was then digested with HaeIII for 30 min.

For both techniques, free triphosphates were separated from the labelled strands by chromatography on 1.0 ml spun columns of Sephadex G50 (Maniatis et al., 1982).

2.11 Hybridization

Nitrocellulose blots of DNA from bacterial cDNA transformants were prehybridized in 4 x SSPE, 50% formamide for 1 hour at 37°C. Hybridization was carried out in 4 x SSPE, 50% formamide containing 1.0 x 10^6 cpm per filter of $[^{125}I]$-labelled RNA at 37°C for at least 12 hours. The filters were washed in two changes of 2 x SSPE, 0.1% SDS followed by two changes of 0.1 x SSPE, 0.1% SDS all at room temperature. Final washes were carried out at 50°C in 0.1 x SSPE, 0.1% SDS.

In the case of double-stranded DNA probes, prehybridization was done in 5 x SSPE, 50% formamide, 5 x Denhardt's reagent, 0.1% SDS and 100 - 200 µg/ml of sheared, denatured E. coli, calf thymus or salmon sperm DNA. Prehybridization was at 42°C for at least 1 hour. Hybridizations were carried out for at least 12 hours at 42°C in the solution described for the prehybridization except that 1 x Denhardt's reagent was used and the denatured $[^{32}P]$-labelled probe was included. Washing of the filters was identical to the conditions described for the RNA probe above.

For 5'-labelled oligodeoxynucleotides, prehybridization and hybridization was done at 37°C for at least 1 hour and 12 hours, respectively, in 6 x SSPE, 2 x Denhardt's reagent and 0.2% SDS. All washes were done in 6 x SSPE. Washing was usually carried out at room temperature for 15 minutes, then 2 x 15 minutes at 37°C followed by 2 final washes at 48°C for 15 minutes each. 1 x SSPE is 0.1 mM EDTA, 10 mM NaH_2PO_4 pH 7.0 and 0.18 M NaCl. 1 x Denhardt's reagent is 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin and 0.02% ficoll.

After washing, nitrocellulose filters from all of the hybridizations described were air dried and fluorographed at -70°C using either Kodak

X-Omat AR or X-Omat RP film, the former being approximately 5-fold more

sensitive.

2.12 Summary of Bacterial Strains Used.

Table IV. Genotypes of Bacterial Strains

	Strain	Genotype
a.	DP50, supF	F^-, tonA53, dapD8, lacY1, glnV44 (supE44), Δ(gal-uvrB)47, λ^-, tyrT48(supF58), gyrA29, Δ(thyA57), hsdS3
b.	LE392	F^-, hsdR514(r_k^-,m_k^-), supE44, supF58, lacY1 or Δ(lacIZY)6, galK2, galT22, metB1, λ^-, trpR55
c.	RRI	F^-, hsdS20(r_B^-, m_B^-), ara-14, proA2, lacY1, galK2, rpsL20 (Sm^r), xyl-5, mtl-1, supE44, λ^-
d.	Q358	$hsdR_k^-$, $hsdM_k^+$, su_{II}^+, 80^R
e.	Q359	$hsdR_k^-$, $hsdM_k^+$, su_{II}^+, 80^R, P2
f.	JM101	Δlacpro, supE, thi, F'traD36, proAB, $lacI^q Z\Delta M15$
g.	JM103	Δlacpro, supE, thi, F'traD36, strA, sbcB15, endA, hspR4, proAB, $lacI^q Z\Delta M15$

a,b,c - taken from Maniatis et al. (1982)

d,e, - taken from Karn et al. (1980)

f,g - taken from Messing (1983)

The hosts JM103 and JM101 were streaked out on plates containing 0.2%

glucose, 0.001% vitamin B1 and minimal salts as described in Messing (1983).

III. RESULTS

3.1 The Heat Shock Response of <u>Caenorhabditis</u> <u>elegans</u> var. Bristol, strain
N2

<u>C</u>. <u>elegans</u> undergoes a typical heat shock response at the elevated
temperature of 35°C, the normal growth temperature being approximately
20°C. As shown in Figure 3, there was a dramatic decrease in general
protein synthesis while a unique set of hsps was induced. From these
in <u>vivo</u> labelling experiments it appeared that hsp81, hsp41, hsp38 and hsp29
are synthesized in control nematodes but they continue to be synthesized at
35°C. These observations are in agreement with Snutch and Baillie (1983)
except for hsp29 which they reported to be synthesized only during heat
shock conditions. A protein with an apparent molecular weight of 36,000
accumulated for the first 3 hours of labelling but was not detectable by 5
hours. The simplest explanation for this is that this particular protein
has a high turnover rate and is preferentially degraded during prolonged
exposure of the nematodes to 35°C.

Hsp70, hsp18 and hsp16 did not appear to be synthesized in nematodes
growing at normal temperature but were induced during heat shock, hsp18 and
hsp16 accumulating to very high levels. The nematode hsp70 is probably
homologous to hsp70 of <u>Drosophila</u> and rainbow trout since <u>C</u>. <u>elegans</u> genomic
DNA hybridizes to a <u>Drosophila</u> hsp70 genomic DNA clone (Snutch and Baillie,
1983) and a trout hsp70 cDNA (Kothary <u>et al</u>., 1984). Hsp18 sometimes
appeared as a doublet as shown in Figure 3, and is probably identical to
hsp19 described by Snutch and Baillie (1983). In addition, Figure 3 shows
the induction of a 50 kd protein in a synchronous population
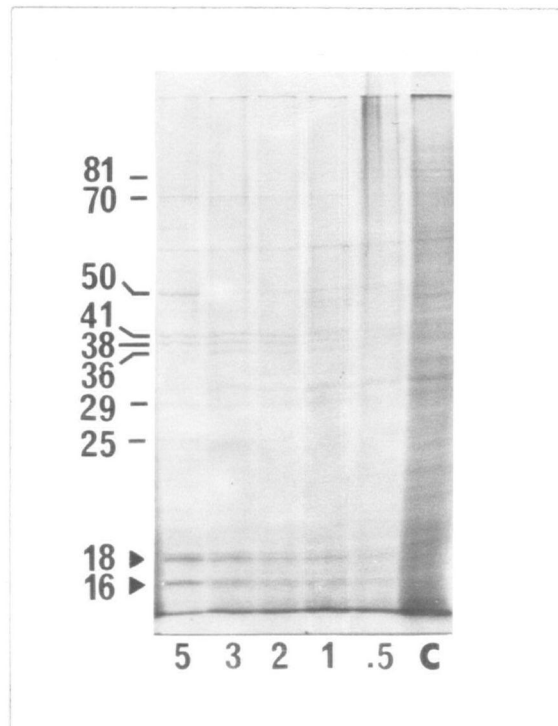
Figure 3.   Induction of heat shock proteins in C. elegans.

Synchronously growing adult worms were incubated at 35°C and labelled in vivo with [^{35}S]-labelled E. coli. The numbers below each lane indicate the duration of exposure, in hours, to the radioactive bacteria before a 30 minute chase at 35°C. The lane on the right (C) represents total protein synthesis at 22°C. The positions of the shsps are indicated by solid arrows; the positions of the other major hsps are indicated by lines. Numbers refer to approximate molecular weights, in kilodaltons.

of adult nematodes. This protein may be analogous to the 50 kd protein that is synthesized in heat shocked dauer larvae (Snutch and Baillie, 1983). They, however, failed to detect hsp50 in normally growing nematodes which had been heat shocked. This discrepancy may be due to the difference in labelling protocols. Snutch and Baillie (1983) exposed the nematodes to radioactive bacteria prior to heat shock. The higher background of normal protein synthesis may have masked the appearance of a heat inducible protein in the 50 kd range.

Experiments were also carried out to determine the stability of pulse labelled hsps. If the nematodes were labelled for two hours at 35°C and allowed to recover at room temperature on unlabelled E. coli, it could be shown that the hsps of C. elegans, especially the characteristic 18 kd and 16 kd proteins, were still present at 24 hours. The results of this experiment are shown in Figure 4.

3.2 Identification of cDNAs Coding for Hsp16

In order to characterize the genes coding for hsp16 and hsp18, a cDNA library constructed from polyA$^+$RNA isolated from heat shocked Bristol nematodes was screened. The probe used was a polyA$^+$RNA fraction (Figure 5, lane D) which was highly enriched for messages coding for hsp16 and hsp18. Total mRNA was fractionated by formamide polyacrylamide gel electrophoresis and monitored by in vitro translation in a rabbit reticulocyte system as an assay for biologically active mRNA coding for the shsps. This enriched fraction was then iodinated with ^{125}I and used to screen for cDNAs specific to this fraction.
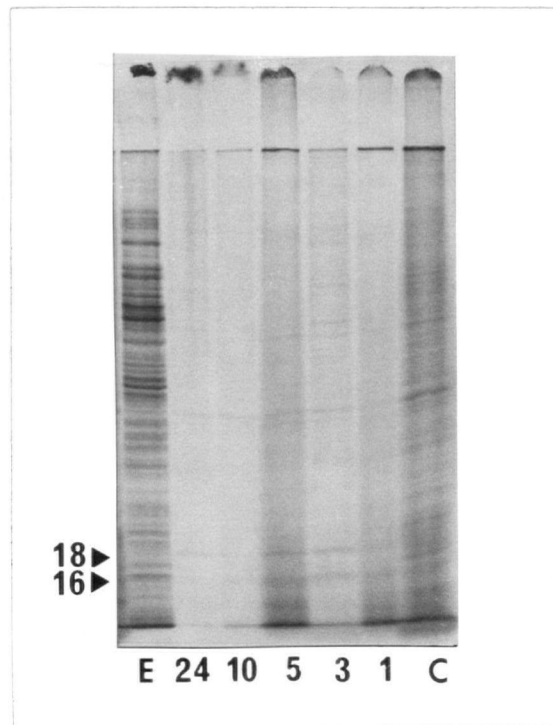
Figure 4. Stability of the hsps of C. elegans.

Synchronously growing cultures consisting of late larval stages were labelled for 2 hours at 35°C. After a chase for 30 minutes at 35°C in the presence of unlabelled bacteria, the worms were allowed to recover at normal temperature for the length of time (hours) shown below each lane. Lane C represents total protein synthesis at 22°C. Lane E is a profile of in vivo labelled proteins of E. coli K12. The positions of the shsps are indicated.
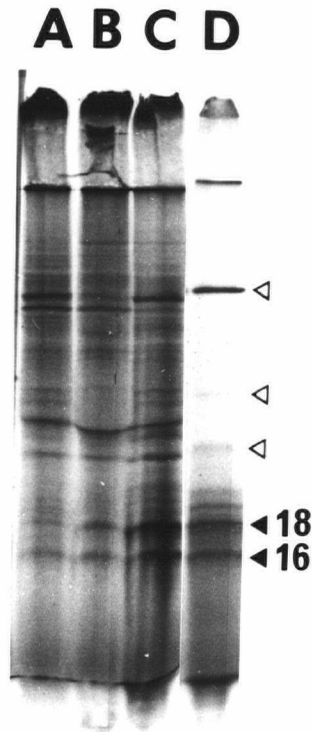
Figure 5. In vitro translation products of mRNA from heat shocked cells, and enrichment of messages coding for the small hsps.

Approximately 1.0 µg of RNA was translated in a rabbit reticulocyte system in the presence of [^{35}S]methionine. Translation products were separated on a 15% SDS-polyacrylamide gel and identified by autoradiography. Lanes are A) total control RNA; B) total heat shock RNA; C) heat shock polyA$^+$RNA; D) enriched fraction of polyA$^+$RNA purified by formamide polyacrylamide gel electrophoresis. Translation of the 18 kd and 16 kd hsps is indicated by the solid arrows. Open arrows indicate endogenously-labelled proteins which are also present when no RNA is added to the cell-free translation system (see Figure 6, lane C).

Approximately 1000 transformants were screened by colony hybridization. This yielded 27 putative positive colonies from which plasmid DNA was prepared. The DNA was restricted with PstI and the liberated cDNA inserts were analyzed on a 1.5% agarose gel. The DNA was transferred from the gel to nitrocellulose and hybridized to the [^{125}I]-labelled probe. Of the original 27 plasmids, 12 contained PstI inserts which hybridized to the radiolabelled RNA (results not shown). To identify hybrid plasmids coding for hsp16 or hsp18, hybridization selection was carried out. Seven plasmids were identified which selected mRNA specifically coding for hsp16. A typical positive selection is shown in Figure 6. Restriction enzyme analysis revealed that the positive cDNA inserts fell into two groups, based on the presence or absence of an EcoRI site. Two cDNA clones, pCEHS48 and pCEHS41, were used for sequence analysis since they contained the longest cDNA inserts representing each group.

In the hybridization selection experiments, the only mRNA selected by the cDNA hybrid plasmids tested were species coding for the 16,000 dalton hsp. Those plasmids which did not select hsp16 mRNA gave completely negative results. This suggests that messages coding for hsp16 are either very abundant, or that their structure is particularly well suited to cDNA synthesis under the conditions used. It is perhaps surprising that no positive recombinant plasmids selected mRNAs coding for hsp18, which appears to be induced to the same high levels as hsp16. The hsp18 messages were not degraded during the hybridization selection, since total polyA^{+}RNA removed from the hybridization mix contained mRNA which was translated into a 18,000 dalton protein as shown in Figure 6 (lane T). Again, it is possible that the hsp18 mRNA may possess some structural feature which, in this case,
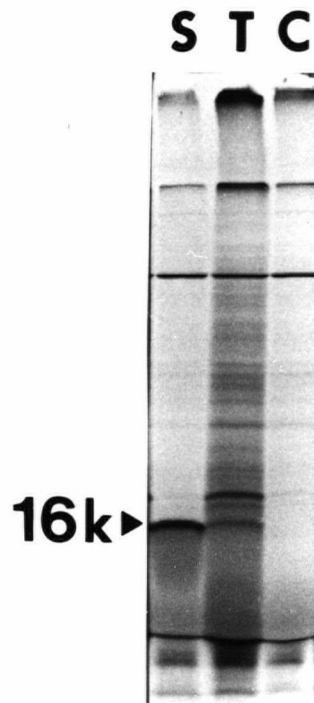
Figure 6. <u>In vitro</u> translation of RNA selected by hybridization with pCEHS41.

Plasmid pCEHS41 was bound to nitrocellulose and hybridized to polyA$^+$RNA from heat shocked nematodes. After hybridization, the RNA selected by the recombinant cDNA clone was eluted and translated <u>in vitro</u>. [^{35}S]-labelled products were analyzed by SDS-polyacrylamide gel electrophoresis and autoradiography. Lanes are C) control with no added RNA; T) non-hybridized or total polyA$^+$RNA and S) selected polyA$^+$RNA. The position of hsp16 is shown.

resulted in poor synthesis of cDNA. This feature is not the lack of a polyA
sequence, resulting in inadequate priming with oligo(dT), since this mRNA
binds to oligo(dT)-cellulose (see Figure 5, lane C). Interestingly, cDNAs
coding for Drosophila hsp22 and hsp27 were not identified under conditions
in which cDNAs coding for hsp23 and hsp26 were (Wadsworth et al., 1980;
Voellmy et al., 1981).

3.3 Two-dimensional Gel Electrophoresis of Hsp16

The finding of two distinctly different cDNAs coding for hsp16
suggested the existence of more than one gene for this protein. The
possible existence of multiple forms of hsp16 was investigated by analyzing
the translation products of mRNA selected by pCEHS48 or pCEHS41 (Figure 6,
lane S) on two-dimensional gels using the system of O'Farrell (1975). The
gel analysis was kindly carried out by Elizabeth Burgess. The results for
pCEHS41 are shown in Figure 7. Five polypeptides with distinctly different
iso-electric points, but identical molecular weights were resolved.
Identical results were obtained when pCEHS48 was used in the selection.
This implies that, under the conditions used in the hybridization selection
experiments, both pCEHS48 and pCEHS41 hybridized to the same transcripts.
This is not surprising since the cDNAs CEHS48 and CEHS41 cross-hybridize
with each other strongly under normal hybridization stringencies (results
not shown).

These results are consistent with the existence of at least two
distinct but related hsp16 sequences as suggested by the cDNA restriction
enzyme patterns.
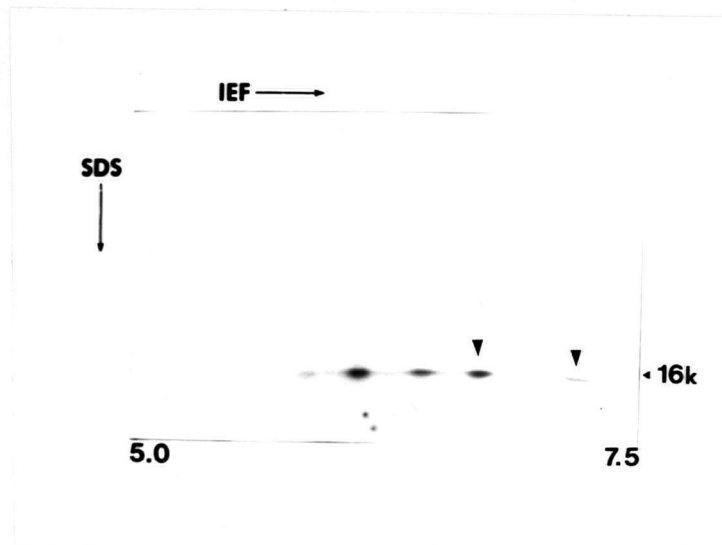
Figure 7. Two-dimensional gel electrophoresis of the products from
translation of hybrid-selected polyA[+]RNA.

PolyA[+]RNA selected by hybridization to pCEHS41 was translated
in vitro (see Figure 6, lane S). The 25 µl translation mixture was
precipitated in 10 volumes of acetone at -20°C overnight. The pellet was
washed with ethanol, dried and resuspended in O'Farrell loading buffer A
(O'Farrell, 1975). The directions of iso-electric focusing and SDS-gel
electrophoresis are shown by arrows. Arrows also show the migration of the
two most basic protein variants which are discussed in section 4.1.

3.4 Messages Coding for Hsp16 Are Not Transcribed in Control Nematodes

In vitro translation of control RNA sometimes showed the existence of what appeared to be hsp16 (see Figure 5, lanes A and B). Hsp16 transcripts might have been induced inadvertently during collection and washing of control batches of nematodes or they may be present under normal conditions. The availability of specific DNA probes for hsp16 mRNA made it possible to address this question. As shown in Figure 8, no detectable hybridization of CEHS41 cDNA to polyA$^+$RNA from control nematodes was seen. Identical results (not shown) were obtained with the CEHS48 cDNA insert as probe. Clearly, a very high differential synthesis of hsp16 mRNA occurs during heat shock, demonstrating the strong inducibility of these genes. Hsp16 transcripts do not display any size heterogeneity under these electrophoresis conditions.

3.5 Sequence Analysis of the cDNAs CEHS48 and CEHS41

The cDNAs were sequenced using Maxam and Gilbert base modification and chemical cleavage reactions on the end-labelled DNA fragments shown in Figure 9. As seen in Figure 10, the nucleotide sequences of CEHS48 and CEHS41 are very similar. Numbering of the nucleotide and predicted amino acid sequences begins at the 5' end of CEHS48. CEHS48 encodes 135 amino acid residues. CEHS41 is aligned with CEHS48, beginning at amino acid 32 since it codes for 31 fewer amino acid residues at the 5' end. One of the most striking features of these sequences is the contrast in degree of homology between different regions. From nucleotides 92 to 112, the two cDNAs differ at all positions except four. Sequence analysis of the hsp16-41 gene (Jones et al., 1986) has revealed that this stretch in CEHS41
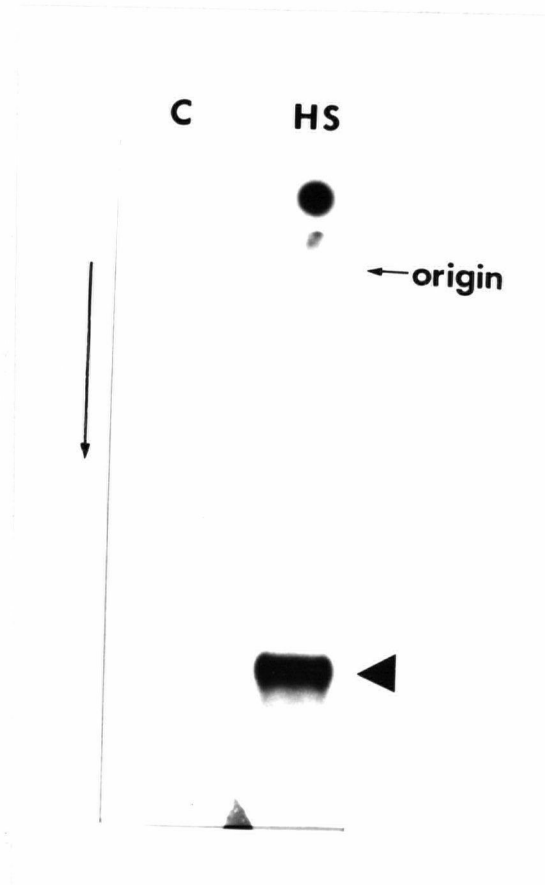
Figure 8. Northern blot and dot-blot analysis of polyA$^+$RNA from control and heat shocked worms.

1.0 µg of polyA$^+$RNA was denatured, separated on a 1.5% agarose gel and transferred to nitrocellulose. A 0.5 µg sample of the same RNA was then applied as a spot to the nitrocellulose filter at the top of the transferred RNA lanes. These blots were then probed for sequences coding for hsp16 with [^{32}P]-labelled CEHS41 cDNA. Control and heat shock lanes are indicated.
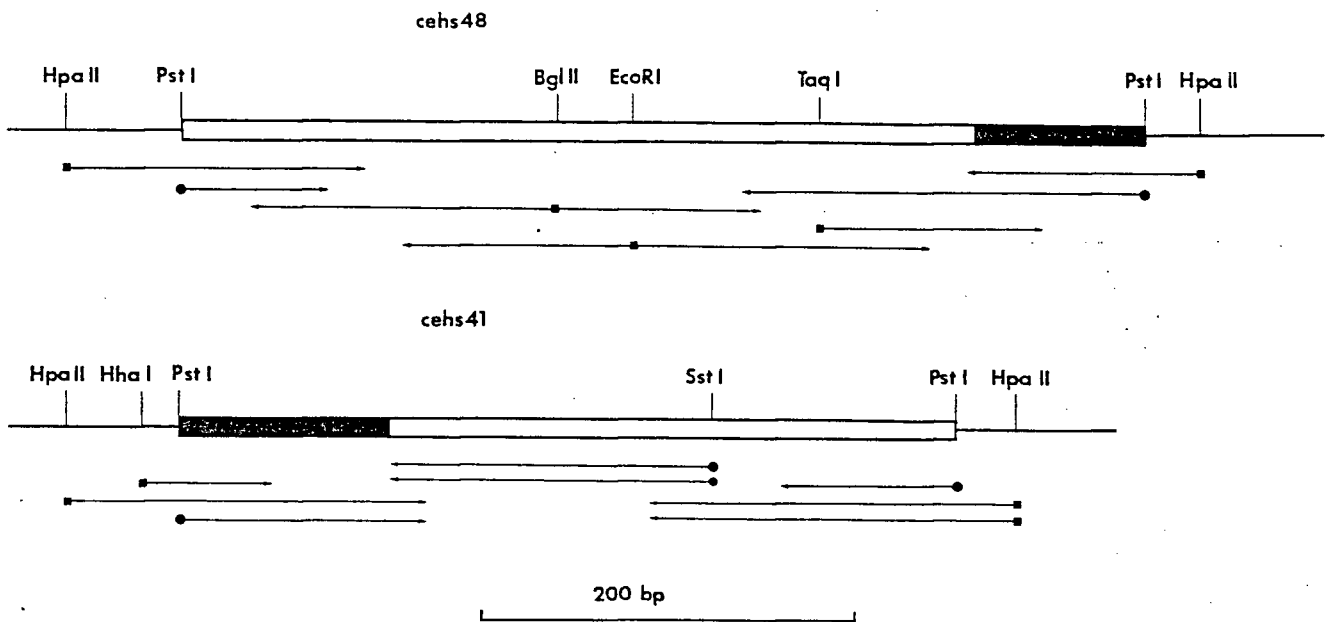
Figure 9. Strategy used to determine the nucleotide sequence of CEHS48 and CEHS41 cDNA inserts.

