ISOLATION AND CHARACTERIZATION OF GENES ENCODING HEAT SHOCK PROTEIN 70S (*hsp70*s) FROM TWO SPECIES OF THE COELACANTH, *LATIMERIA CHALUMNAE* AND *LATIMERIA MENADOENSIS*

A thesis submitted in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

of

RHODES UNIVERSITY

by

KEOAGILE WILLIAM MODISAKENG

August 2006

ABSTRACT

The extant coelacanths have a close resemblance to the coelacanth fossil records dating back to 230mya. Like their predecessors, the extant coelacanths inhabit rocky caves at a depth of 100-300m below sea level. In the Comoros, the water temperature at these depths is estimated to fluctuate between 14-20˚C. High-level adaptation to these environment and lack of competition are thought to have led to the morphological uniformity and slow change throughout the history of the coelacanths. Under stress conditions, proteins unfold or misfold leading to the formation of aggregates. Molecular chaperones facilitate the correct folding of other proteins so that they can attain a stable tertiary structure. In addition, molecular chaperones aid the refolding of denatured proteins and the degradation of terminally misfolded protein after cellular stress. Heat shock proteins form one of the major classes of molecular chaperones. Here we show that, despite high-level adaptation to a unique habitat and slow change, the genome of the coelacanth encodes the major and highly conserved molecular chaperone, Hsp70. *Latimeria menadoensis* and *Latimeria chalumnae* contain intronless *hsp70* genes encoding Hsp70 proteins archetypal of known Hsp70s. Based on the coelacanth codon usage, we have shown that bacterial protein expression systems, particularly *Escherichia coli*, may not be appropriate for the overproduction of coelacanth Hsp70s and coelacanth proteins in general. Also interesting, was the discovery that like the rat Hsc70, the *L. menadoensis* Hsp70 could not reverse thermal sensitivity in a temperate sensitive *E. coli* DnaK mutant strain, BB2362. We also report the successful isolation of a 1.2 kb region of *L. menadoensis hsp70* upstream regulatory region. This region contain three putative heat shock elements, a TATA- box and two CAAT-boxes. This regulatory region resembled the *Xenopus*, mouse, and particularly tilapia *hsp70* promoters, all of which have been shown to drive the expression of reporter genes in a heat dependent manner. Taken together, this data is the first to strongly suggest an inducible Hsp70-base cytoprotection mechanism in the coelacanth. It further provides basis to formulate testable predictions about the regulation, structure and function of Hsp70s in the living fossil, *Latimeria*.

DECLARATION

I declare that this thesis is my own, unaided work. It is being submitted for the degree of Doctor of Philosophy in Rhodes University. It has not been submitted before for any degree or examination in any other university.

_____

_____

This _____ day of _____2006

DEDICATION

This thesis is dedicated to

My family, my loving wife Cynthia and my son Kgolagano.

Also in special memory of David and Samaria Mahlangu, our loving parents who passed away during my PhD studies. Thank you for the unwavering support.

ACKNOWLEDGMENTS

I would like to acknowledge my supervisor, Professor Greg Blatch for his support, love for research and passion for training and developing student.

I would like to acknowledge:

- The Deutscher Akademischer Austauchdienst (DAAD) for the PhD scholarship.
- The National Research Foundation (NRF) for the PhD scholarship.
- The South African Department of Science and Technology and the NRF for funding the project through The African Coelacanth Ecosystem Programme (ACEP).
- Dr C. T. Amemiya (Benaroya Research Institute at Virginia Mason, USA) and Ms M. Jiwaji (Rhodes University, Grahamstown, South Africa) for providing assistance and equipment for construction of the *L. menadoensis* mini-library.
- Dr. B. Bukau (University of Heidelberg, Germany) for the gift of the *E. coli* BB2362 strain
- Dr. W. Burkholder for the gift of *pQE60* and *pBB46* plasmids

I thank all my colleagues in the Chaperone Research Laboratory (Rhode University), past and present. Thank you.

I extend my thanks to Professor Rosemary Dorrington, Mrs Val Hodgson and Lab 417 for their unwavering support.

Special thank to my friends Elijah, Khensani, Dudu, Anthony, Dez, Cecil, Nenekazi, Kennedy, Jabu, Nolwazi, Seale, Helen, Vuyo, Malcom, Bonolo. You all had a hand.

"If any of you lacks wisdom, he should ask God, who gives generously to all without finding fault, and it will be given to him. But when he asks, he must believe and not doubt, because he who doubts is like a wave of the sea, blown and tossed by the win." James 1:5-6

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Literature review

1.1 A COMPREHENSIVE REVIEW OF THE BIOLOGY AND MOLECULAR BIOLOGY OF THE COELACANTH

1.1.1 Coelacanth discoveries and population biology

Coelacanths were thought to have become extinct about 80-70 million years ago (mya), when they disappeared from the fossil record. Preceding this period, there appears to have been around 80 species of coelacanths in 40 genera and four families during the period extending from 350 to 70 mya (Heemstra, 2001). In 1938, a coelacanth was trawled near East London, South Africa and named *Latimeria chalumnae* after its discoverer, Marjorie Courtney-Latimer (Smith, 1939). A second fish was caught off Anjouan Island in the Comoros in 1952 followed by more than one hundred subsequent catches in the region (Smith, 1953; Heemstra, 2001). Coelacanths have also been discovered off the coast of Manado Tua Island, Sulawesi, Indonesia. In September 1997, Mark Erdmann and his wife saw and photographed a fish that was later identified as a coelacanth in the fish market in Manado Tua. Subsequent to this sighting, the second Indonesian coelacanth was caught North of Manado Tua and was named *Latimeria menadoensis* (Erdmann *et al*., 1998; Pouyaud *et al*., 1999). Two further Asian coelacanths were later sighted by Fricke and co-workers southwest of Manado Tua (Fricke *et al*., 2000).

The first discovery of a live coelacanth population off the South African coast occurred in November 2000, when recreational divers found three coelacanths in the underwater Jesser Canyon at Sodwana Bay at the depth of 117 m (Heemstra, 2001). Surveys by the African Coelacanth Ecosystem Program (ACEP) have identified a total of 25 coelacanth individuals (Sink and Scott-Williams, 2005). Other African coelacanth catches have been reported off the coasts of Mozambique, Madagascar, Tanzania and Kenya (Schartl *et al*.,

2005). Investigations into the relatedness of the East African coelacanths using mitochondrial DNA sequence and microsatellite DNA from 47 coelacanth individuals suggested that the scattered groups of African coelacanths might have come from a single remote population, possibly the Comoros or other unknown habitats in the Indian Ocean (Schartl *et al.*, 2005).

1.1.2 Preservation of tissue material

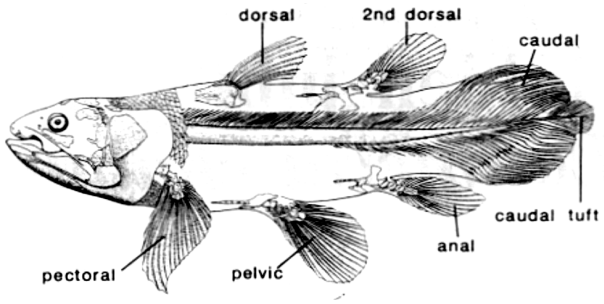
A major difficulty with functional studies on the coelacanth is the lack of coelacanth-specific biological systems. Despite the capture of over 180 coelacanths (Heemstra, 2001, Fricke and Hissmann, 1990), appropriate tissue preservation has been problematic, mainly due to the fact that the habitat and distribution of both African and Indonesian coelacanths coincides with some of the poorest and most undeveloped communities in the world. Consequently, most of the stored tissue specimens are unsuitable for molecular research because of the time lag between an accidental catch and the preservation of the tissue, as well as the use of unsuitable preservation methods. Some of the best-preserved material was reported in 1992 which included tissue from a gravid female with 26 late-term fetuses that was caught off the coast of Mozambique. Frozen tissue samples from this catch were donated to the JLB Smith Institute of Ichthyology in Grahamstown, South Africa, and the Max-Plank-Institut für Verhaltensphysiologie in Seewiesen, Germany (Fricke, 1992)..

1.1.3 Etymology and evolutionary significance

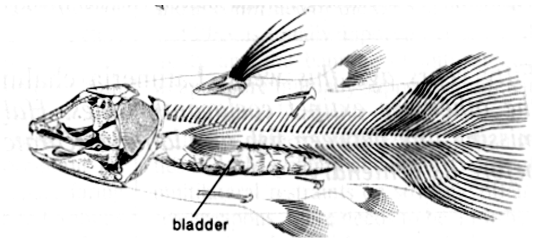
The genus name *Coelacanthus* means “hollow spine” and refers to the hollow neural and haemal spines of the vertebrae that connect to the tubular bones supporting the upper and lower caudal-fin rays (Heemstra, 2001). Notably all coelacanths, extinct and extant, possess these hollow spines, as well as other unique features compared to other living fishes. These include the extra lobe on their tail, paired lobed fins, a vertebral column that is not completely developed, and a fully functional intracranial joint (Forey, 1988). The coelacanth is a good example of a species that has evolved and adapted while

retaining the characteristic features of its ancestors. The skeleton of the extant coelacanth have been found to be almost identical to those of the *Macropoma* (upper cretaceous period: ~80 mya) and *Laugia* (lower triassic period: ~230 mya) (Forey, 1988; Figure 1.1).

(A)



(B)



(C)

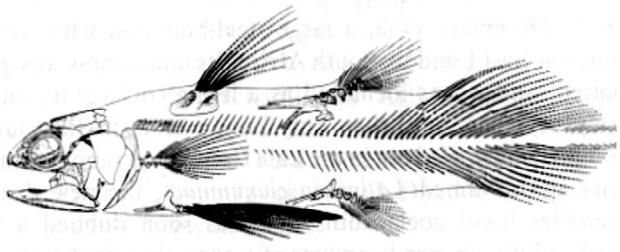


Figure 1.1: The skeletons of coelacanths depicting their conserved morphology. Differences in the morphologies include the increased lobulation in the second dorsal fin and anal fin. (A) *Latimeria* (extant coelacanth); (B) *Macropoma* -80 mya; (C) *Laugia* -230 mya. *Latimeria* (extant coelacanth), has a fat-filled organ in place of the gas-filled bladder found in *Macropoma*. However the basic skeletal structure of the extant coelacanth remains unchanged relative to *Macropoma* and *Laugia* (Figure adapted from Forey, 1988).

Several *ad hoc* hypotheses have sought to explain the slow change throughout the history of the coelacanth. One such explanation came from the Darwinian view that high-level adaptation to certain environments and lack of competition may have led to morphological uniformity (Forey, 1988). Heemstra suggested that *Latimeria* fits Darwin's description of a living fossil, since it was the only extant member of the coelacanths which was abundantly found 290-208 mya (Heemstra, 2001). Furthermore, the coelacanth inhabits rocky caves at 100-300 m below sea level where competition is presumably not severe (Heemstra, 2001). An alternative approach to investigate the slow evolutionary change might be through molecular genetics (Forey, 1988). However, the application of genetics to derive a clear relationship between coelacanth evolution to generation time or differential lineage-specific mutation rate necessitates more research on the coelacanth genome (Forey, 1988).

1.1.4 Phylogenetic studies

The evolutionary relationships of gnathostomes (jawed vertebrates) have been debated for over a century. Gnathostomes fall into two major taxa: the Chondrichthyes (cartilaginous fishes) and the Osteichthytes (bony fishes). Bony fishes include actinopterygians (ray-finned fish) and sarcopterygians (coelacanths, lungfish and tetrapods) (Venkatesh *et al*., 2001).

Currently, there are three contested hypotheses on the phylogenetic position of the coelacanth within the sarcopterygian lineage (Figure 1.2):

(a) the lungfish is the sister group to tetrapods;

(b) the coelacanth is the sister group to tetrapods; or

(c) the coelacanth and lungfish are sister-taxa and equally distant from tetrapods (Venkatesh *et al*., 2001; Brinkmann *et al*., 2004; Takezaki *et al*., 2004).

Most molecular biology research on the coelacanth has been aimed at resolving the phylogenetic classification of the coelacanth. These include studies on genes encoding coelacanth α and β haemoglobin proteins, 18S rRNA and 28S rRNA, cytochrome

oxidase subunit 1, major histocompatibility complex (MHC) class 1, 12S rRNA and 16S rRNA, the enolase protein family and cytochrome b (Gorr *et al*., 1991; Hillis *et al*., 1991, Stock *et al*., 1991; Hedges *et al*., 1993; Yokobari *et al*., 1994; Zardoya and Meyer, 1996; Tracy and Hedges, 2000; Meyer,1995; Stock and Swofford, 1991; and Meyer and Wilson, 1990).

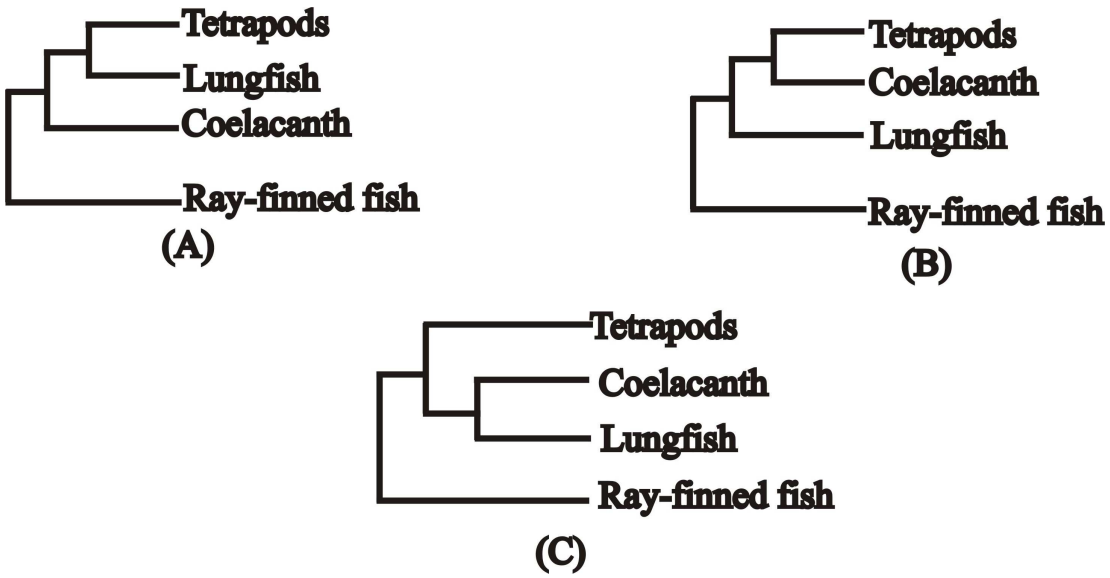


Figure 1.2: Three possible phylogenetic relationships of tetrapod, lungfish and coelacanth lineages. (A) The lungfish is the sister group to tetrapods, (B) the coelacanth is the sister group to tetrapods and (C) The coelacanth and lungfish are sister groups and equally distant from tetrapods. (Figure adapted from Venkatesh *et al*., 2001and Brinkmann *et al*., 2004).

A close tetrapod-coelacanth relationship was initially surmised on the basis of phylogenetic analysis of hemoglobin sequences. Among 522 hemoglobin sequences analyzed, the β chains of the coelacanth and the α chains of the *Rana catesbeiana* tadpoles had the highest match. These data supported a sister-group relationship between the coelacanth and tetrapods (hypothesis (b), Figure 1.2B; Gorr *et al*., 1991), but this conclusion was vigorously rejected by other researchers who reanalyzed the data and came to a contradictory interpretation, where no support was found for the coelacanth-tetrapod relationship (Stock and Swofford, 1991). An attempt to solve the coelacanth phylogenetic position using comparative analysis of 18S rRNA sequences resulted in a tree with none of its relationships significantly supported at 95% confidence. The 18S

rRNA was therefore deemed as insufficient to address the question of coelacanth relations (Meyer, 1995). In contrast, a comparative study of the genes encoding 12S rRNA and cytochrome b from the coelacanth, tetrapods, lungfish and frog supported the sister-group relationship of the lungfish and tetrapods (hypothesis (a), Figure 1.2A; Meyer and Wilson, 1990; Meyer, 1995). Using vertebrate 28S rRNA sequences, Hillis and co-workers provided evidence for a sister-group relationship between the coelacanth and tetrapods; however, this analysis was incomplete since it did not include the lungfish 28S rRNA. The reanalysis of the 28S rRNA data set including the lungfish 28S rRNA supported hypothesis (c) that the coelacanth and lungfish are sister taxa and equally distant from tetrapods (Hillis *et al*., 1991; Zardoya and Meyer, 1996). It was obvious at this stage that a much larger data set would be needed to resolve the phylogenetic classification of the coelacanth.

The complete mitochondrial genome sequences of both coelacanth species have been successfully used to provide evidence for their relatedness. In 1997, Zardoya and Meyer (1997) reported a complete DNA sequence of the *L. chalumnae* mitochondrial genome. This study revealed that the general codon usage of the coelacanth was similar to that of the lamprey, bichir, carp, trout, lungfish and tetrapods with adenosines and cytosines preferentially used at the third codon position. Comparison of the recently completed *L. menadoensis* mitochondrial sequence with that of *L. chalumnae* revealed a 4.28% difference between the two coelacanth species and provided an estimate of the divergence time between the two species at 40-30 mya (Inoue *et al*., 2005). This estimate of divergence time was different from the 6.3-4.7 mya estimate by Holder and co-workers; however, the latter study was only based on mitochondrial fragments containing the *cyt b* and two rRNA genes (Holder *et al*., 1999).

Phylogenetic analyses using the complete *L. chalumnae* mitochondrial DNA sequence could not clearly exclude any of the three hypotheses on the tetrapod, lungfish and coelacanth relationship; however, it did place the coelacanth closer to tetrapods and other sarcopterygians, than to ray-finned fish (Zardoya and Meyer, 1996). The inability of the coelacanth mitochondrial genomes to resolve the tetrapod, lungfish and coelacanth

phylogenetic relationship has intensified the search for answers in the chromosomal genome. Phylogenetic analysis using the amino acid sequence of the myelin DM20 protein supported hypothesis (a) (Figure 1.2A), that the lungfish and not the coelacanth is the closest relative of tetrapods (Tohyama *et al*., 2000). Phylogenetic analysis of molecular markers, *recombination activating genes* 1 and 2 genes (*RAG1* and *RAG2*), also supported a sister relationship between the lungfish and tetrapods (Brinkmann *et al*., 2004). Another recent attempt to clarify the tetrapod, lungfish, and coelacanth relationship came from Takezaki and co-workers who analysed amino acid sequences from 44 protein coding nuclear genes (Takezaki *et al*., 2004). Data from this analysis, however, did not conclusively support any one of the three hypotheses, leading to the conclusion that the divergence between the coelacanth, lungfish and tetrapods probably occurred in a very short period of time thereby making their phylogenetic resolution difficult. The authors further proposed that the short divergence period between the three lineages necessitated an increase in the number of loci used for such analyses. To resolve these phylogenetic relationships with more confidence there is a need to conduct in the short to medium term an analysis of other informative molecular markers (loss and gain of introns, indels, genomic rearrangements such as inversions etc.), and in the long term whole genome sequencing (Venkatesh *et al*., 2001; Brinkmann *et al*., 2004).

1.1.5 The coelacanth nuclear genome

The coelacanth was found to have a 48-chromosome karyotype that contained metacentric, subtelocentric and telocentric chromosomes, as well as microchromosomes. This karyotype was unlike that of the lungfish (which has 36-38 large metacentric chromosomes), but very similar to the 46-karyotype of the *Ascaphus truei*, an ancient frog (Bogard *et al*., 1994). Flow cytometric analysis of Comoran coelacanth (*L. chalumnae)* blood estimated the genome size of this species of the coelacanth at 2.75 pg per 1C (C.T. Amemiya, unpublished data). This estimate was substantially smaller than the 3.6 pg per 1C estimated from prior studies (Bogard *et al*., 1994; Danke *et al*., 2004).

A coelacanth (*L. menadoensis)* genomic library with average insert size of 171 kb has been successfully constructed using the bacterial artificial chromosome (BAC) system. The representation of the library was estimated at ≥ 7 genome-equivalents (Danke *et al*., 2004). The successful construction of this Indonesian coelacanth genomic library is expected to bring new insight to both comparative and functional genomics. Data and lines of investigation that have come out of this resource include the discovery of the *L. menadoensis Hoxa14* gene (Powers *et al*., 2004), the comparative and functional analysis of the *L. menadoensis Hoxc8* early enhancer (Anand *et al*., 2003; Shashikant *et al*., 2004), and the characterization of *L. menadoensis hsp70* genes (Danke *et al*., 2004; Modisakeng, data presented in thesis). The study of the protocadherin-encoding genes was used to evaluate the possible utility of the coelacanth genome sequence to infer the evolutionary history of tetrapod genomes (Noonan *et al*., 2004). The protocadherins are synaptic cell adhesion molecules thought to provide a molecular code involved in the generation of synaptic complexity in the developing brain. The copy number of the protocadherin genes and their sequences in vertebrates could be an indication of adaptive differences in protocadherin function with reference to the level of organism complexity (Noonan *et al*., 2004). The coelacanth was found to have 49 protocadherin-encoding genes clustered in the same manner as the 54 protocadherin-encoding genes in humans. These results, including the statistical comparisons of the rates of molecular evolution, strongly suggested that the coelacanth protocadherin genes have not undergone accelerated rates of molecular evolution as is the case with teleost fishes. If the protocadherin results are indicative of the entire genome, the coelacanth could be used as a genome outgroup in tetrapod comparative sequence analyses, and a valid candidate for whole-genome sequencing (Noonan *et al*., 2004).

1.1.6 *Hox* genes

The *Hox* genes encode transcription factors engaged in specifying the body plans of metazoans (Anand *et al*, 2003). The *Hox* genes are important not only due to their central developmental function, but also because they offer insights into genome duplication events that occurred during the evolution of modern tetrapods and ray-finned fishes

(Minguillón *et al*., 2005). All invertebrates so far investigated have been shown to have a single HOX cluster as opposed to three to eight clusters in the major taxa of vertebrates (Wagner *et al*., 2003; Minguillón *et al*., 2005). It has also been shown that mammals and tetrapods have four gene clusters while the sea lamprey (*Petromyzon marinus*), a jawless vertebrate (Agnatha), has three or four HOX clusters, each cluster containing up to 13 paralogous groups (Figure 1.3; Garcia-Fernàndez and Holland, 1994; Holland and Garcia-Fernàndez, 1996; Force *et al*., 2002; Irvine *et al*., 2002). Comparative analysis of these genes and gene clusters has often been used to exemplify the notion that at least two rounds of genome duplication preceded modern mammals (Aparicio, 2000; Vandepoele *et al*., 2004). Given the evolutionary significance of the coelacanth, Koh and co-workers used an extensive PCR survey with degenerate primers to characterize the *L. menadoensis* HOX clusters and their gene complement in an attempt to partially answer the questions regarding genome duplication in the coelacanth and to probe into the origins of tetrapod limbs (Koh *et al*., 2003). *L. menadoensis* was found to have 33 genes out of a possible 52 genes organized into four clusters. Although the gene complement was lower than the 39 genes normally found in mammals, the coelacanth *Hox* genes were more similar in sequence to mammalian *Hox* genes than to that of ray-finned fish. Even though the strategy used by Koh and co-workers could not confirm the absence of *Hox* genes that were not obtainable in the study or the absolute arrangement of the genes within the respective clusters, it provided good support for the notion of two rounds of genomic duplication during vertebrate evolution (Koh *et al*., 2003). In a separate report, the *L. menadoesis Hoxc8* early enhancer was isolated and shown to bear a close resemblance to the mouse early enhancer in both structure and function (Anand *et al*., 2003; Shashikant *et al*., 2004).

Recent sequence analysis of the *L. menadoensis* HOX-A cluster and the horn shark (*Heterodontus francisci*) HOX-D cluster revealed the presence of additional paralogous *Hox14* genes, named *Hoxa14* and *Hoxd14*, respectively (Powers *et al*., 2004). Although the cephalochordate *amphioxus* (closest extant relative of vertebrates) has a single HOX cluster with a group 14 gene, the discovery of the horn shark and coelacanth *Hox14* genes was the first in vertebrates (Figure 1.3; Powers *et al*., 2004). While the coelacanth and

horn shark *Hox14* genes occur in different clusters (A and D, respectively), the similarity between them is strongly suggestive of the existence of the group 14 genes in the common ancestor of the gnathostomes (Powers *et al*., 2004). This finding strongly challenged the traditional view that the protovertebrates possessed only 13 *Hox* genes per HOX cluster (Powers *et al*., 2004).

In an attempt to answer questions on the possible occurrence of more *Hox* genes (e.g., chordate *Hox15*) an analysis of the amphioxus HOX cluster (650 kb), from *Hox1* to *Evx* was conducted. The data from this study postulated that based on the archetypal character of a single amphioxus HOX cluster (14 genes per cluster, Figure 3A), it is improbable that more paralogous *Hox* genes will be discovered in chordates (Minguillón *et al*., 2005). Recent analyses of complete HOX cluster sequences from the coelacanth and shark have borne this out (C.T. Amemiya, unpublished).

1.1.7 Immune System

The MHC proteins are the first receptors in the specific immune system that comes into contact with peptides of processed foreign protein (Betz *et al*., 1994). Betz and co-workers conducted studies on coelacanth genes encoding MHC class I proteins in view of three characteristics of coelacanths: firstly, its interesting evolutionary and phylogenetic position; secondly, the fact that the Comoran coelacanths live at the depths of ≈200 m where they have no known predators, and very little competition; and lastly, the availability of data from prior studies on vertebrate *MHC* genes made the coelacanth *MHC* genes good candidates for comparative studies when attempting to determine the tempo and nature of changes influencing the slow evolution of the coelacanth (Betz *et al*., 1994).

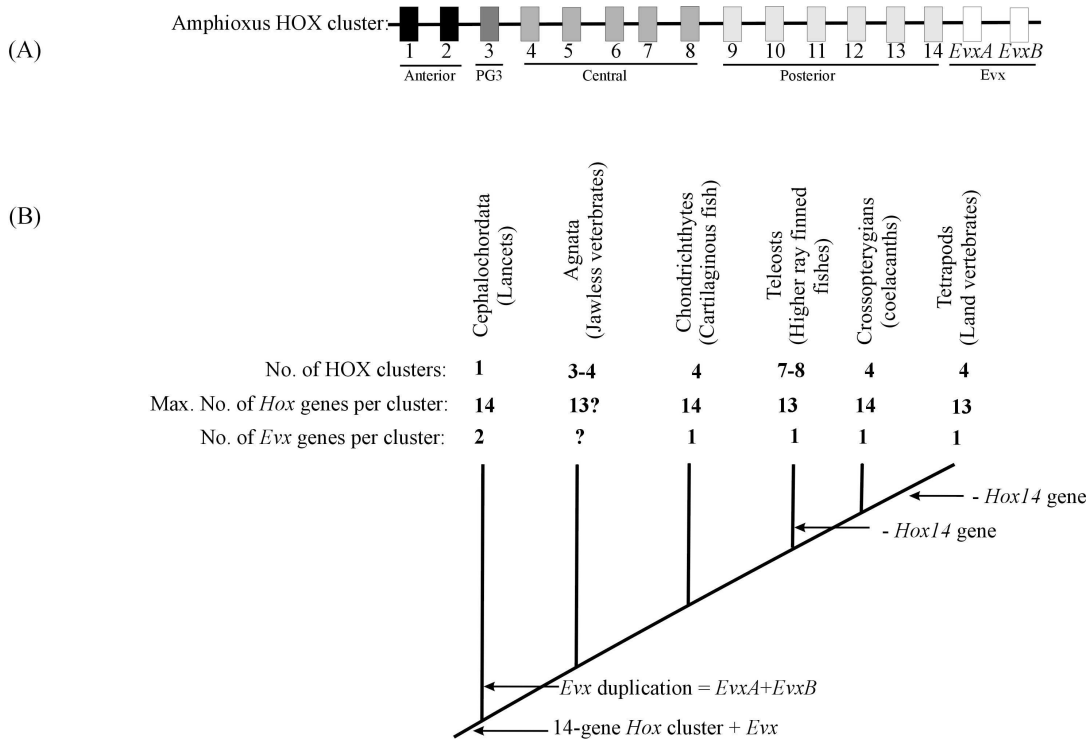


Figure 1.3: Evolution of vertebrate HOX clusters. (A) The structure of the amphioxus HOX cluster. The cluster has 14 *Hox* genes and two *Evx* genes. The genes at the 3′ end of the cluster pattern the anterior of the organism's body while the 5′ and central genes pattern the posterior and central regions of the body, respectively (Minguillón *et al*, 2005). (B) Duplication of HOX clusters and genes in the vertebrate lineage. The occurrence of one cluster in cephalochordates and four in chondrichytes, crossopterygians and tetrapods supports the hypothesis of two rounds of duplication. The occurrence of more than four clusters in higher teleosts is thought to be the result of an independent genome duplication within the teleost radiation (Wagner *et al*., 2003). When assuming the 14-gene HOX cluster in the vertebrate stem lineage, each vertebrate HOX cluster can have a maximum of 14 genes. The amphioxus cluster has no missing genes while vertebrate clusters have lost some genes. Assuming that the phylogeny is correct, teleosts and tetrapods are thought to have independently lost their *Hox14* genes (Minguillón *et al*., 2005).

The coelacanth has at least one functional *MHC* gene encoding a class I1α chain indicative of the possibility that the unique environment of the coelacanth has not led to the underdevelopment of its immune system (Betz *et al*., 1994). The authors further showed that the coelacanth *MHC* class I gene exon-intron layout was similar to that of mammals. In terms of DNA sequence similarity, the coelacanth *MHC* class I sequence was much closer to amphibian *MHC* gene than actinopterygian class I genes. These findings supported prior findings that the coelacanth encoded a functional set of genes encoding immunoglobulin proteins (Amemiya *et al*., 1993). However, the data on the

immunoglobulin heavy chain gene locus suggested a unique arrangement that, while utilizing the cardinal components of the immunoglobulin system, was organized in a completely different fashion than that of any other gnathostome (Amemiya *et al*., 1993; Stavenzer and Amemiya, 2004).

1.1.8 Adaptation of vision to the deep sea environment

In their natural habitat, coelacanths only receive a narrow band of light, approximately 480 nm (Yokoyama *et al*., 1999; Yokoyama, 2000; Yokoyama and Tada, 2000). Both species of the coelacanth have been shown to use rhodopsin (RH) 1 and 2 pigments to detect colour within a narrow range. *RH1* and *RH2* genes were cloned and characterized from species of the coelacanth (Yokoyama *et al*., 1999; Yokoyama, 2000; Yokoyama and Tada, 2000). These studies demonstrated that both species have identical RH1 and RH2 pigments with optimum light sensitivities (λ_{max}) of 485 and 479 nm, respectively, showing detection of a narrow colour range (Yokoyama *et al*., 1999; Yokoyama, 2000; Yokoyama and Tada, 2000). Compared to corresponding orthologous pigments, the maximum absorption of the coelacanth RH1 and RH2 pigments is shifted by 20 nm towards blue (Yokoyama and Tada, 2000). Ten amino acid changes have been shown to shift the maximum absorption of visual pigments by more than 5 nm. Among these, D83N, E122Q, M207L and A292S are prevalent in shifting maximum absorption of the RH1 and RH2 pigments in vertebrates. The E122Q/A292S and E122Q/M207L changes were used to explain the 20 nm shift in the maximum absorption of the coelacanths RH1 and RH2 pigments, respectively (Yokoyama *et al*., 1999; Yokoyama, 2000; Yokoyama and Tada, 2000). Most animals use ultraviolet (UV) light for activities such as foraging, mate selection, and communication (Shi and Yokoyama, 2003). Short wavelength-sensitive type 1 (SWS1) pigments absorbing light maximally at 360 nm mediate UV vision. When UV light is unavailable or not necessary the gene encoding the SWS1 pigments can be lost or become nonfunctional (Shi and Yokoyama, 2003). The coelacanth appears to have lost the SWS1 pigments and further modified the RH1 and RH2 pigments to detect light between 479-485 nm (Yokoyama and Tada, 2000).

1.1.9 The endocrine system

Proopiomelanocortin (POMC) is the precursor of a group of proteins such as adrenocorticotropin (ACTH), β-lipotropin (β-LPH), melatropin and β-endorphin (EP) which are closely associated with the response to stress and environmental adaptation (Takahashi *et al*., 2001). The diversity of POMC is suggested to have occurred as a result of the POMC gene evolution events including POMC gene internal duplication, or insertion and deletion in the region encoding their domains. The evolutionary significance of POMC also comes from the fact that it is the first gene of the adrenohypophysial hormones shown to be present in all classes of vertebrates (Takahashi *et al*., 2001).

Little is known about the components of the coelacanth's endocrine system. However, structural characterization of hormones derived from the POMC of *L. chalumnae* has been reported (Takahashi *et al*., 2003). The interest in the structure of the hormones derived from the coelacanth POMC was sparked by observations that the coelacanth's pituitary basal structure was significantly different from that of the ray-finned fish, cartilaginous fish and lampreys (Takahashi *et al*., 2003). Reverse-phase high performance liquid chromatography, amino acid sequencing and mass spectroscopic analysis of an extract from the pituitary gland identified several melanocyte-stimulating hormones (MSH) including alpha-MSH, N-Des-acetyl-alpha-MSH, beta-MSH, and an N-terminal peptide containing gamma-MSH. In addition, a corticotrophin-like intermediate lobe peptide (CLIP) and N-acetyl-beta-endorphin (END) were identified (Takahashi *et al*., 2003). Phylogenetic analysis of these peptides revealed the coexistence of putative tetrapod-type and fish-type molecules in the coelacanth POMC perhaps indicative of the coelacanth's phylogenetic intermediate position between fish and tetrpods. For instance, the gamma-MSH and CLIP molecules showed high similarity to amphibians and birds, while the beta-END was most similar to the sturgeon peptide and the alpha-MSH was highly similar to mammalian alpha-MSH. The conclusion drawn from the coelacanth POMC study was that the molecular design of the coelacanth putative POMC closely

resembled the tetrapod's POMC while its amino acid sequences is a mixture of types found in fish and tetrapods (Takahashi *et al*., 2003).

1.2 PROTEIN FOLDING AND THE MOLECULAR CHAPERONE MACHINERY

1.2.1 Protein folding *in vivo* and *in vitro*

The highly crowded and complex cellular environment renders most proteins incapable of correct folding, obligating some form of machinery to assist correct protein folding (Ellis, 2001). Using folding assays of chemically denatured protein, Anfinsen (1973) established that the information to direct the folding of a polypeptide chain was embedded in its primary amino acid sequence. The process of protein folding can be described in two steps. In the first step, the folding of the polypeptide chain is driven by the inherent nature of the polypeptide chain to place the hydrophobic amino acid side-chains into the core of the protein removing them from contact with the aqueous solution (Hartl, 1996; Frydman, 2001). The first step of polypeptide folding also involves the formation of secondary structure such as α-helices and β-sheets. In the second step of folding, the interaction of secondary structures results in the formation of tertiary structures and finally a folded and functional protein (Hartl, 1996; Frydman, 2001).

Protein folding *in vitro* is to a large degree not an ideal model for *in vivo* protein folding. However, it has advantages since it can be easily manipulated for a wide range of biophysical parameters to provide crucial clarification about the kinetics of protein folding pathways, (Fink, 1999). These *in vitro* studies on protein folding are done at high dilutions and low temperatures in homogeneous condition that do not relate to the highly heterogeneous interior of a cell (Frydman, 2001). This phenomenon has been exemplified by the finding that *in vitro* refolding of lysozyme decreases by 40% in the presence of crowding agents such as Ficoll when compared to the yield of refolded lysozyme in the absence of crowding agents (van der Berg *et al*., 2000).

The interior of the *Escherichia coli* cell has been shown to contain protein concentrations of up to 300 mg/ml (Zimmerman and Trach, 1991) resulting in the cytoplasm not behaving as an ideal fluid and influencing the kinetics of folding and reactions associated with folding. Exceeding the optimal level of macromolecular content for extended

periods of time results in aggregation of native and/or non-native proteins (Minton, 2000; van der Berg *et al*., 2000).

Another aspect of protein folding in the cell relates to the vectorial synthesis of proteins, wherein the N-terminus of the synthesized polypeptide is the first to emerge from the ribosomes while the C-terminus is synthesized. A group of proteins termed "molecular chaperones" bind the protein as it emerges from the ribosomes to avoid its misfolding and aggregation (Hartl, 1996; Frydman, 2001).

1.2.2 The molecular chaperone

The function of a protein can only be successful when that particular protein is correctly folded and is localized in its cellular compartment or it is secreted (Smith *et al.,* 1998). Failure of the protein to fold correctly may result in patho-physiological condition or even be lethal to the cell (Ellis and Pinheiro, 2002). The misfolding and aggregation of polypeptides into amyloid is characteristic of approximately 20 protein deposition diseases including type II diabetes, Alzheimer's disease and Parkinson's disease (Stefani, 2004).

Laskey and co-workers first used the term "molecular chaperones" to describe proteins called nucleoplasmins (Laskey *et al.,* 1978). The binding nucleoplasmins to histones allows only selected specific interactions to occur between histones and DNA. Subsequently, Ellis (1987) proposed the term "molecular chaperones" to depict a category of cellular proteins which interact with certain other proteins to ensure their correct folding and assembly into oligomeric structures. Heat shock proteins, particularly the 70 kDa heat shock proteins (Hsp70s), were suggested to fall within the class of such proteins (Ellis, 1987). There is now mounting evidence to the role of this class of proteins as a "regulatory element of cellular networks" in addition to their function as molecular chaperones (Sőti *et al.,* 2005). During stress conditions, chaperones deactivate and disable protein, signaling and transcriptional networks to offer more protection to the cell (Sőti *et al.,* 2005). The functions of molecular chaperones are not only restricted to their

role in stress. Under normal conditions, molecular chaperones play an important role in protein translocation, signaling and cell division and differentiation (Sőti *et al.*, 2005). Molecular chaperones, particularly heat shock protein, are also known to facilitate the degradation of highly misfolded protein by delivering to the proteasomal or lysosomal pathways (Shin *et al*., 2005)

Hsp40, Hsp60, Hsp70 and Hsp90, named according to their molecular weights, are major classes of molecular chaperones. Not only do these heat shock proteins interact with their substrate proteins but also interact amongst themselves forming protein-folding networks at some stages of protein folding in the cell (Fink, 1999).

1.2.3 Hsp90 system

The Hsp90 family of proteins is one of the most abundant cytosolic proteins, forming 2% of the yeast cytosolic total protein during growth at normal temperatures and is further up regulated under heat shock condition (Buchner, 1999; Caplan *et al*., 2003). A study conducted in *Drosophila malanogaster* and the recombinant inbred lines of *Arabidopsis thaliana*, have revealed that the inhibition of Hsp90 function resulted in a range of phenotypes, in a manner dependent on underlying genetic variations (Rutherford and Lindquist, 1998; Queltsch *et al*., 2002). Based on these studies Hsp90 was suggested to buffer genetic variations in morphogenetic pathways (Rutherford and Lindquist, 1998; Queltsch *et al*., 2002).

The interaction between Hsp90 and Hsp70 is mediated by a tetratricopeptide (TPR)-containing Hsp70/Hsp90 organizing protein (Hop) (Chen *et al.*, 1996; Hernández *et al.*, 2002; Wagele *et al*., 2003). Hop has three TPR domains of which TPR1 binds Hsp70 while TPR2a binds Hsp90 (Odunuga *et al.*, 2003; Song, *et al*., 2005). The TPR2b domain of yeast Hop has recently been shown to have overlapping or redundant functions with the TPR1 domain (Flom *et al*., 2006). The interaction between Hsp70 and Hsp90 is necessary for the folding and conformational regulation of eukaryotic proteins such as steroid hormone receptors and several signal transduction proteins. These findings

suggested that Hsp90 is a dedicated chaperone to specialized proteins (Young *et al.*, 2001; Basso *et al.*, 2002).

Hop was demonstrated to inhibit the ATPase activity of Hsp90 and simultaneously increase the Hsp70 ATPase activity. The effect of Hop on Hsp90 and Hsp70 ATPase activities is thought to be critical to its role in facilitating the delivery of substrates from Hsp70 to Hsp90 (Richter *et al*., 2003; Wagele *et al*., 2003).

1.2.4 Hsp70 system

Hsp70s exist in almost all organisms and are often found as several homologues, with some restricted to particular cellular compartments. Mammalian Hsp70s include the inducible Hsp72 and the constitutive cytosolic Hsp73 (also known as Hsp70 and Hsc70, respectively), the mitochondrial Hsp70 known as mtHsp70/glucose regulated protein (Grp) 75, and the endoplasmic Hsp70 known as Grp78, and Grp170 (Fink, 1999). The *Saccharomyces cerevisiae* cytosolic Hsp70s include Ssa1-4 and the ribosome-associated Hsp70s called Ssb1, Ssb2, Pdr13, Ssz, and Sse (Table 1.1; Wagele *et al*., 2003). While both the Hsc70 and the inducible Hsp70 are collectively referred to as Hsp70s, different genes encode these two proteins. Hsc70 proteins are encoded by intron-containing *hsp70* genes whereas Hsp70 proteins are to a large extent encoded by inducible, intronless, *hsp70* genes. The intronless feature of inducible *hsp70* genes plays an important role during the stress response since the mRNA splicing machinery is inhibited by stresses, such as heat (Yost and Lindquist, 1986; Vogel *et al.*, 1995).

In protein folding, Hsp70s play three critical roles: (i) prevention of protein aggregation, (ii) promotion of protein folding to attain the native state and (iii) the solubilization and refolding of aggregated proteins (Mayer and Bukau, 2005). To prevent the aggregation of proteins, Hsp70s and their co-chaperones, Hsp40s (Hsp40s are discussed in section 1.2.5), associate with exposed hydrophobic patches of the unfolded substrate thus preventing non-specific intermolecular interactions. This activity of Hsp70s is described as “holder” activity. The ability of Hsp70 to assist protein folding has been described in

two possible mechanisms. In the first model, the recurring binding and release of non-native substrate protein decreases the concentration of unfolded protein in the cell thus decreasing chances of aggregation and allowing unbound protein to fold to the native state (Mayer and Bukau, 2005). In an alternative model, the repetitive binding and release of non-native substrate protein causes the local unfolding of misfolded substrate to allow another attempt at productive folding, an ability described as "local unfolding" activity (Mayer and Bukau, 2005).

The role of Hsp70 in protein folding has been shown to include not only the prevention of protein aggregation but also to dissolving preformed protein aggregates (Schlieker *at al*., 2004). DnaK (prokaryotic Hsp70) has been shown to cooperate with Hsp104 to solubilize malate dehydrogenase aggregates (Schlieker *at al*., 2004). This finding was consistent with a prior proposal by Goloubinoff and co-workers that protein disagregation could be a two-phase reaction. In the first phase, aggregated protein would be solubilized by the combined effort of Hsp104 and the Hsp70 system (DnaK, DnaJ and a nucleotide exchange factor, GrpE). In the second phase, the Hsp70 system aids the folding of the solubilized protein (Goloubinoff *et al.*, 1999).

