National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

NL-339 (r. 88/04) c

Canada

CHARACTERIZATION OF THE HSP70 MULTIGENE SUBFAMILIES FROM *CAENORHABDITIS ELEGANS*

by

Mark Frank Patrick Heschl

B.Sc. (Honors), University of Alberta, 1981

M.Sc., University of Calgary, 1984

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Department

of

Biological Sciences

© Mark Frank Patrick Heschl, 1988

Simon Fraser University

June, 1988

All rights reserved. This work may not be reproduced in whole or in part, by photocopy or other means, without permission of the author.

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-48744-5

APPROVAL

Name: Mark Frank Patrick Heschl

Degree: Doctor of Philosophy

Title of Thesis:

Characterization of the hsp70 multigene subfamilies from *Caenorhabditis elegans*

Examining Committee:

Chairman:

Dr. D.L. Baillie, Professor, Senior Supervisor

Dr. M.J. Smith, Professor

Dr. B.M. Honda, Associate Professor

Dr. A.T. Beckenbach, Associate Professor, Public Examiner

Dr. E.P.M. Candido, Professor, U.B.C., Public Examiner

Dr. A.S. Lee, Associate Professor, Department of Biochemistry, University of Southern California, Los Angeles, Calif., External Examiner

Date Approved 21 June 1988

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

Characterization of the hsp70 multigene subfamilies from Caenorhabditis elegans

Author:

(signature)

Mark Frank Patrick Heschl

(name)

June 21, 1988

(date)

ABSTRACT

The *Caenorhabditis elegans* hsp70 gene family consists of at least nine genes and I have sequenced and characterized five of these genes. The constitutively expressed, heat inducible *hsp-1(IV)* gene is highly homologous to the *Drosophila melanogaster Hsc4* gene and the *Saccharomyces cerevisiae SSA1* gene. The *hsp-1* gene shares 5' sequence elements common to other heat inducible genes. Unlike other heat inducible hsp70 genes, the *hsp-1* coding region is interrupted by introns. The *hsp-2ps(X)* gene is a pseudogene of *hsp-1*. Two stop codons generated near the 5' end of the sequence as well as several frame shift mutations and a large internal deletion confirm the identification of *hsp-2ps* as a pseudogene. The *hsp-2ps* gene duplicates a region of *hsp-1* DNA which corresponds exclusively to the transcribed region retaining the introns. The nucleotide substitution rate of the third codon position was twice that of the first or second codon positions suggesting that *hsp-2ps* was nonfunctional since the time of the transpositional duplication event and is estimated to have occurred approximately 8.5 million years ago. The constitutively expressed hsp70C gene product is highly homologous to the rat endoplasmic reticulum-located grp78 gene product. Therefore, the hsp70C gene is probably the *C. elegans* grp78 gene. The hsp70C gene homolog from a closely related species, *C. briggsae*, shares a high degree of homology with the hsp70C gene in the coding and non-coding regions. The 5' regulatory regions share several blocks of sequences one of

which is conserved in the rat grp78 regulatory region. This sequence may be involved in the glucose-mediated response. Other conserved regions encompass a heat shock element and copies of mammalian viral core enhancer sequences. The highly heat inducible hsp70D gene is closely related to the hsp70C and grp78 genes and is the first heat inducible variety of the grp78 family described. The constitutively expressed, heat inducible hsp70F gene shares more identity with the Escherichia coli dnaK gene than with any eukaryotic hsp70 gene identified to date. The hsp70F protein has a leader sequence characteristic of proteins transported into the mitochondrial matrix.

DEDICATION

to

Susan

ACKNOWLEDGEMENTS

I would like to thank my senior supervisor, Dr. David L. Baillie, as well as Dr. Ann Rose and the members of my supervisory committee, Drs. Michael J. Smith and Barry M. Honda for their suggestions in the course of my research and the preparation of the thesis and manuscripts. I would also like to thank my colleagues Karen Beckenbach, Debashish Bhattacharya, Denise Clark, Shiv Prasad and Raja Rosenbluth for their helpful discussions and encouragement. Thanks to Eric Cabot for his assistance with the computer and the various programs. Finally, I would like to thank my friends and family (old and new) for their support during the highs and lows of the research and writing of the thesis.

This work was supported by a Medical Research Council of Canada Studentship and Simon Fraser University Graduate Fellowships.

TABLE OF CONTENTS

LIST OF FIGURES

GENERAL INTRODUCTION

When organisms or cells in culture are exposed to a sudden elevation of temperature, they respond by synthesizing a small specific set of highly conserved proteins, the heat shock proteins (hsp). The heat shock response was first observed by Ritossa (1962) when he shifted Drosophila hydei larvae from 20°C to 37°C. Examination of the salivary gland polytene chromosomes revealed an alteration in the puffing pattern indicative of an alteration in gene activity. This change in puffing pattern was also induced by exposing the larvae to dinitrophenol or sodium salicylate. The heat shock response, or more appropriately the stress shock response, is a universal response which has been observed in every organism studied including bacteria, yeast, mammals and plants (Schlesinger et al., 1982; Craig, 1985; Lindquist, 1986; Burdon, 1986).

The heat shock response is characterized by a vigorous and transient activation of a small number of specific genes previously either active at low levels or inactive (Craig, 1985; Lindquist, 1986). This increase in expression is accompanied by repressing transcription of normally active genes (Berendes, 1968; Spradling et al., 1975; Jamrich et al., 1977; Findly and Pederson, 1981) and translation of pre-existing mRNAs (McKenzie et al., 1975; Lindquist, 1981; Kelly and Schlesinger, 1982; DiDomenico et al., 1982a). In addition to an increase in ambient temperature the heat shock response can also be induced, in whole or in part, by the presence of denatured or abnormal proteins (Finley et al., 1984; Karlik et

al., 1984; Hiromi and Hotta, 1985; Ananthan et al., 1986), ethanol (Li, 1983; Plesset et al., 1982; Neidhardt et al., 1984), heavy metals (Levinson et al., 1980; Burdon et al., 1982; Courgeon et al., 1984; Key et al., 1985; Li and Laszlo, 1985), glucose starvation (Sciandra and Subjeck, 1983), viral infection (Collins and Hightower, 1982; Nevins, 1982; Notarianni and Preston, 1982; Khandjian and Turler, 1983; Wu and Morimoto, 1985), sodium arsenite (Burdon et al., 1982; Kothary and Candido, 1982; Tanguay and Vincent, 1982; Li, 1983; Key et al., 1985), amino acid analogs (Kelley and Schlesinger, 1978; Hightower, 1980; Thomas and Mathews, 1984), oxygenation following anoxia (Li and Schrieve, 1982; Sciandra et al., 1984) and inhibitors of oxidative phosphorylation (Ashburner and Bonner, 1979; Schlesinger et al., 1982). Pre-treatment with any of the inducers of the heat shock response also induces thermotolerance, i.e. the ability to survive lethal temperatures (Plesset et al., 1982; Li, 1983; Velazquez and Lindquist, 1984; Key et al., 1985).

The exact number and types of proteins synthesized varies from one organism to the next but falls within several well defined groups (Schlesinger et al., 1982; Craig, 1985; Lindquist, 1986). These groups include proteins ranging in size from 80 to 110 kDa (the hsp80/90 class), 62 to 78 kDa (the hsp70 class), 16 to 35 kDa (the low molecular weight class) and 8 kDa (ubiquitin; Bond and Schlesinger, 1985). The heat shock proteins represent some of the most highly conserved proteins throughout evolution (Ingolia et al., 1982; Bardwell and Craig,

1984; Bienz, 1984; Farrelly and Finkelstein, 1984). This is exemplified by the observation that one of the human hsp70 proteins (Hunt and Morimoto, 1985) shares 47% identity at the amino acid level with the Escherichia coli hsp70 equivalent, dnaK (Bardwell and Craig, 1984).

The heat shock proteins are highly conserved both in sequence and in structure. Their presence plays an essential role in the survival of and the recovery from stress. In fact, some of the heat inducible genes are present in multiple copies (Ashburner and Bonner, 1979; Ingolia et al., 1980; Bienz, 1984). Related proteins are important not only during times of cellular stress but also under non-stressed conditions. Several isoforms of the hsps have been found in the unstressed cells. These hsp-like proteins have been termed heat shock cognates or hsc and are important normal cellular constituents (Schlesinger et al., 1982; Hughes and August, 1982; Ingolia and Craig, 1982; Velazquez et al., 1983; Bardwell and Craig, 1984; Lowe and Moran, 1984; Craig and Jacobsen, 1984; Chappell et al., 1986; Munro and Pelham, 1986). Some of these hscs are heat inducible (Ingolia et al., 1982; Craig and Jacobsen, 1984; Palter et al., 1986) and have been termed constitutive genes to distinguish them from the non-heat inducible hscs. Some of the hsp70 isoforms may also be developmentally regulated (Dura, 1981; Sirotkin and Davidson, 1982; Bensuade et al., 1983; Cheney and Shearn, 1983; Zimmerman et al., 1983; Mason et al., 1984; Vitek and Berger, 1984; Kurtz and Lindquist, 1984; Bienz, 1985; Glaser et al., 1986; Kurtz et al., 1986; Palter et al.,

1986). Within each class size the different isoforms are related and as such are members of specific multigene families (Craig, 1985; Lindquist, 1986).

Expression of the heat shock genes is mediated through a heat shock transcription factor, HSTF (Parker and Topol, 1984a, 1984b; Wu et al., 1987). The HSTF recognizes and binds to a heat shock element, or HSE (Parker and Topol, 1984b; Wu, 1984a; Wiederrecht et al., 1987), found in the 5' regulatory region of all heat inducible hsps. The HSE sequence CNNGAANNTTCNNG, is found approximately 20 base pairs upstream of the TATA box, exhibits dyad symmetry and has been found in all eukaryotic heat inducible genes studied (Pelham, 1982, 1985). Proteins such as HSTF that bind to the HSE facilitate the transcription of the hsp genes during heat shock (Wu, 1984a). Differences in transcriptional activity and kinetics of activation of the individual heat shock genes may be controlled by the number of HSEs and the affinity of HSTF for the HSEs (Wu, 1984a). The HSTF binds to the HSE in stressed Drosophila, HeLa and yeast cells (Wu, 1984b; Kingston et al., 1987; Sorger et al., 1987) and unstressed yeast cells (Sorger and Pelham, 1987; Sorger et al., 1987). Although HSTF does not bind to the HSE in unstressed Drosophila and HeLa cells, HSTF is present in the cells (Parker and Topol, 1984b; Wu et al., 1987; Zimarino and Wu, 1987). Activation of the yeast HSTF occurs when HSTF is phosphorylated (Sorger et al., 1987). In E. coli there is no HSTF. Instead, the product of the htpR gene, a 32 kDa sigma factor, replaces the normal 70 kDa sigma factor and facilitates

transcription of the heat inducible genes while under stress conditions (Neidhardt and VanBogelen, 1981; Yamamori and Yura, 1982; Grossman et al., 1984; Grossman et al., 1985; Neidhardt, 1987).

How is HSTF activated during periods of stress? Ubiquitin is part of a protein degradation system (Hershko and Ciechanover, 1982; Hershko, 1983; Hershko et al., 1984). Overloading the ubiquitin-dependent protein degradation system induces the heat shock response (Munro and Pelham, 1985; Ananthan et al., 1986). Therefore, upon heat shock or stress, the amount of damaged, abnormal and denatured protein increases which may result in a transient shortage of free ubiquitin. A lack of ubiquitin is known to induce the heat shock response (Finley et al., 1984). The synthesis of ubiquitin during the heat shock response would be to cope with the demand for ubiquitin by the protein degradation system and to replenish the cellular stocks of ubiquitin. The hsp70 proteins also negatively control their own synthesis (DiDomenico et al., 1982b; Tilly et al., 1983; Craig and Jacobsen, 1984; Grossman et al., 1984). It has been suggested that the hsp70 proteins, once having reached a certain level, might bind to HSTF and inactivate it. Hsp70 has also been reported to have a proteolytic activity (Mitchell et al., 1985) which could then cleave HSTF.

Translation of only hsp mRNAs during heat shock and not of those normally present is mediated through the presence of specific signals within the 5' leader sequence of heat shock

mRNAs (McGarry and Lindquist, 1985). These sequences are thought to adopt a conformation allowing only the heat shock mRNAs to be translated under heat shock conditions. Repression of translation of pre-existing pre-mRNAs may also be mediated through the lack of intron processing (Yost and Lindquist, 1986).

One of the best studied hsp classes is the 70 kDa class. In order to determine the function of the hsp70s it is important to know where they are found before and after heat shock. Protein localization experiments reveal that the heat inducible hsp70s before and after heat shock were distributed between the nucleus and cytoplasm. After heat shock, the concentration of hsp70 protein increases in both compartments (Arrigo et al., 1980; Welch and Feramisco, 1984). The majority of the hsp70 protein in the nucleus is concentrated within the nucleolus and is bound to the chromosomes at the interband regions (Arrigo et al., 1980; Velazquez et al., 1980; Welch and Feramisco, 1984). The chromosome-bound hsp70 protein is resistant to nuclease digestion (Sinibaldi and Morris, 1981; Levinger and Varshavsky, 1981; Welch and Feramisco, 1984). The heterochromatic regions and the chromocenters of Drosophila polytene chromosomes are not bound by the hsp70 proteins (Arrigo et al., 1980; Velazquez et al., 1980). Heat shock also appears to disrupt the nuclear ribonucleoproteins (Maynard and Pederson, 1983). Nuclear proteins become insoluble and operationally part of the nuclear matrix. The hsp70s are bound to the nuclear matrix in a salt-resistant or hydrophobic manner

(Sinibaldi and Morris, 1981; Levinger and Varshavsky, 1981; Welch and Suhan, 1985).

A number of cellular morphological changes occur during heat shock. The Golgi apparatus is disrupted and fragmented concomitant with the appearance of vesicularized membranes and the mitochondria swell (Welch and Suhan, 1985). The intermediate filament network collapses around the nucleus. Associated with the intermediate filaments are hsp70 proteins (Welch and Suhan, 1985; Palter et al., 1986) and the migration of the hsp70 proteins to the nucleus may be the result of the collapse of the intermediate filament network during heat shock (Biessmann et al., 1982; Sanders et al., 1982a, 1982b). Also, the cells become flattened as the number of actin-containing stress fibers increases (Welch and Suhan, 1985). This last feature is a characteristic of growth-arrested cells (Thomas et al., 1982). The nucleoli become less condensed accompanied by an increase in nucleolar size and the number of granular ribonucleoprotein components. This change is concomitant with the binding of hsp70-like proteins to the partially assembled ribosomes (Welch and Suhan, 1985).

The function of the hsp70 proteins has long been an enigma. Recently, identification of some of the hsp70 proteins and their properties has led to a partial understanding of their function. The dnaK protein binds ATP and has a DNA-independent ATPase activity (Zylicz et al., 1984). The dnaK protein is involved in replication of λ DNA. During λ DNA replication, the λO protein recognizes the origin of replication of the λ

DNA. The λP protein binds to both the O protein and the host dnaB helicase protein. The dnaK protein then interacts with the P protein, hydrolyzes ATP and catalyzes the dissociation of the P protein from the helicase which allows movement of the helicase along the DNA (Dodson et al., 1986).

Eukaryotic hsp70s share the ability to bind and hydrolyze ATP (Welch and Feramisco, 1985). An observed tight association of hsp70 with other proteins can only be reversed by the addition or presence of ATP (Lewis and Pelham, 1985). The rat hsc70 or "uncoating ATPase" (Rothman and Schmid, 1986; Chappell et al., 1986; Ungewickell, 1985) uses ATP to disrupt the cage of clathrin triskelions surrounding coated vesicles. However, there is a thirty fold excess of hsc70 over clathrin suggesting that hsc70 is involved in many cellular functions rather than being used exclusively to uncoat clathrin coated vesicles. The endoplasmic reticulum localized glucose regulated protein, grp78, is also a member of the hsp70 family and was previously known as the immunoglobulin heavy chain binding protein "BiP" (Munro and Pelham, 1986). BiP or grp78 protein recognizes the exposed hydrophobic regions of the immunoglobulin heavy chain and binds to these sites until an immunoglobulin light chain is available (Bole et al., 1986). The heavy chain is released from the grp78 protein by the addition of ATP (Munro and Pelham, 1986). Recently, members of the Saccharomyces cerevisiae SSA hsp70 subfamily have been shown to facilitate the translocation of proteins into mitochondria, the lumen of the endoplasmic reticulum (Desaies et al., 1988) and microsomes

(Chirico et al., 1988).

A model for the function of the hsp70-related proteins, based on the above properties, was proposed by Hugh Pelham (1986). He suggested that the hsp70 proteins were involved in the assembly/disassembly of proteins and structures containing proteins under normal and heat shock conditions. During heat shock or stress, proteins become partially denatured or protein synthesis disrupted. This can result in hydrophobic regions becoming exposed which would then interact with other exposed hydrophobic regions to form insoluble aggregates. The hsp70 protein would bind to the hydrophobic surfaces limiting the interaction of the hydrophobic regions. If aggregates did form, the hsp70 protein would then catalyze the disaggregation of the insoluble aggregates. Once hsp70 was bound to the substrate it would use ATP to undergo a conformational change and then release itself from the substrate. Before release the conformational change of the hsp70 would distort the substrate resulting in a weakening or breaking of the interactions of the substrate within the aggregate. After release, the substrate could then refold or reassemble into its pre-heat shock state. The freed hsp70 would then bind a new substrate and undergo the cycle once more. In fact, this process is enhanced if there is an over-production of the hsp70 proteins (Pelham, 1984). Under heat shock conditions, the hsp70 protein would function in the same manner binding to exposed hydrophobic regions preventing the formation of insoluble aggregates or, in an ATP-dependent reaction, dissociating proteins within a protein complex or

aggregate.

The function of the heat shock proteins has been primarily determined by molecular biological and cellular localization studies. Recently, the eukaryotic hsp proteins have been the subject of genetic analysis parallelling the genetic studies of the hsp proteins in E. coli (Neidhardt et al., 1984). As a method of determining the function of various hsps, null mutations of specific heat shock genes have been created in the yeast Saccharomyces cerevisiae. Genetic and functional analysis of the hsp70 multigene family of S. cerevisiae has revealed at least four subfamilies based on their expression characteristics and their ability to complement null alleles of other hsp70 genes (Craig and Jacobsen, 1984, 1985; Craig et al., 1987; Werner-Washburne et al., 1987). These genes, when disrupted, may: 1) have no effect (SSD1 [Craig et al., 1987]); 2) have no effect until combined with one or more mutant genes (the SSA subfamily [Craig and Jacobsen, 1984; Werner-Washburne et al., 1987] and the SSB subfamily [Craig and Jacobsen, 1985]); or 3) have a lethal effect (SSC1 [Craig et al., 1987]). The SSA subfamily has many complex interactions. When either SSA1 or SSA2 are disrupted, no mutant phenotype is observed. However, when null alleles of both SSA1 and SSA2 are present, the yeast cells will not grow at higher than normal temperatures but still retain thermotolerance (Craig and Jacobsen, 1984). During normal growth, the double mutant ssa1 ssa2 induces the synthesis of the heat inducible gene SSA4. When a triple mutant ssa1 ssa2 ssa4 is constructed, the yeast

cell cannot survive. This lethal phenotype can be rescued by placing the other heat inducible gene SSA3 under the control of the SSA2 promoter (Werner-Washburne et al., 1987). SSB1 and SSB2, when either is present as a null allele, appears to have no effect on the cell but when both are present as null alleles the yeast cell does not grow normally at reduced temperatures. The double mutant ssb1 ssb2 cannot be rescued by gene members of the SSA subfamily (Craig and Jacobsen, 1985). These results demonstrate that gene members within a subfamily can compensate and at least partially duplicate the function of the other members of that family but cannot compensate functionally for mutant genes in other families. The compensatory action is most likely due to the hsp70 proteins within a family having a similar function but the non-complementary action is most likely due to the different cellular locations of the hsp70 subfamilies.

Studying the genetics of the heat shock response in a multicellular organism would contribute to the understanding of the complex functions of the heat shock response. Isolation of mutant hsp genes in D. melanogaster has been hampered by the numerous copies of each hsp gene as well as the size and complexity of the organism (E.A. Craig, personal communication). Caenorhabditis elegans offers an alternative to D. melanogaster in the study of the genetics of the heat shock response in multicellular eukaryotes. The self-fertilizing hermaphrodite is an excellent model system for the genetic and biochemical analysis of the eukaryotic genome

(Brenner, 1974; Riddle, 1980). The C. elegans haploid genome consists of five autosomes and one X chromosome (Brenner, 1974) and consists of approximately $8x10^7$ base pairs (Sulston and Brenner, 1974). C. elegans is a relatively simple eukaryote consisting of approximately 1000 somatic cells in the adult nematode (Sulston and Horvitz, 1977). It is differentiated into various tissue types including hypodermis, nerve, gonad, muscle and gut and a complete cell lineage has been determined for C. elegans (Sulston and Horvitz, 1977; Sulston et al., 1983).

The heat shock response of C. elegans has recently been characterized (Snutch and Baillie, 1983; Snutch, 1984; Snutch et al., 1988). Upon heat shock, synthesis of eight major groups of proteins with molecular weights of approximately 81 kDa, 70 kDa, 41 kDa, 38 kDa, 29 kDa, 19 kDa, 18 kDa and 16 kDa is initiated. These proteins first appear at 29°C and continue to be synthesized at temperature shifts up to the lethal temperature of 35°C. Like the heat shock response of other organisms, transcription and translation of non-heat shock mRNAs appears to be repressed. The dauer larva, a facultative larval stage formed when environmental conditions are unfavorable, also produce a set of mRNAs upon heat shock even though they do not appear to be synthesizing non-heat shock mRNAs (Snutch and Baillie, 1983; Snutch, 1984). The hsp70 gene family appears to have at least nine members (Snutch, 1984; Snutch et al., 1988). Sequences corresponding to some of the hsp70 genes (Snutch, 1984; Snutch and Baillie, 1984; Snutch et

al., 1988) and the hsp16 genes (Russnak and Candido, 1985; Jones et al., 1986) have been isolated.

In S. cerevisiae, mutations in the heat shock genes were constructed in vitro using the one-step gene replacement method of Rothstein (1983). There is no analogous system in C. elegans that allows for a targeted replacement of the heat shock genes. Therefore, in order for mutations to be induced in the hsp70 genes, they must first be isolated, their expression characterized and the genes identified. The genes could then be mapped genetically using restriction fragment length differences (RFLD) between two C. elegans strains Bristol and Bergerac. Mutant phenotypes for each of the identified hsp70 genes can be predicted and a genetic screen devised to isolate mutant alleles of the heat shock response. These mutants could then be verified as hsp70 mutants by transforming the mutants with the purified gene (Fire, 1986).

As part of the long term goal in the study of the genetics of the heat shock response in C. elegans, my first objective was to identify the six hsp70 genes previously isolated (Snutch, 1984; Snutch et al., 1988). By sequencing the hsp70 genes I could then determine their homologs from other organisms. These results could then be used to predict a mutant phenotype and then proceed with the long term goals outlined above. Second, I wanted to compare one of the hsp70 genes from C. elegans to its homolog from a closely related species C. briggsae. Snutch (1984) and Prasad (1988) have stated that only genetically important sequences are conserved

between the two species which can be detected by cross-hybridization. I wanted to extend this idea and show that a comparison of the sequences from the 5' regulatory region could potentially reveal sequence elements important for gene expression.

The first gene to be described, hsp-1, is a constitutively expressed, heat inducible hsp70 gene closely related to the SSA genes from S. cerevisiae and the Hsc4 and heat inducible hsp70 genes from D. melanogaster. The second gene, hsp-2ps, is an unusual pseudogene of hsp-1. The third and fourth genes, a heat shock cognate gene, hsp70C, and a highly heat inducible gene, hsp70D, belong to a grp78-related subfamily. And finally, a fifth gene, the constitutively expressed, slightly heat inducible hsp70F gene, is more closely related to the dnaK gene from E. coli than any known eukaryotic hsp70 gene identified to date.

CHAPTER 1

A HEAT INDUCIBLE COGNATE GENE, *hsp-1*

INTRODUCTION

In Drosophila melanogaster, the hsp70 gene family consists of both heat inducible and constitutively expressed genes. A family of genes homologous to hsp70 are expressed under normal growth conditions (Ingolia and Craig, 1982; Craig et al., 1983). These genes are termed heat shock cognate genes or hsc. At least three D. melanogaster hsc70 genes have been identified including Hsc1, Hsc2 and Hsc4 which map at 70C, 87D and 88E respectively. These hsc70 genes do not map at the same location as the heat inducible hsp70 and hsp68 genes which map at 87A,C and 95D, respectively. The mRNA levels of each of the hsc70 genes differs in adults. Hsc4 mRNA levels approach the level of actin mRNA while Hsc1 and Hsc2 mRNA levels are 30- to 60-fold less abundant.

The hsc70 protein is the Hsc4 gene product (Palter et al., 1986). hsc70 is enriched in embryos and the Hsc4 mRNA is abundant in oocytes representing a large percentage of the maternal message stored in the egg. Synthesis of hsc70 still occurs after heat shock although the synthesis is repressed slightly. The presence of the hsc proteins during heat shock suggests that the cognate proteins are not dispensable during heat shock. In contrast to the synthesis of hsc70, the synthesis of Hsc4 mRNA is enhanced 2-fold upon heat shock. In yeast, a family related to the Hsc4 gene, the SSA family, has been identified (Craig and Jacobsen, 1984; Werner-Washburne et al., 1987). SSA1 is both constitutively expressed and heat inducible, SSA2 is constitutively expressed and non-heat

inducible while SSA3 and SSA4 are heat inducible. Some members of the SSA gene subfamily encode proteins which facilitate the translocation of proteins into the mitochondria, the lumen of the endoplasmic reticulum (Desaies et al., 1988) and microsomes (Chirico et al., 1988).

In C. elegans, the hsp-1 gene encodes the hsp70A protein (Snutch, 1984; Snutch et al., 1988). The hsp-1 gene is constitutively expressed and has been mapped to the right end of linkage group IV. hsp-1 transcripts are most abundant in the L1 larva and decrease by about 50% by the adult stages but remain an abundant message. Transcription of hsp-1 mRNA is stimulated 3-fold upon heat shock and remains heat inducible during larval development.

