1993

Cloning, sequencing, and expression of Brucella abortus heat shock 70 gene

Emily C. L. Chin
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Microbiology Commons, and the Molecular Biology Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/10803

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Cloning, sequencing, and expression of *Brucella abortus* heat shock 70 gene

Chin, Emily C. L., Ph.D.

Iowa State University, 1993
Cloning, sequencing, and expression of *Brucella abortus* heat shock 70 gene

by

Emily C. L. Chin

A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department: Zoology and Genetics
Major: Zoology (Molecular Biology)

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1993
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>MATERIALS AND METHOD</td>
<td>18</td>
</tr>
<tr>
<td>RESULTS</td>
<td>28</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>58</td>
</tr>
<tr>
<td>LITERATURE CITED</td>
<td>71</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>79</td>
</tr>
</tbody>
</table>
ABSTRACT

*Brucella abortus* is a gram negative bacteria that can infect and cause serious disease in many mammals including humans. Any organism, including *Brucella*, that can survive as an intracellular parasite must be able to survive the stress of a foreign and often hostile environment of the host. Stress usually results in a dramatic change in gene expression including an elevated synthesis of heat shock proteins.

We have subcloned, sequenced and expressed the *Brucella* HSP70 and the upstream sequences. The gene was put under the control of the strong "tac" promoter. However, when expressed, we saw a high accumulation of a 23 kDa polypeptide. Upon sequencing, this peptide was shown to be from the N-terminus of the protein.

Upstream of the initiation codon of the *Brucella* HSP70 gene is a small ORF complete with it's own initiation and termination codon and a good ribosome binding site. Various subclones were made and western blots from these showed that if the *Brucella* HSP70 gene is expressed without the ORF, the 23 kDa polypeptide accumulates. If the ORF is present, the protein accumulates intact. So, we concluded that this ORF is an important upstream regulatory element. Such a system has not been reported before and seems to be unique to *Brucella*. However, *Brucella* is an unusual parasite in that it can survive and multiply in the macrophage - the cell type of the immune system designed to kill it. It's unknown whether this form of regulation has any relationship to survival as an intracellular parasite.

To facilitate expression we also constructed seven expression vectors, pJE1 - pJE7. All these plasmids were derived from pKK223-3. They all carry the M13 origin of replication to facilitate production of single stranded DNA and an expanded polycloning site to facilitate cloning. pJE1 -6 are *in vivo* constitutive expression vectors with modified promoters such that there is a graded expression with pJE1 carrying the
strongest promoter. pJE7 is designed for \textit{in vitro} regulated expression using the "tac" promoter. This plasmid carries the \textit{lac}I\textsubscript{a} gene to prevent over expression from the "tac" promoter.
INTRODUCTION

Bovine brucellosis, a disease associated with abortions and infertility, is caused by the gram-negative organism \textit{Brucella abortus}. \textit{Brucella} exists in the host primarily as an intracellular facultative parasite, especially of the macrophage (11), and may persist in the host for years. \textit{Brucella abortus} can cause serious disease in humans and is capable of infecting many other mammals.

In general, parasites have evolved a balance with the host to allow continued host survival with the maximum opportunity for horizontal infection. Survival as an intracellular parasite includes surviving in a foreign and often hostile environment. A major question regarding the molecular biology of such organisms is how they cope with and survive rapid changes, such as elevated temperature, toxins, oxidants etc.

It has become clear that all organisms, including intracellular parasites, when stressed share a common molecular response that includes dramatic changes in the pattern of gene expression which includes the elevated synthesis of a family of heat shock or stress-induced proteins. Expression of these proteins is rapid and reaches a maximum level within 5-10 minutes followed by a decline to a new steady state level (59). From the molecular view point, the heat shock response represents a beautiful example of inducible gene expression.

These heat shock proteins (HSPs), also and more appropriately known as stress proteins, constitute some of the most highly conserved protein families in nature. Evolutionarily they are conserved with respect to both function and structure, and are present in both procaryotes and eucaryotes. Antibodies against the major \textit{Escherichia coli} HSPs cross react with those from \textit{Mycobacterium} and even yeast and other eukaryotes (45, 85, 96). Similar results have been reported for \textit{Coxiella burnetti} (90), \textit{Legionella pneumophila} (93), \textit{Treponema pallidum} (30), \textit{Borrelia
burgdorferi (28), and Salmonella typhimurium (32). More recently, it has been shown that many of the HSP family members, including those that do not respond significantly to heat shock, are induced under a variety of other stress conditions (13) whose common denominator may be the accumulation of unfolded or malfolded proteins in cells (3, 36). HSPs ensure survival under stressful conditions that, if left unchecked, would result in irreversible cell damage and ultimately cell death. In the cell, some of these proteins are produced constitutively at low levels and synthesis increases within minutes of temperature up-shift or other forms of stress. This activation response is rapid and intense. In eukaryotes, heat shock factors are activated and bind to the heat shock response element, a specific DNA recognition sequence located in the 5'-flanking sequence of the heat shock genes (13). The heat shock transcriptional response attenuates after prolonged exposure of cells to intermediate heat shock temperature or upon return to physiological temperature. As the name implies, HSPs were first noticed because of their specific induction by elevated temperatures (14, 46).

The exact function of HSPs is not fully understood, and in some cases, it is not known whether induced and constitutively produced HSPs have the same or different cellular functions. Generally, it is accepted that the highly conserved HSPs perform similar functions in all organisms and their importance is based on their capacity to associate with other proteins - that is to modify their destiny and/or functions (20). It has been shown that HSPs are involved with thermotolerance, probably by preventing or reversing the disassembly of oligomeric protein complexes or the unfolding of polypeptides and the acceleration of proteolysis of proteins denatured beyond repair (13). The fact that the majority of the family members are also expressed constitutively and abundantly even in the absence of any stress, indicates that they are probably important for normal cell functions (13, 17). Indeed
some of the major HSPs have been shown to be involved with general housekeeping functions associated with the synthesis, transport, and folding of proteins (13). This is accomplished by recognizing and stabilizing partially folded intermediates during peptide folding, assembly and disassembly (21). To reflect these more general roles as housekeeping proteins, HSPs have also been designated as molecular chaperones.

The majority of the currently identified molecular chaperones belong to three highly conserved protein families designated HSP60, HSP70, and HSP90 (13). Members of these families are widely distributed, and are found in prokaryotes, plants, and animals. In eukaryote cells, different members of each family are found in the various cellular compartments and organelles. Using an in vitro system, Langer (39) was able to mimic the in vivo interaction of E. coli HSP70, HSP60, GrpE, DnaJ, and GroES, and demonstrate their participation in polypeptide folding and tertiary structure formation.

The primary structures of at least some of the HSPs have been conserved between organisms as divergent as bacteria and man, suggesting that the function as well as the fundamental regulation of HSPs may be similar among different species. The molecular principles that govern the regulation of transcription of the heat shock genes are not fully understood. A key control element in E. coli is the gene rpoH (htpR) (26, 38) that encodes a specificity factor of RNA polymerase, sigma 32, which directs core RNA polymerase to recognize the promoters of heat shock genes including the rpoH gene itself (27, 86). RpoH-like genes have been identified in many procaryotes (26). Heat shock promoters in E. coli are recognized solely by RNA polymerase containing sigma 32 (12, 19, 98). During temperature shifts, the regulation of the transcription and translation of the rpoH gene controls the cellular concentration of sigma 32; and changes in the concentration of this sigma
factor are in turn responsible for regulating the transcription of the heat shock genes. There is an immediate but transient increase in the intracellular level of sigma 32 following temperature up-shifts. This appears to be responsible for the enhanced transcription initiation from the heat shock promoters (80), leading to a burst of heat shock gene transcription (40, 74, 80). Mutations in rpoH result in decreased constitutive transcription of heat shock genes and prevent induction of HSP synthesis by temperature up-shift (59, 92). Sigma 32 is an unstable protein (26). In fact, one of the most remarkable features of the sigma 32 is its extremely short half life (t1/2 = 1 min.) under steady-state growth conditions (26, 80). The fact that sigma 32 is unstable means that changes in its rate of synthesis will rapidly result in changes in intracellular concentration. In the normal cell, the sigma 32 gene is expressed constitutively at a low rate. The resultant low intracellular concentration of sigma 32 limits the expression of heat shock proteins during steady-state growth (26). As temperature increases, or when the cell is stressed, the sigma 32 transcription factor is stabilized, resulting in an increase of transcriptional efficiency and accumulation of the sigma 32 protein (79). The molecular mechanisms by which the cell senses the temperature change and promptly accumulates sigma 32 are still poorly understood. This increase in accumulation of sigma 32 factor could be a direct result of an increase in the rate of transcription or it could reflect a decrease in the rate of degradation or both.