In addition to their role in protein folding, Hsp70s have been shown to aid the degradation of terminally misfolded proteins by transferring them to the ubiquitin proteasome degradation system (Lüders *et al*., 2000). CHIP (Carboxy terminus of Hsp70-interacting protein) which contains a TPR domain and a U-box has been shown to cooperate with Hsp70 to induce ubiquitination of Hsp70 substrates proteins (Shin *et al*., 2005). Futhermore, CHIP was reported to be an E3 ubiquitin ligase able of driving the ubiquitination of Hsp70 substrate proteins in a U-box depend manner (Jiang *et al*, 2001). Although the U-Box has been linked to proteasomal protein degradation, there is now evidence also linking the U-box to lysosomal protein degradation pathways (Shin *et al*., 2005). HspBP1 is a nucleotide release factor of Hsp70 that displays expression levels and tissue distribution similar to that of CHIP (Raynes and Guerriero, 1998). Furthermore, HspBP1 has been found to form ternary complexes which include Hsp70 and CHIP. In these complexes, CHIP ubiquitin ligase activity is inhibited resulting in increased folding

of substrate proteins through the Hsp70 or Hsp90 systems (Alberti *et al*., 2004). Taken together, these findings suggested a system wherein CHIP redirected Hsp70 from protein folding to protein degradation and wherein HspBP1 provided a means of balancing protein folding and degradation by inhibiting CHIP (Alberti *et al*., 2004).

The secretion of most mammalian proteins begins with their translocation into the endoplamic reticulum (ER), a process that requires the Sec61p protein complex as a translocase. The ER provides a site for the synthesis, folding, and assembly of the secretory pathway proteins (Tyedmers *et al*., 2003). The ER resident members of the Hsp70 family of chaperones, BiP and Grp170, have been shown to be involved the translocation of pre-proteins into the ER (Hamman *et al*., 1998). Additional to this function, BiP has also been shown to bind the secretory pathway protein to prevent their premature transport from the ER and to aid their folding and assembly (Tyedmers *et al*., 2003). BiP has also been demonstrated to target misfolded proteins for the ER associated protein degradation (Knittler *et al*., 1995). To carry out these functions, BiP has been shown to require co-chaperones and co-factors, interacting with the ER resident Hsp40 family of co-chaperones such ERj1p-ERj5p, and BAP (BiP associated protein) the ER resident nucleotide exchange factor (Chung *et al*., 2002).

During the translocation of mammalian proteins across the ER membrane, BiP is thought to be recruited to the Sec61p complex through its interaction with Erj2p (Sec63p), a member of the Sec61p complex (Brodsky *et al*., 1995; Tyedmers *et al*., 2000). This proposal is interesting in view of the report that Sec63p and Kar2p (BiP homolog in yeast) are required for both co- and post-translational protein translocation into the yeast ER (Brodsky *et al*., 1995; Young *et al*., 2001). Once recruited to the Sec61p complex, BiP has been demonstrated to facilitate the translocation of the protein into the ER by binding to the incoming polypeptide chain to avoid misfolding. BiP has also been shown to maintain the permeability barrier of the Sec61p complex by sealing the luminal end of translocon before and early in protein translocation (Hamman *et al*., 1998; Wirth *et al*., 2003).

1.2.4.1 Structural features of the Hsp70 ATPase domain

Hsp70 is composed of the functionally coupled 44 kDa ATPase domain and the 18 kDa substrate binding domain, and the 10 kDa C-terminal domain. Uncoupling of the ATPase domain and the substrate binding domains results in loss of chaperone function (Moro *et al.,* 2003; Splenkov and Witt, 2003).

Flaherty and co-workers determined the three dimensional structure of the 44 kDa ATPase domains of the bovine Hsc70 (Flaherty *et al*., 1990). This domain was reported to be composed of two large globular sub-domains, I and II. The nucleotide binding cleft occurs between sub-domains I and II which are further divided into sub-domains A and B (IA, IB, IIB, IIB, Figure 1.4). Although the overall structure of bovine Hsc70 was considerably different to that of hexokinase, the bovine Hsc70 ATPase domain nucleotide binding "core" highly resembled that of hexokinases (Flaherty *et al.,* 1990; Bork *et al.,* 1992). The nucleotide-binding site was found to be in its most open state when in the nucleotide free state while the ATP-bound state was in the closed conformation (Gässler *et al.,* 2001).

The chaperone function of Hsp70 depends on the cyclic shift between the ATP-bound state, with low affinity for substrate protein and the ADP-bound form with high affinity for substrate (Frydman, 2001; Mayer and Bukau, 2005). In most Hsp70 systems, ATP hydrolysis is thought to be the rate-limiting step in the ATPase cycle. A few exceptions to this system, such as in *Thermus thermophilus* DnaK (Groemping *et al.,* 2005) have been shown. The hydrolysis of ATP by the ATPase domain is coupled to Hsp70 conformational changes that cause the "locking-in" of substrate protein into the substrate binding domain. While substrate proteins can marginally increase ATP hydrolysis, Hsp40s (Hsp70 co-chaperones discussed in section 1.2.5) are known to act synergistically with the substrates to increase the Hsp70 ATPase activity by more than 2500 fold (Groemping *et al.,* 2005).

Table 1.1. The Hsp70 proteins, partner co-chaperones and associated chaperones

Bacterial	Yeast	Mammalian
Hsp70 **DnaK:** involved in the *de novo* protein folding and recovery from stress (Deuerling *et al*., 1999).	**Ssa1-4:** involved in the *de novo* protein folding and recovery from stress (Deshaies *et al*., 1987). **Ssb1 and 2:** associate with ribosomes and nascent polypeptide chains (Pfund *et al*., 1998). **Prd13/Ssz**: associate with ribosomes (Gautschi *et al*., 2001).	**Hsc70:** constitutive, binds nascent polypetides **Hsp70**: stress inducible (Fink, 1999).
Hsp40 **DnaJ:** Stimulates DnaK ATPase activity (Libereck *et al.*, 1991).	**Ydj1:** has chaperone activity. Stimulates Ssa1 ATPase activity (Caplan and Douglas, 1991). **Sis1**: ribosome associated, chaperone activity, stimulates Ssa1 ATPase activity (Zhong and Arndt, 1993). **Zuotin**: interacts with Prd1/Ssz (Gautschi *et al*., 2001).	**Hdj1 and HdJ2:** bind newly translated polypeptides (Ohtsuka, 1993; Chellaiah *et al*., 1993).
Hsp70-ADP stabilizers		**Hip:** stabilizes the Hsp70 –ADP complex. (Höhfeld *et al*., 1995).
Hsp70-associated chaperones **Trigger factor:** ribosome associated, has propyl isomerase activity, binds nascent polypeptide chains (Deuerling *et al*., 1999).	**GimC/prefoldin:** Possibly a TRiC cofactor (Siegers *et al*., 2003)	**GimC/prefoldin:** binds nascent polypeptide chains (Siegers *et al*., 2003).
Chaperonins/Hsp60 **GrEL/Hsp60:** *de novo* protein folding and recovery from stress (Fenton *and* Horwich, 1996) **GroES/Hsp10:** associates with GroEL (Fenton *et al*., 1996).	**TriC/CCT:** de novo protein folding, binds nascent chains (Frydman *et al*., 1994).	**TriC/CCT:** *de novo* protein folding, binds nascent chains (Dunn *et al*., 2001).
Nucleotide exchange factors **GrpE**: interact with DnaK (Libereck *et al.*, 1991).	**Sil1/Sls1**: enhances Hsp40 mediated ATP hydrolysis (Kabani *et al*., 2003 and 200). **Lsh1**: Hsp70-related protein, enhances of the removal of ATP and ADP (Kabani *et al*., 2003). **Fes1**: catalyzes the release of ADP and ATP from cytosolic Hsp70 (Kabani *et al*., 2003).	**Bag1-Bag5:** Facilitate Hsp70 nucleotide exchange (Höhfeld and Jentsch, 1997). **HspBP1:** Fes1 homolog (Kabani *et al*., 2002).

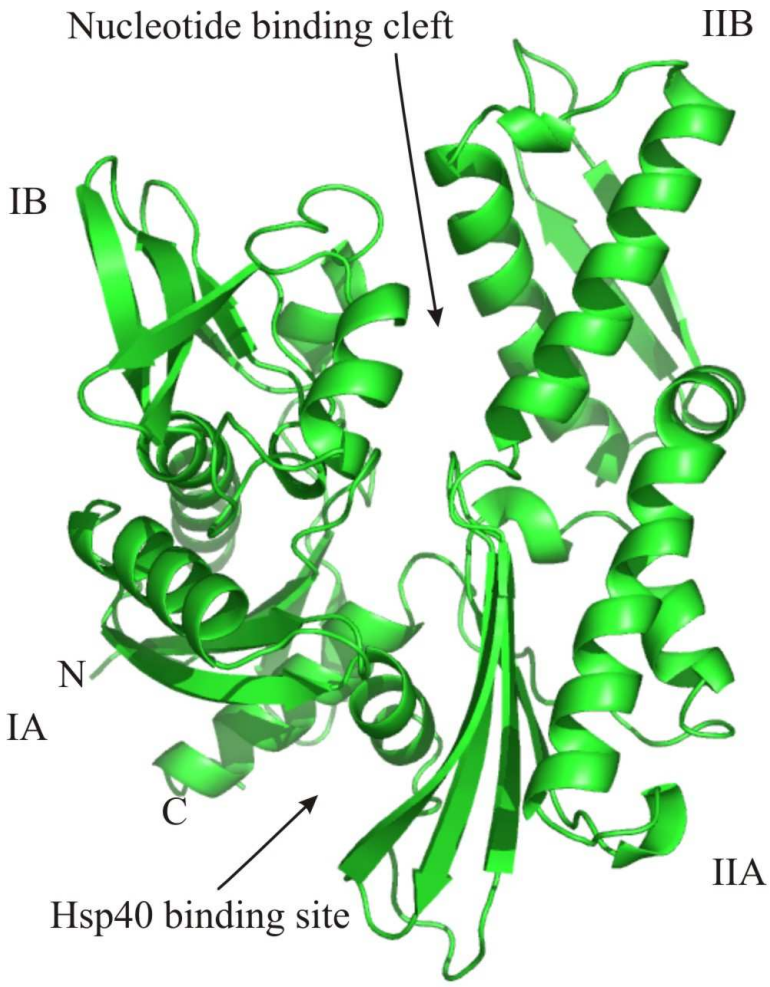


Figure 1.4: Ribbon representation of the crystal structure of the *Bus taurus* Hsc70 ATPase domain (PDB code 3HSC, Flaherty *et al.,* 1990). Subdomains are indicated as IA and IB to form lobe I; IIA and IIB to form lobe II. The amimo- and carboxy-terminal are indicated as N and C, respectively. The ribbon structure was generated using PyMOL (DeLano *et al*., 2003).

Mutational analysis of *E. coli* DnaK to determine the structural determinants for its interaction with DnaJ (prokaryotic Hsp40) showed that the stimulation of ATP hydrolysis by DnaJ requires the ATPase domain to be linked to the adjacent substrate binding domain (Gässler *et al.,* 1998). In addition, DnaJ was found to interact directly with amino acid residues in the lower backside of DnaK's ATPase domain signifying this site as a DnaJ binding site (Gässler *et al.,* 1998). The residues demonstrated to be critical for Hsp40 binding at this side included R171, N170, T173, Y149, N151, D152, E218 and V219 (Suh *et al.,* 1999; Gassler *et al.,* 1998). The residues R155, Y149 and E175 in the

ATPase domain were found to be important for the functional coupling of the ATPase and substrate binding domains, while K71 is important for ATP hydrolysis (O'Brien *et al.*, 1996). T13 was demonstrated to play a role in ATPase domain conformational change arising from the 1A as a result of the ADP/ATP binding (Sousa and McKay, 1998).

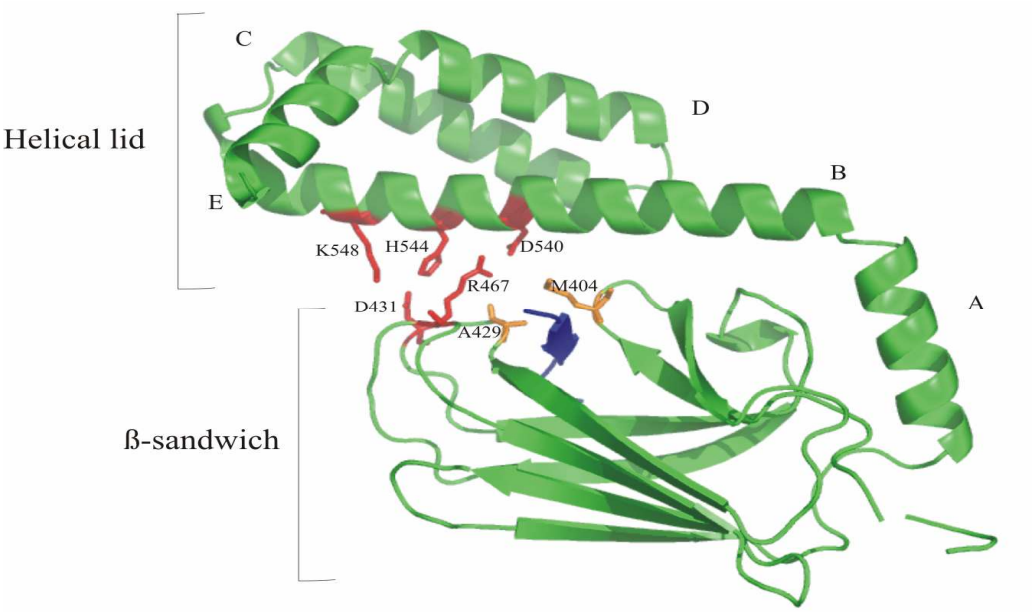


Figure 1.5: Ribbon representation of the structure of the substrate binding domain (residues 384-607) including most of the 10 kDa C-terminal domain of DnaK (PDB file 1DKX, Zhu *et al*., 1996). The bound peptide (NRLLLTG) is depicted in blue. The helical lid is depicted by helices A-E. Amino acid residues depicted in red are involved in the latching of the helical lid to the β-sandwich through a network of hydrogen bonds. Amino acid residues depicted in orange (A429 and M404) are involved in the formation of the hydrophobic arch over the bound peptide. Figure adapted from Slepenkov and Witt, 2002; Rüdiger *et al.*, 2000. The structure was generated using PyMOL (DeLano *et al*., 2003).

1.2.4.2 The substrate binding domain of Hsp70

Substrate binding occurs at the β-strand-rich region of the C-terminal 18 kDa substrate binding domain (Wang *et al.,* 1993; Zhu *et al.,* 1996). This β-strand-rich region was shown to interact directly with the substrate polypeptide (Figure 1.5; Mayer *et al.,* 2000).

Determination of the NMR structure of the substrate binding of the mammalian Hsc70 and its subsequence comparison to the crystal structure of the prokaryotic homologue, DnaK, revealed difference in the position of the helical lid (Zhu *et al.,* 1996; Morshauser *et al.,* 1999). While the position of the helix in the DnaK crystal structure involved a salt-bridge between D526 and R445, the salt bridges in the Hsc70 structure were between R445 and E518, and R414 and E520. The Hsc70 helix therefore lies in the hydrophobic groove, anchored by a salt bridge to the β-sandwich (Morshauser *et al.,* 1999). Residues M404 and A429 which formed the hydrophobic arch of DnaK substrate binding domain were implicated in substrate recognition. The mutation of these two residues to Trp resulted in decreased binding of DnaK to peptides containing stretches of consecutive hydrophobic residues flanked by acidic residues (Rüdiger *et al.,* 2000).

The determination of the Hsp70 binding motif on peptides revealed that Hsp70 recognises polypeptide residues enriched with hydrophobic amino acids, especially peptides enriched in phenylalanine, isoleucine and valine (Rüdiger *et al.,* 1997). The extended conformation of an unfolded polypeptide chain exposes these hydrophobic amino acid residues, which situated are in the interior of the folded protein. Such hydrophobic residues are predicted to be present every 40 residues in all proteins providing a potential binding site for Hsp70s (Rüdiger *et al.,* 1997).

The 10 kDa C-terminal domain of Hsp70s has been shown to form a helical structure which acts as a lid that locks the substrate in the substrate binding domain (Figure 1.5; Mayer *et al.,* 2000). The helical lid and its position must play a role in binding of substrates since its removal increases the dissociation rate of peptides and protein from DnaK (Buczynski *et al.,* 2001; Slepenkov and Witt, 2002). In DnaK the lid is implicated

in coupling the ATPase and substrate binding domains through its helices A and B (Moro *et al.,* 2003). In yeast, the lid-less variant of a mitochondrial Hsp70 (Ssc1) resulted in lethal phenotypes due to the inability to sustain preprotein import in the mitochondria, a function that was recovered in variants with the minimal lid segment, although with reduced substrate affinity (Strub *et al.,* 2003).

The crystal structure of the 10 kDa C-terminal domain of rat Hsc70 revealed that the structure this domain adopts the α-helix B-loop-α-helix C fold, significantly different to that of the equivalent region of DnaK which is composed of α- helix B to α- helix E) (Chou *et al.,* 2003; Ohno *et al.,* 2004; Figure 1.6). The Hsc70 helix C sequence was found to be equivalent to three discrete helices C-E in DnaK and that it (Hsc70 helix C) was required for Hsc70 self-association (Chou *et al.,* 2003).

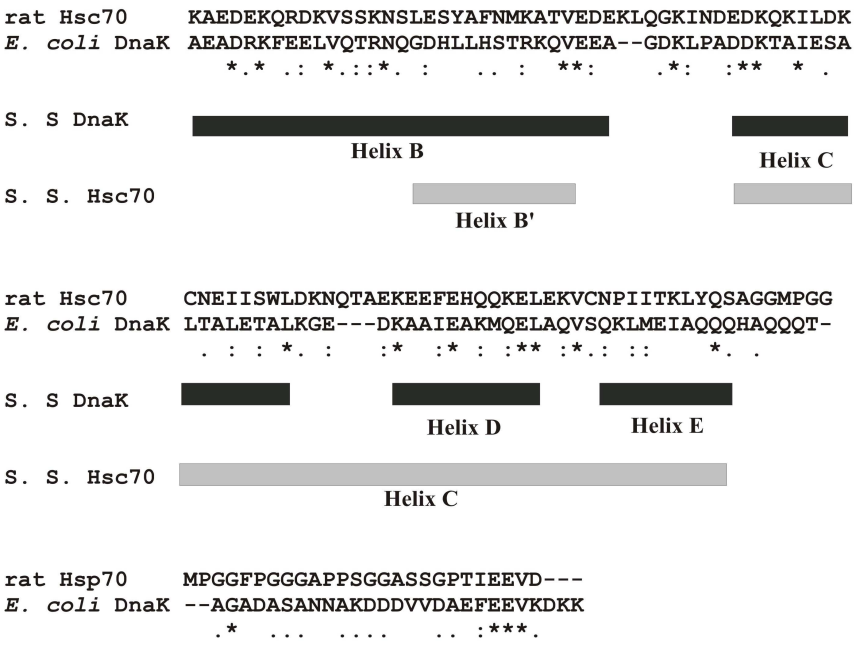


Figure 1.6: Sequence comparison of the rat Hsc70 and *E. coli* DnaK 10 kDA (DnaK residues 523-638) subdomains. Secondary structural elements (s. s) of DnaK are shown in black bars and those of the rat Hsc70 in gray bars. Genebank accession numbers are AAA4154 and BAB96589 for the rat Hsc70 and *E.coli* DnaK, respectively. Figure adapted from Chou *et al*. (2003).

1.2.4.3 Hsp70 interdomain communication

Full understanding of the mechanism of communication between the functional domains of Hsp70 have been held back by the lack of the crystal structure of the complete Hsp70. Recently, the crystal structure of the bovine Hsc70 (amino acids 1-554) that contains the nucleotide binding domain and the substrate binding domain without the 10kDa C-terminal domain was determined (Jiang *et al*., 2005).

Prior studies revealed that the linker region, connecting the ATPase and substrate binding domains (amino acids 384-394) adjacent to the interdomain interface was found in different conformations in the isolated substrate binding domains (Zhu *et al*., 1996). The authors also found that the proteolysis of this linker was reduced upon ATP binding, a finding that implicated ATP binding in the movement of the linker from a solvent exposed to hidden position (Buchberger *et al*., 1995). Consistent with these finding, mutation of the bovine Hsc70 residues L393 and V388 (Figure 1.7) reduced proteolysis of the linker, reduced ATP quenching of Trp fluorescence, and decreased the Hsc70 clathrin disassembly activity. These data indicated that the linker region plays a crucial role in Hsp70 interdomain communication where it passes on the conformational changes in the ATPase domain due to ATP/ADP binding, to the substrate binding domain. The finding that V519C becomes reactive to a thiol-specific dye upon ATP binding of suggested an ATP-dependent disruption of the interdomain interface (Jiang *et al*., 2005).

Helix A of the of the substrate binding domain (amino acids 513-524; numbering by Zhu *et al*. [1996]) was found to rest in the grove between lobes IA and IIA (referred to as the Hsp40 binding site in figures 1.6 and 1.9) of the ATPase domain (Jiang *et al*., 2005; Figure 1.7). The interdomain communication occurs at this interdomain interface and interdomain interactions between residues in the interface include salt bridges, hydrophobic interaction and water mediated hydrogen bonds (Jiang *et al*., 2005; Figure 1.7). Since helix A of the substrate binding domain rests in the Hsp40 binding grove, the binding of Hsp40 would likely result in contact with both the nucleotide binding domain

and the substrate binding domain to influence conformational changes in Hsp70 (Jiang *et al*., 2005).

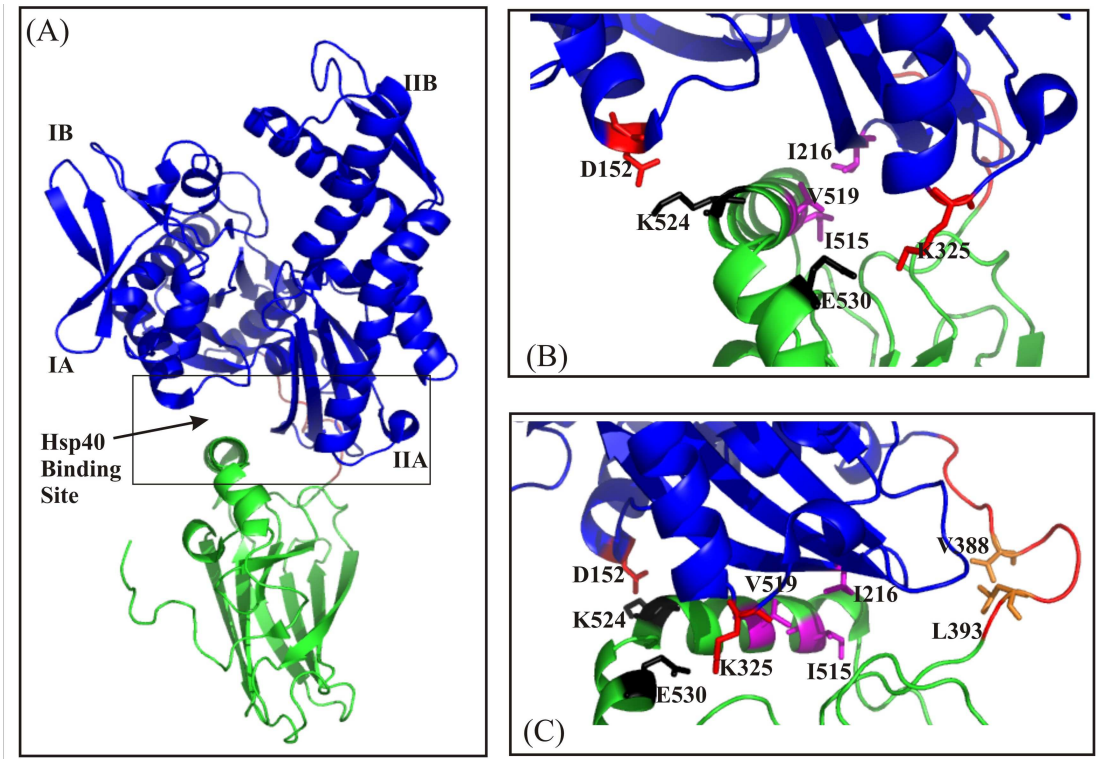


Figure 1.7: Ribbon representation of the bovine Hsc70 structure (bovine Hsc70, PDB 1YUW; Jiang *et al*., 2005). (A) Ribbon representation of the bHsc70 ATPase domain (amino acids 1-383) in blue and the substrate binding domain (amino acids 395-354) in green and the linker region (amino acids 384-394) in red. (B) Expanded view of the interaction between the substrate binding domain helix A (Green) and the groove between subdomains IA and IIA (Hsp40 binding site) in the ATPase domain (blue). Interactions include a salt bridge between residues K524 and D152, and residues E530 and K325. Hydrophobic interactions between residues I216, I515, and V519 are shown in magenta. (C) In addition to the residues in (B), residues V388 and L393 (orange) are shown on the exposed linker region (amino acids 384-394) in red. The ribbon structures were generate using PyMOL (DeLano *et al*., 2003)

1.2.5 The Hsp70 ATP/ADP cycle

Most Hsp70s require the assistance of Hsp40 co-chaperones to stimulate their weak ATPase activity and to trigger their interaction with substrate proteins (Lieberek *et al*., 1991). Members of the Hsp40 family (referred to as DnaJ in prokaryotes) are characterized by a 70 amino acid J-domain. The J-domain is made up of four α-helices with the loop region containing the His33, Pro34 and Asp35 residues (HPD) motif

occurring between helices II and III (Figure 1.8; Qian *et al*., 1996). Hsp40s binds to the interface between ATPase domain and the substrate binding domain of Hsp70s. This thereby increasing the ATPase activity of Hsp70 by specifically enhancing the ATP hydrolysis activity to generate the ADP-bound Hsp70, which has high affinity for unfolded substrate proteins (Figure 1.9; Libereck *et al.,* 1991; Russel *et al*., 1999; Jiang *et al*, 2005).

Following the activation of the Hsp70 ATPase activity, the dissociation of the bound nucleotide and the substrate protein forms the next step of the ATPase cycle which requires the opening of the Hsp70 nucleotide binding cleft. The opening of the ATPase domain of Hsp70s has been demonstrated to be facilitated by additional Hsp70 co-chaperone partners such GprE, Bag-1, or HspBP1, which regulate nucleotide exchange and substrate release (Brehmer *et al*., 2001, Brehmer *et al*., 2004; Shomura *et al*., 2005).

The Hsp70 members are subdivided into Hsp70, HscA and DnaK subfamilies based on the presence or absence of salt bridges in the ATPase domain (Brehmer *et al.*, 2001). In contrast to the DnaK family which requires GrpE to catalyse nucleotide exchange, the Hsp70 family requires Bag-1 as a nucleotide exchange factor. Bag-1 stimulates ADP, but not ATP dissociation from Hsc70 (Brehmer *et al.*, 2001). HscA does not interact with both GrpE and Bag1 (Brehmer *et al.*, 2001).

The ATPase domain of *E. coli* DnaK was demonstrated to interact with a GrpE in a 2:1 stoichiometry ratio resulting in the distortion of the ATPase domain (Libereck *et al.*, 1991; Harrison *et al.,* 1997). The conformational change induced by GrpE binding was shown to cause the mechanical opening of the ATPase domain, the disruption of the nucleotide binding site, and the release of ADP, thereby regulating the lifetime of ADP-bound DnaK (Harrison *et al.,* 1997).

(A)

(B)

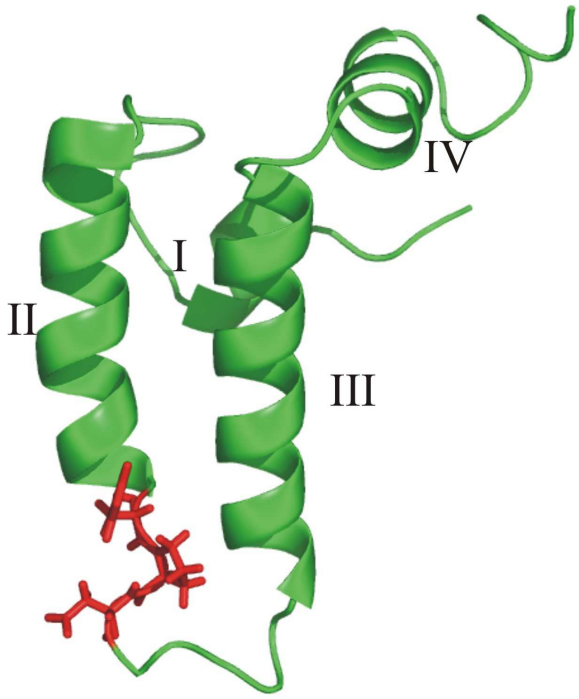


Figure 1.8. Structural representation of Hsp40s and the J-domain. (A) Schematic representation of the conserved regions of DnaJ (Hartl, 1996). G/F represents the glycine/phenylalanine rich domain. (B) Ribbon representation of the structure of the *E. coli* DnaJ J domain showing helices I-IV and the universally conserved tripeptide, His-Pro-Asp, located in the loop between helices II and III of the J-domain is shown in red (PDB 1XBL; Pellechia *et al.*, 1996). The structure was visualized using PyMol (DeLano *et al.*, 2003)

For a while the only well known eukaryotic nucleotide exchange factor was Bag-1 (Bcl2-associated athanogene 1 protein), which promotes the release of ADP from Hsp70 in a GrpE-like reaction. Bag-1 proteins contain a 40-50 amino acid BAG domain with a variable N-terminal domain (Nollen *et al.*, 2001). Resolution of structure of the bovine Hsc70 ATPase domain complex with the BAG domain revealed that the binding of Bag-1 to the ATPase domain induces the movement of residues of the IA and IIB ATPase subdomains involved in the orientation of the adenosine moiety of the nucleotide, thus opening the nucleotide binding cleft (Sondermann *et al.*, 2001). BAP (BiP-associated

protein), the ER resident nucleotide exchange factor was shown to cause nucleotide dissociation from BiP (Chung *et al.*, 2002). The identification of mammalian Hsp70-interacting protein (Hip) suggested a different, more complicated Hsp70 cycle mechanism in eukaryotes (Frydman, 2001). Hip binds to the ATPase domain of Hsp70 and prevents ADP release from Hsp70 (Höhfeld *et al.*, 1995).

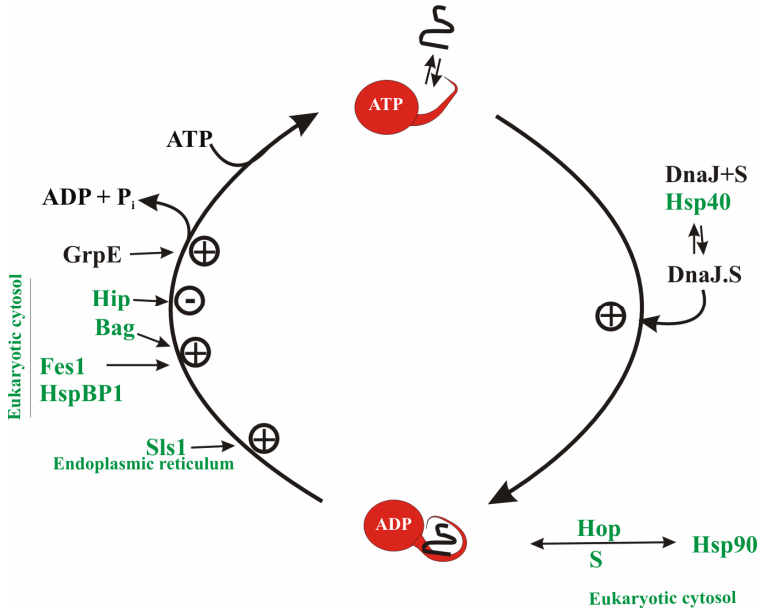


Figure 1.9. Hsp70 ATP/ADP chaperone cycle. The basic Hsp70 cycle elucidated for *E. coli* DnaK is shown in black. Modifications found in eukaryotic cytosol and endoplasmic recticulum are shown in green. Figure adapted from Mayer and Bukau (2005). S= substrate; the twisted line represents an unfolded non-native substrate protein; ATP and ADP = nucleotides; P_i = inorganic phosphate; Hip, Bag, Fes1 HspBP1, and Sls1 = nucleotide exchange factors as described in the text; Hop = Hsp70/Hsp90 organizing protein as described in the text. + = activation of a pathway; - = inhibition of a pathway.

Several other eukaryotic nucleotide exchange factors have been recently identified. In yeast, Sls1p has been identified as a nucleotide exchange factor of the endoplasmic reticulum Hsp70, Kar2 (Kabani *et al.*, 2000). Kabani and co-workers further identified the Sls1p in *D. melanogaster*, *M. musculus* and humans (Kabani *et al.*, 2003). Fes1p, a cytosolic homologue of Sls1p has been shown to bind the ADP-bound cytosolic Hsp70 (ADP-Ssa1p), stimulating its ATPase activity (Kabani *et al.*, 2002). HspBP1, a mammalian homologue of Fes1p and Sls1p was shown to promote nucleotide dissociation from Hsc70 (Figure 1.9; Kabani *et al.*, 2002).

1.3 *hsp70* GENES AND THEIR GENOMIC ORGANIZATION IN FISH

Fish are adapted to live at various ambient temperatures. They also developed mechanisms to cope with daily and seasonal temperature changes. For these reasons, fish have been considered as an ideal model to study cytoprotection, especially thermal response (Ojima and Yamashita, 2004). Most of the reports on the isolation of fish *hsp70* genes are restricted to cDNA sequences. Such sequences have been isolated from several fish species including rainbow trout (Korathy *et al.,* 1984), zebrafish (Graser *et al.,* 1996) and two species of genus *Oryzia* (Arai, 1995).

The studies to determine the genomic organization of *Fugu rubripes hsp70* genes revealed a cluster of five intron-less *hsp70* genes (Lim and Brenner, 1999). These five *Fugu hsp70* genes are combined in a head-to-tail, head-to-head and tail-to-tail arrangement in resemblance of the human *hsp70* gene orientation. On the other hand, the number of *Fugu hsp70* genes per cluster bears resemblance to non-verterbrate species. Putative heat shock elements (HSEs) were found upstream of *Fugu hsp70s.* Although no functional analysis has been done, *Fugu* HSEs has several similarities with HSEs of other organisms. They occur within 250 base pairs (bp) upstream of the start codon and in multiple copies.

Molina and coworkers (Molina *et al.,* 2000) isolated an intronless *hsp70* gene from a tilapia genomic library. The isolated *hsp70* gene included a 5′ 1 kb regulatory region and an open reading frame of 1920 bp encoding the entire Hsp70 protein. The tilapia *hsp70* promoter was able to drive the expression reporter genes in a heat shock dependent manner. Using these experiments, the authors demonstrated that the distal HSE (approximately 800 bp upstream of the transcriptional start site) was responsible for heat shock response of the tilapia *hsp70* promoter (Molina *et al.*, 2001 and 2002). In addition, a zebrafish *hsp70* promoter has been cloned and shown to drive the expression of GFP in zebrafish transgenic lines under heat shock (Halloran *et al.*, 2000).

Lele *et al*. (1997) and Santacruz *et al*. (1997) attempted to establish the number of *hsp70* genes in the zebrasfish genome. A PCR reaction with degenerate primers yielded two *hsp70* gene variants (Lele *et al*., 1997). The two genes resembled an inducible *hsp70* and a constitutively expressed *hsp70* (*hsc70*). The availability of the draft genomic sequences provides an opportunity to investigate the number and organization of *hsp70* genes in zebrafish and *Fugu*. A search by Yamashita *et al.,* (2004) in the zebrafish genome sequence revealed three *hsp70* copies. Two hsp70 genes, labeled *hsp70a* and *hsp70b* on chromosome 16 were found in tandem. Both genes were heat inducible in zebrafish embryos. The third zebrafish *hsp70* copy was found on chromosome 8 and although intronless, was shown to be constitutively expressed in zebrafish embryos. In the same study Yamashita *et al.* (2004) isolated the three different clones encoding two major isoforms of Hsp70, Hsp70-1 and Hsp70-2 and an Hsc70 from the platyfish (*Xiphophorus maculates*) cDNA. By phylogenetic analysis, platyfish Hsp70-1, *Fugu* Hsp70-1 and Hsp70 encoded in the zebrafish chromosome 16 classified together as the fish Hsp70-1 group. The fish Hsp70-2 group contained the platyfish Hsp70-2, Hsp70a and Hsp70b encoded in zebrafish chromosome 16 and the *Fugu* Hsp70-2. It therefore appears that fish have two genetically distinct groups of *hsp70* genes, i.e. *hsp70-1* and *hsp70-2* additional to the *hsc70* gene.

Genomic *hsc71* gene sequences have been obtained from the rainbow trout (Zafarulla *et al.,* 1992) and *Rivulus marmoratus* (Park *et al.,* 2001). Northern blot and primer extension analyses demonstrated that the corresponding mRNA was constitutively abundant in all trout tissues (Zafarulla *et al*., 1992). In contrast, *R. marmoratus hsc71 was* shown to be non-responsive to heat stress but selectively expressed in the muscles. Subsequent studies in *R. marmoratus* identified a novel *hsc71* gene, the promoter of which included putative metal response and muscle specific elements (Lee *et al*., 2004).

The Antarctic notothenoid *Trematotus bernacchii* has been shown to lack Hsp70 heat shock response (Hofmann *et al*., 2000). Three possible reasons for this occurrence were outlined in the study: (i) lack of interaction between HSF and the Hsp70 promoter; (ii) unstable *hsp70* mRNA that could not be translated, and (iii) the hsp70 gene was only

regulated in a constitutive manner (Hofmann *et al*., 2000; Place *et al*., 2004). Although Hsp70 was detectable *in vitro*, thermal stress could not produce a significant increase in mRNA levels (Place *et al*., 2004).

1.4 MECHANISM OF TRANSCRIPTIONAL REGULATION OF INDUCIBLE *hsp70* GENES IN HIGHER EUKARYOTES

Most of the studies on heat shock elongation control have been carried out on *Drosophila*. Although these studies provide a lot of information on heat shock inducibility, it is yet to be found if these systems are the same in all metazoans. The heat inducibility of *hsp70* promoters depends on the binding heat shock factors to the CTxGAAxxTTCxAG sequence called the of *heat shock regulatory element (HSE)* (Pelham 1982; Pelham and Bienz, 1982; Bienz, 1985). Multiple HSEs have been found 400 bp upstream of higher eukaryotic heat shock genes (Figure 1.10), the most proximal HSE occurring as close as 15-18 bp upstream to the TATA box (Bienz and Pelham, 1987). In some genes, the most proximal HSE has been found far from the TATA box and in this case the promoters have been found to contain *cis*-acting element binding sites such as the simian protein 1 (sp1) binding site and the CCAAT, suggesting a role in cell development (Bienz and Palham, 1987; Morimoto, 1992).

Heat shock factors (HSFs) regulate the transcription of heat shock genes in eukaryotes. Heat stimuli cause the trimerization of monomer HSFs, which in turn bind to the HSEs upstream of the *hsp70* promoters activating the transcription of the *hsp70* genes (Bienz and Pelham, 1987; Morimoto, 1992). Evidence for the interaction of TATA box binding protein (TBP) and HSF came from the finding that TBP could be recovered from *Drosophila* nuclear extracts using affinity chromatography with HSF as a ligand. In addition, HSF and TBP were demonstrated to bind cooperatively to DNA (Mason and Lis, 1997).

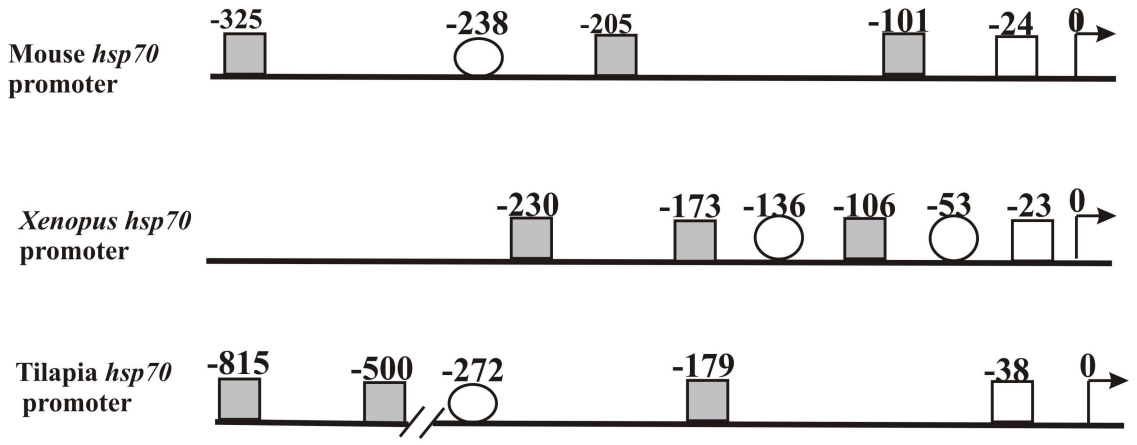


Figure 1.10: **Comparison of inducible *hsp70* promoters**. Comparison of the mouse (Fiszer-Kierzkowska *et al.*, 2003), *Xenopus* (Bienz, 1994) and tilapia (Molina *et al.*, 2000) *hsp70* heat inducible promoter. Not to scale. The filled squares represent the HSEs, the clear square represents the TATA box, the arrow represents the transcriptional start site, the circles represent the CAAT box, and the numbers indicates the position of the indicated elements upstream of the transcription start site.

Both TATA–box binding transcription factor II D (TFIID) and GAGA factor (also known as the Adh transcription factor-2) have been shown to occupy the *hsp26* and *hsp70* promoters prior to heat shock, positioned to aid the recruitment of HSFs and aiding the recruitment of HSFs upon stress (Wu, 1980, Mason and Lis, 1997). A pre-initiation RNA polymerase II complex in elongation mode but paused has also been found on promoters of heat shock genes (Rougvie and Lis, 1990). The T residues of the TATA box were found to be protected from $KMnO_4$ in both uninduced and induced cells suggesting that TFIID was continually bound to the TATA box for fast recruitment of the polymerase (Giardina *et al.*, 1992).

Rasmussen and Lis (1995) further proposed three models describing the pausing of the polymerase complex in uninduced genes and its "escape" in induced genes.

Model 1

Under non-heat shock conditions the elongation competent RNA polymerase is recruited to the promoter through its interaction with promoter bound TFIID and other components of the basal transcription machinery. Although the polymerase II complex is competent, its interactive force with the promoter-bound transcription factors restricts further

movement of the polymerase. HSF binding is therefore thought to out-compete the interactive force between the polymerase and the restrictive factor causing the escape of the polymerase and the transcription of *hsp70* genes (Mason and Lis, 1997).

Model 2

This model is similar to the first model, however, the pausing effect on the RNA polymerase is not due to the restriction placed by interaction of the polymerase with promoter-bound transcription factors. The RNA polymerase II in this case is paused because the RNA polymerase is elongationally incompetent. Support of this model came from the observation by O'Brien *et al.* (1994) and subsequent studies (Komarnitsky *et al.,* 2000; Cho *et al.,* 2001) that paused RNA polymerase complexes in a number of *Drosophila* genes are unphosphorylated, while the transcribing polymerases are hyperphosphorylated.

Model 3

This model resembles the second model with the difference that RNA polymerase II is not modified to the elongationally competent form upon heat shock. Upon heat shock, a new RNA polymerase complex, which is elongationally competent, is assembled at the promoter and the paused elongationally incompetent complex is dissociated or displaced from the promoter (Rasmussen and Lis, 1995).