In this chapter I describe the DNA sequence of the hsp-1 gene. hsp70A shares a high degree of homology with the strictly heat inducible and the constitutively expressed, heat inducible hsp70-like genes from D. melanogaster and S. cerevisiae. Similarities in expression characteristics suggest that the hsp-1 gene is more closely related to the constitutively expressed, heat inducible hsc70 genes. In addition the hsp-1 regulatory region shares features common to the regulatory regions of other eukaryotic heat inducible genes.

MATERIALS AND METHODS

Construction of plasmids for sequencing

The phage containing the hsp-1 gene, hsl 151 (Snutch, 1984; Snutch et al., 1988) was subcloned into the plasmid vector pUC19 (Norrander et al., 1983). The plasmid containing the hsp-1 gene, pCes401, was identified and the orientation of the insert determined by the pattern derived from restriction digests. Isolation of the plasmid pCe6.2 is described elsewhere (Snutch, 1984).

Overlapping plasmid deletions were made using either exonuclease III (Henikoff, 1984) or restriction enzymes.

Plasmid DNA preparation

Plasmid DNA for deletion and sequencing reactions was prepared using the mini alkali lysis method (Maniatis et al., 1982) except that two phenol:sevag extractions were done and after the first ethanol precipitation, the air dried pellet was resuspended in 0.25 M sodium acetate and reprecipitated with ethanol. After digestion with RNase, the plasmid DNA samples were precipitated with polyethylene glycol (Hattori and Sakaki, 1986).

DNA sequencing and sequence analysis

Dideoxy sequencing was performed on denatured plasmid DNA (Chen and Seeburg, 1985; Sanger et al., 1980).

DNA sequences were analyzed using the computer program Microgenie (Beckman). Visual inspection and preparation of the DNA sequences for publication was done with the aid of the computer program ESEE (E. Cabot, personal communication).

RESULTS AND DISCUSSION

Characterization of the hsp-1 gene sequence

A restriction map of the *hsp-1* gene is shown in Figure 1-1. The entire sequence of the *hsp-1* gene and its flanking DNA is presented in Figure 1-2. Within the 3 kb region sequenced an open reading frame corresponding to a *Drosophila hsp70* gene (Ingolia et al., 1980) was found. The open reading frame was interrupted by three short introns of 49, 194 and 55 bp. The position of the first intron in *hsp-1* (aa69) is similar to that of an intron in the non-heat shock inducible *Drosophila Hsc1* (aa66) and *Hsc2* (aa58/59) genes (Craig et al., 1983). The sequence of the intron boundaries (AG/GTAAGT.....TTTCAG/G) matched that found for other *C. elegans* introns (Karn et al., 1983; Spieth et al., 1985b). The positions of the introns were confirmed by sequencing the corresponding regions of a full length cDNA of *hsp-1* (pCe 6.2; Figure 1-1, lower case letters). The pCe6.2 cDNA included the first met residue of the polypeptide and extended 150 bp downstream of the UAA termination codon. The mature *hsp-1* mRNA would encode a 640 amino acid polypeptide of 69,851 daltons.

Within the transcribed leader sequence is a 3' splice site at -6 (Figure 1-2). These 3' splice sites are candidates for a *trans* splicing event (Krause and Hirsh, 1987). At the 5' end of the pCe6.2 cDNA sequence was a stretch of DNA matching the *trans* spliced leader sequence (Figure 1-3). It has recently been demonstrated in primer extension experiments that the *hsp-1* mRNA is *trans* spliced under both heat shock and non-heat

Figure 1-1. Restriction map of the hsp-1 gene. Regions corresponding to the coding sequences are shaded. The transcribed, non-translated regions are unshaded. Introns are shown by breaks in the shaded coding region. B, BamHI; E, EcoRI; H, HindIII; K, KpnI; O, XhoI; S, SalI; X, XbaI.

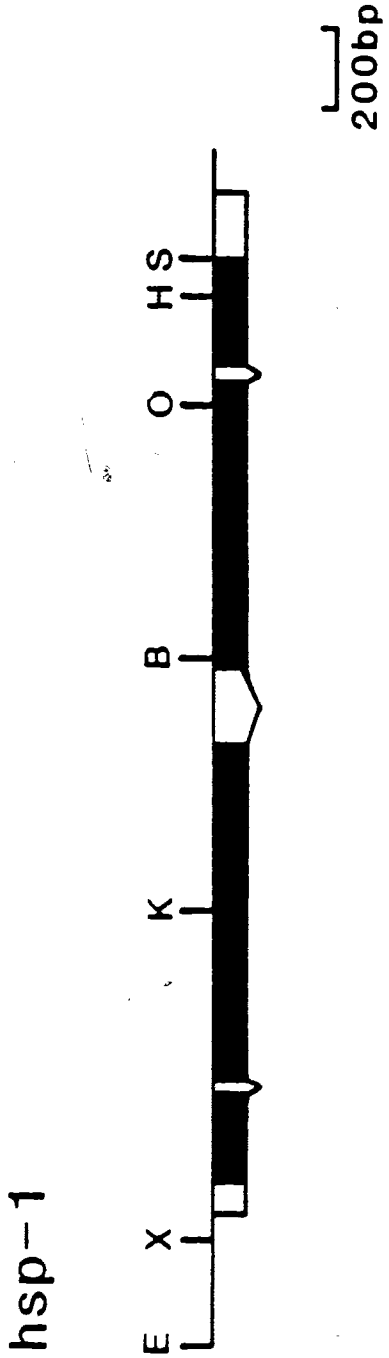

hsp-1
E
X
K
B
O
HS
200bp

Figure 1-2. Nucleotide sequence of the hsp-1 gene. Sequence from the cDNA pCe6.2 is shown on the bottom line and is in lowercase letters. Numbering is with respect to the start of translation. The putative TATA box and polyA addition signal are shown. Also shown are the three heat shock elements (HSE). The amino acid sequence is shown above the DNA sequence. The regions involved in the potential stem and loop structure are indicated with a *. Transcription starts at or near -78 (v). A 3' splice site involved in trans splicing is located at -6.

EcoR1 HSE3
-422 GAATTCGCCGAGTTATTCTTCCTCAAAATGTGGTGCGCGCTCTCTCTCCCCCTTTTCGTCGCGAACATTCTCTGCGAGGCATCTCTCTTCTTTTAATTCA

-322 CAATTCTCAACACTTTTCTGTAGGCAAAACTCTCTATATTGCTCCTTTTTCAGATTTTTGTTCAAACTTTTTTTGTATTTATCCTTGTTCAAGTGTTTTC

HSE2 HSE1 XbaI
-222 CATTCAGCAGTTACAGACTATTTAAGGAAATTTTAGGTTTTTAGCACATTTTTCTAATTTTTTGACGAAATTCGAATTTTCTAGAATCCCGCCACGCCCA
* ***** *****

TATA V
-122 GTCATCTAGTAAATTTGTTGAACTTCATTTCTCTATTTTTAATCATTGTTCTCGACGTCCTAATTTTTTATCTCCATTTGAGTGACTATTTCTTGATTTT
***** ***** *

M S K H N A V G I D L G T T Y S C V G V F M H G K V
-22 TAAATTATTTTTTACAGTAAAAATGAGTAAGCATAACGCTGTTGGAATCGATTTGGGAACTACCTACTCCTGCGTGGGAGTTTTCATGCACGGAAAGGTA
agtaaaaatgagtaagcataacgctgttggaatcgatttgggaactacctactcctgcgtgggagttttcatgcacggaaaggta

E I I A N D Q G N R T T P S Y V A F T D T E R L I G D A A K N Q V
79 GAAATCATTGCCAACGATCAAGGAAACCGTACAACTCCATCATATGTGGCTTTCACCGACACCGAGCGTCTCATCGGAGATGCTGCCAAGAATCAAGTTG
gaaatcattgccaacgatcaaggaaaccgtacaactccatcatatgtggctttcaccgacaccgagcgtctcatcggagatgctgccaagaatcaagttg

A M N P H N T V F D A K R L I G R
179 CCATGAACCCACATAACACTGTTTTCGGTAAGTGCTTAATTTTATATTTTCAACAATCTCAACGTCTGTTTTTCAGATGCCAAACGTCTTATTGGACGCA
ccatgaacccacataacactgttttcg--atgccaaacgtcttattgga

K F D D P A V Q S D M K H W P F K V I S A E G A K P K V Q V E Y K G
279 AGTTCGACGATCCAGCAGTTCAGTCTGACATGAAGCATTGGCCATTCAAGGTCATCTCTGCCGAAGGAGCTAAGCCAAAGGTCCAAGTTGAGTACAAAGG

E N K I F T P E E I S S M V L L K M K K T A E A F L E P T V K D A
379 AGAGAACAAGATCTTCACTCCAGAAGAGATCTCCTCAATGGTTCTGCTGAAGATGAAGAAGACTGCCGAGGCTTTCCTTGAACCGACCGTCAAGGATGCC

V V T V P T Y F N D S Q R Q A T K D A G A I A G L N V L R I I N E
479 GTTGTCACTGTCCCGACTTACTTCAACGACTCGCAGCGTCAAGCCACCAAGGATGCCGGAGCCATCGCTGGACTCAACGTTCTCCGTATCATCAACGAGC

KpnI
P T A A A I A Y G L D K K G H G E R N V L I F D L G G G T F D V S I
579 CAACCGCTGCAGCTATCGCTTACGGACTTGACAAGAAGGGACACGGAGAACGCAACGTTCTTATCTTCGATCTTGGAGGTGGTACCTTCGATGTCTCCAT

L T I E D G I F E V K S T A G D T H L G G E D F D N R M V N H F C
679 TCTTACCATTGAGGACGGAATCTTCGAAGTCAAGTCTACCGCTGGAGACACTCATCTTGGAGGAGAGGACTTCGATAACCGCATGGTGAACCACTTCTGT

A E F K R K H K K D L A S N P R A L R R L R T A C E R A N E T L S
779 GCCGAGTTCAAGCGCAAGGACAAGAAGGATCTTGCTTCCAACCCACGTGCTCTTCGTCGTCTTCGTACCGCCTGCGAGCGCGCAAACGAGACTCTTTCGT
gagactctttcgt

S S C Q A S I E I D S L F E G I D F Y T N I T R A R F E E L C A D L
879 CGTCTTGCCAGGCTTCGATTGAGATCGATTCTCTCTTCGAAGGAATTGACTTCTACACCAACATCACTCGTGCTCGTTTCGAGGAGCTCTGCGCTGATCT
cgtcttgccaggcttcgattgagatcgattctctcttcgaaggaattgacttctacaccaacatcactcgtgctcgtttcgaggagctctgcgctgatct

F R S T M D P V E K S L R D A K M D K S Q
979 CTTCAGATCCACCATGGACCCAGTCGAGAAGTCTCTCCGTGACGCCAAGATGGACAAGAGCCAAGTAAGTTTTCAAGAAAATTTGTGTTGTTAGTCTGCT
cttcagatccaccatggacccagtcgagaagtctctccgtgacgttaagatggacaagagccaa--------------------------

1079 TATCCTCTTCTGTTCCGGCTTCGATTCCCGTAGTTTTTGCCGATATAAAATGAGGAAATATAACTCTGGCATAACCCTACTGTTGCGCTTTTATGTGTTG

BamHI
V H D I V L V G G S T R I P
1179 TCTGTTTCTTAGTCGCCTGTACAATTAATGCAATAAAATCTAATTTGATAATTTTCAGGTTCATGACATCGTCCTTGTCGGAGGATCCACTCGTATCCCA
---gttcatgacatcgtccttgtcggaggatcc

K V Q K L L S D L F S G K E L N K S I N P D E A L A Y G A A V Q A
1279 AAGGTCCAGAAACTTTTGTCCGATCTCTTCTCAGGAAAGGAATTGAACAAGTCCATCAACCCAGATGAGGCGTTAGCCTACGGAGCTGCCGTCCAAGCCG

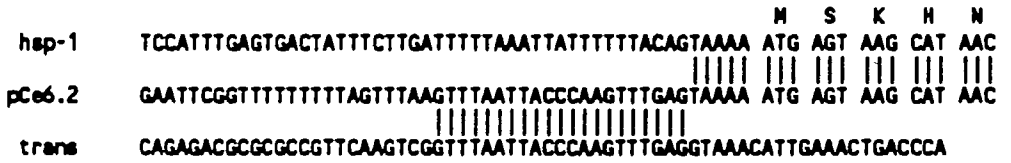
A I L S G D K S E A V Q D L L L L D V A P L S L G I E T A G G V M T
1379 CTATCCTCTCCGGAGACAAGTCTGAGGCTGTCCAGGATCTTCTTCTTCTTGACGTTGCCCCACTTTCCCTTGGTATTGAGACCGCTGGAGGAGTCATGAC

A L I K R N T T I P T K T A Q T F T T Y S D N Q P G V L I Q V Y E
1479 TGCTCTCATCAAGAGAAACACCACCATCCCAACCAAGACCGCTCAGACCTTCACAACCTATTCTGATAACCAACCAGGAGTGTTGATCCAGGTTTACGAA

G E R A M T K D N N L L G K F E L S G I P P A P R G V P Q I E V T
1579 GGAGAACGTGCCATGACCAAGGACAACAACTTGCTCGGAAAGTTCGAGCTCTCCGGAATCCCACCAGCACCACGCGGAGTCCCACAAATCGAAGTCACTT

```
     F  D  I  D  A  M  G  I  L  M  V  S  A  T  D  K  S  T  G  K  A  K  Q  I  T  I  T  M  D  K  D  R  F  S
1679 TCGATATTGACGCCAACGGAATCTTGAACGTCTCTGCCACTGACAAGTCCACCGGAAAGGCAAAACAAATCACCATCACCAACGACAAGGATCGCTTTTC

                                                                                                  XhoI
     K  D  D  I  E  R  M  V  N  E  A  E  K  Y  K  A  D  D  E  A  Q  K  D  R  I  G  A  K  N  G  L  E  S
1779 CAAGGATGACATTGAACGCATGGTCAACGAAGCTGAGAAATACAAGGCTGACGATGAGGCCCAAAAGGACCGTATTGGAGCCAAGAACGGACTCGAGTCA
                                                                                  ggctgacgatgaggcccaaaaggaccgtattggagccaagaacggactcgagtca

     Y  A  F  N  L  K  Q  T  I  E  D  E  K                                                          L  K
1879 TACGCCTTCAACCTTAAGCAGACCATTGAGGACGAGAAGGTTAGTTAATTAATTTATATTTTGTCAACAAGTTTTTAATTTCCATTTTTTTCAGCTCAAG
     tacgccttcaaccttaagcagaccattgaggacgagaag----------------------------------------------------------ctcaag

     D  K  I  S  P  E  D  K  K  K  I  E  D  K  C  D  E  I  L  K  W  L  D  S  N  Q  T  A  E  K  E  E  F
1979 GATAAGATCAGCCCAGAAGACAAGAAGAAGATCGAGGACAAGTGCGACGAGATCTTGAAGTGGCTCGACAGCAACCAGACCGCAGAGAAGGAGGAGTTCG
     gataagatcagcccagaagacaagaagaagatcgaggacaagtgcgacgagatcttgaagtggctcgacagcaaccagaccgcagagaaggaggagttcg

                                                  HindIII
     E  S  Q  Q  K  D  L  E  G  L  A  K  P  D  L  S  K  L  Y  Q  S  A  G  G  A  P  P  G  A  A  P  G  G  A
2079 AGTCACAACAGAAGGATTTGGAAGGATTGGCCAAGCCCGATCTTTCCAAGCTTTACCAGAGTGCCGGAGGAGCCCCACCAGGTGCTGCTCCAGGAGGAGC
     agtcacaacagaaggatttggaaggattggccaagcccgatctttccaagctttaccagagtgccggaggagccccaccaggtgctgctccaggaggagc

                                        SalI
     A  G  G  A  G  G  P  T  I  E  E  V  D  #
2179 CGCCGGAGGAGCTGGAGGACCAACGATCGAGGAGGTCGACTAATTATTTATCTTCTTTTTTGATCTCGGTTTTTATCTTTATTCTCTTTTCCCCCAAATC
     cgccggaggagctggaggaccaacgatcgaggaggtcgactaattatttatcttcttttttgatctcggtttttatctttattctcttttcccccaaatc

                                                                                    poly A
2279 CCCTATATAATCGATCTTCTGCCCCCTTCCCTGTCACCCACCCAAACCGTTATTACGCCTAGAATGGTGAACAATAAAAAAATTAAAATAATCCAAGTTG
     ccctatataatcgatcttctgccccctttccctgtcacccacccaaaccgttattacgcctagaatggtgaacaataaaaaaattaaaataa

2379 TTTGCTTTCGTGTTCTTTCCAGCAGAATCGGTTTTAAGGATGGCAGCTTTTTTTCACAATCGGCATATTTGTGTGACGTCTTGATGCTTCTAATCGTCTT

2479 GACAAGCAAAGAAATGGATAAAATTCAACTCTACACG
```

Figure 1-3. Comparison of the hsp-1 cDNA to the trans spliced leader sequence. The cDNA sequence (pCe6.2) is compared to the genomic DNA (hsp-1) and the trans spliced leader DNA (trans; Krause and Hirsh, 1987). DNA homologies are indicated.


M S K H N
hsp-1 TCCATTTGAGTGACTATTTCTTGATTTTTAAATTATTTTTTACAGTAAAA ATG AGT AAG CAT AAC
pCe6.2 GAATTCGGTTTTTTTTTAGTTTAAGTTTAATTACCCAAGTTTGAGTAAAA ATG AGT AAG CAT AAC
trans CAGAGACGCGCGCCGTTCAAGTCGGTTTAATTACCCAAGTTTGAGGTAAACATTGAAACTGACCCA

shock conditions (K. vanDoren and D. Hirsh, personal communication).

Heat shock genes have previously been considered unique among eukaryotes in their relative lack of introns. The presence of introns in the heat inducible members of the C. elegans hsp16 family (Russnak and Candido, 1985; Jones et al., 1986), and as I report here, in the hsp-1 gene (this chapter), the hsp70C and hsp70D genes (Chapter 3) and the hsp70F gene (Chapter 4), suggests a fundamental difference in gene structure between the heat shock genes of C. elegans and those of other organisms. It has been suggested for Drosophila and other organisms that heat shock genes are relatively devoid of introns in order to circumvent a block in RNA splicing which occurs during a severe heat shock (Yost and Lindquist, 1986). All eight C. elegans heat shock inducible genes sequenced to date (the four hsp16 genes [Russnak and Candido, 1985; Jones et al., 1986] and the four hsp70 genes reported in this thesis) contain at least one intron. Thus, if the theory proposed by Yost and Lindquist (1986) also applies to C. elegans, I would expect that transcripts from these genes would be correctly processed and translated only under conditions of mild or moderate heat shock. It will be interesting to see if the strictly heat inducible members of the hsp70 gene family that are related to hsp-1 have introns or not.

Upstream of the coding region are three sequences which are homologous to the consensus heat shock element (HSE) CNNGAANNTTCNNG (Pelham, 1982); HSE1 (88% homology), HSE2 (75%

homology) and HSE3 (88% homology) (Figure 1-2). It has recently been suggested that the 14 base pair HSE may be better described as a dimer of the 10 base pair sequence NTTCNNGAAN (Xiao and Lis, 1988). This dimer sequence is required for heat induciblity. Further examination of HSE1 revealed a one and a half repeat (cGAAttTTCtaGAAt) which is suggested to be just as functional as the dimer (Xiao and Lis, 1988). An examination of the nucleotide sequence in the HSE1 region also revealed a putative TATA box, TAAATT from -113 to -108, 25 nucleotides downstream of HSE1. Two sequences analogous to the cap site consensus signal in C. elegans (Klass et al., 1988) were found from -80 to -75 and from -98 to -93, 28 and 10 nucleotides downstream of the putative TATA box (Figure 1-2). Flanking the putative TATA box were two repeats, one inverted with respect to the other. The first one, from -118 to -114, has one nucleotide difference from the second sequence, from -103 to -99 (Figure 1-2). The presence of inverted repeats is also seen around the TATA box in a Drosophila heat inducible hsp70 gene (Ingolia et al., 1980). Encompassing the TATA box and HSE1 was a palindromic sequence stretching from -150 to -138 and from -118 to -104 with 11 out of 13 matches (Figure 1-2). These two stretches are capable of pairing with each other to form a stem and loop structure. The resulting loop would be 19 nucleotides long. HSE2 and HSE3 do not have any sequences surrounding them that are capable of forming stem and loop structures. The final sequences of note, flanking HSE1 and HSE2, are three potential CCAAT boxes, CCAAT1 from -125 to

-121 and CCAAT2 from -169 to -165 and CCAAT3 from -205 to -201 (Figure 1-2). CCAAT1 is within the loop of the potential stem and loop structure.

hsp70A is homologous to hsc70 and the SSA1 protein

Overall, hsp70A is highly homologous to both hsp70 and hsc70 from Drosophila and yeast (Figure 1-4). At the amino acid level hsp70A is 75% identical to a Drosophila hsp70 protein (Ingolia et al., 1980) and 77% identical with the yeast hsp70 protein products of the SSA1 and SSA2 genes (Ingolia et al., 1982; M. Slater and E.A. Craig, personal communication). It is also 81% identical to the Drosophila hsc70 (Craig et al., 1983; M. Slater and E.A. Craig, personal communication). Based on the protein similarities, hsc70 is probably the homolog of hsp70A. Figure 1-4 also shows that a similar homology profile exists between hsp-1 and the other hsp70 family members; the N-terminus and the initial approximate 550 amino acids are most highly conserved, while there is marginal conservation at the C-terminal portion of hsp70 related genes. An exception to this, as noted by Hunt and Morimoto (1985), is the conservation between hsp70 related genes of the peptide EEVD at the extreme C-terminus.

The hsp-1 gene of C. elegans is probably the D. melanogaster Hsc4 and S. cerevisiae SSA1 gene equivalent based on protein homologies and similarities in the expression characteristics. Like Hsc4 and SSA1, hsp-1 produces a highly abundant message under non-stress conditions and the synthesis of this mRNA is increased during heat stress. In addition, the

Figure 1-4. Amino acid comparison of the hsp-1 gene product, hsp70A, to hsc70 (Craig et al., 1983; M. Slater and E.A. Craig, personal communication), the SSA1 and SSA2 gene products (Ingolia et al., 1982; M. Slater and E.A. Craig, personal communication) and a heat inducible D. melanogaster hsp70 (Ingolia et al., 1980). Amino acid matches are indicated with a dot (.) and gaps are shown with a dash (-).

```
hsp70A  MSKHNAVGIDLGTTYSCVGVFMHGKVEII-ANDQGNRTTPSYVAFTDTERLIGDAAKNQVAMNPHNTVFDAKRLIGRKFDD   80
hsc70   ...AP.....Q.....-.....TQ.I.....                                                    80
SSA1    ...--.....AH.ANDR.D..-.....F.....A....S.....N.N.                                   78
SSA2    ...--.....AH.SNDR.D..-.....F.G.....A....A.....N.NG                                 80
hsp70   .P---.I.....YQ.....NN.Y.....S...N.EP.....R.....Y..                                 78

hsp70A  PAVQSDMKHWPFKVISAEGAKPKVQVEYKGENKIFTPEEISSMVLLKMKETAEAFLEPTVKDAVVTVPTYFNDSQRQATK  160
hsc70   A.....E.V..D.-...IE.T..D.K.T.F.....T...E....Y.GK..TN..I...A.....                   159
SSA1    .E..A....F...L.DVD.-..QI...F...T.N....Q..P.F.G...E...SY.GAK.N.....A.....           157
SSA2    .E..G....F...L.DVD.-..QI...F...T.N....Q.....G...E...SY.GAK.N.....A.....           157
hsp70   .KIAE.....V.DG.-...IG.....S.R.A.....T...E....Y.GESIT...I...A.....                   157

hsp70A  DAGAIAGLNVLRIINEPTAAAIAYGLDKKGHGERNVLIFDLGGGTFDVSILTIEDGI-FEVKSTAGDTHLGGEDFDNRMVN  240
hsc70   ...T...P..P.....AV.....S.D...-.....L.T                                              239
SSA1    ...T.....KE.H-.....L.F.....-....A.....L..                                          236
SSA2    ...T.....KE.H-.....L.S.....-....A.....L..                                          236
hsp70   ...H.....L.....:NLK.....DE.SL...R.....L.T                                          238

hsp70A  HFCAEFKRKHKKDLASNPRALRRLRTACERANETLSSSCQASIEIDSLFEGIDFYTNITRARFEELCADLFRSTMDPVEK  320
hsc70   ..VQ.....TT.K.....KR.....T.....T....S.....N.....                                   319
SSA1    ..IQ.....N....ST.Q.....KR.....A.T.V.....S.....L.....                              316
SSA2    ..IQ.....N....ST.Q.....KR.....A.T.V.....S.....L.....                              316
hsp70   .LAE.....Y....R.....A...KR.....TE.T....A....Q....KVS.....N...N.LQ....             318

hsp70A  SLRDAKMDKSQVHDIVLVGGSTRIPKVQKLLSDLFSGKELNKSINPDEALAYGAAVQAAILSGDKSEAVQDLLLLDVAPL  400
hsc70   A.....L...VI.....R..Q...N.....V.....H....QE.....T..                               399
SSA1    V.....L.....DE.....VT.Y.N...P.R.....V.....T..E.SKT.....                           396
SSA2    V.....L.....DE.....VT.Y.N...P.R.....V.....T..E.SKT.....                           396
hsp70   A.N.....G.I.....S..QEF.H..N..L.....V.....Q.GKI..V..V.....                          398

hsp70A  SLGIETAGGVMTALIKRNTTIPTKTAQTFTTYSDNQPGVLIQVYEGERAMTKDNNLLGKFELSGIPPAPRGVPQIEVTFD  480
hsc70   .....SV.....QT...I.....                                                            479
SSA1    .....K..P..S..S..KFEI.S..A.....F.....K.....W.....Y                                476
SSA2    .....K..P..S.....KSEV.S..A.....F.....K.....                                        476
hsp70   .....K..E..CR..C.QTK..S.....S.....A..T.D.....                                      478

hsp70A  IDANGILNVSATDKSTGKAKQITITNDKDRFSKDDIERMVNEAEKYKADDEAQKDRIGAKNGLESYAFNLKQTIEDEKLK  560
hsc70   .....T.LER..N.ENK.....G.L..E.....RNE..K..ET.A.....C..M.A.LDEDN..                  559
SSA1    V.S....I...VE.G...SNK.....G.L..E...K..A....F.EE..KESQ..-...Q....I.YS..N..SEAGD.    555
SSA2    V.S.....VE.G...SNK.....G.L..E...K..A....F.EE..KESQ..AS..Q...I.YS..N..SEAGD.        556
hsp70   L.....KEM.....N...K...G.L.QAE.D..... ..ADE..KHRQ..TSR.A....V..V..SV.QAPAG          558

hsp70A  DKISPEDKKKIEDKCDEILKWLDSNQTAEKEEFESQQKDLEGLAKPDLSKLYQSAGGAPPGAAPG------GAAG-----  629
hsc70   T...DS.RTT.L...N.I-....A..L.ARRSTSTARRNGRVC.TRS.PS...G..F-...GM..GGGGMP....AAGAA  637
SSA1    LEQADK.TVTKAEETIS---.....T..S....DDKL.E.QDI.N.IM.....-...-----...------....-----  615
SSA2    LEQADK.AVTKKAEETI--A.....T..T....DD.L.E.QEV.N.IM.....-...------..E------...P-----  617
hsp70   KLDEA-..NSVL...N.TIR.....T.....DHKMEE.TRHCS.IMT.MH.Q--..------------....-----  617

hsp70A  GAG---------------GPTIEEVD                                                          640
hsc70   ...GA----------.....                                                                650
SSA1    ..SGGFGGGAPPAPAPEAE...V....                                                         642
SSA2    ---GGFPGGAPPA--PEAE...V....                                                         639
hsp70   .P.AMCGQQAGGFGGYS--...V....                                                         642
```

Hsc4 gene produces a maternally stored mRNA (Palter et al., 1986). It has been suggested that the hsp-1 gene is expressed in the germ line because of the unusually high amount of DNA polymorphisms surrounding the hsp-1 gene (Snutch and Baillie, 1984). Therefore, it would not be surprising to find that the hsp-1 mRNA is also a maternally stored message. The other hsp-1 homolog, the SSA1 gene, belongs to a multigene subfamily of the S. cerevisiae hsp70 gene family. If the hsp70 gene organization is similar between C. elegans and S. cerevisiae, I believe that there are other members of this C. elegans hsp70 gene subfamily, including the strictly heat inducible hsp70 gene, which need to be isolated.