The cDNA inserts are shown as they are oriented relative to the conventional clockwise numbering of the pBR322 sequence. pBR322 sequences are represented by thin lines. Hsp16 coding regions are shown by open boxes while 3' non-coding regions are shown by shaded boxes. Arrows represent the direction of sequencing from Klenow-labelled fragments (squares) or from terminal transferase-labelled fragments (circles).

Figure 10. Complete nucleotide sequences of the cDNA inserts from pCEHS48

and pCEHS41 with the deduced amino acid sequences.

The coding strands of each cDNA are aligned with each other to

demonstrate sequence homology. Asterisks represent differences in the

sequences at both the nucleotide and amino acid levels. The polyadenylation

signal AATAAA, found 12 base pairs before the poly(A) stretch in pCEHS48 is

also shown.

```
                 1                        20                        40
pCEHS48    G18 TCT GAT TCA AAT GTT CTC GAT CAT TTC TTG GAT GAA ATC ACT GGA TCT GTT CAA TTT
pCEHS41
hsp16-48       Ser Asp Ser Asn Val Leu Asp His Phe Leu Asp Glu Ile Thr Gly Ser Val Gln Phe
hsp16-41
                 1                                        10


         60                        80                        100                       120
    CCA TAT TGG AGA AAT GCT GAT CAC AAC TCA TTC AAT TTT TCC GAC AAT ATT GGA GAG ATT GTA AAT GAC
                                              G14 TA AAA TTG TGT TCT TTT TTT CAG ATT GTA AAT GAT
                                                  ** *** ** *** ** *** ***
    Pro Tyr Trp Arg Asn Ala Asp His Asn Ser Phe Asn Phe Ser Asp Asn Ile Gly Glu Ile Val Asn Asp
                                                  *** *** *** *** *** *** ***
                                                  Lys Leu Cys Ser Phe Phe Gln Ile Val Asn Asp
    20                                  30                                           40


                       140                       160                       180
    GAA TCT AAA TTC TCT GTT CAA CTC GAT GTT TCT CAT TTC AAA CCA GAA GAT CTT AAA ATT GAA TTG GAT
    GAA TCC AAG TTT TCG GTT CAA CTC GAT GTC TCT CAT TTC AAA CCA GAA AAT CTT AAG ATT AAA TTA GAT
           *      *          *                          *      ***        *       ***
    Glu Ser Lys Phe Ser Val Gln Leu Asp Val Ser His Phe Lys Pro Glu Asp Leu Lys Ile Glu Leu Asp
    Glu Ser Lys Phe Ser Val Gln Leu Asp Val Ser His Phe Lys Pro Glu Asn Leu Lys Ile Lys Leu Asp
                                50                                  ***       60


        200                       220                       240                       260
    GGA AGA GAA CTA AAA ATT GAA GGA ATT CAA GAA AAA AAA TCA GAG CAT GGA TAC TCG AAA CGA TCA TTT
    GGA AGA GAG CTC AAA ATT GAA GGG ATT CAA GAA ACA AAA TCG GAA CAT GGA TAC TTG AAA CGC TCA TTT
              *      *          *               ***       *       *                  *
    Gly Arg Glu Leu Lys Ile Glu Gly Ile Gln Glu Lys Lys Ser Glu His Gly Tyr Ser Lys Arg Ser Phe
    Gly Arg Glu Leu Lys Ile Glu Gly Ile Gln Glu Thr Lys Ser Glu His Gly Tyr Phe Lys Arg Ser Phe
                    70                              ***       80          ***


                       280                       300                       320
    TCA AAA ATG ATT CTT CTA CCA GAA GAT GTT GAT TTA ACT TCT GTC AAA TCT GCA ATT TCG AAT GAA GGA
    TCA AAA ATG ATT CTT CTA CCA GAA GAT GCT GAT TTA CCT TCT GTC AAA TCT GCC ATT TCG AAT GAA GGA
                                         *          *                      *
    Ser Lys Met Ile Leu Leu Pro Glu Asp Val Asp Leu Thr Ser Val Lys Ser Ala Ile Ser Asn Glu Gly
    Ser Lys Met Ile Leu Leu Pro Glu Asp Ala Asp Leu Pro Ser Val Lys Ser Ala Ile Ser Asn Glu Gly
                                         ***         ***
    90                                   100                                     110


            340                       360                       380                       400
    AAA CTT CAA ATT GAG GCT CCA AAG AAG ACT AAC TCA TCT CGT TCT ATT CCC ATT AAT TTT GTT GCA AAA
    AAA CTC CAA ATT GAG GCT CCA AAG AAG ACA AAC TCA TCA CGT TCT ATT CCG ATC AAT TTT GTT GCA AAA
          *                          *                *                  *   *
    Lys Leu Gln Ile Glu Ala Pro Lys Lys Thr Asn Ser Ser Arg Ser Ile Pro Ile Asn Phe Val Ala Lys
    Lys Leu Gln Ile Glu Ala Pro Lys Lys Thr Asn Ser Ser Arg Ser Ile Pro Ile Asn Phe Val Ala Lys
                                         120                                     130


                       420                  _____        460                       480
    CAT TAA TCTTTTATTGTATTCCAAATATTCTTAATTTCAATAAAGTCATTAATTTAAAAAAAAAAAAAAAAAAAAAAAAAAC5
    CAT TAA CACTTTTGTTGAAGAGAAGCTACTTATTATTTGTTCTTCTTTTTTCATGTAAAATTTGAAATGTTTCCATTGTGATTCGAGAC18
             ** ** ** ****** ***** *** ***  ***  **  **  ** *** *
    His End
    His End
    135
```

is actually the 3' portion of a 58 bp intron, this particular cDNA being the result of reverse transcription of an unspliced message. CEHS48 corresponds to a correctly spliced transcript, the intron being found between amino acid residues 38 and 39 in Figure 10 (see section 3.9).

From nucleotides 113 to the TAA termination codons, the homology is 91% and results in only 6 amino acid differences out of a total of 97. Non-conservative changes occur at amino acid positions 59 (Asp → Asn), 63 (Glu → Lys), 77 (Lys → Thr), and 101 (Thr → Pro). The 3' noncoding regions are highly divergent compared to the coding region, with a sequence homology of only 30%. CEHS48 contains a polyA stretch of 24 residues, 12 nucleotides after the polyadenylation signal AATAAA (Proudfoot and Brownlee, 1976). The length of the 3' noncoding regions, excluding the polyA stretch, is 49 nucleotides for CEHS48 compared to at least 92 nucleotides for CEHS41 which is incomplete at the 3' end.

At the nucleotide level, the cDNAs have A/T base compositions of 67% and 69%. This high A/T content is reflected in the codon usage where A and T are preferred in the wobble positions. For example, ATT or ATA are used as codons for isoleucine 23 times, whereas the ATC codon is used only once. Similarly, lysine is coded for by AAA 22 times while AAG is used only 7 times. The codon usage for valine, serine, asparagine, aspartic acid, glutamic acid and glycine also reflects this trend.

The deduced amino acid sequences of the shsps of D. melanogaster are very similar in a region of 83 amino acids which shows homology to α-crystallin, the major water-soluble protein component of the vertebrate lens (Ingolia and Craig, 1982b; Ayme and Tissières, 1985). Due to the molecular weight difference and the fact that no other 16,000 dalton hsp had

```
                      AA
DMHSP 27    87- K D G F Q V C M D V S Q F K P N E L T V K V V D N T V V V E G K H E E R G D G
DMHSP 26    87- K D G F Q V C M D V A Q F K P S E L N V K V V D D S I L V E G K H E E R Q D D
DMHSP 23    67- K D G F Q V C M D V S H F E P S E L V V G V Q D N S V L V E G N H E E R E D D
DMHSP 22    61- K D G F K L T L D V K D Y - - S E L K V K V L D G S V L V G G K S E Q Q F A E
A-CRYS B₂   72- K D R F S V N L N V K H F S P E E L K V K V L G D V I E V H G K H E E R Q D E
CEHSP 16-48 43- E S K F S V Q L D V S H F K P E D L K I E L D G R E L K I E G I Q E K K S - E
CEHSP 16-41 43- E S K F S V Q L D V S H F K P E N L K I K L D G R E L K I E G I Q E T K S - E
```

```
                                                                       AA
 H G M I Q R H F V R K Y T L P K G L T P T K V V S T V S S D G V L T L  -159
 H G H I M R H F V R R Y L V P D G Y K A E Q V V S Q L S S D G V L T V  -159
 H G F I T R H F V R R Y A L P P G Y E A D K V A S T L S S D G V L T I  -139
 Q G G Y S R H F L R R F V L P E G Y E A D K V T S T L S S D G V L T I  -133
 H G F I S R E F H R K Y R I P A D V D P L A I T S S L S S D G V L T V  -144
 H G Y S K R S F S K M I L L P E D V D L T S V K S A I S N E G K L Q I  -115
 H G Y L K R S F S K M I L L P E D A D L P S V K S A I S N E G K L Q I  -115
```

Figure 11. Comparison of deduced amino acid sequences of four shsps from D. melanogaster, α-crystallin and two 16,000 dalton hsps from C. elegans.

The region shown corresponds to the 74 amino acid region which was used for comparison by Ingolia and Craig (1982a, Figure 4). The sequence of bovine α-crystallin is that of the B₂ chain (van der Ouderaa, et al., 1973) while the D. melanogaster sequences are taken from Ingolia and Craig (1982b). The regions of each protein used for the comparison are indicated by amino acid numbering. The different levels of homology are shown by open boxes and are discussed in Section 3.5. The single letter amino acid code is A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

been identified in a eucaryotic system, it was surprising to find that

C. elegans hsp16-41 and hsp16-48 contained extensive homology to the shsps

of Drosophila within the conserved block. The boxed regions in Figure 11

show areas of amino acid homology at two different levels. Firstly, amino

acids shared by bovine α-crystallin (B$_2$ chain) and the two hsp16

proteins of C. elegans are indicated. Amino acid numbering of the hsp16

proteins corresponds to that shown in Figure 10. Over the regions shown,

the same amino acid occurs at 26 positions out of 74. In the sub-region of

amino acids 46 to 61 (numbering refers to the C. elegans hsps), the same

amino acid is used at 11 positions out of 16. Also indicated are the

positions at which the same amino acid is used by at least 6 of the 7

proteins. This occurs at 19 positions, 7 falling within the sub-region

discussed above while another 5 are found over a region of 11 amino acids

from positions 103 to 113. Extensive homology among hsp22, 23, 26 and 27 of

Drosophila extends approximately 30 amino acids in a 3' direction from the

74 amino acid domain (see Figure 2). This is also the case with the two

hsp16 sequences where extensive homology is maintained on the 3' side of the

functional domain and, in fact, extends to the carboxyl termini. There are

no similarities between the C. elegans and D. melanogaster sequences in this

region although a proline residue just 3' to the conserved domain shown in

Figure 10 has also been conserved (see Figure 16).

3.6   Isolation of Bristol Genomic DNA Clones

The 0.5 kb cDNA coding for the hsp16-48 variant was used to screen two

separate phage libraries, one a partial EcoRI digest of Bristol DNA in

Charon4 and the other a partial MboI digest in EMBL4. By restriction

enzyme mapping of overlapping recombinant phage clones it was possible to define a region of 30 kb which is shown in Figure 12A. Southern blotting of restriction digests was used to identify those EcoRI fragments which hybridized to the hsp16-48 cDNA. They included two fragments of 3.3 kb which comigrated and which could be distinguished by the presence or absence of a SalI site and will be referred to as 3.3L and 3.3R, respectively (Figure 12A). Another hybridizing EcoRI fragment of 1.0 kb was also identified. For more detailed restriction endonuclease analysis and for bulk preparation of specific DNA fragments, the 3.3L and 1.0 kb fragments were purified from the phage λCharon4 A-1 and subcloned into pBR325. The 3.3R fragment was subcloned into pUC13 after isolation from phage λEMBL4-2. A detailed restriction map of these three adjacent EcoRI fragments, together covering 7.6 kb, is also shown in Figure 12A.

3.7 Sequencing of the Hsp16 Genes and Identification of a Perfect 1.9 kb Inverted Repeat

A contiguous region of approximately 4.4 kb was sequenced using the strategy shown in Figure 12B. Sequence analysis identified the hsp16-48 gene and another open reading frame which defines a gene coding for another 16 kd hsp designated hsp16-1. The two genes are separated by 347 bp and are arranged in different orientations. Detailed analysis of the lambda clones revealed that this gene pair was duplicated nearby as a perfect 1.9 kb inverted repeat. This unusual organization is shown in Figure 12A. The 1.9 kb inverted repeats were separated by 416 bp of unique DNA which would form an unpaired loop, if the otherwise palindromic structure were displayed in a

Figure 12. Organization of the hsp16-48 and hsp16-1 genes of C. elegans, and the DNA sequencing strategy.

(A) Map of overlapping recombinant phage inserts and sequence organization of the small heat shock gene domain. The complete EcoRI and SalI restriction maps of the 30 kb region are shown. Solid circles at the ends of the λEMBL4 phage inserts represent Sau3A sites. The 0.9 kb and 1.6 kb EcoRI fragments at the far left (in parentheses) were identified in λCharon4 B-3, but their orientations have not yet been determined. The 7.6 kb region defined by the 3.3L, 3.3R, and 1.9 kb EcoRI fragments has been enlarged to show the detailed restriction map and gene organization. The arms of the perfect 1.9 kb inverted repeat (IR) and the orientation of the small heat shock genes are indicated. Solid bars, coding sequences; open bars, introns. The 0.9 kb BamHI-XbaI fragment which was used to detect hsp16-1 specific mRNAs is shown above the map. (B) Strategy used to sequence the 4.4 kb region which includes both arms of the inverted repeat, the loop region between them, and the distal boundary regions. All sequencing was done with M13 single-stranded phage. Arrows indicate the direction and extent of the sequence obtained from the restriction sites shown. The unlabelled sites are Sau3A restriction sites.
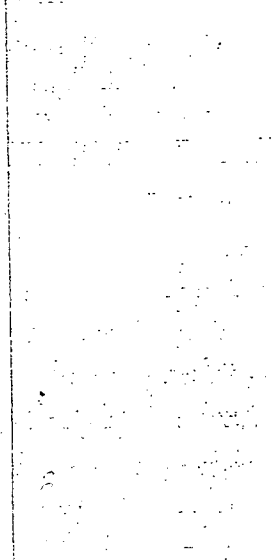
stem-loop configuration. The genes themselves will be discussed in greater detail in section 3.8.

To show that this sequence arrangement is actually present in C. elegans Bristol genomic DNA, extensive Southern blot analysis was carried out with a variety of probes (Figure 13). Probe A, the cDNA that was originally used to screen the phage libraries, identified an intense 3.3 kb EcoRI band which is consistent with the phage map. On the other hand, no evidence of the expected 1.0 kb EcoRI fragment was seen. It seemed possible that the predicted self-complementarity of this fragment might be interfering with hybridization of the cDNA, which is complementary to part of the inverted repeat. To circumvent this problem, an attempt was made to disrupt the inverted repeat structure by digesting Bristol DNA with BclI, a restriction enzyme which cleaves in the loop region. Hybridization to probe B, an M13 clone containing a 150 bp BglII-BclI insert, resulted in genomic BclI fragments of 0.8 and 0.65 kb, again consistent with the phage map. This probe also identified a 1.1 kb BglII band which was more intense since it was derived from the two arms of the inverted repeat.

The possibility that a probe which included the loop region might detect the 1.0 kb EcoRI fragment and help to identify other characteristic fragments from the inverted repeat was also investigated. Probe C, another M13 clone containing a 500 bp DdeI-EcoRI insert, of which 200 bp was specific to the loop region, was used. This probe identified the 1.0 kb EcoRI fragment along with a 1.1 kb BglII fragment, a 1.8 kb PstI fragment, a 3.75 kb SstII fragment, a 4.0 kb BamHI fragment, and a 4.1 kb XhoI fragment, all of which were predicted from the inverted repeat structure presented in Figure 12A.
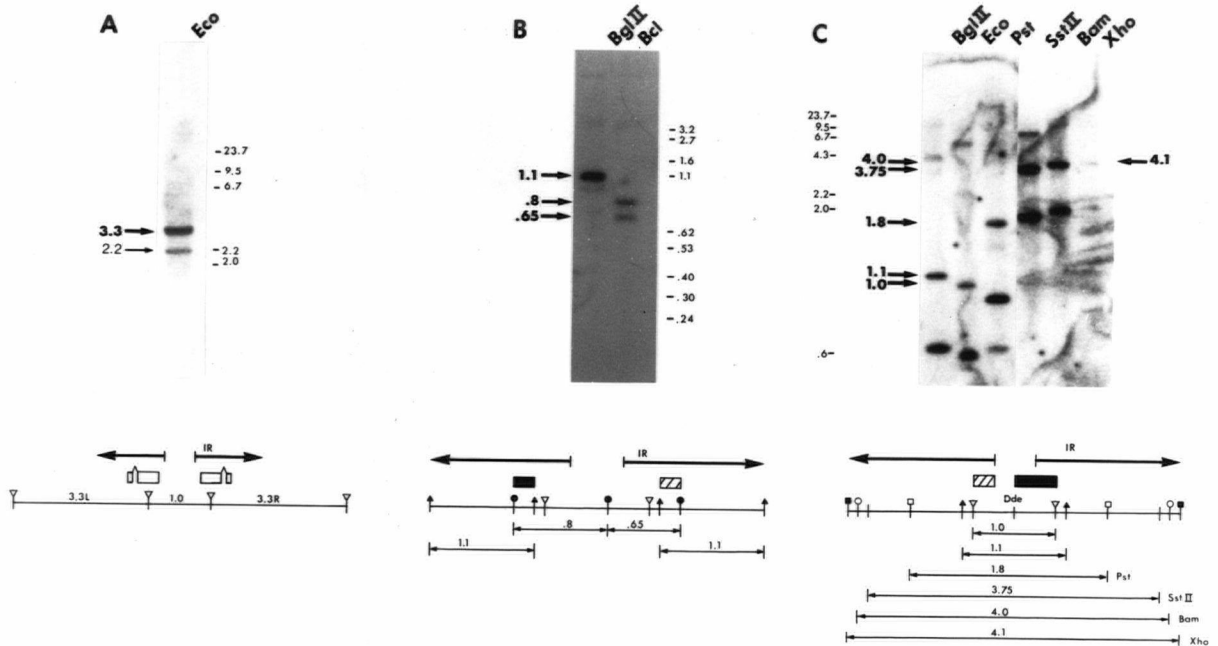
Figure 13. Analysis of C. elegans Bristol genomic DNA.

Approximately 2 μg of DNA was digested with the indicated restriction enzyme, separated by electrophoresis on either 1.5% (probe B) or 0.7% (probes A and C) agarose gels, and transferred to nitrocellulose. Probe A, CEHS16-48 cDNA; probes B and C, inserts in M13 clones. Fragment sizes are shown in kilobases. Below the results for each probe is shown the region that it represents, along with the lengths of the expected hybridizing fragments for each restriction enzyme predicted from the map in Fig. 12A. ■, Regions from which probes B and C were derived; ▨, identical sequences in the other arm of the inverted repeat (IR). The markers used with probes A and C were a HindIII digest of lambda DNA; those used with probe B were a mixture of AccI, HindII, and HpaII digests of pBR322.

In Figure 13, the hsp16-48 cDNA (probe A) hybridized also to a 2.2 kb EcoRI fragment. This is attributed to sequences coding for the hsp16-41 variant which have been cloned and characterized (Jones et al., 1986). The additional faster migrating bands seen in Figure 12C are a result of the palindromic character of this region and can be explained by the formation of foldback structures during the restriction enzyme digestions done at higher temperatures. The ability to form these structures is directly correlated with the degree of nicking or single-strandedness found in the DNA, which varied with different preparations. Thus, these foldback structures are not found in high molecular weight genomic DNA which is highly resistant to S1 nuclease (data not shown). The absence of foldback structures can be seen in Figure 23 in which a different preparation of Bristol DNA was used.

3.8 Identification of the Small Heat Shock Genes

As already mentioned, sequence analysis identified four open reading frames coding for two different hsp16s. Figure 14 shows 4.4 kb of DNA sequence which includes both arms of the 1.9 kb inverted repeat and all four shsp genes. In each case, the presumed initiation site of protein synthesis was taken to be the first ATG in the open reading frame. The gene corresponding to CEHS48 coded for a protein of 143 amino acids, 8 residues longer at the NH_2-terminal end than the protein sequence determined from the cDNA. The molecular weight of this hsp was calculated from the amino acid composition (Table V) to be 16,299, which is in good agreement with the molecular weight derived from the results of SDS-polyacrylamide gel electrophoresis.

Figure 14. Complete sequence of a small heat shock gene cluster of
C. elegans.