When the temperature is down-shifted from 40°C to 30°C a transient repression of HSP synthesis occurs as a result of decreased transcription initiation at heat shock promoters (84). This repression is accompanied by only a small decrease in the level of sigma 32. In one study (80), Western blot analysis showed that the level of sigma 32 decreased only about 2 fold during the first 10 min. after shifting from 42°C to 30°C, while heat shock protein synthesis dropped about 20-fold during the
same time period. The level of sigma 32 continued its slow decline for 60 min.
before reaching its 30°C steady-state level of about 4-fold lower than at 42°C.
This indicates that the reduction in transcription initiation at the heat shock promoters
following a temperature downshift is caused by mechanisms not directly related to
the concentration of sigma 32 (84).

A further indication that reduced HSP synthesis after temperature down shift is
caused by mechanisms other than a simple decrease in the cellular amount of the
sigma 32 is given by two dimensional gel electrophoresis. Labeled spots
representing groEL and dnaK were both dramatically diminished after temperature
down shift. A number of other heat shock proteins were only mildly affected (84).
Several possibilities can be considered to explain this. Even if the regulation is not
directly dependent on sigma 32 concentration, it could involve control of the activity
of the RNA polymerase containing sigma 32. Regulation of this complex could be
brought about through direct modification (i.e. phosphorylation) of the complex, or
through the non covalent interaction of the core or the sigma 32 with other factors,
which allows for the rapid increase in activity during heat shock or other stresses,
and explains the rapid decrease in activity during a rapid temperature downshift.
Other work suggests that the synthesis of sigma 32 protein is normally repressed
during steady-state growth by a mechanism controlling translation of rpoH mRNA,
and that heat shock causes a transient derepression of synthesis. Following a
temperature downshift, translation of rpoH mRNA may already be as repressed as
the mechanism permits, necessitating a different means for further reducing the
expression of the heat shock genes (80). Yet another idea is that an undiscovered
repressor molecule might bind to conserved sequences at heat shock promoters
under the proper conditions and directly block transcription.
It has been observed that some HSPs, such as dnaK (procaryote HSP70), physically associate with the heat shock sigma factor. This was demonstrated by the co-purification of sigma 32 with the HSPs dnaK, dnaJ, and grpE. Such physical associations suggest that these proteins have very close interactions. Indeed, the *E. coli* HSP70 gene, DnaK, has been implicated as a regulator of sigma 32 production (80). The mechanism by which DnaK, DnaJ, and GrpE regulate the activity and stability of sigma 32 is assumed to rely on their activity as chaperones. This activity involves the ATP-dependent binding to substrates of DnaK, and the stimulation of hydrolysis of DnaK-bound ATP by DnaJ and GrpE (44, 45, 75). Thus, auto regulation of the heat shock response in *E. coli* has been proposed to occur in the following way. Under non-heat shock conditions, a large pool of free DnaK protein can bind sigma 32, sequestering it from RNA polymerase and/or rendering it accessible to cellular proteases (88), leading to its rapid degradation. When the cell is stressed, there is an increase of unfolded, or aggregated polypeptides. DnaK associates with these proteins, freeing sigma 32 to associate with RNA polymerase core, and increasing its half life. This, in turn, leads to increased transcription from the heat shock promoters and results in higher levels of heat shock gene expression. As the level of damaged proteins decreases, a larger proportion of the DnaK is unbound, resulting in the 'recapture' of sigma 32 and the dampening of the heat shock response. Generally, the heat shock response is temporary, even at prolonged intervals of high temperature. Work done by Grossman et al. (26) suggests that in *E. coli*, the DnaK gene either directly or indirectly acts to repress translation of rpoH mRNA. Induction of the heat shock response and the corresponding increase in the expression of dnaK results in the translational repression of rpoH. The resulting decrease in sigma 32 synthesis would lead to a
decline in the intracellular level of sigma 32 and ultimately to a decrease in the rate of synthesis of heat shock proteins (including dnaK).

It has been demonstrated that an increase in the rate of synthesis of sigma 32, even in the absence of a temperature increase, is sufficient for the increased expression of HSPs (26). This indicates that heat shock gene expression is regulated at least in part by changes in the intracellular concentration of sigma 32, implying that the sigma protein might be a direct regulator of heat shock gene expression. This model implies that the intracellular concentration of sigma 32 must be able to change rapidly in response to stress, and must be a limiting factor in the expression of the HSPs. However, it has also been noted that constitutive overproduction of sigma 32 fails to cause a proportionate increase of HSP transcription (80). This observation disputes the idea that sigma 32 acts as a direct regulator. Instead, it might simply function in determining promoter specificity while other transcription factors actually control the level of activation or repression of the heat shock genes (39). So the exact role of the sigma 32 protein is uncertain.

DnaK is known to negatively regulate the heat shock response, in part by affecting the synthesis (26) and the stability (86) of the sigma 32 protein. The level of ATP-dependent proteolytic activity in the cell increases in the absence of functional dnaK protein (86). Several hypotheses have been put forward to explain this phenomenon. These include: the stabilization of a protease, increased synthesis of a protease, and decreased level of activity of a protease inhibitor. E. coli selectively degrade proteins with abnormal conformations (24). Only one of the proteases involved in the degradation of abnormal proteins in E. coli, the product of the lon gene, has been identified genetically (9). The Lon protease is also a member of the HSP family whose synthesis is induced by a sudden increase in temperature.
Lon mutants have shown to be deficient in the degradation of many abnormal proteins (25) as well as certain unstable normal proteins (55).

In eucaryotes, the heat shock response is not regulated by a rpoH-like gene product. However, a similar situation exists in that there are multiple positively acting heat shock factors that control the transcriptional level of heat shock gene expression by binding to specific regions upstream of the heat shock promoters (76). Some of these factors have been shown to individually affect the expression of the heat shock response (31).

A 9-bp inverted repeat separated by a 9-bp spacer has been located between the transcription start point and the putative start codon of a number of HSPs from distantly related prokaryotic species including the groE operon of Bacillus subtilis (70), the dnaK and groE operons of Clostridium acetobutylicum (58), and a 10kDa antigen of Mycobacterium tuberculosis (4). A database search for additional genes preceded by this inverted repeat led to the discovery of several more examples (91). In most cases, this inverted repeat is found upstream of heat shock genes. These findings suggest that this sequence might be involved in the regulation of the heat shock response in some species, perhaps by stabilizing the mRNA against degradation or by acting as a binding site for a DNA or RNA binding protein.

Most HSPs are named according to their apparent molecular weight. One of the major HSPs is the HSP70. HSP70 belongs to a large gene family necessary for growth. Procaryotes in general have only one form of HSP70 (dnaK); while eucaryotes usually have several forms. For example, the HSP70 family in HeLa cells includes the major cytoplasmic heat inducible protein HSP70, the constitutively expressed heat shock cognate 70 (HSC70), a 71kDa mitochondrial protein (P71) and BiP/Grp78 found in the endoplasmic reticulum (41). HSP70 is an abundant protein that is highly conserved in evolution. There is 50% amino acid homology.
between \textit{E. coli} HSP70 (DnaK) and human HSP70. In the past several years it has also become evident that there is great similarity between prokaryotic dnaK proteins and the eukaryotic HSP70 homologs localized in the mitochondria, chloroplasts, cytosol, and endoplasmic reticulum. Chloroplast extracts immunoblotted with \textit{E. coli} DnaK antiserum reveal a DnaK-like protein (1).

Comparison of the amino-acid sequences of HSP70 family members reveals the presence of two separate functional domains - a more highly conserved N-terminal 44kDa domain and a more variable C-terminal domain (8, 52). In addition, some HSP70 proteins contain short N-terminal or C-terminal extensions required for targeting to, or retention in, the appropriate cellular compartments (57, 61, 62). In eucaryotes, sequences in the conserved N-terminus appear to include the ATPase domain. The more variable carboxy terminal includes the substrate recognition domain and is thought to play an important role in the specification for compartmentalization of the protein in the cell (41). Sequence variability in the C-terminal region between different family members may determine different specificity of stress-70-peptide interactions. Although the basis for specificity is not understood in any detail, it is likely that each stress-70 protein binds different polypeptide segments with a wide spectrum of affinities (21). Low affinity binding may be reversed quickly and spontaneously, whereas release of peptide segments that are bound with high affinity may involve ATP hydrolysis mediated by the N-terminal domain of the protein. This option may in turn be mediated by other cellular components such as Dna J (21).