Possible mutant phenotypes of the hsp-1gene

The similarities of the hsp-1 gene with the members of the SSA gene subfamily and the Hsc4 gene facilitates predicting a possible phenotype for the hsp-1 gene. It is probable that a lethal mutation of the hsp-1 gene would not be recovered because of the compensating actions of the other proposed members of the hsp-1 gene subfamily. Mutations of individual genes in the SSA subfamily cannot be recovered because other SSA members compensate for the lack of the gene product (Craig and Jacobsen, 1984; Werner-Washburne et al., 1987). Since the hsp-1 gene is very similar to the Hsc4 gene, the hsp-1 gene is most likely a maternally expressed mRNA and as such, one should be able to isolate a maternal effect lethal of the hsp-1 gene.

CHAPTER 2

A HEAT SHOCK PSEUDOGENE, hsp-2ps

INTRODUCTION

Two members of the C. elegans hsp70 gene family, hsp-1 and hsp-2ps (formerly hsp70B; Snutch et al., 1988), cross hybridize under high stringency conditions suggesting that these two genes are closely related (Snutch, 1984; Snutch et al., 1988). The hsp-1 gene has been mapped to linkage group (LG)IV (Snutch, 1984; Snutch et al., 1988) while hsp-2ps has been localized to the left arm of LGX (D. Albertson, personal communication). At the level of transcription, the hsp-1 gene is constitutively expressed and heat inducible while no transcripts have been detected for the hsp-2ps gene (Snutch, 1984; Snutch et al., 1988). Only a homolog of the hsp-1 gene and not the hsp-2ps gene has been found in the closely related species C. briggsae. Based on these features, it was suggested that the hsp-2ps gene was a pseudogene of hsp-1. Several pseudogenes from three different C. elegans multigene families have been identified to date: the hsp-70 family (Snutch, 1984; Snutch et al., 1988), the vitellogenin (vit) family (Spieth et al., 1985a) and the sperm-specific (msp) family (Ward et al., 1988).

Here I describe the DNA sequence of the hsp-2ps gene. The hsp-2ps gene is unusual in that it duplicates only a part of the hsp-1 transcription unit including the introns but does not duplicate any of the flanking region. I present my arguments suggesting that the hsp-2ps gene was nonfunctional from the time of the duplication event.

MATERIALS AND METHODS

Construction of plasmids for sequencing

The phage containing the hsp-2ps gene, hsl 143 (Snutch, 1984; Snutch et al., 1988) was subcloned into the plasmid vector pUC19 (Norrander et al., 1983). The plasmid containing the hsp-2ps gene, pCes440, was identified from the other fragments by its restriction pattern.

Overlapping plasmid deletions were made using either exonuclease III (Henikoff, 1987) or restriction enzymes.

Plasmid DNA preparation

Plasmid DNA for deletion and sequencing reactions was prepared using the mini alkali lysis method (Maniatis et al., 1982) except that two phenol:sevag extractions were done and after the first ethanol precipitation, the air dried pellet was resuspended in 0.25 M sodium acetate and reprecipitated with ethanol. After digestion with RNase, the plasmid DNA samples were precipitated with polyethylene glycol (Hattori and Sakaki, 1986).

DNA sequencing and sequence analysis

Dideoxy sequencing was performed on denatured plasmid DNA (Chen and Seeburg, 1985; Sanger et al., 1980).

DNA sequences were analyzed using the computer program Microgenie (Beckman). Visual inspection and preparation of the DNA sequences for publication was done with the aid of the computer program ESEE (E. Cabot, personal communication).

Gel electrophoresis

Genomic DNAs (2 ug) were digested with 20 units of EcoRI

for 4 hours (Davis et al., 1980). The reactions were stopped by the addition of 1/10 volume 50% sucrose, 25 mM EDTA, 2% BPB and 1xTBE. The DNA was electrophoresed through a 0.7% agarose gel containing 10 ug ethidium bromide/100 ml and 1xTBE (Davis et al., 1980). Marker DNA was λ (cI857 Sam7) HindIII-digested DNA. The DNA was visualized and photographed under a 300 nm wavelength transilluminator.

DNA transfer, nick-translation and hybridizations

The DNA was transferred to nitrocellulose using the bidirectional method of Smith and Summers (1980). The nick-translations were done according to Davis et al. (1980). The reactions were stopped by the addition of sodium dodecyl sulphate and EDTA to a final concentration of 0.3% and 10 mM, respectively. The specific activity of the resulting probes was approximately $1x10^8$ cpm/ug DNA.

Hybridization of the filters to the ^{32}P-labelled probe was done essentially according to the procedure outlined in Davis et al. (1980) except that the carrier DNA and formamide were omitted and 2.5x Denhardt's (1x=0.02% bovine serum albumin, 0.02% Ficoll and 0.02% polyvinyl pyrrolidone) was used. All the hybridizations and post-hybridizations steps were carried out at 68°C. Post-hybridization washes were done using 1xSSPE instead of 2xSSPE. The filters were air dried and exposed to Kodak X-Omat K film at -70°C for 3 days.

RESULTS AND DISCUSSION

Characterization of the hsp-2ps gene sequence

The hsp-2ps gene was sequenced and homology with the hsp-1 gene was determined to be 88%. A comparison of the hsp-2ps gene to the hsp-1 gene revealed that the hsp-2ps sequence homology was truncated and missing the last one-third of the sequence corresponding to the 3' end of the hsp-1 gene (Figure 2-1). Sequence homology between hsp-2ps and hsp-1 began just upstream of the start of translation. The homology extended through the introns to a point two-thirds of the way down the hsp-1 sequence. At this point, the sequence showed no identity (ie. approximately 25% identity) (Figure 2-2). Sequences upstream of the start of transcription, including the hsp-1 gene TATA box, were not conserved. No 3' flanking sequences from the hsp-1 gene were found flanking the hsp-2ps gene (Snutch, 1984; Snutch et al., 1988). There was a large internal deletion of 243 base pairs, from +723 to +966 of the corresponding hsp-1 sequence (Figure 2-1). There was one single nucleotide insertion and a number of single nucleotide deletions and substitutions. These changes are shown in Figure 2-2. Two of the substitutions generated stop codons occurring near the 5' end of the sequence. These features have led me to confirm the identification of the hsp-2ps gene as a pseudogene.

Estimation of the time of gene duplication and inactivation

The rate of nucleotide substitution between the corresponding coding regions was examined. The substitutions were not randomly located with respect to the codon positions.

Figure 2-1. Restriction maps of the hsp-2ps and hsp-1 genes. Limits of homology are defined by the dotted lines (see also Figure 2-2). Regions corresponding to the coding regions are shaded. Transcribed, non-translated regions are unshaded. Introns are represented by breaks in the coding region. B, BamHI; E, EcoRI; H, HindIII; K, KpnI; O, XhoI; S, SalI; X, XbaI.

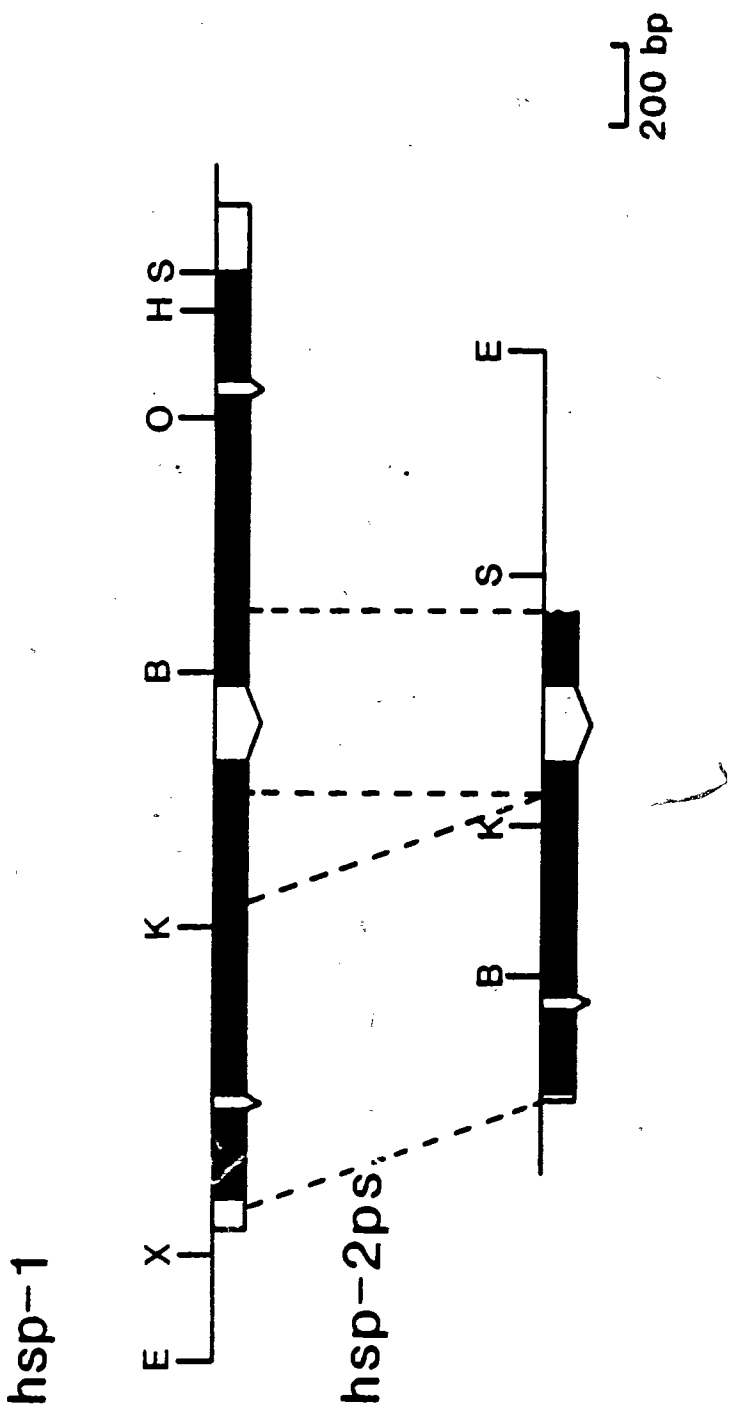
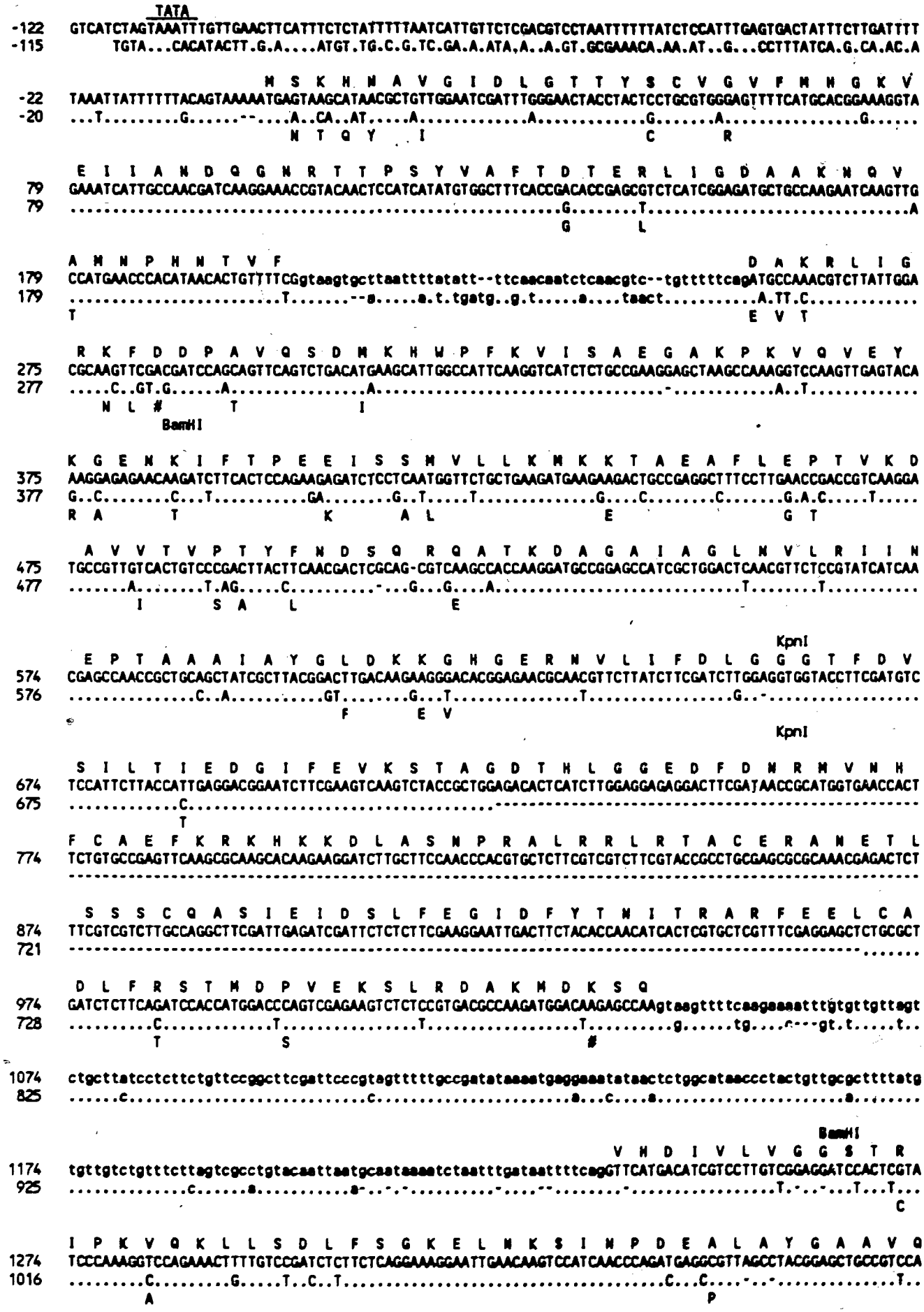

200 bp
hsp-1
E
X
K
B
O
HS
hsp-2ps
B
K
S
E

Figure 2-2. Nucleotide sequence comparison of the hsp-2ps and hsp-1 genes. The hsp-2ps sequence is on the bottom line and the hsp-1 sequence is on the upper line. Introns are shown in lowercase letters. Nucleotide matches are indicated by a dot (.), deletions with a dash (-). The predicted amino acid sequences are shown above or below their respective DNA sequences with only the amino acid differences being shown for the hsp-2ps gene. Inframe stop codons are indicated with a #. The putative TATA box and polyA addition signal of the hsp-1 gene are indicated to show the extent of hsp-1 transcription.


TATA
BamHI
KpnI
KpnI
BamHI

```
          A  A  I  L  S  G  D  K  S  E  A  V  Q  D  L  L  L  L  D  V  A  P  L  S  L  G  I  E  T  A  G  G  V
1374  AGCCGCTATCCTCTCCGGAGACAAGTCTGAGGCTGTCCAGGATCTTCTTCTTCTTGACGTTGCCCCACTTTCCCTTGGTATTGAGACCGCTGGAGGAGTC
1116  ..............................AC.C.A....G.....AC.GGAC..T.CT..A.TT..CA.T.CTCCG.C.GA..ACAA.TTTC.TAG
                                     D     K  E        A  G  H  F  S  T  F  H  F  S  V  G  T  Q  V  S  #

          M  T  A  L  I  K  R  N  T  T  I  P  T  K  T  A  Q  T  F  T  T  Y  S  D  N  Q  P  G  V  L  I  Q  V
1474  ATGACTGCTCTCATCAAGAGAAACACCACCATCCCAACCAAGACCGCTCAGACCTTCACAACCTATTCTGATAACCAACCAGGAGTGTTGATCCAGGTTT
1216  TCA..CT..AATT...GT..TGTTGTGT.G.CAAGCC.A.CT.....GTC.TTGCCGC.TC.A..CAAG.CCTC.TTGTGCCTTA.T..CTC..CCC.GG
                          SalI

          Y  E  G  E  R  A  M  T  K  D  N  N  L  L  G  K  F  E  L  S  G  I  P  P  A  P  R  G  V  P  Q  I  E  V
1574  ACGAAGGAGAACGTGCCATGACCAAGGACAACAACTTGCTCGGAAAGTTCGAGCTCTCCGGAATCCCACCAGCACCACGCGGAGTCCCACAAATCGAAGT
1316  T.ATCTC.TC.ATC.TGTCA.TGCCACTA..ATCG..TAGT.AC.TT..T.TAA.TGAT....A.TT.TGGTG..T.GAGCT.TAGAGCTT.G.T..TTG

          T  F  D  I  D  A  N  G  I  L  N  V  S  A  T  D  K  S  T  G  K  A  K  Q  I  T  I  T  N  D  K  D  R
1674  CACTTTCGATATTGACGCCAACGGAATCTTGAACGTCTCTGCCACTGACAAGTCCACCGGAAAGGCAAAACAGATCACCATCACCAACGACAAGGATCGC
1416  TGAAAGTAG.T.CATT.AAGC.CACCGTGG.TT.A..A.C.A...GAG..CCGTACT.TACTCC.AC..TA.CGAG.TG.CA..TGT.A.ACTATGG.AT
                                                                                                  XhoI
          F  S  K  D  D  I  E  R  M  V  N  E  A  E  K  Y  K  A  D  D  E  A  Q  K  D  R  I  G  A  K  N  G  L
1774  TTTTCCAAGGATGACATTGAACGCATGGTCAACGAAGCTGAGAAATACAAGGCTGACGATGAGGCCCAAAAGGACCGTATTGGAGCCAAGAACGGACTCG
1516  CG..T.CG.CT.ATAT.GCCGTTGTC.CAA...ATTT.CT.TGCCACTGTACTGATTTCGATATGTAT.T.T.TA.C.C.CCCCAGT.TTC.TCA.T.AT

          E  S  Y  A  F  N  L  K  Q  T  I  E  D  E  K
1874  AGTCATACGCCTTCAACCTTAAGCAGACCATTGAGGACGAGAAGgttagttaattaatttatattttgtcaacaagtttttaatttccatttttttcagC
1616  GA.AGCGATT.GC.CCTT.CGTC..CTTT...A.T.TGATTTCTT.C.CC.CT.CT..A.TA...ACTCTTTTTCC..G...TCAGTTCGA...G.A.AT

          L  K  D  K  I  S  P  E  D  K  K  K  I  E  D  K  C  D  E  I  L  K  M  L  D  S  N  Q  T  A  E  K  E  E
1974  TCAAGGATAAGATCAGCCCAGAAGACAAGAAGAAGATCGAGGACAAGTGCGACGAGATCTTGAAGTGGCTCGACAGCAACCAGACCGCAGAGAAGGAGGA
1716  .TGGAATGTCCCAT.C.TTCCTCATGTTC..A..ATCG.GCCT.TTT.TTTTTA.AT...GCTCT..C.ATTT..C.GTTTG....ATT..ATTC
                                                                                                  EcoRI
                                                HindIII
          F  E  S  Q  Q  K  D  L  E  G  L  A  K  P  D  L  S  K  L  Y  Q  S  A  G  G  A  P  P  G  A  A  P  G
2074  GTTCGAGTCACAACAGAAGGATTTGGAAGGATTGGCCAAGCCCGATCTTTCCAAGCTTTACCAGAGTGCCGGAGGAGCCCCACCAGGTGCTGCTCCAGGA
                                     SalI
          G  A  A  G  G  A  G  G  P  T  I  E  E  V  D  #
2174  GGAGCCGCCGGAGGAGCTGGAGGACCAACGATCGAGGAGGTCGACTAATTATTTATCTTCTTTTTTGATCTCGGTTTTTATCTTTATTCTCTTTTCCCCC
                                                                                        poly A
2274  AAATCCCCTATATAATCGATCTTCTGCCCCCTTCCCTGTCACCCACCCAAACCGTTATTACGCCTAGAATGGTGAACAATAAAAAAATTAAAATAATCCA

2374  AGTTGTTTGCTTTCGTGTTCTTTCCAGCAGAATCGGTTTTAAGGATGGCAGCTTTTTTTCACAATCGGCATATTTGTGTGACGTCTTGATGCTTCTAATC

2474  GTCTTGACAAGCAAAGAAATGGATAAAATTCAACTCTACACG
```

Instead, a substitution rate of 7.5% (23/306), 5.6% (17/306) and 13.3% (41/306) at the first, second and third positions of the codons, respectively, was observed. Determination of the nucleotide substitution rate between a pseudogene and the functional gene must take into account the fact that both sequences have been accumulating nucleotide substitutions at different rates. In a functional gene there is very little selection pressure occurring at the third codon position while there is high selection pressure occurring at the first and second codon positions. In a pseudogene there is no selection pressure at any of the three codon positions. Nucleotide substitutions will accumulate randomly (with respect to the ancestral gene before the duplication event) at the third codon positions in both genes whereas nucleotide substitutions will accumulate randomly (with respect to the ancestral gene before the duplication event) at the first and second codon positions in the pseudogene only. Therefore, if inactivation accompanied duplication I would expect the nucleotide substitution rate at the third codon position to be approximately double that of the first or second codon positions. I observed a two-fold difference in the substitution rate at the third codon position with respect to the first and second codon positions.

The nucleotide substitution rate of the introns was also examined. Since most of the intron sequences are not under selective pressure, I would predict that the nucleotide substitution rate of the introns should be similar to that of the third codon position. Taking into account the six

nucleotides at the 5' and 3' ends of the intron which are highly conserved (Karn et al., 1983; Spieth et al., 1985b) the nucleotide substitution rate for the two introns was found to be 12.2% (26/212). The nucleotide substitutions occurred randomly regardless of "codon position". This substitution rate is similar to the substitution rate of 13.3% calculated for the third codon position. Based on these arguments and the structure of the hsp-2ps gene, I conclude that the hsp-2ps gene has been nonfunctional since the duplication event.

To estimate the time since the duplication event, the number of codons that could allow all four nucleotide substitutions without an amino acid change was determined. There are 157 of these codons shared between the hsp-1 and hsp-2ps genes 24 of which are substituted at the third position. A percent divergence of these codons was calculated to be 15.3%. This proportion was corrected for multiple hits by assuming a Poisson distribution of mutations ($-3/4 \ln[1-4/3f]$ where $f=15.3\%$). The corrected percent divergence is 17.1%. A rate of 1% divergence per million years at synonymous sites has been estimated for the hsp82 gene of Drosophila (Blackman and Meselson, 1986) as well as frogs, sea urchins, plants, mammals and bacteria (Ochman and Wilson, 1987 and references therein). If this is used then I estimate that the hsp-1 and hsp-2ps genes have a divergence time of approximately 17 million years. This suggests that the duplication event occurred approximately 8.5 million years ago. G. Poinar has estimated that the time of separation of C. elegans and C. briggsae, based on taxonomy,

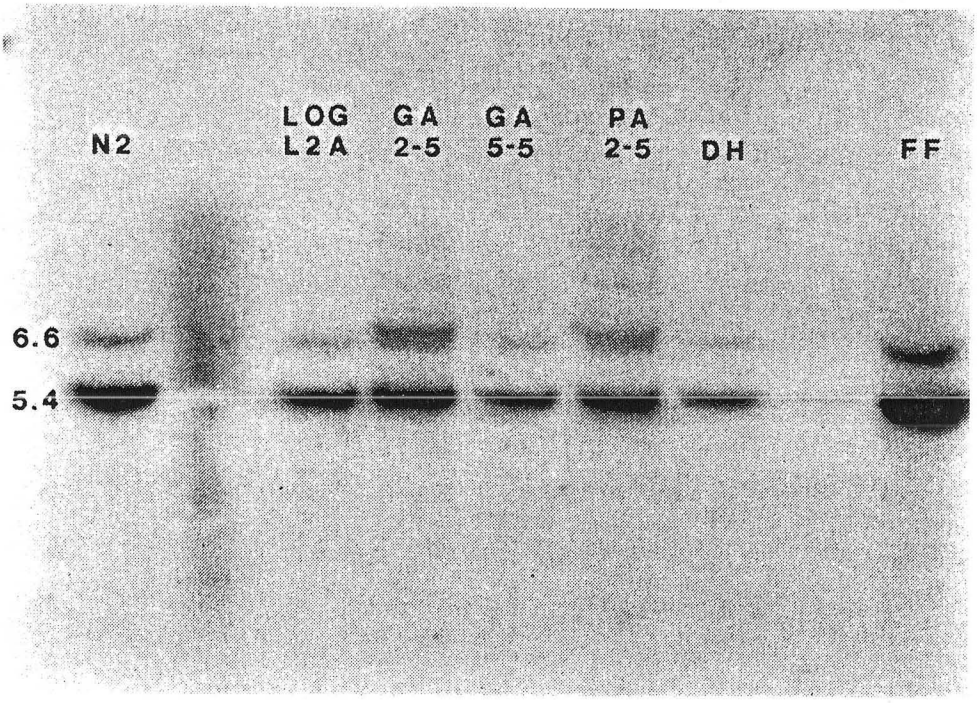
the fossil record and plate tectonics, occurred 20-40 million years ago (personal communication). My estimate of the time of duplication presented here falls well after the estimated point of species divergence and is consistent with the results of DNA hybridizations involving the hsp-1 and hsp-2ps genes to C. briggsae (Snutch, 1984).