The sequence begins 5' to the border of the left arm of the inverted
repeat, which starts at position 114 and ends at position 2005. In this arm
we have underlined the noncoding strand of the hsp16-1 gene, and the coding
strand of the hsp16-48 gene has been translated. The converse is shown in
the right arm, which begins at position 2422 and continues to position
4313. The TATA sequences and the polyadenylation signals for all four
reading frames are boxed. The HIPs of the hsp16-1 and hsp16-48 genes are
underlined and labelled (see text for discussion). The starts of
transcription are shown with arrows over the coding strands of the hsp16-48
(position 1344) and hsp16-1 (position 3339) genes. Potential Z-DNA-forming
regions are shown straddling RsaI sites between the two pairs of genes.
Stretches (12bp) of alternating purine-pyrimidine residues are also shown
within the G + C-rich boundary sequences. They are bordered by
hexanucleotides which form inverted repeats as shown. The one-nucleotide
change between the two inverted repeats is shown at position 4304 in the
boundary sequence of the right arm.

```
                                                                                                      DDEI
              20            40            60            80            100         |   120
CTTTTGAGAAATTAAAAGCTATCCATGTTCCAATTCCTGCAAAGTTTAAATAGAAATGGCCAAAACAAATTACACTATACAAATAAAACCTCTAGCCATGCCAACTGACTCAGATATTTT
                                                                                                                      ┌── IR
                                                                                                                     BAMHI
        Z    140            160            180            200            220            240
                              XHOI
TCGGGGCCGCGCGCACGCAGCCCCGATACGCAAAAATTAAGCTGTTGGCGAATCTCAACTCGAGACTGCCACACAGGTCAGCCCATCCAAAAGTGCAATGGATAGGCCTCACCATGGGGA
 GC Boundary Sequence
              260            280            300            320            340          360
                                                                              SACII
TCCCGCCTTCATGATCGCGAGATAACCCCCAGCCAGGTAAGTTTAAGAAGTGGCGGTAAGAGAGGGTAAGGGTGTTGTATAGCATGACATCTGGCGGGTTCCGCGGACGAATGCAGAATG
                           RSAI
              380            400            420            440            460          480
TGTTAGGATGGGAGGGGCTGTGCAATACCCAAAAATAGGCATTATGCAAGAAGTACATTGCCGGCATTTACATATTTATTACTAACTTTCAAAAAATATCACAAATTCATATTATTTTTA
BGLII
       500            520            540            560            580          600
AATTTTATTGAAACAGAATACTGGAATTTATAGTAATTACATGCATAGTTCAAAAAAAATCATAAAATTACAATAATTATTCAGAAGTTTTTTGTTCAACGGGCGCTTGCTGAATTGGAAT
              620            640            660            680            700            720
AGATCTTCCTTGAATCGCTTCCTTCTTTGGTGCTTCAATTGAAAGTTTTCCATCTTCTGAAAGATTTGAAGCAACTGCACCAACATCAACATCTTCGGGTAGAAGAATAACACGAGAAAA
              740            760            780            800            820            840
TGATTTCTTTGAATATCCATGTTCAGTCTTTAATTCTTGTTCTCCTTGAATTGATAATGTATGTCCATCCAAATTAATTTTCAAATCTTCTGGCTTGAACTGCGAGACATTGAGATTTAT
 BCLI                                                      DDEI
       860            880            900        920            940            960
GGCAAACTTTTGATCATTGTTAACAATCTGAAAGAAATCTTTTTTATACAAACTCTTGAAAAAAAAAATGTGTTACTTACCTCAGAAGATTCAGATGGAGAGCCTCTGCAAACTGGAGTAA
 DDEI
       980            1000           1020           1040           1060           1080
ATTGACGTTCCATCTGAGCCATATCTCTCATGAGATCA CCAAAAACAGAACGTTGAGCTGGACGGAAATAGTGGTAAAGTGACATGATTGTAGTTTGAAGATTTCACAATTAGAGTGAA
                    ◄── HSP16-1
       1100           1120        XBAI   1140           1160           1180           1200
TGTTGTTTGGTTCGGTTTTGTCACTGTATTTATACTCATTTCCACCTTTTTCTAGAACATTCGAGCTGCTTCTTGCAAAAGGAGGGCGACTCACATTCAGAACATTGAGAAATAGTGTGC
RSAI                                        ◄─ HIP                                          Z
       1220           1240           1260           1280  XBAI  HIP ►  1300      PSTI
GTACTGAAGAAACCCAGATACTTTTTCAATCTGCGTCTCTTTGCACCTATGGGGTGTATTTGAAATGAATGCATCTAGGACCTTCTAGAACATTCTAAACGGCTGCAGGATACGGGTAT
                                                  HSP16-48 ────►  M   L   M   L   R   S   P   F   S   D   S   N   V   L   D
       1340           1360           1380           1400           1420           1440
ATAAGCCAATCGTGTTCAGAGGAAACCAATACACTTTGTTCAAGTGCTTACTGTTCATTCTCTAAACTTCAAGAATGCTCATGCTCCGTTCTCCATTTTCTGATTCAAATGTTCTCGATC
 H   F   L   D   E   I   T   G   S   V   Q   F   P   Y   W   R   N   A   D   H   N   S   F   N   F   S   D   N   I   G   E
                                                    BCLI
       1460           1480           1500           1520           1540           1560
ATTTCTTGGATGAAATCACTGGATCTGTTCAATTTCCATATTGGAGAAATGCTGATCACAACTCATTCAATTTTTCCGACAATATTGGAGAGGTAAGAAAATAATCTCTTTTTCAATTGT
         I   V   N   D   E   S   K   F   S   V   Q   L   D   V   S   H   F   K   P   E   D   L   K   I   E   L   D   G   R   E   L
                                                                  BGLII
       1580           1600           1620           1640           1660           1680
TTATTTGTCAAATGTTTTATTTTTCAGATTGTAAATGACGAATCTAAATTCTCTGTTCAACTCGATGTTTCTCATTTCAAACCAGAAGATCTTAAAATTGAATTGGATGGAAGAGAACTA
     K   I   E   G   I   Q   E   K   K   S   E   H   G   Y   S   K   R   S   F   S   K   M   I   L   L   P   E   D   V   D   L   T   S   V   K   S   A   I   S   N
         ECORI
       1700           1720           1740           1760           1780           1800
AAAATTGAAGGAATTCAAGAAAAAAAATCAGAGCATGGATACTCGAAACGATCATTTTCAAAAATGATTCTTCTACCAGAAGATGTTGATTTAACTTCTGTCAAATCTGCAATTTCGAAT
     E   G   K   L   Q   I   E   A   P   K   K   T   N   S   S   R   S   I   P   I   N   F   V   A   K   H   *
       1820           1840           1860           1880           1900           1920
GAAGGAAAACTTCAAATTGAGGCTCCAAAGAAGACTAACTCATCTCGTTCTATTCCCATTAATTTTGTTGCAAAACATTAATCTTTTATTGTATTCCAAATATTCTTAATTTCAATAAAG
       polyA site
          ↓ 1940           1960           1980           2000           2020           2040
TCATTAATTTAATTTATTCATGTTCTCTAGCATAACAAAAACATCAAATCCGACTTTCCAATTCAAATATTTCAAAACAACATAATTGAAGTTATTCCAGAAACTTTTTATGCAAAAAAT
                                                                          IR ◄─┘
       2060           2080           2100           2120           2140           2160
TTATTAGTCTCAATAAATGTTTTAGCTTGAATTTATGCTTAAAACAAAAAAACATAAAAAAATGTTTTAAAAAAATTACAGTGCGTGCAACTTCTACCAGGGCCCTCATAAAATAGGTTCTT
                                 DDEI                                                              BCLI
       2180           2200           2220           2240           2260           2280
CAAGAGAAAAATCAGAAAATATTTTTAAGCTACCAGCTACTTCGGCTTCTCAGAAAAACTCCTGCGACTTGATAAAAACACTTAAAAACTGTCAAAAGCTGTTATAAAGGTGCAGTTGATCA
```

```
          2300          2320          2340          2360          2380          2400
ACACAGTTGTGAGCGAAATTAAGCTGAAAAACTGAATTTTGAAAAAGTTGCAAAAGTCGTTTGAAACGATGCAGGAATTTTTATAGCTAAACGTTTAACGGTTGGTAAAACGTAGCATGA

          2420          2440          2460          2480          2500          2520
GTATGCTGTAAAGTTGAGCCGTTATGTTGTTTTGAAATATTTGAATTGGAAAGTCGGATTTGATGTTTTTGTTATGCTAGAGAACATGAATAAATTAAATTAATGAC TTTATT GAAATTA
                    └──→IR

          2540          2560          2580          2600          2620          2640
AGAATATTTGGAATACAATAAAAGATTAATGTTTTGCAACAAAATTAATGGGAATAGAACGAGATGAGTTAGTCTTCTTTGGAGCCTCAATTTGAAGTTTTCCTTCATTCGAAATTGCAG

                                                                              ECORI
          2660          2680          2700          2720          2740          2760
ATTTGACAGAAGTTAAATCAACATCTTCTGGTAGAAGAATCATTTTTGAAAATGATCGTTTCGAGTATCCATGCTCTGATTTTTTTTCTTGAATTCCTTCAATTTTTAGTTCTCTTCCAT

         BGLII
          2780          2800          2820          2840          2860          2880
CCAATTCAATTTTAAGATCTTCTGGTTTGAAATGAGAAACATCGAGTTGAACAGAGAATTTAGATTCGTCATTTACAATCTGAAAAATAAAACATTTGACAAATAAACAATTGAAAAAGA

                       BCLI
          2900          2920          2940          2960          2980          3000
GATTATTTTCTTACCTCTCCAATATTGTCGGAAAAATTGAATGAGTTGTGATCAGCATTTCTCCAATATGGAAATTGAACAGATCCAGTGATTTCATCCAAGAAATGATCGAGAACATTT

                                                                              PSTI
          3020          3040          3060          3080          3100          3120
GAATCAGAAAATGGAGAACGGAGCATGAGCATTCTTGAAGTTTAGAGAATGAACAGTAAGCACTTGAACAAAGTGTATTGGTTTCCTCTGAACACGATTGG TTATATA CCCGTATCCTG

     XBAI        ←──HSP16-48                                                RSAI
          3140          3160          3180          3200          3220          3240
CAGCCGTTTAGAATGTTCTAGAAGGTCCTAGATGCATTCATTTCAAAATACACCCCATAGGTGCAAAGAGACGCAGATTGAAAAAGTATCTGGGTTTCTTCAGTACGCACACTATTTCTC
  ←HIP                                                                        ‾‾‾‾
                                                                              Z
                         XBAI
          3260          3280          3300          3320          3340          3360
AATGTTCTGAATGTGAGTCGCCCTCCTTTTGCAAGAAGCAGCTCGAATGTTCTAGAAAAAAGGTGGAAATGAG TATAAAT ACAGTGACAAAACCGAACCAAACAACATTCACTCTAATTGT
                          HIP →
HSP16-1 ─────→  M  S  L  Y  H  Y  F  R  P  A  Q  R  S  V  F  G  Y  L   M  R  D  M  A  Q  M  E  R  Q  F  T  P  V  C
          3380          3400          3420          3440          3460          3480
GAAATCTTCAAACTACAATCATGTCACTTTACCACTATTTCCGTCCAGCTCAACGTTCTGTTTTTGG TGATCTCATGAGAGATATGGCTCAGATGGAACGTCAATTTACTCCAGTTTGC
                                                                              DDEI
 R  G  S  P  S  E  S  S  E                                      I  V  N  N  D  Q  K  F  A  I  N  L  N  V
     DDEI                                                                     BCLI
          3500          3520          3540          3560          3580          3600
AGAGGCTCTCCATCTGAATCTTCTGAGGTAAGTAACACATTTTTTTTTCAAGAGTTTGTATAAAAAAGATTTCTTTCAGATGTTGTTAACATGATCAAAAGTTTGCCATAAATCTCAATGT

 S  Q  F  K  P  E  D  L  K  I  N  L  D  G  H  T  L  S  I  Q  G  E  Q  E  L  K  T  E  H  G  Y  S  K  K  S  F  S  R  V  I
          3620          3640          3660          3680          3700          3720
CTCGCAGTTCAAGCCAGAAGATTTGAAAATTAATTTGGATGGACATACATTATCAATTCAAGGAGAACAAGAATTAAAGACTGAACATGGATATTCAAAGAAATCATTTTCTCGTGTTAT

 L  L  P  E  D  V  D  V  G  A  V  A  S  N  L  S  E  D  G  K  L  S  I  E  A  P  K  K  E  A  I  Q  G  R  S  I  P  I  Q  Q
                                                                      BGLII
          3740          3760          3780          3800          3820          3840
TCTTCTACCCGAAGATGTTGATGTTGGTGCAGTTGCTTCAAATCTTTCAGAAGATGGAAAACTTTCAATTGAAGCACCAAAGAAGGAAGCGATTCAAGGAAGATCTATTCCAATTCAGCA

 A  P  V  E  Q  K  T  S  E  •
          3860          3880          3900          3920          3940          3960
AGCGCCCGTTGAACAAAAAACTTCTGAATAATTATTGTAATTTTTATGATTTTTTTGAACTATGCATGTAATTACTATAAATTCCAGTATTCTGTTTC AATAAAA TTTAAAAATAATATGA

                                                  RSAI
          3980          4000          4020          4040          4060          4080
ATTTGTGATATTTTTTGAAAGTTAGTAATAAATATGTAAATGCCGGCAATGTACTTCTTGCATAATGCCTATTTTTGGGTATTGCACAGCCCCTCCCATCCTAACACATTCTGCATTCGT
SACII                                                                         BAMHI
          4100          4120          4140          4160          4180          4200
CCGCGGAACCCGCCAGATGTCATGCTATACAACACCCTTACCCTCTCTTACCGCCACTTCTTAAACTTACCTGGCTGGGGGTTATCTCGCGATCATGAAGGCGGGATCCCCATGGTGAGG

               XHOI                                                   Z   4300   ▼
          4220          4240          4260          4280          4300          4320
CCTATCCATTGCACTTTTTGGATGGGCTGACCTGTGTGGCAGTCTCGAGTTTGAGATTCGCCAACAGCTTAATTTTTGCGTATCGGGGGCTGCGTGCGCGCGGCCCTGAAAAATATATTAATA
                       RSAI                                           GC Boundary Sequence ‾‾‾‾‾‾‾‾‾‾
                                                                                           IR
     4335          4350          4365          4380          4395          4410
CTGTTTTCAATAGTAATTGGCATAACGAATCAGTTTCAAGTAAAACGTACGTCAATTTCAGTAGCAGAACGTTCAAACAGTTTTCTATCTGATC
```

Table V. Amino Acid Compositions of Hsp16-1 and Hsp16-48

Amino Acid	No. of residues	
	hsp16-1	hsp16-48
Ala	8	4
Arg	7	5
Asn	6	9
Asp	7	11
Cys	1	0
Gln	11	4
Glu	13	12
Gly	8	6
His	3	5
Ile	9	11
Leu	11	12
Lys	10	13
Met	4	3
Phe	6	9
Pro	8	6
Ser	16	19
Thr	4	3
Trp	0	1
Tyr	4	2
Val	9	8

The other gene codes for a protein of 145 amino acid residues with a calculated molecular weight of 16,301. To determine whether this gene is transcribed under heat shock conditions, a 0.9 kb BamHI-XbaI fragment containing the gene for hsp16-1 (Figure 12A) was used to probe dot blots of polyA$^+$RNA purified from control and heat shocked nematodes. The results of this experiment are shown in Figure 15. Transcripts specific to the hsp16-1 gene were abundant in mRNA isolated from heat shocked nematodes but were not detectable in normally growing nematodes. This is similar to the high inducibility demonstrated for mRNA sequences complementary to CEHS48 and CEHS41 (Figure 8). Thus, two distinct, heat inducible genes have been identified in this region, and this gene pair has been duplicated to form an inverted repeat.

Hsp16-1 shows extensive homology with hsp16-48 and hsp16-41, and the three predicted amino acid sequences have been aligned in Figure 16. The NH$_2$-terminal regions of hsp16-48 and hsp16-1 are completely different. It is evident that this divergent region defines an exon at the level of gene organization (section 3.9). The rest of the polypeptide chain which defines the second exon, is very similar among the three C. elegans hsps, 55 of 98 positions being identical. This homologous region contains the approximately 80 amino acid domain that shows homology to the shsps of Drosophila melanogaster and to the α-crystallin proteins. In this domain the same amino acid is used by at least four of the five proteins in 32 of 80 positions. Also shown are the 20 positions that are conserved between the proteins shown and the shsps of Drosophila.
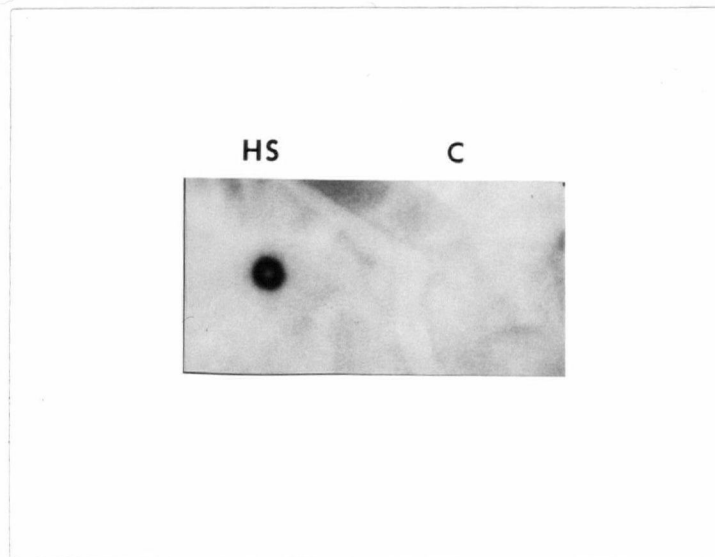
Figure 15.   RNA dot-blot analysis using a probe specific for the hsp16-1 gene.

0.5 μg of polyA$^+$RNA purified from control or heat shocked worms was applied to nitrocellulose before baking and hybridizing.   A 0.9 kb BamHI-XbaI fragment containing a hsp16-1 gene was used as the probe.

Figure 16. Comparison of the deduced amino acid sequences of three shsps of

$\underline{C}$. elegans, the mouse αA_2-crystallin chain, and the bovine

αB_2-crystallin chain.

The proteins have been aligned to demonstrate sequence homology and to

emphasize the similar locations of introns. The location of the intron in

the hsp16-41 gene is taken from Jones et al. (1986). Most of the mouse

αA_2-crystallin amino acid sequence is taken from King et al. (1982), and

the NH_2-terminal portion and the location of the two introns are taken

from King and Piatigorsky (1983). The protein sequence of bovine

αB_2-crystallin was included to illustrate the relationship between the

two α-crystallin classes and is taken from van der Ouderaa et al. (1973).

It is not known at this time whether the bovine αB_2-crystallin gene

contains introns, although the hamster αB_2-crystallin gene has

maintained the exon/intron organization shown by the murine

αA_2-crystallin gene (see section 3.9). Brackets show the borders of the

74 amino acid region which was used in Figure 11. Positions of homology in

which the same amino acid is shared by at least four of the five proteins

are boxed. Solid circles indicate positions that have been conserved

between these five proteins and the shsps of $\underline{D}$. melanogaster.

```
HSP16-41
HSP16-48                                                     (M) L M L R S P F S D S N V L D H F L D E I
HSP16-1                                                    (M) S L Y H Y R F P A Q R S V F G Y
alpha-A2 (mouse)    (M) D V T I Q H P W F K R A L G P F - Y P S R L F D Q F F G E G L F E Y D L L P F L S S
alpha-B2 (bovine)   (M) D I A I H H P W I R R P F F P F H S P S R L F D Q F F G E H L L E S D F L P - A S T


                                                                        •         •    • •      •
                              - 58 bp Intron        - I V N D [E S |K F| S V Q |L D| V |S H F| K
T G S V Q F P Y W R N A D H N S F N F S D N I G E   - 55 bp Intron    - I V N D  E S |K F| S V Q |L D| V |S H F| K
L M R D M A Q M E R Q F T P V C R G S P S E S S E   - 52 bp Intron    - I V N N  D Q |K F| A I N |L N| V |S Q F| K
T I S P Y Y - R Q - S L F R - - T V L D S G I S E   - 1376 bp Intron  - V R S D  R D |K F| V I K |L D| V |K H F| S
S L S P F Y L R P P S F L R A P S W I D T G L S E   - Intron?         - M R L E |K D| R F S V N |L L| N V |K H F| S


  •      •           •           •   •                                  • •      •   •            •
P E N |L K| I K L D |G R E L K |I E |G I Q E| T K S  - No Intron      | - - |E H G Y L K R |S F| S K M I L |L P E D A D| L
P E D |L K| I E L D |G R E L K |I E |G I Q E| K K S  - No Intron        - - |E H G Y S K R |S F| S K M I L |L P E D V D| L
P E D |L K| I N L D |G H T L S |I Q |G E Q E| L K I  - No Intron        - - |E H G Y S K K |S F| S R V I L |L P E D V D| V
P Q D |L T| V K V L |E D F V E |I H |G K H N| E R Q  - 2nd Intron      - D D |H G Y I S R E |F H| R R Y R L |P S N V D| Q
P E E |L K| V K V L |G D V I E |V H |G K H E| E R Q  - Intron?         - D E |H G F I S R E |F H| R K Y R I |P A D V D| P


      •      •      •       •      •                      •
P S V K |S A I S| N E |G K |L Q I E A |P K K| T N S S R S I P I N F V A K H
T S V K |S A I S| N E |G K |L Q I E A |P K K| T N S S R S I P I N F V A K H
G A V A |S N L S| E D |G K |L S I E A |P K K| E A I Q G R S I P I Q Q A P V E Q K T S E
S A L S |C S L S| A D |G M |L T F S G |P K V| Q S G L D A G H S E R A I P V S R E E K P - - S S A P S S
L A I T |S S L S| S D |G V |L T V D G |P R K| Q - - - - A S G P E R T I P I T R E E K P A V T A A P K K
```

3.9 The shsp Genes of C. elegans Contain a Single Intron

Comparison of the hsp16-48 gene with the CEHS48 cDNA revealed a short intron of 55 bp which divided the coding region into two exons. The hsp16-1 gene had a corresponding intron of 52 bp. The exon-intron boundary sequences match well with the consensus sequence described by Mount (1982). The lack of long pyrimidine tracts preceding the 3' splice junction is common to other C. elegans introns (Spieth et al., 1985) and may simply be a consequence of size restraints, the average intron being approximately 50 bp.

The homology between the shsps of C. elegans and vertebrate α-crystallin was also reflected in their gene organization. The position of the intron in the nematode heat shock genes was precisely analogous to the position of the first intron in the αA_2-crystallin genes of mouse (King and Piatigorsky, 1983) and hamster (van den Heuvel et al., 1985) or the hamster αB_2-crystallin gene (Quax-Jeuken et al., 1985). The αA_2-crystallin genes contain within their first intron a 23 amino acid insertion sequence which is expressed in 10% of the messages due to alternative splicing. Interestingly, all of the crystallin genes mentioned above contain a second intron, not found in the hsp genes, which interrupts the conserved domain. Introns have not been found in the related shsp genes from Drosophila (Ingolia and Craig, 1982b; Southgate et al., 1983; Ayme and Tissières, 1985), Xenopus (Bienz, 1984b) or soybean (see Table I). On the other hand, the functional gene coding for human hsp25 contains two introns (Hickey et al., 1986). Although human hsp25 shows more similarity to the α-crystallins than to any other shsps, the locations of the two introns have not been conserved. In the Caenorhabditis hsp16 genes and the vertebrate α-crystallin genes, the first splice junction precedes the

region which shows the greatest homology between the shsps and the α-crystallins. In the human hsp gene, the second exon corresponds to the highly hydrophilic region of 23 amino acids which is found within the larger 80 amino acid conserved domain.