The role of HSP70 proteins during the heat-shock response has been studied extensively for many years, but only recently has the importance of these proteins in normal cellular processes such as protein folding, assembly, disassembly and degradation become widely appreciated (64). Members of the HSP70 family have
been implicated in the stabilization of unfolded protein precursors prior to assembly of multisubunits complexes in the eucaryotic cytosol, or prior to translocation into organelles, in the stabilization of newly translocated polypeptides, in the rearrangement of protein oligomers, in the dissolution of protein aggregates, and in the degradation of rapidly turned-over cytosolic proteins (10). Experiments using microinjection of anti-HSP70 antibody (68) to lower the intracellular concentration of HSP70, indicate that HSP70 may participate in protecting cells against thermal stress but do not indicate that HSP70 expression alone is sufficient for thermal protection. Production of HSP70 is only part of the program of protein biosynthesis initiated after heat shock, and other components of the response also enhance survival.

The signal transduction pathway that converts environmental stress to specific alterations in the transcription of heat shock genes remains unclear. There is some evidence to suggest that the intracellular concentration of aberrant proteins is a major determinant of the cellular concentration of HSPs in E. coli as well as in eukaryotic cells (24, 36, 63). DnaK binds foreign eukaryotic proteins expressed in E. coli. As the local concentration of HSP70 protein increases, there is an increased probability that misfolding can be reversed. Perhaps the cell has evolved mechanisms to sense increased amounts of nascent or unfolded proteins in different cellular compartments and to respond by inducing the transcription of the appropriate HSP70 gene. In E. coli, it has been demonstrated that accumulation of unfolded proteins in the cytosol can cause an increase synthesis of DnaK (and other heat-shock proteins) (24, 63). In eukaryotic cells, it has been shown that accumulation of unfolded proteins (3) or secretary precursors (61) in the cytosol results in induction of HSP70 and/or HSC70 proteins, whereas accumulation of unfolded proteins in the ER causes induction of BiP (35, 61). It has been suggested
that HSP70 proteins assist correct polypeptide folding not only through their 'anti-folding' function but also by disentangling malfolded or aggregated proteins using the energy released during ATP hydrolysis (42, 64).

An attractive general model is that by sequestering HSP70 through its binding to aberrant proteins, induction of the heat shock response is triggered (14). Perhaps sequestering HSP70 prevents the heat shock transcription factor from interacting with HSP70, which in turn activates the heat shock gene transcription (14). Although genetic evidence supporting this model exists, its key prediction, the existence of physical interactions between HSP70 proteins and eukaryotic heat shock transcription factors, remains to be demonstrated. This model was re-evaluated by Gamer (20), he suggested that one of several possible signal pathways is that upon stress, both DnaJ and DnaK get increasingly bound to damaged proteins and thus become less available for binding to sigma 32 (14). This would account for the transient stabilization of sigma 32 during the heat shock response. An alternate signal pathway that he proposed is that temperature differentially affects the availability of DnaJ and DnaK for binding to sigma 32. Upon heat shock, DnaK might become bound to damaged proteins and thereby be less available for association with sigma 32, while DnaJ might retain its availability for binding sigma 32. Association of sigma 32 with DnaJ in the absence of DnaK after heat shock, as opposed to the presence of DnaK at 30°C, may increase the stability of sigma 32 by preventing its degradation.

Although in no case is the interaction between an individual stress-70 (HSP70) protein and its target polypeptide understood in detail, available evidence reveals several common features that point to a conserved mechanism for the action of these ubiquitous proteins (21). These features include 1) specific recognition of target polypeptides and modulation of their conformation or state of assembly, 2)
involvement of ATP binding and/or hydrolysis, 3) interaction with other HSPs or cellular factors, and 4) accumulation of unfolded protein in appropriate cellular locations is responsible for the induction of synthesis of individual stress-70 family members.

All HSP70 family members bind ATP (21), and a number of them have weak ATPase activities that can be elicited by appropriate substrates. Adenine nucleotide binding apparently causes conformational changes in HSP70 proteins that result in altered sensitivity to proteases (34), or an alteration of their oligomeric state (70). In addition to E. coli DnaK several of the members of the HSP70 family, such as bovine and human HSC70, have been shown to be phosphorylated \textit{in vivo} and the level of phosphorylation appears to be regulated to respond to stress and other factors(19). Recently it has been shown that DnaK and the eukaryotic HSP70 homologs localized in mitochondria, chloroplasts and the endoplasmic reticulum undergo a Ca$^{2+}$ stimulated autophosphorylation reaction (41). The widespread occurrence of phosphorylation and the similarity of the Ca$^{2+}$ stimulated autophosphorylation among members of the HSP70 group suggest that modification of these proteins may have physiological significance. In all members of the HSP70 family, the sequence DGGGTFD occurs in about the same location. The threonine in this sequence is known to be autophosphorylated in human cells (49). The high degree of conservation makes it extremely likely that autophosphorylation of many other HSP70 and HSC70 proteins occurs on the corresponding threonine of these proteins. Welch (1990) has proposed that the release of a target protein by HSP70 is coupled to ATP hydrolysis through a conformational change induced by an autophosphorylation event.

Under normal growth conditions, approximately 5% of the total DnaK in the cell is phosphorylated on threonine residues (99). The amino acid corresponding to
Thr199 of DnaK is conserved in every member of the HSP70 family of proteins including human HSC70 (18). It was shown that Thr199 was the amino acid that becomes phosphorylated and it was also demonstrated that threonine199 was critical for the ATPase activity of DnaK (49). Based on the X-ray crystal structure, Thr199 is near the midpoint of a highly conserved loop structure that lies between two highly conserved beta-sheets (18, 49). Both ATPase and autophosphorylation activities appear to be involved in the phosphorylation of other proteins in vivo, including certain aminoacyl-tRNA-synthetases (35). Site directed mutational analysis has identified the functional domains of DnaK for ATP binding, ATPase and autophosphorylation activity. Consistent with other HSP70 proteins, the ATPase activity resides in the amino terminal domain. Truncation mutants which show diminished ability to bind ATP also have diminished autophosphorylation activity (41). The ATPase activity of DnaK can be stimulated 50-fold by the presence of DnaJ and GrpE proteins (44).

In E. coli, autophosphorylation and ATPase activities of DnaK are both highly temperature dependent (49). Autophosphorylation activity was found to increase over the range of temperature that is physiologically relevant to the growth of E. coli. The temperature stimulation of DnaK ATPase and autophosphorylation activities might permit DnaK to function more efficiently as a molecular chaperone at high temperatures. At least in vitro, DnaK directly senses temperature within a physiologically relevant range. This leads to the hypothesis that, in vivo, DnaK may function as a cellular thermometer that directly senses temperature (49). Since HSP70s have been implicated in regulation of other heat shock genes in prokaryotes, it is possible that direct sensing of temperature by HSP70 could play a role in regulation of the heat shock response in eukaryotes as well. DnaK could serve to integrate the direct and indirect effects of temperature. It seems reasonable
that proteins would unfold as an organism shifts to temperatures that are near or above the upper limit of its growth range and, under such circumstances, the sensing of such partially denatured proteins by DnaK, as proposed by Craig and Gross (14), would contribute to the stress response. Direct sensing of the environmental temperature by DnaK could account for the rapid induction of heat shock gene expression after temperature upshift, where maximal induction is achieved approximately 5 min. after the shift (81). Thus, the strategy for *E. coli* survival at higher temperatures might include not only increasing the expression of DnaK, but also, increasing the activity of DnaK.

Very little is known about the mechanism and regulation of the degradation of the HSP70 protein, especially among procaryotes. Almost all of our knowledge about the regulation of the prokaryotic heat shock response is derived from *E. coli*, mostly from *in vitro* experiments. Instability has been reported (54), indicating that it is disposed of *in vivo* by some mechanism that apparently does not affect most other proteins. It has been variously suggested that there is a component that acts specifically to remove the protein or that it is responsible for its own degradation. In *Drosophila*, the *in vitro* degradation of the HSP70 has been demonstrated to be caused by self-proteolysis (54). It is proposed that this self proteolysis is determined by the structure of the protein itself. For example, the carboxyl end of the *Drosophila* HSP70 coding sequence contains the arrangement: ser-cys-his-asp, typically found in "serine" type proteases. This, however, is only a proposed mechanism, and there is no direct evidence for enzymatic activity of this sequence (54).