The possibility of cooperation between the different models to regulate the induction of Hsp70s cannot be excluded. Transcription in eukaryotes has distinct regulatory points at which control can be exerted. The finding that RNA polymerase interaction with general transcriptional factors additional to HSFs could be indicative of the requirement for multiple levels of control of the induction of *hsp70* (Mason and Lis, 1997). Another interesting result came from a study wherein the yeast Gal4 protein (Gal4p) was shown to induce transcription of the *Drosophila hsp70* promoter in which the Gal4p binding site was substituted for a GAGA factor and the HSF binding site upstream of the TATA element (Tang *et al.,* 2000). In addition, Gal4p also caused the escape of RNA

polymerase II from its paused position (Tang *et al.,* 2000). This data support of the proposal that the pausing and escape of the RNA polymerase II depends on its interaction with regulatory elements upstream of the *hsp70* core promoter (Tang *et al.,* 2000).

There is growing evidence in support of model 2. The RNA polymerase paused in the *hsp70* promoter is in the IIa state and is converted to the IIo state shortly after heat stress. Kinases such as the positive transcription elongation factor b (P-TEFb) or a general transcription factor such as transcription factor IID (TFIIH) are thought to play a role in the phosphorylation of RNA polymerase II (O'Brien *et al*., 1994; Lis *et al*., 2000; Fan et al, 2006)

Wu and co-workers proposed a model in which the recently discovered the 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) sensitivity-inducing factor (DSIF) and the negative elongation factor (NELF) inhibit the IIa phosphorylation. The binding of P-TEFb is suggested to displace NELF from the complex and to relieve complex inhibition by phosphorylating the polymerase. While DSIF and NELF localize to the uninduced *hsp70* promoter, only DSIF and the polymerase are recruited to the chromosomal puff following heat shock treatment (Wu *et al*., 2003). It is further proposed that the binding of a HSF causes the dissociation of NELF and subsequence phosphorylation of the polymerase (Wu *et al*., 2003, Figure 1.11). Although, no direct interaction between HSF and P-TEFb has been observed, P-TEFb has been shown to interact with the *hsp70* gene during heat shock (Lis *et al.,* 2000).

In contrast to the well recognised HSF-based induction of inducible *hsp70*s, very little is known about the developmental regulation of *hsp70* promoters by transcriptional factors such as B-myc, C-myc, or p53. The need to elucidate these mechanisms necessitates an in depth study of the DNA regions flanking inducible *hsp70* genes (Fiszer-Kierzkowska *et al*., 2003).

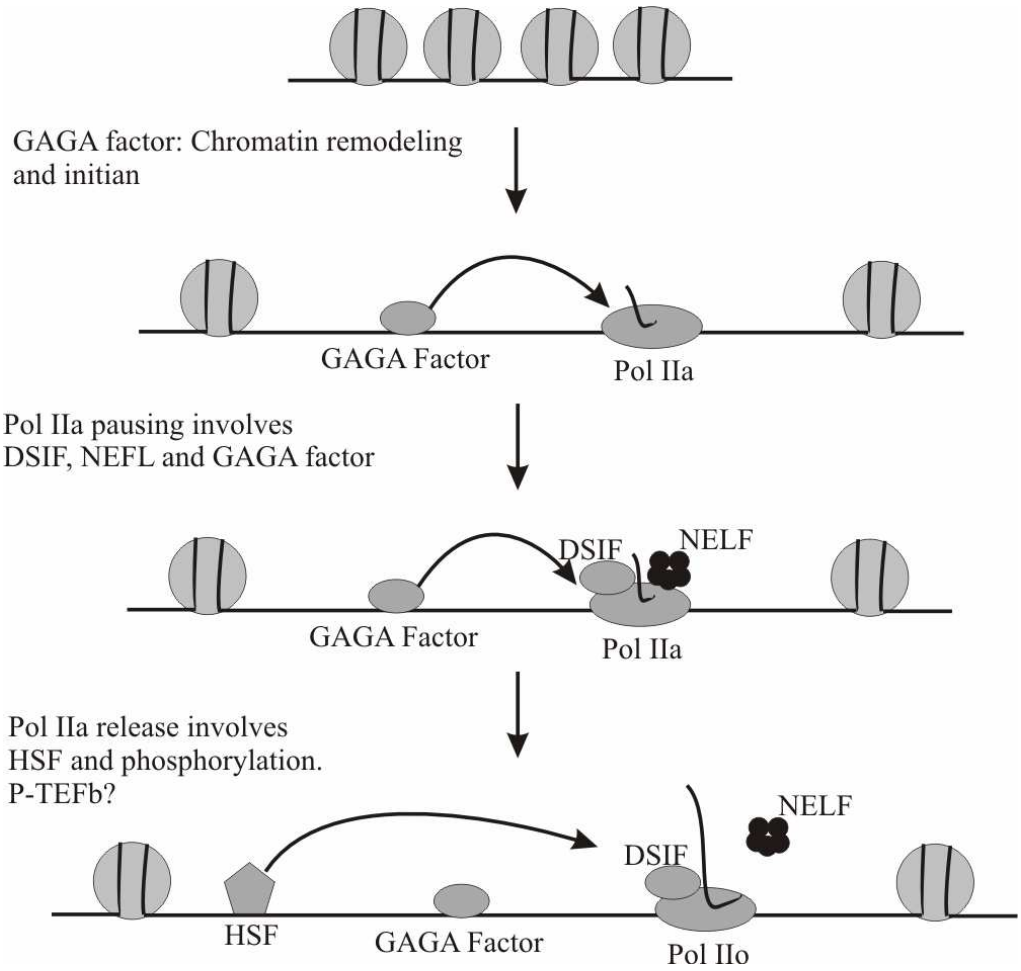


Figure 1.11: Model for *hsp70* transcriptional regulation. The binding of the GAGA factor causes chromatin remodelling and initiation. GAGA factor is also possibly involved in Pol IIa pausing. DSIF and NEFL associate with the Pol IIa shortly after initiation and NELF interacts with the transcript when it is 20-40 nucleotide long, restricted further elongation by Pol IIa. Binding of HSF cause the release of NEFL from the complex and mediates the phosphorylation of Pol IIa by P-TEFb to form the elongation competent Pol IIo. The grey spheres represent histones with the black line joining and wrapping around the shares as chromosomal DNA. (Adapted from Wu *et al*., 2003).

1.5 MOTIVATION AND RESEARCH HYPOTHESIS

At present, there is no information, both at DNA and protein levels, regarding the existence of heat shock proteins as molecular chaperones in the coelacanth. The presence/absence of a heat shock based cytoprotection mechanism in the coelacanth is of particular interest since the coelacanth has occupied its present habitat for a long period of time with minimal challenges, a fact that could have lead to genetic uniformity. In addition there appears to be little evolutionary pressure on the coelacanth as suggested by its ability to have survived extinction for millions of years with minimal changes to its biology. The role of Hsp70s, essential but not limited to *de novo* protein folding, protein degradation, and protein translocation motivates for the existence of the genes encoding these molecular chaperones in the genome of the coelacanth. In order to establish the existence of a heat shock protein based cytoprotection mechanism in the coelacanth, we embarked on a study of genes encoding heat shock protein 70s in the genome of both species of the coelacanth. This study is designed to investigate genes encoding Hsp70s in the coelacanth based on the following hypothesis:

'The coelacanth has highly conserved inducible and constitutive Hsp70s, the structure and function of which will have certain specialized features that will reflect the unique biology and habitat of the coelacanth'.

The broad specific objectives emanating from this hypothesis are:

(i) To isolate inducible *hsp70* genes or parts thereof from *L. chalumnae* and *L. menadoensis*, to facilitate an investigation of the structure and function of the genes and gene products.

(ii) To isolate the regulatory regions upstream of these *hsp70*s to facilitate an investigation of the essential promoter elements.

Chapter 2

Isolation of partial fragments of *Latimeria chalumnae hsp70* genes (*Lchsp70s*)

2.1 Introduction

Several genes encoding Hsp70s in fish have been cloned and characterized including a complete 1920 bp intronless coding region of the tilapia *hsp70* (Molina *et al*., 2000). This intronless characteristic is important for stress-induced genes since RNA-splicing is inhibited during stress (Yost and Lindquist, 1986; Vogel *et al*., 1995). Another interesting feature of *hsp70* genes is that they occur in multiple copy numbers raising questions about their functional significance. A cluster of five intronless *hsp70* genes has been identified on the genome of the pufferfish, *Fugu rubripes* (Lim and Brenner, 1999). In addition, constitutive *hsp70* (*hsc70)* genes have been cloned from the zebrafish (Santacruz *et al*., 1997; Graser *et al*., 1996), rainbow trout (Zafarulla *et al*., 1992) and *Rivulus marmoratus* (Park *et al*., 2001), while inducible *hsp70* genes have been cloned from two fish species of the genus *Oryzias* (Arai *et al*., 1995), zebrafish embyos (Lele *et al*., 1997) and rainbow trout (Korathy *et al*., 1984)

The fact that fish are ectothermic and occupy a highly temperature conductive aquatic environment makes their ability to respond to temperature variation an important factor influencing their physiology (Basu *et al*., 2002). For this reason fish could be suitable model organisms to investigate the regulation, functional significance and the evolutionary genomics of heat shock proteins (Basu *et al*., 2002).

Studies conducted on aquatic organisms that have been adapted to low temperature environments have revealed a general loss of the ability to acquire thermotolerance (Brennecke *et al*., 1998). Studies in the cold water adapted *Hydra oligactis* revealed that this species transcribed much less *hsp70* mRNA during heat stress as opposed to the wide temperature adapted *Hydra bulgaris* or *Hydra magnipapillata* species (Gellner *et al*., 1992). Despite the difference in *hsp70* gene induction, Brennecke and

co-workers demonstrated that the *H. magnipapillata hsp70* regulatory region was able to drive the expression of luciferase from a reporter gene construct in *H. oligactis* and *H. magnipapillata* polyps in a heat inducible manner (Brennecke *et al*., 1998). It was therefore concluded that the *H. oligactis* heat shock factors interacted productively with the mentioned *H. magnipapillata hsp70* regulatory region. An alternative mechanism responsible for differential expression of *hsp70* genes in these *Hydra* species was confirmed by the finding that the half-life of *hsp70* mRNA in heat-shocked *H. oligactis* was significantly shorter than in *H. magnipapillata* (Brennecke *et al*., 1998).

An equally interesting study was conducted in the highly cold-adapted, Antarctic teleost fish, *Trematomus bernacchii* (Hofmann *et al*., 2000). This study revealed that although the constitutive *hsp70* was detectable without heat shock treatment, heat stress did not result in the upregulation of the heat inducible *hsp70* gene (Hofmann *et al*., 2000). The authors speculated that the lack of thermal response in this fish was due to the absence of positive selection during evolution at stable sub-zero temperatures and that this could have been due to loss of functional *hsp70* genes and heat shock factors or the instability of the produced mRNA (Hofmann *et al*., 2000, Place *et al*., 2004).

Latimeria is described as a living fossil since it is the only extant member of the coelacanths living in rocky caves 100-300 m below sea level (Heemstra, 2001). Since *Latimeria* species have survived million of years with little morphological change, the questions relating to the rate of change in coelacanth genomes and cell biology remain intriguing. Given the evolutionary and phylogenetic significance of the coelacanth, studies on the genomic structure of genes encoding heat shock proteins in the coelacanth could be stepping stone to set testable prediction about the function and regulation of *hsp70*s in fish. Since it was understandable from the abovementioned studies that it cannot be assumed that a typical *hsp70* system occurs in aquatic organisms including the coelacanth, the aim of this study was therefore to isolate and analyse by bioinformatics the coelacanth inducible *hsp70*s to make predictions about their functions.

2.2 Experimental procedures

2.2.1 Isolation of high molecular weight DNA from *L. chalumnae* skin tissue

High molecular weight DNA (> 200 kb) was isolated from frozen *L. chalumnae* skin tissue obtained by the proteinase K-formamide DNA isolation method with minimal sheering (Sambrook and Russell, 2001). This procedure involves proteinase K digestion of the tissue followed by formamide treatment to effect the dissociation of DNA-protein complex. The proteinase K and formamide were subsequently removed by extensive dialysis. Detailed method is provided in Appendix A1.

The confirmation that the DNA obtained was isolated from skin tissue of *L. chalumnae* origin was carried out by polymerase chain reaction (PCR) based isolation of the 220 bp of the coelacanth *Hoxa-11* sequence. The *L. chalumane Hoxa-11* specific forward primer (Hoxa-11F/2: 5'-CCATTGCTACTCAACAGG–3') and reverse primer (Hoxa-11R/2: 5'- AGAGTGGTTTTCTGTGC - 3') based on the published *Hoxa-11* sequence (AF287139; Chui *et al*, 2000) were used in the PCR.

2.2.2 Polymerase chain reaction (PCR)

The Expand High Fidelity PCR kit (Roche, Germany) was used for all experiments (Appendix A3). Thermal cycling was performed with a GeneAmp PCR System 9700 (ABI, USA). The amplification conditions were as follows: 5 minutes denaturation at 95˚C followed by 35 cycles of 1 minute denaturation at 95˚C, 1 minute annealing at 45˚C, 2 minutes extension at 72˚C and a final extension at 72˚C for 10 minutes.

All PCR products were ligated into the pGEM-T Easy vector (Promega, USA), transformed into *Escherichia coli* XLI Blue and plated onto 2xYT (16 g/L tryptone, 10 g/L yeast extract, and 5 g/L NaCl) agar plates containing 100 µg/ml ampicillin (Appendix A4 and A5).

Plasmid DNA was isolated from ampicillin-resistant transformants using the QIAprep plasmid isolation kit (Qiagen, Germany). The presence of inserts was determined by *Eco* RI restriction analysis (Appendix A6). T7 and Sp6 sequencing primers were used

to generate DNA sequences using the chain termination-based cycle sequencing reactions with BigDye V3.1 (ABI, USA). Cycle sequencing products were electrophoresed using the Applied Biosystems 3100 Genetic Analyzer (ABI, USA) according to the manufacturer's instructions.

2.2.3 *L. chalumnae* codon usage table

Seven *L. chalumnae* coding regions were obtained from the protein sequence data bank at the National Center for Biotechnology Information (NCBI): α-enolase and β-enolase (AY005154 and AY005155; Tracy and Hedges, 2000); MHC1 mRNAs (U08036 and U08037; Betz *et al*., 1994); olfactory receptor genes (AJ233783 and AJ233784; Freitag *et al*., 1998); and the *Hoxa-11* gene (AF287139; Chui *et al*, 2000). The average codon usage of the obtained coding regions was obtained and a *L. chalumnae* codon usage table was compiled.

2.2.4 Fish Hsp70 alignments and Hsp70 primer design

To facilitate the design of primers to amplify coelacanth *hsp70* coding regions, an alignment of Hsp70 proteins from *Fugu rubripes* (CAA69894); tilapia (CAA04673); zebrafish (BAB72170); rainbow trout (P08108); and bovine Hsc70 (P19120) was carried out. All the protein sequences were obtained from the protein sequence data bank at the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nih.gov). Conserved regions were identified and the amino acid sequence of the selected regions was back translated to nucleotide sequence using the *L. chalumnae* codon usage designed in this thesis.

2.2.5 Sequence and bioinformatic analysis

Sequences were analyzed with the Vector NTI sequence analysis software (InforMax, USA) and searches were performed using the Basic Local Alignment Search Tool (BLAST) software (Altschul *et al*., 1990) at the NCBI website (http://www.ncbi.nih.gov/blast). Multiple sequence alignments were carried out using the ClustalW software (Thompson *et al*., 1994) at the BCM search launcher browser (http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html).

Ribbon representation of the LcHsp70 homology model was generated using SWISSMODEL software (Peitsch *et al*., 1996) using bovine Hsc70 (PDB code: 1YUW; Jiang *et al*., 2005) as template. The structure was visualized using PyMol software (DeLano *et al*, 2005).

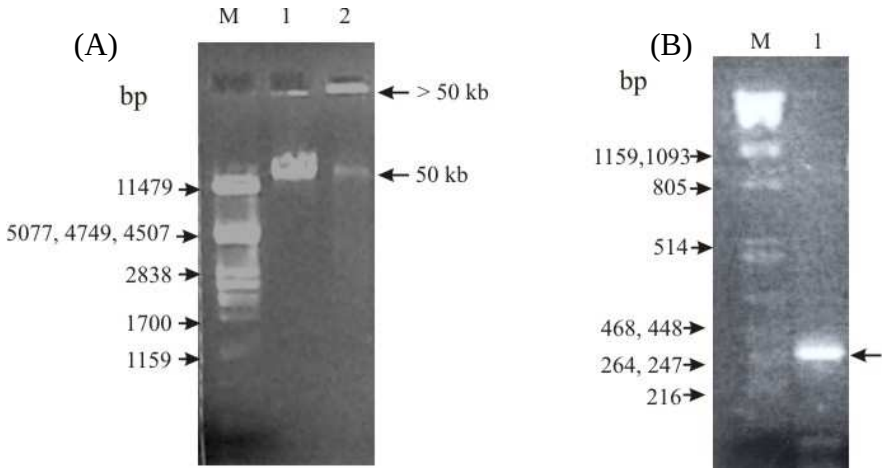
2.2.6 Phylogenetic analysis

The Hsp70 protein sequences used were obtained from NCBI under the following accession numbers: Japanese medaka, AAF91485.1; southern platyfish, BAB72168.1; tilapia CAA04573.1; fugu, CAA69894; zebrafish, BAB72170; rainbow trout, BAB72233.1, Chinook salmon, AAA78276.1; frog, CAA 25576.1; rat, NP_114177.1; bovine, NP_776975; human, NP_002146; *E. coli*, NP_285706; and *Shigela flexneri*, NP_835755.1. The Hsp70 protein sequences were aligned using CLUSTALW (Thompson *et al*., 1994) at default settings. The region from I28 to L650 (according to the numbering to the Bovine Hsc70) was used to generate the tree. The alignment was visually inspected and all proteins were adjusted to the same starting point. A Neighbor-Joining (NJ) tree was constructed with MEGA3 (Molecular Evolutionary Genetics Analysis 3; Nei, 2004).

2.3 Results

2.3.1 Isolation of *L. chalumnae* high molecular weight DNA

High molecular weight genomic DNA, 50 kb and larger, was successfully isolated from *L. chalumnae* frozen tissue (Figure 2.1A, lane 2). The PCR isolation of the *L. chalumnae Hoxa-11* fragment with the isolated DNA as template yielded a 220 bp fragment (Figure 2.1B, lane 1). The DNA sequence obtained for the 220 bp *L. chalumnae Hoxa-11* fragment demonstrate a 100% identity to the same region of the published *L. chalumnae Hoxa-11* sequence (Figure 2.1C).



(C)

```
Pub  1    CCATTGCTACTCAACAGAGGAGATTCTGCACAGGGACTGCCTAGCAACCACCACCGCTTC  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PCR  201  CCATTGCTACTCAACAGAGGAGATTCTGCACAGGGACTGCCTAGCAACCACCACCGCTTC  260

Pub  61   AAGCATAGGAGAAATCTTTGGGAAAGGCAACGCTAACGTCTACCATCCTGGCTCCAGCAC  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PCR  261  AAGCATAGGAGAAATCTTTGGGAAAGGCAACGCTAACGTCTACCATCCTGGCTCCAGCAC  320

Pub  121  CTCTTCTAATTTCTATAACACAGTGGGTAGAAACGGGGTCCTACCGCAAGCCTTTGACCA  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PCR  321  CTCTTCTAATTTCTATAACACAGTGGGTAGAAACGGGGTCCTACCGCAAGCCTTTGACCA  380

Pub  181  GTTTTTCGAGACGGCTTATGGCACAACAGAAAACCACTCT  220
          |||||||||||||||||||||||||||||||||||||||
PCR  381  GTTTTTCGAGACGGCTTATGGCACAACAGAAAACCACTCT  420
```

Figure 2.1: Isolation and validation of *L. chalumnae* high molecular genomic DNA. (A) A 0.6% agarose gel resolution of *L. chalumnae* high molecular weight chromosomal DNA. Lane M = λ *Pst* I size marker; lane 1 = uncut λ DNA; lane 2 = *L. chalumnae* high molecular weight DNA (50 kb and above). (B) A 2% agarose gel resolution of the PCR products amplified from 100ηg of *L. chalumnae* high molecular weight DNA (shown in Figure IA). Lane M = λ *Pst* I size marker; lane 1 = 220 bp *Hoxa-11* gene fragment is indicated by the arrow. (C) Alignment of the published coelacanth *Hoxa-11* (Pub) and the 220 bp *Hoax-11* fragment isolated from *L. chalumnae* frozen skin tissue (PCR).

2.3.2 Primer design to isolate the gene encoding the *L. chalumnae* Hsp70

The in-frame usage of each codon in the coding region of the α-enolase and β-enolase (AY005154 and AY005155; Tracy and Hedges, 2000); MHC1 mRNAs (U08036 and U08037; Betz *et al*., 1994); olfactory receptor genes (AJ233783 and AJ233784; Freitag *et al*., 1998); and the *Hoxa-11* gene (AF287139; Chui *et al*, 2000) was determined. The average codon usage was expressed in frequency/thousand base pairs (Table 2.1).

Table 2.1. *L. chalumnae* codon usage table expressed in frequency/thousand base pairs.

Amino acid	L						I			V				S						W	M
Codon	UUA	UUG	CUU	CUC	CUA	CUG	AUU	AUC	AUA	GUU	GUC	GUA	GUG	UCU	UCC	UCA	UCG	AGU	AGC	UGG	AUG
Gene																					
Alpha Enolase	5.49	19.23	16.5	8.24	0	38.46	38.5	13.7	0	30.2	19.2	5.49	30.22	16.48	10.99	13.7	0	11	2.75	8.24	16.5
Beta Enolase	0	0	16.7	8.33	8.33	25	41.7	41.7	8.33	33.3	0	16.7	33.33	16.67	8.33	0	0	0	16.7	25	8.33
MHC1.TMC2g mRNA	0	25.42	0	25.4	0	42.37	59.3	17	8.47	50.9	33.9	0	8.47	0	0	0	17	8.47	17	25.42	25.4
MHC1.TMC2a mRNA	0	25.42	0	25.4	0	50.85	67	8.47	17	42.4	33.9	0	8.47	0	0	8.47	8.47	8.47	25.4	25.42	25.4
Olifactory Gene 1	4.55	31.82	22.7	59.1	4.55	22.73	18.2	45.5	9.09	9.09	27.3	4.55	45.45	13.64	36.36	22.7	0	18.2	13.6	4.55	36.4
Olifactory Gene2	18.75	18.75	12.5	12.5	25	37.5	37.5	43.8	43.8	34.3	18.8	12.5	6.25	37.5	18.75	25	0	0	0	12.5	25
Hoxa-11 Gene	4.95	9.9	0	4.95	14.9	4.95	9.9	4.95	9.9	14.9	9.9	0	9.9	54.46	34.95	19.8	19.8	19.8	34.7	4.95	4.95
Average usage	4.82	18.65	9.77	20.6	7.53	31.694	38.9	25	13.8	30.7	20.4	5.6	20.3	19.82	15.63	12.8	6.46	9.42	15.7	15.154	20.3

Amino acid	P				T				A				Q		N		K		Y		
Codon	CCU	CCC	CCA	CCG	ACU	ACC	ACA	ACG	GCU	GCC	GCA	GCG	CAA	CAG	AAU	AAC	AAA	AAG	UAU	UAC	
Gene																					
Alpha Enolase	8.24	6.41	22	2.74	22	5.49	11	0	55	13.7	33	2.75	5.49	16.48	30.22	24.7	60.4	30.2	11	19.23	
Beta Enolase	8.33	0	16.7	0	25	25	33.3	0	8.33	16.7	25	0	16.67	8.33	33.33	33.3	33.3	16.7	0	25	
MHC1.TMC2g mRNA	8.47	11.26	17	5.63	8.47	50.85	33.9	0	17	8.47	17	8.47	0	25.42	16.95	25.4	25.4	17	8.47	25.42	
MHC1.TMC2a mRNA	0	14.08	17	8.45	8.47	50.85	17	0	25.4	0	8.47	17	0	25.42	25.42	0	17	8.47	8.47	25.42	
Olifactory Gene 1	18.18	0	0	2.05	22.7	27.27	18.2	0	31.8	36.4	18.2	0	4.55	18.18	4.55	9.09	13.6	13.6	31.8	18.18	
Olifactory Gene2	31.25	10.26	25.5	0	18.8	18.75	25	0	0	6.25	6.25	0	18.75	0	6.25	25	18.8	12.5	37.5	25	
Hoxa-11 Gene	19.8	8.21	9.9	10.3	9.9	34.65	39.6	4.95	15	9.9	19.8	0	19.8	9.9	29.75	29.7	19.8	19.8	19.8	34.65	
Average usage	13.47	7.174	15.4	4.16	16.5	30.409	25.4	0.71	21.8	13.1	18.2	4.02	9.323	14.82	20.92	21	26.9	16.9	16.7	24.7	

Amino acid	R						H		G				C		D		E		F		
Codon	CGU	CGC	CGA	CGG	AGA	AGG	CAU	CAC	GGU	GGC	GGA	GGG	UGC	UGU	GAU	GAC	GAA	GAG	UUU	UUC	
Gene																					
Alpha Enolase	8.24	10.99	0	0	8.24	0	5.49	5.49	44	13.7	24.7	5.49	8.24	5.49	43.96	22	35.7	33	22	16.48	
Beta Enolase	8.33	8.33	0	8.33	8.33	0	41.7	0	50	16.7	16.7	8.33	8.33	16.67	50	25	16.7	33.3	16.7	25	
MHC1.TMC2g mRNA	0	0	17	0	8.47	25.42	8.47	17	8.47	0	33.9	17	8.47	8.47	42.37	25.4	8.47	42.4	8.47	16.95	
MHC1.TMC2a mRNA	0	0	8.47	0	8.47	25.42	8.47	17	0	0	33.9	17	8.47	8.47	42.37	59.4	17	50.9	8.47	16.95	
Olifactory Gene 1	0	9.09	4.55	0	0	27.27	0	27.3	22.7	9.09	9.09	18.2	27.27	22.73	18.18	18.2	9.09	4.55	18.2	40.91	
Olifactory Gene2	6.25	0	0	6.25	6.25	12.5	25	18.8	12.5	6.25	0	12.5	25	37.5	12.5	6.25	6.25	6.25	43.8	37.54	
Hoxa-11 Gene	4.95	4.95	0	4.95	19.8	29.7	14.9	19.8	15	19.8	4.95	9.9	24.75	0	9.9	34.7	24.8	49.5	19.8	19.8	
Average usage	3.967	4.766	4.28	2.79	8.51	17.187	14.9	15	21.8	9.36	17.6	12.6	15.79	14.19	31.33	27.3	16.8	31.4	19.6	24.804	

An alignment of *F. rubripes*; tilapia, zebrafish, rainbow trout Hsp70s and bovine Hsc70 protein sequences revealed high level of conservation. A conserved amino acid sequence of GIDLGTT was identified at the N-terminal region while the C-terminus had a conserved GPTIEEVD (Figure 2.2).

Two degenerate DNA PCR primers, Fish70F/2 and Fish70R/2, were designed based on these GIDLGTT and GPTIEEVD amino acid sequences, respectively. A *L. chalumnae* codon usage table designed from eight *L. chalumnae* coding region (Table 2.1) was used to back translate the amino acid sequences (Figure 2.2).

Figure 2.2: The design of degenerate fish *hsp70* primers from the consensus sequence of aligned *Fugu*, tilapia, zebrafish, rainbow trout and bovine 70 kDa heat shock proteins. The conserved amino acids at the N- and C- termini of the Hsp70s were back translated to DNA sequences using the *L. chalumnae* codon usage table. (A) Hsp70 N-terminus was used to design a degenerate Fish70F/2 primer. (B) Hsp70 C-terminus was used to design a degenerate Fish70R/2 primer. W=T/A; Y= T/C; S=G/C; M= C/A; R= A/G; V=A/G/C; and K=G/T. Identical or similar residues are indicated by * or : or . .

Additional to these regions, a conserved LLQDFFNG amino acid sequence was also identified between the regions encoding the ATPase and substrate-binding domain, based on which two unique guessmer internal primers, Fish70/IntF2 and Fish70/IntF2_R were designed (Table 2.2).

Table 2. 2. Primers for the PCR isolation of *L. chalumnae hsp70* coding regions

Name of primer	Amino acid sequence	Nucleotide sequence
***Lchsp70* primers based on predicted sequence**		
Fish 70F/2	GIDLGTT	5'-GGWATYGAYCTSGGWACMAC-3'
Fish 70/intF2	QDFFNG	5'-CCGTTGAAGAAGTCT-3'
Fish 70/intF2_R	LLQDFFN	5'-CAGGATTTCTTCAACCG-3'
Fish 70R/2	GPTIEEVD	5'-TTARTCVACYTCRATKGT-3'
***Lchsp70* primers based on known sequence**		
Fish70F/3.2		5'-AGATCATCACCAATGACCAG-3'
Fish70R/3.1		5'-TTAGTCGACTTCTTCAATGGT-3

The *hsp70* 5' fragments encoding the N-terminal ATPase domain was amplified with primers Fish70F/2 and Fish70/IntF2_R while the *hsp70* 3' fragment encoding the C-terminal domains was amplified with primers Fish70/IntF2 and Fish70R/2. Degenerate primers Fish70F/2 and Fish70R/2 were used to amplify the coding region of the inducible *hsp70* encoding most of the N-terminal ATPase domain and the C-terminal substrate-binding domain of Hsp70 (referred to as partial full-length *hsp70* gene or LcHsp70 protein from here on).

2.3.3 Isolation of *L. chalumnae hsp70* fragments

The PCR to isolate the *hsp70* 5' fragment encoding the ATPase domain yielded a 1048 bp *hsp70* fragment (Figure 2.3A, lane 1) while a 873 bp fragment was obtained from the reaction to isolate the *hsp70* 3' fragment encoding the Hsp70 C-terminal region (Figure 2.3A, lane 2). The recombinant pGEM-T construct containing the *hsp70* 5' PCR fragment product was named *pGEM-T_LcN*. Similarly, the recombinant pGEM-T construct containing the *hsp70* 3' PCR fragment was named *pGEM-T_LcC*.

Since the full-length coding region could not be obtained using the degenerate primers Fish70F/2 and Fish70R/2, a set of *L. chalumnae hsp70* specific primers was designed based on the sequence of the insert fragments of the *pGEM-T_LcN* and *pGEM-T_LcC* plasmid constructs. The Fish70F/3.2 primer was based on the nucleotide sequence of

the *L. chalumnae* 5' *hsp70* fragment internal to the position of degenerate primers Fish70F/2. The reverse primer, Fish70R/3.1 was based on the C-terminal end of the *hsp70* 3' fragment. The combination of this Fish70F/3.2 and Fish70R/3.1 primers yielded a 1840 bp fragment encoding a partial full-length Hsp70 (Figure 2.3B, lane 1 and 2). The recombinant product pGEM-T construct containing the partial full-length *hsp70* was named *pGEM-T_LmHsp70*.

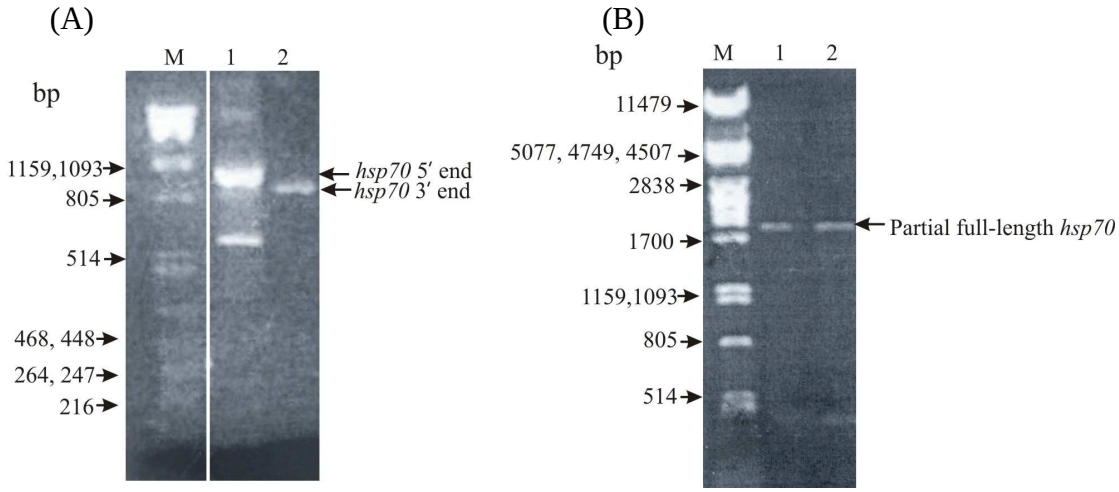


Figure 2.3: Isolation of fragment of the *L. chalumnae hsp70* gene encoding the ATPase domain, substrate binding domain, and the partial full-length Hsp70. (A) Isolation of fragments of the *hsp70* coding regions *from L. chalumnae* genomic DNA. Lane M = λ *Pst* I size marker; Lane 1 = 1048 bp PCR product of the *L. chalumnae hsp70,* encoding the N-terminal ATPase domain; lane 2 = 873 bp PCR product of the *L. chalumnae hsp70* encoding the substrate-binding domain. (B) Isolation of the partial full-length *hsp70* coding region from high molecular weight DNA using *L. chalumnae hsp70* specific primers. Lane M = λ *Pst* I size marker; lane 1 and 2 = 1840 bp PCR product encoding the partial full-length *hsp70*.

The predicted amino acid sequence of the *L. chalumnae* partial full-length Hsp70 (LcHsp70) had 82 % identity to the bovine Hsc70 followed by the *Fugu*, zebrafish and human Hsp70s at 81 %, 80 % and 76 % identity, respectively. At the amino acid level, there were considerably more differences between the partial full-length sequence and the N-terminal sequence (9 %) than between the partial full-length sequence and the C-terminal sequence (less than 1 %) (Figure 2.4 and 2.5). This is particularly interesting in the context that the more conserved region of Hsp70 is the N-terminal ATPase domain. It is therefore possible that the three sequences, especially the partial full-length and N-terminal sequences, represent different *hsp70* genes (Figure 2. 4 and 2.5).

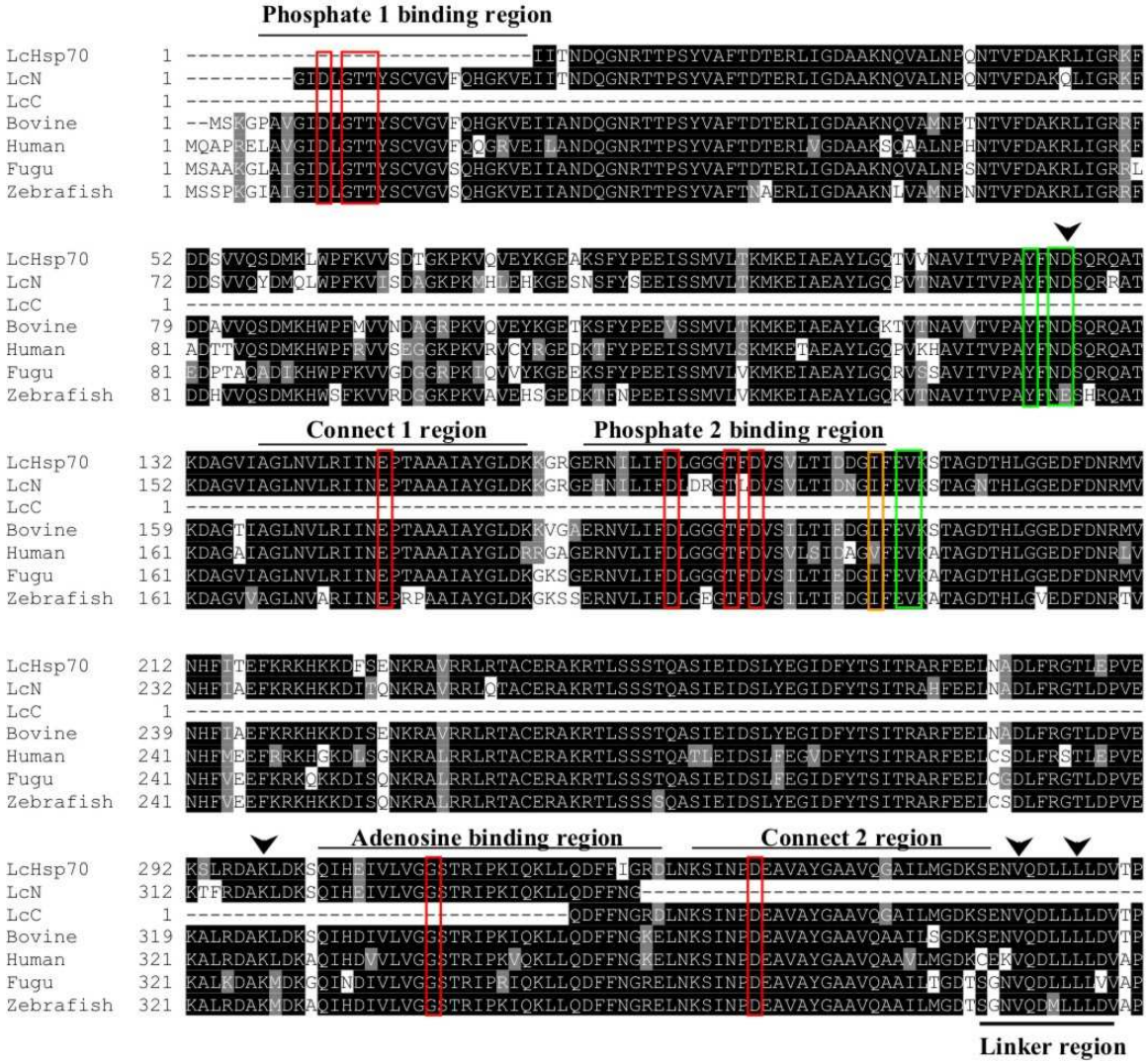


Figure 2.4: Alignment of the primary amino acid sequences of the ATPase domains of the bovine Hsc70 (Bovine), human Hsp70 (Human), *Fugu* Hsp70 (Fugu), zebrafish Hsp70 (Zebrafish), an *L. chalumnae* partial full-length Hsp70 (LcHsp70), an *L. chalumnae* Hsp70 N-terminal region (LcN) and an *L. chalumnae* Hsp70 C-terminal region (LcC). The N-terminal ATPase domain has the following functional and structural elements: ATP/ADP phosphate binding sites; connect 1 and connect 2 and region of adenosine binding (Bork *et al*., 1992). Residues boxed in red are involved in the interaction with ATP/ADP molecules (Flaherty *et al*., 1994). Residues boxed in green are critical for ATP hydrolysis. The linker region (underlined with a solid black line) joins the ATPase and the substrate-binding domain (shown in Figure 2.5). Residues of the linker region shown by arrows are important for Hsp70 chaperone function (Jiang *et al*., 2005). Residue boxed in orange and residues indicated by arrows (other than those in the linker region) form hydrophobic interaction and salt bridges at the subdomain interface, respectively (Jiang *et al*., 2005). The black shading indicates identical residues, the grey shading indicates similar residues, and the white shading highlights differences.

Both the *L. chalumnae* N-terminal and partial full-length Hsp70 protein sequences had conserved amino acid residues typically found within the Hsp70 ATPase domain, that are known to play a role in ATP binding, hydrolysis and nucleotide exchange

mediated by this domain (Figure 2. 4). The alignment (Figure 2.4) revealed a 100% conservation of residues involved in the interaction of Hsp70s with ADP/ATP molecules. The LcN and LcHsp70 protein sequences were 7 and 27 residues shorter than the bovine Hsc70, respectively (Figure 2. 4). Residues Asp^{10}, Thr^{13}, Thr^{14}, Tyr^{15}, Glu^{125}, Asp^{199}, Thr^{204}, Asp^{206}, Gly^{339} and Asp^{366} of the bovine Hsc70 (Flaherty *et al*., 1994) were 100 % conserved in corresponding residues of the LcHsp70. Also conserved in the LcHsp70 and LcN ATPase domain were residues equivalent to and Try^{149}, Asn^{151}, Asp^{152}, Glu^{218} and Val^{219} in bovine Hsc70. These residues were found to be equivalent to the residues Try^{145}, Asn^{147}, Asp^{148}, Glu^{217} and Val^{218} in DnaK, which have been shown to form the site of interaction with Hsp40 (Gässler *et al*., 1998; Suh *et al*., 1998).

In the substrate-binding domain residues of bovine Hsc70, residues Thr^{405}, Ala^{406}, Phe^{428}, Val^{438}, Ile^{440} and Ile^{474} form a hydrophobic cluster for interaction the with peptide substrate (Morshauser *et al*., 1999). Equivalent residues in LcHsp70 and LcC were also conserved (boxed in blue in figure 2.5). Also conserved were residues Ala^{379} and Tyr^{404} in LcHsp70, and Ala^{79} and Tyr^{81} in LcC which were found equivalent to Ala^{406} and Tyr^{431} in bovine Hsc70. These two residues form a bridge over the peptide binding channel of Hsc70 (Morshauser *et al*., 1999).

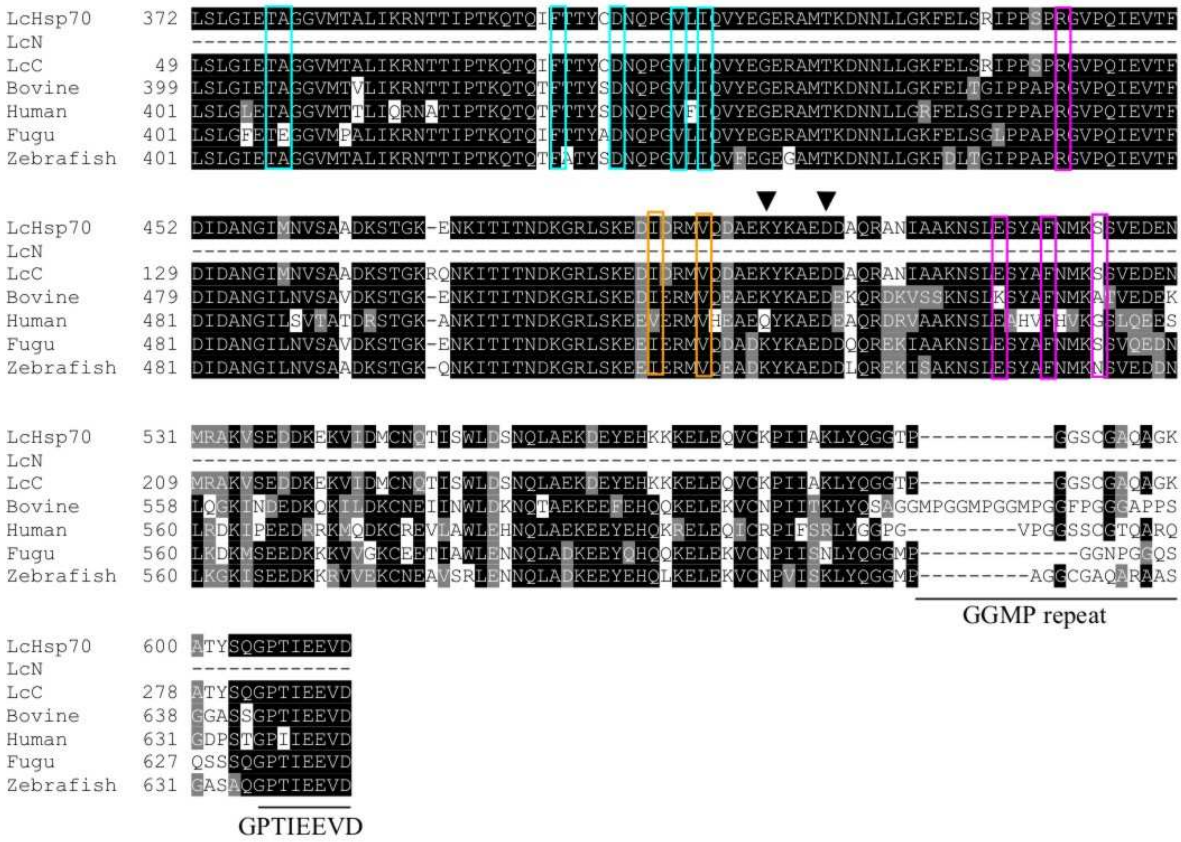


Figure 2.5: Alignment of the primary amino acid sequences of the substrate binding domain (including the C-termianal domain) of the bovine Hsc70 (Bovine), human Hsp70 (Human), *Fugu* Hsp70 (Fugu), zebrafish Hsp70 (Zebrafish), an *L. chalumnae* partial full-length Hsp70 (LcHsp70), an *L. chalumnae* Hsp70 N-terminal region (LcN) and an *L. chalumnae* Hsp70 C-terminal region (LcC). All the fish Hsp70s and the bovine Hsc70 terminates with the GPTIEEVD motif involved in the interaction of Hsp70 with Hsp70/Hsp90 organizing protein (Scheufler *et al*., 2000). Residues of the substrate binding domain boxed in blue interact with the peptide substrate (Zhu *et al*., 1996). Interaction between the substrate binding and helical lid occurs through residues boxed in magenta (Zhu *et al*., 1996). Residues boxed in orange and residues indicated by arrows are predicted to form hydrophobic interaction and salt bridges at the subdomain interface (Jiang *et al*., 2005). The numbering is continued from Figure 2.4. The black shading indicates identical residues, the grey shading indicates similar residues, and the white shading highlights differences.