3.10 Location of the Starts of Transcription

The hsp16-1 and hsp16-48 genes are clustered within a relatively short region of DNA. Furthermore, they are transcribed in opposite directions, with the result that only 347 bp separate their respective ATG initiation codons. Therefore, it is likely that all signals required for accurate heat inducible transcription will be located in this region.

S1 nuclease protection experiments (Figure 17) suggested that the transcripts began with adenine residues located at positions -42 and -51 for the hsp16-1 and hsp16-48 genes, respectively (relative to their ATG initiation codons). In Figure 17, the 5' flanking sequences of the two genes have been aligned to compare the regions upstream from the putative transcription starts. In both genes there is a Goldberg-Hogness or TATA box (M.L. Goldberg, Ph.D. thesis, Stanford University, Stanford, CA, 1979), the first thymine of which fell exactly 26 bp from the transcription start. Approximately 30 bp further upstream were found excellent matches to the consensus sequence CT-GAA--TTC-AG, which was derived by comparing the 5' flanking regions of the D. melanogaster heat shock genes (Pelham, 1982). A synthetic promoter containing this sequence has been shown to confer heat inducibility on non-heat inducible genes such as the Herpes simplex virus thymidine kinase gene (Pelham and Bienz, 1982). The hsp16-48 gene contained

Figure 17. S1 nuclease protection analysis of transcription start sites of

the hspl6-48 and hspl6-1 genes, and comparison of the 5' flanking regions.

(A) S1 nuclease mapping. Specific 5'-end-labelled double-stranded DNA

fragments were allowed to anneal to heat shock poly(A)$^+$RNA under the

hybridization conditions described. S1 nuclease-resistant fragments were

separated on denaturing polyacrylamide gels along with the Maxam and Gilbert

cleavage reactions of the same end-labelled fragment. RNA-, Control lanes

in which the DNA fragment was taken through the same hybridization and S1

digestion reactions in the absence of mRNA. (B) The 5' flanking sequences

of both small heat shock genes were aligned to show the relationship between

the transcription starts (arrows), the TATA boxes, and the HIPs. For each

gene, the A in the first ATG after the mRNA start is numbered + 1. This

numbering has also been used to designate the restriction fragments which

were used in the S1 mapping experiments (section 2.3.3). The HIP elements

have been underlined to demonstrate their palindromic nature, solid vertical

lines representing the dyad axis. The hspl6-48 gene has two overlapping HIP

elements, the second indicated with dots. The dashed vertical line shows

the central axis of a large 22 bp palindrome with only three mismatches.

A

HSP 16-1

GC
&&
GATC   RNA
        + −

5: CGAACCAAACAACATTCACTCT :3
3: GCTTGGTTTGTTGTAAGTGAGA :5

HSP 16-48

CG
&&
CTAG   RNA
        + −

5: TCAGAGGAAACCAATACACTTT :3
3: AGTCTCCTTTGGTTATGTGAAA :5

B

```
          -120                              -100        HIPs              -80
HSP16-1   G C C C T C C T T T T T G C A A G A A G C A G C T C G A A T G T T C T A G A A A A A G G T G G A
HSP16-48  A A A T G A A T G C A T C T A G G A C C T T C T A G A A C A T T C T A A A C G G C T G C A G G
          -130                                -110                              -90

                     -60                                        -40
          A A T G A G T A T A A A T A C A G T G A C A A A A C C G A A C C A A A C A A C A T T C A C T C
          A T A C G G G T A T A T A A G C C A A T C G T G T T C A G A G G A A A C C A A T A C A C T T T
                                           -70

                  -20                             -1
          T A A T T G T G A A A T C T T C A A A C T A C A A T C A T G T C A C T T T A C C A C T A T T T
          G T T C A A G T G C T T A C T G T T C A T T C T C T A A A C T T C A A G A A T G
                     -30                             -10                    -1
```

two of these heat inducible promoters (HIPs), which overlapped each other to form a 22 bp inverted repeat with only three mismatches (Figure 17).

Only 128 bp separated the HIP sequences of the hsp16-1 and hsp16-48 genes. Within this region is a 10 bp stretch of alternating purine-pyrimidine residues overlapping the RsaI site (Figure 14). Such sequences can potentially adopt a left-handed Z-DNA conformation (Arnott et al., 1980), and have been implicated in the regulation of transcription of simian virus 40 as well as other DNA viruses and retroviruses (Nordheim and Rich, 1983). Adjacent to this region was a 35 bp palindromic structure with the potential to form a stem-loop structure. This structure would contain a 12 bp stem with 2 mismatches and an 11 bp looped-out region. This region is also shown in Figure 27 and will be discussed in more detail in section 4.4.

The leaders of the hsp16-1 and hsp16-48 transcripts were 42 and 51 nucleotides respectively, the differences in length being due to insertions/deletions (see also Figure 26, section 4.4). When aligned to compensate for these events, the leaders show a substantial degree of similarity, the sequences of which may be conserved to maintain selective translation at higher temperature.

3.11 The 3' Flanking Regions

The polyadenylation signal AATAAA is found in the 3' noncoding region of both the hsp16-1 and hsp16-48 genes (Figure 14). In the hsp16-48 gene it was found 33 bp downstream from the TAA termination codon . Sequence analysis of the cDNA placed the polyadenylation site 12 nucleotide downstream of the signal (Figure 10), and 74 bp before the end of the

inverted repeat. The predicted length of the hsp16-48 transcript was 533 nucleotides after splicing and processing have been completed, not including the poly(A) tail.

The hsp16-1 gene contains two polyadenylation signals, one beginning at position 3937 and the second at position 3987 of Figure 14. Although a cDNA corresponding to hsp16-1 was not identified, S1 nuclease mapping data suggests that the proximal one is functional in nematodes (Rob Kay, personal communication), the site of polyadenylation occurring at position 3961. This results in a 3' untranslated region of 90 nucleotides and a spliced transcript with a predicted length of 570 nucleotides.
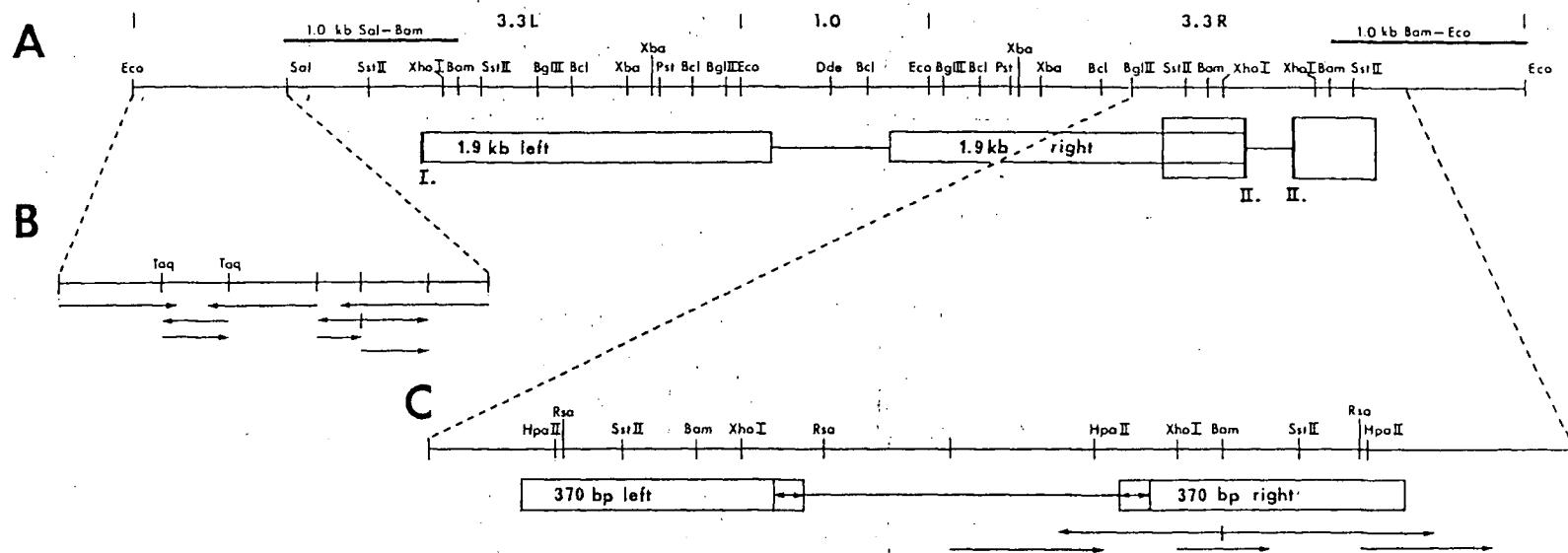
3.12 Organization of Inverted Repeats in the Region Containing Genes Hsp16-1/48

Unusual sequences were found at the distal boundaries of the 1.9 kb inverted repeat. This G + C-rich boundary sequence (Figure 14) is characterized by a stretch of 12 alternating purine-pyrimidine residues, flanked on either side by a G + C hexanucleotide in an inverted repeat orientation. The single nucleotide difference between the two arms of the 1.9 kb inverted repeat was found in this boundary sequence. It falls on one of the flanking hexanucleotides, where a cytosine is changed to a thymine. Further sequence analysis of 3.3R revealed the presence of another 370 bp inverted repeat which overlaps the one containing the heat shock genes. Its organization is shown in Figure 18. Unlike the larger one, this inverted repeat was not perfect, having diverged by approximately 4%. Its right arm, however, contained a G + C-rich boundary sequence which shared perfect homology to its counterpart. In the smaller inverted repeat, the boundary

Figure 18. Organization of inverted repeats and G + C-rich boundary

sequences in locus hsp16-1/48.

A) Restriction map of the 7.6 kb region consisting of 3.3L, 3.3R, and

the 1.0 kb EcoRI fragment. The location of the 1.9 kb inverted repeat

containing the heat shock genes and the overlapping 370 bp inverted repeat

are indicated by open boxes. The two variant forms of the G + C-rich

boundary sequences are shown at the borders of the inverted repeats by heavy

lines. Also shown, above the restriction map, are the regions which were

analyzed for RNA transcription. B) Strategy used to determine the DNA

sequence of the 0.9 kb EcoRI-SalI fragment. C) Detailed map of the 370 bp

inverted repeat structure showing the palindromic boundary sequences, the

strategy used to sequence the right arm, and the locations of the RsaI and

HpaII restriction sites within the loop region which were used to disrupt

the foldback structure. B) and C) All sequencing was done using M13

single-stranded phage, the direction and extent of the sequence obtained

from each restriction site shown with arrows. All of the unlabelled

restriction sites are Sau3A sites.

A

1.0 kb Sal—Bam   3.3L   1.0   3.3R   1.0 kb Bam—Eco

Eco   Sal   SstⅡ   XhoⅠ Bam SstⅡ   BgⅢ Bcl   Xba   Pst Bcl BgⅢEco   Dde Bcl   Eco BgⅢ Bcl Pst   Xba   Bcl BgⅢ SstⅡ Bam XhoⅠ   XhoⅠ Bam SstⅡ   Eco

Xba   Xba

1.9 kb left   1.9 kb right

I.   II. II.

B   Taq   Taq

C   Rsa   Rsa

HpaⅡ   SstⅡ   Bam   XhoⅠ   Rsa   HpaⅡ   XhoⅠ Bam   SstⅡ   HpaⅡ

370 bp left   370 bp right

sequences were located on the proximal borders, being separated by

approximately 500 bp of unique sequence. The distal boundaries were

characterized by a complete divergence of sequence, the left one

interestingly located precisely at the functional polyadenylation signal of

the hsp16-1 gene in the right arm of the 1.9 kb inverted repeat.

Since the boundary sequences of the larger inverted repeat precisely

defined a locus of transcriptional activity, an effort was made to locate

any genes that may have been to the right of the smaller inverted repeat.

Northern blot analysis showed that there were no detectable messages in

control polyA$^-$, control polyA$^+$, or heat shock polyA$^+$RNA using as a

probe the 3.3R-derived 1.0 kb BamHI-EcoRI fragment (not shown). Similar

negative results were obtained using the 1.0 kb BamHI-SalI fragment of 3.3L

which flanks the large inverted repeat on the other side. The locations of

both of these probes are shown in Figure 18. Thus it would appear that

there are no transcribed regions immediately adjacent to the inverted

repeats of locus hsp16-1/48 unless the transcripts accumulate to levels

below the limits of detection. Furthermore, the completed sequence of the

0.9 kb EcoRI-SalI segment in 3.3L (Figure 18) contained no open reading

frames of substantial length in either direction.

To examine the frequency of occurrence of the boundary sequences in the

C. elegans genome, oligodeoxynucleotides specific for the two variant forms

were used as probes. These oligodeoxynucleotides included the alternating

purine/pyrimidine stretch but contained only one of the hexanucleotide

repeats to avoid self-complementarity. The single base difference between

the two oligodeoxynucleotides is located in this hexanucleotide. The

results, using variant I as probe, are shown in Figure 19. Under conditions
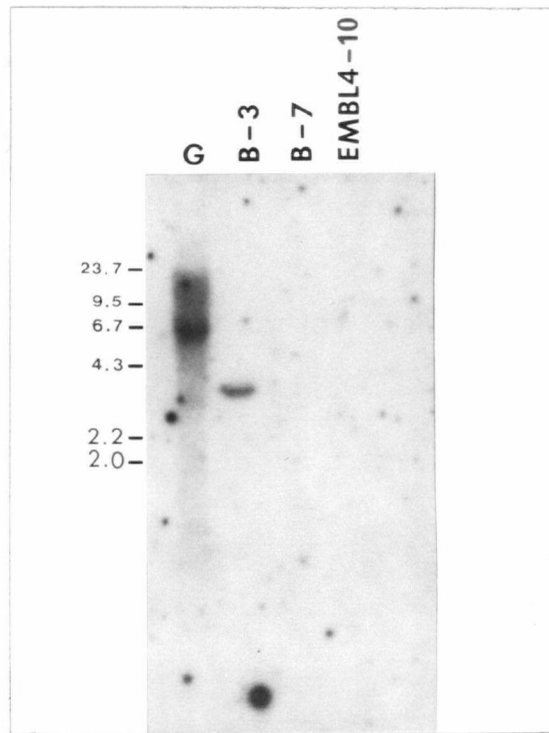
Figure 19. Analysis of G + C-rich boundary sequences in various recombinant phage and Bristol genomic DNA.

Approximately 2.0 µg of Bristol genomic DNA(G) and 0.5 µg of DNA from the indicated phage were digested with EcoRI, transferred to nitrocellulose after agarose gel electrophoresis and hybridized to oligodeoxynucleotide variant I. The markers are a HindIII digest of lambda DNA. Phage clones B-3 and B-7 contain the left and right arms of the 3.3 kb inverted repeat structure, respectively. Clone B-7 includes also the overlapping 370 bp inverted repeat structure. Phage EMBL4-10 contains DNA representing the hsp16-2/41 locus in which the hsp16 genes are not duplicated.

in which only 3.3L from phage λCharon4 B-3 (Figure 12) hybridized to the

oligodeoxynucleotide specific to the boundary sequence in that fragment,

genomic Bristol DNA digested with EcoRI gave a heterodisperse series of

bands which indicates that such sequences are abundant in the nematode

genome. The signal was not abolished at washing temperatures as high as

53°C in 6 X SSPE.

This oligodeoxynucleotide, surprisingly, did not hybridize to 3.3R from

phage λCharon4 B-7 which contains at least two copies of the variant

form. To ensure that the lack of hybridization was not due to the single

base pair mismatch, an oligodeoxynucleotide specific for the boundary

sequences found in 3.3R was used as a probe (variant II). Identical results

to those shown in Figure 19 were obtained even at low washing temperatures

of 43°C. Considering the difficulties encountered in the genomic DNA

analysis of the 1.9 kb inverted repeat, it was possible that the lack of

hybridization to 3.3R was once again a result of the formation of foldback

structures derived from the smaller 370 bp inverted repeat. To examine this

possibility, the inverted repeat of 3.3R was disrupted with various

restriction enzymes and probed again with the oligodeoxynucleotide. As

shown in Figure 20, there is relatively little hybridization to the 3.3R

EcoRI fragment or the derived 700 bp BamHI fragment in which the inverted

repeat structure is maintained. Digestion of the same amount of DNA with

either HpaII or RsaI results in strong hybridization to two fragments of 700

bp and 350 bp which contain the boundary sequence.

Neither oligodeoxynucleotide hybridized to any EcoRI fragments derived

from λEMBL4-10, the recombinant phage which contains a pair of related

hsp16 genes (Jones et al., 1986). Although this suggests that sequences
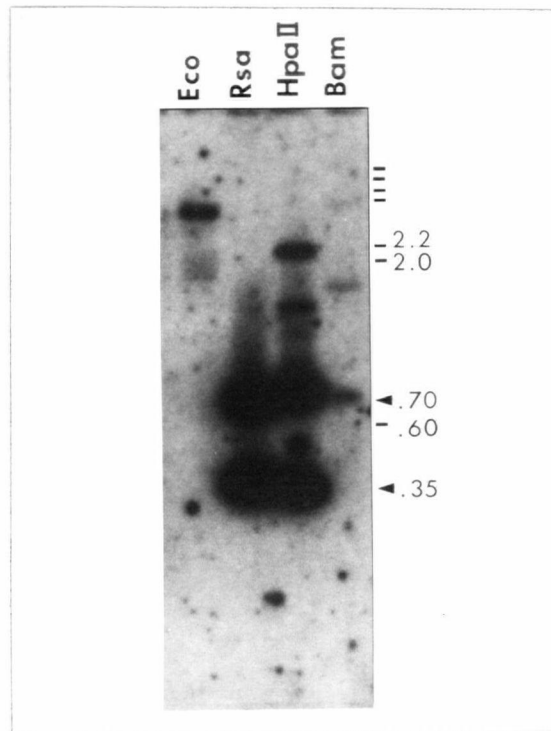
Figure 20. Detection of boundary sequences in 3.3R by disruption of the 370

bp inverted repeat.

Approximately 0.5 μg of a recombinant pUC13 plasmid containing the

3.3R EcoRI fragment was digested with the restriction enzymes indicated.

After electrophoresis and Southern transfer, the DNA was hybridized to

oligodeoxynucleotide variant II. The markers are a HindIII digest of lambda

DNA.

homologous to the boundary sequences at locus hsp16-1/48 may not be

associated with the other heat shock gene locus, it is possible that their

association with foldback structures dictates a more rigorous analysis with

different restriction enzymes.

In summary, G + C-rich sequences similar or identical to the boundary

sequences described here are abundant in the nematode genome. Their

association with inverted repeat structures in the present instance is

interesting but whether they can be generally correlated with other regions

of similar structure remains to be seen.

3.13 Differential Expression of Hsp16 Genes in C. elegans

A pair of related hsp16 genes in C. elegans has been characterized by

sequencing (Jones et al., 1986). The two genes are arranged in divergent

orientations similar to the hsp16-48 and hsp16-1 genes. The two genes are

designated hsp16-2 and hsp16-41, the latter corresponding to the cDNA

CEHS41, and both are highly expressed during heat shock. The gene pair at

locus hsp16-2/41 is not duplicated to form an inverted repeat but is flanked

by repetitive elements. To determine the relative levels of transcription

from the four hsp16 genes, M13 probes were used that showed minimal

cross-hybridization between homologous genes. Probes corresponding to the

second exons were avoided since the carboxy-terminal halves of the hsp16-48

and hsp16-41 genes, represented by their respective cDNAs, cross-hybridize

strongly. Thus probes corresponding to either the 3' untranslated regions

(genes 1 and 2) or exon 1 (genes 48 and 41) were used. These M13 clones are

listed in the legend to Figure 21. The specificities of these probes were

verified by hybridizing them to Southern transfers containing EcoRI digests
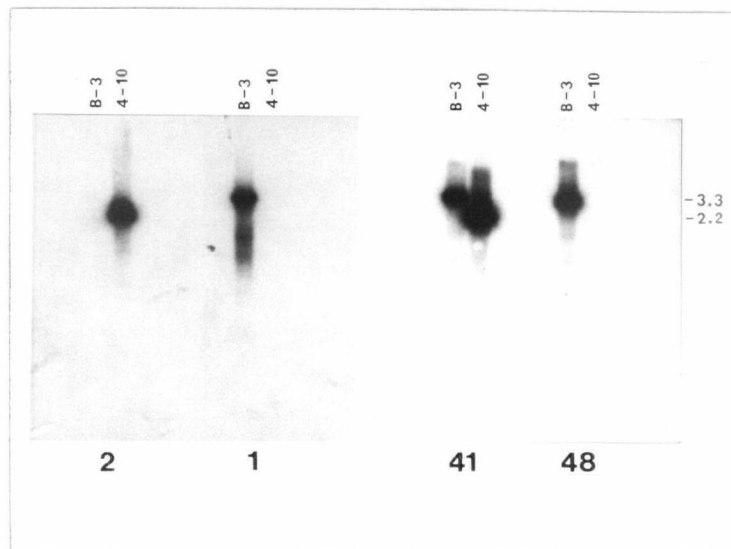
Figure 21. Determination of the specificity of various probes used in the mRNA expression studies.

Approximately 500 ng of EcoRI-digested λCharon4 B-3 and λEMBL4-10 phage DNA was separated on agarose gels, transferred to nitrocellulose and hybridized to probes derived from the hsp16 genes which are indicated. The probes were derived from M13 templates by primer extension and include the following sequences: 16-1, Sau3A (253) - BglII (601); 16-48, BglII (1647) - BclI (1494); 16-2, TaqI (1974) - TaqI (1910); 16-41, TaqI (933) - TaqI (1169). For the hsp16-1 gene and hsp16-48 gene probes, the numbering is taken from Figure 14 while the numbering for the hsp16-2 and hsp16-41 gene probes is taken from Jones et al. (1986). Phage λCharon4 B-3 contains a 3.3 kb EcoRI fragment which codes for hsp16-1/48 while λEMBL4-10 contains a 2.2 kb EcoRI fragment which codes for hsp16-2/41. 2×10^{6} Cerenkov counts were used in each hybridization.

of the phage λCharon4 B-3 and λEMBL4-10 which correspond to the 1/48 and

2/41 loci, respectively (Figure 21). This experiment also verified that the

probes were of comparable specific activities. Of the four probes used,

only the one selected for the hsp16-41 gene showed some degree of

cross-hybridization to both phage clones but as the mRNA expression data

show, this was of no consequence.

When these probes were applied individually to Northern transfers of

polyadenylated RNA from heat shocked and control nematodes, mRNA levels from

the hsp16-2/41 genes appeared to be 10 to 20 times higher than those from

the genes hsp16-48 and hsp16-1 (Figure 22). Since there are presumably two

functional copies of the latter gene pair, the relative transcription level

per gene at the hsp16-2/41 locus is approximately 20 - 40 times that of the

hsp16-1/48 locus. Also, these blots indicated that the polyadenylated hsp16

transcripts are between 600 and 700 nucleotides in length, as expected.