The most intensively studied member of the HSP70 family is the *E. coli* DnaK protein. This protein purifies predominantly as a monomer (86). Alignment of the dnaK proteins from various bacterial species revealed a substantial overall homology, which also declines towards the C-terminal end (91). DnaK was originally
identified as being essential for bacteriophage lambda DNA replication, but has now been shown to be essential for viability of *E. coli* at any temperature (17). DnaK functions in the modulation of the heat shock response as well as performing fundamental roles in normal cellular physiology including a major role in folding and unfolding of polypeptides. DnaK is the only stress-70 protein for which unfolding activity has been directly demonstrated *in vitro*. Aggregates of heat-inactivated RNA polymerase were solubilized in the presence of DnaK and ATP (75). This process was dependent on the presence of hydrolyzable ATP. Some of the regulatory events leading to heat shock response in *E. coli* have been elucidated. The heat shock response of *E. coli* is characterized by a transient increase in the rate of synthesis of about 20 proteins. The rate of synthesis of each HSP reaches a maximum between 5-10 minutes after temperature upshift (60). However, this high level of synthesis is transient, and when stress is removed, repression of synthesis occurs, mRNA synthesis is turned off, and the mRNA itself is degraded. The molecular and biochemical basis underlying these effects of HSP70 expression are unknown, although hypotheses have been advanced. Detailed *in vitro* studies have now illustrated the ability of DnaK to interact with both fully assembled or unfolded polypeptide substrates (21). Moreover, DnaK, which often associates with *E. coli* RNA polymerase, can stabilize the polymerase at elevated temperatures and protect the enzyme from heat inactivation *in vitro*. ATP is not required for this protective effect. Increased synthesis of DnaK occurs as a result of the accumulation in cells of unfolded polypeptides (63). Mutations in the DnaK gene result in temperature-sensitive growth of *E. coli*, overproduction of other heat shock proteins even at permissive temperatures, impaired synthesis of DNA and RNA, and a generalized defect in proteolysis (35). All these strongly suggest that the DnaK protein plays a role in the degradation of abnormal proteins, either by itself possessing a proteolytic
function directed against abnormal (misfolded or truncated) proteins, or by positively regulating a protein or system of proteins that possesses such activity. These observations all indicate the involvement of dnaK in modulating many protein-protein interactions in vivo.

Foreign HSPs are among the most immunodominant antigens recognized by the immune system, and have been shown to be the targets of immune responses in many nonviral infections. It has been shown that HSP70 can elicit strong humoral and cellular responses. Several lines of evidence have led to the proposal that the mammalian immune system is primed to respond to foreign heat shock proteins (94, 96). Immunized animals have been shown to mount an antibody response to the mycobacterial HSP70 protein despite the strong sequence homology with the mammalian HSP70s (91). Perhaps this strong immune response could be elicited because the protein is so highly conserved and is produced by all pathogens. The body could have immunological memory for cross reactive determinants from previous unrelated infections. However, it has also been shown that antibodies stimulated by members of the HSP70 family from a wide range of parasites and bacteria are directed mainly against the non conserved regions of the protein. Under most circumstances there is no reason to believe that this response is detrimental to the host, but it is possible that normal regulation may break down in some infections. Triggering of an auto immune response by the foreign HSP could result in suppression of the immune response to the pathogen, and contribute to other immunopathological aspects of infection (91). It has been proposed that an auto immune response of this type is associated with several diseases (reactive arthritis, Lyme disease, tuberculoid leprosy), but experimental verification of this is incomplete (95).
There is also suggestive evidence for an immune response against these proteins during *Brucella abortus* infection. Western blots of ammonium sulfate fractionated protein extracts from *B. abortus* probed with sera from infected cattle reveal a characteristic pair of bands which runs at 60kDa and 70kDa (6). It seems likely, but is not yet proven, that these bands correspond to HSP70 (dnaK) and HSP60 (groEL). Recent evidence has also shown that high expression of stress proteins including groEL and dnaK is induced upon phagocytosis of *S. typhimurium* by macrophages (16).

The starting point of this project was the body of data developed by Tabatabai and co-workers using a high salt extract of methanol killed *Brucella abortus* cells, called *Brucella* Cell Surface Proteins (BCSP). Rabbit antisera made against the BCSP was used to screen a library in lambda bacteriophage for expression of *Brucella* proteins. Clones expressing five different proteins were identified. These proteins were named BCSP20, BCSP31, BCSP45, BCSP58, and BSCP65 according to their apparent molecular weights. For reasons which were not apparent at the time BCSP65 expressing clones were the most commonly identified.

Tabatabai also explored the reaction of sera from vaccinated and infected cattle with the BCSP protein extract. Sera from either infected or vaccinated and challenged cattle reacted strongly with two bands in the 60-70 kDa range (6). It seemed likely that the cattle sera tested by Dr. Tabatabai reacted with *Brucella* HSP60 and HSP70 proteins. This would parallel what had been observed during *Mycobacterium* infection (94, 96). Rabbit antisera against high molecular weight BCSP was prepared and used to again screen the genomic *Brucella* DNA lambda library. This resulted in the isolation of several more clones expressing BCSP65, but no other proteins.
MATERIALS AND METHODS

Bacterial Strains and Plasmids

The *E. coli* strains used in this work are shown in Table 1. The plasmids used in this study which are not constructed in this study are listed in Table 2. The pUC plasmids and pKK223-3 are commercially available plasmids. Plasmids 11-8 and 11-4 were constructed by Moore (56) in Dr. J. Mayfield's lab.

Table 1: *E. coli* strains used for transformation

<table>
<thead>
<tr>
<th>strain</th>
<th>genotype</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hb101</td>
<td>Δ(gpt-proA)62 leuB6 thi1 lacY1 hsdSB20 recA rpsL20 (Str') ara-14 galK2 xyl-5 mtl-1 supE44</td>
<td>7</td>
</tr>
<tr>
<td>JM105</td>
<td>F' traD36 proA+ proB+ lacI9 lacZΔM15/Δ(lac-pro)X111 thi rpsL (Str') endA sbcB supE hsdR</td>
<td>93</td>
</tr>
<tr>
<td>Y1090</td>
<td>ΔlacU169 proA+ Δ (lon) araD139 strA supF trpC22::Tn10 mcrA/pMC9</td>
<td>53</td>
</tr>
<tr>
<td>TG1</td>
<td>supE hsdΔ5 thiΔ(lac-proAB) F'[traD36 proAB+ lacI9 lacZΔM15]</td>
<td>22</td>
</tr>
<tr>
<td>GM161</td>
<td>dam−</td>
<td>Yale Univ. Stock Center</td>
</tr>
</tbody>
</table>
Table 2: Plasmids used in this study for cloning and protein expression

<table>
<thead>
<tr>
<th>plasmid</th>
<th>reference</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC18</td>
<td>51, 62, 93</td>
<td>popular high copy cloning vector containing polycloning site within the alpha region of the lacZ gene, and carrying the beta lactamase gene.</td>
</tr>
<tr>
<td>pUC19</td>
<td>51, 62, 93</td>
<td>same as pUC18, but with the polycloning site in the reverse orientation.</td>
</tr>
<tr>
<td>pUC118</td>
<td>51, 62, 89, 93</td>
<td>same as pUC18 but with the M13 origin of replication</td>
</tr>
<tr>
<td>pUC119</td>
<td>51, 62, 89, 93</td>
<td>same as pUC19, but with the M13 origin of replication</td>
</tr>
<tr>
<td>pKK223-3</td>
<td>16, 2</td>
<td>created by Brosius and coworkers and purchased from Pharmacia. Contains the artificial &quot;tac&quot; (trp-lac fusion) promoter followed by a polycloning site, in turn, by the strong rRNA transcriptional terminator. The plasmid also includes the beta lactamase gene.</td>
</tr>
<tr>
<td>11-8</td>
<td>56</td>
<td>6.3kb SmaI <em>Brucella abortus</em> DNA insert in pUC19. The insert includes the BCSP65 (dnaK) gene.</td>
</tr>
<tr>
<td>11-4</td>
<td>56</td>
<td>12.5kb SmaI <em>Brucella abortus</em> DNA insert in pUC19. The insert is adjacent to the 11-8 insert on the Brucella chromosome and contains the dnaJ gene.</td>
</tr>
</tbody>
</table>
Plasmids 11-8 and 11-4 are Sma1 subclones of lamda clone 10-2-9 isolated from a λ1059 (33) *Brucella abortus* strain 19 DNA library (56). 10-2-9 was isolated because it expressed the BSCP65 protein which, in turn, was identified by antiserum prepared against the BCSP proteins. BCSP proteins are those proteins extracted by 1M NaCl plus 0.1M sodium citrate from methanol killed *Brucella* cells (82, 83). Fig. 1 shows restriction maps of lambda clone 10-2-9 (56).