A homology model of LcHsp70 using the crystal structure of the bovine Hsc70 as template revealed high level of similarity to the bovine Hsc70 (Figure 2.6). Residues Ile^{216}, Ile^{515} and Val^{519} which form a hydrophobic interaction, salt bridges between Glu^{530} and Lys^{325} and between Lys^{524} and Asp^{152} at the ATPase and substrate binding domain interface have recently been shown to play a role in bovine Hsc70 interdomain communication (Jiang *et al*., 2005). Equivalent residues Ile^{189}, Ile^{488} and Val^{492} which form a hydrophobic interaction, salt bridges between Asp^{503} and Lys^{298} and between Lys^{497} and Asp^{125} at the ATPase and substrate binding domain interface of LcHsp70 were also conserved (figure 2.4 and figure 2.5). However, Glu^{530} was

equivalent to Asp[503]. In the linker region residues were also conserved. LcHsp70 and LcC sequences terminated with a C-terminal EEVD motif, known to be the site of interaction with the Hsp70/Hsp90 organizing protein, Hop (Scheufler *et al*., 2000; Figure 2.5).

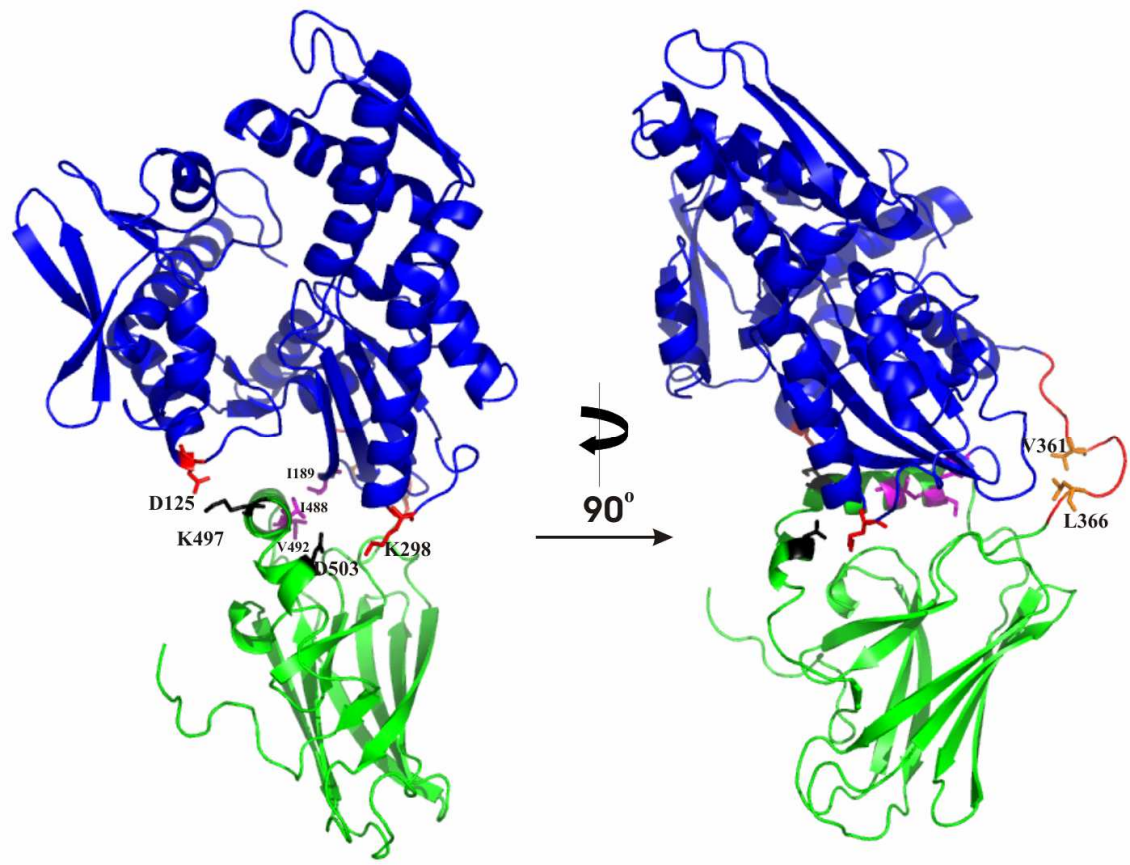


Figure 2.6: The ribbon representation of a homology model of LcHsp70. The model was generated using SWISSMODEL browser in First Approach Mode (http://swissmodel.expasy.org/ [Peitsch, 1996]) with bovine Hsc70 (PDB code: 1YUW; Jiang *et al*., 2005) as template. The structure was visualized using PyMol (DeLano *et al*., 2003). Salt bridges are predicted to occur between D125 and K497, and K298 and D503 (black and red residues). Predicted hydrophobic interactions are predicted to occur between I189, I488 and V492 (magenta residues). V361 and L366 in the linker region are shown in orange. The ATPase domain is shown in blue, the linker region in red and the substrate binding domain in green.

The neighbour-joining method (NJ [figure 2.7]) tree revealed that LcHsp70 formed a monophyletic branch A2 within the fish Hsp70 cluster A. It was particularly interesting that although the LcHsp70 clustered with fish it did not fall into the distinctive fish cluster A1. Notwithstanding this finding, LcHsp70 was found to be more closely related to fish than amphibian, tetrapod and bacterial Hsp70s. These data

did not diverge from the theme that the lungfish and not the coelacanth is the closest relative to tetrapods.

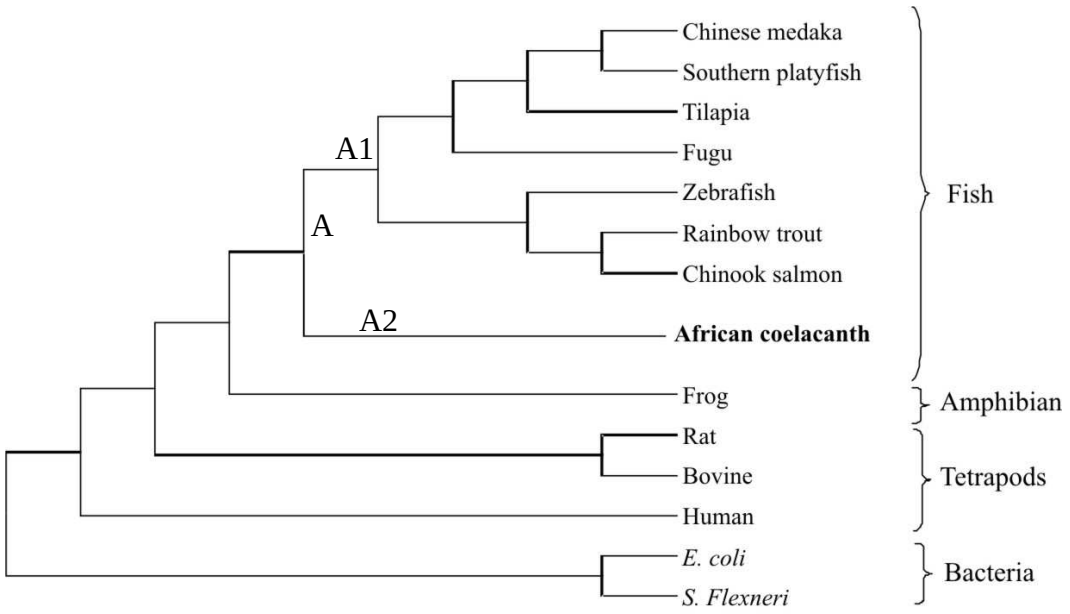


Figure 2.7: NJ tree representative of the phylogenetic relationship of *L. chalumnae* (African Coelacanth) to fish, amphibians, tetrapods and bacteria. The NJ tree was generated using Mega3 software (Nei, 2004) and Hsp70 protein sequences from the organisms indicated. A= fish cluster; A1= typical fish cluster; and A2=monophyletic LcHsp70 branch.

2.4 Discussion

The successful cloning of three distinct novel *hsp70* related sequences (5'-sequence, 3'-sequence and partial full-length sequence) strongly suggests the existence of at least *hsp70* genes on the *L. chalumnae* genome. This is finding is consistent with the discovery of a cluster of five *hsp70* genes in *F. rubripes* (Lim and Brenner, 1999). In addition, the discovery that the partial full-length coding region is intronless strongly suggests the presence of at least one inducible *hsp70* gene in the *L. chalumnae* genome. It is therefore likely that an inducible Hsp70-based cytoprotection mechanism exists in the coelacanth, as has been found for other fish species as mentioned in section 2.1. This suggestion should however, be made in light of the finding that in the highly cold-adapted, Antarctic teleost fish, *Trematomus bernacchii*, heat stress did not result in the upregulation of the heat inducible *hsp70* gene (Hofmann *et al*., 2000). It therefore becomes important to investigate the structure and

function of the promoter/regulatory region driving the expression of these genes, and to ascertain their response to thermal stress in a reporter gene expression system. The presence of the conserved regions of a typical ATPase domain suggests the expression of an Hsp70 protein that is functionally regulated by an ADP/ATP cycle. Furthermore, the C-terminal EEVD motif suggests a possible chaperone mechanism involving Hop and Hsp90 (Scheufler *et al*., 2000).

The partial full-length coding sequence will be completed and validated as part of a strategy to obtain a full coelacanth *hsp70* gene using a genomic library. It will also be interesting to investigate if the coelacanth Hsp70 can functionally replace Hsp70s in Hsp70-deleted strains of suitable yeast or bacterial systems, thereby enabling an *in vivo* characterization of the coelacanth chaperone machinery.

Chapter 3

Isolation of a *Latimeria menadoensis* intronless *hsp70* gene (*Lmhsp70*)

3.1 Introduction

Fish are adapted to live at various ambient temperatures and have also developed mechanisms to cope with daily and seasonal temperature changes. For these reasons, fish have been considered as an ideal model to study cytoprotection, especially thermal response (Ojima and Yamashita, 2004). Heat-induced expression of *hsp70* genes is regulated at the transcription level by heat shock factor 1 (HSF1) (Morimoto, 1992; Morimoto, 1998). Heat stress causes the trimerization and binding of HSF1 to the heat shock elements HSEs on the *hsp70* promoter (Morimoto, 1992; Morimoto, 1998). The HSE sequence has been determined as AGAAxxTTC, where xx represents any possible nucleotide (Pelham and Bienz, 1982; Bienz, 1985; Morimoto, 1992; Morimoto, 1998).

The 1.5 kb regulatory region of the zebrafish *hsp70-4* coding region has been isolated and shown to contain a TATAA, CCAAT and HSEs, all within 638 bp upstream of the initiation methionine (Halloran *et al.,* 2000). Both the 1.5 kb and 638 bp fragments were demonstrated to drive to expression of GFP in a reporter construct in zebrafish transgenic lines under heat shock (Halloran *et al.,* 2000).

Molina and coworkers isolated an intronless *hsp70* gene including an upstream 1 kb regulatory region and an open reading frame of 1920 bp encoding the entire tilapia Hsp70 protein (Molina *et al.,* 2000). In contrast to the zebrafish *hsp70* regulatory region, the TATA box, CCAAT box and three HSEs were all within the 1 kb from the initiation methionine. The HSEs occurred at -815 bp, -500 bp, and -179 bp upstream of the transcriptional start, respectively. Although the entire regulatory region was able to drive the expression of reporter genes in a heat shock dependent manner, the distal HSE (-800 bp) was shown to be important for the heat shock response of the tilapia *hsp70* promoter (Molina *et al.*, 2001 and 2002).

The aim of the study presented in this chapter was to isolate the complete *Lmhsp70* coding region as part of the strategy to identify the *Lmhsp70* upstream regulatory region. It was also the aim of the study to determine the putative *hsp70* promoter elements and to set testable predictions on the heat inducibility of this promoter.

3.2 Experimental Procedures

3.2.1 Screening of the *L. menadoensis* bacterial artificial chromosome (BAC) library with an *L. chalumnae hsp70* 5′-probe

Detailed protocols on the labelling of probes with α-^{32}P-dCTP, high density colony filters and hybridisation conditions are outlined in Appendix D.

Danke and co-workers successfully constructed a *pCCBACIE* vector-based genomic library with an estimated insert size of 171 kb (Danke *et al*., 2004). The pCCBACIE vector (Epicentre, USA) has a single copy *E. coli* F-factor replicon and a high-copy origin of replication called "*ori*V". Other features of the vector includes a chloramphenicol-resistance gene as an antibiotic selectable marker, primer binding sites for BAC-end sequencing, a *Not* I sites surrounding the *Bam* HI, *Hin*d III and *Eco*R I cloning sites, and a bacteriophage P1 *lox*P site for Cre-recombinase cleavage. The vector is linearized at a unique restriction enzyme recognition site (*Bam*H I, *Hin*d III or *Eco*R I), dephosphorylated, and highly purified to ensure very low background (pCCBACIE plasmid map provided Appendix Figure B2.3)

For screening of the *L. menadoensis* genomic library, a 220 bp probe was isolated by PCR from the *pGEM-T_LcN* plasmid using the forward primer (Probe1F: TTG AAG TAA AGT CTA CAG C) and reverse primer (Probe1R: AGA GTC GAT TTC AAT ACT GG). The probe will be referred to as the *Lchsp70* 5′-probe from hereon (Appendix Figure D4.1). *Lchsp70* 5′-probe was labelled with α-^{32}P dCTP as outlined in Appendix D2 and subsequently used to hybridize colony filters of 110 592 transformants of the *L. menadoensis* BAC genomic library. BAC DNA was isolated from twenty-six putative *hsp70* containing library transformants and further

confirmed by PCR using the same primers used to isolate the *Lchsp70* 5′-probe. The isolated putative *hsp70* containing BAC clones were restricted with *Not* I to release the BAC inserts and to determine the average insert size (Danke *et al*., 2004).

We were supplied with the twenty-six *E. coli* DH10BT transformants containing the putative *hsp70* BACs for further analysis.

3.2.2 Isolation of the *L. menadoensis* partial full-length *hsp70* coding region

The twenty-six *Escherichia coli* DH10BT transformants housing the BAC phagemid (called BAC1 to BAC26) containing putative *hsp70* sequences were grown overnight in 5 ml 2x YT broth supplemented with 34 μg/ml chloroamphenicol at 37°C and shook at 200 rpm. BAC DNA was isolated from overnight cultures using the QIAprep large construct isolation kit (Qiagen, Germany). Degenerate primers, Fish70F/2 and Fish70R/2 (degenerate fish *hsp70* primers designed to amplify the *hsp70* partial-full fragment as described in section 3.3.2) were used in PCR with the isolated phagemid DNA as template. Each reaction was carried out in a 50 μl as described in section 2.2.2 and analysed as described in section 2.2.5. The PCR product from the reaction with BAC25 DNA as template was gel purified and ligated into pGEM-T Easy vector as described in Appendix A4. The resultant construct was named *pGEM-T_Lmhsp70*. The PCR and cloning of the BAC25 PCR product was carried out three independent times to account for PCR errors. This *pGEM-T_Lmhsp70* sequence was used to design the BAC25/70F and BAC25/70Rsp which were used to amplify the *L. menadoensis* specific *hsp70* probe (*Lmhsp70* 5′ probe) (described in 3.2.3) and also the BAC25/70R primer used to isolate the full-length *Lmhsp70* fragment (described in section 3.2.5).

3.2.3 Construction and screening of the BAC25 mini-library

BAC25 DNA (1 μg) was partially digested with *Sau* 3A (Roche, Germany) and samples were resolved on a 0.8% agarose gel (Appendix A9). The digested BAC25 DNA fragments between 1-1.7 kb were gel purified from the sample with most of the restricted DNA at 1-1.7 kb using the GFX DNA purification kit (Amersham,

Sweeden). Purified DNA was ligated into the *Bam* HI site of the pBK-CMV phagemid (plasmid map provided in Appendix Figure B2.4) using the mass excision protocol of the λZAP Express vector kit (Stratagene, USA). The recombinant pBK-CMV phagemid was transformed into *E. coli* XLI Blue cells and plated onto 2x YT agar plates containing 50 μg/ml of kanamycin (Appendix A5). Colonies were inoculated into 5 ml of 2x YT broth supplemented with 50 μg/ml of kanamycin and incubated at 37ºC overnight. Phagemid DNA was isolated using the smart buffer plasmid isolation protocol (Appendix A12). The insert size of the library was estimated by digesting the phagemid DNA with *Eco* RI and *Pst* I to release the inserts (Appendix A6). Since the size of BAC25 was approximately 178 kb, a minimum of 600 ligated fragments of 1 kb in size were calculated to be necessary to have a 95% confidence that any 1 kb fragment will be represented at least once in the library.

To screen the BAC25 mini-library, 3000 transformants were robotically inoculated into eight 384 well plates containing 50 μl of 2x YT broth supplemented with 50 μg/ml of kanamycin and the plate were incubated at 37ºC overnight. Overnight cultures were robotically spotted on Hybond^{+} hybridization membrane (Amersham, Sweeden) to make a high density membrane which was incubated at 37ºC overnight. The colonies on the membrane were lysed using the lysis solution (2x SSC; 5% SDS) followed by a proteinase K treatment (Proteinase K buffer: 50 mM Tris [pH 8]; 50 mM EDTA; 100 mM NaCl; 1% N-lauryl sarcosine; 10 μg/ml proteinase K). UV (ultraviolet) crosslinking was carried out using the Stratalinker 1800 UV Crosslinker (Stratagene, USA) using the "autocrosslink" setting while the filter was still damp. Hybridisation was carried out with the α-^{32}P dCTP labelled *Lmhsp70* 5′ probe (detailed protocol is supplied in the Appendix D).

An *L. menadoensis* specific *hsp70* probe (*Lmhsp70* 5′ probe, Appendix Figure D4.1) was isolated from the *pGEM-T_Lmhsp70* plasmid DNA using the High Fidelity Expand PCR kit (Promega, USA) with BAC25 *hsp70* specific forward primer BAC25/70F (5′-ACC TCG GTA CCA CTT ACT C-3′) and reverse primer BAC25/70Rsp (5′-AGA GTT TCA TGT CTG ATT GG). The position of the primers was designed so that the amplified probe occurred at the most 5' portion of the coding region. The practical significance of this is that hybridization with a probe at the most

5' portion of the coding region as opposed to the *Lchsp70* 5' probe (Appendix Figure D4.2) would increase the probability to hybridize fragments containing both the upstream regulatory region and a portion of the coding region. The *Lmhsp70* 5′-probe was used to probe the Hybond^{+} hybridization membrane containing the 3000 transformants isolated from the BAC25 mini-library.

Positive colonies were inoculated into 5 ml of 2x YT supplemented with 50 µg/ml of kanamycin. Plasmid DNA was isolated from overnight cultures using the QIAprep miniprep kit (Qiagen, Germany), sequenced and analysed as outlined in section 2.2.4. The pBK-CMV phagemid mini-library recombinant identified to contain a regulatory region upstream of the *Lmhsp70* coding region of was named *pBK-CMV_BAC25prom.* The plasmid construct was sequenced and use to design the FL_Lm70 forward primer as described section 3.2.5.

3.2.4 Pulsed field gel electrophoresis (PFGE)

Putative *hsp70* containing BAC DNA was isolated from positive colonies identified by screening the *L. menadoensis* BAC genomic library with the *L. chalumnae hsp70* specific 220 bp probe described in section 3.2.1 and Appendix D (*Lchsp70* 5' probe). The average insert size was estimated by *Not* I restriction digests of the isolated BAC DNA and its subjection to PFGE conducted in 1 % agarose gel (0.5X TBE: 45 mM Tris; 45 mM Borate; 1 mM EDTA) using a CHEF DRIII unit (Biorad, USA) at 0.1-40 s switching time, 6 V/cm field strength, 120º reorientation angle, 14ºC chamber temperature, and 16 h run time. Low range PFGE molecular weight markers (New England Biolabs, United Kingdom) were used to determine the size of the insert (Danke *et al*., 2004).

3.2.5 Isolation of the full-length *Lmhsp70* coding region

An *Lmhsp70* specific forward primer FL_Lm70 (5′-ATG TCT GCA GCC AAG GGA G-3′) was designed based on the most 5' portion of *Lmhsp70* encoding the LmHsp70 beginning with the initiation methionine as determined from the *pBK-CMV_BAC25prom* phagemid construct. The *L. menadoensis hsp70* specific reverse

primer BAC25/70R (5′- TTA ATC GAC CTC TTC AAT G-3′) was designed based on the 3' terminal portion of partial full-length *Lmhsp70* as determined in the nucleotide sequence of *pGEM-T_Lmhsp70.* PCR with the FL_Lm70 and BAC25/70R primer combination was used to isolate the 1926 bp full-length *Lmhsp70* PCR fragment from the BAC25 DNA. PCR was carried out as described in Appendix A3. The PCR product was gel purified and ligated into pGEM-T Easy as outlined in Appendix A4-A6. Three independent PCRs were carried out to eliminate PCR errors. The resultant plasmid construct was named *pGEM-T_FLLmhsp70*.

3.2.6 Bioinformatic Analyses

The prediction of putative heat shock element occurring in the regulatory region at the 5'-end of the *Lmhsp70* coding region was carried out using the Transcription Factor Search software (TFSEARCH; Heinemeyer *et al*., 1998) at default settings. TFSEARCH software searches the highly correlated sequence fragments against the transcription factor binding sites profiles database in the “TRANSFAC” database. The TRANSFAC database provides information on the genomic binding sites of eukaryotic factors and the binding protein (Heinemeyer *et al*., 1998). The software was accessed on the 07 October 2005.

Putative promoter sequences in the same region were predicted using the Berkeley Drosophila Genome Project (BDGP): Neural Network Promoter Prediction Software at default settings (www.fruitfly.org; Reese, 2001). This promoter prediction system uses a technique that combines neural networks with weight pruning. While the neural networks recognises a promoter until the lowest minimum, the pruning function removes prediction that are of lowest prediction value. This process is repeated until the lowest error is achieved. By analysing the remainder of the weight the pruned neural networks provide information on the specific positions of the promoter elements. The software was accessed on the 07 October 2005.

3.3 Results

3.3.1 Isolation of the *L. menadoensis* partial full-length *hsp70* coding region

Not I restriction analysis of the BAC DNA isolated from the twenty-six putative *hsp70*-containing transformants of the *L. menadoensis* genomic library yielded an average insert size of 170 kb (Figure 3.1; Danke *et al*., 2004).

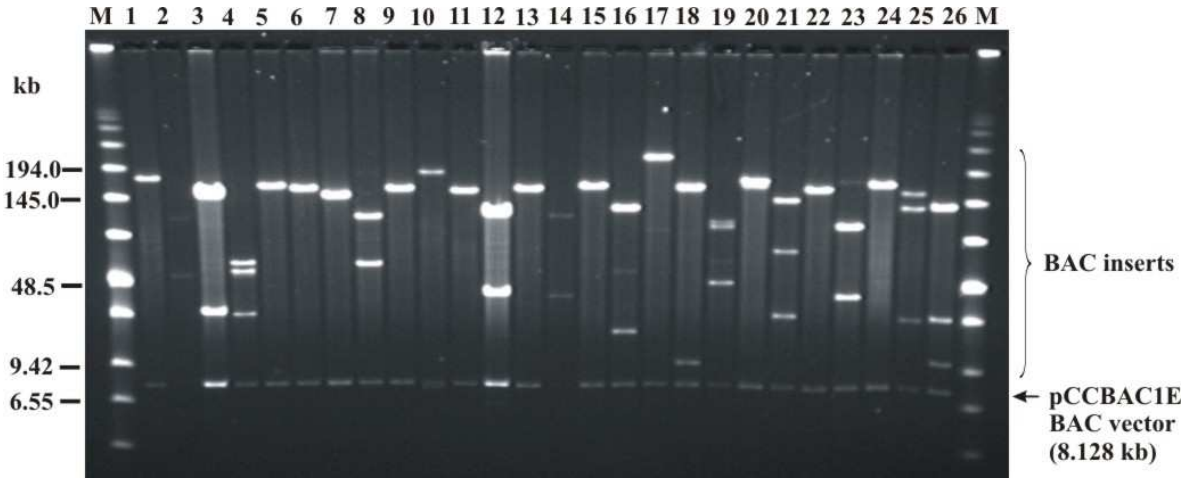


Figure 3.1: Pulse field gel electrophoresis (PFGE) of twenty-six putative *hsp70* containing *L menadoensis* BAC plasmids. The average insert size was estimated at 170 kb upon digestion of the isolated BAC DNA with *Not* 1 and its subjection to PFGE electrophoresis (lanes 1-26) on 1% agaose. M= low range PFGE molecular weight markers (New England Biolabs, United Kingdom). The Not 1 digest released the BAC genomic inserts from the pCCBAC1E BAC vector.

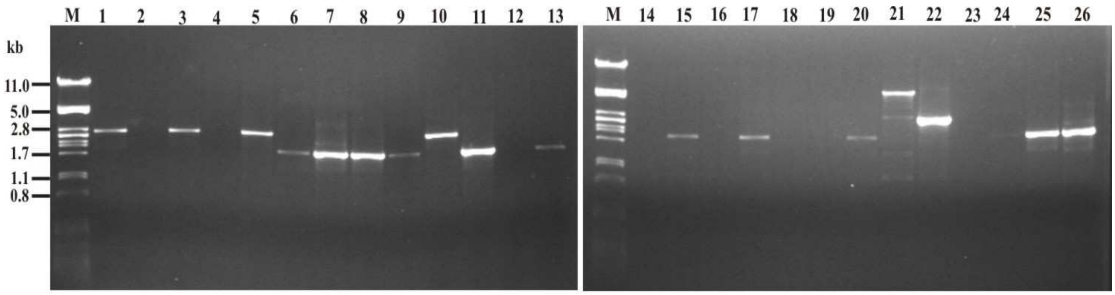


Figure 3.2: PCR scanning of the putative *hsp70* containing *L. menadoensis* BAC plasmid. PCR was conducted in reactions containing BAC1-BAC26 as templates and degenerate fish *hsp70* primers described in section 3.3.2 to amplify the regions corresponding to the partial full-length *hsp70* coding region described in chapter 2. The PCR products were resolved using 0.8% agarose gel electrophoresis. M= λ *Pst* I molecular size marker; 1-26= PCR product obtained form BAC 1-26.

The presence of partial full-length *hsp70* coding region in all the twenty-six putative *hsp70* containing BACs was investigated by PCR screening with degenerate primers Fish70F/2 and Fish70R/2 (as described in section 2.3.2.). These primers were shown to amplify a partial full-length fragment of the intronless *L. chalumnae hsp70* coding region.

Eighteen of the twenty-six BACs yielded PCR products ranging between 1.7 and 5.0 kb (Figure 3.2). While no PCR products were obtained from BACs 2; 4; 14; 16; 18; 19 and 23, a 1.7 kb PCR product was obtained from BACs 6-9; 15; 17 and 20. PCR from BACs 11; 13; 24; 25 and 26 yielded a 1.9 kb fragment while BACs 1; 3; 5; 10 and 22 yielded a 2.8 kb fragment. A much larger 5.0 kb fragment was obtained from PCR on BAC21.

Sequence analysis of the *pGEM-T_Lmhsp70* plasmid construct containing the PCR product obtained from BAC25 yielded an 1899 bp intronless *hsp70* segment encoding a 633 amino acids partial full-length Hsp70 protein (Figure 3.3). This *L. menadoensis* partial full-length *hsp70* fragment had a 99.6% and 99.1% identity to the nucleotide and amino acid sequences of *Lchsp70* and LcHsp70, respectively.

3.3.2 Construction and screening of the BAC25 mini-library

To isolate the complete *L. menadoensis hsp70* gene including the 5′- regulatory region, a mini-library of BAC25 was constructed. A BAC25 partial digest was carried out (Figure 3.4A) and the fragments between 1-1.7 kb in lane 3 in Figure 3.4A were gel purified and ligated into the pBK-CMV phagemid vector. The *Bam* HI and *Pst* I digestion of the phagemid DNA isolated from the resultant transformants yielded an estimated size of the inserts in the library at 1-1.7 kb (Figure, 3.4B). A total of 3000 transformants were selected to make a library and hybridisation filters.

Screening of the BAC25 mini-library with the *Lmhsp70* 5′ probe described in section 3.2.3 yielded 20 positives (Figure 3.5). Phagemid DNA was isolated from eight randomly selected transformants and sequenced with T7 promoter and T3 promoter based sequencing primers.

```
GGAATTGACCTCGGTACCACTTACTCCTGCGTGGGTGTCTTCCAGCATGGCAAAGTGGAGATCATCGCCAATGACCAGGGCAACAGGACG 90
 G  I  D  L  G  T  T  Y  S  C  V  G  V  F  Q  H  G  K  V  E  I  I  A  N  D  Q  G  N  R  T

ACTCCCAGCTACGTGGCTTTCACTGACACGGAGAGGCTCATTGGAGACGCGGCCAAGAACCAGGTGGCCCTGAATCCCCAGAACACTGTC 180
 T  P  S  Y  V  A  F  T  D  T  E  R  L  I  G  D  A  A  K  N  Q  V  A  L  N  P  Q  N  T  V

TTCGATGCCAAGAGACTGATTGGGAGGAAATTTGACGATTCAGTGGTCCAATCAGACATGAAACTCTGGCCTTTTAAAGTCGTGAGTGAC 270
 F  D  A  K  R  L  I  G  R  K  F  D  D  S  V  V  Q  S  D  M  K  L  W  P  F  K  V  V  S  D

ACTGGGAAGCCCAAGGTGCAGGTGGAATACAAGGGAGAAGCGAAGAGCTTCTACCCAGAAGAGATCTCTTCCATGGTATTGACCAAGATG 360
 T  G  K  P  K  V  Q  V  E  Y  K  G  E  A  K  S  F  Y  P  E  E  I  S  S  M  V  L  T  K  M

AAAGAGATTGCAGAAGCTTACCTGGGCCAAACAGTGGTCAATGCTGTCATCACTGTCCCCGCCTACTTCAATGATTCCCAGCGTCAGGCT 450
 K  E  I  A  E  A  Y  L  G  Q  T  V  V  N  A  V  I  T  V  P  A  Y  F  N  D  S  Q  R  Q  A

ACCAAAGATGCTGGAGTGATCGCTGGTCTGAATGTCCTGAGAATCATCAATGAGCCCACAGCAGCTGCCATTGCATATGGACTGGATAAA 540
 T  K  D  A  G  V  I  A  G  L  N  V  L  R  I  I  N  E  P  T  A  A  A  I  A  Y  G  L  D  K

AAGGGCAGAGGAGAGCGCAATATTCTTATCTTTGACCTGGGTGGAGGCACCTTTGACGTCTCTGTCCTCACCATTGATGATGGTATCTTT 630
 K  G  R  G  E  R  N  I  L  I  F  D  L  G  G  G  T  F  D  V  S  V  L  T  I  D  D  G  I  F

GAAGTAAAGTCTACAGCTGGTGATACCCACCTGGGAGGAGAAGACTTTGACAATCGCATGGTCAACCACTTTATCACAGAGTTCAAGAGA 720
 E  V  K  S  T  A  G  D  T  H  L  G  G  E  D  F  D  N  R  M  V  N  H  F  I  T  E  F  K  R

AAGCACAAGAAGGACTTTAGTGAGAACAAAAGGGCAGTCAGGAGGCTTCGGACAGCCTGCGAGAGAGCTAAGAGGACTCTGTCTTCCAGC 810
 K  H  K  K  D  F  S  E  N  K  R  A  V  R  R  L  R  T  A  C  E  R  A  K  R  T  L  S  S  S

ACTCAGGCCAGTATTGAAATTGACTCTTTGTATGAGGGTATTGACTTCTATACATCTATTACCAGGGCTCGCTTTGAAGAGCTGAATGCC 900
 T  Q  A  S  I  E  I  D  S  L  Y  E  G  I  D  F  Y  T  S  I  T  R  A  R  F  E  E  L  N  A

GACCTCTTTAGGGGAACCCTGGAACCTGTAGAGAAATCCCTCCGAGATGCTAAATTAGACAAGTCTCAGATCCATGAGATCGTCTTGGTT 990
 D  L  F  R  G  T  L  E  P  V  E  K  S  L  R  D  A  K  L  D  K  S  Q  I  H  E  I  V  L  V

GGAGGATCTACCCGCATTCCCAAGATCCAGAAACTGCTGCAAGACTTCTTCAACGGCAGAGACCTGAACAAGAGCATCAACCCTGATGAA 1080
 G  G  S  T  R  I  P  K  I  Q  K  L  L  Q  D  F  F  N  G  R  D  L  N  K  S  I  N  P  D  E

GCCGTGGCTTATGGTGCTGCTGTCCAGGGAGCCATTCTGATGGGGGACAAATCTGAGAACGTGCAGGACCTGCTGCTGCTGGATGTCACT 1170
 A  V  A  Y  G  A  A  V  Q  G  A  I  L  M  G  D  K  S  E  N  V  Q  D  L  L  L  L  D  V  T

CCGCTGTCCCTGGGTATAGAGACAGCAGGAGGGGTAATGACTGCCCTCATCAAACGCAACACCACCATCCCCACCAAGCAAACTCAGATT 1260
 P  L  S  L  G  I  E  T  A  G  G  V  M  T  A  L  I  K  R  N  T  T  I  P  T  K  Q  T  Q  I

TTCACCACCTACTGTGACAATCAGCCCCGGTGTCCTGATCCAAGTCTATGAGGGGGAAAGAGCCATGACCAAAGATAACAACCTGCTGGGA 1350
 F  T  T  Y  C  D  N  Q  P  G  V  L  I  Q  V  Y  E  G  E  R  A  M  T  K  D  N  N  L  L  G

AAGTTTGAGCTGAGCAGAATTCCTCCTTCCCCTCGTGGCGTTCCACAGATCGAGGTCACCTTTGACATTGACGCCAATGGCATCATGAAC 1440
 K  F  E  L  S  R  I  P  P  S  P  R  G  V  P  Q  I  E  V  T  F  D  I  D  A  N  G  I  M  N

GTGTCGGCTGCGGACAAGAGCACCGGCAAAGAGAACAAGATCACCATCACCAACGACAAGGGCAGGCTGAGCAAAGAGGATATCGACAGG 1530
 V  S  A  A  D  K  S  T  G  K  E  N  K  I  T  I  T  N  D  K  G  R  L  S  K  E  D  I  D  R

ATGGTGCAGAATGCGAGAAAATACAAAGCAGAAGATGATGCACAGAGGGCGAATATAGCTGCCAAGAACTCCCTGGAGTCCTACGCCTTC 1620
 M  V  Q  N  A  R  K  Y  K  A  E  D  D  A  Q  R  A  N  I  A  A  K  N  S  L  E  S  Y  A  F

AACATGAAGAGTTCAGTGGAGGACGAGAACATGAGAGCGGAGGTTAGCGTAGACGATAAGGAGAAAGTGATTGACATGTGTAACCAGACC 1710
 N  M  K  S  S  V  E  D  E  N  M  R  A  E  V  S  V  D  D  K  E  K  V  I  D  M  C  N  Q  T

ATCTCCTGGCTGGACAGCAACCAGCTGGCTGAGAAGGACGAGTATGAGCACAAGAAGAAGGAACTGGAGCAAGTGTGCAAACCCATCATC 1800
 I  S  W  L  D  S  N  Q  L  A  E  K  D  E  Y  E  H  K  K  K  E  L  E  Q  V  C  K  P  I  I

GCCAAGCTGTACCAGGGGGGCACGCCAGGAGGCTCTTGTGGAGCCCAAGCAGGCAAAGCTACCTACAGCCAAGGACCCACCATTGAAGAG 1890
 A  K  L  Y  Q  G  G  T  P  G  G  S  C  G  A  Q  A  G  K  A  T  Y  S  Q  G  P  T  I  E  E

GTCGATTAA 1899
 V  D  *
```

Figure 3.3: Nucleotide and amino acid sequences of the partial full-length *Lmhsp70* isolated from BAC25 plasmid DNA using *hsp70* degenerate primers Fish70/2 and Fish70R/2. The nucleotide sequence is the top sequence with the amino acid shown in the below. The underlined nucleotide sequence indicates the region PCR amplified to generate the *Lmhsp70* 5′ probe used to screen the BAC 25 mini-library in section 3.2.3 and Appendix Figure D4.1. The numbering is based on the nucleotide sequence. Nucleotides number 3 (A), 12 (C) and 15 (T) differed to corresponding *Lmhsp70* full-length nucleotides (Figure 3.7) 30 (C), 40 (G) and 42 (C), respectively. In both sequences, the said differences were at the third base because of the degeneracy on the Fish70F/2.

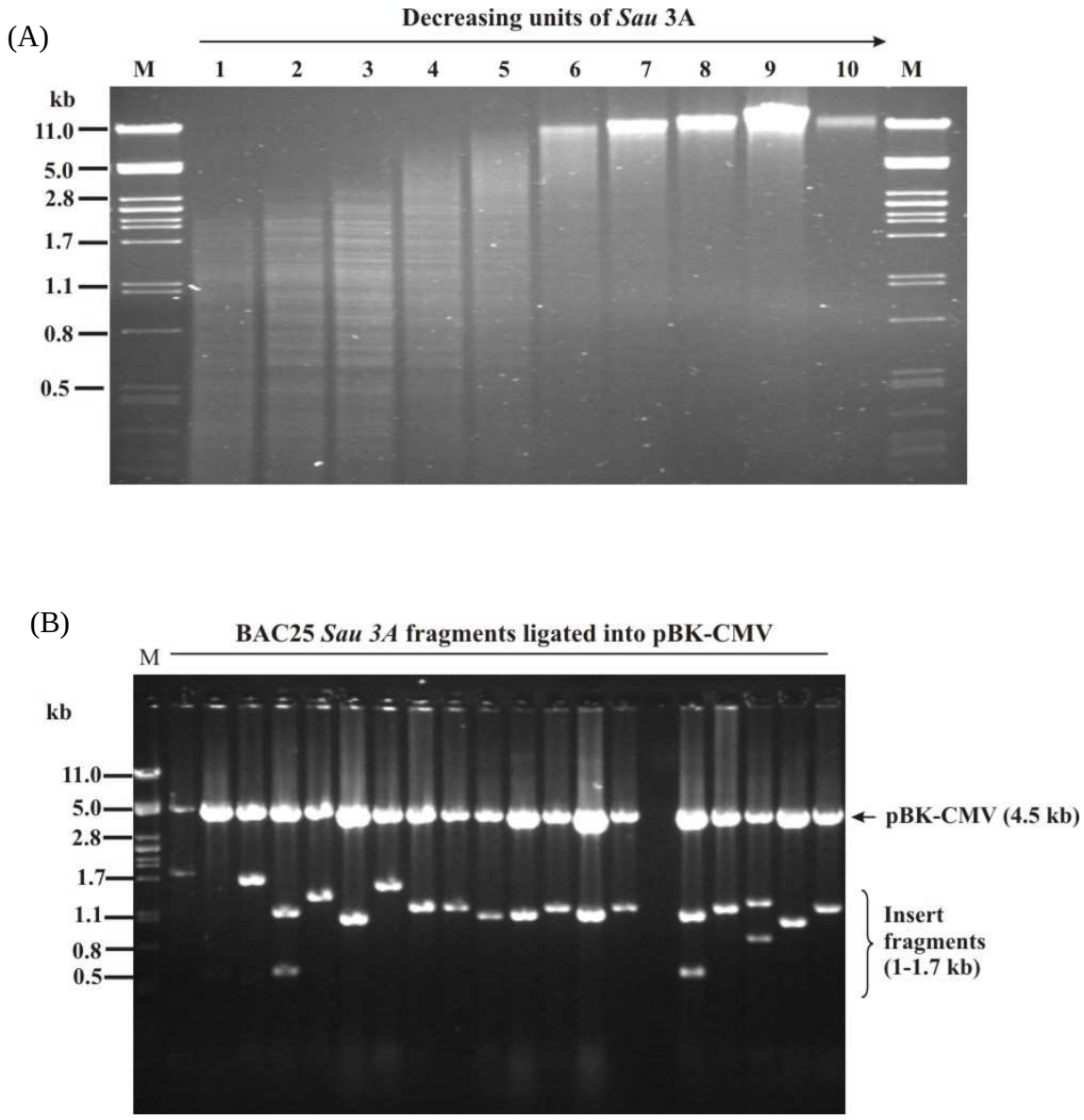


Figure 3.4: Mini-library of the *hsp70*- containing *L. menadoensis* BAC 25 into pBK-CMV phagemid vector. (A) A 1% agarose gel of the *Sau* 3a partial digest of BAC25 DNA. Fragments between 1 and 1.7 kb in lane 3 were gel purified and ligated into pBK-CMV phagemid using the ZAP express (detailed protocol of the *Sau* 3A digest is in Appendix A8). M= λ *Pst* I size marker; 1= 10 units *Sau* 3A; 2 = 5 units *Sau* 3A; 3 = 2.5 units *Sau* 3A; 4 = 1.25 units *Sau* 3A; 5 = 0.625 units *Sau* 3A; 6 = 0.313 units *Sau* 3A; 7 = 0.165 units *Sau* 3A; 8 = 0.078 units *Sau* 3A; 9 = 0.039 units *Sau* 3A; 10 = 0 units *Sau* 3A. (B) A representative 0.8 % agarose gel of recombinant pBK-CMV plasmids digested with *Pst* I and *Eco* RI. The average insert size was estimated at 1-1.7 kb. The pBK-CMV vector is indicated at 4.5 kb and M= λ *Pst* I size marker.