3.14 Comparison of Locus Hsp16-1/48 in Caenorhabditis elegans Bristol and

Bergerac Strains

The related strains Bristol and Bergerac can interbreed and thus any

restriction fragment length differences (RFLDs) can be used as genetic

markers to determine linkage. This can be done by making use of

Bristol-Bergerac hybrid populations homozygous for a single particular

Bristol or Bergerac chromosome, the remaining chromosomes being heterozygous

(Rose et al., 1982). The two strains differ at their nucleotide level by

approximately 1% (Emmons et al., 1979; Rose et al., 1982). The observed

frequency of RFLDs (Rose et al., 1982) is consistent with the occurrence of
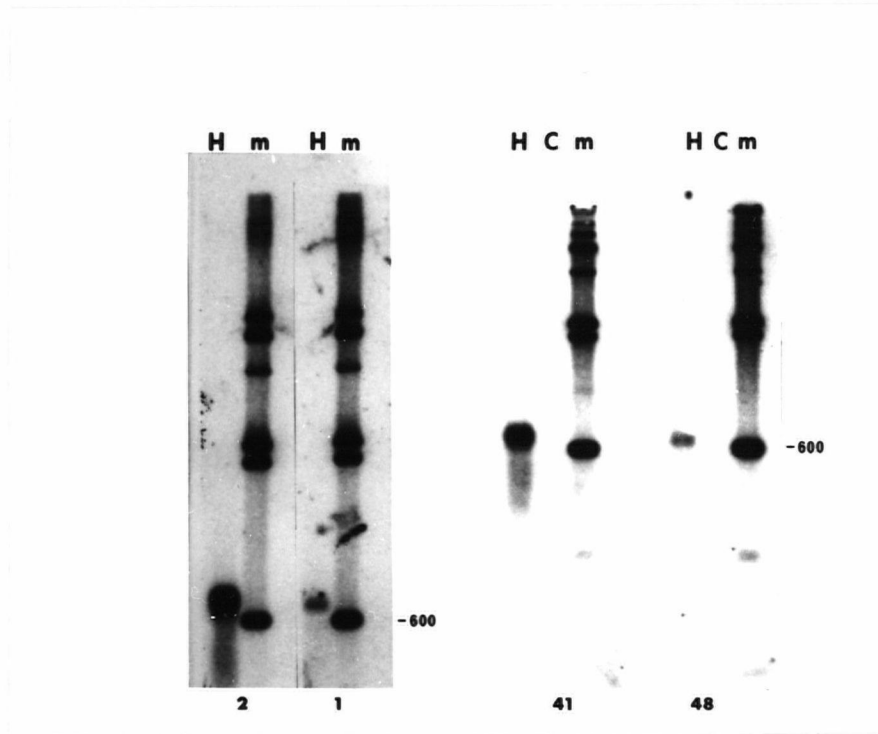
Figure 22.   Northern blot analysis of hsp16 mRNA levels using gene-specific

probes.

Approximately 1.0 µg of polyA$^+$RNA from either control (C) or heat

shocked (H) worms was separated on a 1.2% agarose-formaldehyde denaturing

gel, transferred to nitrocellulose and hybridized to the probes indicated.

The markers (m) are a HindIII digest of lambda DNA.   Fragments were

end-labelled by the filling-in reaction using Klenow polymerase and

transferred along with the RNA to nitrocellulose.

a base pair change per 120 bp or one insertion/deletion rearrangement per 52,000 bp.

In an attempt to determine the linkage of locus hsp16-48/1, several purified EcoRI fragments derived from λCharon4 A-1 or A-4 were used to probe filters containing both Bristol and Bergerac genomic DNA digested with EcoRI. No differences were seen in the migration of hybridizing bands between the two strains (Ann Rose, personal communication). However, only a total of approximately 16,000 bp was analyzed, which is well below the average distance required to detect an insertion/deletion event (Rose et al., 1982). Snutch and Baillie (1984) have proposed that the high mutation rate (10% sequence divergence) associated with a C. elegans hsp70 gene is due to its status as a highly inducible gene. This does not appear to be the case with the hsp16 genes which are also highly inducible.

A detailed comparison of the 1.9 kb inverted repeat was also undertaken. Bristol and Bergerac DNA was digested with 12 different restriction enzymes which cut within each arm of the perfect inverted repeat. As a result 136 bp of DNA was analyzed for mutations. As seen in Figure 23, no restriction fragment polymorphisms could be detected, emphasizing the sequence conservation between the two strains within this particular locus. The probe used in this set of experiments was the same as that used in Figure 13C. Foldback structures were not seen in these particular DNA preparations in which it could be shown that there was a much lower degree of nicking.

Since these two strains cannot be distinguished morphologically, the specificity of the Bristol and Bergerac DNA was verified by probing nitrocellulose filters containing genomic DNA with the plasmid pCeh2 which
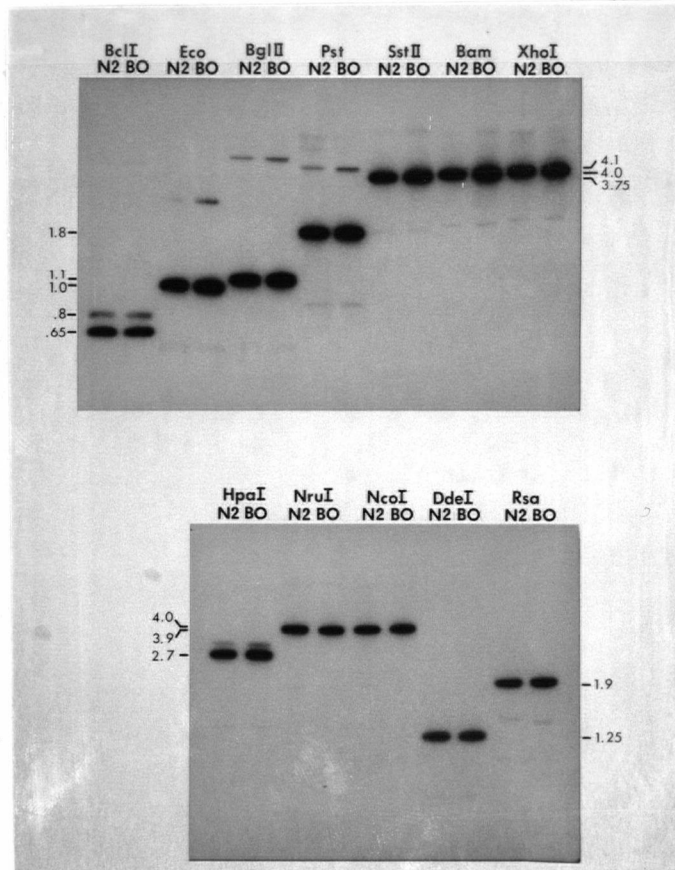
Figure 23. Detailed comparison of the 1.9 kb inverted repeat structure in the closely related strains Bergerac and Bristol.

Approximately 2 μg of Bristol (N2) and Bergerac (BO) genomic DNA was digested with the indicated restriction enzymes, separated by electrophoresis on 1.0% agarose gels and transferred to nitrocellulose. The probe used was an M13 clone containing a 500 bp DdeI-EcoRI insert of which 200 bp were specific to the loop region of the inverted repeat (Probe C, Figure 13C). The size of the hybridizing band in each case is shown in kilobase pairs.

contains a copy of Tcl, a transposable element of C. elegans. This element
was isolated from Bristol DNA and was kindly provided by Linda Harris,
University of British Columbia. When hybridized to genomic DNA, 20 - 30
bands could be observed with Bristol DNA but a more complicated pattern was
found with Bergerac DNA since the latter contains approximately 10 times as
many copies of the transposable element (Emmons et al., 1983; Liao et al.,
1983). These results are not shown.

IV. DISCUSSION

4.1 The Hsp16 Gene Family of C. elegans

The sequences presented in this thesis along with those of Jones et al.
(1986) define a family of four related hsp16 genes. They are arranged in
divergently transcribed pairs at two separate loci. The hsp16-48/41 genes
code for one class of hsp16, 143 amino acid residues long while the
hsp16-1/2 genes encode the other class, which is two amino acid residues
longer. Thus each locus and each gene pair codes for the two major types of
hsp16. The complete amino acid sequences of all 4 proteins are aligned in
Figure 24 to show the relatedness of hsp16-41 to hsp16-48 and of hsp16-2 to
hsp16-1. The basis for their classification into two distinct groups is
most evident in the comparison of their first exons. For example, only
three amino acid changes occur between the related proteins in this region.
Between the two classes of proteins, the regions encoded by exon 1 show no
homology.

In the region encoded by the second exon, 61 of the possible 98 amino
acid positions which can be compared are identical in at least three of the
four proteins as shown in Figure 24. The second exons of hsp16-41 and
hsp16-48 contain only 6 codon changes while the homology in the same region
of hsp16-1 and hsp16-2 is comparable, with only 7 codon differences. Less
similarity is seen between genes of the same locus. For example, a
comparison of the second exons of hsp16-48 and hsp16-1 shows that there are
37 codon differences within 98 amino acids. Thus there is a highly variable
protein domain represented by the first exon while the remainder of the
protein contains a conserved domain which shows homology to the shsps of

```
HSP16-41  MLMLRSPYSDSNALDHFLDELTGSVQFPYWRNADHNSFNFSD
HSP16-48  MLMLRSPFSDSNVLDHFLDEITGSVQFPYWRNADHNSFNFSD
HSP16-1     MSLYHYFRPAQRSVFGDLMRDMAQMERQFTPVCRGSPS
HSP16-2     MSLYHYFRPAQRSVFGDLMRDMALMERQFAPVCRISPS


HSP16-41  NIGE- 58bp INTRON -IVNDESKFSVQLDVSHFKPENLK
HSP16-48  NIGE- 55bp INTRON -IVNDESKFSVQLDVSHFKPEDLK
HSP16-1   ESSE- 52bp INTRON -IVNNDQKFAINLNVSQFKPEDLK
HSP16-2   ESSE- 46bp INTRON -IVNNDQKFAINLNVSQFKPEDLK


HSP16-41  IKLDGRELKIEGIQETKSEHGYLKRSFSKMILLPEDADLPSV
HSP16-48  IELDGRELKIEGIQEKKSEHGYSKRSFSKMILLPEDVDLTSV
HSP16-1   INLDGHTLSIQGEQELKTEHGYSKKSFSRVILLPEDVDVGAV
HSP16-2   INLDGRTLSIQGEQELKTDHGYSKKSFSRVILLPEDVDVGAV


HSP16-41  KSAISNEGKLQIEAPKKTNSS-RSIPINFVAKH
HSP16-48  KSAISNEGKLQIEAPKKTNSS-RSIPINFVAKH
HSP16-1   ASNLSEDGKLSIEAPKKEAIQGRSIPIQQAPVEQKTSE
HSP16-2   ASNLSEDGKLSIEAPKKEAVQGRSIPIQQAIVEEKSAE
```

Figure 24.   A comparison of the proteins encoded by the C. elegans 16 kd

heat shock genes.

In the boxed regions, at least three of the four proteins have

identical amino acids.   Also shown are the highly variable and highly

conserved domains which are precisely defined by a short intron.   The amino

acid sequences of hsp16-2 and hsp16-41 are taken from Jones et al. (1986).

Drosophila, Xenopus, soybean and man. The structure of this domain has been

maintained in the vertebrate eye lens protein α-crystallin.

Two-dimensional gel electrophoresis of the in vitro translation

products of RNA hybrid-selected by either CEHS48 or CEHS41 revealed at least

five electrophoretic variants of hsp16 which had significantly different

isoelectric points. The possiblility that another hsp16 gene exists cannot

be excluded although it is also possible that a given gene product may exist

in more than one electrophoretic form since acetylation of proteins has been

shown to occur in a rabbit reticulocyte system (Palmiter, 1977; Garrels and

Hunter, 1979; Rubenstein and Deuchler, 1979).

The shsps of Drosophila, Xenopus, plants and man are also encoded by

multigene families. At the protein level, shsps are always present in at

least two distinct isoforms on two-dimensional gels, the only exception

occurring in quail (Atkinson et al., 1981). The vertebrate α-crystallins

are characterized by two primary gene products which have been designated as

B_2 types (basic) and A_2 types (acidic) due to their isoelectric points

(Schoenmakers and Bloemendal, 1968). In hamster, these proteins are encoded

by single copy genes which are located on different chromosomes (Quaz-Jeuken

et al., 1985). In the vertebrate lens, the A_2 and B_2 crystallins form

large aggregates with their deamidation products which have been designated

A_1 and B_1. With age, all of the α-crystallin chains undergo

C-terminal degradation (for a review see Bloemendal 1982). Similarly, the

shsps of Drosophila can be grouped according to their net charges: hsp27

and hsp26 are basic whereas hsp23 and hsp22 are acidic (Mirault et al.,

1978; Storti et al., 1980).

It is possible that the two classes of hsp16 in C. elegans correspond to acidic and basic types. Since each heat shock locus codes for both types and since transcripts from one locus accumulate to levels much higher than those corresponding to the other, one could expect to see a poorly expressed acidic variant and a poorly expressed basic variant. Similarly there would be a more abundant acidic and a more abundant basic isoform. This is what is observed on the two-dimensional gel shown in Figure 7. For example, the relative abundance of the two most basic variants (shown by arrows) is consistent with the relative accumulation of transcripts from the two loci, one being 20 to 40 times higher than the other.

Based on the overall proportion of acidic amino acid residues (glutamate and aspartate) to basic amino acid residues (lysine and arginine) as predicted by the derived amino acid sequences, it is not possible to assign the two classes of hsp16 into distinct isoelectric forms. However, there are some interesting charge differences within localized regions or domains. The first exon of both hsp16-41 and hsp16-48 has a predicted net charge of -3 while the corresponding region of hsp16-2 and hsp16-1 has a net charge of +1. The second exons of hsp16-41 and hsp16-48 have a net charge of +3 and 0, respectively, while the second exons of both hsp16-2 and hsp16-1 have a net charge of -4. These differences in net charge between corresponding domains may be important depending on the distribution and spatial arrangement of charged groups as a result of protein folding. The availability of specific cloned sequences now makes it possible to assign a specific isoelectric variant to a particular gene using hybridization selection under stringent conditions and subsequent in vitro translation.

4.2 Evolution of the Hsp16 Gene Family and Its Relationship to Other shsps and to Vertebrate α-Crystallins

The presence of homologous multigene families coding for shsps of many organisms suggests that an ancestral heat shock gene has undergone a number of duplication events. The same is true for the homologous α-crystallin genes in vertebrates. It has been postulated that the progenitor heat shock gene and the progenitor α-crystallin gene arose from the duplication of a single ancestral gene (van den Heuvel et al., 1985). It is proposed that this progenitor α-crystallin gene then evolved through an internal duplication within the sequences represented by the first exon, followed by a duplication of the entire gene which gave rise to the αA and αB type genes. An alternative proposal can be made in which a single ancestral gene duplicated to form a pair of genes representing A and B types. This pair of genes gave rise to the α-crystallin genes and the hsp16 gene family. This scheme is illustrated in Figure 24. The argument in favor of this scheme comes from the comparison of the C. elegans hsp16 amino acid sequences with the α-crystallins of the dogfish, Squalus acanthias. The α-crystallins from this species are considered to be primitive relative to those of other vertebrates based on immunological comparisons (Manski and Malinowsky, 1978; De Jong, 1982). This primitive α-crystallin also exists in acidic (A) and basic (B) forms. Computer analysis (Jones et al., 1986) using the program described by Delaney (1982) with a MINBLOCK parameter set of 2 revealed that hsp16-2 resembled the A form of dogfish crystallin more than the B form. Conversely, hsp16-41 showed greater similarity to the B type crystallin of Squalus.
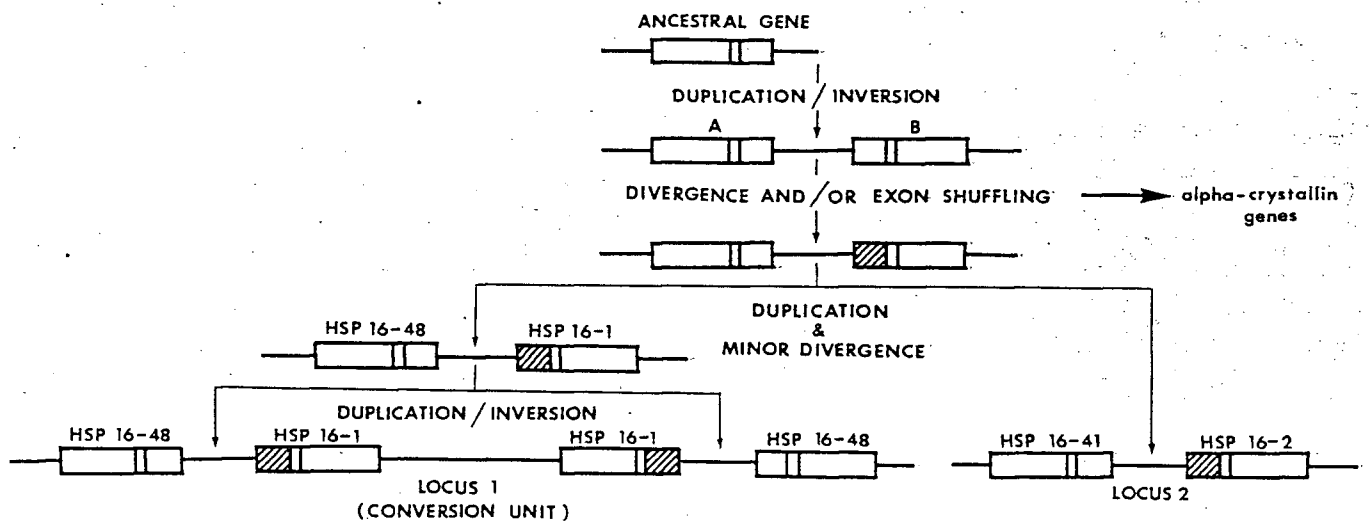
Figure 25. A model for the evolutionary origin of the hsp16 gene loci of

C. elegans and the α-crystallin genes of vertebrates.

Open boxes show the separate exons, and hatching indicates a major

divergence in the first exon sequence.

The evolution of an ancestral gene into A and B types must have been followed by a major divergence in the sequence of the first exon since in C. elegans, genes within the same pair share little or no homology in this region. This divergence may also have been brought about by exon shuffling (van den Heuvel et al., 1985) since the diverged region is demarcated by an intron in the C. elegans genes. Furthermore, this event probably took place after the divergence of plants and animals since the shsps of soybean are highly homologous to each other in all regions of the proteins including the amino-terminal portion (Nagao, 1985); however, the amino acid sequences of only 4 out of at least 10 shsps have been determined. The shsps of Drosophila, on the other hand, show little homology in their amino-terminal regions (see Figure 2).

In C. elegans, this period of divergence must have been followed by a second duplication event which is represented by the homologous gene pairs found at different chromosomal locations. Following the proposed second duplication, only minor changes have occurred in the coding regions of the genes. Also, over the latter period the sequence of the first exon evidently changed at the same rate as that of the second. Perhaps the acquisition of function at this stage fixed the rate of mutation within the first exon. A third duplication and inversion event has occurred within the hsp16-48/1 locus, resulting in a 1.9 kb inverted repeat. The perfect identity of the two "arms" of this repeat at the nucleotide sequence level implies that the duplication event was very recent or that sequence identity has been maintained by frequent intralocus gene conversion events. The argument for a gene conversion mechanism will be presented in section 4.3.

Thus, the evolution of the hsp16 genes of C. elegans has been characterized by a series of duplications, at least two of which were accompanied by inversions. Leigh-Brown and Ish-Horowicz (1981) have suggested that the ancestral organization of the Drosophila hsp70 genes was as inverted repeats for both the 87A and 87C loci. The two hsp70 genes at 87A are approximately 1.7 kb apart in opposite orientations (Goldschmidt-Clermont, 1980). The copies at 87C are present in two domains, a single proximal sequence separated from tandem distal sequences by about 38 kb (Ish-Horowicz and Pinchin, 1980). In the related species, D. simulans and D. mauritiana, this 38 kb insert is missing, resulting in two hsp70 genes which are in a head to head configuration like those at locus 87A. Similarly, at locus 67B, if the orientation of gene 2 (which is not known) were opposite to that of gene 3, the shsp genes of Drosophila would exist in three gene pairs, the origins of which could be explained by inverted duplications (see Figure 1). Inverted duplication has been shown to occur during the amplification of a cellular oncogene (c-myc) in a variety of tumor cell lines and during the amplification of the CAD gene in N-(phosphonacetyl)-L-aspartate-resistant cell lines (Ford and Fried, 1986). CAD is an acronym for the multifunctional enzyme which catalyzes the first three steps in uridine biosynthesis. These authors suggest that the wide occurrence of inverted duplications is related to the mechanism of gene amplification.

The shsps of C. elegans resemble the α-crystallins of vertebrates more than they do the shsps of D. melanogaster. This is clearly demonstrated by the amino acid comparisons shown in Figures 11 and 16. For example, in Figure 16, there are 32 positions conserved between the shsps of

C. elegans and the α-crystallins while there are only 20 positions

conserved between the shsps of C. elegans and Drosophila. Furthermore, the

30 amino acid residues at the carboxy-terminal of the α-crystallins show

no obvious homology to corresponding regions in the Drosophila shsps

sequences although there is a slight similarity with the C-terminal 20 amino

acid residues of nematode hsp16 (Wistow, 1985). Also, each of the hsp16

genes contains a small intron in a position analogous to that of the first

intron of the α-crystallin genes. This suggests that some common property

of shsps and α-crystallins has been more rigorously conserved in

C. elegans than in Drosophila.

Human hsp25 is even more homologous to α-crystallin than any of the

shsps including those of C. elegans (Hickey et al., 1986). Also, regions of

amino acid identity between hsp25 and α-crystallin are found along almost

the entire length of α-crystallin. The lack of a divergent amino terminal

region, as demarcated by an intron in the hsp16 genes of C. elegans,

suggests that the human hsp genes evolved from an ancestral pair of genes

which did not experience major divergence and/or exon shuffling. This is

supported by the fact that neither of the two introns of the human hsp25

gene is in a position analogous to the single introns of the C. elegans

hsp16 genes or to the first introns of the α-crystallin genes.

The shsps of Drosophila (except hsp22) contain a hydrophobic N-terminal

region of 14 amino acid residues. If this domain has a property which is

required for the function of these proteins, then it must have evolved after

the divergence of nematodes and insects since the N-termini of the shsps of

C. elegans do not show this degree of hydrophobicity. Interestingly, the

shsps of soybean contain an N-terminal hydrophobic region of approximately

10 amino acid residues (Nagao et al., 1985). Human hsp25 shows no degree of hydrophobicity within its amino terminal portion (Hickey et al., 1986).