Hybridization of the hsp-2ps gene to seven C. elegans strains collected from several geographical locations indicates that the hsp-2ps gene is present in all laboratory strains tested (Figure 2-3). Thus, it appears these strains have originated since the duplication event occurred.

Generation of the hsp-2ps gene

The hsp-2ps gene is unlike the vitellogenin pseudogene, vit-1 (Spieth et al., 1985a), and the msp pseudogenes (Ward et al., 1988; Klass et al., 1988). These pseudogenes are found within their respective multigene family clusters. The 5' flanking DNA corresponding to the regulatory region is also conserved between the pseudogenes and their functional counterparts. I propose two possible models for the origin of the hsp-2ps gene. 1) The hsp-2ps gene could have arisen through an RNA intermediate since it has only transcribed DNA sequences from the ancestral gene. It is known that in Drosophila, if the heat stress is great enough introns of the transcripts are not excised (Yost and Lindquist, 1986). Based on this, it is possible that the ancestral genome underwent a heat stress great enough to induce heat shock transcription of the hsp-1 gene without intron excision. The unprocessed pre-

Figure 2-3. Hybridization of hsp-2ps to different C. elegans strains. The hsp-2ps plasmid, pCes440, was hybridized to EcoRI digested genomic DNAs of various C. elegans laboratory strains. The strain designations are indicated on the autoradiograph. The 5.4 kb hsp-2ps band and the 6.6 kb hsp-1 band are indicated. The higher molecular weight bands represent partial digests. Hybridizations were done at 68°C in 5xSSPE, 0.3% SDS and 2.5xDenhardt's (Davis et al., 1980).


N2
LOG L2A
GA 2-5
GA 5-5
PA 2-5
DH
FF
6.6
5.4

mRNA would then be available to be reversed transcribed. Integration of the cDNA into the DNA of the X-chromosome would result in the duplication of the hsp-1 gene. 2) The hsp-2ps gene could have arisen by a DNA duplication event involving a transposition of a copy of the hsp-1(LGIV) DNA to LGX. A high amount of restriction fragment length polymorphisms have been observed around the hsp-1 gene between the Bristol and Bergerac strains (Snutch and Baillie, 1984). This suggests that the hsp-1 gene region is prone to nucleotide changes and DNA damage. As such, a stretch of DNA corresponding only to the hsp-1 gene may have been excised and then transposed to the X chromosome. This may also account for the large internal deletion seen in the hsp-2ps gene. Either model could explain the duplication of the hsp-1 gene. However, due to the lack of a proven reverse transcriptase activity in C. elegans, the simpler hypothesis is that the hsp-2ps gene arose as an unusual DNA-mediated transposition/duplication event of the hsp-1 gene.

CHAPTER 3

THE GRP78 GENE SUBFAMILY

INTRODUCTION

Proteins related to the stress inducible hsp70 proteins are normally found in unstressed cells and organisms. These proteins have been called hsc70 (heat shock cognates; Ingolia and Craig, 1982; Craig et al., 1983). Some of the functions of the hsp70-related proteins have been determined. Recently, one of the constitutively expressed hsc70 proteins has been identified as a clathrin-uncoating ATPase and appears to be involved in disrupting the protein-protein interactions of the clathrin triskelions (Chappell et al., 1986).

A second constitutively expressed protein has been identified as a glucose-regulated protein (grp; Munro and Pelham, 1986). Synthesis of the grp78 protein is enhanced in rat cells when the cells are deprived of glucose or stimulated with calcium ionophores (reviewed in Lee, 1987). The rat grp78 protein also appears to be slightly heat inducible in hamster fibroblasts (Attenello and Lee, 1984; Lin et al., 1986) and HeLa cells (Watowich and Morimoto, 1988). The rat grp78 protein is closely related to a wide variety of hsp70 proteins sharing 57-62% identity. In addition, the grp78 protein may be identical to the immunoglobulin heavy chain binding protein "BiP" (Munro and Pelham, 1986). The grp78 protein has a hydrophobic secretory leader sequence and is transported into the endoplasmic reticulum (ER) with subsequent removal of the hydrophobic leader sequence (Munro and Pelham, 1986). The carboxy terminal sequence, KDEL, is important for the retention of the grp78 protein in the lumen of the ER (Munro and Pelham,

1987). It has been suggested that the grp78 protein binds to abnormal, underglycosylated proteins in glucose starved cells. Under normal conditions, the grp78 protein may also associate with partially assembled protein complexes until assembly can be completed and the proteins secreted or incorporated into the membrane (Pelham, 1986).

The C. elegans hsp70C gene is constitutively expressed and non-heat inducible under the conditions tested. The hsp70C mRNA is found at maximum levels in the L1 larval stage and decreases to about 15% that level in the adult nematode. The hsp70D gene has low basal mRNA levels and heat induction of this gene results in the increased synthesis of mRNA 8 to 50 fold (Snutch et al., 1988).

In this chapter I present the sequence of the hsp70C gene and the 3' half of the hsp70D gene from C. elegans. I find that the hsp70C and hsp70D genes share a high degree of homology with the rat grp78 gene and appear to be the C. elegans grp78 equivalents. The hsp70D gene is a highly heat inducible member of the grp78 subfamily. Comparison of the coding region of the hsp70C gene to its homolog from a closely related species, C. briggsae, reveals a high degree of similarity. Likewise, analysis of the 5' regulatory region reveals that there are several stretches of nucleotide homology between C. elegans and C. briggsae. One of these blocks of homology is found in the rat grp78 regulatory region. I suggest that this conserved element may be involved in the transcriptional regulation of the glucose-regulated proteins.

MATERIALS AND METHODS

Construction of plasmids for sequencing

The phage containing the hsp70 genes (hsp70C-hs1 140 and hsp70D-hs1 B9 [Snutch, 1984; Snutch et al., 1988]; *C. briggsae* hsp70C-CB10 [T.P. Snutch, personal communication]) were digested with *Eco*RI or *Hin*dIII and subcloned into the plasmid vectors pUC19 (Norrander et al., 1983) or Bluescript$^+$ (Stratagene). The plasmids containing the hsp70 genes were identified (*C. elegans* hsp70C 5' end, pCes433 and pCes434, hsp70C 3' end, pCes428; hsp70D pCes403; *C. briggsae* hsp70C gene, pCbs1, and flanking DNA, pCbs2) and their orientations determined based on the pattern derived from restriction digests. The hsp70C(Cb) coding region was delineated by hybridizations back to the hsp70C(Ce) gene.

Overlapping plasmid deletions were made using either exonuclease III (Henikoff, 1987) or restriction enzymes.

Plasmid DNA preparation

Plasmid DNA for deletion and sequencing reactions was prepared using the mini alkali lysis method (Maniatis et al., 1982) except that two phenol:sevag extractions were done and after the first ethanol precipitation, the air dried pellet was resuspended in 0.25 M sodium acetate and reprecipitated with ethanol. After digestion with RNase, the plasmid DNA samples were precipitated with polyethylene glycol (Hattori and Sakaki, 1986).

DNA sequencing and sequence analysis

Dideoxy sequencing was performed on denatured plasmid DNA

(Chen and Seeburg, 1985; Sanger et al., 1980).

DNA sequences were analyzed using the computer program Microgenie (Beckman). Visual inspection and preparation of the DNA sequences for publication was done with the aid of the computer program ESEE (E. Cabot, personal communication).

Screening of genomic libraries

The lambda libraries were screened according to Benton and Davis (1977) and the phage purified on a 0.75 g cesium chloride per millimeter equilibrium gradient. Phage DNA was purified as described by Davis et al. (1980). Nick translations and hybridizations of the filters were done as described in Chapter 2.

RESULTS AND DISCUSSION

Characterization of the hsp70C gene from C. elegans

A restriction map of the C. elegans hsp70C gene [hsp70C(Ce)] is shown in Figure 3-1. The complete sequence of the hsp70C(Ce) gene and flanking DNA is presented in Figure 3-2. The hsp70C(Ce) gene contains three introns (Figure 3-1 and Figure 3-2) of 46, 238 and 104 nucleotides. The intron boundaries were determined by gaps and shifts in the amino acid sequence when compared to the hsp-1 gene sequence and predicted amino acid sequence. As well, similarity to the 5' and 3' intron splice sites, characterized by the consensus sequences AG/GTAAGT and TTTTCAG/G (Karn et al., 1983; Spieth et al., 1985b) was used to aid in the identification of the hsp70C intron boundaries. One of the intron positions was confirmed by sequencing over the appropriate region surrounding the third intron from a cDNA. None of the hsp70C introns were in the same position as the hsp-1 introns (Snutch et al., 1988).

There is a long 3' untranslated sequence of 623 nucleotides. Typically, in C. elegans, the 3' untranslated sequences are much shorter (approximately 150 nucleotides). I estimate the size of the mRNA to be approximately 2.7 kb based on the positioning of the polyA addition signal (Figure 3-2, block B) and an approximation of the transcriptional start site (see below). This is in good agreement with the 2.6 kb size predicted from Northern blots (Snutch, 1984; Snutch et al., 1988).

Within this trailer sequence is the septemer TTTTTTC

Figure 3-1. Restriction maps of the hsp70C genes from C. elegans [hsp70C(Ce)] and C. briggsae [hsp70C(Cb)] and the hsp70D gene from C. elegans. The genes are aligned to show regions of homology. The coding regions are shaded and the transcribed, untranslated regions are unshaded. Introns are shown by breaks in the coding region. B, BamHI; E, EcoRI; H, HindIII; K, KpnI; P, PstI; S, SalI; Sa, SacI; X, XbaI.

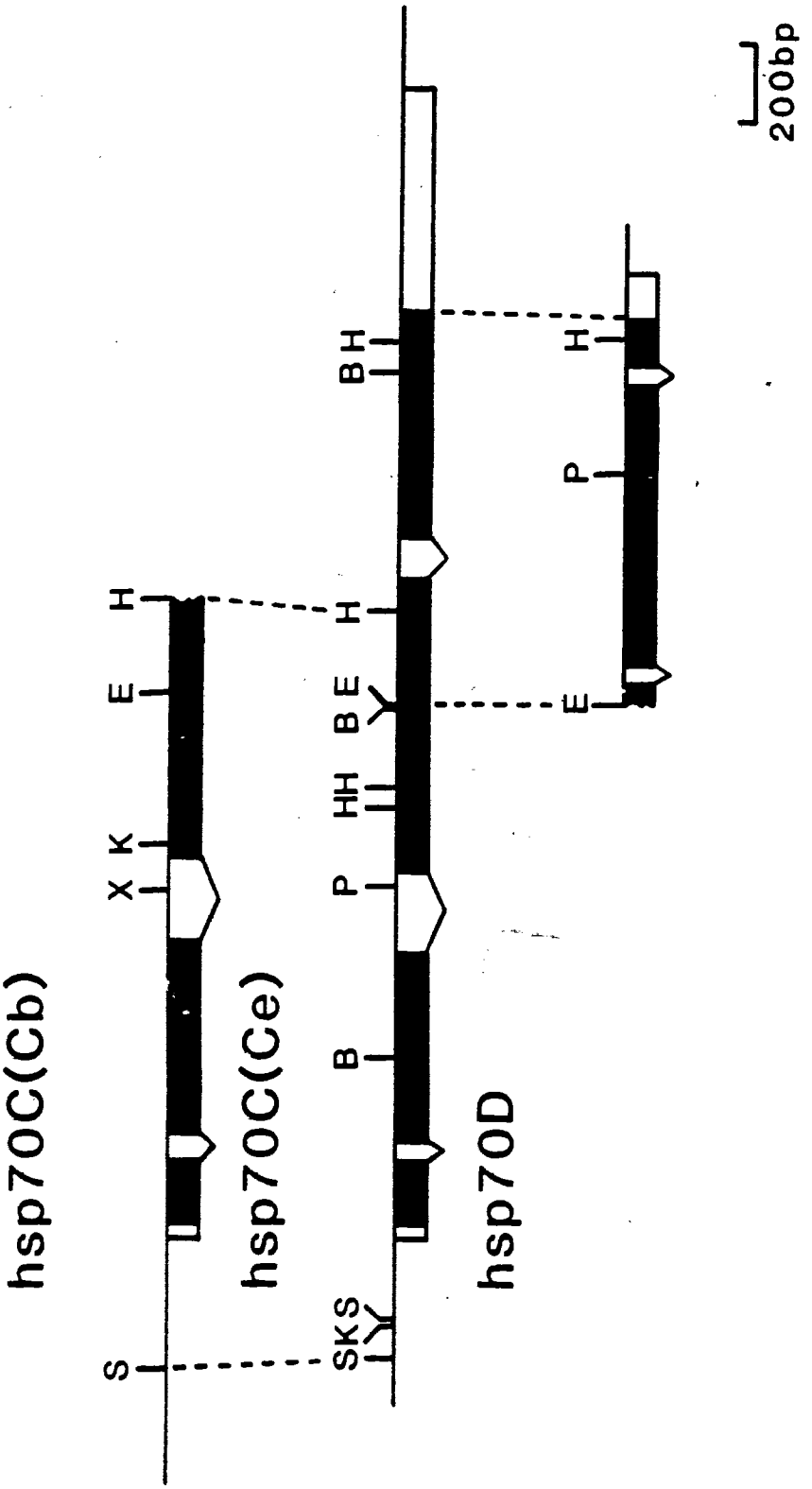
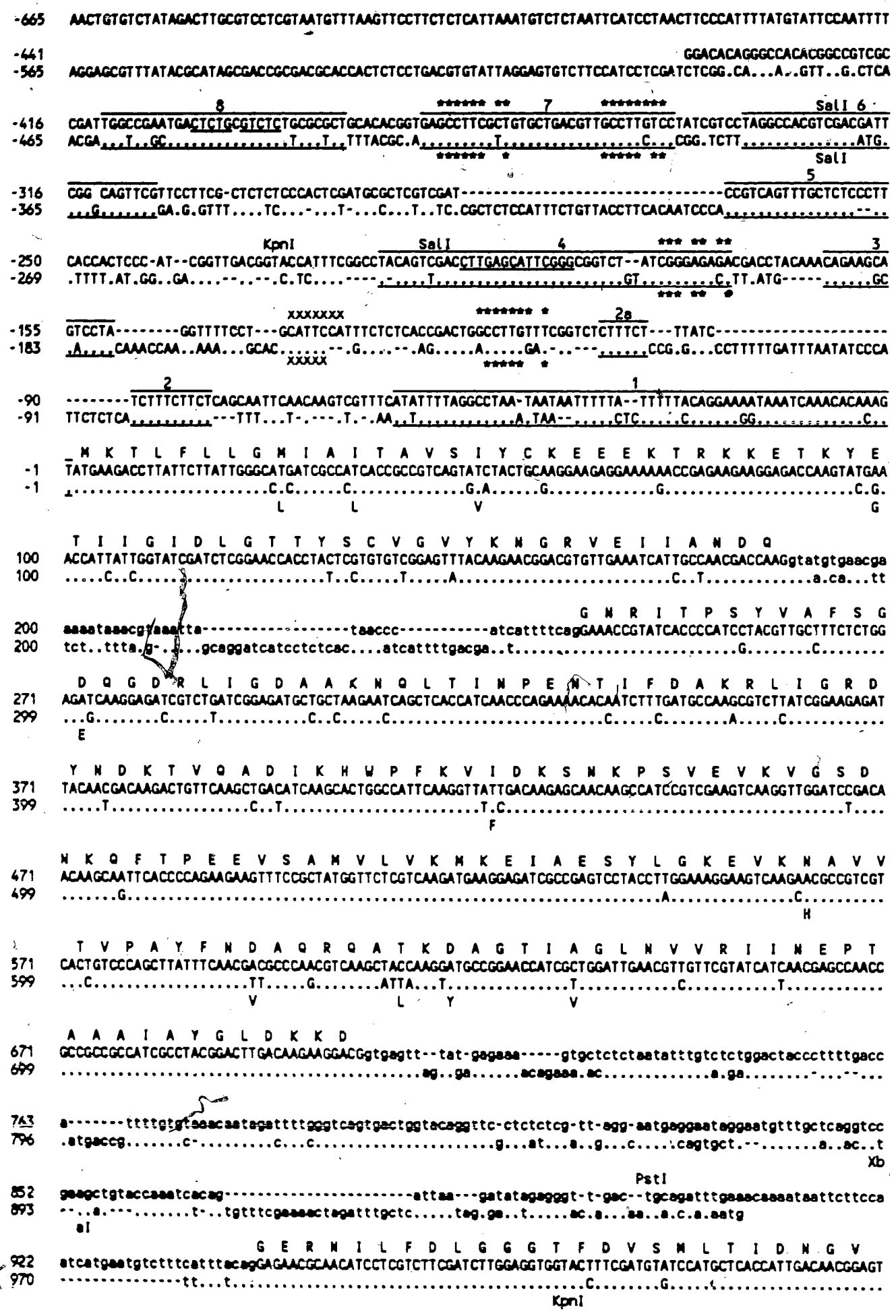
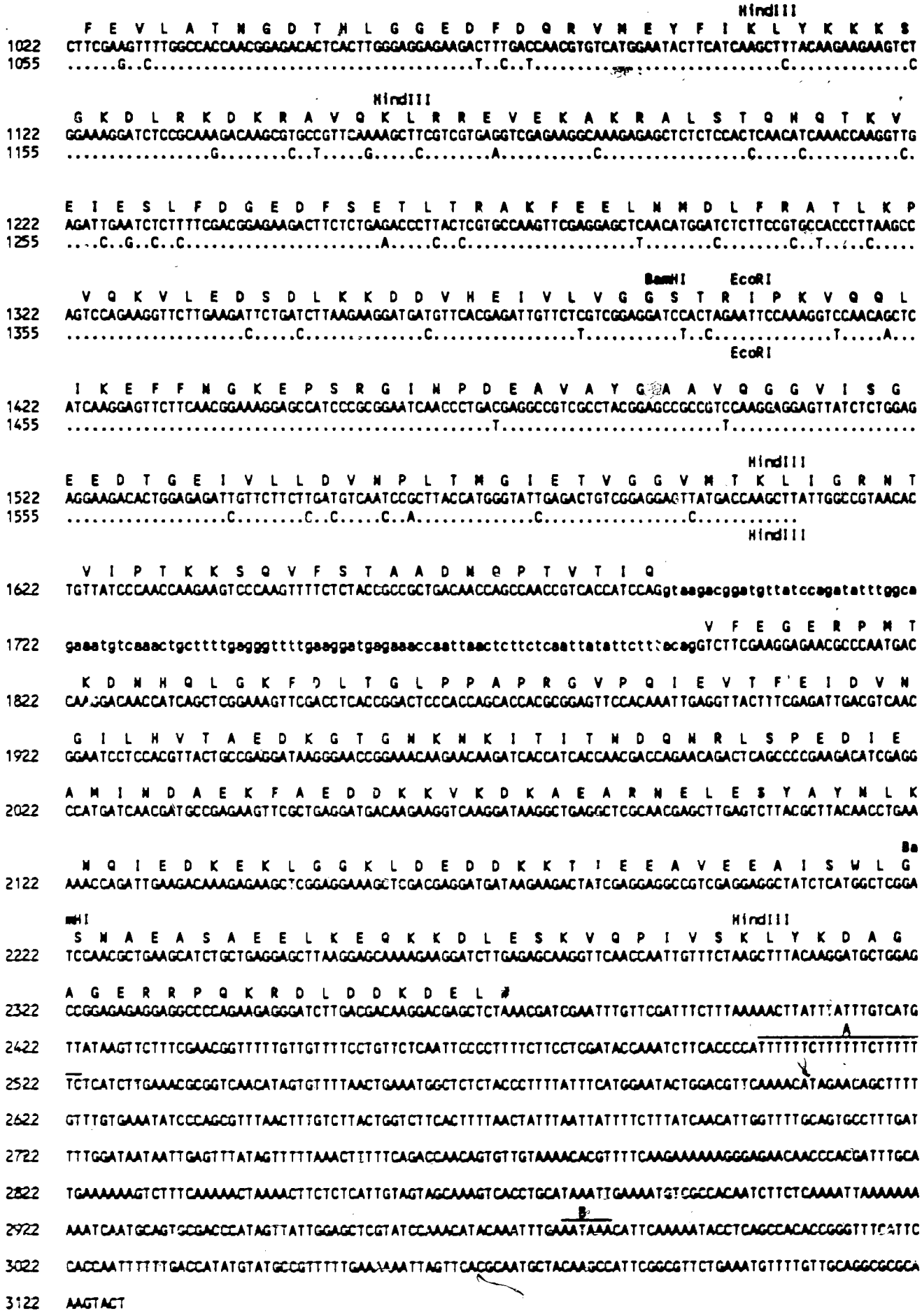

hsp70C(Cb)
hsp70C(Ce)
hsp70D
200bp
H
E
X K
S
B H
H
B E
H H
P
B
S K S
H
P
E

Figure 3-2. Nucleotide sequences of the hsp70C(Ce) and hsp70C(Cb) genes. The sequences are aligned for maximum homology. Introns are shown in lower case letters. Numbering is with respect to the start of translation. Dots (.) indicate a nucleotide match and gaps (-) in the sequence are shown to maintain maximum homology of the sequences. The amino acid sequences are shown above and below their respective genes with only the amino acid changes shown for the hsp70C(Cb) gene. Sequences homologous to the E1A core enhancer sequence ($^{A}_{C}$GGAAGTG$^{A}_{C}$; Hearing and Shenk, 1983) are indicated with a * and the SV40 core enhancer sequence (GTGG$^{AAA}_{TTT}$G; Weiher et al., 1983) are indicated with an x. Blocks 1 to 8 represent regions conserved between C. elegans and C. briggsae. Block 8 contains an inverted repeat and the region between block 8 and block 7 contains an alternating Pu/Py stretch thought to be involved in the formation of Z DNA. Block 5 is an imperfect repeat of block 6. Block 4 contains an HSE while blocks 2 and 2a represent regions that are TC-rich. Block 1 represents the region of transcription initiation and the transcribed/untranslated leader sequence. Within this region, at -23, is a 3' splice site which could serve as a trans splice site (Krause and Hirsh, 1987). Block A represents the three time repeated heptemer sequence TTTTTTC and block B represents the polyA addition signal.