(A1)

2 ●	6	2 ●	3
1	4	7	5
7	8	3	6
1	5	4	8

(A2)

2	6	2	3
1 ●	4	7	5
7	8	3	6
1 ●	5	4	8

(B)

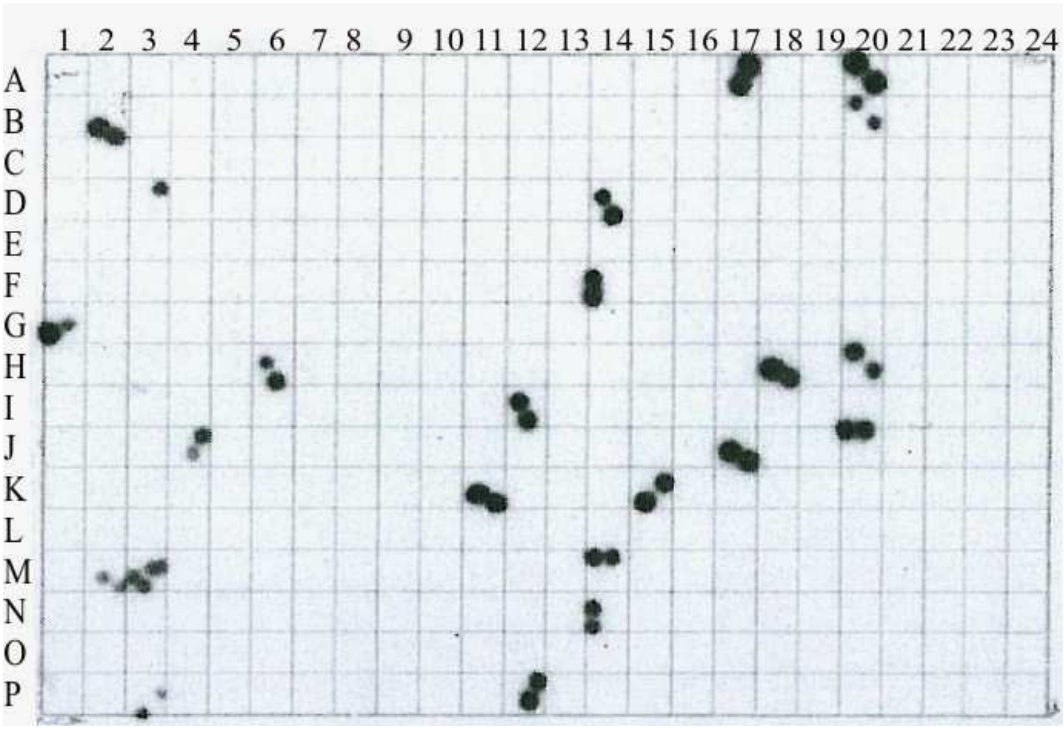


Figure 3.5: Successful screening of BAC25 mini-library for constructs encoding the 5'-end of *Lmhsp70*. (A1) and (A2) are examples and explanations of how results present in (B) are read. Each colony on each plate of the library is robotically spotted twice on the membrane in a different position. In our case 8 plates were place on the same membrane, each plate spotted twice. The large square in (A1) and (A2) represents a single square on a 384 microtiter plate in (B). The same numbers in a square represents position of the first spot relative to the second spot, which means that there will be a total of sixteen spots per small square in (B) or a large square in (A1) and (A2). The spots are therefore used to determine the plate from which the clone was selected. Example: if (A1) and (A2) represents squares 14M and 14N, then the positive colony in (A1) and (A2) would have been selected from plates 2 and 1, respectively. (B) An X-ray film after a 1 hour exposure to a nylon membrane hydridized with the ^{32}P-labelled *Lmhsp70* 5′ probe overnight. Each set of black spots per square represent a positive colony at that particular position (e.g. 20B) in a particular plate (e.g. the plate number for spots in 20B is 6; 14M is 2; 14N is 1; etc. A total of twenty positives was clearly identified.

Sequence analysis of the recombinant plasmids revealed that all the eight clones had the same insert and one was chosen for further analysis. The 1.313 kb insert consisted of the *Lmhsp70* 1.227 kb upstream regulatory region and an 86 bp region encoding the N-terminal end of the LmHsp70 (Figure 3.6). One such recombinant pBK-CMV phage was name pBK-*CMV_BAC25prom* was selected for further analysis.

A search for transcription factor binding sites in the *Lmhsp70* 1.227 kb upstream regulatory region using TFSEARCH software revealed three putative HSEs. The most proximal HSE1 was found at 411 bp upstream of the *Lmhsp70* coding region. The intermediate HSE 2 was found at 486 bp with only 75 nucleotides separating it from HSE1. HSE3 occurred at 972 bp upstream of the *Lmhsp70* coding region (Figure 3.6). Also of interest was the finding that the predicted HSEs contained two nGAAn repeats.

A search for promoters using the Neural Network Promoter Prediction software revealed six (6) predicted promoters. Predicted promoters 1 and 2 occurred between the most proximal HSE 1 and the *Lmhsp70* coding region while predicted promoters 4 and 5 occurred at a position between HSE 2 and the distal HSE3. Promoter 6 occurred at a position beyond HSE3 (Figure 3.6). Although no classical CCAAT boxes (Chae *et al*., 2005; Molina *et al*., 2000) were found in the regulatory region, two CAAT boxes were found (Figure 3.6). A TATA-box was found at 46 bp upstream of the translation start of promoter 1.

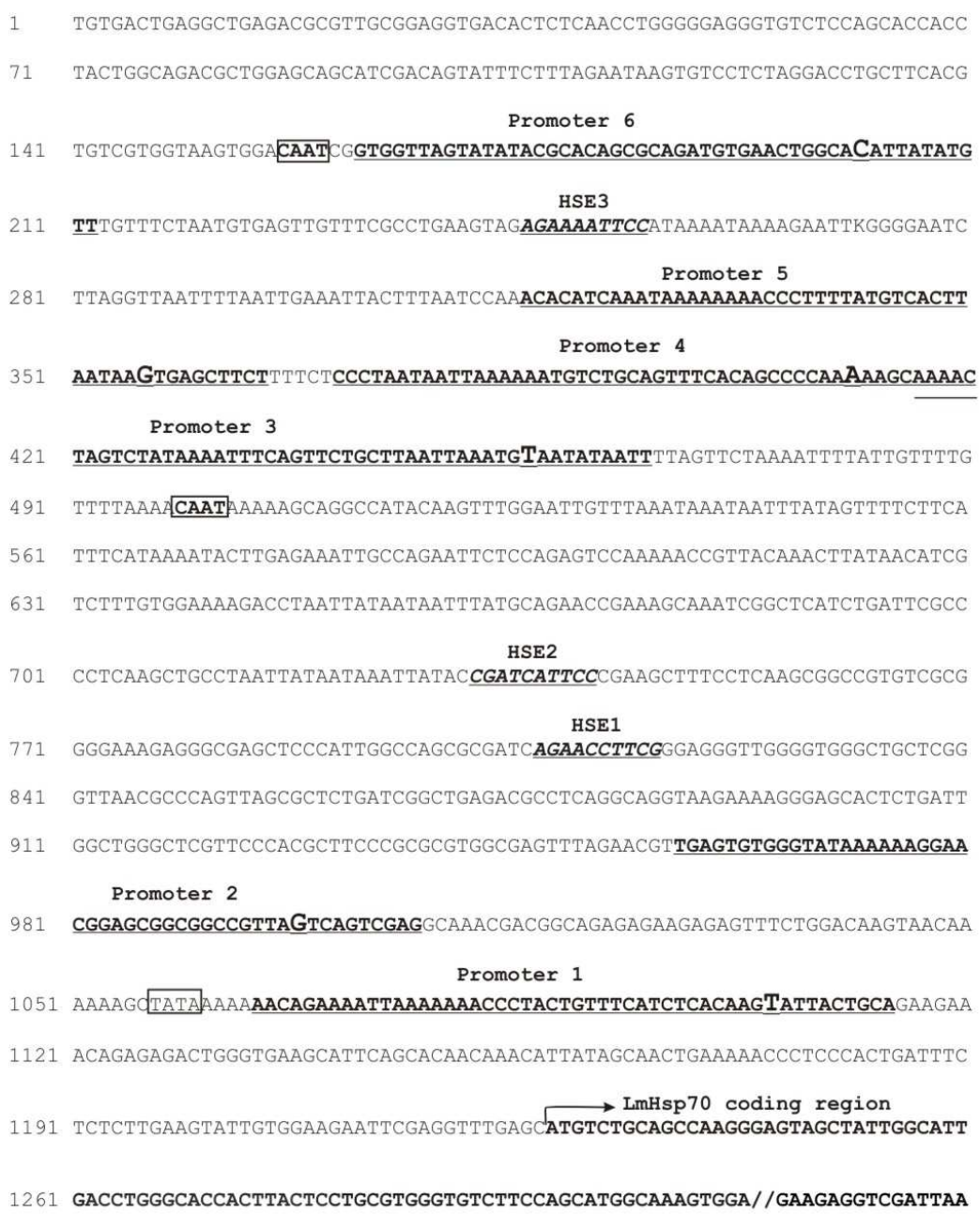


Figure 3.6: Annotation of the 1.2 kb regulatory region upstream of the *Lmhsp70* coding region. Nucleotide sequence of the 5' *Lmhsp70* regulatory region including the beginning and end of the region encoding the LmHsp70 protein (complete coding region is shown in Figure 3.7). Three predicted HSEs (HSE1, HSE2 and HSE3: bold, italics and underlined) occur at 411(position 804-816), 486 (position 732-741) and 972 (position 246-255) nucleotides upstream of the start codon (ATG) of the *Lmhsp70* open reading frame, respectively. Predicted promoters 1-6 (bold and underlined) were found at positions 1065-1114, 958-1008, 416-466, 371-421, 316-366 and 163-213, respectively. Putative TATA-box and CAAT-boxes are boxed. Transcriptional starts in each promoter are shown in a larger font. An arrow indicates the start of the *Lmhsp70* encoding region.

```
ATGTCTGCAGCCAAGGGAGTAGCTATTGGCATTGACCTGGGCACCACTTACTCCTGCGTGGGTGTCTTCCAGCATGGCAAAGTGGAGATC 90
 M  S  A  A  K  G  V  A  I  G  I  D  L  G  T  T  Y  S  C  V  G  V  F  Q  H  G  K  V  E  I

ATCGCCAATGACCAGGGCAACAGGACGACTCCCAGCTACGTGGCTTTCACTGACACGGAGAGGCTCATTGGAGACGCGGCCAAGAACCAG 180
 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

GTGGCCCTGAATCCCCAGAACACTGTCTTCGATGCCAAGAGACTGATTGGGAGGAAATTTGACGATTCAGTGGTCCAATCAGACATGAAA 270
 V  A  L  N  P  Q  N  T  V  F  D  A  K  R  L  I  G  R  K  F  D  D  S  V  V  Q  S  D  M  K

CTCTGGCCTTTTAAAGTCGTGAGTGACACTGGGAAGCCCAAGGTGCAGGTGGAATACAAGGGAGAAGCGAAGAGCTTCTACCCAGAAGAG 360
 L  W  P  F  K  V  V  S  D  T  G  K  P  K  V  Q  V  E  Y  K  G  E  A  K  S  F  Y  P  E  E

ATCTCTTCCATGGTATTGACCAAGATGAAAGAGATTGCAGAAGCTTACCTGGGCCAAACAGTGGTCAATGCTGTCATCACTGTCCCCGCC 450
 I  S  S  M  V  L  T  K  M  K  E  I  A  E  A  Y  L  G  Q  T  V  V  N  A  V  I  T  V  P  A

TACTTCAATGATTCCCAGCGTCAGGCTACCAAAGATGCTGGAGTGATCGCTGGTCTGAATGTCCTGAGAATCATCAATGAGCCCACAGCA 540
 Y  F  N  D  S  Q  R  Q  A  T  K  D  A  G  V  I  A  G  L  N  V  L  R  I  I  N  E  P  T  A

GCTGCCATTGCATATGGACTGGATAAAAAGGGCAGAGGAGAGCGCAATATTCTTATCTTTGACCTGGGTGGAGGCACCTTTGACGTCTCT 630
 A  A  I  A  Y  G  L  D  K  K  G  R  G  E  R  N  I  L  I  F  D  L  G  G  G  T  F  D  V  S

GTCCTCACCATTGATGATGGTATCTTTGAAGTAAAGTCTACAGCTGGTGATACCCACCTGGGAGGAGAAGACTTTGACAATCGCATGGTC 720
 V  L  T  I  D  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

AACCACTTTATCACAGAGTTCAAGAGAAAGCACAAGAAGGACTTTAGTGAGAACAAAAGGGCAGTCAGGAGGCTTCGGACAGCCTGCGAG 810
 N  H  F  I  T  E  F  K  R  K  H  K  K  D  F  S  E  N  K  R  A  V  R  R  L  R  T  A  C  E

AGAGCTAAGAGGACTCTGTCTTCCAGCACTCAGGCCAGTATTGAAATTGACTCTTTGTATGAGGGTATTGACTTCTATACATCTATTACC 900
 R  A  K  R  T  L  S  S  S  T  Q  A  S  I  E  I  D  S  L  Y  E  G  I  D  F  Y  T  S  I  T

AGGGCTCGCTTTGAAGAGCTGAATGCCGACCTCTTTAGGGGAACCCTGGAACCTGTAGAGAAATCCCTCCGAGATGCTAAATTAGACAAG 990
 R  A  R  F  E  E  L  N  A  D  L  F  R  G  T  L  E  P  V  E  K  S  L  R  D  A  K  L  D  K

TCTCAGATCCATGAGATCGTCTTGGTTGGAGGATCTACCCGCATTCCCAAGATCCAGAAACTGCTGCAAGACTTCTTCAACGGCAGAGAC 1080
 S  Q  I  H  E  I  V  L  V  G  G  S  T  R  I  P  K  I  Q  K  L  L  Q  D  F  F  N  G  R  D

CTGAACAAGAGCATCAACCCTGATGAAGCCGTGGCTTATGGTGCTGCTGTCCAGGGAGCCATTCTGATGGGGGACAAATCTGAGAACGTG 1170
 L  N  K  S  I  N  P  D  E  A  V  A  Y  G  A  A  V  Q  G  A  I  L  M  G  D  K  S  E  N  V

CAGGACCTGCTGCTGCTGGATGTCACTCCGCTGTCCCTGGGTATAGAGACAGCAGGAGGGGTAATGACTGCCCTCATCAAACGCAACACC 1260
 Q  D  L  L  L  L  D  V  T  P  L  S  L  G  I  E  T  A  G  G  V  M  T  A  L  I  K  R  N  T

ACCATCCCCACCAAGCAAACTCAGATTTTCACCACCTACTGTGACAATCAGCCCGGTGTCCTGATCCAAGTCTATGAGGGGGAAAGAGCC 1350
 T  I  P  T  K  Q  T  Q  I  F  T  T  Y  C  D  N  Q  P  G  V  L  I  Q  V  Y  E  G  E  R  A

ATGACCAAAGATAACAACCTGCTGGGAAAGTTTGAGCTGAGCAGAATTCCTCCTTCCCCTCGTGGCGTTCCACAGATCGAGGTCACCTTT 1440
 M  T  K  D  N  N  L  L  G  K  F  E  L  S  R  I  P  P  S  P  R  G  V  P  Q  I  E  V  T  F

GACATTGACGCCAATGGCATCATGAACGTGTCGGCTGCGGACAAGAGCACCGGCAAAGAGAACAAGATCACCATCACCAACGACAAGGGC 1530
 D  I  D  A  N  G  I  M  N  V  S  A  A  D  K  S  T  G  K  E  N  K  I  T  I  T  N  D  K  G

AGGCTGAGCAAAGAGGATATCGACAGGATGGTGCAGAATGCGGAGAAATACAAAGCAGAAGATGATGCACAGAGGGCGAATATAGCTGCC 1620
 R  L  S  K  E  D  I  D  R  M  V  Q  N  A  E  K  Y  K  A  E  D  D  A  Q  R  A  N  I  A  A

AAGAACTCCCTGGAGTCCTACGCCTTCAACATGAAGAGTTCAGTGGAGGACGAGAACATGAGAGCGGAGGTTAGCGTAGACGATAAGGAG 1710
 K  N  S  L  E  S  Y  A  F  N  M  K  S  S  V  E  D  E  N  M  R  A  E  V  S  V  D  D  K  E

AAAGTGATTGACATGTGTAACCAGACCATCTCCTGGCTGGACAGCAACCAGCTGGCTGAGAAGGACGAGTATGAGCACAAGAAGAAGGAA 1800
 K  V  I  D  M  C  N  Q  T  I  S  W  L  D  S  N  Q  L  A  E  K  D  E  Y  E  H  K  K  K  E

CTGGAGCAAGTGTGCAAACCCATCATCGCCAAGCTGTACCAGGGGGGCACGCCAGGAGGCTCTTGTGGAGCCCAAGCAGGCAAAGCTACC 1890
 L  E  Q  V  C  K  P  I  I  A  K  L  Y  Q  G  G  T  P  G  G  S  C  G  A  Q  A  G  K  A  T

TACAGCCAAGGACCCACCATTGAAGAGGTCGATTAA 1926
 Y  S  Q  G  P  T  I  E  E  V  D  *
```

Figure 3.7: Nucleotide and amino acid sequences of the full-length *L. menadoensis hsp70*. The coding region was isolated with FL_Lm70 and Bac25/70R primers. The nucleotide sequence is indicated with lower case; the amino acid sequence is indicated with upper case. The numbering is based on the nucleotide sequence.

3.3.3 Bioinformatic analysis of the full-length *L. menadoensis* Hsp70

A PCR reaction to isolate the full-length *L. menadoensis hsp70* coding region yielded a 1. 926 kb fragment (Figure 3.7). The fragment encoded a full-length Hsp70 protein with 84%, 81%, 80% and 76% identity to the bovine Hsc70, *Fugu*, zebrafish and human Hsp70s, respectively.

The alignment (Figure 3.8) revealed a 100 % conservation of residues involved in the interaction of Hsp70s with ADP/ATP molecules. The LmHsp70 protein sequence had two extra residues at the N-terminus and eleven at the C-terminus than the bovine Hsc70. Therefore, residues Asp^{10}, Thr^{13}, Thr^{14}, Tyr^{15}, Glu^{125}, Asp^{199}, Thr^{204}, Asp^{206}, Gly^{339} and Asp^{366} of the bovine Hsc70 (Flaherty *et al*., 1994) corresponded to residues Asp^{12}, Thr^{15}, Thr^{16}, Tyr^{17}, Glu^{127}, Asp^{201}, Thr^{206}, Asp^{208}, Gly^{341} and Asp^{368}.

Also conserved in the ATPase domain were residues equivalent to Try^{149}, Asn^{151}, Asp^{152}, Glu^{218} and Val^{219} in bovine Hsc70. These residues were found to be equivalent to the residues Try^{145}, Asn^{147}, Asp^{148}, Glu^{217} and Val^{218} in DnaK, which have been shown to form the site of interaction with Hsp40 (Gässler *et al*., 1998; Suh *et al*., 1998).

In the substrate binding domain, residues equivalent to residues Thr^{405}, Ala^{406}, Phe^{428}, Val^{438}, Ile^{440} and $Ile^{474}$of bovine Hsc70, which form a hydrophobic cluster for interaction with the peptide substrate were also conserved (boxed in blue in Figure 3.8; Morshauser *et al*., 1999). Also conserved were residues equivalent to Ala^{406} and Tyr^{431} in bovine Hsc70. These two residues form a bridge over peptide binding channel the Hsc70 (Morshauser *et al*., 1999). Residues Lys^{544} and Ala^{552} shown to be involved in the interaction of the helical lid and the substrate binding domain of Hsc70 (Zhu *et al*., 1996) were equivalent to Glu^{546} and Ser^{554} in LmHsp70. LmHsp70 sequence terminated with a C-terminal EEVD motif, known to be the site of interaction with the Hsp70/Hsp90 organizing protein, Hop (Scheufler *et al*., 2000; Figure 3.8).

The homology model of LmHsp70 using the crystal structure of the bovine Hsc70 as template revealed high level of similarity to the bovine Hsc70 (Figure 3.9). Residues

Ile^{216}, Ile^{515} and Val^{519} which form a hydrophobic interaction, salt bridges between Glu^{530} and Lys^{325} and between Lys^{524} and Asp^{152} at the ATPase and substrate binding domain interface have recently been shown to play a role in bovine Hsc70 interdomain communication (Jiang *et al*., 2005). Equivalent residues at the domain interface of LmHsp70 were also conserved (Figure 3.8 and Figure 3.9). However, Hsc70 residue Glu^{530} was replaced by Asp^{532} in LmHsp70. LmHsp70 residues V^{390} and L^{395} equivalent to V^{388} and L^{393}, in the Hsc70 linker region, where also conserved.

Figure 3.8 on page 72

Figure 3.8: Alignment of the primary amino acid sequences of bovine Hsc70 (bHsc70), LmHsp70 and LcHsp70. The N-terminal ATPase domain (underlined with a blue line) has the following functional and structural elements: phosphate binding sites; connect 1 and connect 2 and region of adenosine coordination (Bork *et al*., 1992). Residues boxed in red are involved in the interaction with ATP/ADP molecules (Flaherty *et al*., 1994). The linker region (underlined with a broken line) joins the ATPase and the substrate binding domain (underlined in red). Residues of the linker region shown in bold and italics are important for Hsp70 chaperone function (Jiang *et al*., 2005). The C-terminal polypeptide domain (underlined in green) forms a helical lid (Zhu *et al*., 1996; Chou *et al*., 2003) that terminates with the EEVD motif (boxed in black) involved in the interaction of Hsp70 with Hop (Scheufler *et al*., 2000). Residues of the substrate binding domain boxed in blue interact with the peptide substrate (Zhu *et al*., 1996). Interaction between the substrate binding and helical lid occurs through residues highlighted in magenta (Zhu *et al*., 1996). Bold residues and bold underlined residues form salt bridges and hydrophobic interaction at the subdomain interface, respectively (Jiang *et al*., 2005). Identical (*****) and similar (**.** and **:**) residues are indicated. The black shading indicates identical residues, the grey shading indicates similar residues, and the white shading highlights differences.

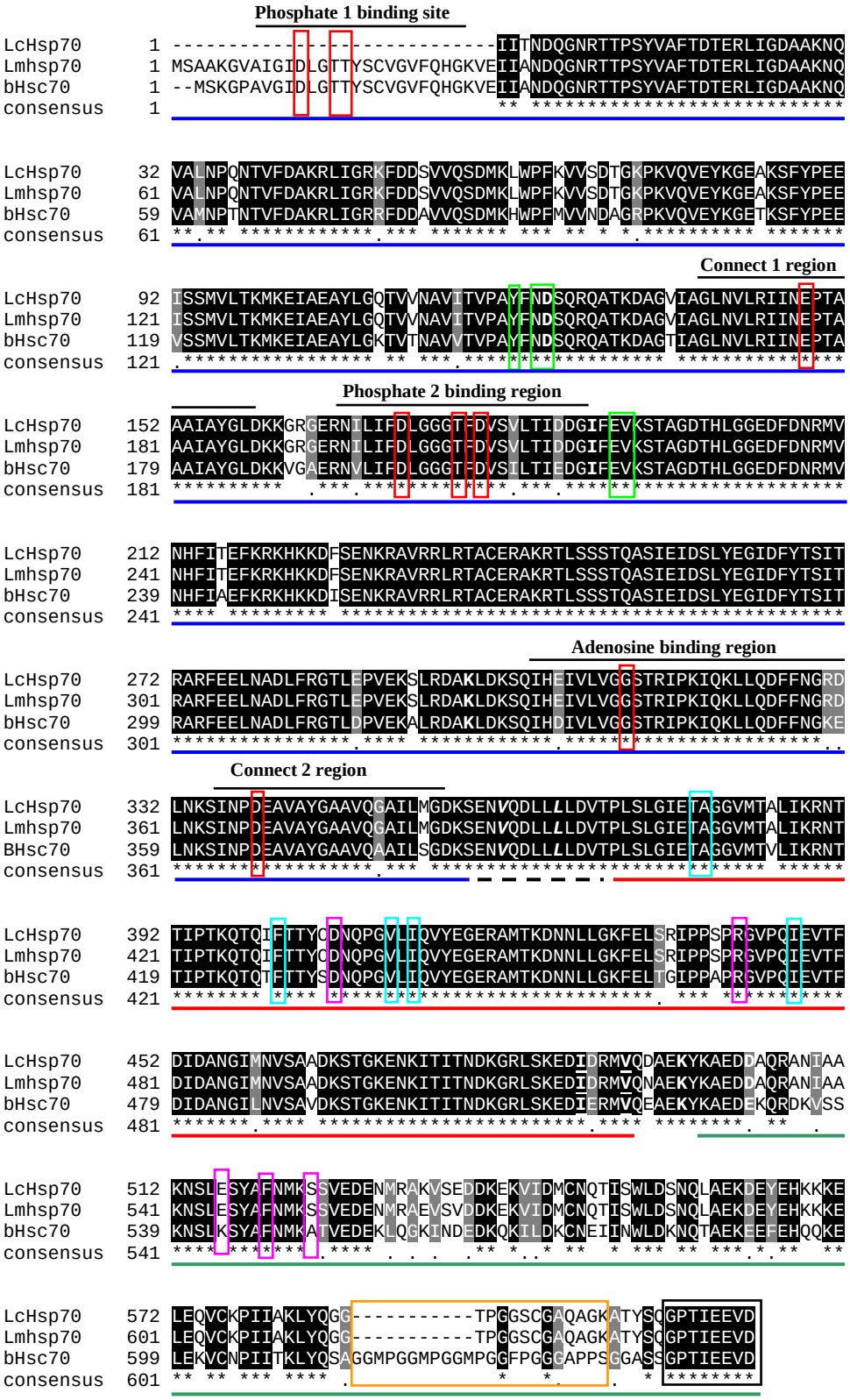

Phosphate 1 binding site
Connect 1 region
Phosphate 2 binding region
Adenosine binding region
Connect 2 region
LcHsp70
Lmhsp70
bHsc70
BHsc70
consensus

Figure 3.8.

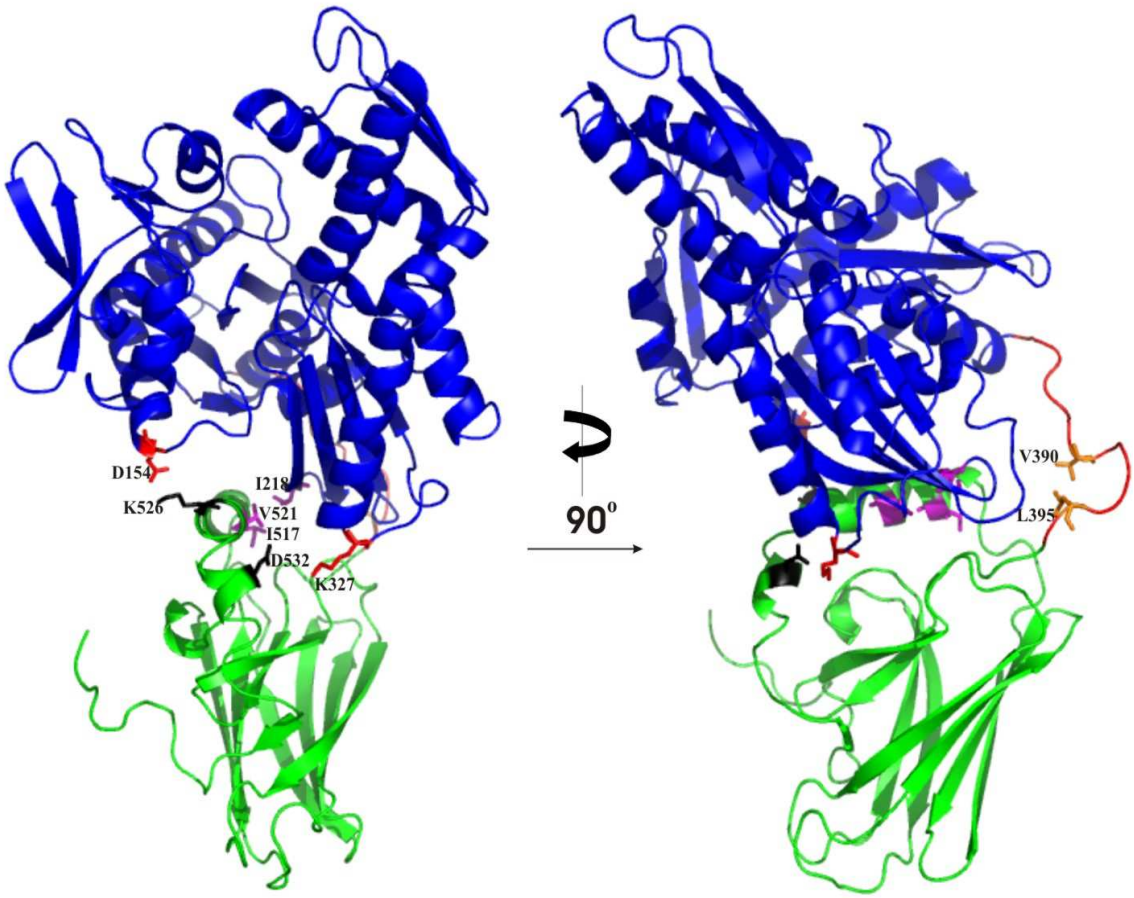


Figure 3.9: Ribbon representation of the homology model of LmHsp70. The model was generated using SWISSMODEL browser in First Approach Mode (http://swissmodel.expasy.org/ [Peitsch, 1996]) with bovine Hsc70 (PDB code: 1YUW; Jiang *et al*., 2005) as template. The structure was visualized using PyMol (DeLano *et al*., 2003). Potential salt bridges occur between D152 and K526, and K327 and D532 (black and red). Potential hydrophobic interactions occur between I218, I517 and V521 (magenta). V390 and L395 in the linker region are shown in orange. The ATPase domain is shown in blue, the linker region in red and the substrate binding domain in green.

3.4 Discussion

The isolation of an intronless 1.926 kb *hsp70* fragment encoding for an Hsp70 protein highly identical to the *L. chalumnae* partial full-length Hsp70 (LcHsp70) supports the suggestion that the coelacanth genome does encode an inducible Hsp70-based cytoprotection mechanism. The presence of conserved regions of a typical Hsp70 ATPase domain as also found in the LcHsp70 further strengthens the suggestion that the coelacanth Hsp70 expressed by these genes is functionally regulated by an ADP/ATP cycle (Bork *et al*., 1992, Flaherty *et al*., 1994). As also found in LcHsp70, Lmhsp70 had a conserved C-terminal EEVD which suggests possible chaperone cooperation with Hsp90 coordinated by Hop (Scheufler *et al*., 2000). The successful isolation of the full-length *Lmhsp70* coding region makes it possible to investigate if the coelacanth Hsp70s can be over-produced in heterologous expression system. It will also be interesting to investigate if LmHsp70 can functionally replace Hsp70s in a heterologous system in which the *hsp70* genes have been deleted or mutated.

The isolation of the 1.227 kb upstream of the LmHsp70 encoding region comprising of predicted promoter and heat shock element further suggests that the gene we have obtained is most likely functional and more particularly inducible. While the *Lmhsp70* regulatory region contained the predicted HSEs and promoters within a 1 kb region upstream of the putative methionine start, the zebrafish *hsp70* regulatory region was reported to containing the same motifs with a 640 bp region upstream of the ATG encoding the putative initiation methionine (Halloran *et al*., 2000). The sequences of the HSEs in the zebrafish *hsp70* promoter, the proximity to each other and to the TATA-box or CCAAT-box were not reported.

Lmhsp70 proximal HSE1 was found at 411 bp upstream of the initiation methionine. While the intermediate HSE2 and distal HSE3 were found at 486 and 972 bp upstream of ATG encoding the initiation methionine, respectively (Figure 3.6). Although the number of HSEs in the *Lmhsp70* regulatory region corresponded to the three of HSEs in the tilapia *hsp70* regulatory region (Molina *et al*., 2000), their position where found to differ. The three tilapia *hsp70* HSEs have been found at 872

bp, 671 bp and 350 bp upstream of the ATG encoding for the initiation methionine, respectively (Molina *et al*., 2000).

When using the *Lmhsp70* transcription start on the putative promoter 1, the transcription start would be 122 bp upstream of the ATG codon encoding the initiation methionine and the most proximal *Lmhsp70* HSE would be at 289 bp upstream of the transcription start. This was interesting since the tilapia *hsp70* most proximal HSE was found 179 upstream of the transcription start.

The position of the most proximal HSE in both the tilapia *hsp70* and *Lmhsp70* promoters differed significantly to the position of the most proximal HSEs in both the mouse *hsp68* (Adám *et al*., 2003) and the *Xenopus hsp70* (Adám *et al*., 2003) promoters which occurred at 98 a 106 bp upstream of the transcription start. Although the number of HSEs correlated across the *Lmhsp70,* tilapia *hsp70,* mouse *hsp68,* and *Xenopus hsp70* promoters, all the HSEs, TATA-box and CAAT-box occurred within 300 bp in the mouse *hsp68* promoter and 230 bp in the *Xenopus hsp70* promoter. The TATA-box in Lmhsp70 upstream regulatory region was found at 46 bp upstream of the transcription start in promoter 1. Also found in this region were two CAAT-boxes at 603 and 946 bp upstream of the transcription start in promoter 1.

Given the structure of this 5´-promoter/regulatory region, it is plausible that it could drive the expression of reporter gene under heat stress conditions as a measure of thermal response in a suitable experimental system.

Chapter 4

Heterologous expression and functional analysis of *Latimeria menadoensis* and *Latimeria chalumnae* Hsp70 proteins in *Escherichia coli*

4.1 Introduction

The lack of experimental systems (e.g. cell/tissue cultures) for some organisms dictates the use of heterologous systems to produce proteins for biochemically or biophysical characterization. While there are several prokaryotic and eukaryotic heterologous expression systems, bacterial systems, particularly the *E. coli* system, are widely used (Baneyx, *et al.,* 1999). The *E. coli* system is favourable for two main reasons: firstly, its genome and genetics have been well studied and secondly, it has high growth rates (Baneyx, *et al.,* 1999). Other reasons for the wide usage of *E. coli* system include the wide variety of expression vector systems engineered for production of high levels of target protein in a controlled manner (Nicoll *et al*., 2006). Nicoll and co-workers have reported general successful usage of the T5 promoter based protein expression vectors such the Qiagen *pQE30* vector for the production of target chaperone proteins. (Nicoll *et al*., 2006).

In most organisms there is a perceptible preference for certain codons some of which may be rarely used by *E. coli*. These so-called rare codons present a problem with regards to the heterologous overproduction of proteins in the *E. coli* system where they diminish the levels of full-length proteins due to translational stalling (Matambo *et al*., 2004). Codons AGG, AGA encoding arginine, AUA encoding isoleucine and GGA encoding glycine, have been shown to be particularly problematic when overproducing a heterologous protein in an *E. coli* host (Baca *et al*., 2000).

Nicoll and co-workers (Nicoll *et al.,* 2006) reviewed three methods that have been successfully used to overproduce molecular chaperones in *E. coli* hosts:

(i) The co-expression of arginyl (AGA/AGG), isoleucyl (AUA), and glycyl (GGA) tRNAs from the RIG plasmid constructed by Baca and co-workers (Baca *et al*., 2000) has been used to significantly increase the yield of *Plasmodium falciparum* full-length Hsp70 in *E. coli* (Matambo *et al.,* 2004).

(ii) Codon optimization provides a method of re-synthesizing the coding region of the target gene to introduce frequently used codons for the host organism (Nicoll *et al.,* 2006). Codon optimisation has been used to re-synthezise a *P. falciparum* Hsp40 gene to obtain high-level expression in *E. coli* (Nicoll *et al.,* 2006).

(iii) Codon homonization is a further development of codon optimisation wherein the codons in the target coding sequence are converted to codons used by the host strain with similar codon usage frequencies thereby preserving the fluctuations in the original target coding region (Hillier *et al.,* 2005).

A few reports are published on the expression of coelacanth proteins in heterologous systems. The reported cases include the expression of wild type and mutant *L. menadoensis* and *L. chalumnae* RH1 and RH2 visual pigments in Cos1 cells to study their light sensitivity (Yokoyama *et al*., 1999; Yokoyama and Tada, 2000).

Although eukaryotic and prokaryotic Hsp70s have very similar structures, they cooperate with different cochaperones to carry out their functions. The ATPase cycle of the prokaryotic Hsp70, DnaK, requires interaction with Hsp40 and the nucleotide exchange factor GrpE (Brehmer *et al*., 2001, Suppini *et al*., 2004). Although eukaryotic Hsp70s interact with Hsp40s, they do not interact with GrpE. Eukaryotic Hsp70s interact with eukaryotic specific nucleotide exchange factors such as Bag-1 (Brehmer *et al*., 2001). Suppini and co-worker found that the Hsp70 protein chimeras consisting of the *E. coli* DnaK ATPase domain and the rat Hsc70 could not reverse thermosensitivity of a thermosensitive *E. coli* DnaK mutant strain (Suppini *et al*., 2004). In contrast, Shonhai and co-workers have demonstrated that *Plasmodium falciparum* Hsp70 can functionally reverse thermosensitivity of an *E. coli* strain with a mutant DnaK. This provides some

eukaryotic Hsp70 substrate binding domains do recognise prokaryotic protein substrates (Shonhai *et al*., 2005). The functional differences between prokaryotic and eukaryotic Hsp70s necessitate the investigation on the ability of eukaryotic Hsp70 proteins to functionally replace prokaryotic Hsp70s (Suppini *et al*., 2004).

The aim of the study presented in this chapter was to investigate the suitability of the *E. coli* protein expression system to overproduce coelacanth chaperone proteins, especially the Hsp70s from a suitable vector. The expression of this protein in *E. coli* is desirable in accordance with the reasons mentioned above and the fact that there are no coelacanth experimental systems that could be employed for this purpose. It was also the aim of the study to determine the ability of the coelacanth Hsp70 (particularly the *L menadoensis* Hsp70) to reverse thermosensitivity of a thermosensitive *E. coli* strain with a dysfunctional DnaK.

4.2 Experimental procedures

4.2.1 Plasmid constructs

The *Lchsp70* coding region was polymerase chain reaction (PCR) amplified from *pGEM-T_Lchsp70* construct using a forward primer *Bam* HI_Lchsp70F: **GGA TCC** ATC ATC ACC AAT GAC (*Bam* HI site indicated in bold) and a reverse primer *Pst* I_Lchsp70R: **CTG CAG** ATT AGT CGA CTT CTT (*Pst* I site indicated in bold). The PCR reactions and ligation reactions into pGEM-T were carried out as described in Appendix A3-A6 and further analyzed as in section 2.2.5. The insert fragment from the appropriate plasmid DNA was gel purified using the GFX gel purification kit (Amersham, Sweeden) subsequent to a restriction digestion with *Bam HI* and *Pst* I (as described in Appendix A6). The gel purified fragment was ligated into a *pQE30* protein expression vector (Qiagen, Germany) digested with *Bam* HI and *Pst* I (as described in Appendix A6). The resultant construct was named *pQE30_ Lchsp70.*

Similarly the full-length *Lmhsp70* was PCR amplified from *pGEM-T_FLLmhsp70* construct using primers *Bam* HI_Bac25_70F: **GGA TCC** ATG TCT GCT GCA GCC (*Bam* HI site indicated in Bold) and *Kpn* I_BAC25_70R: **GGT ACC** TTA ATC GAC CTC TTC AAT G (*Kpn* I site indicated in bold) and the PCR product was ligated into pGEM-T as outline above. DNA sequencing revealed an error in the *Kpn* I restriction site engineered at the 5´-end of the *Kpn* I_BAC25_70R primer. Since a *Sal* I site was found in the multiple cloning site of pGEM-T easy vector 3′ to the *Lmhsp70* coding region, the insert fragment was released by *Bam* HI and *Sal* I restriction digest and gel purified as outlined above. The gel purified fragment was ligated into a *pQE30* protein expression vector digested with *Bam* HI and *Sal* I and the resultant construct was named *pQE30_Lmhsp70.*

The RIG plasmid was kindly provided by Professor W. G. J. Hol (Biomolecular Structure Center, University of Washington, USA). The *pQE60* and pBB46 (Amp^R and encoding

the *E. coli* DnaK) plasmids were kindly provided by Dr. W. Burkholder (Stanford University, USA).

4.2.2 Codon usage tables

Latimeria menadoensis, *Homo sapiens*, *Saccharomyces cerevisiae* and *Escherichia coli* codon usage tables were accessed at www.kazusa.or.jp/codon on 14 August 2005

4.2.3 Protein expression study

pQE30_Lchsp70 and *pQE30_Lmhsp70* where transformed into *E. coli* XLI Blue and plated onto 2x YT agar plates supplemented with 100 μg/ml ampicillin, followed by overnight incubation at 37ºC. Single colonies were inoculated into 25 ml 2x YT broth supplemented with 100 μg/ml ampicillin, followed by an overnight incubation at 37ºC with shaking. The overnight 25 ml cultures were inoculated into 225 ml 2x YT broth supplemented with 100 μg/ml ampicillin and incubated at 30ºC with shaking until mid log phase (OD_{600} 0.4-0.6 units).

Similarly, co-transformation of the *pQE30_Lchsp70* or *pQE30_Lmhsp70* constructs with the RIG plasmid was carried out. The transformants were grown on 2x YT agar plates or in 2x YT broth cultures supplemented with 100 μg/ml ampicillin and 34 μg/ml chloroamphenicol.

At mid-log phase samples (2 x 1ml) were collected from each culture. One sample was used to determine the A_{600} reading and the second sample was centrifuged at 16000 x g to pellet the cells which were resuspended in phosphate buffered saline (PBS: 137 mM NaCl; 27 mM KCl; 4.3 mM Na_2HPO_4; 1.4 mM KH_2PO_4). The resuspension of the pellets in the standard manner (A600/0.5 x 150 μl) ensured that an equal number of cells was analyzed for each sample.

Following the collection of pre-induction samples, protein production was induced by the addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). Samples were collected every hour over 6 hours after which overnight samples were collected as per the method outlined above. The samples were analyzed by SDS-PAGE and Western analysis descried in Appendix A10 and A11.

4.2.4 LmHsp70 functional analysis by *in vivo* complementation assays

E. coli BB2362 strain was used to determine the ability of LmHsp70 to functionally replace *E. coli* DnaK under heat shock conditions. *E. coli* BB2362 is sensitive to growth temperatures above 40˚C. The temperature sensitivity of *E. coli* BB2362 is due to three glycine-to-aspartate mutations in DnaK (Buchberger *et al*., 1999). The G32D mutationdisturbs optimal effects of GrpE binding while the G455D and G468D mutations are in the substrate binding domain (Buchberger *et al*., 1999). In a complementation assay, the ability of an Hsp70 to functionally replace the mutated endogenous DnaK is investigated. The plasmids *pQE30* (Qiagen Germany) and *pQE60* (Qiagen Germany) were used as negative controls while *pBB46* plasmid (Burkholder *et al*., 1996) was used as a positive control. The untransformed *E. coli* BB2362 was used as a strain control.