4.3 Gene Conversion Within Locus Hsp16-48/1

The hsp16-1 and hsp16-48 genes, including their coding and flanking sequences, are located within a 1.9 kb region which, because of its compactness, functional unity, and repetition, might be referred to as a module. Two such modules, duplicated perfectly with only a single base pair change, exist in the genome in an inverted orientation relative to each other and are separated by 416 bp of unrelated sequence. This novel gene organization was confirmed by three lines of evidence. Firstly, extensive genomic DNA analysis has been carried out using 12 different restriction enzymes which cut in the region of interest. The fragments obtained were entirely consistent with the predicted structure. Secondly, overlapping phage clones were isolated from two independently constructed libraries which contained both the 3.3L and 3.3R EcoRI fragments. Thirdly, both the right and left modules were sequenced completely, and a single base pair difference was identified. The palindromic region has the potential to adopt a structure in which intra-strand pairing results in a 1.9 kb double-stranded stem portion with a looped-out single-stranded region of approximately 400 nucleotides. This foldback form remains stable under otherwise denaturing conditions. DNA fragments containing this alternate configuration migrate faster in agarose gels as seen in the genomic DNA analysis. More direct proof of the existence of foldback DNA at this locus comes from the observation that the predicted 1.9 kb stem structure is

resistant to S1 nuclease digestion after samples of genomic DNA have been denatured (Don Jones, personal communication).

Evidence at the molecular level for gene conversion has arisen from the sequencing of duplicated nonallelic genes such as the tRNA (Amstutz et al., 1985) and cytochrome C (Ernst et al., 1981) genes of yeast, the α-globin genes of goats (Schon et al., 1982) and humans (Liebhaber et al., 1980, 1981), the human γ-globulin genes (Slightom et al., 1980; Shen et al., 1981), the human immunoglobin λ2 chain genes (Ollo and Rougeon, 1983), and the high-cysteine chorion genes of the silk moth, Bombyx mori (Iatrou et al., 1984). In each of these cases, there are regions of greater homology between genes than would be expected from normal rates of sequence divergence since the time of the original duplication event. In other words, these regions have not evolved independently.

Gene conversion may be involved in maintaining microdiversity among the immunoglobulin variable-region gene families (Cohen et al., 1982; Dildrop et al., 1982) and among the class I (Lalanne et al., 1982; Mellor et al., 1983; Schulze et al., 1983) and class II (Widera and Flavell, 1984; Denaro et al. 1984) major histocompatability complex antigens. As an example, a converted region of only 39 bp has been discovered in the gene family encoding the variable regions of human immunoglobulin kappa light chains (Bentley and Rabbits, 1983). This stretch accounts for 7 of 10 base substitutions over a length of 940 bp and is evidence that gene conversion may lead to the shuffling of gene segments.

The bovine prepro-AVP-NPII (vasopressin) and prepro-OT-NPI (oxytocin) genes are related and probably arose from a duplication event. In this case there is a stretch of 332 bp which has perfect sequence identity between the

two genes. The similarity drops to 81% in the preceding 60 bp, and the other side can be only poorly aligned (Ruppert et al., 1984). These authors have suggested the occurrence of two successive conversion events leading to the present sequence organization, the second one being very recent.

The perfect homology between the two 1.9 kb modules containing the hsp16-1 and hsp16-48 genes could be explained if the duplication event were fortuitously very recent. Alternatively, the homology could have been continuously maintained after the second duplication by repeated gene conversion events. The latter explanation is favored for probabilistic reasons, because the sequence identity would be in a steady state and therefore detectable over a much longer period of evolution. This conversion process might be facilitated by the proximity of the two modules and by the unusual sequences which have been found at either end of the inverted repeat structure. This mechanism must also be precise since the borders of the "conversion unit" are very sharply defined in the present example, contrasting 100% sequence homology within the converted 1.9 kb region with complete divergence on either side of it. It is interesting that a procaryotic transposable element such as transposon Tn5, which has two IS50 modules of 1,530 bp in an inverted repeat, can maintain the sequence identity of these modules. Perhaps in the hsp gene cluster the elements occupy stable positions in the genome (unlike transposable elements) but undergo strand exchange and mismatch repair frequently. It has been demonstrated that intrachromosomal, nonreciprocal transfer or gene conversion can occur at a high frequency in cultured mouse cells (Liskay and Stachelek, 1983) and in yeast (Klar and Strathern, 1984; Klein, 1984).

Given the postulate that the hsp16-1/48 gene cluster is a hot spot for gene conversion, there should be sequences which are involved in facilitating and delimiting this phenomenon. The G + C-rich sequences discussed in Section 3.12 are intriguingly situated at the distal borders of the inverted repeat. This sequence is 24 bp in length and is characterized by a stretch of 12 alternating purine-pyrimidine residues, flanked on either side by a G + C hexanucleotide in an inverted repeat orientation. In fact, the entire 24 bp stretch has the potential to adopt a palindromic structure, the left variant containing 2 mismatches in a 12 bp stem, the right variant containing 3 mismatches. All of the base pairs in these structures would be G - C pairs, theoretically maximizing their stability.

Similar G + C-rich sequences are found at the 3' border of a conversion event between the human immunoglobulin α1 and α2 constant region genes (Flanagan et al., 1984) (Table VI). Strikingly, G + C hexanucleotides identical to the ones in the heat shock locus flank a 40 bp stretch of alternating purine-pyrimidine residues and are oriented in an inverted repeat. These sequences are shown in upper case letters. The similarity even extends to a cytosine to thymine change which breaks up the inverted repeat and which is characteristic of the boundary sequence variant in the right module of the hsp16-1/48 locus.

A 2.1 kb region of DNA containing a pair of chicken histone genes (H4 and H2A) has been duplicated and inverted, the two arms being 97% homologous and separated by approximately 2.0 kb of unrelated sequence which contains a single histone H3 gene (Wang et al., 1985). Once again the borders are defined by 10 bp G + C-rich sequences, the core of which is identical to the hexanucleotides found in locus hsp16-48/1 (Table VI). In this case,

however, the sequences are found at both the proximal and distal borders where they occur as inverted repeats and direct repeats, respectively. They are separated by DNA stretches which do not contain alternating purine/pyrimidine residues or, in the case of the proximal borders, do not contribute to a large palindromic structure. A C/G to A/T transition has also occurred in an analogous position in one of the direct repeats at the right distal border.

A computer search of DNA sequences in proximity to 13 recombinatorial breakpoints presumed to be a result of gene conversion has located in each case a palindromic sequence (Krawinkel et al., 1986). These sequences have stem structures ranging in length from 9 to 16 bps with loop sizes ranging from 0 to 28 nucleotides. A correlation of conversion boundaries with direct repeats was not observed. These results, in conjunction with those discussed above suggest that palindromic sequences may promote gene conversion and that they may serve as recognition sites for one or more enzymes involved in genetic recombination. The characterization of other similar G + C-rich sequences, which appear to be abundant within the genome of C. elegans, should help in understanding their function. Both the 1.9 kb and the 370 bp inverted repeat structures are associated with these sequences. The availability of specific oligodeoxynucleotide probes should facilitate the identification of other cloned foldback structures to determine whether they are associated with functional genes which have also undergone gene conversion processes.

Table VI. Association of G + C-Rich Sequences with Palindromic
Structures Located at Gene Conversion Boundaries

Palindromic Structure		Location	
Watson-Crick pair/ Stem Length	Loop Size	Gene	Boundary
CGGGGCcgcgcg GCCCCGacgcac	10/12	0	hsp16-1/48 locus	left distal
CAGGGCcgcgcg GCCCCGacgcac	9/12	0	hsp16-1/48 locus	right distal
ggCGGGGCgg ccGCCCCGcc	10/10	21	chicken histone gene locus	left and right proximal
CAGGGC GCCCCG	5/6	40	mouse Ig-constant α1 and α2 genes	3"-boundary

4.4   The Heat Shock Response of C. elegans

C. elegans undergoes a typical response when exposed to hyperthermic conditions.   In vivo labelling of nematodes at 35°C shows that most of the pre-existing protein synthesis declines while approximately 10 proteins are abundantly synthesized.   These experiments cannot distinguish between those proteins whose synthesis is induced at the higher temperature and those whose previous synthesis under normal temperatures is unaffected by heat shock.   For example, hsp70 consists of at least two primary gene products, one which is synthesized only under heat shock conditions while the other is a product of an hsp70 cognate gene which is active at normal temperatures (Snutch and Baillie, 1984).   Hsp16, however, appears to be newly synthesized at the elevated temperature and this has been verified by the hybridization of specific cloned DNA sequences to polyA^{+}RNA isolated from heat shocked nematodes but not to a similar RNA fraction purified from normally growing nematodes.   Thus the hsp16 genes, of which there are at least four, are under a tight transcriptional control mechanism.   Although the transient nature of the heat shock response in C. elegans has not been investigated, it appears that the resumption of normal protein synthesis does not occur within 5 - 6 hours if the nematodes are maintained at 35°C (Figure 3). This is similar to the situation seen in HeLa cells in which normal protein synthesis does not resume for at least 10 hours if the cells are maintained at 42°C (Hickey and Weber, 1982).   In contrast, resumption of existing protein synthesis can occur within 10 - 20 minutes at 36°C in yeast (McAlister and Finkelstein, 1980) or within 2 - 3 hours at 37°C in Drosophila (DiDomenico et al., 1982b).   Assuming that the heat shock proteins serve a protective role during cellular stress, then it would be beneficial for them to possess

relatively extended half lives. The importance of this property is further emphasized by the presumably transient nature of the heat shock response: it has been found that in <u>Drosophila</u>, hsp70 represses its own synthesis upon continued stress. As shown in Figure 4, the hsps of <u>C</u>. <u>elegans</u>, which were pulse-labelled at 35°C, are detectable in cells 24 hours following heat shock, at levels comparable to those observed during or immediately after the shock.

In <u>vitro</u> translation of polyA^{+}RNA from heat shocked nematode cultures demonstrated that a translational control also operates in nematodes at high temperatures. Although pre-existing messages are still present in heat shocked cells as detected by translation <u>in</u> <u>vitro</u>, they are not translated <u>in</u> <u>vivo</u> (compare Figure 3 and Figure 5, lane C). In <u>Drosophila</u>, the leader sequences of <u>hsp70</u> (McGarry and Lindquist, 1985; Klemenz <u>et al</u>., 1985) and <u>hsp22</u> (Hultmark <u>et al</u>., 1986) mRNA are responsible for their selective translation, although it is not yet known how this is accomplished. The mechanism involved does not appear to have specific sequence (McGarry and Lindquist, 1985) or length (McGarry and Lindquist, 1985; Klemenz <u>et al</u>., 1985; Hultmark <u>et al</u>., 1986) requirements. The leader sequences of all 4 sequenced <u>hsp16</u> genes are shown in Figure 26. When aligned to compensate for deletions/insertions, the leaders of the <u>C</u>. <u>elegans</u> <u>hsp16</u> genes display a great deal of similarity with an average length of approximately 45 nucleotides. The involvement of these leaders in the ability of hsp16 transcripts to be recognized by heat shocked ribosomes awaits experiments in which mutated leader regions are assayed for selective translation.

Due to the divergent directions of transcription, the 5' noncoding regions of the <u>hsp16</u> gene pairs are precisely defined. The intergenic

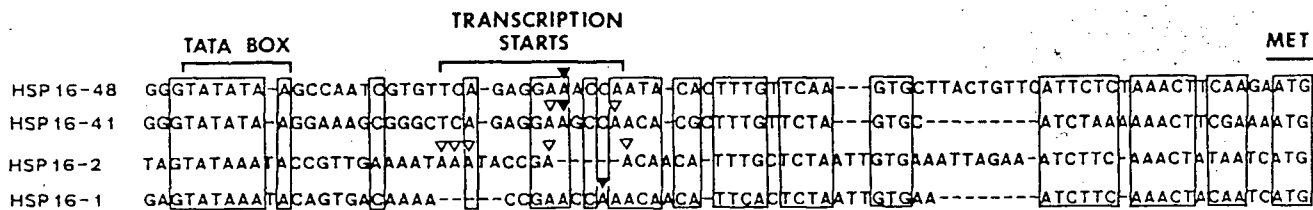Figure 26.   The leader sequences of the hsp16 genes of C. elegans

The TATA motifs and the transcription starts deduced from S1 nuclease

protection analysis are indicated.   Positions where at least three out of

four nucleotides are identical are boxed.   Gaps were introduced to give the

best alignment.   The hsp16-2 and hsp16-41 sequences and start sites are

taken from Jones et al. (1986).

regions of the hsp16-41/2 and hsp16-48/1 gene pairs are aligned in Figure

27. This region shows an overall homology of 85% and contains only 26

nucleotide differences. The known functionally important 5' sequences of

the hsp16 genes have been conserved. These include the TATA sequences and

the HIP sequences; two of the latter overlap each other upstream of the

hsp16-48 and hsp16-41 genes. The conserved regions include sequence

elements the significance of which is unknown, such as the stretch of

alternating purine-pyrimidine residues situated adjacent to a prominent

region of dyad symmetry. Transcriptional studies of the hsp16-48 and

hsp16-1 gene pair in mouse fibroblast cells have demonstrated that heat

inducibility as well as arsenite inducibility is dependent on the HIP

sequences (Kay et al., 1986). A single HIP sequence can function

bidirectionally, inducing the transcription of both genes when placed

between the TATA boxes. In this case, however, the efficiency is reduced 10

fold relative to the wild type gene pair. Placing four overlapping promoter

elements between the genes resulted in inducible bidirectional transcription

at levels higher than those found with the wild type gene pair.

The high degree of similarity between the intergenic regions of the two

heat shock loci implies the existence of functionally important sequences in

addition to the heat shock promoters. Although none have been identified as

yet, it is possible that other metabolic stressors or physiological states

activate the hsp16 genes through alternative cis-acting elements. Based on

observations made in Drosophila, there exists the possibility that hsp16 may

serve some function during nematode development, i.e. the hsp16 genes or a

subset thereof may be activated in a stage-specific manner. The

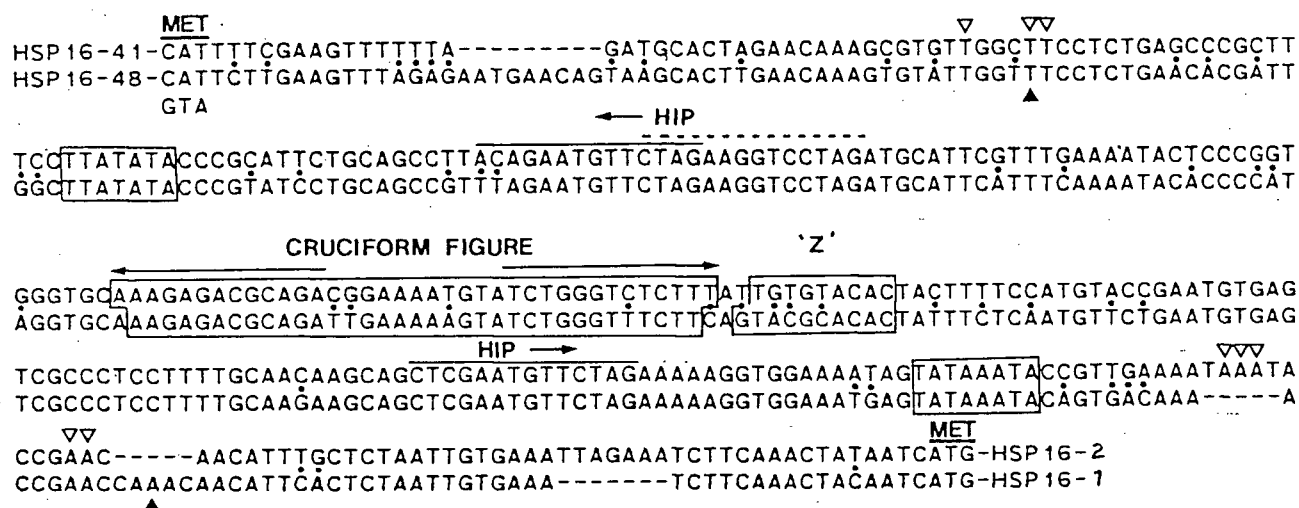availability of specific gene probes now allows a detailed analysis of RNA

```
        MET ·                                                              ▽      ▽▽
HSP16-41-CATTTTCGAAGTTTTTTTA---------GATGCACTAGAACAAAGCGTGTTGGCTTCCTCTGAGCCCGCTT
HSP16-48-CATTCTTGAAGTTTAGAGAATGAACAGTAAGCACTTGAACAAAGTGTATTGGTTTCCTCTGAACACGATT
        GTA                                                                ▲
                                                  ◄─── HIP
TCCTTATATACCCGCATTCTGCAGCCTTACAGAATGTTCTAGAAGGTCCTAGATGCATTCGTTTGAAAATACTCCCGGT
GGCTTATATACCCGTATCCTGCAGCCGTTTAGAATGTTCTAGAAGGTCCTAGATGCATTCATTTCAAAATACACCCCAT

              CRUCIFORM FIGURE                            'Z'
GGGTGCAAAGAGACGCAGACGGAAAATGTATCTGGGTCTCTTTATTGTGTACACTACTTTTCCATGTACCGAATGTGAG
AGGTGCAAAGAGACGCAGATTGAAAAAGTATCTGGGTTTCTTCAGTACGCACACTATTTCTCAATGTTCTGAATGTGAG
                            HIP ───►                                         ▽▽▽
TCGCCCTCCTTTTGCAACAAGCAGCTCGAATGTTCTAGAAAAAAGGTGGAAAATAGTATAAATACCGTTGAAAATAAATA
TCGCCCTCCTTTTGCAAGAAGCAGCTCGAATGTTCTAGAAAAAAGGTGGAAATGAGTATAAATACAGTGACAAA-----A
    ▽▽                                             MET
CCGAAC-----AACATTTGCTCTAATTGTGAAATTAGAAATCTTCAAACTATAATCATG-HSP16-2
CCGAACCAAACAACATTCACTCTAATTGTGAAA-------TCTTCAAACTACAATCATG-HSP16-1
    ▲
```

Figure 27.   A comparison of the intergenic regions of the two hsp16 gene

loci of C. elegans.

Dots indicate positions of sequence differences and gaps were

introduced to obtain the best alignment. The heat inducible promoters (HIPs)

and TATA boxes are indicated, as is the position of the large cruciform

figure with the potential to adopt a stem-loop structure and its adjoining

purine-pyrimidine sequence (Z).   Transcription start sites are indicated by

triangles, with the major sites shown as filled triangles.   The hsp16-2/41

intergenic region is taken from Jones et al. (1986).

purified from different developmental stages including eggs, adults and the four intermediate larval forms. Although the control RNA used in this thesis was prepared from mixed cultures, it is likely that the transcription of a particular hsp16 gene during one specific stage out of the six major identifiable ones would have gone undetected due to the limited sensitivity of Northern analysis. In vivo protein labelling experiments have shown that hsp16 is heat inducible in all of the above-mentioned developmental stages (Snutch and Baillie, 1983).

Remarkably, the two hsp16 gene pairs of C. elegans have radically different expression levels during heat shock. This may reflect a difference in their rates of transcription or in their relative message stabilities. The intergenic regions of the two gene pairs are very similar but not identical. It is possible that some minor sequence variation could affect the relative levels of heat inducible transcription. For example, a cytosine to thymine transition within the centrally located palindromic structure introduces a second base pair mismatch into the predicted 12 bp stem structure between the hsp16-1/48 gene pair, which would be expected to significantly affect its stability. Alternatively, three major features outside the intergenic regions distinguish the two different gene pairs. These include nonhomologous 3' noncoding sequences, the presence of repetitive elements flanking the hsp16-2/41 gene pair and the inverted duplication of the hsp16-1/48 gene pair.

The 3' noncoding regions of the two related genes hsp16-48 and hsp16-41 are radically different in length (38 versus 94 nucleotides) and show no sequence similarity. The other two related genes, hsp16-1 and hsp16-2, show homology extending 45 bp past the polyadenylation signal. Perhaps more

significant is the presence of an A + T-rich sequence in the 3' noncoding

region of the hsp16-2 gene which could form a perfect 7 bp hairpin. It

occurs in a region which has been deleted in the corresponding less actively

expressed hsp16-1 gene. The other highly expressed gene, hsp16-41, has a

perfect 11 bp A + T-rich hairpin located adjacent to the polyadenylation

signal. It is conceivable that these secondary structures may be involved

in transcript processing, polyadenylation or stability.

Three copies of the 200 bp repetitive element family CeRep-16 flank the

more active hsp16 locus (Jones et al., 1986). An EcoRI-EcoRV fragment

containing the proximal half of the single repetitive element at the 3' end

of the hsp16-41 gene was hybridized to phage clones λCharon4 B-3 and

λCharon4 B-7, which represent the entire 30 kb region shown in Figure 12.

Under conditions in which the related repetitive elements from the 3' end of

the hsp16-2 gene could be detected, there was no hybridization to any of the

EcoRI fragments derived from the above-mentioned phage. These results (not

shown) indicate that homologous sequences are not present within at least 10

kb on either side of the less active locus containing the hsp16-1/48 gene

pairs. Members of the CeRep-16 element contain multiple repeats of the

sequence of G$^{A}_{G}$TTTGC, which is very similar to part of the "enhancer

core" sequence GGTTTG found in a variety of viral and cellular enhancers

(Serfling et al., 1985). Thus it is possible that an enhancer-like

mechanism may be regulating the expression of the heat shock genes at this

locus. Unfortunately, the differential expression of these gene pairs

cannot be reproduced when transfected into mouse cells (Rob Kay, personal

communication) suggesting that the factors that mediate 20 fold higher

transcript levels of the hsp16-2/41 gene pair relative to the hsp16-1/48

gene pair in C. elegans are not conserved in mouse cells. The use of a homologous transformation system, which has recently been developed for C. elegans (Stinchcomb et al., 1985), may prove to be helpful in demonstrating the existence and nature of such factors.

Another reason for the observed differential expression may simply lie in the possibility that the two heat shock loci are situated in different chromosomal environments or domains. The loci have not yet been genetically mapped but even if they were linked, they would be separated by at least 10 kb of DNA.

In addition to transcriptional and translational control, the presence of introns in heat inducible genes introduces the possibility of regulating gene expression at the level of splicing. It initially appeared that the presence of intervening sequences was confined to genes which were transcribed at normal growth temperatures. These include the constitutively expressed cognate genes of the Drosophila hsp70 gene family as well as the Drosophila hsp83 and human hsp25 genes which are expressed under normal conditions but are induced to much higher levels during heat shock. The hsp16 genes of Caenorhabditis comprise a third group of intron-containing genes, being inactive at normal temperatures.