-665 AACTGTGTCTATAGACTTGCGTCCTCGTAATGTTTAAGTTCCTTCTCTCATTAAATGTCTCTAATTCATCCTAACTTCCCATTTTATGTATTCCAATTTT
-441 GGACACAGGGCCACACGGCCGTCGC
-565 AGGAGCGTTTATACGCATAGCGACCGCGACGCACCACTCTCCTGACGTGTATTAGGAGTGTCTTCCATCCTCGATCTCGG.CA...A..GTT..G.CTCA
8
7
SalI 6
-416 CGATTGGCCGAATGACTCTGCGTCTCTGCGCGCTGCACACGGTGAGCCTTCGCTGTGCTGACGTTGCCTTGTCCTATCGTCCTAGGCCACGTCGACGATT
-465 ACGA...T..GC.....T...T..TTTACGC.A.....T.....C...CGG.TCTT.....ATG.
SalI
5
-316 CGG CAGTTCGTTCCTTCG-CTCTCTCCCACTCGATGCGCTCGTCGAT----------CCGTCAGTTTGCTCTCCCTT
-365 ...G.....GA.G.GTTT....TC...-...T-...C...T..TC.CGCTCTCCATTTCTGTTACCTTCACAATCCCA.....--..
KpnI
SalI
4
3
-250 CACCACTCCC-AT--CGGTTGACGGTACCATTTCGGCCTACAGTCGACCTTGAGCATTCGGGCGGTCT--ATCGGGAGAGACGACCTACAAACAGAAGCA
-269 .TTTT.AT.GG..GA...--..--.C.TC....----.-....T.....GT.....C.TT.ATG-----.....GC
xxxxxxx
2a
-155 GTCCTA--------GGTTTTCCT---GCATTCCATTTCTCTCACCGACTGGCCTTGTTTCGGTCTCTTTCT---TTATC----------
-183 .A....CAAACCAA..AAA...GCAC.....--.G....--.AG.....A.....GA.-..---.....CCG.G...CCTTTTTGATTTAATATCCCA
xxxxx
2
1
-90 --------TCTTTCTTCTCAGCAATTCAACAAGTCGTTTCATATTTTAGGCCTAA-TAATAATTTTTA--TTTTTACAGGAAAATAAATCAAACACAAAG
-91 TTCTCTCA.....---TTT...T-.---.T.-.AA..T.....A.TAA--.....CTC.....C.....GG.....C..
M K T L F L L G M I A I T A V S I Y C K E E E K T R K K E T K Y E
-1 TATGAAGACCTTATTCTTATTGGGCATGATCGCCATCACCGCCGTCAGTATCTACTGCAAGGAAGAGGAAAAAACCGAGAAGAAGGAGACCAAGTATGAA
-1 .....C.C.....C.....G.A.....G.....G.....C.G.
L L V G
T I I G I D L G T T Y S C V G V Y K N G R V E I I A N D Q
100 ACCATTATTGGTATCGATCTCGGAACCACCTACTCGTGTGTCGGAGTTTACAAGAACGGACGTGTTGAAATCATTGCCAACGACCAAGgtatgtgaacga
100 .....C..C.....T..C.....T.....A.....C..T.....a.ca...tt
G N R I T P S Y V A F S G
200 aaaataaacgtaaatta----------taaccc----------atcattttcagGAAACCGTATCACCCATCCTACGTTGCTTTCTCTGG
200 tct..ttta.g-.....gcaggatcatcctctcac....atcattttgacga..t.....G.....C.....
D Q G D R L I G D A A K N Q L T I N P E N T I F D A K R L I G R D
271 AGATCAAGGAGATCGTCTGATCGGAGATGCTGCTAAGAATCAGCTCACCATCAACCCAGAAAACACAATCTTTGATGCCAAGCGTCTTATCGGAAGAGAT
299 ...G.....C.....T.....C..C.....C.....C.....C.....A.....C.....
E
Y N D K T V Q A D I K H W P F K V I D K S N K P S V E V K V G S D
371 TACAACGACAAGACTGTTCAAGCTGACATCAAGCACTGGCCATTCAAGGTTATTGACAAGAGCAACAAGCCATCCGTCGAAGTCAAGGTTGGATCCGACA
399 .....T.....C..T.....T.C.....T....
F
N K Q F T P E E V S A M V L V K M K E I A E S Y L G K E V K N A V V
471 ACAAGCAATTCACCCCAGAAGAAGTTTCCGCTATGGTTCTCGTCAAGATGAAGGAGATCGCCGAGTCCTACCTTGGAAAGGAAGTCAAGAACGCCGTCGT
499 .....G.....A.....C.....
H
T V P A Y F N D A Q R Q A T K D A G T I A G L N V V R I I N E P T
571 CACTGTCCCAGCTTATTTCAACGACGCCCAACGTCAAGCTACCAAGGATGCCGGAACCATCGCTGGATTGAACGTTGTTCGTATCATCAACGAGCCAACC
599 ...C.....TT.....G.....ATTA...T.....T.....C.....T.....
V L Y V
A A A I A Y G L D K K D
671 GCCGCCGCCATCGCCTACGGACTTGACAAGAAGGACGgtgagtt--tat-gagaaa-----gtgctctctaatatttgtctctggactacccttttgacc
699 .....ag..ga.....acagaaa.ac.....a.ga.....-...--...
763 a-------ttttgtgtaaacaatagatttt gggtcagtgactggtacaggttc-ctctctcg-tt-agg-aatgaggaataggaatgtttgctcaggtcc
796 .atgaccg.....c-.....c...c.....g...at...a..g...c.....cagtgct.--.....a..ac..t
Xb
PstI
852 gaagctgtaccaaatcacag----------attaa---gatatagagggt-t-gac--tgcagatttgaaacaaaataattcttcca
893 --..a.---.....t-..tgtttcgaaaactagatttgctc.....tag.ga..t.....ac.a...aa..a.c.a.aatg
aI
G E R N I L F D L G G G T F D V S M L T I D N G V
922 atcatgaatgtctttcatttacagGAGAACGCAACATCCTCGTCTTCGATCTTGGAGGTGGTACTTTCGATGTATCCATGCTCACCATTGACAACGGAGT
970 ----------tt...t.....C.....G.....t.....
KpnI


HindIII
F E V L A T N G D T N L G G E D F D Q R V M E Y F I K L Y K K K S
1022 CTTCGAAGTTTTGGCCACCAACGGAGACACTCACTTGGGAGGAGAAGACTTTGACCAACGTGTCATGGAATACTTCATCAAGCTTTACAAGAAGAAGTCT
1055 ......G..C.....T..C..T.....C.....C
HindIII
G K D L R K D K R A V Q K L R R E V E K A K R A L S T Q M Q T K V
1122 GGAAAGGATCTCCGCAAAGACAAGCGTGCCGTTCAAAAGCTTCGTCGTGAGGTCGAGAAGGCAAAGAGAGCTCTCTCCACTCAACATCAAACCAAGGTTG
1155 .....G.....C..T.....G.....C.....A.....C.....C.....C.....C.
E I E S L F D G E D F S E T L T R A K F E E L M M D L F R A T L K P
1222 AGATTGAATCTCTTTTCGACGGAGAAGACTTCTCTGAGACCCTTACTCGTGCCAAGTTCGAGGAGCTCAACATGGATCTCTTCCGTGCCACCCTTAAGCC
1255 ...C..G..C..C.....A.....C..C.....T.....C.....C..T..C.....
BamHI EcoRI
V Q K V L E D S D L K K D D V H E I V L V G G S T R I P K V Q Q L
1322 AGTCCAGAAGGTTCTTGAAGATTCTGATCTTAAGAAGGATGATGTTCACGAGATTGTTCTCGTCGGAGGATCCACTAGAATTCCAAAGGTCCAACAGCTC
1355 .....C.....C.....C.....T.....T..C.....T.....A...
EcoRI
I K E F F M G K E P S R G I M P D E A V A Y G A A V Q G G V I S G
1422 ATCAAGGAGTTCTTCAACGGAAAGGAGCCATCCCGCGGAATCAACCCTGACGAGGCCGTCGCCTACGGAGCCGCCGTCCAAGGAGGAGTTATCTCTGGAG
1455 .....T.....T.....
HindIII
E E D T G E I V L L D V M P L T M G I E T V G G V M T K L I G R N T
1522 AGGAAGACACTGGAGAGATTGTTCTTCTTGATGTCAATCCGCTTACCATGGGTATTGAGACTGTCGGAGGAGTTATGACCAAGCTTATTGGCCGTAACAC
1555 .....C.....C..C.....C..A.....C.....C.....
HindIII
V I P T K K S Q V F S T A A D N Q P T V T I Q
1622 TGTTATCCCAACCAAGAAGTCCCAAGTTTTCTCTACCGCCGCTGACAACCAGCCAACCGTCACCATCCAGgtaagacggatgttatccagatatttggca
V F E G E R P M T
1722 gaaatgtcaaactgcttttgagggttttgaaggatgagaaaccaattaactcttctcaattatattcttttcagGTCTTCGAAGGAGAACGCCCAATGAC
K D M H Q L G K F D L T G L P P A P R G V P Q I E V T F E I D V N
1822 CAAGGACAACCATCAGCTCGGAAAGTTCGACCTCACCGGACTCCCACCAGCACCACGCGGAGTTCCACAAATTGAGGTTACTTTCGAGATTGACGTCAAC
G I L H V T A E D K G T G N K N K I T I T N D Q N R L S P E D I E
1922 GGAATCCTCCACGTTACTGCCGAGGATAAGGGAACCGGAAACAAGAACAAGATCACCATCACCAACGACCAGAACAGACTCAGCCCCGAAGACATCGAGG
A M I N D A E K F A E D D K K V K D K A E A R N E L E S Y A Y N L K
2022 CCATGATCAACGATGCCGAGAAGTTCGCTGAGGATGACAAGAAGGTCAAGGATAAGGCTGAGGCTCGCAACGAGCTTGAGTCTTACGCTTACAACCTGAA
Ba
N Q I E D K E K L G G K L D E D D K K T I E E A V E E A I S W L G
2122 AAACCAGATTGAAGACAAAGAGAAGCTCGGAGGAAAGCTCGACGAGGATGATAAGAAGACTATCGAGGAGGCCGTCGAGGAGGCTATCTCATGGCTCGGA
mHI HindIII
S M A E A S A E E L K E Q K K D L E S K V Q P I V S K L Y K D A G
2222 TCCAACGCTGAAGCATCTGCTGAGGAGCTTAAGGAGCAAAAGAAGGATCTTGAGAGCAAGGTTCAACCAATTGTTTCTAAGCTTTACAAGGATGCTGGAG
A G E R R P Q K R D L D D K D E L #
2322 CCGGAGAGAGGAGGCCCCAGAAGAGGGATCTTGACGACAAGGACGAGCTCTAAACGATCGAATTTGTTCGATTTCTTTAAAAACTTATTTATTTGTCATG
2422 TTATAAGTTCTTTCGAACGGTTTTTGTTGTTTTCCTGTTCTCAATTCCCCTTTTCTTCCTCGATACCAAATCTTCACCCCATTTTTTTCTTTTTTCTTTTT
2522 TCTCATCTTGAAACGCGGTCAACATAGTGTTTTAACTGAAATGGCTCTCTACCCTTTTATTTCATGGAATACTGGACGTTCAAAACATAGAACAGCTTTT
2622 GTTTGTGAAATATCCCAGCGTTTAACTTTGTCTTACTGGTCTTCACTTTTAACTATTTAATTATTTTCTTTATCAACATTGGTTTTGCAGTGCCTTTGAT
2722 TTTGGATAATAATTGAGTTTATAGTTTTTAAACTTTTTCAGACCAACAGTGTTGTAAAACACGTTTTCAAGAAAAAAGGGAGAACAACCCACGATTTGCA
2822 TGAAAAAAGTCTTTCAAAAACTAAAACTTCTCTCATTGTAGTAGCAAAGTCACCTGCATAAATTGAAAATGTCGCCACAATCTTCTCAAAATTAAAAAAA
2922 AAATCAATGCAGTGCGACCCATAGTTATTGGAGCTCGTATCCAAACATACAAATTTGAAATAAACATTCAAAAATACCTCAGCCACACCGGGTTTCATTC
3022 CACCAATTTTTTGACCATATGTATGCCGTTTTTGAAAAATTAGTTCACGCAATGCTACAAGCCATTCGGCGTTCTGAAATGTTTTGTTGCAGGCGCGCA
3122 AAGTACT

repeated three times tandemly (Figure 3-2, block A). This septemer was searched for in other sequences. The sequence of the C. elegans transposable element Tc1 (Rosenzweig et al., 1983) was examined and the TTTTTTC sequence was found several times within the terminal arms of the Tc1 element but not in the open reading frame. A hexamer sequence of similar nature (TTTTTC) is tandemly repeated four times in the 3' untranslated regions of the Trypanosoma cruzi hsp83-like genes (Dragon et al., 1987). The significance of these repeats, however, is not known.

The predicted hsp70C(Ce) amino acid sequence was compared to the predicted hsp-1 amino acid sequence (Figure 3-3) and grp78 from rat (Munro and Pelham, 1986) in order to identify the hsp70C gene product. This comparison revealed a striking degree of similarity with the rat grp78 amino acid sequence (77%). This is in contrast to the degree of identity shared with hsp70A (59%) and the heat inducible hsp70 (55%). The hsp70C carboxy terminus has the characteristic sequence KDEL found at the rat grp78 carboxy terminus. This sequence is required for retention of the protein in the ER (Munro and Pelham, 1987). Overall, the last 70 amino acids, excluding the last four, exhibit marginal conservation between hsp70C and grp78. This has been observed with other members of the hsp70 family. The amino terminal segment when compared to hsp70A is much longer (Figure 3-3). The leader sequence has features characteristic of secretory leader sequences. It contains a positively charged amino terminus followed by a hydrophobic

Figure 3-3. Comparison of the amino acid sequences of hsp70C(Ce), hsp70C(Cb), hsp70D, hsp-1 (Chapter 1; Snutch et al. 1988) and the rat grp78 (Munro and Pelham, 1986). Amino acid matches are indicated by a dot (.) and amino acid gaps are shown with a dash (-).

```
hsp70C(Ce)  MKTLFLLGMIAITAVSIYCKEEEKTRKKETKYET-IIGIDLGTTYSCVGVYKNGRVEIIANDQGNRITPSYVAFSGDQGDRLIGDAAKNQ  89
hsp70C(Cb)  ........L..L....V.............G.-.....................................E.....................  89
grp78       ..FTVAAALLLLC..RA---...DKKEDVGT---VV.....F.....TPE-.E.,,.....                               81
hsp70A                                    MSKMMAV.....FMM.K.....T.....T-.T-E.....                       59

hsp70C(Ce)  LTIMPENTIFDAKRLIGRDYNDKTVQADIKMWPFKVIDKS-NKPSVEVKVGS-DNKQFTPEEVSAMVLVKMKEIAESYLGKEVKNAVVTV  177
hsp70C(Cb)  ..................................F...-.....-.....................................H.....  177
grp78       ..S.....V.........TW..PS..Q...FL....VE.K-T..YIQ.DI.GGQT.T.A...I.....T....T..A....K.TH.....  178
hsp70A      VAM..H..V.........KFD.PA..S.M........SAEGA..K.Q.EYKG-E..I.....I.S...L...KT..AF.EPT..D.....  148

hsp70C(Ce)  PAYFNDAQRQATKDAGTIAGLNVVRIINEPTAAAIAYGLDKKD-GERNILVFDLGGGTFDVSMLTIDNGVFEVLATNGDTHLGGEDFDQR  266
hsp70C(Cb)  .....V.....L.Y....V.....................-.....................................................  266
grp78       .........................M.....................RE-..K.....L.........V.....................  267
hsp70A      .T....S.........A......L.....................GH....V.I...........I...ED.I...KS.A...........N.  238

hsp70C(Ce)  VMEYFIKLYKKKSGKDLRKDKRAVQKLRREVEKAKRALSTQHQTKVEIESLFDGEDFSETLTRAKFEELNMDLFRATLKPVQKVLEDSDL  356
hsp70C(Cb)  ..........................................................................................  356
grp78       ...H........T...V...N.....................S...ARI....F.E.....................S.M.............  357
hsp70A      MVNH.CAEF.R.HK..ASNP..LRR..TAC.R.NET..SSC.ASI..D...E.I..YTNI...R....CA....S.MD..E.S.R.AKM  328

hsp70C(Ce)  KKDDVHEIVLVGGSTRIPKVQQLIKEFFNGKEPSRGINPDEAVAYGAAVQGGVISGE-EDT-GEIVLLDVNPLTMGIETVGGVMTKLIGR  444
hsp70C(Cb)  ..........................................................-...-.........................  441
hsp70D                        .........DY.....................A...G.V-.N.-.DV.........L.................  74
grp78       ..S.ID.............I...V.........................A..L..D-Q..-.DL.....C...L.............P.  445
hsp70A      D.SQ..D.............K.LSDL.S...LNKS......L........AAIL..DKSEAVQDLL....A..SL....A.....A..K.  418

hsp70C(Ce)  NTVIPTKKSQVFSTAADNQPTVTIQVFEGERPMTKDNHQLGKFDLTGLPPAPRGVPQIEVTFEIDVNGILHVTAEDKGTGNKNKITITND  534
hsp70D      ............P....S.SA.S.VIY......VM...K..N..V..I.....................S...........L.....  164
grp78       ...V......I....S........K.Y.....L.....L..T.....I.....................R...................  535
hsp70A      ..T....TA.T.T.YS....G.L...Y....A.....NL....E.S.I.............D..A....N.S.T..S..KAKQ......  508

hsp70C(Ce)  QNRLSPEDIEAMINDAEKFAEDDKKVKDKAEARNELESYAYNLKNQIEDKEKLGGKLDEDDKKTIEEAVEEAISWLGSNAEASAEELKEQ  624
hsp70D      H.........R.....D...A..QAQ.E.V.S.....A...QI.T..A.........TDE..VS..S...R..E.....QD..T..N...  254
grp78       ....T..E..R.V........E...L.ERIDT.........S.....G.........SPE..E.M.K....K.E..E.HQD.DI.DF.AK  625
hsp70A      KD.F.KD...R.V.E...YKA..EAQ..RIG.K.G.....F...QT...-...KD.ISPE...K..DKCD.ILK..D..QT.EK..FEH.  597

hsp70C(Ce)  KKDLESKVQPIVSKLYKDAGAGERRPQKRDLDDKDEL                                                      661
hsp70D      ..E...V.........-S..GQGEQASEEPSE.H...                                                      290
grp78       ..E..EI....I....GSG.PPPT-GEE-.TSE....                                                      660
hsp70A      Q....GLAN..I....QS..GAPPGAAPGGAAGGAGGPTIEEVD                                               641
```

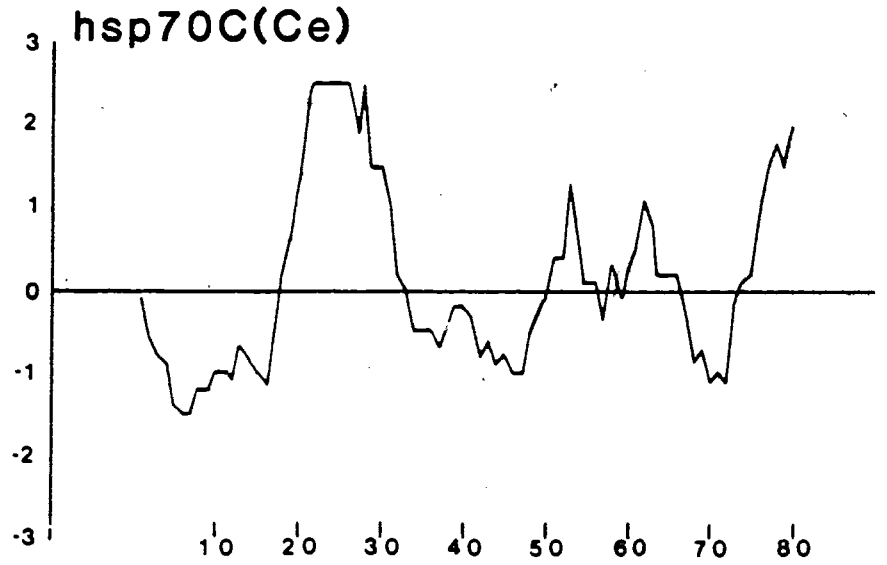
central region (Figure 3-3 and Figure 3-4) followed by two small nonpolar residues with a single residue between (Figure 3-3). Cleavage probably occurs between I17 and Y18 after the two nonpolar residues. Although the structural features may be conserved between the hsp70C and grp78 proteins, the amino acids in the leader sequence are not conserved (Figure 3-3 and Figure 3-4). The size of the hsp70C protein before cleavage has been estimated to consist of 661 amino acids with a molecular weight of 73,339 daltons. After cleavage, the hsp70C protein would have a predicted molecular weight of 71,579 daltons. The hsp70C protein does not have any glycosylation sites (Asp-X-Ser/Thr) in agreement with the rat grp78 results (Munro and Pelham, 1986).

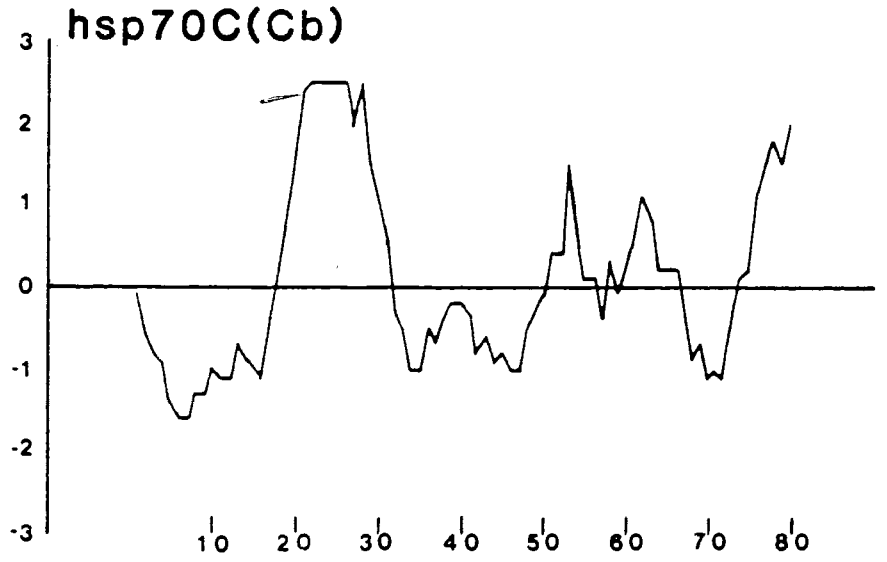
The high degree of similarity shared with the rat grp78 protein in both structural characteristics and amino acid sequence has led me to conclude that the hsp70C gene is probably the grp78 gene equivalent in C. elegans.

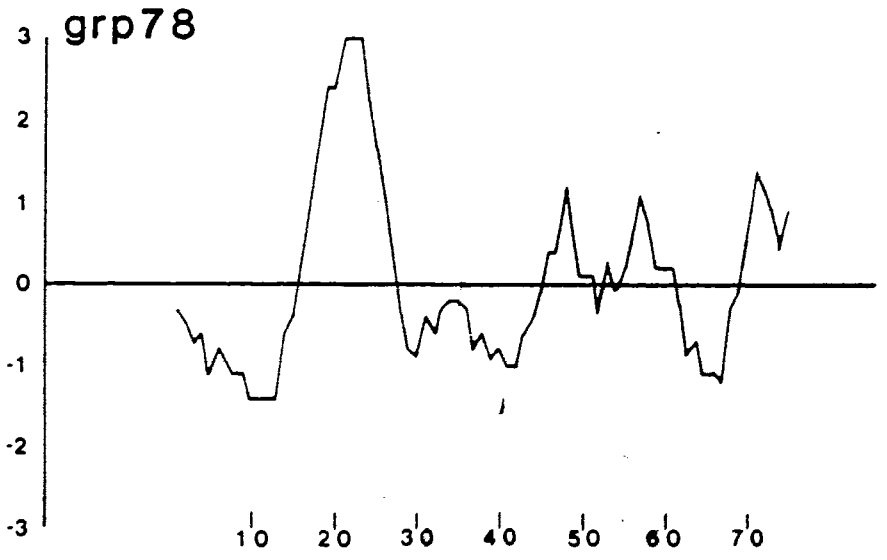
Characterization of the hsp70C gene from a closely related species, C. briggsae

It has been suggested that only sequences important for the survival of the nematode would be conserved between C. elegans and a closely related species C. briggsae (Snutch, 1984; Prasad, 1988). Sequences important for gene regulation and sequences representing genes have been identified using DNA sequence comparisons (Blackman and Meselson, 1986; Fischer and Maniatis, 1986; Henikoff and Eghtedarzadeh, 1987). Based on this, I decided to compare the extent of similarities of the

Figure 3-4. A hydropathy plot of the first 80 amino acids of the hsp70C(Ce) and hsp70C(Cb) proteins and the corresponding region of the grp78 protein. The hydropathy of the protein sequence was determined using Microgenie (Beckman).


hsp70C(Ce)
3
2
1
0
-1
-2
-3
1'0
2'0
3'0
4'0
5'0
6'0
7'0
8'0


hsp70C(Cb)
3
2
1
0
-1
-2
-3
1'0
2'0
3'0
4'0
5'0
6'0
7'0
8'0


grp78
3
2
1
0
-1
-2
-3
1'0
2'0
3'0
4'0
5'0
6'0
7'0

C. elegans and C. briggsae hsp70C genes in 1) the coding regions, 2) 5' flanking regions and 3) the introns of the two species. Two phage of one type containing the C. briggsae hsp70C gene [hsp70C(Cb)] were isolated from a partial EcoRI genomic library in Charon 4 (Snutch, 1984). Only the first two-thirds of the hsp70C(Cb) gene was represented in the phage isolated. These data are summarized in Figure 3-1. The sequence of the hsp70C(Cb) gene and 5' flanking DNA is presented in Figure 3-2 and is aligned against the hsp70C(Ce) gene for maximum homology.

Comparison of the hsp70C genes from C. elegans and C. briggsae

The hsp70C coding region appears to be highly conserved between the two species. The two hsp70C genes share 93% homology at the nucleotide level (Figure 3-2) and 98% homology at the amino acid level (Figure 3-3). The hsp70C(Cb) protein also has the same hydrophobic leader sequence as the hsp70C(Ce) protein (Figure 3-3 and Figure 3-4). An estimate of the time of divergence between C. elegans and C. briggsae was made by determining the percent divergence of the third position of the codons that can accept all four nucleotide substitutions without changing the amino acid. Of 203 such codons, 44 or 21.6% have substitutions at the third position. However, this figure must be corrected to take into account multiple mutation events. Assuming a Poisson distribution of mutations ($-3/4 \ln[1-4/3f]$ where f=21.6%), the corrected percent divergence is 25.5%. An estimate of 1% divergence per million years has been calculated for the hsp82 gene in Drosophila

(Blackman and Meselson, 1986) as well as other organisms (Ochman and Wilson, 1987). Based on this, I predict that C. elegans and C. briggsae have been diverging for approximately 25.5 million years. This would imply the last common ancestor that C. elegans and C. briggsae shared was approximately 13 million years ago. This is less than an estimate of 20-40 million years ago based on taxonomy, the fossil record and plate tectonics (G. Poinar, personal communication). My estimate probably underestimates the time of divergence of C. elegans and C. briggsae because 1) there is some bias in the use of individual codons but the codon usage is similar between these two species and 2) it has been suggested that there are functional constraints placed on the secondary structure of the hsp70 mRNAs or at the DNA level (Hunt and Morimoto, 1985).

The hsp70C 5' regulatory regions share strong similarities with the rat grp78 regulatory region

The 5' regulatory regions were searched for sequences known to act as regulatory elements in other genes. I detected one copy of an HSE (Pelham, 1982) from -205 to -195 in C. elegans and -230 to -217 in C. briggsae (Figure 3-2). This would suggest that the hsp70C gene should be heat inducible although no increase in the mRNA concentration has been detected under the conditions tested (Snutch et al., 1988). It may be that the hsp70C gene is transiently expressed for only a short time after heat shock. The presence of an HSE is consistent with the observation that the grp78 genes are slightly heat inducible (Attenello and Lee, 1984; Lin et al., 1986).

Several sequences similar to the enhancer core sequences of E1A (Hearing and Shenk, 1983) and SV40 (Weiher et al., 1983) are also found in the 5' regulatory region of these two genes (Figure 3-2). Similarities to E1A and SV40 core enhancer sequences are also observed in the rat grp78 regulatory region (Lin et al., 1986). There appears to be a TATA box from -58 to -50 and a CCAAT box from -78 to -74 (Figure 3-2).