The plasmids *pBB46*, *pQE30*, *pQE60* and *pQE30_Lmhsp70* were transformed into *E. coli* BB2362 (*dnaK756 recA::Tc*^{R} *pDM1*; kindly provided by Prof. B. Bukau [University of Heidelberg, Germany]) and plated onto 2x YT agar plates supplemented with 50 μg/ml kanamycin, 100 μg/ml ampicillin and 15 μg/ml tetracycline. The untransformed *E. coli* BB2362 was also streaked onto 2x YT agar plates supplemented with 50 μg/ml kanamycin and 15 μg/ml tetracycline. Colonies were inoculated into 10 ml 2x YT broth supplemented with appropriate antibiotics and incubated overnight at 37°C with shaking. The overnight cultures were used to inoculate 90 ml 2x YT broth supplemented with appropriate antibiotics. These cultures were shook at 37°C with shaking until OD_{600} of 2 was reach. At OD_{600} of 2 units, the cultures were diluted to OD_{600} of 0.2 and further serial dilutions from 10^{0} to 10^{-5} were prepared. A sample of each dilution (2 μl) was plated on duplicate 2x YT agar plates supplemented with appropriate antibiotics and 50 μM IPTG.

The plates were incubated at a normal growth temperature of 37°C while a duplicate set of plates at a heat shock temperature of 42°C. To investigate the effect of inducing LmHsp70 prior to heat shock 0.5 mM IPTG was added to the cultures at an OD_{600} reading of 1.0 and further treated as described above.

4.3 Results

4.3.1 Codon usage analysis

Baca and co-workers revealed that parasite genes frequently used the codons AGA/G (Arg), AUA (Ile), and GGA (Gly) which are rarely used in *E. coli* (Baca *et al*., 2000). The usage of said codons results in translational stalling when the parasite genes are expressed in *E. coli* (Baca *et al*., 2000). The "*RIG*" plasmid containing the genes that encode three tRNAs (Arg, Ile, Gly) was constructed to overcome this potential codon-bias problem. The tRNAs recognize rare codons found in parasite genes when the *RIG* plasmid is co-transformed into *E. coli* along with expression plasmids containing parasite genes (Baca *et al*., 2000).

Comparison of codon usages from the *L. menadoensis*, *H. sapiens*, *S. cerevisiae* and *E. coli* revealed that the coelacanth generally followed the eukaryotic codon usage (Figure 4.1). As seen in Figure 4.1, the *L. menadoensis* used the AU (U/C) codons encoding Ile less frequently as compared to *E. coli* and had a high preference for AUA to code for the same amino acid. *L. menadoensis* further displayed a much higher usage of AG (A/G) and GGA encoding for Arg and Gly, respectively. Particularly interesting was the finding that the *L. menadoensis* codon usage displayed a high frequency usage of the AGG codon for Arg that was almost not used in *E. coli* (Figure 4.1). In general *L. menadoensis* codon usage was overly closer to the two other eukaryotes in the comparison (Figure 4.1). The same usage was found in the coding region of the Lc*hsp70*. These findings were to be considered when overproducing coelacanth Hsp70 in *E. coli*.

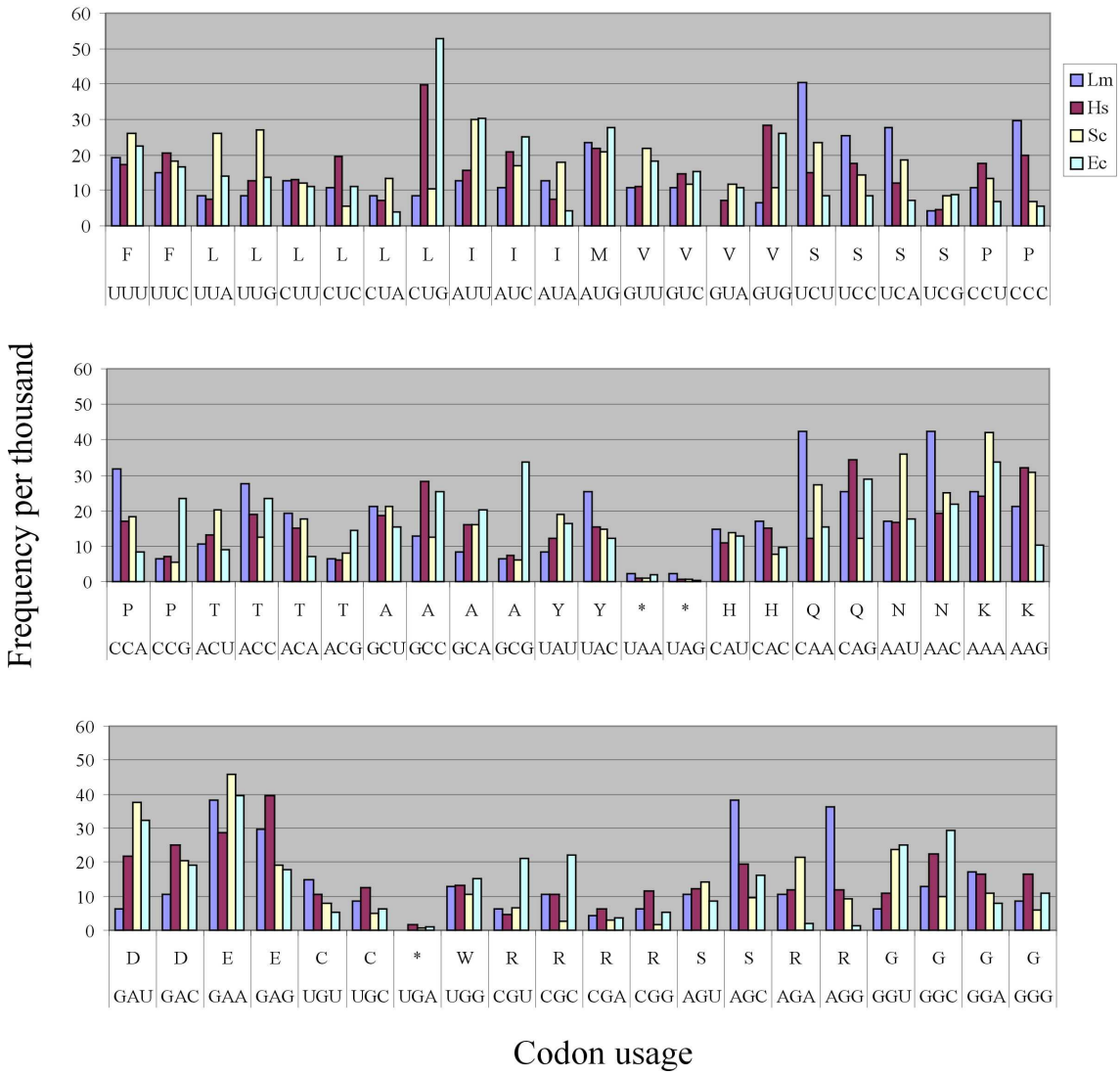


Figure 4.1: Comparison of codon usages from *L. menadoensis* (Lm), *Human* (Hs), *Saccharomyces cerevisiae* (Sc) and *E. coli* (Ec). Amino acids are indicated as solitary capital letters; the codon encoding the amino acids are indicated as the triplet capital letters; the * indicates stop codons. The coelacanths showed a relatively lower usage of AU (U/C) codons encoding Ile, but a relatively higher usage of AG (A/G) and GGA encoding for Arg and Gly, respectively. *L. menadoensis*, *Homo sapiens*, *Saccharomyces cerevisiae* and *E. coli* codon usage tables were accessed at www.kazusa.or.jp/codon on 14 August 2005.

Although the coelacanth codon used was found generally similar to that of eukaryoutes, the bacterial protein expression system, particularly the *E. coli* system was investigated for the overproduction of coelacanth Hsp70. *E. coli* was favourable because its genome and genetics have been well studied and it has high growth rates (Baneyx, *et al.*, 1999). In

addition there is wide variety of expression vector systems engineered for production of high levels of target protein in a controlled manner in *E. coli* (Nicoll *et al*., 2006).

4.3.2 Heterologous overproduction of the LmHsp70 and LcHsp70

The *pQE30* vector is a T5 promoter-based bacterial expression vector which allows for the insertion of a target coding region in frame with upstream codons for six histidines (Appendix Figure B2.2). The protein produced off the *pQE30* vector is therefore histidine tagged and allows for Western analysis using an anti-histidine primary antibody and an appropriate secondary antibody. To facilitate their overproduction in *E. coli*, *Lchsp70* and *Lmhsp70* coding region were ligated into the *pQE30* protein expression vectors as described in sections 4.2.1 (Figure 4.2). Subsequent to the transformation of the *pQE30_Lchsp70* and *pQE30_Lmhsp70* into *E. coli* XLI-blue with or without the RIG plasmid, a protein induction study was carried out as described in section 4.2.3 (Figure 4.3 and Figure 4.4).

When both the LcHsp70 and LmHsp70 were produced in *E. coli* XLI-Blue without the RIG plasmid, no detectable overproduction of protein was observed on SDS-PAGE analysis (upper panel Figure 4.3[A], lane 5-10 and upper panel Figure 4.4[A], lanes 5-8). However, chemiluminescence-based Western analysis revealed that full-length LcHsp70 and some lower molecular mass species were produced (lower panel Figure 4.3[A], lanes 5-10). A similar observation was made for LmHsp70 (lower panel Figure 4.4[A], lanes 5-8).

In the presence of the RIG plasmid, full-length LcHsp70 and LmHsp70 were overproduced in *E. coli* XLI-blue as detected by SDS-PAGE (upper panel Figure 4.3[B], lane 5-10 and upper panel Figure 4.4[B], lanes 5-8). Chemiluminescence-based Western analysis confirmed the overproduction of full-length LcHsp70 with the occurrence of some lower molecular mass species (lower panel Figure 4.3[B], lanes 5-10). A similar observation was made for LmHsp70 (lower panel Figure 4.4[B], lanes 5-8). Production of LmHsp70 in the *E. coli* JM109 strain did not yield a different result.

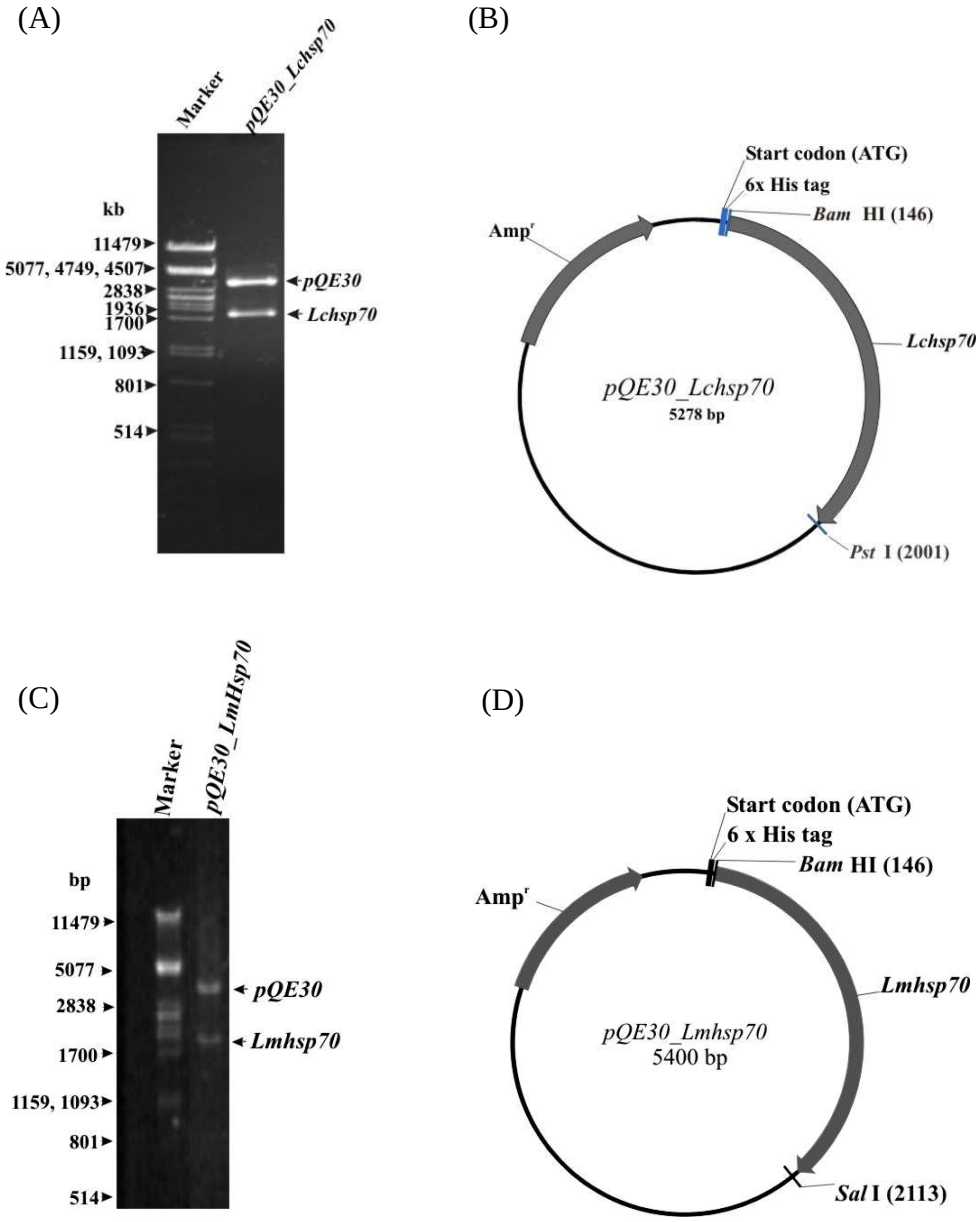


Figure 4.2: *pQE30_Lchsp70* and *pQE30_Lmhsp70* constructs for the overproduction of LcHsp70 and LmHsp70 in *E. coli* XLI-Blue. (A) A 0.8% agarose gel showing the *Bam* HI and *Pst* I restriction digest of the *pQE30_Lchsp70* construct releasing the *pQE30* vector and the *Lchsp70* coding region. (B) The plasmid map of the *pQE30_Lchsp70* construct highlighting the *Bam* HI and *Pst* I sites and the coding region for the 6xHis-tag. (C) A 0.8% agarose gel showing the *Bam* HI and S*al* I restriction digest of the *pQE30_Lmhsp70* construct releasing the *pQE30* vector and the *Lmhsp70* coding region. (D) The plasmid map of the *pQE30_Lmhsp70* construct highlighting the *Bam* HI and *Sal* I sites and the coding region for the 6xHis-tag. In (B) and (D), the grey arrows indicate the ampicillin resistance (Amp^r) gene and *Lchsp70*/*Lmhsp70* coding region. Marker = λ *Pst* I

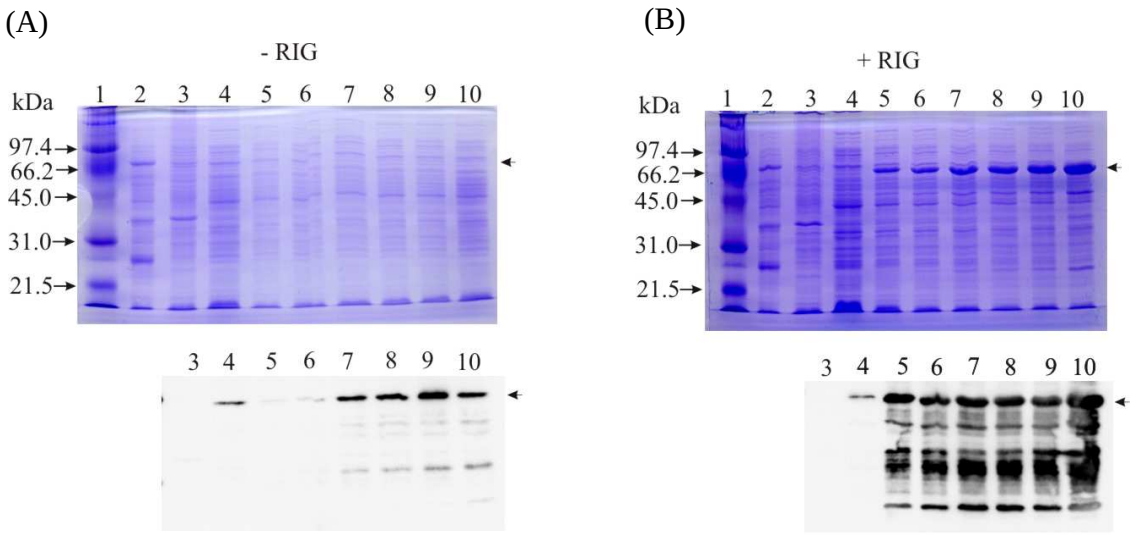


Figure 4.3: Induction study of *L. chalumnae* Hsp70 (LcHsp70) from an *pQE30_Lchsp70* construct in *E. coli* XLI-Blue with (+RIG) and without (-RIG) the RIG plasmid. (A) Upper panel: 12% SDS-PAGE analysis of *E. coli* [–RIG and *pQE30_Lchsp70*] total cell lysates; Lower panel Western analysis of the total lysates resolved in top panel with an anti-His antibody. (B) Upper panel panel: 12% SDS-PAGE analysis of *E. coli* [+RIG and *pQE30_Lchsp70*] total cell lysates. Lower panel: Western analysis of the total lysates resolved in top panel with an anti-His antibody. 1= molecular weight marker; 2= positive control (to mark the 70kDa protein); 3= negative control (extract from *E. coli* [–RIG or +RIG] 4= pre induction; 5-9= 1 hourly sample post induction; 10 = overnight sample post induction. The study was carried out at 1mM IPTG in 2xYT broth at 30°C. The arrows indicate the 70 kDa His-tagged LcHsp70.

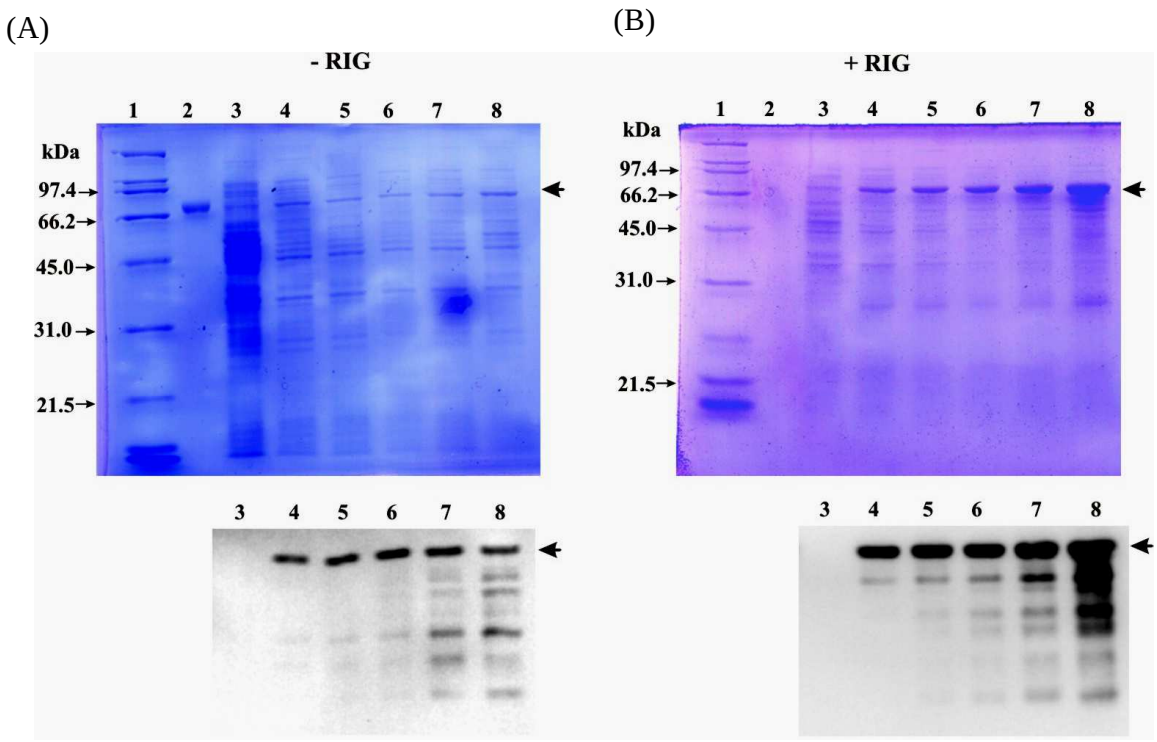


Figure 4.4: Induction study of *L. menadoensis* Hsp70 (LmHsp70) from an *pQE30_Lmhsp70* construct in *E. coli* XLI-Blue with (+RIG) and without (-RIG) the RIG plasmid. (A) Upper panel: 12% SDS-PAGE analysis of *E. coli* [–RIG and *pQE30_Lmhsp70*] total cell lysates; Lower panel Western analysis of the total lysates resolved in top panel with an anti-His antibody. (B) Upper panel: 12% SDS-PAGE analysis of *E. coli* [+RIG and *pQE30_Lmhsp70*] total cell lysates. Lower panel: Western analysis of the total lysates resolved in top panel with an anti-His antibody. 1= molecular weight marker; 2= positive control (to mark the 70kDa protein); 3= negative control (extract from *E. coli* [–RIG or +RIG] 4= pre induction; 5= two hours post induction; 6= four hours post induction; 7= six hours post induction; 8 = overnight sample post induction. The study was carried out at 1mM IPTG in 2xYT broth at 30°C. The arrows indicate the 70 kDa His-tagged LcHsp70.

4.3.3 Functional analysis LmHsp70 by complementation assay in *E. coli* BB2362

E. coli BB2362 strain is sensitive to growth temperatures above 40˚C. The temperature sensitivity of *E. coli* BB2362 is due to three glycine-to-aspartate mutations in DnaK (Buchberger *et al*., 1999). Complementation assays measures the ability of a test protein to functionally replace the dysfunctional protein thereby restoring the normal phenotype. The ability of LmHsp70 to reverse thermal sensitivity of *E. coli BB2362* was used a test for LmHsp70 chaperone activity.

As can be seen in the 37°C panel of Figure 4.5A the untransformed *E. coli BB236, the E. coli BB2362* transformed with *pBB46, pQE30, pQE60 and pQE30_ Lmhsp70* all grew at 37°C. In the 42°C panel of Figure 4.5A, the untransformed *E. coli BB236, E. coli* transformed with *pQE30, pQE60 and pQE30* (negative controls) displayed thermosensitivity. However, *E. coli BB2362* transformed with *pBB46* could reverse thermosensitivity of the strain while *E. coli BB2362* transformed with *pQE30_Lmhsp70* plasmid displayed thermosensitivity.

The investigation of the effect of pre-inducing Hsp70s prior to heat shock is shown in Figure 4.5B. The pre-induction of broth cultures with 0.5mM IPTG displayed no significant difference to data displayed in Figure 4.5B while pre-induction with IPTG concentrations above 0.5mM resulted in DnaK toxicity in the *E. coli BB2362* transformed with *pBB46.*

To confirm that failure of LmHsp70 to reverse thermal sensitivity *E. coli BB2362* was not due to lack of protein production, an LmHsp70 induction study was carried out in BB2362. As seen in the upper panel of Figure 4.6, lanes 5-7 no overproduction of LmHsp70 was detectable by SDS-PAGE analysis. However a chemiluminescence-based Western analysis revealed that full-length LmHsp70 was produced (lower panel of Figure 4.6, lanes 5-7). Some low molecular mass species were observed after 6 hour of induction (lower panel of Figure 4.6, lanes 6 and 7).

SDS-PAGE analysis of the pellet and soluble fractions of *E. coli BB2362* transformed with *pQE30_Lmhsp70* showed some protein in both fractions (upper panel of Figure 4.6, lanes 8 and 9). Chemiluminescence-based Western analysis revealed that full-length LmHsp70 was mostly in the soluble fraction (lower panel of Figure 4.6, lanes 8 and 9).

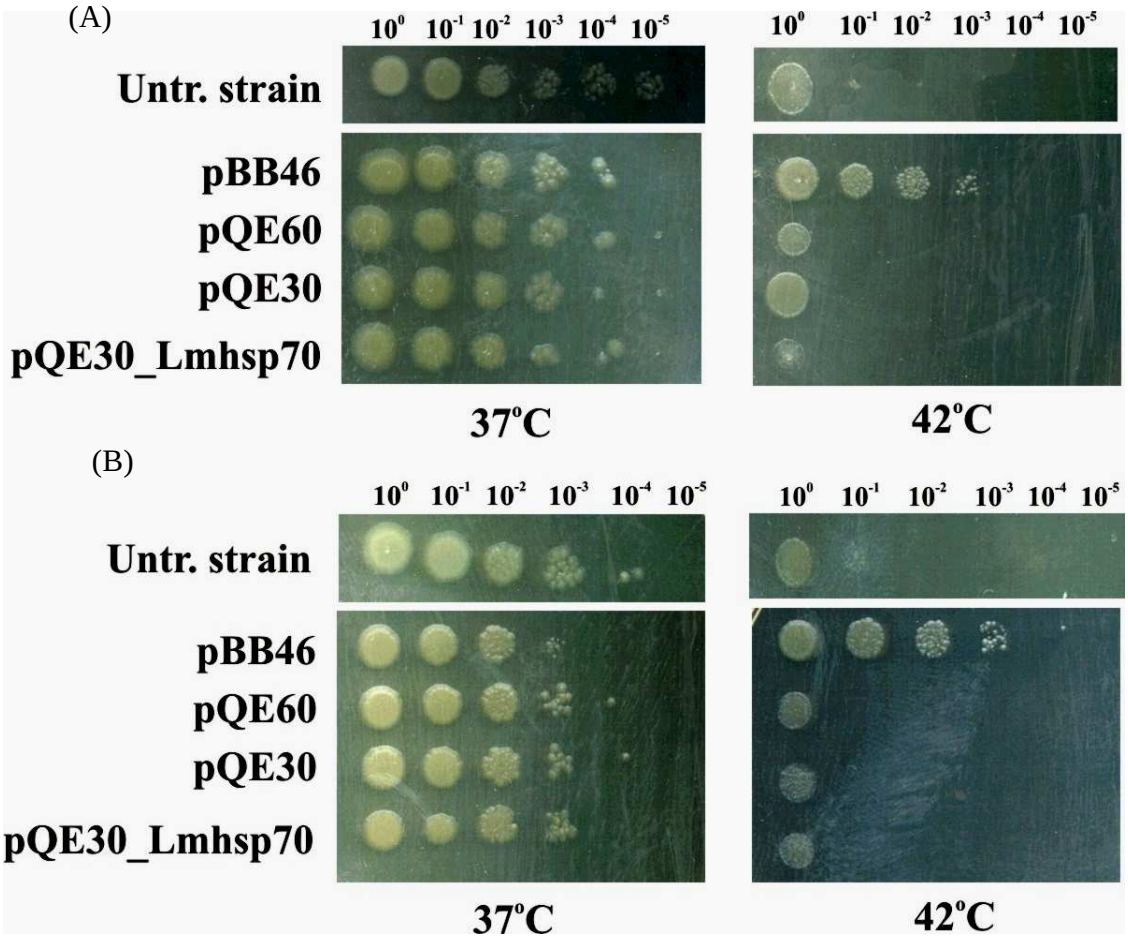


Figure 4.5: LmHsp70 failed to suppress thermal sensitivity of *E. coli* BB2362 strain. Plasmids *pBB46* (positive control), *pQE30* and *pQE60* (negative controls), *pQE30_Lmhsp70* were transformed into *E. coli BB2362*, and their ability to complement for lack of a fully functional DnaK was investigated at 42°C. (A) No IPTG was added to the broth culture prior to plating on 2x YT agar supplemented with 50 μM IPTG and appropriate antibiotics. (B) IPTG (0.5 mM) was added to the broth culture prior to plating on 2x YT agar supplemented with 50 μM IPTG and appropriate antibiotics. The names of plasmid used to transform *E. coli BB2362* are indicated on the left side of the panels. The incubation temperatures are indicated below each panel. The plated culture dilution are indicated above each panel.

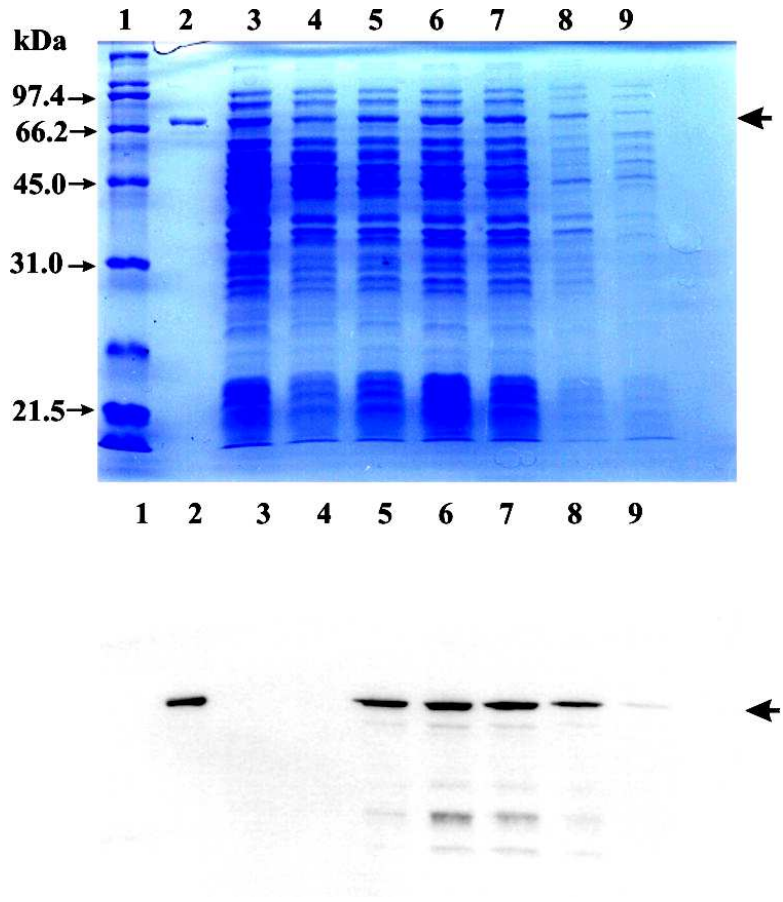


Figure 4.6: Successful production of LmHsp70 in *E. coli* BB2362 without the RIG plasmid. Upper panel: 12% SDS-PAGE analysis of *E. coli* BB2362 [*pQE30_Lmhsp70*] total cell lysates. Lower panel Western analysis of the total lysates resolved in top panel with an anti-His antibody. 1= molecular weight marker; 2= positive control (to mark the 70kDa protein); 3= negative control (total extract from *E. coli* BB2362); 4= pre induction; 5-7: 2 hourly sample post induction. A solubility analysis is depicted in lanes 8 and 9. 8= soluble fraction and lane 9= pellet fraction. The arrows indicate the 70 kDa His-tagged LcHsp70. The study was carried out at 1mM IPTG in 2x YT broth at 30°C.

4.5 Discussion

We hypothesized that the low level production of both the LcHsp70 and LmHsp70 proteins was due to the high level usage of the AUA, AGG and AGA coelacanth codons which are rarely used in *E. coli*. The usage of the RIG plamids, previously shown to relieve the translational stalling caused by these rare codons, resulted in a significant increase in the production of the partial full-length LcHsp70 and full-length LmHsp70. Our finding corresponded to the findings that the RIG plasmid improved the overproduction of *P. falciparum* full-length Hsp70 from a *pQE30* based protein expression vector, *pQE30_PfHsp70,* in *E. coli* (Matambo *et al*, 2004). Although Matambo and co-workers reported that the usage of the RIG plasmid decrease the production of the low molecular weight species, we have found a corresponding increase in the production of full-length protein and low molecular weight species. Our finding may suggest the use of alternative expression vectors to facilitate protein purification or the use of a eukaryotic expression system.

The *L. menadoensis* Hsp70 was unable to reverse thermosensitivity of the thermosensitive *E. coli* strain, BB2362. The expectation that the LmHsp70 would functionally replace *E. coli* DnaK assumed that the LmHsp70 had an ATPase activity and that its substrate binding domain could recognise *E. coli* substrate proteins.

Suppini and co-worker found that the Hsp70 protein chimeras consisting of the *E. coli* DnaK ATPase domain and the eukaryotic Hsp70 substrate binding domain, particularly the rat Hsc70, could not reverse thermosensitivity of an *E. coli* DnaK mutant strain (Suppini *et al*., 2004). However, Hsp70 chimeric proteins consisting of the rat Hsc70 ATPase domain and the DnaK substrate binding domain could successfully reverse thermosensitivity of an *E. coli* DnaK mutant strain (Suppini *et al*., 2004). Since the rat Hsc70 ATPase lacks the site for GrpE interaction, this finding strongly suggested the DnaK substrate binding domain to be the essential in the protection of *E. coli* at high temperatures (Suppini *et al*., 2004). These data also suggested that the eukaryotic Hsp70

substrate domain could not recognise prokaryotic protein substrates (Suppini *et al*., 2004).

Conversely, full-length *P. falciparum* Hsp70 and a chimeric protein consisting of the *E. coli* DnaK ATPase domain and the *P. falciparum* Hsp70 substrate binding domain were found to display an ability to reverse thermosensitivity to an *E. coli* DnaK mutant strain (Shonhai *et al*., 2005). This finding led to the speculation that some eukaryotic Hsp70s substrate binding domains have broad range protein substrate specificity.

The discovery that the LmHsp70 cannot reverse thermal sensitivity of *E. coli* DnaK mutant strain raises question about the LmHsp70 domain responsible for this observation. These findings thus necessitate an in-depth study of the functional and specificity aspects of LmHsp70.

Chapter 5

Discussion, conclusion and future work

5.1 Discussion

5.1.1 The intronless *Lchsp70* and *Lmhsp70* genes: the implications

The coding regions of most eukaryotic inducible *hsp70*s such as human (Bond, 1988), *Drosophila* (Yost and Lindquist, 1986), yeast (Yost and Lindquist, 1991), and fish (Lim and Brenner, 1999; Molina et al., 2000) have been found to be intronless as opposed to the constitutive intron-containing *hsp70* genes. RNA splicing has been shown to be inhibited during heat stress and Hsp70, has also been found to cooperate with Hsp104 to reactivate RNA splicing in cell extracts that had been heat-inactivated *in vitro* (Vogel *et al*., 1998, Yost and Lindquist, 1991). Based on these findings, the intronless characteristic of inducible *hsp70* genes was proposed to be connected to their expression and function during stress. A few exceptions wherein inducible *hsp70* genes in other organisms such as *Blastocladiella emersonii* contain introns have been shown. However, the splicing machinery in these fungi is highly thermoresistant (Stefani and Gomes, 1995). This finding therefore suggests that the intronless character of inducible *hsp70* genes is not a general rule (Černila *et al*., 2003). The absence of introns in the *Lchsp70* and *Lmhsp70* (Figures 2.4, 2.5 and 3.3) coding regions therefore strongly suggest the existence of an inducible Hsp70-based cytoprotection mechanism in the coelacanth which likely resembles the eukaryotes one. However, the restriction of intronless *hsp70* genes to heat stress induction can not be assumed since some inducible *hsp70s* such as the stress-inducible zebrafish *hsp70* gene have been shown to exhibit a strong but brief expression during a normal embryonic lens formation under nonstress conditions (Blechinger *et al*., 2002). The structural and functional analysis of the regulatory region upstream of *Lchsp70* and *Lmhsp70* genes therefore becomes important in a strategy to establish the suitability of these genes for heat shock induction.

Our bioinformatic analysis predict that the 1.2 kb upstream regulatory region of the *Lmhsp70* coding region contain three putative HSEs, six putative promoters and two CAAT-boxes (Figure 3.6). This *Lmhsp70* regulatory region displays some similarities to *Xenopus* (Bien and Pelham, 1986), zebrafish (Halloran *et al*., 2000), mouse (Fiszer-Kierzkowska *et al*., 2003) and tilapia (Molina *et al*., 2000), at least with regards to the number and structure of HSE (Figure 5.1).

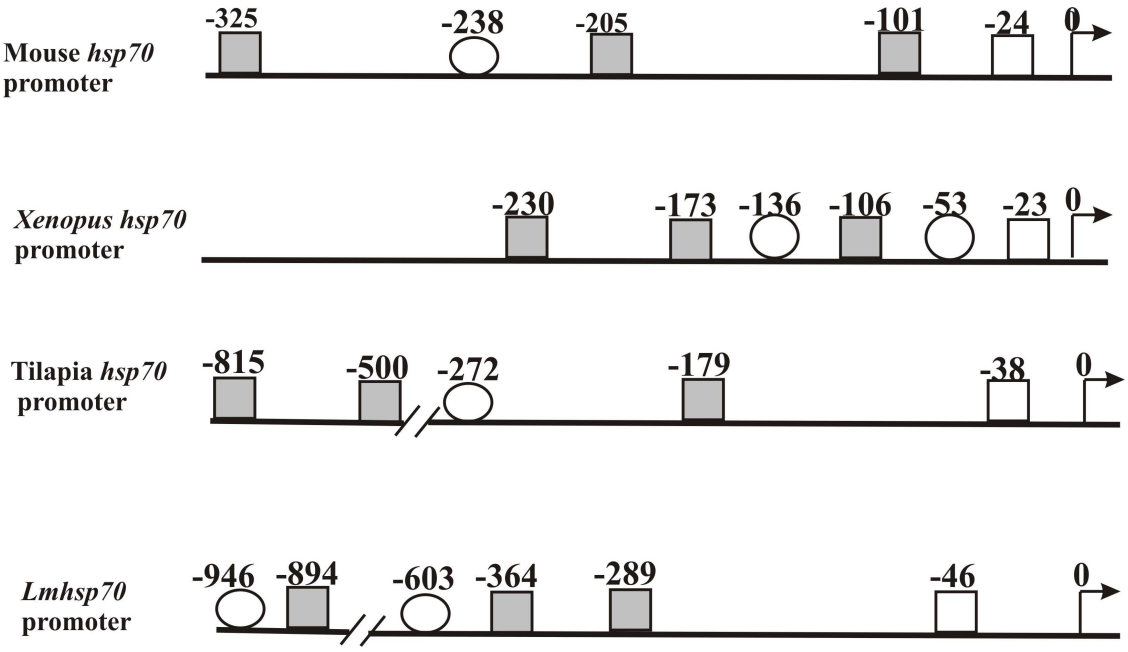


Figure 5.1: Comparison of inducible *hsp70* promoters. Comparison of the mouse (Fiszer-Kierzkowska *et al.*, 2003), *Xenopus* (Bien and Pelham, 1986), tilapia (Molina *et al.*, 2000) and *L. menadoensis hsp70* (predicted in this thesis) heat inducible promoters. Figure is not to scale. The filled squares represent the HSEs, the clear square represents the TATA box, the arrow represents the transcriptional start site, the circles represent the CCAAT-box, and the numbers indicates the position of the indicated elements upstream of the transcription start site.

The *Lmhsp70* regulatory region closely resembled the tilapia *hsp70* promoter. The HSEs in this promoter were dispersed within 800 bp upstream of the transcription start. In contrast, the mouse and *Xenopus hsp70* HSEs, TATA boxes and CCAAT boxes are all within at most 325 bp upstream of the transcription start. Despite the differences in the promoters, the mouse, *Xenopus* and tilapia *hsp70* promoters have been shown to drive the

expression of reporter genes under heat shock condition (Adám *et al*., 2000; Molina *et al*., 2000).

The *Lmhsp70* upstream regulatory region contained two CAAT (-603 and -946) elements as opposed to the CCAAT element. The position of these elements are greatly distal to the TATA box as compared to their relative position in the mouse, tilapia, and more particularly, the *Xenopus hsp70* promoters. Although the CCAAT and TATA elements have been shown to be important for *hsp70* basal expression only the TATA element have been found to be crucial for induced *hsp70* expression (Williams and Morimoto, 1990). Despite these differences, the *Lmhsp70* upstream regulatory region does strongly resemble the heat inducible eukaryotic *hsp70* promoters in terms of the essential elements..

5.1.2 The LcHsp70 and LmHsp70 protein structure and function

LcHsp70 and LmHsp70 were found to have a high level of sequnce identity to the human, bovine, *Fugu* and zebrafish Hsp70s. The predicted homology model of the LcHsp70 and LmHsp70 determined using the bovine Hsc70 crystal structure as template revealed a high level similarity in conformation of 3D-structure including the ATPase domain, substrate binding domain and the linker region. The ATPase domains of both coelacanth species show a high level conservation of regions and residues involved in ATP/ADP binding and hydrolysis equivalent to bovine Hsc70 (Figure 2.5 and 3.8). The coelacanth Hsp70s also displayed residues described as important for binding Hsp40, an Hsp70 cochaperone (Figure 2.5 and 3.8). Also conserved were residues equivalent to V388 and L393 bovine Hsc70 in the linker region that have been demonstrated to play a role in interdomain communication (Figure 2.7 and Figure 3.9). It was also interesting that the residues equivalent to residues involved in hydrophobic bonds and salt bridges in bovine Hsc70 domain interface were also conserved (Figure 2.7 and Figure 3.9). Bovine Hsc70 residues shown to play a role in substrate binding were also conserved in both LcHsp70 and LmHsc70 (Figure 2.6 and 3.8).

Despite this high level identity to the mentioned Hsp70s, LmHsp70 could not functionally replace DnaK since it failed to reverse thermosensitivity of *E. coli* BB2362 strain. Assuming that LmHsp70 does have an ATPase activity in *E. coli*, the ability of the substrate binding domain to recognise *E. coli* substrates becomes an important determinant of LmHsp70 chaperone function. The Hsp70 interaction with substrates has been shown to depend on specific determinants in the β-sandwich of the substrate binding domain (Rüdiger *et al*., 1997; Rüdiger *et al*., 2000, Suppini *et al*., 2004) The ability of Hsp70 to productively interact with their regulatory co-chaperones, Hsp40s could also be important since Hsp40 is thought to deliver substrate proteins to Hsp70 (Hennessy *et al*., 2005). The possible interpretation of our data could therefore be that the LmHsp70 does not have the capacity to interact physically with *E. coli* substrates or *E. coli* Hsp40 protein, DnaJ. It is also possible that LmHsp70 requires specific nucleotide exchange factors to enable it to carry out its chaperone function.

5.1.3 Heterologous production of LcHsp70 and LmHsp70 proteins

The finding that the overproduction of both full length LcHsp70 and LmHsp70 was vastly improved in *E. coli* XLI-blue having the RIG plasmid, was consistent with the predictions that were made in the codon usage analysis (Figure 4.1). Although the RIG tRNAs improved the yield of full-length LcHsp70 and LmHsp70, there was a corresponding increase in the production of lower molecular weight species. *E. coli* is more biased towards a G-C rich genome having differing codon usage, therefore resulting in degradation or aborted translation products (Baca *et al*., 2000). This was particularly interesting since Matambo and co-workers (Matambo *et al*., 2004) had found a decrease in truncation derivatives and an increase in the production of full-length protein. Alternative overproduction methods are discussed in section 5.3.2.

5.2 Conclusion

The unique habitat of the coelacanth does not seem to have caused significant changes in the *hsp70* coding region and gene product. Here, we have presented data that strongly supports that the genome of both extant species of the coelacanth does encode Hsp70s, the structure which are archetypal of known Hsp70 proteins. We have also shown that the *Lmhsp70* upstream regulatory region contained all putative elements (HSEs, CAAT-box, TATA-box) that are an essential for heat inducible promoter. The varied predicted structure of this regulatory region may have an impact the mechanism of transcriptional regulation in a manner perhaps indicative of the coelacanth habitat. We have therefore provided first evidence of the Hsp70-based cytoprotection mechanism in the extant coelacanth. The data presented in this thesis, thus provides a foundation to formulate testable prediction on the regulation and function of *hsp70* genes and gene products in the extant coelacanth.