**Antisera**

Rabbit antisera R222 and R213 were made by Dr. L. Tabatabai and co-workers at the National Animal Disease Center, Ames, Iowa by repeated injection of rabbits with the BCSP protein extracts (82, 83). Antiserum BA23 was prepared by Zhengyu Sha by eluting the 23kDa N-terminal peptide of cloned *Brucella* HSP70 from SDS-PAGE gels and subcutaneously injecting the polypeptide at multiple sites on the back of a rabbit.

**Enzymes, Plasmid DNA Purification and Cloning**

Enzymes were purchased from commercial sources and used according to the manufacturers' recommended conditions. Procedures for plasmid isolation and transformation into host strains using calcium chloride were as described in Maniatis *et al.* (47). Most subcloning was then done in pUC118 and pUC119 and transformed into the *E. coli* host TG1. These plasmids contain the beta-lactamase (ampicillin resistance) and beta-galactosidase genes, and colonies were plated out on agar containing ampicillin and overlaid with 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-gal) and isopropyl-beta-D-thiogalactoside (IPTG). White colonies were then further checked for the presence of appropriate size plasmids by a plasmid screening procedure, as described by Sekar (72), in which the colonies
Fig. 1: Restriction endonuclease site maps of lambda 10-2-9 DNA (ref. 56). The locations of the BamH1 sites present in the lambda 1059 vector genome are represented below each linear map as solid triangles, and indicate the boundaries of the inserted Brucella abortus DNA. The 6.8 kb Sma1 fragment was subcloned into pUC 19 to give us clone 11-8 which carries the dnaK gene. Clone 11-4 which carries the dnaJ gene was a subclone of the 12.5kb Sma1 fragment.
were suspended in a protoplasting buffer of 30 mM Tris HCl (pH 8.0), 5 mM EDTA, 50 mM NaCl, 20% sucrose, 50 ug/ml RNAse, and 50 ug/ml lysozyme. Five microliters (ul) of each sample was loaded onto wells containing 2 ul of lysis buffer (90% TBE, 0.05% SDS, 5% sucrose and bromo phenol blue). This was electrophoresed for 3 hours in TBE with 0.05% SDS at 120 volts in an 0.8% agarose gels with 0.05% SDS. The plasmid DNA released migrates according to size.

Appropriate size plasmids were then extracted from overnight cultures using the alkaline-SDS method as previously described (47). The DNA was then restriction digested at 37°C and separated on agarose gels to demonstrate the presence and size of appropriate inserts.

DNA fragments were digested with appropriate restriction endonuclease at 37°C to excise the sequence of interest from the vector and other flanking sequences. These fragments were either directly ligated into the appropriated vectors or gel purified by electrophoresing in appropriate percent agarose gel, poured and run in 50mM Tris-borate buffer (pH 8.2), visualized by staining in ethidium bromide (1 mg/ml), and the band excised from the gel in as small a volume as possible. The DNA was extracted from the agarose using SpinBind DNA extraction units from FMC Corp. Procedures for extraction were according to the instruction manual provided by the manufacturer. Fragments were eluted with TE, quantitated, and ligated overnight at 12°C to appropriate vectors using T4 DNA ligase (BRL). The DNA was then transformed into host bacteria. The high expression vector pKK223-3 and derived clones were propagated in E. coli JM105.

E. coli. Y1090, was cured of the episome with 100 ug/ml acridine orange. The cultured was grown overnight in the dark in LB titrated to a pH of 7.6. Cultures were streaked out on LB plates with no ampicillin. Single colonies were replicated onto LB
and LB with ampicillin (LB-A) plates. Colonies that grew on LB and not on LB-A media was used for preparation of competent cells.

**Sequencing and Autoradiography**

Single stranded DNA from subclones in pUC118 and pUC119 was made by co-infecting early log phase cell cultures with M13 helper phage K07. Infected cells were selected with ampicillin and kanamycin (ampicillin resistance from the plasmid in the bacteria and kanamycin resistance from the helper phage). Phage particles were precipitated from the overnight culture supernatants, using 5% polyethyleneglycol and 0.5M NaCl. After centrifugation, the DNA pellet was resuspended in TE, phenol-chloroform extracted, and ethanol precipitated. The single stranded DNA was then quantitated spectrophotometrically to determine the concentration. Sequencing was done manually using the didioxy method labeling with S\(^{35}\) according to procedure provided with the Sequenase Version 2.0 kit from United States Biochemical. The universal primer was used whenever possible, otherwise primers were synthesized at the Iowa State University Nucleic Acid Facility. Sequences were resolved with buffered gradient gels, with a final concentration of 8M urea (Aldrich), 6% acrylamide (Aldrich), 0.3% bis-acrylamide (Aldrich), and 0.5X TBE for the top gel. The final concentration for the bottom gel was 8M urea, 6% acrylamide, 0.3% bis, 2.5X TBE, 10% sucrose and 0.01% bromophenol blue. The gels were polymerized with ammonium persulphate catalyzed with N,N,N',N'-Tetramethylethyl-enediamine (TEMED). A BRL sequencing gel apparatus was used. After electrophoresis, gels were dried with a Biorad gel dryer, according to manufacturer's instructions. The dried gels were exposed to Kodak X-ray film (XAR-5) at -70°C overnight and developed as stated in the instructions supplied by the manufacturer.
Construction of *Brucella* HSP70 Clones

The *Brucella* HSP70 gene including the flanking sequences when identified from the lambda library was in a 6.5 Kb fragment ligated into pUC19 at the Sma1 site, giving us 11-8, a 9.2 Kb plasmid. This 6.5 Kb *Brucella* DNA fragment was introduced into pKK223-3 by digesting 11-8 with Xma1 (same recognition site as Sma 1) to excise the *Brucella* DNA, and EcoR1 to inactivate pUC19. The fragments were then ligated together with pKK223-3 after the vector had been digested with Xma1 and the 5' phosphates removed with calf intestinal phosphatase (CIP) (Boehringer Mannhein) as described in Maniatis (47). Treatment with CIP minimizes circle closing of the vector. The ligated plasmids were then transformed into *E. coli* JM105. Ampicillin resistant colonies were screened (rapid screen, ref.72) for the proper sized plasmids. Several of these plasmids were re-digested with EcoR1, religation (circle closing) was done to make sure that only one copy of the linker remained. Digestion with HindIII and Kpn1 determined the orientation of the insert, and the clone with the insert in the forward reading direction was called 11-8PKA. The plasmid with the insert in the reverse orientation was named 11-8PKB.

Next, 11-8PKA was digested with Cla1 and EcoR1. This DNA was ligated together with a complementary pair of 25 nucleotide double stranded primers which included a 5' terminal EcoR1 overhang, the N-terminal 20 nucleotides of the *dnaK* coding sequence, and a 3' terminal Cla1 overhang which matched the Cla1 site in the *dnaK* gene. The resultant clone was screened by size and called HSP70PK. Clone HSP70PK11 was generated from 11-8PKA by digestion with Rsrl and EcoR1. The ends were ligated together using a synthetic linker and transformed into JM105. Clone HSP70PKC was constructed by moving HSP70PK from JM105 into GM161 (dam-). The isolated plasmid was then digested with Cla1 and ligated to
circle close the plasmid. A similar clone, but lacking the ORF, was constructed from HSP70PK11. The \textit{Brucella} insert DNA was excised from HSP70PK11 with BamH1 and EcoR1 and cloned into BamH1 and EcoR1 digested pJE7, and the 5' phosphate removed with CIP. This clone identified as HSP70JE7 was then digested with Cla1 and EcoR1 and ligated together with linkers to give clone HSP70JE7C. A similar construct to HSP70PK11 was also made. Clone 11-8PKA was digested with Cla1 and EcoR1 and ligated with linkers (oligo# 174, 175), so that there was no duplication of the initiation sequence of the \textit{Brucella} Hsp70 gene.

Expression from 'tac' Promoters

Cultures were grown overnight in LB with ampicillin at 37°C with shaking, diluted 50 fold and returned to the shaker until they reached an O.D.\textsubscript{600} of 0.7. IPTG, to a final concentration of 3 mM, was added and the cultures grown for from for 30 minutes to overnight before total protein was extracted.