Hsp83 has a pattern of expression which is different from any of the other hsps in Drosophila (Yost and Lindquist, 1986). As already mentioned, it is produced at substantial levels at $25^{\circ}C$. Maximum expression is observed between $33^{\circ}C$ and $35^{\circ}C$, decreasing to barely detectable levels by $38^{\circ}C$. In contrast hsp70 is highly induced at $38^{\circ}C$. Although the amount of hsp83 mRNA remains constant at temperatures above $35^{\circ}C$, the decreased expression of hsp83 is due to the reversible thermal sensitivity

of the splicing apparatus, virtually all of the hsp83 mRNA being found as

unspliced precursor at 38°C (Yost and Lindquist, 1986). Similar

observations were made by Stellar and Pirrotta (1985) in a transformed

Drosophila strain carrying the white gene which had been linked to an hsp70

promoter. At 37°C, the white gene transcripts, which contained a large

intron, were inefficiently spliced. Proper splicing resumes upon recovery

at 25°C (Steller and Pirrotta, 1985; Yost and Lindquist, 1986). If a mild

35°C heat shock, which induces the synthesis of hsps, is administered

prior to a severe 38°C heat shock, hsp83 transcript processing occurs

under otherwise restrictive conditions (Yost and Lindquist, 1986). It is

suggested by these authors that one of the functions of the heat shock

response is to protect the splicing apparatus.

It remains to be seen whether RNA processing is similarly more

thermo-labile in organisms other than Drosophila. It is possible that a

similar phenomenon occurs in C. elegans. Snutch and Baillie (1983) have

observed that hsp16 is not expressed at higher temperatures. Also, the cDNA

pCEHS41 corresponds to an unspliced transcript. Although S1 mapping

indicated that hsp16-41 transcripts were completely spliced (Jones et al.,

1986), the mRNA used in these studies was isolated independently from the

RNA used in the construction of the cDNA library and which was kindly

provided by Terry Snutch, Simon Fraser University. Thus the nematodes from

which RNA was isolated were subjected to different heat shock conditions.

C. elegans hsp70 is expressed under severe heat shock conditions in which

hsp16 is not (Snutch and Baillie, 1983). Although one of the members of the

hsp70 gene family contains an intron (Mark Heschl, personal communication),

the presence of intervening sequences in the other related genes has not

been determined. Thus the synthesis of hsp70 at higher temperatures may take place on transcripts which have bypassed the splicing mechanism. Human hsp25 mRNA is completely spliced at $42^{o}C$ in HeLa cells (Hickey et al., 1986) but this temperature may be below the threshold of splicing inactivation for a mammalian cell.

The increased temperature sensitivity of splicing relative to transcription and translation may reflect a physiologically relevant control mechanism in which the expression of transcripts containing intervening sequences is inhibited during severe stress. Also, the accumulation of unspliced transcripts during heat shock results in the delayed expression of these RNAs into protein products since splicing and subsequent translation can only occur during recovery at lower temperatures or upon adaptation of the splicing mechanism during maintained stress.

V.   REFERENCES

1. Altschuler, M. and J.P. Mascarenhas. 1982. Plant Mol. Biol. 1: 103-115.

2. Amin, J., R. Mestril, R. Lawson, H. Klapper, and R. Voellmy. 1985. Mol. Cell. Biol. 5: 197-203.

3. Amstutz, J., P. Munz, W.-D. Heyer, U. Leupold, and J. Kohli. 1985. Cell 40: 879-886.

4. Ananthan, J., A.L. Goldberg, and R. Voellmy. 1986. Science 232: 522-524.

5. Arnott, S., R. Chandrasekaran, D.L. Birdsall, A.G.W. Leslie and R.L. Rattliff. 1980. Nature 283: 743-745.

6. Arrigo, A.P., S. Fakan, and A. Tissières. 1980. Dev. Biol. 78: 86-103.

7. Arrigo, A.-P., and C. Ahmad-Zadeh. 1981. Mol. Gen. Genet. 184: 73-79.

8. Arrigo, A.-P., J.-L. Darlix, E.W. Khandjian, M. Simon, and P.-F. Spahr. 1985. EMBO J. 4: 399-406.

9. Artavanis-Tsakonas, S., P. Schedl, M.-E. Mirault, L. Moran, and J. Lis. 1979. Cell 17: 9-18.

10. Ashburner, M. 1970. Chromosoma 31: 356-376.

11. Ashburner, M., and J.J. Bonner. 1979. Cell 17: 241-254.

12. Atkinson, B.G. 1981. J. Cell Biol. 89: 666-673.

13. Atkinson, B.G., and M. Pollock. 1982. Can. J. Biochem. Cell Biol. 60: 316-327.

14. Atkinson, B.G., T. Cunningham, R.L. Dean, and M. Somerville. 1983. Can. J. Biochem. Cell Biol. 61: 404-413.

15. Atkinson, B.G., T. Cunningham, R.L. Dean, and M. Somerville. 1983. Can. J. Biochem. Cell Biol. 61: 404-413.

16. Atkinson, B.G., and R.L. Dean. 1985. in Changes in Eukaryotic Gene Expression in Response to Environmental Stress (Atkinson, B.G., and D.B. Walden, Eds.) pp. 159-181, Academic Press, New York.

17. Atkinson, T., and M. Smith. 1984. in Oligonucleotide Synthesis, a Practical Approach (Gait, M.J., eds.) pp. 35-81, IRL Press Ltd., Oxford.

18. Aviv, H., and P. Leder. 1972. Proc. Nat. Acad. Sci. U.S.A. 69: 1408-1412.

19. Ayme, A., and A. Tissières. 1985. EMBO J. 4: 2949-2954.

20. Baker, T.A., A.D. Grossman, and C.A. Gross. 1984. Proc. Nat. Acad. Sci. U.S.A. 81: 6779-6783.

21. Ballinger, D.G., and M.L. Pardue. 1983. Cell 33: 103-114.

22. Bardwell, J.C.A., and E. Craig. 1984. Proc. Nat. Acad. Sci. U.S.A. 81: 848-852.

23. Barnett, T., M. Altschuler, C.M. McDaniel, and J.P. Mascarenhas. 1980. Dev. Gen. 1: 331-340.

24. Baszczynski, C.L., D.B. Walden, and B.G. Atkinson. 1982. Can. J. Biochem. Cell Biol. 60: 569-579.

25. Baszczynski, C.L., D.B. Walden, and B.G. Atkinson. 1985. in Changes in Eukaryotic Gene Expression in Response to Environmental Stress (Atkinson, B.G., and D.B. Walden, eds.) pp. 349-371, Academic Press, New York.

26. Beckman, D., and S. Cooper. 1973. J. Bacteriol. 116: 1336-1342.

27. Bensaude, O., and M. Morange. 1983. EMBO J. 2: 173-177.

28. Bentley, D.L., and T.H. Rabbitts. 1983. Cell 32: 181-189.

29. Benton, W.D., and R.W. Davis. 1977. Science 196: 180-182.

30. Berger, E.M., and M.P. Woodward. 1983. Exp. Cell. Res. 147: 437-442.

31. Berk. A.J., and P.A. Sharp. 1977. Cell 12: 721-732.

32. Bienz, M. 1982. in Heat Shock: From Bacteria to Man (Schlesinger, M.J., M. Ashburner, and A. Tissières, eds.) pp. 177-181, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

33. Bienz, M., and H.R.B. Pelham. 1982. EMBO J. 1: 1583-1588.

34. Bienz, M., and J.B. Gurdon. 1982. Cell 29: 811-819.

35. Bienz, M. 1984a. Proc. Nat. Acad. Sci. U.S.A. 81: 3138-3142.

36. Bienz, M. 1984b. EMBO J. 3: 2477-2483.

37. Birnboim, H.C., and J. Doly. 1979. Nucleic Acids Res. 7: 1513-1523.

38. Bloemendal, H. 1982. Crit. Rev. Biochem. 12: 1-39.

39. Bloemendal, H., T. Berns, A. Zweers, H. Hoenders, and E.L. Benedetti. 1972. Eur. J. Biochem. 24: 401-409.

40. Bond, U., and M.H. Schlesinger. 1985. Mol. Cell. Biol. 5: 949-956.

41. Bouche, G., F. Amalric, M. Caizergues-Ferrer, and J.P. Zalta. 1979. Nucleic Acids Res. 7: 1739-1747.

42. Bourouis, M., and G. Richards. 1985. Cell 40: 349-357.

43. Brenner, S. 1974. Genetics 77: 71-94.

44. Bretscher, M.S., and A.E. Smith. 1972. Anal. Biochem. 47: 310-312.

45. Brevet, A., P. Plateau, M. Best-Belpomme, and S. Blanquet. 1985. J. Biol. Chem. 260: 15566-15570.

46. Burke, J.J., J.L. Hatfield, Robert R. Klein, and J.E. Mullet. 1985. Plant Physiol. 78: 394-398.

47. Buzin, C.H., and N.S. Peterson. 1982. J. Mol. Biol. 158: 181-201.

48. Buzin, C.H., and N. Bournias-Vardiabasis. 1984. Proc. Nat. Acad. Sci. U.S.A. 79: 855-859.

49. Catelli, M.G., N. Binart, J.R. Feramisco, and D.M. Helfman. 1985a. Nucleic Acid Res. 13: 6035-6047.

50. Catelli, M.G., N. Binart, I. Jung-Testas, J.M. Renoir, E.E. Baulieu, J.R. Feramisco, and W.J. Welch. 1985b. EMBO J. 4: 3131-3135.

51. Chappell, T.G., W.J. Welch, D.M. Schlossman, K.B. Palter, M.J. Schlesinger, and J.E. Rothman. 1986. Cell 45: 3-13.

52. Cheney, C.M., and A. Shearn. 1983. Develop. Biol. 95: 325-330.

53. Chirgwin, J.M., A.E. Przybyla, R.J. MacDonald, and W.J. Rutter. 1979. Biochemistry 18: 5294-5299.

54. Ciechanover, A., D. Finley, and A. Varshavsky. 1984. Cell 37: 57-66.

55. Cohen, J.B., K. Effron, G. Rechavi, Y. Ben-Neriah, R. Zakut, and D. Givol. 1982. Nucleic Acids Res. 10: 3353-3370.

56. Cohen, R.S., and M. Meselson. 1985. Cell 43: 737-746.

57. Collins, P.L., L.E. Hightower, and L.A. Ball. 1980. J. Virology 35: 682-693.

58. Commerford, S.L. 1971. Biochemistry 10: 1993-1999.

59. Compton, J.L., and B.J. McCarthy. 1978. Cell 14: 191-201.

60. Cooper, P., T.D. Ho, and R.M. Hauptmann. 1984. Plant Physiol. 75: 431-441.

61. Corces, V., R. Holmgren, R. Freund, R. Morimoto, and M. Meselson. 1980. Proc. Nat. Acad. Sci. U.S.A. 77: 5390-5393.

62. Courgeon, A.M., C. Maisonhaute, and M. Best-Belpomme. 1984. Exp. Cell Res. 153: 515-521.

63. Cowing, D.W., J.C.A. Bardwell, E.A. Craig, C. Woolford, R.W. Hendrix, and C.A. Gross. 1985. Proc. Nat. Acad. Sci. U.S.A. 82: 2679-2683.

64. Craig, E.A., B.J. McCarthy, and S.C. Wadsworth. 1979. Cell 16: 575-588.

65. Craig, E.A., and B.J. McCarthy. 1980. Nucleic Acids Res. 8: 4441-4457.

66. Craig, E.A., T.D. Ingolia, and L.J. Manseau. 1983. Dev. Biol. 99: 418-426.

67. Craig, E.A., and K. Jacobsen. 1984. Cell 38: 841-849.

68. Craig, E.A., and K. Jacobsen. 1985. Mol. Cell. Biol. 5: 3517-3524.

69. Craig, E.A., M.R. Slater, W.R. Boorstein, and K. Palter. 1985. UCLA Symp. Mol. Cell Biol. 30: 659-668.

70. Czarnecka, E., L. Edelman, F. Schöffl, and J.L. Key. 1984. Plant Mol. Biol. 3: 45-58.

71. Czarnecka, E., W.B. Gurley, R.T. Nagao, L.A. Mosquera, and J.L. Key. 1985. Proc. Nat. Acad. Sci. U.S.A. 82: 3726-3730.

72. Dean, R.L., and B.G. Atkinson. 1983. Can. J. Biochem. Cell Biol. 61: 472-479.

73. De Jong, W.W. 1982. in Macromolecular Sequences in Systematic and Evolutionary Biology (Goodman, M., eds.) pp. 75-114, Plenum Press, New York.

74. Delaney, A.D. 1982. Nucleic Acids Res. 10: 61-67.

75. Denaro, M., U. Hammerling, L. Rask, and P.A. Peterson. 1984. EMBO J. 3: 2029-2032.

76. Denisenko, O.N. 1984. FEBS Lett. 178: 149-152.

77. DiDomenico, B.J., G.E. Bugaisky, and S. Lindquist. 1982a. Proc. Nat. Acad. Sci. U.S.A. 79: 6181-6185.

78. DiDomenico, B.J., G.E. Bugaisky, and S. Lindquist. 1982b. Cell 31: 593-603.

79. Dildrop, R., M. Brüggemann, A. Radbruch, R. Rajewsky, and K. Beyreuther. 1982. EMBO J. 1: 635-640.

80. DiNocera, P.P., and I.B. Dawid. 1983. Proc. Nat. Acad. Sci. U.S.A. 80: 7095-7098.

81. Dudler, R., and A. Travers. 1984. Cell 38: 391-398.

82. Duncan, R., and J.W.B. Hershey. 1984. J. Biol. Chem. 259: 11882-11889.

83. Dura, J.-M. 1981. Mol. Gen. Genet. 184: 381-385.

84. Ellgaard, E.G. 1972. Chromosoma 37: 417-422.

85. Emmons, S.W., M.P. Klass, and D. Hirsch. 1979. Proc. Nat. Acad. Sci. U.S.A. 76: 1333-1337.

86. Emmons, S.W., L. Yesner, K. Ruan, and D. Katzenberg. 1983. Cell 32: 55-65.

87. Epstein, P., R. Reddy, D. Henning, and H. Bush. 1980. J. Biol. Chem. 255: 8901-8906.

88. Ernst, J.F., J.W. Stewart, and F. Sherman. 1981. Proc. Nat. Acad. Sci. U.S.A. 78: 6334-6338.

89. Ernst, V., E. Zukofsky-Baum, and P. Reddy. 1982. in Heat Shock: From Bacteria to Man (Schlesinger, M.J., M. Ashburner, and A. Tissières, eds.) pp. 215-225, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

90. Farrell-Towt, J., and M.M. Sanders. 1984. Mol. Cell. Biol. 4: 2676-2685.

91. Farrelly, F.W., and D.B. Finkelstein. 1984. J. Biol. Chem. 259: 5745-5751.

92. Faulkner, F.G., and H. Biessmann. 1980. Nucleic Acids Res. 8: 943-955.

93. Faulkner, F.-G., H. Saumweber, and H. Biessman. 1981. J. Cell Biol. 91: 175-183.

94. Fink, K., and E. Zeuthen. 1980. Exp. Cell Res. 128: 23-30.

95. Finley, D., A. Ciechanover, and A. Varshavsky. 1984. Cell 37: 43-55.

96. Flanagan, J.G., M.-P. Lefranc, and T.H. Rabbitts. 1984. Cell 36: 681-688.

97. Ford, M., and M. Fried. 1986. Cell 45: 425-430.

98. Francis, D., and L. Lin. 1980. Dev. Biol. 79: 238-242.

99. Garnier, J., D.J. Osguthorpe, and B. Robson. 1978. J. Mol. Biol. 120: 97-120.

100. Garrels, J.I., and T. Hunter. 1979. Biochim. Biophys. Acta. 564: 517-525.

101. Glover, C.V.C. 1982a. in Heat Shock: From Bacteria to Man (Schlesinger, M.J., M. Ashburner, and A. Tissières, eds.) pp. 227-234, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

102. Glover, C.V.C. 1982b. Proc. Nat. Acad. Sci. U.S.A. 79: 1781-1785.

103. Goff, S.A., L.P. Casson, and A.L. Goldberg. 1984. Proc. Nat. Acad. Sci. U.S.A. 81: 6647-6651.

104. Goff, S.A., and A.L. Goldberg. 1985. Cell 41: 587-595.

105. Goldschmidt-Clermont, M. 1980. Nucleic Acids Res. 8: 235-252.

106. Grosschedl, R., and M.L. Birnstiel. 1980. Proc. Nat. Acad. Sci. U.S.A. 77: 1432-1436.

107. Grossman, A.D., J.W. Erickson, and C.A. Gross. 1984. Cell 38: 383-390.

108. Grunstein, M., and D.S. Hogness. 1975. Proc. Nat. Acad. Sci. U.S.A. 79: 3961-3965.

109. Guedon, G., D. Sovia, J.P. Ebel, N. Befort, and P. Remy. 1985. EMBO J. 4: 3743-3749.

110. Guidice, G., M.C. Roccheri, and M.G. DiBernardo. 1980. Cell. Biol. Int. Rep. 4: 69-74.

111. Gurley, W.B., E. Czarnecka, R.T. Nagao, and J.L. Key. 1986. Mol. Cell. Biol. 6: 559-565.

112. Guttman, S.D., C. Glover, C.D. Allis, and M.A. Gorovsky. 1980. Cell 22: 299-307.

113. Hackett, R.W., and J.T. Lis. 1983. Nucleic Acids Res. 11: 7011-7030.

114. Hall, B.G. 1983. J. Bacteriol. 156: 1363-1365.

115. Harada, F., N. Kato, and S. Nishimura. 1980. Biochim. Biophys. Res. Commun. 95: 1332-1340.

116. Heikkila, J.J., G.A. Schultz, K. Iatrou, and L. Gedamu. 1982. J. Biol. Chem. 257: 12000-12005.

117. Heikkila, J.J., M. Kloc, J. Bury, G.A. Schultz, and L.W. Browder. 1985. Dev. Biol. 107: 483-489.

118. Hickey, E.D., and L.A. Weber. 1982. Biochemistry 21: 1521-1529.

119. Hickey, E., S.E. Brandon, R. Potter, G. Stein, J. Stein, and L.A. Weber. 1986. Nucleic Acids Res. 14: 4127-4145.

120. Hightower, L.E. 1980. J. Cell. Physiol. 102: 407-427.

121. Hightower, L.E., P.T. Guidon, S.A. Whelan, and C.N. White. 1985. in Changes in Eukaryotic Gene Expression in Response to Environmental Stress (Atkinson, B.G., and D.B. Walden, eds.) pp. 197-210, Academic Press, New York.

122. Hiromi, Y., and Y. Hotta. 1985. EMBO J. 4: 1681-1687.

123. Hiromi, Y., H. Okamoto, W.J. Gehring, and Y. Hotta. 1986. Cell 44: 293-301.

124. Hodgetts, R.B., B. Sage, and J.D. O'Connor. 1977. Dev. Biol. 60, 310-317.

125. Hoffman, E.P., and V.G. Corces. 1984. Mol. Cell. Biol. 4: 2883-2889.

126. Hoffman, E.P., and V.G. Corces. 1986. Mol. Cell. Biol. 6: 663-673.

127. Holmgren, R., K. Livak, R. Morimoto, R. Freund, and M. Meselson. 1979. Cell 18: 1359-1370.

128. Holmgren, R., V. Corces, R. Morimoto, R. Blackman, and M. Meselson. 1981. Proc. Nat. Acad. Sci. U.S.A. 78: 3775-3778.

129. Hultmark, D., R. Klemenz, and W.J. Gehring. 1986. Cell 44: 429-438.

130. Hunt, C., and R.I. Morimoto. 1985. Proc. Nat. Acad. Sci. U.S.A. 82: 6455-6459.

131. Iatrou, K., S.G. Tsitilou, and F.C. Kafatos. 1984. Proc. Nat. Acad. Sci. U.S.A. 81: 4452-4456.

132. Ingolia, T.D., E.A. Craig, and B.J. McCarthy. 1980. Cell 21: 669-679.

133. Ingolia, T.D., and E.A. Craig. 1982a. Proc. Nat. Acad. Sci. U.S.A. 79: 525-529.

134. Ingolia, T.D., and E.A. Craig. 1982b. Proc. Nat. Acad. Sci. U.S.A. 79: 2360-2364.

135. Ingolia, T.D., M.J. Slater, and E.A. Craig. 1982. Mol. Cell. Biol. 2: 1388-1398.

136. Ireland, R.C., and E. Berger. 1982. Proc. Nat. Acad. Sci. U.S.A. 79: 855-859.

137. Ireland, R.C., E. Berger, K. Sirotkin, M.A. Yund, D. Osterbur, and J. Fristrom. 1982. Dev. Biol. 93: 498-507.

138. Ish-Horowicz, D., and S.M. Pinchin. 1980. J. Mol. Biol. 142: 231-245.

139. Johnston, D., H. Oppermann, J. Jackson, and W. Levinson. 1980. J. Biol. Chem. 255: 6975-6980.

140. Jones, D., R.H. Russnak, R.J. Kay and E.P.M. Candido. 1986. J. Biol. Chem., 261: 12006-12015.

141. Kapoor, M. 1983. Int. J. Biochem. 15: 639-649.

142. Karin, M., A. Haslinger, H. Holtgreve, R.I. Richards, P. Krauter, H.M. Westphal, and M. Beata. 1984. Nature 308: 513-519.

143. Karlik, C.C., M.D. Coutu, and E.A. Fyrberg. 1984. Cell 38: 711-719.

144. Karn, J., S. Brenner, L. Barnett, and G. Cesareni. 1980. Proc. Nat. Acad. Sci. U.S.A. 77: 5172-5176.

145. Kay, R.J., R.J. Boissy, R.H. Russnak, and E.P.M. Candido. 1986. Mol. Cell. Biol., 6: 3134-3143.

146. Kelley, P.M., and M.J. Schlesinger. 1978. Cell 15: 1277-1285.

147. Kelley, P.M., G. Aliperti, and M.J. Schlesinger. 1980. J. Biol. Chem. 255: 3230-3233.

148. Kelley, P.M., and M.J. Schlesinger. 1982. Mol. Cell. Biol. 2: 267-274.

149. Kennedy, I.M., R.H. Burdon, and D.P. Leader. 1984. FEBS Lett. 169: 267-273.

150. Key, J.L., C.Y. Lin, and Y.M. Chen. 1981. Proc. Nat. Acad. Sci. U.S.A. 78: 3526-3530.

151. Key, J.L., E. Czarnecka, C.-Y. Lin, J. Kimpel, C. Mothershed, and F. Schöffl. 1983. in Current Topics in Plant Biochemistry and Physiology, Vol. 2 (Randall, D.D., D.G. Blevins, R.L. Larson, and R.J. Rapp., eds.) pp. 107-117, University of Missouri Press, Columbia.