5.3 Future work

5.3.1 *Lmhsp70* promoter studies

Molina *et al.* (2000) studied the expression of the *lacZ* gene driven by the tilapia *hsp70* promoter/*lacZ* reporter construct in carp cells. The authors reported that heat shock conditions resulted in 16-fold increase in the β-gal activity in cell transfected with the reporter construct. The control cells (non-transfected cells) were not affected. This work thus showed that fish *hsp70* promoters could be used to regulate the expression of a reporter gene in mammalian cell lines in a heat inducible manner. Using the same cell line, Molina *et al.* (2001, 2002), showed that the same tilapia *Hsp70* promoter was able to drive the expression of *green fluorescence protein (GFP)-luciferase* reporter genes. The same approach could be used to determine the heat inducibility of the *Lmhsp70* promoter in mammalian cell lines. Deletion mutants containing the *Lmhsp70* upstream regulatory region up to the -289 bp HSE, -369 bp HSE or -894 HSE, respectively, could be employed to determine the minimal regulatory region required for the heat inducibility of

the *Lmhsp70* promoter. Point mutations in each functional HSE could be used to determine the minimal *Lmhsp70* HSE sequence required to drive the expression of a reporter gene in a temperature dependent manner. The point mutations could also be used to asses the significance of the CAAT-box and TATA-box on the basal and inducible activity of the *Lmhsp70* promoter in reporter constructs.

5.3.2 Heterologous production of LcHsp70 and LmHsp70 proteins

In accordance with the codon usage analysis presented in Figure 4.1, the coelacanth codon usage is generally similar to the eukaryotic codon usage. In view of this consideration, perhaps a eukaryotic system, such as yeast or mammalian systems may be better suited to investigate the functioning and regulation of the coelacanth Hsp70. The COS1 cells have been succefully used to overproduce *L. chalumnae* and *L. menadoensis* RH1 and RH2 pigments using the pMT expression vector (Yokoyama and Tada, 2000). The use of such systems could reduce the degradation or aborted translation products due to differences in codon preference. Vectors containing the *Drosophila* MT promoter have been described for high-level, inducible expression of the heterologous proteins in *Drosophila* S2 (or D.Mel-2) cells. The promoters in this vector are inducible by addition of copper sulfate or cadmium chloride to the culture medium (Invitrogene, USA).

5.3.3 *LmHsp70* functional analysis

The chaperone activity of Hsp70s may require interactions with co-chaperones, such as Hsp40 and nucleotide exchange factors (Lieberek e*t al.*, 1991; Shomura *et al*., 2005). In this regard the question is whether the coelacanth Hsp70 can productively interact with prokaryotic co-chaperones to form functional chaperone-cochaperone partnerships. The inability of the LmHsp70 to reverse thermosensitivity of a temperature sensitive *E. coli* BB2362 strain may necessitate further *in vitro* and *in vivo* analyses. To facilitate such analyses DnaK-LcHsp70/LmHsp70 recombinant constructs encoding the following chimeric proteins could be constructed:

(i) Hsp70 chimera containing DnaK ATPase domain and LcHsp70/Lmhsp70 substrate binding domain

(ii) Hsp70 chimeras containing LcHsp70/Lmhsp70 ATPase domain and DnaK substrate binding domain

(iii) Wild type DnaK

(iv) Wild type LcHsp70/LmHsp70.

5.3.3.1 LmHsp70 *in vitro* functional analysis

The Hsp70 chaperone activity is directly associated with its nucleotide-bound state, its basal ATPase activity, and the stimulation of its ATPase activity by co-chaperones (Freeman *et al*., 1995). It also depends on the communication of the ATPase domain conformational changes induced by ATP/ADP cycling to the substrate binding domain (Jiang *et al*., 2005). The basal ATPase activity of LcHsp70 and LmHsp70 and how it is impacted by eukaryotic Hsp70 co-chaperones needs to be investigated. Future studies could include investigation on the stimulation of LcHsp70 and LmHsp70 ATPase activities by Hsp40s. Three fish genes belonging to the Hsp40 family have recently been isolated from *Paralichthys olivaceus* (Dong *et al*., 2006). The effect of such Hsp40s together with eukaryotic nucleotide exchange factors such as Bag-1 could also be investigated to establish their effect on the coelacanths Hsp70 ATP/ADP cycle. The ATP hydrolysis assay described by Freeman and co-workers measures the rate of ATP hydrolysis by measuring the release of ^{32}PPi from α-^{32}P-ATP (Freeman *et al*., 1995).

The chaperone activity of LcHsp70 or LmHsp70 could also be assayed by their ability to aid the refolding of denatured protein substrates. Freeman and co-workers have described a refolding assay wherein a β-galactosidase is unfolded in a buffer containing high concentrations of β-mercaptoethanol and 6M guanidine-HCl. The refolding of the protein was assayed in the presence of Hsp70 and ATP. Different Hsp70 co-chaperones were also added to assay their effect on protein folding (Freeman *et al*., 1996).

5.3.3.2 LmHsp70 *in vivo* functional analysis

The ability of the wild type and chimeric proteins (described in section 5.3.3) to reverse thermosensitivity of temperature sensitive *E. coli* strains in complementation assays as described in section 4.2.4 could be used to determine the LcHsp70 and LmHsp70 chaperone activity *in vivo*. Complementation assays in yeast strains such as *Saccharomyces cerevisiae* could be more challenging since these strains contain multiple Hsp70s (Table 1.1). *S. cerevisiae* genome encodes at least 14 different Hsp70 proteins grouped into five subclasses (Nelson *et al*., 2004). Of these classes the cytosolic subfamilies, *SSA* and *SSB*, share 60% amino acid identity but cannot functionally substitute for one another (Nelson *et al*., 2004). Ssa1 and -2 are constitutively expressed, while Ssa3 and -4 are heat-inducible (Nelson *et al*., 2004). *S. cerevisiae* strain wherein *hsp70* genes with a close resemblance to the *Lchsp70* and *Lmhsp70* genes (preferably inducible *hsp70*s) are knockout could be a valuable tools to carry out complementation assays.

References

Adám, A., Bártfai, R., Lele, Z., Krone, P. H. and Orbán, L. (2000). Heat-inducible expression of a reporter gene detected by transient assay in zebrafish. *Exp. Cell. Res*. **256**, 282-290.

Alberti, S., Böhse, K., Arndt, V., Schmitz, A. and Höhfeld, J. (2004). The cochaperone HspBP1 inhibits the CHIP ubiquitin ligase and stimulates the maturation of the cystic fibrosis transmembrane conductance regulator. *Mol. Biol. Cell* **15**, 4003-4010.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol*. **215**, 403-410.

Amemiya, C. T., Ohta, Y., Litman, R. T., Rast, J. P., Haire, R. N. and Litman, G. W. (1993). V_H gene organization in a relict species, the coelacanth *Latimeria chalumnae*: evolutionary implications. *Proc. Natl. Acad. Sci. USA.* **90,** 6661-6665.

Anand, S., Wang, W. C. H., Powell, D. R., Bolanowski, S. A., Zhang, J., Ledje, C., Pawashe, A. B., Amemiya, C. T. and Shashikant, C. S. (2003). Divergence of *Hoxc8* early enhancer parallels diverged axial morphologies between mammals and fishes. *Proc. Natl. Acad. Sci. USA.* **100**, 15666-15669.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.

Aparicio, S. (2000). Vertebrate evolution recent perspectives from fish. *Trends Genet.* **16**, 54-56.

Arai, A., Naruse, K., Mitani, H. and Shima, A. (1995). Cloning and Characterization of cDNA for 70- kDA heat-shock protein (hsp70) from two fish species of genus *Oryzias*. *Jpn. J. Genet.* **70**, 423-433.

Baca, A. M. and Hol, W. G. J. (2000). Overcoming codon bias: A method for high-level overexpression of *Plasmodium* and other AT-rich parasite genes in *Escherichia coli*. *Int. J. Parasitol*. **30**, 113-118.

Baneyx, F. (1999). Recombinat protein expression in *Escherichia. coli* . *Curr. Opin. Biotechnol*. **10**, 411-421.

Basso, A. D., Solit, D. B., Chiosis, G., Giri, B., Tsichlis, P. and Rosen, N. (2002). Akt forms an intracellular complex with heat shock protein 90 and Cdc37 and is destabilized by inhibitors of Hsp90 function. *J. Biol. Chem*. **277**, 39858-39866.

Basu, N., Todgham, A. E., Ackerman, P. A., Bibeau, M. R., Nakano, K., Schulte, P. M., and Iwama, G. K. (2002). Heat shock protein genes and their functional significance in fish. *Gene* **295**, 173–183.

Betz, U. A. K., Mayer, W. E. and Klein, J. (1994). Major histocompatibility complex class I genes of the coelacanth *Latimeria chalumnae*. *Proc. Natl. Acad. Sci. USA.* **91**, 11065-11069.

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

Bienz, M. and Pelham, H. R. (1986). Heat shock regulatory elements function as an inducible enhancer in the *Xenopus hsp70* gene and when linked to a heterologous promoter. *Cell*, **45**, 753-760.

Bienz, M. and Palhan, H. R. B. (1987). Mechanism of heat-shock gene activation in higher eukaryotes. *Adv. Genet*. **24**, 31-66.

Blechinger, S. R, Evans, T. G., Tang, P. T., Kuwada, J. Y., Warren, J. T. Jr., and Krone, P. H. (2002). The heat-inducible zebrafish *hsp70* gene is expressed during normal lens development under non-stress conditions. *Mech. Dev.* **112**, 213-215.

Bogard, J. P., Balon, E. K. and Bruton, M. N. (1994). The chromosomes of the living coelacanth and their remarkable similarity to those of one of the most ancient frogs. *J. Hered.* **85**, 322-325.

Bond, U. (1988). Heat shock but not other stress inducers leads to the disruption of a sub-set of snRNPs and inhibition of *in vitro* splicing in HeLa cells. *EMBO J.* **7**, 3509–3518.

Bork, P., Sander, C. and Valencia, A. (1992). An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and Hsp70 heat shock proteins. *Proc. Natl. Acad. Sci. USA* **89**, 7290- 7294.

Brehmer, D., Rüdiger, S., Gässler, C. S., Klostermeier, D., Packschies, L., Reinstein, J., Mayer, M., P. and Bukau, B. (2001). Turning of chaperone activity of Hsp70 proteins by modulation of nucleotide exchange. *Nat. Struct. Biol*. **8**, 427-432.

Brehmer, D., Gässler, C. S., Rist, W., Mayer, M. P. and Bukau, B. (2004). Influence of GrpE on DnaK-substrate interactions. *J. Biol. Chem.* **279,** 27957-27964.

Brennecke. K., Gellner, K. and Bosch, T. C. (1998). The lack of a stress response in *Hydra oligactis* is due to reduced *hsp70* mRNA stability. *Eur. J. Biochem*. **255**, 703 – 709.

Brinkmann, H., Venkatesh, B., Brenner, S. and Meyer, A. (2004). Nuclear protein-coding genes support lungfish and not the coelacanth as the closest living relatives of land vertebrates. *Proc. Natl. Acad. Sci. USA.* **101**, 4900-4905.

Brodsky, J. L., Goeckeler, J. and Schekman R. (1995). BiP and Sec63p are required for both co- and posttranslational protein translocation into the yeast endoplasmic reticulum. *Proc. Natl. Acad. Sci. USA*. **92**, 9643-9646.

Buchberger, A., Gässler, C. S., Büttner, M., McMacken, R. and Bukau, B. (1999). Functional Defects of the DnaK756 Mutant chaperone of *Escherichia coli* indicate distinct roles for amino- and carboxyl-terminal residues in substrate and co-chaperone interaction and interdomain communication**.** *J. Biol. Chem*. **274,** 38017 **–** 38026**.**

Buchberger, A., Theyssen, H., Schröder, H., McCarty, J. S., Virgallita, G., Milkereit, P., Reinstein, J. and Bukau, B. (1995). Nucleotide-induced conformational changes in the ATPase and substrate binding domains of the DnaK chaperone provide evidence for interdomain communication. *J. Biol Chem*. **270**, 16903-16910.

Buchner, J. (1999). Hsp90 & co. – a holding for folding. *Trents Biochem. Sci*. **24**, 136-141

Buczynski, G., Slepenkov, S. V., Sehorn, M. G. and Witt, S. N. (2001), Characterization of a lidless form of the molecular chaperone DnaK. *J. Biol. Chem*. **276**, 27231-27236.

Burkholder, W. F., Zhao, X., Zhu, X., Hendrickson, W. A., Gragerov, A. and Gottesman, M. E. (1996). Mutations in the C-terminal fragment of DnaK affecting peptide binding *Proc. Natl. Acad. Sci. USA.* **93,** 10632–10637.

Caplan, A. and Douglas, M. (1991).Characterization of YDJ1: A yeast homologue of the bacterial dnaJ protein. *J. Cell Biol*. **114**, 609-662.

Caplan, A. J., Jackson, S. and Smith, D. (2003). Hsp90 reaches new heights. *EMBO J.* **4**, 126-130.

Cěrnila, B., Črešnar, B. and Breskvar, K. (2003). Molecular characterization of genes encoding cytosolic Hsp70s in the zygomycete fungus *Rhizopus nigricans. Cell Stress Chap*. **8**, 317-328.

Chae, H-D., Yun, J and Shin, Y. D. (2005). Transcription repression of a CCAAT-binding transcription factor CBF/HSP70 by p53. *Exp. Mol. Med.* **37**, 488-491.

Chellaiah, A., Davis, A. and Mohanakumar, T. (1993). Cloning of a unique human homologue of the *Escherichia coli* DnaJ heat shock protein. *Biochim. Biophys. Acta.* **1174**, 111–113.

Chen, S. and Smith, D. F. (1998). Hop is an adapter in the heat shock protein 70 (Hsp70) and Hsp90 chaperone machinery. *J. Biol. Chem*. **273**, 35194-35200.

Cho, E. J., Kobor, M. S., Kim, M., Greenblatt, J. and Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser2 of the RNA polymerase II C-terminal domain. *Genes Dev*. **15**, 3319–3329.

Chou, C., Forouhar, F., Yeh, Y., Shr, H., Wang, C. and Hsiao, C. (2003). Crystal structure of the C-terminal 10-kDa subdomain of Hsc70. *J. Biol. Chem* **278**, 30311-30316.

Chui, C., Nonaka, D., Xue, L., Amemiya, C. T. and Wagner, G. P. (2000). Evolution of *hoxa-11* in lineages physiologically positioned along the fin-limb transition. *Mol. Phyl. Evol.* **17**, 305–306.

Chung, K. T., Shen, Y., and Hendershot, L. M. (2002). BAP, a mammalian BiP-associated protein, is a nucleotide exchange factor that regulates the ATPase activity of BiP. *J. Biol. Chem.* **277**, 47557-47563.

Danke, J., Miyake, T., Powers, T., Schein, J., Shin, H., Bosdt, I., Erdmann, M, Caldwell, R. and Amemiya, C. T. (2004). Genome resource for the Indonesian coelacanth, *Latimeria menadoensis*. *J. Exp. Zool*. **301A**, 228-234.

DeLano, W. L. and Bromberg, S. (2003). PyMOL Reference Manual. DeLano Scientific LLC, San Carlos, CA.

Deshaies, R. and Schekman, R. (1987). A yeast mutant defective at an early stage in import of secretory protein precursors into the endoplasmic reticulum. *J. Cell Biol.* **105**, 633-645.

Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., and Bukau, B. (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* **400**, 693-696.

Dittman, K. D., Hutchison, K. A., Owens-Grllo, J. K. and Pratt, W. B. (1996) Reconstitution of the steroid receptor-Hsp90 heterocomplex assembly system of the rabbit reticulocyte lysate. *J. Biol. Chem*. 271, 12833-12839.

Dunn, A. Y., Melville, M. W. and. Frydman, J. (2001). Review: cellular substrates of the eukaryotic chaperonin TRiC/CCT. *J. Struct. Biol*. **135,**176-184.

Dong, C. W., Zhang, Y. B., Zhang, Q. Y. and Gui, J. F.(2006) Differential expression of three Paralichthys olivaceus Hsp40 genes in responses to virus infection and heat shock. *Fish Shellfish Immunol*.**21**, 146-58

Ellis, J. (1987). Proteins as molecular chaperones. *Nature* **328**, 378-379.

Ellis, R. J and Pinheiro, T. J. P. (2002). Danger-misfolding proteins. *Nature* **416**, 483-484.

Ellis, R. J. (2001). Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.* **26**, 597-604.

Erdmann, M. V., Caldwell, R. L. and Moosa, M. K. (1998). The Indonesian "king of the sea" discovered. *Nature* **395**, 335.

Fan, X., Shi., H. and Lis, J. T. (2005). Distinct transcriptional responses of RNA polymerase I, II and III to aptamers that bind TBP. *Nucleic Acid Res*. **33**, 838-845.

Fenton, W. A. and Horwich A. L. (1997). GroEL-mediated protein folding. *Protein Sci.* **6**, 743–760.

Fenton, W. A., Weissman, J. A., and Horwich, A. L. (1996). Putting a lid on protein folding: structure and function of the co-chaperonin, GroES. *Chem. Biol.* **3**, 157–161.

Fink, A. L. (1999) Chaperone-mediated folding. *Physiol. Rev.* **79**, 425-449.

Fiszer-Kierzkowska, A., Wysocka, A., Jarzab, M., Lisowska, K. and Krawczyk, Z. (2003). Structure of gene flanking regions and functional analysis of sequences upstream of the rat hsp70.1 stress gene. *Biochim. Biophys. Acta* 1625, 77– 87.

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

Flaherty, K. M., Wilbanks S. M., DeLuca-Flaherty, C. and McKay, D. B. (1994). Structural basis of the 70-kilodalton heat shock cognate protein ATP hydrolytic activity. Structure of the active site with ADP or ATP bound to wild type and mutant ATPase fragment. *J. Biol. Chem*. **269**, 12899-12907.

Flom, G., Weekes, J., Williams, J. J. and Johnson, J. L. (2006). Effect of the tetratricopeptide repeat and aspatate-proline 2 domains of Sti1 on Hsp90 signalling and interaction in *Saccharomyces cerevisiae*. *Genetics* **172**, 41-51.

Force, A., Amores, A. and Postlethwait, J. H. (2002). *Hox* cluster organization in the jawless vertebrate *Petromyzon marinus*. *J. Exp. Zool. (Mol. Dev. Evol.)* **294**, 30-46.

Forey, P.L. (1988). Golden jubilee for the coelacanth *Latimeria chalumnae*. *Nature* **336**, 727-732.

Freeman, B. C., Myers, M. P., Schumacher, R. and Morimoto, R. I. (1995). Identification of a regulatory motif in Hsp70 that affects AYPase activity, substrate binding and interaction with HDJ-1. *EMBO J*. 14, 2281-2292.

Freeman, B. C. and Morimoto, R. I. (1996). The human cytosolic molecular chaperones Hsp90, Hsp70 (Hsc70) and Hdj-1 have distinct roles in the recognition of non-native protein and protein folding. *EMBO J*. **15**, 2969-2979.

Freitag, J., Ludwig, G., Andreini, I., Rössler, P. and Breer, H. (1998). Olfactory receptors in aquatic and terrestrial vertebrates. *J. Comp. Physiol.* **183**, 635-650.

Fricke, H. (1992). Coelacanth tissue bank. *Nature* **357**, 105.

Fricke, H. and Hissmann, K. (1990). Natural habitat of coelacanths. *Nature* **346**, 323-324.

Fricke, H. and Hissmann, K. (1994). Home range and migrations of the living coelacanth *Latimeria chalumnae*. *Mar. Biol*. **120,** 171-180.

Fricke, H., Hissmann, K., Schauer, J., Erdmann, M., Moosa, M. K. and Plante, R. (2000). Biogeography of the Indonesian coelacanths. *Nature* **403**, 38.

Frydman, J. (2001). Folding of newly translated proteins in vivo: The role of molecular Chaperones. *Ann. Rev. Biochem.* **70**, 603-647.

Frydman, J., Nimmesgern, E., Ohtsuka, K. and Hartl, F. U. (1994). Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. *Nature* **370**, 111–117.

Garcia-Fernàndez, J. and Holland, P. W. H. (1994). Archetypal organization of the amphioxus *Hox* gene cluster. *Nature* **370**, 565-566.

Gässler C. S., Buchberger, A., Laufen, T., Mayer, M. P., Schröder, H., Valencia, A. and Bukau, B. (1998). Mutations in the DnaK chaperone affecting interaction with the DnaJ cochaperone. *Proc. Natl. Acad. Sci. USA.* **95**, 15229-15234.

Gässler C. S., Wiederkehr, T., Brehmer, D. Bakau, B. and Meyer M. P. (2001) Bag-1M accelerates nucleotide release for human Hsc70 and Hsp70 and can act concentration-dependent as a positive and negative cofactor. *J. Biol. Chem*. 276, 32538-32544.

Gautschi, M., Lilie, H., Funfschilling, U., Mun, A., Ross, S., Lithgow, T., Rucknagel, P. and Rospert, S. (2001). RAC, a stable ribosome-associated complex in yeast formed by the DnaK-DnaJ homologs Ssz1p and zuotin. *Proc. Natl. Acad. Sci. USA*. **98,** 3762–3767.

Gellner, K., Praetzel, G. and Bosch, T. C. (1992). Cloning and expression of a heat-inducible *hsp70* gene in two species of *Hydra* which differ in their stress response. *Eur. J. Biochem*. **210**, 683 – 691.

Giardina, C., Perez-Riba, M. and Lis, J. T. (1992). Promoter melting and TFIID complexes on *Drosophila* genes *in vivo*. *Genes Develop*. 6, 2190-2200.

Goloubinoff, P., Mogk, A., Zvi, A., Tomoyasu, T. and Bakau, B. (1999) Peres Ben sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. *Proc. Natl. Acad. Sci. USA*. **96**, 13732-13737.

Gorr, T., Kleinschmidt, T. and Fricke, H. (1991). Close tetrapod relationships of the coelacanth *Latimeria* indicated by haemoglobin sequences. *Nature* **351**, 394-397.

Graser, R., Malner-Dragojevic, D. and Vincek, V. (1996). Cloning and characterization of a 70 kd heat shock protein cognate (*hsc70*) gene from the zebrafish (*Danio rerio*). *Genetica* **98**, 273-276.

Groemping, R., Seidel, Y. and Reinstein, J. (2005). Balance of ATPase stimulation and nucleotide exchange is not required for efficient refolding activity of the DnaK chaperone. *FEBS Lett*. **579**, 5713-5717.

Halloran, M. C., Sato-Maeda, M., Warren, J. T. Jr., Su, F., Lele, Z., Krone, P. H., Kuwada, J. Y. and Shoji, W. (2000). Laser-induced gene expression in specific cells of transgenic zebrafish. *Development* **127**, 1953-1960.

Hamman, B. D., Hendershot, L. M. and Johnson, A. E. (1998). BiP maintains the permeability barrier of the ER membrane by sealing the lumenal end of the translocon pore before and early in translocation. *Cell* **92**, 747–758.

Harrison, C. J., Hayer-Hartl, M., Liberto, M., Hartl, F. U. and Kuriyan, J. (1997). Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. *Science* **276**, 413-435.

Hartl F. U. (1996). Molecular chaperone in cellular protein folding. *Nature* **381**, 571-579.

Hedges, S. B., Hass, C. A. and Maxson L. R. (1993). Relations of fish and tetrapods. *Nature* **363**, 501-502.

Heemstra, P. (2001). New coelacanth discovery: Sodwana's living fossils. *Archimedes* **43**, 13-17.

Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., *et al*. (1998). Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acids Res* **26**, 362-367.

Hennessy, F., Nicoll, W., Zimmermann, R., Cheetham, M. E., and Blatch, G. L. (2005). Not all J domains are created equal: implication for the specificity of Hsp40-Hsp70 interactions. *Protein Sci*. **15**, 1697-1709.

Hernández, M. P. Sullivan, W. P. and Toft, D. O. (2002). The assembly and intermolecular properties of the Hsp70-Hop-Hsp90 molecular chaperone complex. *J. Biol. Chem*. **277**, 38294-38304.

Hillier, C. J., Ware, L. A., Barbosa, A., Angov, E., Lyon, J. A., Heppner, D. G. and Lanar, D. E. (2005). Process development and analysis of liver-stage antigen-1, a pre-erythrocyte-stage protein-based vaccine for *Plasmodium falciparum*. *Imfect. Immun*. **73**, 2109-2115.

Hillis, D. M, Dixom, M. T. and Ammerman L. K. (1991). The relationship of the coelacanth *Latimeria chalumnae*: evidence from sequences of vertebrate 28S ribosomal RNA genes. *Environ. Biol. Fishes* **32**, 119-130.

Hofmann, G. E., Buckley, B. A., Airaksinen, S., Keen, J. E. and Somero, G. N. (2000). Heat-shock protein expression is absent in the Antarctic fish *Trematomus brenacchii* (family Nototheniidae). *J. Exp. Biol*. **203**, 2331-2339.

Höhfeld, J. and Jentsch, S (1997). GrpE-like regulation of the Hsc70 chaperone by the anti-apoptotic protein BAG-1. *EMBO J*. **16**, 6209-6216.

Höhfeld, J., Minami, Y. and Hartl, F.U. (1995). Hip, a novel cochaperone involved in the eukaryotic Hsc70/Hsp40 reaction cycle. *Cell* **83**, 589-598.

Holder, M. T., Erdmann, M. V., Wilcox, T. P., Caldwell, R. L. and Hillis, D. M. (1999). Two living species of coelacanths? *Proc. Natl. Acad. Sci. USA.* **96**, 12616-12620.

Holland, P. W. H. and Garcia-Fernàndez, J. (1996). *Hox* genes and chordate evolution. *Dev. Biol*. **173**, 382-395.

Inoue, J. G., Miya, M., Venkatesh, B. and Nishida, M. (2005). The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and the divergence time estimation between the two coelacanths. *Gene* **349**, 227-235.

Irvine, S. Q., Carr, J. L., Bailey, W. J., Kawasaki, K., Shimizu, N., Amemiya, C. T. and Ruddle, F. H. (2002). Genomic analysis of Hox clusters in the sea lamprey, *Petromyzon marinus*. *J. Exp. Zool. (Mol. Dev. Evol.)* **294**, 47-62.

Jiang, J., Ballinger, C, A., Wu, X., Dai, Q., Cyr, D. M., Höhfeld, J., and Patterson, C. (2001). CHIP is a U-box-dependent E3 ubiquitin ligase. *J. Biol. Chem.* **276**, 42938-42944.

Jiang, J., Prasad, K., Lafer, E. M. and Sousa, R. (2005). Structural basis of interdomain communication in the Hsc70 chaperone. *Mol. Cell*. **20**, 513-524.

Kabani, M, McLellan C., Raynes, D. A., Guerriero, V., and Brodsky, J. L. (2002). HspBP1, a homologue of the yeast Fes1 and Sls1 proteins, is an Hsc70 nucleotide exchange factor. *FEBS Lett*. **531**, 339-342.

Kabani, M., Beckerich J. M. and Gaillardin, C. (2000). Sls1p stimulates Sec63p-mediated activation of Kar2p in a conformation-dependent manner in the yeast endoplasmic reticulum. *Mol. Cell. Biol*. *20*, 6923-6934.

Kabani, M., Beckerich, J. M. and Brodsky, J. L. (2002). Nucleotide exchange factor for the yeast Hsp70 molecular chaperone Ssa1p. *Mol. Cell. Biol*. **22**, 4677-4689.

Kabani, M., Beckerich, J. M. and Brodsky, J. L. (2003). The yeast Sls1p and Fes1p proteins define a new family of Hsp70 nucleotide exchange factors. *Curr. Genomics* **4**, 465-473.

Koh, E. G. L., Lam, K., Christoffels, A., Erdmann., M. V., Brenner, S. and Venkatesh, B. (2003). Hox gene clusters in the Indonesian coelacanth, *Latimeria menadoensis*. *Proc. Natl. Acad. Sci. USA.* **100**, 1084-1088.

Komarnitsky, P., Cho, E. and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* **14**, 2452-2460.

Knittler**,** M. R., Dirks, S. and Haas, I. G. (1995). Molecular chaperones involved in protein degradation in the endoplasmic reticulum: quantitative interaction of the heat shock cognate protein BiP with partially folded immunoglobulin light chains that are degraded in the endoplasmic reticulum *Proc. Natl. Acad. Sci. USA.* **92**, 1764-1768.

Korathy, R. K., Jones, D. and Candido, P. M. (1984). 70-Kilodalton heat shock polypeptide from the rainbow trout: characterization of the cDNA sequences. *Mol. Cell. Biol*. **4**, 1785-1791.

Laskey, R. A., Honda, B. M., Mills, A. D. and Finch, J. T. (1978). Nucleosomes are assembled by an acid protein, which bind histones and transfers them to DNA. *Nature* **275**, 416-420.

Lebedeva, L. A, Nabirochkina, E. N., Kurshakova, M. M., Rorbet, F., Krasnov, A., Evgen´ev, M. B., Kadonaga, J. T., Georgieva, S. G. and Tora, L. (2005). Occupancy of the *Drosophila hsp70* promoter by a subset of basal transcriptional factors diminishes upon transcriptional activation. *Proc. Natl. Acad. Sci. USA.* **102**, 18087-18092.

Lee, J. (2004). Genomic cloning of a Heat shock cognate 71-1 gene (*hsc71-1*) from the Hermaphroditic fish *Rivulus marmoratus* (Cyprinodontiformes, Rivulidae). *DNA Sequence* **15**, 33-38.

Lele, Z., Engel, S. and Krone, P. (1997). *hsp47* and *hsp70* gene expression is differentially regulated in a stress- and tissue- specific manner in zebrafish embyos. *Dev. Genet.* **21**, 123-133.

Liberek, K., Skowyra, D., Zylicz, M., Johnson, C. and Georgopoulos., C (1991). The *Escherichia coli* DnaK chaperone, the 70-kDa heat shock protein eukaryotic equivalent, changes conformation upon ATP hydrolysis, thus triggering its dissociation from a bound target protein. *J. Biol.Chem.* **266**, 14491-14496.

Lim, E. H. and Brenner, S. (1999). Short-range linkage relationships, genomic organization and sequence comparisons of a cluster of five *HSP70* genes in *Fugu rubripes. Cell. Mol. Life Sci.* **55**, 668-678.

Lis, J. T., Mason, P., Peng, J, Prince, D. H. and Werner, J. (2000). P-TEFb kinase recruitment and function at heat shock loci. *Genes Dev.* **14**, 792-803.

Lüders, J., Demand, J. and Höhfeld, J. (2000). The ubiquitin-related BAG-1 provided a link between the molecular chaperone Hsc70/Hsp70 and the proteasome. *J. Biol. Chem.* **275**, 4613-4617.

Mason P. B. Jr. and Lis J. T. (1997). Cooperative and competitive protein interaction at the Hsp70 promoter. *J. Biol. Chem*. **272**, 33227-33233.

Matambo, T, S., Odunuga, O. O., Boshoff, A. and Blatch, G. L. (2004). Overproduction, purification and characterization of the *Plasmodium falciparum* heat shock protein 70. *Prot. Expr. Purif*. **33**, 214-222.

Mayer , M. P. and Bukau, B. (2005) Hsp70 chaperone: cellular function and molecular mechanism. *Cell. Mol. Life. Sci*. **62**, 670-684.

Mayer, M. P., Schröder, H., Rüdiger, S., Paal, K., Laufen, T. and Bukau, B. (2000). Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nat. Struct. Biol*. **7**, 586-593

Meyer, A. (1995). Molecular evidence on the origin of tetrapods and the relationships of the coelacanth. *Tree* **10**, 111-116.

Meyer, A. and Wilson, A. C. (1990). Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. *J. Mol. Evol*. **31**, 359-364.
Minguillón, C., Gardenyes, J., Serra, E., Castro, L.F.C., Hillforce, A., Holland, P., Amemiya, C.T. and Garcia-Fernàndez, J. (2005). No more than 14: the end of the amphioxus Hox cluster. *Int. J. Biol. Sci*. **1**, 19-23.

Minton, A. P. (2000). Implication of macromolecular crowding for protein assembly. *Curr Opin Struct Biol*. **10**, 34-39.

Molina, A., Biemar, F., Müller, F., Iyengar, A., Prunet, P., Maclean, N., Martial, J.A. and Muller, M. (2000). Cloning and expression analysis of an inducible *HSP70* gene from tilapia fish. *FEBS Lett.* **474**, 5-10.

Molina, A., Di Martino, E., Martial, J. A. and Muller, M. (2001). Heat shock stimulation of a tilapia heat shock protein 70 promoter is mediated by a distal element. *Biochem. J.* **356**, 353-359.

Molina, A., Carpeaux, R., Martial, J. A. and Muller, M. (2002). A transformed fish cell line expressing a green fluorescent protein-luciferase fusion gene responding to cellular stress. *Toxicol. in vitro* 16, 201-207.

Morimoto, R. I. (1998). Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. *Genes Dev*. **27**, 3788-3796.

Morimoto, R.I., Sarge, K.D. and Abravaya, K. (1992). Transcriptional regulation of heat shock genes. *J. Biol. Chem*. **267**, 21987-21990.

Moro, F., Fernández, V. and Muga, A. (2003). Interdomain interaction through helices A and B of DnaK peptide binding domain. *FEBS Lett*. **533**, 119-123.

Morshauser, R. C., Hu, W., Wang, H., Pang, Y., Flynn, G. C. and Zuiderweg E. R. P. (1999). High resolution solution structure of the 18 kDa substrate-binding domain of the mammalian chaperone protein Hsc70. *J. Mol. Biol*. **289**, 1387-1403.

Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform.* **5**, 150–163.

Nicoll, W. S., Boshoff, A., Ludewig, M. H., Hennessy, F., Jung, M. and Blatch, G. L. (2006). Approaches to the isolation and characterization of molecular chaperones. *Prot. Expr. Purif*. **46**, 1-15.

Nollen, E. A. A., Kabakov, A. E., Brunsting, J. F., Kanon, B., Höhfeld, J. and Kampinga, H. H. (2001). Modulation of *in vivo* Hsp70 chaperone activity by Hip and Bag-1. *J. Biol. Chem*. **276**, 4677-4682.

Noonan, J. P, Grimwood, J., Schmutz, J., Dickson, M. and Myers, R. M. (2004). Gene conversion and the evolution of protocadherin gene cluster diversity. *Genome Res*. **14**, 354–366.

Noonan, J. P., Grimwood, J., Danke, J., Schmutz, J., Dickson, M., Amemiya C. T. and Myers, R. M. (2004). Coelacanth genome sequence reveals the evolutionary history of vertebrate genes. *Genome Res.* **14**, 2397-2405.

O'Brien, M. C. Flaherty, K. M. and McKay, D. B. (1996). Lysine 71 of the chaperone protein Hsc70 is essential for ATP hydrolysis. *J. Biol. Chem*. **271**, 15874-15878.

O'Brien, T., Hardin, S., Greenleaf, A. and Lis, J. T. (1994). Phosphorylation of RNA polymerase II C-terminal domain and transcriptional elongation. *Nature* **370**, 75-77.

Odunuga, O. O., Hornby, J. A., Bies, C., Zimmermann, R., Pugh, D. and Blatch, G. L. (2003). Tetratricopeptide Repeat Motif-mediated Hsc70-mSTI1 Interaction. Molecular characterization of the critical contacts for useful binding and specificity**.** *J Biol. Chem***. 278,** 6896-6904.

Ohno, M., Kitabakate, N., Tani, F. (2004). The role of the C-terminal region of mouse inducible Hsp72 in the recognition of peptide substrate for chaperone activity. *FEBS Lett*. **576**, 381-386.

Ohtsuka, K. (1993). Cloning of a cDNA for heat-shock protein hsp40, a human homologue of bacterial DnaJ. *Biochem. Biophys. Res. Commun.* **197**, 235–240.

Ojima, N. and Yamashita, M. (2004). Cloning and characterization of two distinct isoforms of the rainbow trout heat shock factor 1: evidence for heterotrimer formation. *Eur. J. Biochem*. **271**, 703-712.

Peitsch, M. C. (1996). ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochem. Soc. Trans*. **24**:274-279.

Pelham, H. R. (1982). A regulatory upstream promoter element in the Drosophila *hsp 70* heat-shock gene. *Cell* **2**, 517-28.

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

Park, J. H., Lee, J. J., Yoon, S., Lee, J., Choe, S. Y., Choe, Y., Choe, J., Park, E. and Kim, C. G. (2001). Genomic cloning of the *Hsc71* gene in the hermaphroditic teleost *Rivulus marmoratus* and analysis of its expression in skeletal muscles: identification of a novel muscle-preferred regulatory element. *Nucleic Acids Res*. **29**, 3041-3050.

Pellechia, M., Szyperski, T., Wall, D., Georgopoulos, C. and Wüthrich, K. (1996). NMR structure of the J- domain and the Gly/Phe-rich region of the *Escherichia coli* DnaJ chaperone. *J. Mol. Biol*. *260*, 236-250.

Pfund, C., Lopez-Hoyo, N., Ziegelhoffer, T., Schilke, B. A., Lopez-Buesa, P., Walter, W. A., Wiedmann, M. and Craig, E. A. (1998). The molecular chaperone Ssb from *Saccharomyces cerevisiae* is a component of the ribosome–nascent chain complex. *EMBO J*. **17,** 3981–3989.

Place, S. P., Zippy, M. L. and Hofmann, G. E. (2004). Constitutive roles for inducible genes: evidence for the alteration in expression of the inducible *hsp70* gene in Antarctic notothenoid fishes. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **287**, 429-436.

Pouyaud, L., Wirjoatmodjo, S., Rachmatica, I., Tjakrawidjaja A., Hadiaty, R. and Hadie, W. (1999). A new species of the coelacanth. *C. R. Acad. Sci.* **322**, 261-267.

Powers, T. P. and Amemiya, C.T. (2004). Evidence for a Hox14 paralog group in vertebrates. *Curr. Biol.* **14**, R183-R184.

Qian, Y. Q., Patel, D., Hartl, F. U. and McColl D. J. (1996). Nuclear Magnetic Resonance solution of the human Hsp40 (HDJ-1) J-domain. *J. Mol. Biol*. **260**, 224-235.

Queltsch, C., Sangster, T. A. and Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618-624.

Rasmussen, E. B. and Lis, J. T. (1995). Short transcripts of the ternary complex provide insight into RNA polymerase II elongational pausing. *J. Mol. Biol*. **252**, 522-535.

Raynes, D. A. and Guerriero, V. Jr. (1998). Inhibition of Hsp70 ATPase activity and protein renaturation by a novel Hsp70-binding protein. *J. Biol. Chem*. **273**, 32883-32888.

Reese, M. G. (2001). Application of a time-delay neural network to promoter annotation in *Drosophila melanogaster* genome. *Compt. Chem*. **26**, 51-6.

Richter, K., Muschler, P., Hainzl, O., Reinstein, J. and Buchner, J. (2003). Sti1 is a non-competitive inhibitor of the Hsp90 ATPase. *J. Biol. Chem*. **278**, 10328-10333.

Rougvie, A. E. and Lis, J. T. (1990). Post-initiation transcriptional control in *Drosophila melanogaster*. *Mol Cell Biol*. **10**, 6041-6045.

Rüdiger, S., Meyer, M. P., Schneider-Merger, J. and Bukau, B. (2000). Modulation of specificity of the Hsp70 DnaK chaperone by alteration the hydrophobic arch. *J. Mol. Biol*. **304**, 245-251.

Rüdiger, S., Germeroth, L., Schneider-Mergener, J. and Bukau, B. (1997). Substrate specificity of the DnaK molecular chaperone determined by screening cellulose-bound peptide libraries. *EMBO J*. **16**, 1501-1507.

Rutherford, S. I. and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336-342.

Sambrook, J. and Russell, D. W. (2001). Molecular Cloning: A Laboratory Manual. Cold Spring Harbour Laboratory Press.

Santacruz, H., Vriz, S. and Angelier, N. (1997). Molecular characterization of a heat shock cognate cDNA of zebrafish, *hsc70*, and developmental expression of corresponding transcripts. *Dev. Genet.* **21**, 223-233.

Schartl, M., Hornung, U., Hissmann, K., Schauer, J. and Fricke, H. (2005). Relatedness among east African coelacanths. *Nature* **435**, 901.

Scheufler, C., Brinker, A., Bourenkov. G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F. U. and Moarefi, I. (2000). Structure of the TPR domain-peptide complexes: critical elements in the Hsp70-Hsp90 multichaperone machine. *Cell* **101**, 199-210.

Schlieker, C., Tews, I., Bukau. B. and Mogk, A. (2004). Solubilization of aggregated proteins by ClpB/DnaK relies on the continuous extraction of unfolded polypeptides. *FEBS Lett*. **578**, 351-356.

Shashikant, C., Bolanowski, S. A., Danke, J. and Amemiya, C. T. (2004). *Hoxc8* early enhancer of the Indonesian coelacanth, *Latimeria menadoensis*. *J. Exp. Zool*. **302B**, 557-563.

Shi, Y. and Yokoyama, S. (2003). Molecular analysis of the evolutionary significance of ultraviolet vision in vertebrates. *Proc. Natl. Acad. Sci. USA.* **100**, 8308-8313.

Shin, Y., Klucken., J., Patterson, C., Hyman, B. T. and McLean, P. J. (2005). The co-chaperone carboxyl terminus of Hsp70-interacting protein (CHIP) mediates α-synuclein degradation decisions between proteosomal and lysosomal pathways. *J. Biol. Chem.* **280**, 23727-23734.

Shomura, Y., Dragovic, Z., Chang, H. C., Tzvetkov, N., Young, J. C., Brodsky, J. L., Guerriero, V., Hartl, F. U. and Bracher, A. (2005). Regulation of Hsp70 Function by HspBP1: structural analysis reveals an alternate mechanism for Hsp70 nucleotide exchange. *Mol. Cell*. **17**, 367-379.

Shonhai, A., Boshoff, A., Blatch, G. L. (2005). Plasmodium falciparum heat shock protein 70 is able to suppress the thermosensitivity of an *Escherichia coli* DnaK mutant strain. *Mol. Genet. Genomics* **274**, 70-78.

Siegers, K., Bölter, B., Schwarz, J. P., Böttcher, U. M. K., Guha, S. and Hartl, F. U. (2003). TRiC/CCT cooperates with different upstream chaperones in the folding of distinct protein classes**.** *EMBO J*. **22**, 5230-5240.

Sink, K.J. and Scott-Williams, N. (2005). African Coelacanth Ecosystem Program-ROV *Trial Report*. 25 May 2005 (http://acep.co.za/content/view/276/118/).

Slepenkov, S. V. and Witt, S. N. (2002). Kinetic analysis of the interdomain coupling in a lidless variant of the molecular chaperone DnaK: DnaK's lid inhibits transition to the low affinity state. *Biochem*. **41** 12224-12235.

Slepenkov, S. V. and Witt, S. N. (2003). Detection of a concerted conformational change in the ATPase domain of DnaK triggered by peptide binding. *FEBS Lett*. **539**, 100-104.

Smith, D. F., Whitesell, L. and Katsanis, E. (1998). Molecular chaperones: biology and prospects for pharmaceutical intervention. *Pharmacol. Rev*. **50**, 493-513.

Smith, J.L.B. (1939). A living fish of Mesozoic type. *Nature* **3620**, 455-456.

Smith, J.L.B. (1953). The second coelacanth. *Nature* **4342**, 99-101.

Sondermann, H., Scheufler, C., Schneider, C., Höhfeld, J., Hartl, F. U. and Moarefi, I. (2001). Structure of a Bag/Hsp70 complex: convergent functional evolution of Hsp70 nucleotide exchange factors. *Science* 219, 1553-1557.