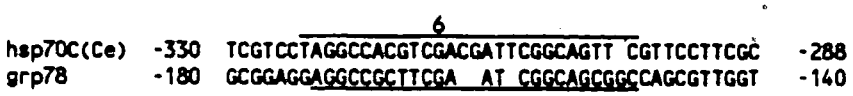
The 5' regulatory regions of hsp70C(Cb) and hsp70C(Ce) were further compared visually. Allowing for insertions and deletions, eight distinct blocks of homology were observed and are outlined and numbered 1 to 8 in Figure 3-2. The fourth block corresponds to the HSE and surrounding nucleotides. The first block is adjacent to the hsp70C translated region and probably represents the region surrounding the TATA box, the transcription start site and the 5' transcribed/untranslated leader sequence. Within this block, at -23, is a 3' splice site which may be involved in a <u>trans</u> splicing reaction (Krause and Hirsh, 1987). The fourth and seventh blocks contain sequences similar to core enhancer sequences from E1A and SV40. A pyrimidine rich region is represented by the second block and is repeated further upstream. The fifth block is an imperfect repeat of the sixth block. Two inverted sequences are represented by the eighth block. Between the seventh and eighth blocks is a region of alternating purines and pyrimidines, these sequences are thought to be involved in the formation of Z DNA. In the rat grp78 regulatory region, an alternating Pu/Py region is seen and divides two tandem repeats

of the promoter/enhancer sequences (Lin et al., 1986). However, this repeated promoter motif is not seen in the regulatory regions of the Caenorhabditis sequences.

My results are similar to those seen in a comparison of the hsp82 genes of several Drosophila sp. (Blackman and Meselson, 1986). Several blocks of nucleotides were conserved between four closely related Drosophila species. Some of these blocks have known regulatory functions. Others are conserved but no function has yet been assigned to them (Blackman and Meselson, 1986).

It has been reported that the 5' regulatory region of the rat grp78 gene has an enhancer-like activity (Lin et al., 1986). I reasoned that if the HSEs are highly conserved between evolutionarily distant species (Pelham, 1982, 1985) then, if the grp78 enhancer-like activity is important, the "glucose responsive element" should be conserved between Caenorhabditis and rat. Seven of the eight blocks of homology, excluding the 5' untranslated region, were compared to the rat grp78 regulatory region (Lin et al., 1986) for any regions of identity. Disregarding homologies to the viral enhancers, the sixth block shares identity with a sequence in the rat grp78 regulatory region (Figure 3-5). Within this block there is 80% identity. This block of identity in the rat is located within the restriction fragment reported to contain the enhancer-like activity (Lin et al. 1986; Chang et al., 1987). I believe that this element is important for the regulation of the grp78 genes because of the high degree of conservation within this block.

Figure 3-5. A comparison of the regulatory regions of the hsp70C(Ce) gene and the rat grp78 gene (Lin et al., 1986). Numbering in hsp70C(Ce) is with respect to the start of translation. Numbering of the rat grp78 gene begins at the start of transcription (Lin et al., 1987). Block 6 represents the stretch of nucleotide identity shared between the hsp70C regulatory region and the rat grp78 regulatory region.


6
hsp70C(Ce) -330 TCGTCCTAGGCCACGTCGACGATTCGGCAGTT CGTTCCTTCGC -288
grp78 -180 GCGGAGGAGGCCGCTTCGA AT CGGCAGCGGCCAGCGTTGGT -140

Recently, nuclease protection studies using the rat grp78 genehave revealed that this block of identity is protected by a protein (E. Resendez Jr., S.K. Wooden and A.S. Lee, manuscript submitted). My observations of the conservation of blocks of sequence in the 5' region suggest that putative regulatory elements could be detected through a comparison of sequence data obtained from C. elegans and C. briggsae.

Intron sequences are highly conserved between C. elegans and C. briggsae

The two introns that are represented in the hsp70C(Cb) gene sequence are found at corresponding positions when compared to the first two introns of the hsp70C(Ce) gene sequence. The second introns of the hsp70C(Cb) gene and the hsp70C(Ce) gene are approximately the same size (243 vs. 238 nucleotides respectively) while the first introns are not (74 vs. 46 nucleotides respectively). The similar sizes of the second intron facilitated a comparison of the two introns. It was expected that the introns, having little or no known important functional sequences, would not be similar. To my surprise, there were several long blocks of homology present between the two introns. The significance of these sequence homologies is not known.

A heat inducible grp78-like gene

One of the heat inducible genes previously described, hsp70D, is represented only by the 3' half (Snutch, 1984; Snutch et al., 1988). The hsp70D EcoRI fragment was used to screen both a partial EcoRI genomic library in Charon 4 (Snutch

et al., 1988) and a partial MboII genomic library in EMBL4 (a gift of M. Krause). Using high stringency hybridization conditions (Snutch et al., 1988), several positives were isolated. Characterization of these phage revealed a restriction pattern identical to the hsp70C(Ce) gene and not the hsp70D gene. Since the hsp70D gene hybridized strongly to the hsp70C gene sequences I felt that the hsp70D gene might be a heat inducible variety of the hsp70C gene. The sequence of the hsp70D gene is shown in Figure 3-6 and corresponds to the last half of the hsp70C gene from a shared EcoRI site (Figure 3-1).

Two introns are found in the 3' half of the hsp70D gene. The intron positions are not conserved with respect to the hsp70C gene (Figure 3-1). In addition, the 3' untranslated region is not as long as the 3' untranslated region of the hsp70C gene.

Analysis of the predicted amino acid sequence reveals a high degree of identity between the hsp70D and hsp70C genes (76%; Figure 3-3). The hsp70D gene shares 71% identity with the rat grp78 amino acid sequence. When compared to the hsp-1 amino acid sequence, an identity of 54% was calculated. Like hsp70C, no glycosylation sites were detected in the last half of the hsp70D protein.

The predicted carboxy terminal residues of the hsp70D polypeptide has the sequence HDEL. This is similar to the KDEL sequence found at the carboxy terminus of the predicted hsp70C and rat grp78 polypeptides. Although it is not identical,

Figure 3-6. Sequence of the 3' half of the hsp70D gene. Intron sequences are in lower case letters. Amino acid sequence of the hsp70D gene is shown. The polyA addition signal is shown.

EcoRI

```
      I  P  K  V  Q  Q  L  I  K  D  Y  F  N  G  K  E  P  S  R
1     GAATTCCAAAGGTTCAACAACTTATCAAGGACTACTTCAACGGCAAGGAGCCATCTCGTGgtaagagtaaaaattaactttttatatttctaagtttggta

        G  I  N  P  D  E  A  V  A  Y  G  A  A  V  Q  A  G  V  I  G  G  V  E  N  T  G  D  V  V  L  L  D
101   tttcagGAATCAACCCTGACGAAGCAGTTGCTTACGGAGCTGCTGTTCAAGCAGGAGTAATCGGAGGTGTCGAAAATACCGGAGATGTTGTCCTTCTTGA

        V  N  P  L  T  L  G  I  E  T  V  G  G  V  N  T  K  L  I  G  R  N  T  V  I  P  T  K  K  S  Q  V  F
201   CGTCAATCCACTCACTCTGGGTATTGAGACTGTCGGAGGAGTCATGACTAAGTTGATTGGAAGAAATACCGTTATCCCAACAAAGAAGAGCCAAGTTTTT

        P  T  A  A  D  S  Q  S  A  V  S  I  V  I  Y  E  G  E  R  P  N  V  N  D  N  H  K  L  G  N  F  D  V
301   CCTACCGCTGCCGATAGTCAGAGCGCTGTATCAATCGTTATTTATGAAGGTGAACGACCAATGGTTATGGACAATCACAAGCTCGGAAACTTCGATGTCA

        T  G  I  P  P  A  P  R  G  V  P  Q  I  E  V  T  F  E  I  D  V  N  G  I  L  H  V  S  A  E  D  K  G  T
401   CCGGAATCCCACCAGCTCCACGCGGTGTTCCACAAATTGAGGTTACATTCGAAATTGATGTCAACGGAATATTGCACGTAAGCGCCGAAGACAAGGGAAC

        G  N  K  N  K  L  T  I  T  N  D  H  N  R  L  S  P  E  D  I  E  R  N  I  N  D  A  D  K  F  A  A  D
501   TGGAAACAAGAACAAGCTCACAATCACCAACGACCACAATCGTCTCAGTCCAGAAGACATCGAGCGCATGATCAATGATGCTGACAAGTTTGCTGCAGAT

        D  Q  A  Q  K  E  K  V  E  S  R  N  E  L  E  A  Y  A  Y  Q  I  K  T  Q  I  A  D  K  E  K  L  G  G
601   GATCAAGCCCAAAAAGAGAAGGTTGAGTCACGAAATGAGCTGGAAGCATATGCCTATCAGATCAAAACTCAAATCGCCGACAAGGAAAAACTCGGAGGAA

        K  L  T  D  E  D  K  V  S  I  E  S  A  V  E  R  A  I  E  W  L  G  S  N  Q  D  A  S  T  E  E  N  K  E
701   AGCTCACTGACGAAGACAAGGTGTCCATCGAGTCTGCCGTCGAGAGAGCTATTGAGTGGCTCGGAAGCAACCAAGATGCCTCTACTGAAGAAAACAAGGA

        Q  K  K                                                                E  L  E  S  V  V  Q  P  I  V  S  K  L  Y
801   GCAAAAGAAGgtttgttttcatcaattttctaaacgaaaatataatatttgtttttcagGAACTTGAATCTGTTGTTCAACCAATCGTCTCCAAACTTTA

                          HindIII
        S  A  G  G  Q  G  E  Q  A  S  E  E  P  S  E  D  H  D  E  L  #
901   CTCGGCGGGAGGACAAGGAGAACAAGCTTCTGAGGAGCCATCGGAGGATCATGATGAACTGTAAAATATTAATTGCCTTCAACTACTTGCTGCCATCGGG
                                                                                                    polyA
1001  TTCCTGTTTATTGCTTTATCCCCCCTTCATTCTTTCCCAACCCAACCCCACCATTCATTTTTACGAGTTTAGTGAAATTTTTTATTTAATAAATTATTTA

1101  TTATTTATCATGATATCCTGTATATTTGAGTTGAACATACAGAACAAGCTAGATCGATCAGTTTTGGGTATTAAAAAAGTTAATCATGATTTTCATATGG

1201  CCTTATCAGAAGTTCTTTCAAGATCAAAA
```

histidine is a conservative substitution of lysine and may be able to function in a similar manner. It is possible that although the hsp70D gene is heat inducible the sequence HDEL may signal retention of this polypeptide in the ER or ER-derived structures but only under heat shock or stress conditions. The constitutively expressed hsp70C and the heat inducible hsp70D members of the grp78 family may interact and complement each other, functioning together or separately as the situation dictates.

The hsp70C(Ce) gene has been tentatively mapped within approximately 150 kb of cosmid DNA (A. Coulson and J. Sulston, personal communication). Using this DNA, it should now be possible to find some restriction fragment length differences and map the hsp70C(Ce) gene to a specific chromosomal region. Assignment of the hsp70D gene to cosmids and eventually a chromosomal region awaits the isolation of phage DNA containing the hsp70D gene and flanking DNA. Once the C. elegans grp78 genes have been localized to a specific region of a chromosome, it should be possible to generate strains mutant for the grp78 genes.

Possible mutant phenotypes of the grp78 genes

Mutations of the hsp70C and hsp70D genes may be difficult to isolate; however, there are several different scenarios possible. First, the grp78 proteins, being located in the ER, are probably essential proteins. If there is an interaction between the two grp78 genes as seen in the SSA and SSB families of S. cerevisiae (Craig and Jacobsen, 1984; Werner-Washburne et

al., 1987) then no lethal mutation could be recovered for at least the hsp70C and possibly the hsp70D gene. Since the hsp70D gene is highly heat inducible a mutation which renders the hsp70D gene product nonfunctional might be sensitive to heat stress i.e. might not survive a prolonged heat stress or take longer to recover from a heat stress. Second, mutations in the hsp70C and hsp70D genes may be recovered by looking for temperature-sensitive mutations that are defective in secreting proteins (for example, cuticle proteins). Third, the grp78 genes are induced when glucose sources are depleted. In the case of C. elegans, this would be when the food source is depleted. This is followed by the appearance of a specialized larval stage, the dauer larvae, in which the larvae become resistant to most environmental conditions and the external openings are sealed until the larvae encounter favorable nutritional environments. It may be that the hsp70C and hsp70D genes are involved in the decision to become dauer larvae. A decrease in food would lead to a decrease in glucose and enhance the synthesis of the grp78 proteins triggering dauer larva formation. In this scenario, mutations of the grp78 genes may be recovered as dauer defective genes or daf. Over-production of either protein may force the nematode to enter the dauer larva life cycle even though food is abundant. Under-production of either protein may prevent the nematode from entering the dauer life cycle. A second possibility is that the grp78 genes, once turned on could not be turned off forcing the nematode to stay a dauer larva.

CHAPTER 4

A dnaK-LIKE GENE, hsp70F

INTRODUCTION

The hsp70 proteins have been highly conserved throughout evolution. The Escherichia coli dnaK gene product is one of fourteen polypeptides whose synthesis is enhanced upon heat shock. Induction of the heat shock response is under control of the htpR gene (Neidhardt and VanBogelen, 1981). The dnaK protein was previously identified as the B66.0 protein (Georgopoulos et al., 1982) and is an abundant protein under normal growth conditions. The abundance of the dnaK protein increases as the temperature and growth rate increases. Isolation of dnaK mutants revealed that the dnaK protein is necessary for λ DNA replication (Georgopoulos and Herskowitz, 1971; Sunshine et al., 1977; Georgopoulos, 1977; Saito and Uchida, 1977) and probably interacts with the P protein of λ (Georgopoulos and Herskowitz, 1971; Georgopoulos, 1977). However, the role of dnaK in bacterial functions is not as well defined. When the temperature-sensitive mutant dnaK756 is present DNA synthesis is inhibited after a shift to non-permissive temperatures (Saito and Uchida, 1977). Further analysis showed that RNA synthesis is inhibited more rapidly than DNA synthesis after the temperature shift (Itikawa and Ryu, 1979). The lack of dnaK alleles that are not temperature-sensitive suggests that the dnaK gene produces a protein essential for cell viability.

The E. coli hsp70 equivalent, the dnaK gene, when compared to a human hsp70 gene (Hunt and Morimoto, 1985) and a Drosophila hsp70 gene (Craig et al., 1982) shares 47% and 48%

identity at the amino acid level, respectively, and 50% and 57% identity at the nucleotide level, respectively. However, the *Drosophila hsp70* gene and the human *hsp70* gene share 73% identity at the amino acid level and 72% identity at the nucleotide level (Hunt and Morimoto, 1985).

The *C. elegans* hsp70F gene is transcribed under non-stress conditions and upon heat shock, the synthesis of the hsp70F mRNA is enhanced two-fold (Snutch, 1984; Snutch et al., 1988). In this chapter, I describe the sequencing and features of the hsp70F gene. The hsp70F gene product shares a significant amount of homology with the *dnaK* gene product but only moderate amounts of identity with previously identified eukaryotic hsp70 proteins. The hsp70F protein has a leader sequence characteristic of proteins imported into the mitochondrial matrix.

MATERIALS AND METHODS

Construction of plasmids for sequencing

The phage containing the hsp70F gene, hsl B4 (Snutch, 1984; Snutch et al., 1988) was digested with EcoRI and subcloned into the plasmid vector pUC19 (Norrander et al., 1983). The plasmid containing the hsp70F gene, pCes405, was identified and the orientation determined based on the pattern derived from restriction digests and hybridizations to the hsp-1 gene.

Overlapping plasmid deletions were made using either exonuclease III (Henikoff, 1987) or restriction enzymes.

Plasmid DNA preparation

Plasmid DNA for deletion and sequencing reactions was prepared using the mini alkali lysis method (Maniatis et al., 1982) except that two phenol:sevag extractions were done and after the first ethanol precipitation, the air dried pellet was resuspended in 0.25 M sodium acetate and reprecipitated with ethanol. After digestion with RNase, the plasmid DNA samples were precipitated with polyethylene glycol (Hattori and Sakaki, 1986).

DNA sequencing and sequence analysis

Dideoxy sequencing was performed on denatured plasmid DNA (Chen and Seeburg, 1985; Sanger et al., 1980).

DNA sequences were analyzed using the computer program Microgenie (Beckman). Visual inspection and preparation of the DNA sequences for publication was done with the aid of the computer program ESEE (E. Cabot, personal communication).

RESULTS AND DISCUSSION

Characterization of the hsp70F gene

A restriction map of the 2.9 kb EcoRI restriction fragment containing the hsp70F gene is shown in Figure 4-1. The region of the hsp70F gene sequenced represents the first two-thirds of an hsp70 gene based on comparison to the hsp-1 gene. The sequence of the partial hsp70F gene is presented in Figure 4-2. Within the hsp70F coding region there are two introns of 66 and 49 nucleotides. These two introns are not in the same position as the introns of the other three C. elegans hsp70 genes described in this thesis. In the 5' flanking region is an HSE (Pelham, 1982, 1985) from -316 to -303 (Figure 4-2). This HSE is also part of a two dimer 10 base pair sequence NTTCNNGAAN required for full heat inducibility of the heat shock inducible genes (Xiao and Lis, 1988). The HSE is flanked by the heptamer sequence TTTTTTC (Chapter 3). There are several more copies of this sequence as well as several degenerate copies of this sequence upstream of the coding region. Downstream of the HSE is a region centered around -278 to -272 which could function as a TATA box. Approximately 30 nucleotides downstream of this region is a sequence that shares identity with a transcription initiation/capsite sequence (derived from an analysis of the C. elegans major sperm protein gene family, CATAATCTTCA where A is the probable site of transcription initiation [Klass et al., 1988]) from -246 to -236 and -216 to -206. There is a 3' splice site at -6 which is a feature of the trans splicing event (Krause and Hirsh, 1987).

Figure 4-1. Restriction map of the hsp70F gene. The coding region is shaded and the transcribed, untranslated region is unshaded. Introns are shown by breaks in the coding region. B-BamHI; E-EcoRI; H-HindIII; O-XhoI; P-PstI; Sa-SacI.

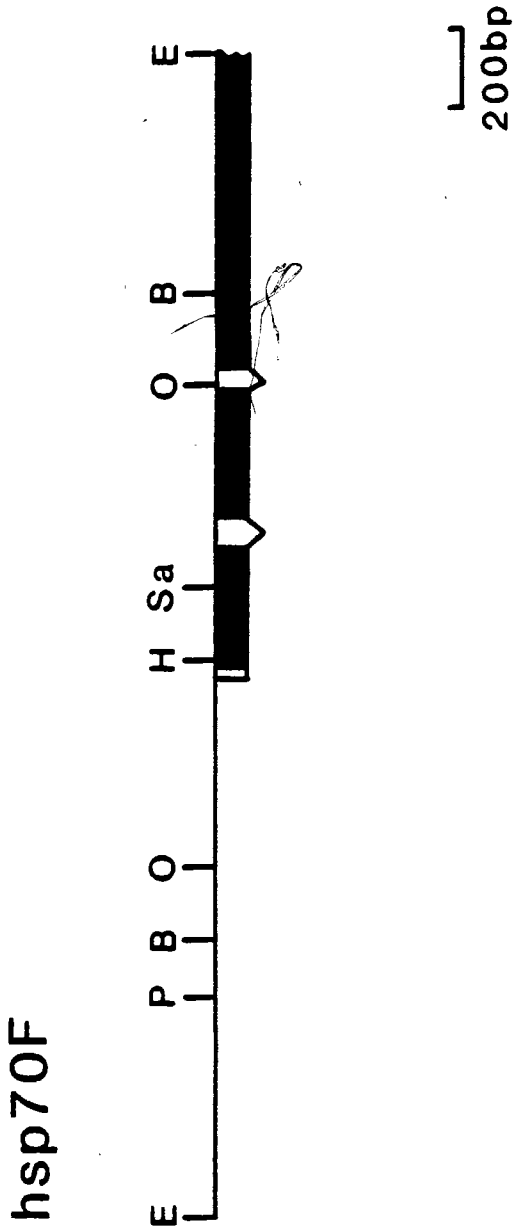

hsp70F
E
P
B
O
H
Sa
O
B
E
200bp

Figure 4-2. Nucleotide sequence of the hsp70F gene. Introns are shown in lowercase letters. Numbering is with respect to the start of translation. A putative TATA box is shown as well as potential transcription start sites (v). A dimer of the HSE (Xiao and Lis, 1988) is indicated.

```
        EcoRI
-1307   GAATTCTTGCATCAATTCCCGTCGTTTTGCAATAAGTTCCGACGCGGAAAGATGCGAAAAATCTTCCCGATGTGGCTCATCCTCTTCTCCTTCTTCTTCT

-1207   TCATCACCGCTTTTTTCTTCTTCGTCTTCCTTAACTTCTTTTTCTCGTGTCGGACGAATCAGTGTTCCATCTTTTAGCTTAATCGGCAGAAGTTCTTCAT

-1107   GGCCTTCTTCGAGTAGTCCTGCGAATTTTCTCTTTTTCTTCTCAACATCATCTTCATCGTCCTGTCTGATTCTACACCATTCTTTCTGCAATTTAACGTT

-1007   ATTAAGTTTTATTAACTAATTTTCAAACAACAATTACTTTTCTCAAGCCGCCTTTTTGCGTTGGCAAATGCGCTGTGTCCCTTCGATGTCGTCGTCCATC

-907    ATGTCCAATGCGAAACTTCATCACTATTCTCATAATGAGTCTTGTCTGTCGGCGATGAAATCATCTTCAATCTGAAAATTATTGGATAAATTGTTTTATT

                                                    PstI
-807    TTCGAAAAATTGCTCACATTAACAGCTCTCTGCCTCAAATCTGCAGATTTTCTTGATTTAATATCTGCCATTCCATCGCGAATCTCTCTCGACAGCTTTT

                                                                                                     BamHI
-707    TCCGCTTTGCAAGGCGATTTAAATGTTTCGTAGTTTTTCTTGTTTTATTTGTCTTCAGCATTTTCAGCTTCTCTTCGCGAGAAGCGGAGACAGGATCCCA

-607    TGTTTTTCTGCAAAATAATCCATTGATTTAACACCTCGTAAATAATTTAAAAAAGAGTTAAATTTAATTGCAACCCTATTTGTAAAAAGAAAACTCATTT

-507    TCGCCAAAAATAAAGCAAAAATAATTCAAGAGAAAAACGCGCCGGTGTTGCGATTGGGGCGTAACTGCAATGTGTGCGCACACAATCTCAACAAGCGCTG

-407    CGAGACCCGCCGCCTGACCGTAATGTGAAATGGGCGGAGACGAGAAGTTTTTTTCTGTTTGAAAGTTGATGCAAAAGCCCGTGATTCTTTTTTTCGAGAA

        HSE                         TATA                                V            V
-307    ATTTCTCGAGTTTTTTCCAACGAAAAATTCATTAAATTTAAACCTTTTAGCTCTCCTTTCCAATCTTTTGCATCATTATTCTCCTAAACTTGGCATATTC

-207    AGTGGAAATGATGCAAAATGACCCTGACTTTTGTTATCAAAAATAACAAGAAAATTGTCCCGTTTAACGGTTGAAAAGCAAATTTTGTGTCATTTTGTTT

-107    AGGAAATGTCAAAATAAGCTCAAAAACCGATTACAAATTATATTTTACTGCTTTTTATCCTATTTTCTCGCGTTTTCGTTCATGAATGCAATTTTCTTTC

                                                                                            HindIII
                M  L  S  A  R  S  F  L  S  S  A  R  T  I  A  R  S  S  L  M  S  A  R  S  L  S  D  K  P  K  G
-7      AGGCACTATGCTTTCCGCACGATCATTCTTGTCTTCCGCTCGCACAATTGCTCGATCGAGCTTGATGAGCGCCAGAAGCTTGTCGGACAAACCAAAGGGA

         H  V  I  G  I  D  L  G  T  T  N  S  C  V  S  I  M  E  G  K  T  P  K  V  I  E  N  A  E  G  V  R  T
94      CATGTTATCGGAATTGATCTTGGAACCACAAACTCCTGTGTCAGTATCATGGAAGGAAAGACTCCAAAAGTTATCGAGAACGCAGAAGGAGTCAGAACCA

                                                  SacI
        T  P  S  T  V  A  F  T  A  D  G  E  R  L  V  G  A  P  A  K  R  Q  A  V  T  N  S  A  N  T  L  F  A  T
194     CTCCATCGACGGTTGCTTTCACTGCTGACGGTGAGCGTCTTGTTGGAGCTCCAGCCAAAAGACAGGCCGTTACCAACTCTGCCAATACTCTTTTCGCCAC

          K  R  L  I  G  R  R  Y  E  D  P  E  V  Q  K  D  L
294     AAAGAGATTGATCGGAAGAAGATACGAAGATCCAGAGGTTCAAAAGGACTTgtaaggactcttttcactaaatttcttttaaaaggactcttttcact

                             K  V  V  P  Y  K  I  V  K  A  S  N  G  D  A  W  V  E  A  Q  G  K  E  Y  P  P  S
394     aaattttcattttttcagAAAGGTCGTTCCATACAAGATTGTCAAAGCCAGCAACGGAGACGCGTGGGTTGAGGCTCAAGGAAAAGAGTATCCCCCATCTC

        Q  V  G  A  F  V  L  M  K  M  K  E  T  A  E  S  Y  L  G  T  T  V  N  N  P  V  V  T  V  P  A  Y  F  N
494     AGGTTGGAGCATTCGTTCTGATGAAGATGAAGGAAACTGCCGAAAGCTATTTGGGAACCACCGTCAACAACCCCGTTGTTACAGTTCCAGCTTACTTCAA

          D  S  Q  R  Q  A  T  K  D  A  G  Q  I  S  G  L  N  V  L  R  V  I  N  E  P  T  A  A  A  L  A  Y  G
594     CGATTCTCAGCGTCAAGCTACTAAGGATGCTGGACAAATCTCTGGTCTTAACGTTCTTCGTGTCATCAACGAGCCAACTGCTGCCGCTCTCGCCTATGGA

         L  D  K  D  A  G  D  K  I                                                  I  A  V  Y  D  L  G  G
694     TTGGATAAGGACGCTGGAGATAAGATgtaggctaagcgctcgagtaatacttttcacaatatatattttttttagCATCGCTGTCTACGATCTTGGAGGTG

        G  T  F  D  V  S  I  L  E  I  Q  K  G  V  F  E  V  K  S  T  N  G  D  T  F  L  G  G  E  D  F  D  H  A
794     GTACTTTCGATGTGTCAATTCTTGAAATCCAAAAGGGCGTCTTCGAGGTCAAGTCCACCAACGGAGATACATTCCTCGGAGGAGAAGACTTCGATCACGC

                                                             BamHI
          L  V  H  H  L  V  G  E  F  K  K  E  Q  G  V  D  L  T  K  D  P  Q  A  M  Q  R  L  R  E  A  A  E  K
894     TCTCGTCCATCACCTCGTTGGAGAGTTCAAGAAGGAGCAAGGAGTTGATCTTACCAAGGATCCACAGGCCATGCAGAGACTTCGTGAAGCCGCCGAGAAG

          A  K  C  E  L  S  S  T  T  Q  T  D  I  N  L  P  Y  I  T  M  D  Q  S  G  P  K  H  L  N  L  K  L  T
994     GCCAAGTGCGAACTTTCATCCACCACCCAGACCGACATTAATCTTCCATACATCACCATGGATCAATCTGGACCAAAACATCTTAACTTGAAGCTCACCA

        R  A  K  F  E  Q  I  V  G  D  L  I  K  R  T  I  E  P  C  R  N  V  L  H  D  A  E  V  K  S  S  Q  I  A
1094    GAGCCAAGTTCGAGCAGATTGTCGGAGATCTCATCAAGAGAACCATTGAGCCATGCCGTAACGTCCTTCACGACGCTGAAGTCAAGTCCTCCCAAATCGC

          D  V  L  L  V  G  G  M  S  R  M  P  K  V  Q  A  T  V  Q  E  I  F  G  K  V  P  S  K  A  V  N  P  D
1194    CGATGTTCTTCTCGTAGGAGGAATGAGCAGAATGCCAAAGGTGCAAGCCACTGTTCAAGAAATCTTCGGAAAAGTTCCATCAAAGGCTGTCAACCCAGAC
```

```
      E  A  V  A  M  G  A  A  I  Q  G  A  V  L  A  G  D  V  T  D  V  L  L  L  D  V  T  P  L  S  L  G  I
1294  GAGGCCGTCGCCATGGGAGCTGCTATTCAAGGAGCCGTCTTGGCCGGAGACGTCACCGATGTTCTTCTCCTTGATGTCACTCCACTTTCCCTTGGTATCG

      E  T  L  G  G  I  M  T  K  L  I  T  R  M  T  T  I  P  T  K  K  S  Q  V  F  S  T  A  A  D  G  Q  T  Q
1394  AGACTCTCGGAGGAATTATGACCAAACTCATCACCAGAAACACTACCATCCCAACTAAGAAGTCCCAAGTCTTCTCTACCGCCGCCGATGGACAGACTCA

                                                                                                  EcoRI
      V  Q  I  K  V  F  Q  G  E  R  E  M  A  T  S  N  K  L  L  G  Q  F  S  L  V  G  I
1494  AGTGCAGATCAAGGTGTTCCAAGGAGAACGTGAAATGGCCACCAGCAACAAACTTCTCGGTCAATTCTCGCTTGTCGGAATTC
```