Total Protein Extraction and Electrophoresis

Sodium dodecylsulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was done as described by Laemmli (37). Bacterial cultures were pelleted and resuspended with the equal volumenof sample buffer, boiled for one minute, and then microfuged for 15 minutes. One hundred microliter of the upper supernatant was transferred to another tube with 4 ul of bromophenol blue. Forty microliters of each sample was then loaded onto the gel. Eleven percent acrylamide gels were used and electrophoresed at 100-150 volts with cooling, for 3-4 hours using 10 liters of running buffer (3 gm Tris, 14.4 gm glycine and 2.5 ml 20% SDS per liter water) in a Hoefer apparatus. After electrophoresis, the gels were stained with 0.1% Coomassie blue R-250 dissolved in 25% 2-propanol and 10% glacial acetic acid
(gHOAc) and destained in 25% 2 propanol and 10% gHOAc to visualize protein bands. Molecular markers (SDS-7) were purchased from Sigma.

Western Blot

Total proteins separated by SDS-PAGE were transferred from slab gels to nitrocellulose by standard procedures (87) using a Hoefer "Transphor" device (Hoefer Scientific Instruments). The transfer was done in buffer containing 12.4mM Tris base, 76mM glycine, 0.05% SDS, and 10% methanol, for one hour at 100 volts, with cooling. The nitrocellulose was then blocked for 15 minutes with 1% bovine serum albumin dissolved in PBS (145mM NaCl, 10mM NaPO₄ pH7.4), washed with PBS containing 0.05% Tween-80 (PBS -Tween) and then again with PBS. This was then probed at 37°C for 2 hours with a 1:200 dilution of rabbit antiserum. The nitrocellulose was then washed 5 more times with PBS-Tween, blocked for another 15 min. with PBS-BSA, and then incubated for 1 hour at room temperature with secondary antibody, horseradish-peroxidase-conjugate goat anti-rabbit IgG (Cooper Biomedical) diluted 1:200 (final concentration of 60 ug/ml of antibody protein). This was followed with five washes with PBS-Tween and then with PBS before developing the color with a solution of 0.06% 4-chloro, 1-naphthol and 0.012% hydrogen peroxide in 80% PBS-20% ethanol.

A molecular weight marker lane was cut from the rest of the lanes after gel transfer and was stained overnight with India ink (15 ul ink in 100 ul TBE).

N-Terminal Amino Acid Sequence Analysis

The N-terminal amino acid sequence was obtained by separating proteins by SDS-PAGE. The gel was then blotted onto Immobilon (Micron Separation Incorp.),
the desired protein spot cut out and the amino acid sequence analyzed by Dr. L. Tabatabai.

Synthesis of Oligonucleotides

Oligonucleotides were synthesized using the Biosearch 8700 DNA synthesizer, Waters Inc. by the Iowa State University Nucleic Acid Facilities. Those oligonucleotides used for construction of \textit{B. abortus} clones are listed in Table 3:

Table 3: Oligonucleotides used to construct \textit{B. abortus} clones

<table>
<thead>
<tr>
<th>oligo #</th>
<th>sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>701</td>
<td>AATTCATGGCTAAAGTTATTGGTAT</td>
</tr>
<tr>
<td>702</td>
<td>CGATACCAATAACTTTAGCCATG</td>
</tr>
<tr>
<td>905</td>
<td>GTCGACCACGCG</td>
</tr>
<tr>
<td>906</td>
<td>AATTCGCGTGTC</td>
</tr>
<tr>
<td>174</td>
<td>AATTCGAGCTCGGGTTAACAT</td>
</tr>
<tr>
<td>175</td>
<td>CGATGTTAACCCGAGCTCG</td>
</tr>
</tbody>
</table>
RESULTS

Part 1: The dnaK Gene and its Expression

Cloning and Sequencing of the Brucella abortus dnaK Gene

William Moore (56), identified a lambda clone which expressed the Brucella abortus protein designated as BCSP65. Moore also created a variety of subclones of the Brucella DNA inserts of this clone, several of which were later shown to weakly express the protein. One of these, 11-8, was selected for intensive study. Clone 11-8 (9.2kb) was digested (Fig. 2), restriction mapped, and the approximate location of the coding sequence identified by subcloning and western blot. The coding sequence was further subcloned as HindIII and KpnI fragments into pUC118 and pUC119 and sequenced in both orientations (Fig. 3). The universal primer was used to prime the sequencing reactions from the ends of the subclones, and, when necessary, primers based on earlier sequence runs were synthesized and used to extend previously determined sequences. This was necessary, since approximately 400 bases could be determined from each priming site. The complete sequencing strategy is shown in Fig. 3. A minimum of three runs in one direction and two in the reverse direction were done.

The complete nucleotide sequence of the BSCP65 gene and upstream flanking region from Brucella abortus strain19 is shown in Fig. 4. The sequence reveals a large open reading frame (ORF). Comparison of this sequence against Genbank sequences showed extensive homology to known heat shock protein 70 (Hsp70) sequences and demonstrated that the gene is an analog of E. coli dnaK. The nucleic acid sequence is 68% identical to E. coli dnaK and the predicted amino acid sequence shows 70% identity to the dnaK peptide. Immediately upstream is a small
Fig. 2: DNA gel of restriction digests of *Brucella abortus* dnaK gene (clone 11-8) stained with ethidium bromide. HindIII: H, Kpn1: K, HindIII digest of λ DNA: λ H. Molecular marker scale is given in Kb
Fig. 3: Summary of sequencing strategy for *Brucella* dnaK and dnaJ. Small arrows indicate multiple sequencing runs. Heavy arrows indicate reading frame direction from the promoter region. The restriction sites shown are: BamH1 (B), HindIII (H), Kpn1 (K), and Sma1 (S). Clone 11-8 contains sequence to the right of the Sma1 site, and clone 11-4 contains sequence to the left of the Sma1 site.
Fig. 4: *Brucella abortus* Hsp70 sequence including the upstream small ORF. The putative promoter sequence and initiation and termination codons are underlined. Probable SD sequences are matched to the corresponding *E. coli* rRNA sequence.
ORF complete with the Shine Delgarno (SD) sequence. If expressed, this ORF would produce an eleven amino acid peptide.

In *E. coli*, *dnaK* exists as an operon together with *dnaJ*, which codes for a 37 kDa protein (95). The terminal SmaI site of clone 11-8 is only 25 nucleotides from the termination codon of *dnaK*. So, the sequence of clone 11-8 does not reveal whether or not *dnaJ* follows *dnaK* in Brucella DNA. The map in Fig. 1 indicates that sequences following *dnaK* should be located near the end of clone 11-4. This clone was restriction mapped (Fig. 5), BamH1 subclones created, and a small amount of sequence determined from the 4.1 kb BamH1 subclone as indicated in Fig. 3. In Fig. 6, the sequence is compared to a portion of the published *E. coli dnaJ* sequence. Clearly there is strong homology. Furthermore, there is very good homology in a region (underlined) that is particularly highly conserved in other dnaJ genes (67).

**Expression of the *B. abortus* dnaK Gene**

The dnaK protein was only very weakly expressed from clone 11-8 as determined by western blot. So, the entire gene sequence was subcloned into the high expression plasmid pKK223-3. This plasmid is based on pBR322, contains the strong "tac" (trp-lac) promoter followed by a ribosome binding sequence followed by the pUC8 (88) polylinker sequence which in turn is followed by the strong transcription terminator *rrnB*. The EcoRI site of the polylinker is separated by five nucleotides from the ribosome binding sequence. Inspection of the N-terminal nucleotide sequence of the *Brucella dnaK* gene reveals a unique ClaI site (ATCGAT) 20 nucleotides downstream from the initiation codon (see Fig. 4). A pair of complementary oligonucleotides (oligo# 701, 702) were synthesized which would reconstruct the beginning of the gene and serve as a linker between the ClaI site of the *Brucella dnaK* gene and the EcoRI site in pKK223-3. The *Brucella* DNA from
Fig. 5: Restriction digest of 11-4-4A. (5A) Restriction digest of 11-4-4A with BamH1 (B). The 4.1 Kb fragment represents the Brucella insert in the plasmid carrying part of the dnaJ gene. (5B) Restriction digest of 11-4-4A with HindIII (H), Kpn1 (K), and EcoR1 (E). Lane 2 represents the uncut (U) plasmid.
Fig. 6: Comparison of *Brucella abortus* dnaJ sequence (top sequence) with its *E. coli* analog (bottom sequence)
clone 11-8 was then inserted into pKK223-3 at the Xma1 (Sma1) site. Fig 7A demonstrates that the plasmid yields the expected 6.1 and 5.1 Kb fragments when digested with Kpn1 and EcoR1. The resultant plasmid, 11-8PKA, was then cut with EcoR1 and Cla1 and the oligonucleotide linkers (oligo# 701, 702) inserted. The details of this cloning are given in the methods section. This construction juxtaposes the 'tac' promoter to the dnaK coding sequence. This 6.8 kb construct was named Hsp70PK and figure 7B demonstrates that when digested with Kpn1 and EcoR1, it yields the expected fragments. When placed in a lac Iq expressing host, such as JM105, the cloned gene should not be expressed unless a lactose analog such as IPTG is added to the medium. Figure 11 shows these clones in schematic form.