Song, Y. and Madison, D. C. (2005). Independent regulation of Hsp70 and Hsp90 chaperones by Hsp70/Hsp90-organizing protein Sti1 (Hop). *J. Biol. Chem*. **280**, 34178 – 34185.

Söti, C., Pál, C., Papp, B. and Csermely, P. (2005) Molecular chaperones as regulatory elements of cellular networks *Curr. Opin. Cell. Biol.* **17**, 210-215

Sousa, M. C. and McKay, D. B. (1998). The hydroxyl of threonine 13 of the bovine 70-kDa heat shock cognate protein is essential for transducing the ATP-induced confromational change. *Biochem* **37**, 15391-15399.

Stavnezer, J. and Amemiya C. T. (2004). Evolution of isotype switching. *Semin. Immunol*. **16**, 257-275.

Stefani, M. (2004). Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. *Biochemica et biophysica acta* **1739**, 5-25.

Stefani, R. M. and Gomes, S. L. (1995). A unique intron-containing *hsp70* gene induced by heat shock and during sporulation in the aquatic fungus *Blastocladiella emersonii*. *Gene* **152**, 19–26.

Stock D. W. and Swofford, D. L. (1991). Coelacanth's relationships. *Nature* **353**, 217-218.

Stock, D. W., Moberg, K. D., Maxson, L. R. and Whitt, G. S. (1991). A phylogenetic analysis of the 18S ribosomal RNA sequence of the coelacanth *Latimeria chalumnae*. *Environ. Biol. Fishes* **32**, 99-117.

Strub, A., Zufall, N. and Voos, W. (2003). The putative helical lid of the Hsp70 peptide-binding domain is required for efficient preprotein translocation into mitochondria. *J. Mol. Biol.* **334**, 1087-1099.

Suh, W. C., Burkholder, W. F., Lu, C. Z., Zhao, X., Gottesman, M. E. and Gross, C. A. (1998). Interaction of the Hsp70 molecular chaperone, DnaK, with its cochaperone DnaJ. *Proc. Natl. Acad. Sci. USA.* ***95***, 15223-15228.

Suh, W. C., Lu, C. Z. and Cross C. A. (1999). Structural features required for the interaction of the Hsp70- molecular chaperone DnaK with the its cochaperone DnaJ. *J. Biol. Chem*. **274**, 30534-30539

Suppini, J. P., Amor, M., Alix, J. H., and Ladjimi, M. M. (2004)**.** Complementation of an *Escherichia coli* DnaK defect by Hsc70-DnaK chimeric proteins. *J. Bacteriol*. **186**, 6248-6253.

Takahashi, A., Amemiya Y., Nozaki, M., Sower, S. A. and Kawauchi, H. (2001). Evolutionary significance of proopiomelanocortin in agnatha and chondrichthyes. *Comp. Biochem. Physiol*. [B] **129**, 283-289.

Takahashi, A., Yashuda, A., Sullivan, C. V. and Kawauchi, H. (2003). Identification of proopiomelanocortin-related peptides in the rostral pars distalis of the pituitary in the coelacanth: evolutional implications. *Gen. Comp. Endocr.* **130**, 340-349.

Takezaki, N., Figueroa, F., Zaleska-Rutczynska, Z., Takahata, N. and Klein, J. (2004). The phylogenetic relationship of tetrapod, coelacanth, and lungfish revealed by the sequences of fourty-four nuclear genes. *Mol. Biol. Evol*. **21**, 1512-1524.

Tang, H., Liu, Y., Madabusi., L. and Gilmour, D. S. (2000). Promoter-proximal pausing on the hsp70 promoter in *Drophisola melanogaster* depends on the upstream regulator. *Mol. Cell. Biol*. **20**, 2569-2580.

Tohyama, Y., Ichimiya, T., Kasama-Yoshidi, H., Cao, Y., Hasegawa, M., Kojima, H., Tamai, Y. and Kurihara, T. (2000). Phylogenetic relation of lungfish indicated by the amino acid sequence of myelin DM20. *Mol. Brain Res*. **80**, 256-259.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. **22**, 4673-4680.

Towbin, H., Staehelin, T. and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Natl. Acad. Sci. USA. 76*, 4350-4354.

Tracy, M. R. and Hedges, S. B. (2000). Evolutionary history of the enolase gene family. *Gene* **259**, 129-138.

Tyedmers, J., Lerner, M., Bies, C., Dudek, J., Skowronek, M. H., Haas, I. G., Heim, N., Nastainczyk, W., Volkmer, J. and Zimmermann, R. (2000). Homologs of the yeast Sec complex subunits Sec62p and Sec63p are abundant proteins in dog pancreas microsomes**.** *Proc. Natl. Acad. Sci. USA.* **97**, 7214-7219.

Tyedmers, J., Lerner, M., Wiedmann, M., Volkmer, J and Zimmermann, R. (2003). Polypeptide-binding proteins mediate completion of co-translational protein translocation into the mammalian endoplasmic reticulum. *EMBO rep.* **4**, 505–510**.**

van der Berg, B., Wain, R., Dobson, C, M. and Ellis, R. J. (2000). Molecular crowding perturbs protein refolding kinetics: implications for folding inside the cell. *EMBO J*. **15**, 3870-3875.

Vandepoele, K., De Vos, W., Taylor, J.S., Meyer, A. and Van der Peer, Y. (2004). Major events in the genome evolution of vertebrates: Paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc. Natl. Acad. Sci. USA.* **101**, 1638-1643.

Venkatesh, B., Erdmann, M.V. and Brenner, S. (2001). Molecular synapomorphies resolve evolutionary relationship of extant jawed vertebrates. *Proc. Natl. Acad. Sci. USA.* **98**, 11382-11387.

Vogel, J. L., Parsell D. A. and Lindquist, S. (1995) Heat-shock proteins Hsp104 and Hsp70 reactivate mRNA splicing after heat shock. *Curr. Biol*. **5**, 306-317.

Wagner, G. P., Amemiya, C. and Ruddle, F. (2003). Hox cluster duplications and the opportunity for evolutionary novelties. *Proc. Natl. Acad. Sci. USA.* **100**, 14603-14606.

Wang, T. F., Chang, J. and Wang, C. (1993). Identification of the peptide binding domain of hsc70. 18-Kilodalton fragment located immediately after ATPase domain is sufficient for high affinity binding. *J. Biol. Chem*. **268**, 26049-26051.

Wegele, H., Haslbeck, M., Reinstein, J., and Buchner, J. (2003). Sti1 Is a Novel Activator of the Ssa Proteins. *J. Biol. Chem*. **278**, 25970-25976.

Williams, G. T. and Morimoto, R. I. (1990). Maximal stress-induced transcription from the human HSP70 promoter requires interactions with the basal promoter elements independent of rotational alignment. *Mol. Cell. Biol***. 6,** 3125-3136.

Wirth, A., Jung, Y., Bies, C., Frien, M., Tyedmers, J., Zimmermann, R. and Wagner, R. (2003). The Sec61p complex is a dynamic precursor activated channel. *Mol. Cell* **12**, 261–268.

Wu, C. (1980). The 5' ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. *Nature* **286**, 854-860.

Wu, C., Yamaguchi, Y., Benjamin, L. R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A, Handa, H. and Gilmour, D. (2003). NELF and DSIF cause promoter proximal pausing on the *hsp70* promoter in *Drosophila*. *Genes, Dev*. **17**, 1402-1414.

Yamashita, M., Hirayoshi, K. and Nagata, K. (2004). Characterization of multiple members of the Hsp70 family in Platyfish culture cells: molecular evolution of stress protein Hsp70 in vertebrates. *Gene* **336**, 207-218.

Yokobari, S., Hasegawa, M., Ueda, T., Okada, N., Nishikawa, K. and Watanabe, K. (1994). Relationship among coelacanths, lungfishes, and tetrapods: a phylogenetic analysis based on mitochondrial cytochrome oxidase I gene sequences. *J. Mol. Evol.* **38**, 602-609.

Yokoyama, S. (2000). Color vision of the coelacanth (*Latimeria chalumnae*) and adaptive evolution of rhodopsin (RH1) and rhodopsin-like (RH2) pigments. *J. Hered.* **91**, 215-220.

Yokoyama, S. and Tada, T. (2000). Adaptive evolution of the African and Indonesian coelacanths to deep-sea environments. *Gene* **261**, 35-42.

Yokoyama, S., Zhang, H., Radlwimmer, B.F. and Blow, N.S. (1999). Adaptive evolution of color vision of the Comoran coelacanth (*Latimeria chalumnae*). *Proc. Natl. Acad. Sci. USA.* **96**, 6279-6284.

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

Yost H. J and Lindquist S. (1991). Heat shock proteins affect RNA processing during the heat shock response of *Saccharomyces cerevisiae*. *Mol Cell Biol*. **11**, 1062–1068.

Young, B. P., Craven, R. A., Reid, P. J., Willer, M. and Stirling, C. J. (2001). Sec63p and Kar2p are required for the translocation of SRP-dependent precursors into the yeast endoplasmic reticulum *in vivo*. *EMBO J.* **20**, 262–271.

Zafarulla, M., Wisniewski, J., Shieman, S., Shworak, N, Misra, S. and Gedamu, L. (1992). Molecular cloning and characterization of a constitutively expressed heat shock cognate *hsc71* gene from rainbow trout. *Eur. J. Biochem.*, **204**, 893-900.

Zardoya, R. and Meyer, A. (1996). Evolutionary relationships of the coelacanth, lungfishes, and tetrapods based on the 28S ribosomal RNA gene. *Proc. Natl. Acad. Sci. USA.* **93**, 5449-5454

Zardoya, R. and Meyer, A. (1997). The complete DNA sequence of the Mitochondrial genome of a "living fossil," the coelacanth (*Latimeria chalumnae*). *Genetics* **146**, 995-1010.

Zhong, T. and Arndt, K. T. (1993). The yeast SIS1 protein, a DnaJ homolog, is required for the initiation of translation. *Cell* **18**, 1175-1186

Zhu, X., Zhao, X., Burkholder, W. F., Gragerov, A., Ogata, C. M., Gottesman, M. E. and Hendrickson, W. A. (1996). Sructural analysis of substrate binding by the molecular chaperone DnaK. *Science* **272**, 1606-1615.

Zimmerman S. B. and Trach, S. O. (1991) Estimation of macromolecule concentrations and the excluded volume effects for the cytoplasm of *Escherichia coli*. *J. Mol. Biol*. **222**, 599-620.

APPENDIX A

STANDARD PROTOCOLS

A1. Isolation of high molecular weight DNA from mammalian cell using proteinase K and formamide

Approximately 1 g of tissue was dropped into a mortar with liquid nitrogen and ground to fine powder using the pestle. The liquid nitrogen was allowed to evaporate and the ground powder was transferred and submerged into a 50 ml falcon tube with ~ 10 volumes (w/v) lysis buffer (10 mM Tris-Cl, pH 8.0; 0.1 M EDTA, pH8; 0.5% SDS; 20μg/ml DNase free RNase) and incubated at 37°C for 1 hour. A 50 μl aliquot of proteinase K (20 mg/ml) was added and mixed with a sterile glass rod and incubated at 50°C for 3 hours. The lysate was subsequently cooled to 15°C followed by the addition of 1.25 ml of formamide denaturing buffer (20 mM Tris-Cl; 0.8 M NaCl; 80% Formamide) for every 1 ml of lysate and mixed gently with a glass rod. The lysate was incubated at 15°C for 12 hours. The viscous solution was decanted into a dialysis bag (10 kDa molecular weight cut off) and dialyzed against 2 liters buffer 1 (20 mM Tris-Cl pH 8.0; 0.1 M NaCl; 10 mM EDTA pH 8.0) for 1 hour at 4°C. The buffer was replaced with fresh buffer 1 and dialyzes for four hours followed by a further 4 hours in the third 2 liters of buffer 1. Buffer 1 was subsequently replaced by two liters of buffer 2 (10 mM Tris-Cl pH 8.0; 10 mM NaCl; 0.5 mM EDTA pH 8.0) which was changed three times with dialysis allowed for at least 1 hours for each buffer change. DNA was resolved by electrophoresis on a 0.6% agarose gel. The concentration was determined by diluting 20 μl of DNA solution in 480 μl of TE buffer and read at 260 and 280 ηm.

A2. Isolation of plasmid DNA

Plasmid DNA was isolated using the QiaPrep kit (Qiagen, Germany). The kit was used according to manufactures instructions.

A3. Polymerase chain reaction

Polymerase chain reaction was performed using the Expand High fidelity PCR system (Roche, Germany). A 50 µl reaction volume contained the following reagents (unless otherwise stated):

Reagents	**Volume (µl)**
Sterile deionised H_20	38
dNTP mix (10 mM each)	1
Forward primer (300 nM)	2
Reverse primer (300 nM)	2
Template DNA (100 ng)	1
10 x Reaction buffer	5
Expand High Fidelity Enzyme (3.5 U/ µl)	1
Total volume	50

The reaction was resolved on a suitable percentage agarose gel (generally 0.8% agarose).

A4. Agarose Gel purification of DNA fragments and ligation reaction

Agarose gel purification of DNA fragments was carried out using the GFX PCR product and nucleic acid purification kit (Amersham, Germany). The purified fragments we ligated into appropriate plasmids.

The standard protocol for inserting PCR products into pGEM-T Easy vector (Promega, USA) was followed as per manufacturer's instructions. The PCR product (100 ng) was added to 2 x rapid ligation buffer (60 mM Tris-HCl, pH 7.8, 20 mM $MgCl_2$, 20 mM DTT, 2 mM ATP, 10 % polyethylene glycol) and pGEM-T Easy

vector (50 ng). T4 DNA ligase was added (0.3 U). The mixture was incubated at 4°C overnight and transformed into competent *E. coli* XL1-Blue (Appendix, Section B1). Cells were plated on 2xYT agar (supplemented with antibiotics where required) and colonies were selected for DNA isolation, as described previously. The isolated plasmid DNA was restricted with *Eco* RI or *Not* I, which release inserted fragments.

Similarly, T4 DNA ligase (Roche, Germany) was used for ligation of DNA fragments into vectors such as pQE30 (Qiagen, Germany). Appropriate amounts of the DNA fragments (100 ng) to be ligated were incubated with the 10 x ligase buffer (660 mM Tris-HCl; 50 mM $MgCl_2$; 50 mM dithiothreitol (DTT); 10 mM ATP, pH 7.5) and 1 U of the ligase, to a final volume of 30 µl. All ligation reactions were performed overnight at 4°C, or at room temperature for four hours. Resultant products were transformed into appropriate competent cells as described in Appendix A5.

A5. Preparation of competent cells and transformations

Appropriate *E. coli* cells were inoculated into 2 x 5 ml LB broth and incubated overnight with shaking at 37°C. The overnight cultures were diluted 200 times with LB broth and incubated with shaking at 37°C until to early log phase. The cultures were then centrifuged at 5000 g for 5 min at 4°C and kept on ice. The cell pellet was resuspended in one culture volume of ice cold 0.1 M $MgCl_2$ and was left on ice for 20 minutes followed by centrifugation at 5000 g for 5 minutes at 4°C. The pellet was resuspended in a half culture volume of ice cold 0.1M $CaCl_2$ and left on ice for 1-2 hours before the cell were pelleted as mentioned. One tenth culture volume of $CaCl_2$ and one volume of 30% glycerol were added to the cells followed by storage at -70°C in suitable aliquots.

Transformations were performed as follows. Ligation reactions or plasmid DNA to be transformed were added to 100 µl of the competent cells. The cells were left on ice for 20 minutes. The mixture was heat shocked at 42°C for 2 minutes. The cells were returned to ice for 1 minute, and 0.9 ml YT broth was added. The cells were left to produce antibiotic resistance for 1 hour at 37°C and subsequently plated on YT-agar plates (15 g agar per 1 l broth) containing appropriate antibiotics. Selected colonies

were grown in YT broth and plasmid DNA was recovered as described previously (Appendix A4).

A6. Restriction digests

Plasmid DNA was isolated as described in Appendix A2. Reaction digest contained plasmid DNA (200 ηg), appropriate restriction buffer (2 µl) and 2U of appropriate restriction enzyme, up to a total volume of 20 µl. The reaction was incubated for 4 hours at appropriate temperature. The restriction digest was resolved by gel electrophoresis using a 0.8 % agarose gel.

A7. BAC DNA isolation

BAC DNA was isolated using the large DNA construct isolation kit (Qiagen, Germany) according to manufacture's instructions.

A8. *Sau*3a partial digests

Ten eppendorf tubes were labelled 1 - 10. A mixture containing 20 µL 10X Sau3A buffer (New England Biolabs, United Kingdom) and 2 µL BSA (10 mg/mL; New England Biolabs) and 200 µL of a BAC DNA (0.1 µg/µL prepared as described in Appendix A7). The mixture 40 µL was dispensed into tube 1 and 20 µL was transferred into tubes 2 - 10. The tubes were kept on ice, and 1 µL of *Sau*3A (10 units/µL; New England Biolabs, United Kingdom) was added to tube 1 and mixed gently. A 20 µL aliquot of the mixture in tube 1 was subsequently transferred into tube 2 which was serially diluted up to tube 9. No dilution was made into 10 which served as a no-digest control. All ten tubes were incubated for 1 hour at 37˚C. The restriction enzyme was inactivated at 65˚C for 20 minutes. The reactions were resolved on 0.8% agarose gel.

A9. Resolving of DNA on agarose gels

Appropriate amount of molecular grade agarose was melted in 0.5 x Tris-Borate-EDTA buffer (TBE; 45 mM Tris; 45 mM Borate; 1mM EDTA) to give an appropriate percentage (w/v) agarose. Subsequent to cooling and addition of ethidium bromide to a total of 0.5 µg/ml, the gel was left to set in a sealed platform. A 6x DNA gel loading buffer 2µl (0.25% bromophenol blue; 30% glycerol) was added per 10µl of DNA. The mixture was loaded on the gel. The marker used was λ DNA (Promega, USA) restricted with *Pst* I. Agarose gels were run at 100 V for 1 hour, and the DNA was visualised using ultra-violet (UV) light.

A10. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Preparation of protein samples was carried out by the addition of SDS-PAGE sample buffer (0.0625 M Tris, pH 6.8; 10% glycerol, 2% SDS; 5% β-mercaptoethanol; 0.05% bromophenol blue) to the sample in a ratio 1 part buffer to 4 parts sample. The mixture was then boiled for 5 minutes.

The BioRad minigel system was used for the running of all gels described. A 12 % acrylamide separating gels (0.375 M Tris, pH 8.8; 0.1 % SDS; 12 % acrylamide) and a 4 % stacking gels (0.125 M Tris, pH 6.8; 0.1 % SDS; 4 % acrylamide) were prepared. Polymerization of both the stacking and separating gels was induced by the addition of 0.05 % ammonium persulphate (APS) and 0.005 % N,N,N'N'-tetramethylethylenediamine (TEMED). The gels were resolved in 1x SDS running buffer (25 mM Tris; 192 mM glycine; 1 % SDS) at 200V for 1 hour. The polyacrylamide gel were stained in Commassie stain (40% methanol; 0.7% acetic acid; 0.075% Commassie dye) and distained using Commassie destain (40% methanol; 0.7% acetic acid).

A11. Western analysis

Transfer of proteins onto nitrocellulose was performed as originally described by Towbin and coworkers (Towbin *et al.*, 1979). Protein transfers were carried out using the BioRad western apparatus (BioRad, USA) with freshly prepared transfer buffer (25 mM Tris; 192 mM glycine; 20% methanol). The transfer was performed at 4°C for 1 hour at 100 V. Proteins transferred onto nitrocellulose were visualized using Ponceau S stain (0.5% Ponceau S; 1% glacial acetic acid). Destaining was performed using distilled water.

Chemimiluminescent-based immunodetection of histidine-tagged proteins was carried out ECL chemimiluminescent immunodetection kit (Amersham, Germany). Subsequent to the transfer of protein onto the nitrocellulose membrane, the membrane was rinsed twice with Tris buffered saline (TBS: 50 mM Tris; 150 mM; Nacl; pH 7.5) The nitrocellulose membrane blot was then left in blocking solution (5% fat free milk in TBS) at 4˚C overnight. The membrane was incubated in murine produced monoclonal anti-His antibody (Amersham, Germany) diluted 1 in 5000 in the blocking solution, for 1 hour at room temperature. The membrane was then washed 3 times in Tris buffered saline-Tween 20 (TBST: 0.1% Tween 20 in TBS) for 20 minutes. Following the washes, the membrane was incubated in horseradish-peroxidase conjugate anti-mouse antibody (Amersham, Germany) diluted 1 in 5000 in blocking solution for 1 at room temperature. The membrane was washed 3 times in TBST for 20 minute each. The ECL chemiluminescence reagents (ECL chemiluminescence immunodetection kit; Amersham, Germany) were prepared by mixing reagent A to reagent B in a 1 to 1 ratio. The membrane was placed in the BioRad ChemiDoc EQ (BioRad, USA) and the mixed reagent were place on the protein side of the membrane. Images were capture using the auto expose settings.

A12. Smart buffer plasmid isolation protocol

An overnight culture 1.5 ml was placed in an eppendorf tube and centrifuged for 2 minutes at maximum speed (16 000 x g). The supernatant was removed and the pellet was resupended in 50 µl of smart buffer (10mM Tris pH8.0; 1 mM EDTA; 15% (v/w)

sucrose; 200 µg/ml RNase; 100 µg/ml BSA; 2 mg/ml lysozyme). The mixture was incubated at 37°C for 30 minutes then boiled for 90 seconds. The mixture was place on ice for 1 minute and centrifuged for 10 minutes at maximum speed (16 000 x g). 10 µl of the supernatant was used per restriction digest.

APPENDIX B

STRAINS OF *ESCHERICHIA COLI* AND PLASMIDS

B1. Strains of *E. coli* used in this thesis

Table B1 Genotype of strains of *E. coli* used in this thesis

Strain	Genotype
XLI-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F′ proAB lacIqZΔM15 Tn10 (Tetr)*
JM-109	*e14-(McrA-) rec A1 endA1 gyrA96 thi-1 hsdR17(rk-mK+ supE44 relA1 Δ(lac-proAB) [F′ traD36 proAB lacIqZΔM15]*
BB2362*	*dnaK756 recA::Tc^{R}pDM1*

* The BB2362 strain (dnaK756) is resistant to bacteriophage lambda and unable to grow above 40°C. This strain expresses a DnaK with three amino acid substitutions, one of which reduces its affinity for GrpE, whilst the two other substitutions elevate the basal ATPase activity of DnaK (Buchberger *et al*., 1999).

B2. Vectors and plasmids used in this thesis

Table B2 Vectors and plasmids vectors used in this thesis

Plasmid	Description	Supplier
pGEM-T Easy	Cloning and sequencing of PCR products.	Promega, USA.
pGEM-T_LcN	Cloning and sequencing of the PCR fragment encoding the LcHsp70 ATPase domain.	Generated in this thesis.
pGEM-T_LcC	Cloning and sequencing of the PCR fragment encoding the LcHsp70 substrate binding domain.	Generated in this thesis.
pGEM-T_LmHsp70	Cloning and sequencing of the PCR fragment encoding the partial full-length LmHsp70.	Generated in this thesis.
pGEM-T_LcHsp70	Cloning and sequencing of the PCR fragment encoding the partial full-length LcHsp70.	Generated in this thesis.
pGEM-T_ FLLmhsp70	Cloning and sequencing of the PCR fragment encoding the full-length LmHsp70.	Generated in this thesis.
pQE30	T5 promoter-based expression vector; allows insertion of a coding region in frame with and downstream of the codons for six histidines.	Qiagen, Germany
pQE30_ Lchsp70	pQE30 containing the LcHsp70 partial full-length coding region	Generated in this thesis.
pQE30_ Lmhsp70	pQE30 containing the LcHsp70 partial full-length coding region.	Generated in this thesis.
PQE60	T5 promoter-based expression vector used; allows insertion of a coding region in frame with and upstream of the codons for six histidines.	Qiagen, Germany
pBB46	pQE60 containing the *E. coli* DnaK coding region. The DnaK protein produced from this plasmid was not Histidine tagged since its coding region contained a stop codon upstream of the codons coding for the six Hisdines.	Burkholder *et al.*, 1996.
pCCBAC1E	BAC vector used for the construction of the *L. menadoensis* genomic library.	Epicentre, USA
pBK-CMV	Construction of the *L. menadoensis* BAC 25 minilibrary.	Stratagene, USA

B2.1 Plasmid maps of plasmids used in this thesis

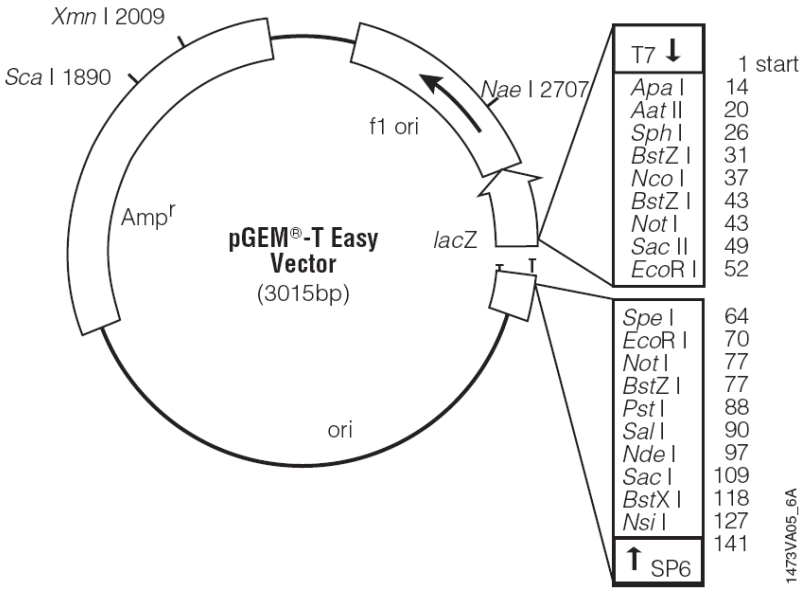


Figure B2.1 Plasmid map of the pGEM-T easy vector

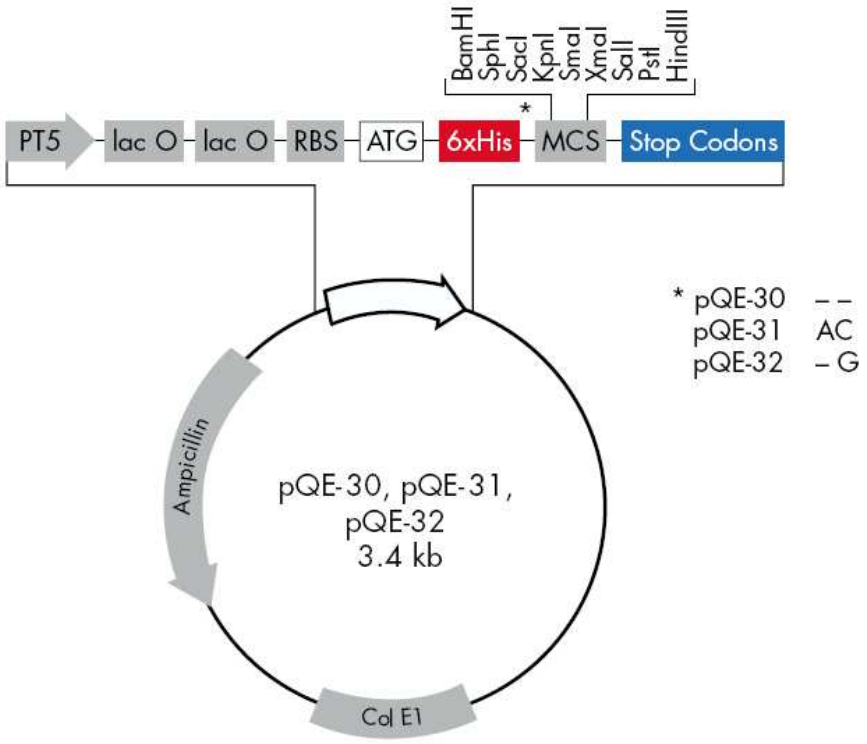


Figure B2.2 Plasmid map of the pQE30 bacterial protein expression vector (Qiagen, Germany).

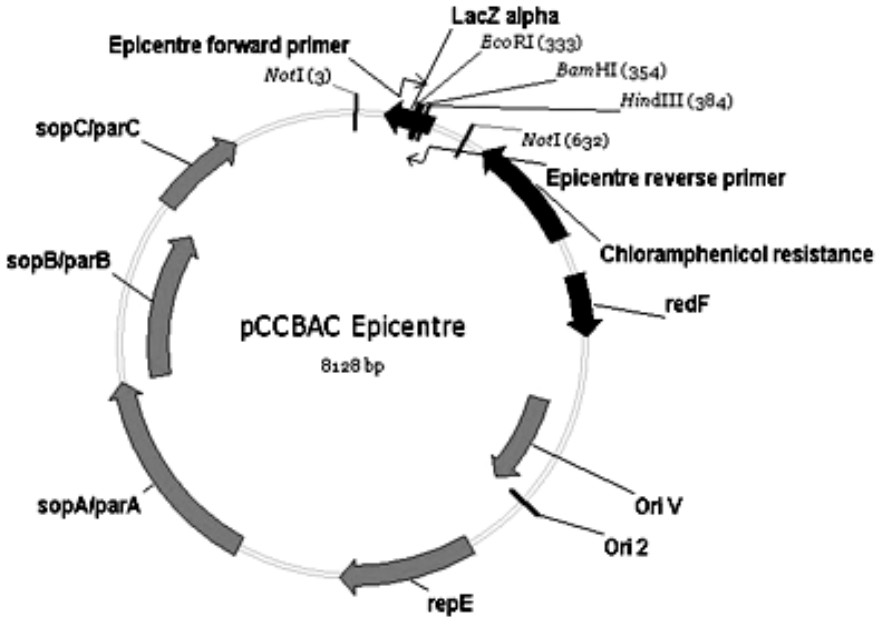


Figure B2.3 The plasmid map of pCCBAC1E BAC vector (Epicentre, USA).

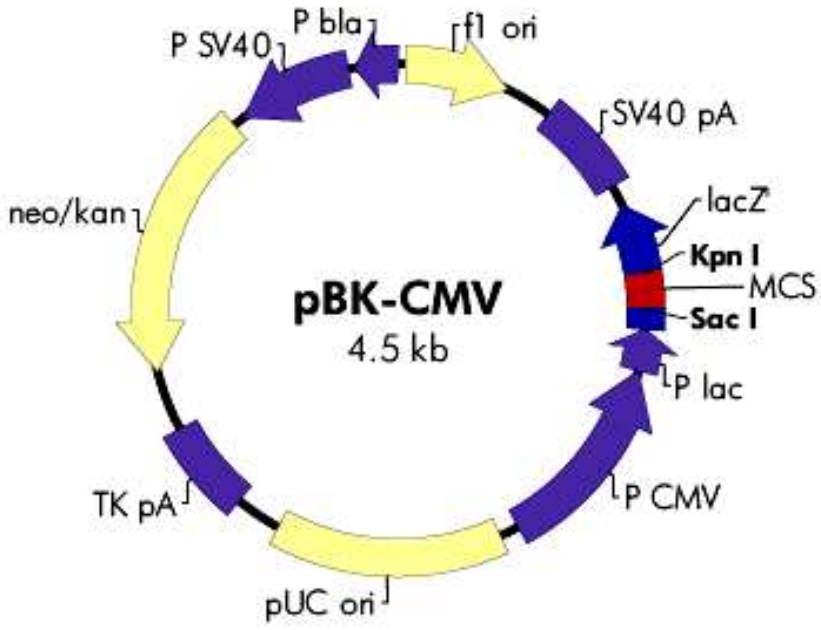


Figure B2.4 The plasmid map of pBK-CMV vector (Promega, USA).

APPENDIX C

PCR PRIMERS

Table C1 PCR primers used in this thesis

Primer	Sequence	T_m
Hoxa-11F/2	5'-CCATTGCTACTCAACAGG–3'	53.8
Hoxa-11R/2	5'- AGAGTGGTTTTCTGTGC - 3'	54.5
Fish 70F/2	5'-GGWATYGAYCTSGGWACMAC-3'	54.8
Fish 70/R2	5'-TTARTCVACYTCRATKGT-3'	53.0
Fish 70/intF2	5'-CCGTTGAAGAAGTCT-3'	50.1
Fish 70/intF2_R	5'-CAGGATTTCTTCAACCG-3'	51.7
Fish70F_3.1	5′-GGAATCGATCTGGGTACCAC-3′	55.5
Fish70F_3.2	5'-AGATCATCACCAATGACCAG-3'	53.0
Fish70R_3.1	5'-TTAGTCGACTTCTTCAATGGT-3	52.6
Fish70R_3.2	5′-TAGCTTTGCCTGGTTGGG-3′	56.2
Probe1F	5′-TTGAAGTAAAGTCTACAGC-3′	53.8
Probe1R	5′-AGAGTCGATTTCAATACTGG-3′	55.6
BAC25/70F	5′-ACCTCG GTA CCA CTT ACT C-3′	52.0
BAC25Rsp	5′-AGAGTTTCATGTCTGATTGG-3′	53.0
FL_Lm70	5′-ATGTCTGCAGCCAAGGGAG-3′	57.6
BAC25/70R	5′- TTAATCGACCTCTTCAATG-3′	54.3
Bam HI_Lchsp70F	5’-GGATCCATCATCACCAATGAC-3’	59.0
Pst I_Lchsp70R	5’-CTGCAGATTAGTCGACTTCTT-3’	58.5
Bam HI_Bac25_70F:	5’-GGATCCATGTCTGCTGCAGCC-3’	59.5
Kpn I_BAC25_70R:	5’-GGTACCTTAATCGACCTCTTCAATG-3’	58.5

APPENDIX D

PREPARATION AND SCREEN OF HIGH DENSITY COLONY HYBRIDIZATION FILTERS

D1. Processing of high density bacterial colony filters for hybridization

This method was used for preparing colony filters from BAC/PAC clones. The filters generated are suitable for hybridization with radiolabeled or biotin-labeled probes.

Using the BioGrid BG600 (BioRobotics, USA) transformants were picked and applied to four 22 cm^2 hybond$^+$ nylon membrane overlaid onto 300 ml of LB-agar (+ CAM or KAN) in a bioassay dish. The membrane was incubated overnight at 37°C. A piece of 3 mm filter paper slightly larger than nylon membrane was placed onto a large square glass plate which is placed within a cafeteria dining tray[1]. The 3 mm paper was saturated with lysis solution (2x SSC; 5% SDS) and flattened to remove bubbles. The colony membranes were place on 3 mm filter paper, colony-side up and allowed to sit for 3 minutes at room temperature or until the colonies appeared very mucoid. The square glass was placed onto a turntable in a microwave oven and microwaved at maximum power for 3 minutes or until the membrane was completely dry. Proteinase K buffer (50 mM Tris (pH 8); 50 mM EDTA; 100 mM NaCl; 1% N-lauryl sarcosine; 10 µg/ml proteinase K), approximately 100 ml per four filters, was placed in a large Pyrex baking dish. The membranes were submerged individually, swirled in the dish so the filter was completely wetted with the solution. The dish was subsequently covered with saran wrap and incubated at 37°C for 2 hours, rocking the dish occasionally. After this incubation, the colonies appeared as indentations in the filter and there were no bacterial debris. The filters rinsed in a large volume of 2x SSC for 10 minutes, air dried and UV (ultraviolet) crosslinked using the "autocrosslink" setting on the Stratalinker 1800 UV Crosslinker (Stratagene, USA)while the filter was still damp.

[1] A large tray/dish may be used for this purpose as it merely serves as a sink for excess lysis buffer.

D2. Labelling of probe with α-^{32}P- dCTP

The random priming reaction was carried out using the NEBlot kit (New England Biolabs, United Kingdom). 25 ηg of probe DNA was dissolved in 28 µl of nuclease free H_2O and denatured in a boiling water bath for 5 minutes. The denatured DNA was immediately placed on ice for 5 minutes and briefly centrifuged. The following reagents were added:

5 µl of 10 x labelling buffer containing octadeoxyribonucleotides (New England Biolabs, United Kingdom)

6 µl of dNTP mixture (dATP, dGTP, dTTP;10 mM each; New England Biolabs, United Kingdom)

5 µl of α-^{32}P dCTP (50µCi; New England Biolabs, United Kingdom)

1 µl of DNA polymerase I Klenow fragment (3.5 units/ µl; New England Biolabs, United Kingdom).

The reaction was incubated at 37°C for an hour.

D3. Hybridization method for membrane-bound nucleic acids

This is a simple method that can be used for hybridization of phage lifts (including primary library lifts), dot blots, as well as Southern and Northern blots. The method employs a hybridization solution largely consisting of Sodium chloride, EDTA and Tris (SET).

The one membrane prepared in Appendix D1 was placed in a hybridization container with 10-15 ml of SET solution (0. 6 M NaCl, 0.02 M EDTA, 0.2 M Tris, 0.5% SDS, 0.1% Sodium pyrophosphate). Prehybridization was carried out for four hours. The probe prepared in section D2 was added directly to the hybridization dish containing SET solution and filters. The container was sealed and placed in a rotating water bath overnight at 65°C. The membrane was removed from the hybridization solution and placed in a small volume of general purpose wash solution (1X SSC, 0.1% SDS, 0.025% Sodium pyrophosphate) for 10 minutes at 52°C. The membranes were then washed in a large volume of high stringency wash solution (0.1X SSC, 0.1% SDS, 0.05 % Sodium pyrophosphate) for one hour at 52°C. The wash solution was changed

frequently. The membrane blot surface was covered with one layer of Saran wrap and exposed to an X-ray film at -70°C with intensifying screen three hours. Longer exposure times did not change the results.

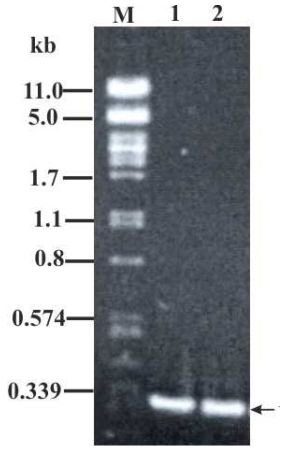


Figure D4.1: Lane 1=Isolation of the *L. chalumnae hsp70* specific DNA probe (*Lchsp70 5´*probe). PCR amplification of the 220 bp probe from the *L. chalumnae* 5' *hsp70* fragment encoding the N-terminal ATPase domain. M= λ Pst 1 size marker; 220 bp PCR probe.

Lane 2 = Isolation of the *L. menadoensis hsp70* specific DNA probe (*Lm* probe). PCR amplification of the 247 bp probe from the *L. menadoensis* 5' *hsp70* region of the partial full length BAC25 *Lmhsp70*.

APPENDIX E

COMPOUNDS AND SUPPLIERS

Table D1 Compounds and suppliers used in this thesis

Compound	Supplier	Country of Origin
Acrylamide	Sigma	Germany
Agar (Bacteriological)	Biolab (Merck)	South Africa
Agarose	WhiteSci	South Africa
Ammonium acetate	Saarchem	South Africa
Ammonium persulphate	Saarchem	South Africa
Ammonium sulphate	Saarchem	South Africa
Ampicillin	Roche	Germany
Bis-Acrylamide	Sigma	Germany
Boric Acid	Saarchem	South Africa
Bromophenol Blue	Sigma	Germany
Calcium Chloride	Saarchem	South Africa
Chloroform:Isoamyl alcohol (24:1)	Sigma	Germany
Coomassie Brilliant Blue G250	USB	USA
EDTA	Saarchem	South Africa
Ethidium Bromide	Sigma	Germany
Glycine	Sigma	Germany
IPTG	Roche	Germany
Kanamycin sulphate	Roche	Germany
Lambda DNA	Promega	USA
Lysozyme	Roche	Germany
Magnesium chloride	Saarchem	South Africa
Mercaptoethanol	Merck	Germany
Nucleotide mix	Roche	Germany
Ponceau S	Sigma	Germany
Potassium acetate	Saarchem	South Africa
Potassium chloride	Saarchem	South Africa

Potassium hydroxide	BDH	Germany
Potassium phosphate; K_2H	Merck	Germany
Potassium phosphate; KH_2	Merck	Germany
Proteinase K	Roche	Germany
Sodium acetate	Merck	Germany
Sodium chloride	Saarchem	South Africa
Sodium dodecyl sulphate	BDH	Germany
Sodium hydroxide	Saarchem	South Africa
Sodium Phosphate; Na_2H	Saarchem	South Africa
Sodium Phosphate; NaH_2	BDH	Germany
Sucrose	Sigma	Germany
TEMED	Sigma	Germany
Tris	Sigma	Germany
Tryptone powder	Oxoid	England
Tween 20	Saarchem	South Africa
Yeast extract	Oxoid	England

Restriction endonucleases were purchased from Amersham (Sweden) or from Roche (Germany).

Primers were purchased from Integrated DNA Technologies (USA) and Inqaba Biotechnology (South Africa).

APPENDIX F

LIST OF ABBREVIATIONS

α	Alpha
β	Beta
λ	Lambda
µg	Microgram(s)
µl	Microlitre(s)
µM	Micromolar
µmol	Micromole(s)
°C	Degrees Centigrade
A	Absorbance
ADP	Adenosine diphosphate
Amp^R	Ampicillin resistance
APS	Ammonium persulphate
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphatase
Bag-1	Bc12-associated athanogene 1
BAP	BiP associated protein
BiP	Binding protein
BLAST	Basic Local Alignment Search Tool
bp	Base pair(s)
BSA	Bovine serum albumin
CbpA	Curved DNA binding protein
DNA	Deoxyribonucleic acid
DnaJ	Prokaryotic Hsp40
DnaK	Prokaryotic Hsp70
EDTA	Ethylene diamine tetra-acetic acid
ER	Endoplasmic reticulum
g	gram
G/F region	Glycine-phenylalanine rich region
Hdj	Human DnaJ-like protein
Hip	Hsc 70 interacting protein
Hop	Hsp70 /Hsp90 organising protein
HPD motif	Histidine, proline and aspartic acid motif
His Tag	6 x Histidine Tag
Hsc	Heat shock cognate protein
Hsp	Heat shock protein

Hsp40	Hsp of 40 kDa
Hsp60	Hsp of 60 kDa
Hsp70	Hsp of70 kDa
Hsp90	Hsp of 90 kDa
IPTG	Isopropyl βD-thiogalactoside
kDa	Kilo Dalton
M	Molar
mg	Milligram
ml	Millilitre
MSA	Multiple sequence alignment
NCBI	National Center for Biotechnology Information
NMR	Nuclear magnetic resonance
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PDI	Protein disulphide isomerase
Pi	Inorganic phosphate
RNA	Ribonucleic acid
rpm	Revolutions per minute
SAP	Shrimp alkaline phosphatase
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel eletrophoresis
TBE buffer	Tris Borate EDTA buffer
TBS	Tris buffered saline
TBST	TBS containing Tween 20
TE buffer	Tris-EDTA buffer
TEMED	N,N,N'N'-tetramethylethylenediamine
TPR	Tetratricopeptide repeat
Tris	Tris-2-amino-2-(hydroxymethyl)-1,3,propandiol
w/v	Weight per volume
Ydj1	Yeast DnaJ
YT agar	Yeast tryptone agar
YT broth	Yeast tryptone broth

The IUPAC-IUBMB three and one letter codes for amino acids were used, and single letter codes were used for nucleotides.