152. Kim, Y.-J., J. Shuman, M. Sette, and A. Przybyla. 1983. J. Cell Biol. 96: 393-400.

153. Kim, Y.-J., J. Shuman, M. Sette, and A. Przybyla. 1984. Mol. Cell. Biol. 4: 468-474.

154. King, C.R., T. Schinohara, and J. Piatigorsky. 1982. Science 215: 985-987.

155. King, C.R., and J. Piatigorsky. 1983. Cell 32: 707-712.

156. Kirk, M.M., and D.L. Kirk. 1985. Cell 41: 419-428.

157. Klar, A.J.S., and J.N. Strathern. 1984. Nature 310: 744-748.

158. Klein, H.L. 1984. Nature 310: 748-753.

159. Klemenz, R., D. Hultmark, and W.J. Gehring. 1985. EMBO J. 4: 2053-2060.

160. Klemenz, R., and W.J. Gehring. 1986. Mol. Cell. Biol. 6: 2011-2019.

161. Kodama, R., and G. Eguchi. 1983. Dev. Growth Differ. 25: 261-270.

162. Kothary, R.K., and E.P.M. Candido. 1982. Can. J. Biochem. Cell. Biol. 60: 347-355.

163. Kothary, R.K., D. Jones, and E.P.M. Candido. 1984. Mol. Cell. Biol. 4: 1785-1791.

164. Krawinkel, U., G. Zoebelein, and A.L.M. Bothwell. 1986. Nucleic Acids Res. 14: 3871-3882.

165. Kruger, C., and B.-J. Benecke. 1981. Cell 23: 595-603.

166. Kurtz, A., J. Rossi, L. Petko, and S. Lindquist. 1986. Science 231: 1154-1157.

167. Laemmli, U.K. 1970. Nature 227: 680-685.

168. Lalanne, J.L., F. Brégégère, C. Delarbre, J.P. Abastado, G. Gachelin, and P. Kourilsky. 1982. Nucleic Acids Res. 10: 1039-1049.

169. Lamarche, S., P. Chretien, and J. Landryl. 1985. Biochem. Biophys. Res. Commun. 131: 868-876.

170. Landick, R., V. Vaughn, E.T. Lau, R.A. Van Bogelen, J.W. Erickson, and F.C. Neidhart. 1984. Cell 38: 175-182.

171. Lawson, R., R. Mestril, P. Schiller, and R. Voellmy. 1984. Mol. Gen. Genet. 198: 116-124.

172. Lee, P.C., B.R. Bochner, and B.N. Ames. 1983. Proc. Nat. Acad. Sci. U.S.A. 80: 7496-7500.

173. Leicht, B.G., H. Biessmann, K.B. Palter, and J.J. Bonner. 1986. Proc. Nat. Acad. Sci. U.S.A. 83: 90-94.

174. Leigh-Brown, A.J., and D. Ish-Horowicz. 1981. Nature 290: 677-682.

175. Lemaux, P.G., S.L. Herendeen, P.L. Bloch, and F.C. Niedhardt. 1978. Cell 13: 427-434.

176. Levinson, W., H. Oppermann, and J. Jackson. 1978a. Biochim. Biophys. Acta 518: 401-412.

177. Levinson, W., P. Mikelens, H. Oppermann, and J. Jackson. 1978b. Biochim. Biophys. Acta 519: 65-75.

178. Levinson, W., J. Idriss, and J. Jackson. 1979. Biol. Trace Elements Res. 1: 15-23.

179. Levinson, W., H. Opperman, and J. Jackson. 1980. Biochim. Biophys. Acta 606: 170-180.

180. Li, G.C., and Z. Werb. 1982. Proc. Nat. Acad. Sci. U.S.A. 79: 3218-3222.

181. Li, G.C., and A. Lazlo. 1985. J. Cell. Physiol. 122: 91-97.

182. Liao, L.W., B. Rosenweig, and D. Hirsh. 1983. Proc. Nat. Acad. Sci. U.S.A. 80: 3585-3589.

183. Liebhaber, S.A., M.J. Goossens, and Y.W. Kan. 1980. Proc. Nat. Acad. Sci. U.S.A. 77: 7054-7058.

184. Liebhaber, S.A., M.J. Goossens, and Y.W. Kan. 1981. Nature 290: 26-29.

185. Lindquist, S.L. 1980. J. Mol. Biol. 137: 151-158.

186. Lindquist, S. 1981. Nature 293: 311-314.

187. Liskay, R.M., and J.L. Stachelek. 1983. Cell 35: 157-165.

188. Livak, K.F., R. Freund, M. Schweber, P.C. Wensink, and M. Meselson. 1978. Proc. Nat. Acad. Sci. U.S.A. 75: 5613-5617.

189. Loomis, W.F., and S.A. Wheeler. 1980. Dev. Biol. 79: 399-408.

190. Loomis, W.F., and S.A. Wheeler. 1982. Dev. Biol. 90: 412-418.

191. Lowe, D.G., and L.A. Moran. 1984. Proc. Nat. Acad. Sci. U.S.A. 81: 2317-2321.

192. Lowe, D.G., and L.A. Moran. 1986. J. Biol. Chem. 261: 2102-2112.

193. Maniatis, T., A. Jeffrey, and H. van de Sande. 1975. Biochemistry 14: 3787-3794.

194. Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

195. Manski, W., and K. Malinowsky. 1978. Immunochem. 15: 781-786.

196. Mason, P.J., L.M.C. Hall, and J. Gausz. 1984. Mol. Gen. Genet. 194: 73-78.

197. Matsumoto, Y., H. Yasuda, T. Marunouchi, and M. Yamada. 1983. FEBS Lett. 151: 139-142.

198. Maxam, A., and W. Gilbert. 1980. Methods Enzymol. 65: 499-560.

199. McAlister, L., and D.B. Finkelstein. 1980a. J. Bacteriol. 143: 603-612.

200. McAlister, L., and D.B. Finkelstein. 1980b. Biochim. Biophys. Res. Commun. 93: 819-824.

201. McCormick, W., and S. Penman. 1969. J. Mol. Biol. 39: 315-333.

202. McGarry, T.J., and S. Lindquist. 1985. Cell 42: 903-911.

203. McGinnis, W., A.W. Shermoen, J. Heemskerk, and S.K. Beckendorf. 1983. Proc. Nat. Acad. Sci. U.S.A. 80: 1063-1067.

204. McKenzie, S.L., S. Henikoff, and M. Meselson. 1975. Proc. Nat. Acad. Sci. U.S.A. 72: 1117-1121.

205. McKenzie, S.L., and M. Meselson. 1977. J. Mol. Biol. 117: 279-283.

206. McMaster, G.K., and G.G. Carmichael. 1977. Proc. Nat. Acad. Sci. U.S.A. 74: 4835-4838.

207. Mellor, A.L., E.H. Weiss, K. Ramachandran, and R.A. Flavell. 1983. Nature 306: 792-795.

208. Messing, J. 1983. Meth. Enzymol. 101: 20-78.

209. Mestril, R., P. Schiller, J. Amin, H. Klapper, J. Ananthan, and R. Voellmy. 1986. EMBO J. 5: 1667-1673.

210. Miller, M.J., N. Xueng, and Geiduschek. 1979. Proc. Nat. Acad. Sci. U.S.A. 72: 1117-1121.

211. Minton, K.W., P. Karmin, G.M. Hahn, and A.P. Minton. 1982. Proc. Nat. Acad. Sci. U.S.A. 79: 7107-7111.

212. Mirault, M.-E., M. Goldschmidt-Clermont, L. Moran, A.P. Arrigo, and A. Tissières. 1978. Cold Spring Harbor Symp. Quant. Biol. 42: 819-827.

213. Mirault, M.-E., R. Southgate, and E. Delwart. 1982. EMBO J. 1: 1279-1285.

214. Mitchel, R.E.J., and D.P. Morrison. 1982. Radiat. Res. 90: 284-291.

215. Mitchell, H.K., and L.S. Lipps. 1975. Biochem. Genet. 13: 585-602.

216. Mitchell, J.K., G. Moller, N.S. Petersen, and L. Lipps-Sarmiento. 1979. Dev. Genet. 1: 181-192.

217. Moran, L., M.-E. Mirault, A.P. Arrigo, M. Goldschmidt-Clermont, and A. Tissières. 1978. Phil. Trans. Roy. Soc. Lond. B 283: 391-406.

218. Moran, L., M.-E. Mirault, A. Tissières, J. Lis, P. Schedl, S. Artavanis-Tsakonas, and W.J. Gehring. 1979. Cell 17: 1-8.

219. Mount, S.M. 1982. Nucleic Acids Res. 10: 459-472.

220. Muskavitch, M., and D. Hogness. 1982. Cell 29; 1041-1051.

221. Nazario, M., and J.A. Evans. 1974. J. Biol. Chem. 249: 4934-4942.

222. Neidhardt, F.C., and R.A. Van Bogelson. 1981. Biochim. Biophys. Res. Commun. 100: 894-900.

223. Neidhardt, F.C., R.A. Van Bogelen, and V. Vaughn. 1984. Annu. Rev. Genet. 18: 295-329.

224. Nordheim, A., and A. Rich. 1983. Nature 303: 674-678.

225. Notarianni, E.L., and C.M. Preston. 1982. Virology 123: 113-122.

226. Nover, L., K.-D. Scharf, and D. Neumann. 1983. Mol. Cell. Biol. 3: 1648-1655.

227. O'Farrell, P.H. 1975. J. Biol. Chem. 250: 4007-4021.

228. Ohshima, Y., N. Okada, T. Tani, Y. Itoh, and M. Itoh. 1981. Nucleic Acids Res. 9: 5145-5157.

229. Ollo, R., and F. Rougeon. 1983. Cell 32: 515-523.

230. Olsen, A.S., D.F. Triemer, and M.M. Sanders. 1983. Mol. Cell. Biol. 3: 2017-2027.

231. O'Malley, K., A. Mauron, J.D. Barchas, and L. Kedes. 1985. Mol. Cell. Biol. 5: 3476-3483.

232. Oppermann, H., W. Levinson, and J.M. Bishop. 1981. Proc. Nat. Acad. Sci. U.S.A. 78: 1067-1071.

233. Palmiter, R.D. 1977. J. Biol. Chem. 252: 8781-8783.

234. Palter, K.B., M. Watanabe, L. Stinson, A.P. Mahowald, and E.A. Craig. 1986. Mol. Cell. Biol. 6: 1187-1203.

235. Parker, C.S., and J. Topol. 1984. Cell 37: 273-283.

236. Pelham, H.R.B., and R.J. Jackson. 1976. Eur. J. Biochem. 67: 247-256.

237. Pelham, H.R.B., and M. Bienz. 1982. EMBO J. 1: 1473-1477.

238. Pelham, H.R.B. 1982. Cell 30: 517-528.

239. Petko, L. and S. Lindquist. 1986. Cell 45: 885-894.

240. Phillips, T.A., R.A. Van Bogelen, and F.C. Neidhart. 1984. J. Bacteriol. 159: 283-287.

241. Plesset, J., C. Palm, and C.S. McLaughlin. 1982. Biochim. Biophys. Res. Commun. 108: 1340-1345.

242. Proudfoot, N.J., and G.G. Brownlee. 1976. Nature 263: 211-214.

243. Quax-Jeuken, Y., W. Quax, G. van Rens, P.M. Khan, and H. Bloemendal. 1985. Proc. Nat. Acad. Sci. U.S.A. 82: 5819-5823.

244. Rapaport, E., S.K. Svihovec, and P.C. Zamecnik. 1975. Proc. Nat. Acad. Sci. U.S.A. 72: 2653-2657.

245. Rensing, L. 1973. Cell Differentiation 2: 221-228.

246. Richter, W.W., K.D. Zang, and O.-G. Issinger. 1983. FEBS Lett. 153: 262-266.

247. Riddihough, G. and H.R.B. Pelham. 1986. EMBO J. 5: 1653-1658.

248. Rigby, P.W.J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. J. Mol. Biol. 113: 237-251.

249. Ritossa, F.M. 1962. Experientia 18: 571-573.

250. Ritossa, F.M. 1963. Drosophila Information Service 37: 122-123.

251. Rodenhiser, D., J.H. Jung, and B.G. Atkinson. 1985. Can. J. Biochem. Cell Biol. 63: 711-722.

252. Rose, A.M., D.L. Baillie, E.P.M. Candido, K.A. Beckenbach, and D. Nelson. 1982. Mol. Gen. Genet. 188: 286-291.

253. Rosen, E., A. Sivertsen, R.A. Firtel, S. Wheeler, and W.F. Loomis. 1985. in _Changes in Eukaryotic Gene Expression in Response to Environmental Stress_ (Atkinson, B.G., and D.B. Walden, eds.) pp. 257-278, Academic Press, New York.

254. Rubenstein, P., and J. Deuchler. 1979. J. Biol. Chem. _254_: 11142-11147.

255. Ruppert, S., G. Scherer, and G. Schütz. 1984. Nature _308_: 554-557.

256. Sanders, M.M. 1981. J. Cell. Biol. _91_: 579-583.

257. Sanders, M.M., D.F. Triemer, and A.S. Olsen. 1986. J. Biol. Chem. _261_: 2189-2186.

258. Sanger, F., S. Nicklen, and A.P. Coulson. 1977. Proc. Nat. Acad. Sci. U.S.A. _74_: 5463-5467.

259. Scharf, K.-D., and L. Nover. 1982. Cell _30_: 427-437.

260. Schedl, P., S. Artavanis-Tsakonas, R. Stewart, W.J. Gehring, M.-E. Mirault, J. Goldschmidt-Clermont, L. Moran, and A. Tissières. 1978. Cell _14_: 921-929.

261. Schlesinger, M.J., G. Aliperti, P.M. Kelley. 1982. Trends Biochem. Sci. _7_: 222-225.

262. Schlesinger, M.J. 1985. in _Changes in Eukaryotic Gene Expression in Response to Environmental Stress_ (Atkinson, B.G., and D.B. Walden, eds.) pp. 183-195, Academic Press, New York.

263. Schmid, H.P., O. Akhayat, C. Martin De Sa, F. Puvion, K. Koehler, and K. Scherrer. 1984. EMBO J. _3_: 29-34.

264. Schoenmakers, J.G.G., and H. Bloemendal. 1968. Nature _220_: 790-791.

265. Schöffl, F., and J.L. Key. 1982. J. Mol. Appl. Genet. _1_: 301-314.

266. Schöffl, F., E. Rashke, and R.T. Nagao. 1984. EMBO J. _3_: 2491-2497.

267. Schon, E.A., S.M. Wernke, and J.B. Lingrel. 1982. J. Biol. Chem. _257_: 6825-6835.

268. Schuldt, C., and P.-M. Kloetzel. 1985. Dev. Biol. _110_: 65-74.

269. Schulze, D.H., L.R. Pease, S.S. Geier, A.A. Reyes, L.A. Sarmiento, R.B. Wallace, and S.G. Nathenson. 1983. Proc. Nat. Acad. Sci. U.S.A. _80_: 2007-2011.

270. Scott, M.P., J.M. Fostel, and M.L. Pardue. 1980. Cell _22_: 929-941.

271. Scott, M.P., and M.L. Pardue. 1981. Proc. Nat. Acad. Sci. U.S.A. 78: 3353-3357.

272. Serfling, E., M. Jasin, and W. Schaffner. 1985. Trends Genet. 1: 224-230.

273. Shen, S.-H., J.L. Slightom, and O. Smithies. 1981. Cell 26: 191-203.

274. Simcox, A.A., C.M. Cheney, E.P. Hoffman, and A. Shearn. 1985. Mol. Cell. Biol. 5: 3397-3402.

275. Simon, J.A., C.A. Sutton, R.B. Lobell, R.L. Glaser, and J.T. Lis. 1985. Cell 40: 805-817.

276. Sinibaldi, R., and P.W. Morris. 1981. J. Biol. Chem. 256: 10735-10738.

277. Sirotkin, K., and N. Davidson. 1982. Dev. Biol. 89: 196-210.

278. Slater, A., A.C.B. Cato, G.M. Sillar, J. Kieussis, and R.H. Burdon. 1981. Eur. J. Biochem. 117: 341-346.

279. Slightom, J.L., A.E. Blechl, and O. Smithies. 1980. Cell 21: 627-638.

280. Snutch, T.P., and D.L. Baillie. 1983. Can. J. Biochem. Cell. Biol. 61: 480-487.

281. Snutch, T.P., and D.L. Baillie. 1984. Mol. Gen. Genet. 195: 329-335.

282. Southern, E.M. 1975. J. Mol. Biol. 98: 503-518.

283. Southgate, R., A. Ayme, and R. Voellmy. 1983. J. Mol. Biol. 165: 35-57.

284. Southgate, R., M.-E. Mirault, A. Ayme and A. Tissières. 1985. in Changes in Eukaryotic Gene Expression in Response to Environmental Stress (Atkinson, B.G., and D.B. Walden, eds.) pp. 1-30, Academic Press, New York.

285. Spieth, J., K. Denison, E. Zucker, and T. Blumenthal. 1985. Nucleic Acids Res. 13: 7129-7138.

286. Spradling, A., S. Penman, and M.L. Pardue. 1975. Cell 4: 395-404.

287. Spradling, A., M.L. Pardue, and S. Penman. 1977. J. Mol. Biol. 109: 559-587.

288. Steller, H., and V. Pirrotta. 1985. EMBO J. 4: 3765-3772.

289. Stinchcomb, D.T., J.E. Shaw, S.H. Carr, and D. Hirsh. 1985. Mol. Cell. Biol. 5: 3484-3496.

290. Storti, R.V., M.P. Scott, A. Rich, and M.L. Pardue. 1980. Cell 22: 825-834.

291. Sulston, J.E., and S. Brenner. 1974. Genetics 77: 95-104.

292. Sulston, J.E., E. Schierenberg, and J.G. White. 1983. Dev. Biol. 100: 64-119.

293. Tanguay, R.M., and M. Vincent. 1981. Can. J. Biochem. Cell Biol. 59: 67-73.

294. Tanguay, R.M., and M. Vincent. 1982. Can. J. Biochem. Cell Biol. 60: 306-315.

295. Thomas, G.P., and M.B. Mathews. 1982. in Heat Shock: From Bacteria to Man (Schlesinger, M.J., M. Ashburner, and A. Tissières, eds.) pp. 207-213, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

296. Thomas, P.S. 1980. Proc. Nat. Acad. Sci. U.S.A. 77: 5201-5205.

297. Tilghman, S.M., P.J. Curtis, D.C. Tiemeier, P. Leder, and C. Weissmann. 1978. Proc. Nat. Acad. Sci. U.S.A. 75: 1309-1313.

298. Tilly, K., N. McKittrick, M. Zylicz, and C. Georgopoulos. 1983. Cell 34: 641-646.

299. Tissières, A., H.K. Mitchell, and U.M. Tracy. 1974. J. Mol. Biol. 84: 389-398.

300. Topol. J., D.M. Ruden, and C.S. Parker. 1985. Cell 42: 527-537.

301. Tu, S.-P.D., and S.N. Cohen. 1980. Gene 10: 177-183.

302. Ungewickell, E. 1985. EMBO J. 4: 3385-3391.

303. Van Bogelen, R.A., V. Vaughn, and F.C. Neidhardt. 1983. J. Bacteriol. 153: 1066-1068.

304. van den Heuvel, R., W. Hendriks, W. Quax, and H. Bloemendal. 1985. J. Mol. Biol. 185: 273-284.

305. van der Ouderaa, F.J., W.W. de Jong, A. Hilderink, and H. Bloemendal. 1973. Eur. J. Biochem. 39: 207-211.

306. Velazquez, J.M., B.J. DiDomenico, and S. Lindquist. 1980. Cell 20: 679-689.

307. Vincent, M., and R.M. Tanguay. 1979. Nature 281: 501-503.

308. Vitek, M.P., and E.M. Berger. 1984. J. Mol. Biol. 178: 173-189.

309. Voellmy, R., M. Goldschmidt-Clermont, R. Southgate, A. Tissières, R. Levis, and W. Gehring. 1981. Cell 23: 261-270.

310. Voellmy, R., and P.A. Bromley. 1982. Mol. Cell. Biol. 2: 479-483.

311. Voellmy, R., A. Ahmed, P. Schiller, P. Bromley, and D. Rungger. 1985. Proc. Nat. Acad. Sci. U.S.A. 82: 4949-4953.

312. Wadsworth, S., E.A. Craig, and B.J. McCarthy. 1980. Proc. Nat. Acad. Sci. U.S.A. 77: 2134-2137.

313. Walsh, C. 1980. J. Biol. Chem. 255: 2629-2632.

314. Walter, M.F., and H. Biessmann. 1984. J. Cell Biol. 99: 1468-1477.

315. Wang, A.H.-J., G.J. Quigley, F.J. Kolpak, J.L. Crawford, J.H. van Boom, G. van der Marel, and A. Rich. 1979. Nature 282, 680-686.

316. Wang, C., R.H. Gomer, and E. Lazarides. 1981. Proc. Nat. Acad. Sci. U.S.A. 78: 3531-3535.

317. Wang, S.-W., A.J. Robins, R. d'Andrea, and J.R.E. Wells. 1985. Nucleic Acids Res. 13: 1369-1386.

318. Wannemacher, C.F., and A. Spector. 1968. Exp. Eye Res. 7: 623-625.

319. Welch, W.J. 1985. J. Biol. Chem. 260: 3058-3062.

320. Welch, W.J., and J.R. Feramisco. 1985. Mol. Cell. Biol. 5: 1571-1581.

321. White, C.N., and L.E. Hightower. 1984. Mol. Cell. Biol. 4: 1534-1541.

322. Widelitz, R.B., B.E. Magun, and E.W. Gerner. 1986. Mol. Cell. Biol. 6: 1088-1094.

323. Widera, G.T., and R.A. Flavell. 1984. EMBO J. 3: 1221-1225.

324. Wistow, G. 1985. FEBS Lett. 181: 1-6.

325. Wu, B., C. Hunt, and R. Morimoto. 1985. Mol. Cell. Biol. 5: 330-341.

326. Wu, C. 1984. Nature 309: 229-234.

327. Xiao, C.-M., and J.P. Mascarenhas. 1985. Plant Physiol. 78: 887-890.

328. Yamamori, T., K. Ito, Y. Nakamura, and T. Yura. 1978. J. Bacteriol. 134: 1133-1140.

329. Yamamori, T., and T. Yura. 1982. Proc. Nat. Acad. Sci. U.S.A. 79: 860-864.

330. Yost, H.J., and S. Lindquist. 1986. Cell 45: 185-193.

331. Young, E.T., T. Mattson, G. Selzer, G. Van Houwe, A. Bolle, and R. Epstein. 1980. J. Mol. Biol. 138: 423-445.

332. Yura, T., T. Tobe, K. Ito, and T. Osawa. 1984. Proc. Nat. Acad. Sci. U.S.A. 81: 6803-6807.

333. Zhimulev, I., and V.E. Grafadatskaya. 1974. Drosophila Information Service 51: 96.

334. Zimmerman, J.L., W. Petri, and M. Meselson. 1983. Cell 32: 1161-1170.

335. Zoller, M.J., and M. Smith. 1983. Meth. Enzymol. 100: 468-500.