Comparison of the hsp70F gene to other hsp70-like genes

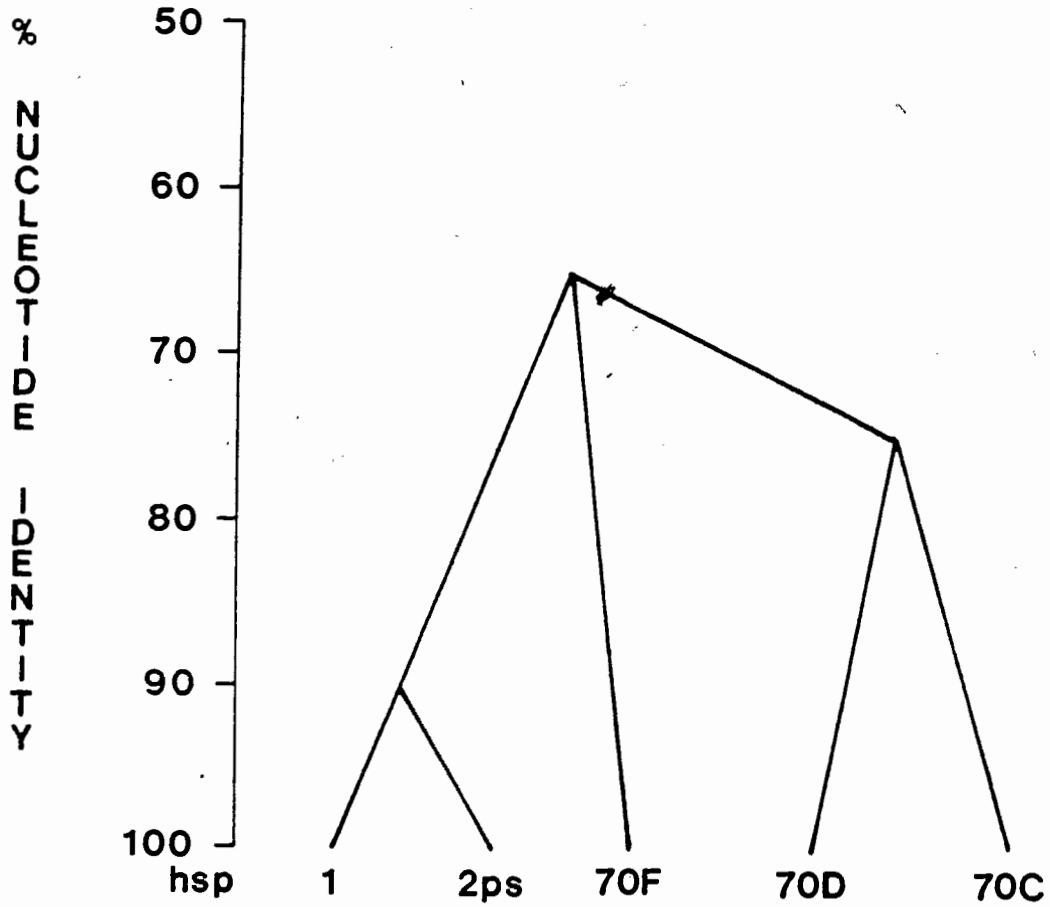
To determine the identity of the hsp70F gene, a comparison of the hsp70F amino acid sequence to several other known hsp70 amino acid sequences was made (Figure 4-3). The hsp70F gene shares limited amino acid identity with hsp70A (53%; Chapter 1; Snutch et al., 1988) and the hsp70C gene product (53%; Chapter3) from C. elegans, the SSB1 gene product from S. cerevisiae (43%; M. Slater and E.A. Craig), and the Hsc1 (48%), Hsc2 (43%; Craig et al., 1983) and a heat inducible hsp70 (52%; Ingolia et al., 1980) gene products from D. melanogaster. Surprisingly, a comparison of the hsp70F protein to the predicted dnaK protein (Bardwell and Craig, 1984) revealed a 67% identity. The hsp70 proteins listed above, over the same region covered by the hsp70F gene, share approximately 48-54% identity at the amino acid level with the dnaK protein and 61-86% identity at the amino acid level amongst themselves. Therefore, I conclude that the hsp70F gene is more like the dnaK gene from E. coli than any eukaryotic hsp70 gene identified to date.

At the nucleotide level, the hsp70F gene shares 63-65% identity with the other C. elegans hsp70 genes (Figure 4-4). This is in comparison to the 75% identity shared among the two grp78-like genes, hsp70C and hsp70D. In a comparison to the dnaK gene, hsp70F shares 65% identity at the nucleotide level while the remaining functional hsp70 genes share 55-57% identity at the nucleotide level. The shared identity between the dnaK gene and the C. elegans hsp70 genes is slightly higher

Figure 4-3. Comparison of the hsp70F amino acid sequence to the protein products of dnaK, hsp-1 (hsp70A), hsp70C, SSB1, Hsc1, Hsc2 and hsp70. Matches are indicated with a dot (.) and gaps with a dash (-).

```
hsp70F   MLSARSFLSSARTIARSSLNSARSLSDKPKGHVIGIDLGTTNSCVSI-MEGKTPKVIENAEGVRTTPSTVAFTAD-GE   76
dnaK                                    M.KI.....A.-.D.T..R.L.....D.....II.Y.Q.-..   47
hsp70A                                  MS.HNAV.....Y...GVF.H..VE-I.A.DQ.N.....Y....-.-T.   48
hsp70C   MKTLFLLGMIAITAVSIYCKEEEKTRKKETKYETI.....Y...GVYKN.RVE-I.A.DQ.N.I...Y...SG.Q.D   79
SSB1                                  MAEGVFQGA.....Y...AT-Y.SSVE-I.A.EQ.N.V...F....PE--.   50
hsc1                                    M..LPAV.....Y...GVFQN..VE-I.A.DQ.N.....Y....-E-S.   48
hsc2                                    MG.IPA.....Y...GV-WQNSKVEI.A.DQ.N.....Y...N-E-T.   48

hsp70F   RLVGAPAKRQAVTNSANTLFATKRLIGRRYEDPEVQKDLKVVPYKIV-KASNGD-AW-VEAQ--G--KEYPPSQVGAFVL  149
dnaK     T...Q.....PQ....,I.....FQ.E...R.VSIM.F..I-A.D...-..-..VK--.-Q.MA..-.IS.E..  120
hsp70A   ..I.DA..N.VAM.PH..V.DA.....KFD..A..S.M.MV.F.VI-S.EGAK-PK-.QVEYK.EN.IFT.EEISSN..  125
hsp70C   ..L.DA..N.LTI.PE..I.DA.....D.N.KT..A.I.MV.F.VID.SNKPSVEVK.GSD--N--.QFT.EE.S.M..  155
SSB1     ..I.DA..N..AL.PR..V.DA.....FD.ES..N.M.TV.F.VIDVDG.PVIEV--QYLEET--.TFS.QEIS.M..  126
hsc1     ..I.DA..N.VAM.PN..I.DA.....FD.AT..S.M.MV.FEAF--.G..K-PR-IR                       104
hsc2     ..I.D...N.VAM.AK..V.DA.....KFD..KI.E...LV.F.VI--NEK.K-PK-I.                       104

hsp70F   MKMKETAESYLGTTVNNPVVTVPAYFNDSQRQATKDAGQISGLNVLRVINEPTAAALAYGLD--KDA-GDKIIAVYDLGG  226
dnaK     K...K...D...EP.TEA.I.....A.....R.A..E.K.I.....--.GT-.NRT.....  197
hsp70A   L...K...AF.EP..KDA.....T.....A.A.....I.....I.....--.KGM.ERNVLIF....  203
hsp70C   V....I.....KE.K.A.....A.....T.A....V.I.....I.....--.KD-.ERN.L.F....  232
SSB1     T....I..AKI.KK.EKA.I.....A.....A.....I.....I....GAG.SEKERNVLIF.....  206

hsp70F   GTFDVSILEI--QKG--VFEVKSTNGDTFLGGEDFDHALVHHLVGEFKKEQGVDLTKDPQAMQRLREAAEKAKCELSSTT  302
dnaK     ....I..I..DEVD.EKT...LA.....H.....SR.INY..E....D..I..RN..L.....K.....I....AQ  277
hsp70A   .....T.--ED.--I.....A...H.....NRM.N.FCA...RKHKK..ASN.R.LR...T.C.R.NET...SC  279
hsp70C   ......M.T.--DN.--....LA.....H.....QRVMEYFIKLY..KS.K..R..KR.V.K..REV....RA..TQH  308
SSB1     ......L.H.--AG.--.YT....S.N.H...Q...TN.LE.FKA....KT.L.ISD.AR.LR...T...R..RT...V.  282

hsp70F   QTDINLPYITMDQSGPKHLNLKLTRAKFEQIVGDLIKRTIEPCRNVLHDAEVKSSQIADVLLVGGMSRMPKVQATVQEIF  382
dnaK     ...V.....A.AT....M.I.V....L.SL.E..VN.S...LKVA.Q..GLSV.D.D..I....QT...M..KK.A.F.  357
hsp70A   .AS.EID-SLFE--.IDFYT-NI...R..ELCA..FRS.MD.VEKS.R..KMDK..VH.IV....ST.I....KLLSDL.  355
hsp70C   ..KVEIE-SLF.--.ED-FSET.....ELNM..FRA.LK.VQK..E.SDL.KDDVHEIV....ST.I....QLIK.F.  384
SSB1     ..TVEVD-SLF.--.ED-FESS....R..DLNAA.F.S.L..VEQ..K..KISK...DE.V....ST.I....KLLSDF.  359

hsp70F   -GKVPSKAVNPDEAVAMGAAIQGAVLAGD----V-T-DVLLLDVTPLSLGIETLGGIMTKLITRNTTIPTKKSQVSFTAA  455
dnaK     -..E.R.D.....I...V..G..T..----.-K-.....M..V..T..AK.....H.....E  430
hsp70A   S..ELN.SI.....L.Y...V.A.I.S..KSEA.-Q-.L.....A.....A..V..A..K.....TA.T.T.YS  433
hsp70C   N..E..RGI.....Y...V..G.IS.E----ED.GEIV....N..TM....V..V.....G...V.....  460
SSB1     D..QLEKSI.....Y...V...I.T.QSTSDETK-.L.....A.....VGMQ.DMFGIVVP....V..I.RRTFT.C.  438

hsp70F   DGQTQVQIKVFQGEREMATSNKLLGQFSLVGI                                              487
dnaK     .N.SA.T.H.L....KR.AD..S....N.D..                                              462
hsp70A   .N.PG.L.Q.YE...A.TKD.N...K.E.S..                                              465
hsp70C   .N.PT.T.Q..E...P.TKD.HQ..K.D.T.L                                              492
SSB1     .N..T..FP.Y....VNCKE.T...E.D.KN.                                              470
```

Figure 4-4. A comparison of the percent nucleotide identities of the *C. elegans* hsp70 genes. Nucleotide identity of the DNA sequences was determined using Microgenie (Beckman).


% NUCLEOTIDE IDENTITY
50
60
70
80
90
100
hsp
1
2ps
70F
70D
70C

than the identity shared between the dnaK gene and the Drosophila (Ingolia et al., 1980) and human (Hunt and Morimoto, 1985) hsp70 genes.

The hsp70F protein has a 29 amino acid leader sequence when compared to the dnaK protein (Figure 4-3). The hsp70F leader sequence is composed primarily of uncharged amino acids with a few hydrophobic and basic amino acids but no acidic amino acids. Within this 29 amino acid leader sequence ten of the residues are serine and threonine. Since there are few hydrophobic residues this sequence would not be as hydrophobic as the hsp70C leader sequence. Instead, the hsp70F leader sequence is quite similar to the mitochondrial matrix import leader sequence. These sequences are characterized by their lack of acidic amino acids and the presence of basic amino acids as well as extensive stretches of uncharged amino acids and a high content of serine and threonine residues (van Loon et al., 1986; Colman and Robinson, 1986). Therefore, it seems likely that the hsp70F protein is transported into the mitochondrial matrix. This would explain the high degree of homology shared between hsp70F and the bacterial dnaK protein since it is believed that mitochondria arose as a symbiotic relationship between bacteria and the primitive eukaryotic cell. Recently, it has been determined that the S. cerevisiae SSC1 protein is imported into the mitochondria (E.A. Craig, personal communication). It is likely that the SSC1 gene product will be homologous to both the dnaK and hsp70F gene products.

Possible mutant phenotypes of the hsp70F gene

Only temperature-sensitive mutations of the dnaK gene have been isolated (Saito and Uchida, 1977; Itikawa and Ryu, 1979; Neidhardt et al., 1984) suggesting that the dnaK gene product is essential for cell growth. The hsp70F gene is most likely a dnaK-like gene because mRNA levels under stress and non-stress conditions are similar and there is a high degree of homology between the two genes at the amino acid level. A survey of the S. cerevisisae hsp70 genes reveals that the SSC1 gene, when disrupted, prevents growth (Craig et al., 1987) suggesting that the SSC1 gene product is essential for cell viability. This is consistent with the genetics of the dnaK gene. There are three possible mutant phenotypes for the hsp70F gene. The first type of mutation would be an early embryonic lethal, the second type would be a temperature-sensitive mutation that, when shifted to the non-permissive temperature, would block growth and the third type would be a hypomorphic mutant with a slow growth phenotype.

SUMMARY AND CONCLUSIONS

The work presented in this thesis involved DNA sequencing of five members of the Caenorhabditis elegans hsp70 multigene family. The primary goal of this thesis was to identify the functional homologs of the hsp70 genes and then predict the mutant phenotypes so that, eventually, isolation of mutant hsp70 alleles can be done. The second goal was to compare at least the 5' flanking sequences of two homologs, one from each of the two closely related species, C. elegans and C. briggsae, to identify any possible regulatory elements based on nucleotide sequence conservation. If there was any nucleotide sequence conservation, and if some of these conserved regions contained known regulatory elements, then I would be able to extend the hypothesis of Snutch (1984) and Prasad (1988). They suggested that cross-hybridizing DNA sequences representing functionally important elements are conserved between the two species. Therefore, blocks of nucleotide sequence homology determined using DNA sequence comparisons should also represent functionally important (i.e. regulatory) sequence elements.

The constitutively expressed, heat inducible gene, hsp-1 (Chapter 1), was found to be homologous to the Hsc4 gene from Drosophila and the SSA1 gene, a member of the essential SSA subfamily from S. cerevisiae. The hsp-2ps gene was identified as an unusual pseudogene duplicating only a part of the hsp-1 transcription unit that had been transposed from LGIV to LGX (Chapter 2). The pattern of nucleotide substitutions led me to suggest that the hsp-2ps gene was non-functional from the time

of the duplication event which occurred approximately 8.5 million years ago. The hsp70C and hsp70D genes were identified as members of the C. elegans grp78 gene subfamily (Chapter 3). The hsp70D gene is the first description of a highly heat inducible grp78 gene. The last gene, the constitutively expressed, heat inducible hsp70F gene was found to be more closely related to the dnaK gene from E. coli than any other identified eukaryotic hsp70 gene (Chapter 4). The hsp70F protein may be transported into the mitochondrial matrix.

A comparison of the 5' flanking sequences of the hsp70C genes from C. elegans and C. briggsae revealed several long stretches of nucleotide sequence conservation (Chapter 3). Within these conserved blocks were homologies to sequences known to function as enhancer sequences and heat shock promoter elements. One of the conserved sequence blocks was detected in the rat grp78 5' flanking sequence which has subsequently been shown to be protected by a protein during nuclease footprinting studies (E. Resendez, Jr., S.K. Wooden and A.S. Lee, submitted). My observations suggest that in a comparison of nucleotide sequences, flanking regions that are conserved between C. elegans and C. briggsae are potentially regulatory sequence elements.

Hsp70 multigene families consisting of eight or more members have been identified in S. cerevisiae, Drosophila (see Craig, 1985 and Lindquist, 1986), humans (Mues et al., 1986) and C. elegans (Snutch, 1984; Snutch et al., 1988). Some of the Drosophila and human hsp70 genes have been sequenced,

identified and studied in detail (Craig, 1985; Lindquist, 1986) but only the S. cerevisiae hsp70 multigene family has been extensively analyzed. This includes an analysis of the expression of each individual hsp70 gene and a genetic characterization of the hsp70 gene family (Craig and Jacobsen, 1984, 1985; Craig et al., 1987; Werner-Washburne et al., 1987). There are at least eight genes in the yeast hsp70 multigene family (Ingolia et al., 1982). Four of these genes belong to the essential SSA subfamily (Craig and Jacobsen, 1984; Werner-Washburne et al., 1987) which includes two heat inducible genes SSA3 and SSA4, a constitutively expressed gene SSA2 and a constitutively expressed, heat inducible gene SSA1. Two genes belong to the cold sensitive SSB subfamily (Craig and Jacobsen, 1985). One gene, SSC1, is an essential gene and the last gene, SSD1, cannot be assigned to a subfamily because of the lack of a mutant phenotype (Craig et al., 1987).

A comparison of the nucleotide sequences of the C. elegans hsp70 genes to the other hsp70 gene sequences revealed that the C. elegans hsp70 genes could also be assigned to gene subfamilies. The hsp-1 gene represents one gene family analogous to the SSA gene family in yeast. Within this subfamily is the pseudogene hsp-2ps. Isolation and further analysis of the remaining C. elegans hsp70 genes should allow identification of hsp70 genes analogous at least to the heat inducible varieties of the SSA subfamily and a constitutively expressed SSA gene. The grp78 subfamily is a separate family not yet defined in the S. cerevisiae system although a grp78

homolog has been identified (R.C. Nicholson and L.A. Moran, submitted). The hsp70F gene probably represents the SSC family based on the mitochondrial location of SSC1 and the proposed transport of hsp70F into the mitochondrial matrix. No homologs of the cold sensitive SSB family have been identified yet in C. elegans and remain to be isolated.

PROPOSALS FOR FURTHER RESEARCH

1) Isolate and sequence the remaining fragments of the hsp70D and hsp70F genes and complete their characterization.

2) Map the hsp70C, hsp70D and hsp70F genes with the aid of RFLPs.

3) Isolate mutants of the hsp70 genes:

   a) hsp-1: around the dpy-4 region, look for a maternal effect lethal

   b) hsp70C and hsp70D: look for temperature-sensitive mutants that are defective in secreting proteins or intracellular protein transport

   c) hsp70F: look for an early embryonic lethal effect mutant, a temperature-sensitive mutant which arrests growth at the non-permissive temperature or a mutant which exhibits a slow growth phenotype.

   Verify the isolation of mutants of these genes by microinjecting the purified gene DNA to rescue the mutant phenotype.

4) Isolate and identify the remaining hsp70 genes including the strictly heat inducible and the cold sensitive hsp70 genes.

5) Using DNA microinjection techniques, study the effects of deletions in the 5' flanking DNAs on expression of the hsp70 genes.

6) Using DNA microinjection techniques, study the effects of deletions of the heat inducible 3' untranslated sequences on the stability of the mRNAs during heat shock.

REFERENCES

Ananthan, J., A.L. Goldberg and R. Voellmy. 1986. Abnormal proteins serve as eukaryotic stress signals and trigger the activation of heat shock genes. Science 232: 522-524.

Arrigo, A.P., S. Fakan and A. Tissieres. 1980. Localization of the heat shock induced proteins in Drosophila melanogaster tissue culture cells. Dev. Biol. 78: 86-103.

Ashburner, M. and J.J. Bonner. 1979. The induction of gene activity in Drosophila by heat shock. Cell 17: 251-254.

Attenello, J.W. and A.S. Lee. 1984. Regulation of a hybrid gene by glucose and temperature in hamster fibroblast. Science 226: 187-190.

Bardwell, J.C.A. and E.A. Craig. 1984. Major heat shock gene of Drosophila and the Escherichia coli inducible dnaK gene are homologous. Proc. Natl. Acad. Sci. U.S.A. 81: 848-852.

Bensuade, D., L. Babinet, M. Morange and F. Jacob. 1983. Heat shock protein, first major products of zygote gene activity in mouse embryo. Nature 305: 331-333.

Benton, W.D. and R.W. Davis. 1977. Screening of recombinant clones by hybridization to single plaques in situ. Science 196: 180-182.

Berendes, H.D. 1968. Factors involved in the expression of gene activity in polytene chromosomes. Chromosoma 24: 418-437.

Bienz, M. 1984. Developmental control of the heat shock response in Xenopus. Proc. Natl. Acad. Sci. U.S.A. 81: 3138-3142.

Bienz, M. 1985. Transient and developmental activation of heat shock genes. Trends Biochem. Sci. 10: 157-161.

Biessmann, H., F.-G. Falkner, H. Saumweber and M.F. Walter. 1982. Disruption of the vimentin cytoskeleton may play a role in heat-shock response. In Heat Shock From Bacteria to Man. M.J. Schlesinger, M. Ashburner and A. Tissieres, eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) pp. 275-281.

Blackman, R.K. and M. Meselson. 1986. Interspecific nucleotide sequence comparisons used to identify regulatory and structural feature of the Drosophila hsp82 gene. J. Mol. Biol. 188: 499-515.

Bole, D.G., L.M. Hendershot and J.F. Kearney. 1986. Post-translational association of immunoglobulin heavy chain binding protein with nascent heavy chains in nonsecreting

and secreting hybridomas. J. Cell. Biol. 102: 1558-1566.

Bond, U. and M.J. Schlesinger. 1985. Ubiquitin is a heat shock protein in chicken embryo fibroblasts. Mol. Cell. Biol. 5: 949-956.

Brenner, S. 1974. The genetics of Caenorhabditis elegans. Genetics 77: 71-94.

Burdon, R.H. 1986. Heat shock and the heat shock proteins. Biochem. J. 240: 313-324.

Burdon, R.H., A. Slater, M. McMahon and A.C.B. Cato. 1982. Hyperthermia and the heat shock proteins of HeLa cells. Br. J. Cancer 45: 953-963.

Chang, S.C., S.K. Wooden, T. Nakaki, Y.K. Kim, A.Y. Lin, L. Kung, J. Attenello and A.S. Lee. 1987. Rat gene encoding the 78kDa glucose regulated protein GRP78: its regulatory sequences and the effect of protein glycosylation on its expression. Proc. Natl. Acad. Sci. U.S.A. 84: 680-684.

Chappell, T.G., W.J. Welch, D.M. Schlossman, K.B. Palter, M.J. Schlesinger and J.E. Rothman. 1986. Uncoating ATPase is a member of the 70kDa family of stress proteins. Cell 45: 3-13.

Chen, E.J. and P.H. Seeburg. 1985. Supercoil sequencing: a fast and simple method of sequencing plasmid DNA. DNA 4: 165-170.

Cheney, C.M. and A. Shearn. 1983. Developmental regulation of Drosophila imaginal disc proteins: synthesis of a heat shock protein under non-heat shock conditions. Dev. Biol. 95: 325-330.

Chirico, W.J., M. G. Waters and G. Blobel. 1988. 70K heat shock related proteins stimulate protein translocation into microsomes. Nature 332: 805-810.