When the proteins produced by these clones were analyzed by western blot, bands at several molecular weights were seen with an accumulation of an unexpected but stable 23 kDa band (fig. 8, 9) when the gene expression is induced with IPTG. This 23 kDa polypeptide was the dominant feature on coomassie blue stained gels, and was transferred from SDS-PAGE gels to immobilon for amino acid sequencing. The sequence data showed that the peptide was derived from the N-terminus of the protein predicted from the DNA sequence from the Brucella HSP70 protein. Antibody against this 23 kDa fragment was made, and it reacts to the intact Brucella HSP70 protein and to the 23 kDa polypeptide. The basis of the apparent truncation or degradation of Hsp70 is unknown, but a similar sized 23 kDa polypeptide has been identified by western blot in B. abortus extracts (Z. Sha, personal communication).

Examination of the sequence (Fig. 4) revealed that the Cla1 site that was the cornerstone of our cloning strategy includes the sequence 'GATC' which should be methylated by the dam methylase in wild type E. coli. Such methylation would inhibit the action of the Cla1 enzyme. When the plasmid Hsp70PK was isolated from strain
Fig. 7: Restriction digest of *Brucella abortus* plasmids. (7A) Restriction digest of the 11.2 kb, 11-8PKA plasmid with Kpn1 (K) and EcoR1 (E). (7B) Restriction digest of plasmid Hsp70PK with Kpn1 (K) and EcoR1 (E)
Fig. 8: Western blot demonstrating expression of *Brucella abortus* proteins probed with rabbit antisera R222. (A) Expression of intact 70K protein from three randomly chosen 11-8 isolates. (B) Over expression of the 70K and 23K proteins from three random HSP70PK isolates.
Fig. 9: Western blot of *Brucella abortus* dnaK protein probed with rabbit antisera R213. (9A) Expression of the 70K protein from 3 randomly chosen 11-8 clones. (9B) Over expression of the 23K protein with clone Hsp70PK. Late log phase cells were treated with 1mM IPTG for 30 min (lane A), 45 min (lane B), 1 hr. (lane C), 2 hrs. (lane D), 3 hrs. (lane E), 4 hrs. (lane F), and 5 hrs. (lane G), respectively.
GM161 (dam⁻), the clone appeared to be about 300 base pairs smaller than when digested with EcoR1 (Fig. 10). Apparently, the initial cloning had utilized a second, previously unknown, Cla1 site about 300 base pairs upstream from the targeted site. This is illustrated in Figure 11. Sequencing confirmed that we had indeed ligated the 701 and 702 oligonucleotides to a second Cla1 site about 300 base pairs upstream from the first. The complete upstream sequence of clone Hsp70PK is given in figure 12.

Why this clone should produce a degraded or truncated gene product was not at all obvious. Clones 11-8 and 11-8PKA both expressed the intact polypeptide at low levels, whereas clone Hsp70PK produced large amounts of degraded or truncated gene product. The most likely explanation was that over expression led to degradation. Arguing against this was the observation that uninduced (the "tac" promoter is "leaky") expression (not shown) or early times after induction (fig. 9B) also showed large amounts of the 23kDa product. Other possibilities were that the plasmid Hsp70PK contains a nonsense point mutation resulting in a truncated product, or that plasmid 11-8 included an upstream gene whose product stabilized the Brucella dnaK protein.

In an effort to express the intact protein, we attempted to move the plasmid Hsp70PK to a Lon⁻ (protease deficient) E. coli strain, Y1090, which had been cured of its normal plasmid. This attempt was unsuccessful, probably because the high level of expression of the Hsp70 protein killed the host. The Y1090 strain does not carry the lacI⁹ gene and it is known that the "tac" promoter expresses constitutively under these conditions. Hsp70PK was also introduced into the dam⁻ E. coli strain GM161 so that we might create the originally intended construction from the first Cla1 site. This too resulted in very poor survival of the host bacteria. Only a few ampicillin resistant bacteria survived, and those that did formed very small colonies
Fig. 10: Restriction digest of 3 Hsp70PK isolates purified from *E. coli* GM161. GM161 does not methylate Cla1 sites. Cla1 (C) digests are about 300 bp smaller than EcoR1 (E) digests. Both enzymes should produce the full length linear form of the plasmid.
Fig. 11: Diagram summarizing the construction of Hsp70PKC
Fig. 12: Machine sequencing of clone Hsp70PK. Nucleotide 13-37 shows the sequence of the oligo used to ligate the Brucella DNA to the vector. This sequence duplicates the first twenty-four nucleotides of the Brucella Hsp70 gene. The SD sequence (AGGA), and initiation codon (ATG) of the small ORF are underlined for reference. Also underlined is the Cla1 site (ATCGAT)
even after a 24 hour incubation. One of the transformations did, however, yield the desired unmethylated Hsp70PK DNA.

**Function of dnaK Upstream Sequences**

According to our sequence data, (refer to Fig. 4), immediately upstream of the *Brucella* Hsp70 initiation codon is a 33 base pair ORF preceded by a good ribosome binding site. Current knowledge of prokaryotic gene expression would indicate that this eleven amino acid peptide should be expressed whenever the dnaK gene is expressed. The presence of such a structure upstream of a gene is very unusual. Because of its location it seemed likely that it must, somehow, be involved in the regulation of the dnaK gene expression.

In order to explore the functions of this upstream sequence several subclones were constructed in which the small ORF sequence was either present or absent. Fig. 13 is a flow diagram outlining the origin of each of the clones constructed. Fig. 14 shows the upstream sequences of each clone. Before some of the clones could be constructed, it was necessary to create an expression vector which included the lacI^q gene, so that the clones could be transferred to *E. coli* strain GM161 in order to reliably produce un-(dam)-methylated plasmid DNA. This vector is called pJE7 and its construction is described later in this thesis.

The construction of the plasmids listed in figure 13 is described in the Material and Methods section. Figures 12 and 15 document critical sequences. The relevant structural features of these clones are as follows: all the plasmids contain the entire dnaK coding sequence. Plasmids 11-8 and 11-pKA include approximately 4 kb of *Brucella* sequence upstream from the dnaK coding sequence. Plasmid Hsp70PK includes 273 base pairs (bp) of upstream sequences plus a duplication of 25 bp representing the beginning of the dnaK gene (oligonucleotides# 701 and 702).
Hb101 (11-8)  
(Hsp70 gene + flanking sequences in cloning vector)

↓

JM105 (11-8PKA)  
(Hsp70 gene + flanking sequences in expression vector)  
**expresses 70K protein**

↓

JM105 (Hsp70PK)  
(includes 273 bp of upstream seq.)  
**expresses the 23kDa polypeptide**

↓

JM105 (Hsp70PK11)  
(includes 109 bp of upstream seq)  
**expresses the 70K protein**

↓

GM161 (Hsp70JE7C)  
(no upstream seq.)  
**expresses the 23kDa polypeptide**

GM161 (Hsp70PKC)  
(no upstream seq.)  
**expresses the 23kDa polypeptide**

JM105 (Hsp70PKO)  
(includes 273 bp upstream seq.)  
**expresses the 70K protein**

Fig. 13: Flow diagram outlining the origin of each of the *Brucella abortus* dnaK clones constructed in this study
Fig. 14: Upstream sequences of the various Brucella Hsp70 deletion clones. The bold print indicates the duplication of the beginning of the dnaK gene. Restriction sites and start codons are underlined and the ribosome binding sites are in italics. Clone Hsp70JE7C is not listed since it has the same sequence as Hsp70PKC.
Fig. 15: Sequencing of clone Hsp70PK11. Nucleotide 13 to 28 shows the sequence of the oligo used to ligate the Brucella DNA to the vector. The SD sequence (AGGA), initiation (ATG), and termination (TGA) codons of the small ORF are underlined. The second initiation codon underlined indicates the beginning of the HSP70 coding sequence. Also underlined is the methylated Cla1 site (ATCGAT) within the HSP70 gene.
Plasmid Hsp70PKO includes the same 273 bp of upstream sequence as Hsp70PK, but does not include the duplication. Plasmid Hsp70PK11 includes 109 bp of upstream sequence (Fig. 16). Plasmid Hsp70PKC and Hsp70JE7C contain no upstream Brucella sequences. All the plasmids except Hsp70PKC and Hsp70JE7C include the complete small ORF and its associated ribosome binding sequence.