Collins, P.C. and L.E. Hightower. 1982. Newcastle disease virus stimulates the cellular accumulation of stress (heat shock) mRNAs and proteins. J. Virol. 44: 703-707.

Colman, A. and C. Robinson. 1986. Protein import into organelles: hierarchial targeting signals. Cell 46: 321-322.

Courgeon, A.M., C. Maisonhaute and M. Best-Belpomme. 1984. Heat shock proteins are induced by cadmium in Drosophila cells. Exp. Cell. Res. 153: 515-521.

Craig, E.A. 1985. The heat shock response. Crit. Rev. Biochem. 18: 239-280.

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

Craig, E., T. Ingolia, M. Slater, L. Manseau and J. Bardwell. 1982. Drosophila, yeast and E. coli genes related to the Drosophila heat shock genes. In Heat Shock From Bacteria to Man. M.J. Schlesinger, M. Ashburner and A. Tissieres, eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) pp. 11-18.

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

Craig, E.A. and K. Jacobsen. 1985. Mutations in cognate genes of Saccharomyces cerevisiae hsp70 result in reduced growth rates at low temperatures. Mol. Cell. Biol. 5: 3517-3524.

Craig, E.A., J. Kramer and J. Kosic-Smithers. 1987. SSC1, a member of the 70-kDa heat shock protein multigene family of Saccharomyces cerevisiae, is essential for growth. Proc. Natl. Acad. Sci. U.S.A. 84: 4156-4160.

Davis, R.W., D. Botstein and J.R. Roth. 1980. A Manual For Genetic Engineering: Advanced Bacterial Genetics. (Cold Spring Harbor Laboratory, Cold Spring Harbor NY).

Deshaies, R.J., B.D. Koch, M. Werner-Washburne, E.A. Craig and R. Schekman. 1988. A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides. Nature 332: 800-805.

DiDomenico, B.J., G.E. Bugaisky and S. Lindquist. 1982a. Heat shock and recovery are mediated by different translational mechanisms. Proc. Natl. Acad. Sci. U.S.A. 79: 6181-6185.

DiDomenico, B.J., G.E. Bugaisky and S. Lindquist. 1982b. The heat shock response is self-regulated at both the transcriptional and post-transcriptional levels. Cell 31: 593-603.

Dodson, M., H. Echols, S. Wickner, C. Alfano, K. Mensa-Wilmot, B. Gomes, J. LeBowitz, J. Roberts and R. McMacken. 1986. Specialized nucleoprotein structures at the origin of replication of bacteriophage λ: localized unwinding of duplex DNA by a six-protein reaction. Proc. Natl. Acad. Sci. U.S.A. 83: 7638-7642.

Dragon, E.A., S.R. Sias, E.A. Kato and J.D. Gabe. 1987. The genome of Trypanosoma cruzi contains a constitutively expressed, tandemly arranged multicopy gene homologous to a major heat shock protein. Mol. Cell. Biol. 7: 1271-1275.

Dura, J.M. 1981. Stage dependent synthesis of heat shock induced proteins in early embryos of Drosophila melanogaster. Mol. Gen. Genet. 184: 381-385.

Farrelly, F.W. and D.B. Finkelstein. 1984. Complete sequence of the heat shock-inducible hsp90 gene of Saccharomyces cerevisiae. J. Biol. Chem. 259: 5745-5751.

Findly, R.C. and T. Pederson. 1981. Regulated transcription of the genes for actin and heat shock proteins in cultured Drosophila cells. J. Cell Biol. 88: 223-228.

Finley, D., A. Ciechanover and A. Varshavsky. 1984. Thermolabiltiy of ubiquitin-activating enzyme from the mammalian cell cycle mutant ts85. Cell 37: 43-55.

Fire, A. 1986. Integrative transformation of Caenorhabditis elegans. EMBO J 5: 2673-2680.

Fischer, J.A. and T. Maniatis. 1986. Regulatory elements involved in Drosophila Adh gene expression are conserved in divergent species and separate elements mediate expression in different tissues. EMBO J 5: 1275-1289.

Georgopoulos, C.P. 1977. A new bacterial gene (groPC) which affects lambda DNA replication. Mol. Gen. Genet. 151: 35-39.

Georgopoulos, C.P. and I. Herskowitz. 1971. Escherichia coli mutants blocked in lambda DNA synthesis. In The Bacteriophage Lambda. A.D. Hershey, ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) pp. 553-564.

Georgopoulos, C.P., K. Tilly, D. Drahos and R. Hendrix. 1982. The B66.0 protein of Escherichia coli is the product of the dnaK$^+$ gene. J. Bacteriol. 149: 1175-1177.

Glaser, R.L., M.F. Wolfner and J.T. Lis. 1986. Spatial and temporal pattern of hsp26 expression during normal development. EMBO J 5: 747-754.

Grossman, A.D., J.W. Erickson and C.A. Gross. 1984. The htpR gene product of E. coli is a sigma factor for heat-shock promoters. Cell 38: 383-390.

Grossman, A.D., W.E. Taylor, Z.F. Burton, R.R. Burgess and C.A. Gross. 1985. Stringent response in Escherichia coli induces expression of heat shock proteins. J. Mol. Biol. 186: 357-365.

Hattori, M. and Y. Sakaki. 1986. Dideoxy sequencing method using denatured templates. Anal. Biochem. 152: 232-238.

Hearing, P. and T. Shenk. 1983. The adenovirus type 5 E1A

transcriptional control region contains a duplicated enhancer element. Cell 33: 695-703.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates target points for DNA sequencing. Gene 28: 351-359.

Henikoff, S. 1987. Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods Enzymol. 155: 156-165.

Henikoff, S. and M.K. Eghtedarzadeh. 1987. Conserved arrangement of nested genes at the Drosophila Gart locus. Genetics 117: 711-725.

Hershko, A. 1983. Ubiquitin: roles in protein modification and breakdown. Cell 34: 11-12.

Hershko, A. and A. Ciechanover. 1982. Mechanisms of intracellular protein breakdown. Ann. Rev. Biochem. 51: 335-364.

Hershko, A., E. Leshinsky, D. Ganoth and H. Heller. 1984. ATP-dependent degradation of ubiquitin protein conjugates. Proc. Natl. Acad. Sci. U.S.A. 81: 1619-1623.

Hightower, L.E. 1980. Cultured animal cells exposed to amino acid or puromycin rapidly synthesize several polypeptides. J. Cell. Physiol. 102: 407-427.

Hiromi, Y. and Y. Hotta. 1985. Actin gene mutations in Drosophila-heat shock activation in the indirect flight muscles. EMBO J 4: 1681-1687.

Hughes, E.N. and J.T. August. 1982. Coprecipitation of heat shock proteins with a cell surface glycoprotein. Proc. Natl. Acad. Sci. U.S.A. 79: 2305-2309.

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

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

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

Ingolia, T.D., M.R. Slater and E.A. Craig. 1982. Saccharomyces

cerevisiae contains a complex multigene family related to the major heat shock-inducible gene of Drosophila. Mol. Cell. Biol. 2: 1388-1398.

Itikawa, H. and J.I. Ryu. 1979. Isolation and characterization of a temperature-sensitive dnaK mutant of Escherichia coli. J. Bacteriol. 138: 339-344.

Jamrich, M., A.L. Greenleaf and E.K.F. Bautz. 1977. Localization of RNA polymerase in polytene chromosomes of Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.A. 79: 2079-2083.

Jones, D., R.H. Russnack, R.J. Kay and E.P.M. Candido. 1986. Structure, expression, and evolution of a heat shock gene locus in Caenorhabditis elegans that is flanked by repetitive elements. J. Biol. Chem. 261: 12006-12015.

Karlik, C.C., M.D. Coutu and E.A. Fyrberg. 1984. A nonsense mutation within the Act88F actin gene disrupts myofibril formation in Drosophila indirect flight muscles. Cell 38: 711-719.

Karn, J., S. Brenner and L. Barnett. 1983. Protein structural domains in the Caenorhabditis elegans unc-54 myosin heavy chain gene are not separated by introns. Proc. Natl. Acad. Sci. U.S.A. 80: 4253-4257.

Kelley, P. and M.J. Schlesinger. 1978. The effect of amino-acid analogues and heat shock on gene expression in chicken embryo fibroblasts. Cell 15: 1277-1286.

Kelley, P.M. and M.J. Schlesinger. 1982. Antibodies to two major chicken heat shock proteins cross-react with similar proteins in widely divergent species. Mol. Cell. Biol. 2: 267-274.

Key, J.L., J. Kimpel, E. Vierleng, C.Y. Lin, R.T. Nagao, E. Czarnecka and F. Schoffl. 1985. Physiological and molecular analyses of the heat shock response in plants. In Changes in Eukaryotic Gene Expression in Response to Environmental Stress. B.G. Atkinson and D.B. Walden, eds. (Academic Press, London) pp. 327-348.

Kingston, R.E., T.J. Schuetz and Z. Larin. 1987. Heat-inducible human factor that binds to a human hsp70 promoter. Mol. Cell. Biol. 7: 1530-1534.

Klass, M., D. Ammons and S. Ward. 1988. Conservation in the 5' flanking sequences of transcribed members of the Caenorhabditis elegans major sperm protein gene family. J. Mol. Biol. 199: 15-22.

Kothary, R.K. and E.P. Candido. 1982. Induction of a novel set

of polypeptides by heat shock of sodium arsenite in cultured cells of rainbow trout, Salmo gairdneri. Can. J. Biochem. 60: 347-355.

Krause, M. and D. Hirsh. 1987. A trans-spliced leader sequence on actin mRNA in C. elegans. Cell 49: 753-761.

Kurtz, S. and S. Lindquist. 1984. Changing patterns of gene expression during sporulation in yeast. Proc. Natl. Acad. Sci. U.S.A. 81: 7323-7327.

Kurtz, S., J. Rossi, L. Petko and S. Lindquist. 1986. An ancient developmental induction: heat-shock proteins induced in sporulation and oogenesis. Science 231: 1154-1157.

Lee, A.S. 1987. Coordinated regulation of a set of genes by glucose and calcium ionophores in mammalian cells. Trends Biochem. Sci. 12: 20-23.

Levinger, L. and A. Varshavsky. 1981. Heat shock proteins of Drosophila are associated with nuclease-resistant, high salt-resistant nuclear structures. J. Cell Biol. 90: 793-796.

Levinson, W., H. Opperman and J. Jackson. 1980. Transition series metals and sulfhydril reagents induce the synthesis of four proteins in eukaryotic cells. Biochim. Biophys. Acta. 606: 170-180.

Lewis, M.J. and H.R.B. Pelham. 1985. Involvement of ATP in the nuclear and nucleolar functions of the 70kd heat shock protein. EMBO J. 4: 3137-3143.

Li, G.C. 1983. Induction of thermotolerance and enhanced heat shock protein synthesis in Chinese hamster fibroblasts by sodium arsenite and by ethanol. J. Cell. Physiol. 115: 116-122.

Li, G.C. and A. Laszlo. 1985. Amino acid analogs while inducing heat shock proteins sensitize CHO cells to thermal damage. J. Cell. Physiol. 122: 91-97.

Li, G.C. and D.C. Schrieve. 1982. Thermal tolerance and specific protein synthesis in Chinese hamster fibroblasts exposed to prolonged hypoxia. Exp. Cell Res. 142: 464-468.

Lin, A.Y., S.C. Chang and A.S. Lee. 1986. A calcium ionophore-inducible cellular promoter is highly active and has enhancerlike properties. Mol. Cell. Biol. 6: 1235-1243.

Lindquist, S. 1981. Regulation of protein synthesis during heat shock. Nature 293: 311-314.

Lindquist, S. 1986. The heat shock response. Ann. Rev.

Biochem. 55: 1151-1191.

Lowe, D,G. and L.A. Moran. 1984. Proteins related to the mouse L-cell major heat shock protein are synthesized in the absence of heat shock gene expression. Proc. Natl. Acad. Sci. U.S.A. 81: 2317-2321.

Maniatis, T., E.F. Fritsch and J. Sambrook. 1982. Molecular Cloning, A Laboratory Manual. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

Mason, P.J., L.M. C. Hall and J. Gausz. 1984. The expression of heat shock genes during normal development in Drosophila melanogaster. Mol. Gen. Genet. 194: 73-78.

Maynard, S. and T. Pederson. 1983. Heat shock alters ribonucleoprotein assembly in Drosophila cells. Mol. Cell. Biol. 3: 161-171.

McGarry, T.J. and S, Lindquist. 1985. The preferential translation of Drosophila hsp70 mRNA requires sequences in the untranslated leader. Cell 42: 903-911.

McKenzie, S.L., S. Henikoff and M. Meselson. 1975. Localization of RNA from heat induced polysomes to puff sites in Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.A. 72: 1117-1121.

Mitchell, H.K., N.S. Peterson and C.H. Biezin. 1985. Self-degradation of heat shock proteins. Proc. Natl. Acad. Sci. U.S.A. 82: 4969-4973.

Mues, G.I., T.Z. Munn and J.D. Raese. 1986. A human gene family with sequence homology to Drosophila melanogaster Hsp70 heat shock gene. J. Biol. Chem. 261: 874-877.

Munro, S. and H.R.B. Pelham. 1985. What turns on heat shock genes? Nature 317: 477-478.

Munro, S. and H.R.B. Pelham. 1986. An hsp70-like protein in the endoplasmic reticulum identity with the 78 kD glucose-regulated protein and immunoglobulin heavy chain binding protein. Cell 46: 291-300.

Munro, S. and H.R.B. Pelham. 1987. A C-terminal signal prevents secretion of luminal ER proteins. Cell 48: 899-907.

Neidhardt, F.C. 1987. What the bacteriologists have learned about heat shock. Genes Devel. 1: 109-110.

Neidhardt, F.C. and R.A. VanBogelen. 1981. Positive regulatory gene for temperature-controlled proteins in Escherichia coli. Biochem. Biophys. Res. Commun. 100: 894-900.

Neidhardt, F.C., R.A. VanBogelen and V. Vaughn. 1984. The genetics and regulation of heat shock proteins. Ann. Rev. Genet. 18: 295-329.

Nevins, J.R. 1982. Induction of the synthesis of a 70,000 dalton mammalian heat-shock protein by the adenovirus E1A gene product. Cell 29: 913-919.

Norrander, J., T. Kempe and J. Messing. 1983. Construction of improved M13-vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26: 101-106.

Notarianni, E.L. and C.M. Preston. 1982. Activation of cellular stress protein genes by herpes simplex virus temperature-sensitive mutants which overproduce early polypeptides. Virology 123: 113-122.

Ochman, H. and A.C. Wilson. 1987. Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. 26: 74-86.

Palter, K.B., M. Watanabe, L. Stinson, A.P. Mahowald and E.A. Craig. 1986. Expression and localization of Drosophila melanogaster hsp70 cognate proteins. Mol. Cell. Biol. 6: 1187-1203.

Parker, C.S. and J.A. Topol. 1984a. A Drosophila RNA polymerase II transcription factor contains a promoter-region-specific DNA binding activity. Cell 36: 357-369.

Parker, C.S. and J.A. Topol. 1984b. Drosophila RNA polymerase II transcription factor specific for the heat shock gene binds to the regulatory site of an hsp70 gene. Cell 37: 273-283.

Pelham, H.R.B. 1982. A regulatory upstream promoter element in the Drosophila hsp70 heat-shock gene. Cell 30: 517-528.

Pelham, H.R.B. 1984. Hsp70 accelerates the recovery of nucleolar morphology after heat shock. EMBO J 3: 3095-3100.

Pelham, H.R.B. 1985. Activation of heat shock genes in eukaryotes. Trends Genet. 1: 31-35.

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

Plesset,J., C. Palm and C.S. McLaughlin. 1982. Induction of heat shock proteins and thermotolerance by ethanol in Saccharomyces cerevisiae. Biochem. Biophys. Res. Commun. 108: 1340-1345.

Prasad, S.S. 1988. Molecular analysis within the dpy-20 to unc-

22 region on chromosomeIV of Caenorhabditis elegans. Ph. D. thesis. Simon Fraser University, Burnaby, British Columbia.

Riddle, D.L. 1980. Developmental genetics of Caenorhabditis elegans. In Nematodes as Biological Models. B.M. Zuckerman, ed. (Academic Press, New York) pp. 263-283.

Ritossa, F. 1962. A new puffing pattern induced b temperature shock and DNP in Drosophila. Experientia 18: 571-573.

Rosenweig, B., L.W. Liao and D. Hirsh. 1983. Sequence of the C. elegans transposable element Tc1. Nucleic Acids Res. 11: 4201-4209.

Rothman, J.E. and S.L. Schmid. 1986. Enzymatic recycling of clathrin from coated vesicles. Cell 46: 5-9.

Rothstein, R.J. 1983. One step gene disruption in yeast. Methods Enzymol. 101: 202-211.

Russnack, R.H. and E.P.M. Candido. 1985. Locus encoding a family of small heat shock genes in Caenorhabditis elegans: Two genes duplicated to form a 3.8-kilobase inverted repeat. Mol. Cell. Biol. 5: 1268-1278.

Saito, H. and H. Uchida. 1977. Initiation of the DNA replication of bacteriophage lambda in Escherichia coli K12. J. Mol. Biol. 113: 1-25.

Sanders, M.M., D. Feeney-Triemer, A.S. Olsen and J. Farrell-Towt. 1982a. Changes in protein phosphorylation and histone H2b disposition in heat shock in Drosophila. In Heat Shock From Bacteria to Man. M.J. Schlesinger, M. Ashburner and A. Tissieres, eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) pp. 235-242.

Sanders, M.M., D.F. Triemer and B. Brief. 1982b. Heat shock changes the cytoskeleton in Drosophila. J. Cell Biol. 95: 237a.

Sanger, F., A.R. Coulson, B.G. Barell, A.J.H. Smith and B.A. Rose. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143: 161-178.

Schlesinger, M.J., M. Ashburner and A. Tissieres. 1982. Heat Shock From Bacteria to Man. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY)

Sciandra, J.J. and J.R. Subjeck. 1983. The effects of glucose on protein synthesis and thermosensitivity in Chinese hamster ovary cells. J. Biol. Chem. 258: 12091-12093.

Sciandra, J.J., J.R. Subjeck and C.S. Hughes. 1984. Induction of glucose-regulated proteins during anaerobic exposure and

of heat shock proteins after reoxygenation. Proc. Natl. Acad. Sci. U.S.A. 81: 4843-4847.

Sinibaldi, R.M. and P.W. Morris. 1981. Putative function of D. melanogaster heat shock proteins in the nucleoskeleton. J. Biol. Chem. 256: 10735-10738.

Sirotkin, K. and N. Davidson. 1982. Developmentally regulated transcription from Drosophila melanogaster chromosomal site 67B. Dev. Biol. 89: 196-210.

Smith, G.E. and M.O. Summers. 1980. The bidirectional transfer of DNA and RNA to nitrocellulose or DBM paper. Anal. Biochem. 109: 123-129.

Snutch, T.P. 1984. Molecular and genetic analysis of the heat shock response of Caenorhabditis elegans. Ph. D. thesis. Simon Fraser University, Burnaby, British Columbia.

Snutch, T.P. and D.L. Baillie. 1983. Alterations in the pattern of gene expression following heat shock in the nematode Caenorhabditis elegans. Can. J. Biochem. Cell Biol. 61: 480-487.

Snutch, T.P. and D.L. Baillie. 1984. A high degree of DNA strain polymorphism associated with the major heat shock gene in Caenorhabditis elegans. Mol. Gen. Genet. 195: 329-335.

Snutch, T.P., M.F.P. Heschl and D.L. Baillie. 1988. The Caenorhabditis elegans hsp70 gene family: a molecular genetic characterization. Gene, in press.

Sorger, P.K., M.J. Lewis and H.R.B. Pelham. 1987. Heat shock factor is regulated differently in yeast and HeLa cells. Nature 329: 81-84.

Sorger, P.K. and H.R.B. Pelham. 1987. Purification and characterization of a heat shock element binding protein from yeast. EMBO J 6: 3035-3041.

Spieth, J., K. Denison, S. Kirtland, J. Cane and T. Blumenthal. 1985a. The C. elegans vitellogenin genes: short sequence repeats in the promoter regions and homology to the vertebrate genes. Nucleic Acids Res. 13: 5283-5295.

Spieth, J., K. Denison, E. Zucker and T. Blumenthal. 1985b. The nucleotide sequence of a nematode vitellogenin gene. Nucleic Acids Res. 13: 7129-7138.

Spradling, A., S. Penman and M.L. Pardue. 1975. Analysis of Drosophila mRNA by in situ hybridization: sequences transcribed in normal and heat shocked cultured cells. Cell 4: 395-404.

Sulston, J.E. and S. Brenner. 1974. The DNA of Caenorhabditis elegans. Genetics 77: 95-104.

Sulston, J.E. and H.R. Horvitz. 1977. Post-embryonic cell lineages of the nematode Caenorhabditis elegans. Dev. Biol. 56: 110-156.

Sulston, J.E., E. Schierenberg, J.G. White and J.N. Thomson. 1983. The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev. Biol. 100: 64-119.

Sunshine, M., M. Feiss, J. Stuart and J. Yochem. 1977. A new host gene (groPC) necessary for lambda DNA replication. Mol. Gen. Genet. 151: 27-34.

Tanguay, R.M. and M. Vincent. 1982. Intracellular translocation of cellular and heat shock induced proteins upon heat shock in Drosophila Kc cells. Can. J. Biochem. 60: 306-315.

Thomas, G.P., W.J. Welch, M.B. Mathew and J.R. Feramisco. 1982. Molecular and cellular effects of heat shock and related treatments of mammalian tissue culture cells. Cold Spring Harbor Symp. Quant. Biol. 46: 985-996.

Thomas, P.G. and M.B. Mathews. 1984. Alterations of transcription and translation in HeLa cells exposed to amino acid analogs. Mol. Cell. Biol. 4: 1063-1072.

Tilly, K., N. McMittrick, M. Zylicz and C. Georgopoulos. 1983. The dnaK protein modulates the heat shock response in Escherichia coli. Cell 34: 641-646.

Ungewickell, E. 1985. The 70-kD mammalian heat shock proteins are structurally and functionally related to the uncoating protein that releases clathrin triskelia from coated vesicles. EMBO J 4: 3385-3391.

VanLoon, A.P.G.M., A.W. Brandli and G. Schatz. 1986. The presequences of two imported mitochondrial proteins contain information for intracellular and intramitochondrial sorting. Cell 44: 801-812.

Velazquez, J.M., B.J. DiDomenico and S. Lindquist. Intracellular localization of heat shock proteins in Drosophila. Cell 20: 679-689.

Velazquez, J.M. and S. Lindquist. 1984. Hsp70: nuclear concentration during environmental stress: cytoplasmic storage during recovery. Cell 36: 655-662.

Velazquez, J.M., S. Sonoda, G. Bugaisky and S. Lindquist. 1983. Is the major Drosophila heat shock protein present in cells that have not been heat shocked? Cell Biol. 96: 286-290.

Vitek, M.P. and E.M. Berger. 1984. Steroid and high-temperature induction of the small heat shock protein genes in Drosophila. J. Mol. Biol. 178: 173-189.

Ward, S., D.J. Burke, J.E. Sulston, A.R. Coulson, D.G. Albertson, D. Ammons, M. Klass and E. Hogan. 1988. Genomic organization of major sperm protein genes and pseudogenes in the nematode Caenorhabditis elegans. J. Mol. Biol. 199: 1-13.

Watowich, S.S. and R.I. Morimoto. 1988. Complex regulation of heat shock- and glucose-responsive genes in human cells. Mol. Cell. Biol. 8: 393-405.

Weiher, H., M. Konig and P. Gruss. 1983. Multiple point mutations affecting the simian virus 40 enhancer. Science 217: 626-631.

Welch, W.J. and J.R. Feramisco. 1984. Nucleolar localization of the 72,000 dalton heat shock protein in mammalian cells. J. Biol. Chem. 259: 4501-4513.

Welch, W.J. and J.R. Feramisco. 1985. Rapid purification of the mammalian 70,000 dalton stress proteins: affinity of the proteins for nucleotides. Mol. Cell. Biol. 5: 1229-1237.

Welch, W.J. and J.P. Suhan. 1985. Morphological study of the mammalian stress response: characterization of changes in cytoplasmic organelles, cytoskeleton, and nucleoli, and appearance of intranuclear actin filaments in rat fibroblasts after heat shock treatment. J. Cell Biol. 101: 1198-1211.

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

Wiederrecht, G., D.J. Shuey, W.A. Kibbe and C.S. Parker. 1987. The Saccharomyces and Drosophila heat shock transcription factors are identical in size and DNA binding properties. Cell 48: 507-515.

Wu, B.J. and R.I. Morimoto. 1985. Transcription of the human hsp70 gene is induced by serum stimulation. Proc. Natl. Acad. Sci. U.S.A. 82: 6060-6074.

Wu, C. 1984a. Two protein-binding sites in chromatin implicated in the activation of heat shock genes. Nature 309: 229-234.

Wu, C. 1984b. Activating protein factor binds in vitro to upstream control sequences in heat shock gene chromatin. Nature 311: 81-84.

Wu, C., S. Wilson, B. Walker, I. Dawid, T. Paisley, V. Zimarino and H. Ueda. 1987. Purification and properties of Drosophila heat shock activator protein. Science 238: 1247-1253.

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

Yamomori, T. and T. Yura. 1982. Genetic control of heat shock protein synthesis and its bearing on growth and thermal resistance in Escherichia coli K12. Proc. Natl. Acad. Sci. U.S.A. 79: 860-864.

Yost, H.J. and S. Lindquist. 1986. RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. Cell 45: 185-193.

Zimarino, V. and C. Wu. 1987. Induction of sequence-specific binding of Drosophila heat shock activator protein without protein synthesis. Nature 327: 727-730.

Zimmerman, J.L., W. Petri and M. Meselson. 1983. Accumulation of a specific subset of Drosophila melanogaster heat shock mRNAs in normal development without heat shock. Cell 32: 1161-1170.

Zylicz, M., J. LeBowitz, R. McMacken and C. Georgopoulos. 1984. The dnaK protein of Escherichia coli possesses an ATPase and autophosphorylating activity and is essential in a in vitro DNA replication system. Proc. Natl. Acad. Sci. U.S.A. 80: 6431-6435.