Surprisingly, when expression of dnaK from these plasmids in E. coli was examined by SDS-PAGE, a curious pattern emerged which implies that the small ORF plays a critical role in the stability of the dnaK protein (see Figs. 17 and 18). As indicated in figure 13, Hsp70PKC and Hsp70JE7C (no upstream sequences) produce the truncated 23kDa polypeptide. Hsp70PK11 (109 bp of upstream sequence), Hsp70PKO (273 bp of upstream sequence), and 11-8PKA and 11-8 all produce the full length polypeptide. These results suggest that some sequence within the first 109 bp upstream is necessary for dnaK protein stability. The obvious structure within this sequence is the small ORF. Why Hsp70PK yields the truncated polypeptide is a mystery. Apparently, the 25 bp duplication somehow negates the effect of the small ORF.

Part II: Construction of Expression Vectors

In order to carry out the previous project and to support several other projects in the lab, it was desirable to construct two different types of expression vectors. To construct these vectors we modified the commercially available regulated expression plasmid pKK223-3 constructed by Brosius and co-workers (16, 100). The first type is an improved regulated vector for high expression of heterologous proteins. This vector carries the lacIq gene to control over-expression in various hosts. The second type of expression plasmid contains artificial unregulated promoters of various
Fig. 16: Restriction digest of *Brucella* clones. Lanes labeled 1 represent restriction digest of clone Hsp70PK11 and those labeled 2 represent restriction digest of Hsp70PKC. Both clones were digested with EcoRI (E), BamH1 (B), or double digested with EcoRI and BamH1 (R+B). The size differences seen in the lower band indicate that the insert Hsp70PKC is about 110 base pairs smaller than Hsp70PK11.
Fig. 17: Western blot of *E. coli* clones containing various constructs which express *Brucella* dnaK (refer to Fig. 13). Total proteins of late log phase cultures were challenged with 2 mM IPTG for 5 hours. The blot was probed with rabbit antisera R222

<table>
<thead>
<tr>
<th>Lane</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11-8</td>
</tr>
<tr>
<td>2</td>
<td>11-8PKA</td>
</tr>
<tr>
<td>3</td>
<td>Hsp70PK</td>
</tr>
<tr>
<td>4</td>
<td>Hsp70PKC</td>
</tr>
<tr>
<td>5</td>
<td>Hsp70PKO</td>
</tr>
<tr>
<td>6</td>
<td>Hsp70PK11</td>
</tr>
<tr>
<td>7</td>
<td>TG1 (host cell)</td>
</tr>
<tr>
<td>8</td>
<td>PKK223-3 (vector)</td>
</tr>
<tr>
<td>9</td>
<td>pJE7 (vector)</td>
</tr>
<tr>
<td>10</td>
<td>Hsp70JE7</td>
</tr>
<tr>
<td>11</td>
<td>Hsp70JE7C</td>
</tr>
</tbody>
</table>
Fig. 18: Western blot of *E. coli* clones containing various constructs which expresses *Brucella* dnaK (refer to Fig. 13). Total proteins of late log phase cultures were challenged with 2mM IPTG for 5 hours. The blot was probed with monoclonal antibody against *E. coli* dnaK protein (Stress Gene incorp.)

Lane 1: Hsp70PK11  
Lane 2: Hsp70PKO  
Lane 3: Hsp70PKC  
Lane 4: Hsp70PK  
Lane 5: 11-8PKA  
Lane 6: 11-8  
Lane 7: Hsp70JE7C  
Lane 8: Hsp70JE7C
strengths followed by a polycloning sequence and the rnb terminator. We created six different promoters.

**Construction of Regulated Expression Vector (pJE7)**

This plasmid is based on pKK223-3 and the construction is summarized in Figure 19. The first step was to cut bacteriophage M13 RF DNA with HgiA1 and Aha111 to excise the origin of replication, this was blunt ended with S1 nuclease, and ligated, as described by Maniatis (47), into the Pvu II site of pKK223-3. The plasmid was then digested with HindIII and EcoR1 to remove the polycloning sequence, and a HindIII/EcoR1 adapter oligonucleotide ligated in. The plasmid was then cut with BamH1 and Sal1 and a Kpn1/Nsi1 adapter with BamH1 and Sal1 compatible ends was ligated to the opened plasmid (see Fig. 19 for the sequence of the adapter). The lacIq gene from a commercially (Amersham) available plasmid pMJR1560 (78), was excised with Kpnl and Pst1, and ligated into the Kpnl and Nsi1 sites (Nsi1 and Pst1 create compatible ends). The reconstructed plasmid was then cut with HindIII and EcoR1 and the polycloning sequence from pUC118 inserted. The procedure had the effect of destroying the extra Sal1 and BamH1 sites in the vector and introducing a new Kpn1 site.

**Construction of Unregulated Plasmids (pJE1-pJE6)**

These plasmids were also based on pKK223-3. The promoter sequences were created by oligonucleotide synthesis. The plasmid construction is summarized in Figure 20. The M13 origin of replication was inserted into pKK223-3 in the same manner as for pJE7, described above. Then the plasmid was cut with Hind111 and EcoR1 to remove the existing polycloning sequence. The Hind111 and EcoR1 sites were then preserved by inserting a Hind111/EcoR1 oligonucleotide adapter. The
Fig. 19: Summary of construction of regulated expression plasmid pJE7
Fig. 20: Summary of construction of unregulated expression plasmids pJE1-pJE6
plasmid was cut with SalI and BamHI and a "-35" double stranded oligonucleotide ligated in (oligo# 1, 2, or 3 in Table 4). These oligonucleotides all have a "sticky" ends which when ligated to the SalI site in the plasmid destroy the SalI site. The next step involved cutting the three plasmids with BamHI and EcoRI and a "-10" double stranded oligonucleotide ligated into place (oligo# 4 or 5 in Table 4).

Finally, each plasmid was cut with EcoR1 and HindIII and the polylinker sequence from pUC118 inserted. Each step of the construction was checked by single strand DNA sequencing through the promoter region to be certain unexpected changes did not occur. The plasmids were named pJE1, pJE2, pJE3, pJE4, pJE5, and pJE6. The expected phenotypes are shown in Table 5.

Testing the Plasmids in Bacteria

The two plasmid types were tested by inserting the BCSP31 gene at the HindIII site of the multiple cloning site of each plasmid. This gene and gene product have been well characterized (48, 77), the protein seems to be very stable in E. coli and specific polyclonal antisera against the protein is available. Figure 21 demonstrate that the gene sequence is present in each plasmid. Controlled expression from pJE7 was tested by growing E. coli in the presence and absence of inducer, IPTG. Protein expression was determined by western blot (Fig. 22) (85). Single strand synthesis was tested by transforming the plasmid into E. coli JM105 and co-infecting with M13 helper phage K07. Single stranded DNA was isolated and sequenced. Qualitatively, the plasmids function as expected.
Table 4: Oligonucleotides used in the construction of pJE1 - pJE6

<table>
<thead>
<tr>
<th>seq#</th>
<th>-35 oligo sequence</th>
<th>-10 oligo sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TCGATTGACAATCGGTAG</td>
<td>AACTGTTAGCCATCCTAG</td>
</tr>
<tr>
<td>2</td>
<td>TCGATTACAATCGGTAG</td>
<td>AAATGTTAGCCATCCTAG</td>
</tr>
<tr>
<td>3</td>
<td>TCGATTTAAAATCGGTAG</td>
<td>AAATTTTAGCCATCCTAG</td>
</tr>
<tr>
<td>4</td>
<td>GATCGCTCGTATGATGTTCGCATG</td>
<td>CGAGCATACCTACAGCGTACTTAA</td>
</tr>
<tr>
<td>5</td>
<td>GATCGCTCGTATGGTGTTCGCATG</td>
<td>CGAGCATACCACAAGCGTACTTAA</td>
</tr>
</tbody>
</table>

Table 5: Expected phenotypes of pJE1 - pJE6

<table>
<thead>
<tr>
<th>name</th>
<th>-35 oligo</th>
<th>-10 oligo</th>
<th>expected expression phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>pJE1</td>
<td>seq#1</td>
<td>seq#4</td>
<td>stronger than BCSP31</td>
</tr>
<tr>
<td>pJE2</td>
<td>seq#1</td>
<td>seq#5</td>
<td>same as BCSP31</td>
</tr>
<tr>
<td>pJE3</td>
<td>seq#2</td>
<td>seq#4</td>
<td>somewhat weaker than BCSP31</td>
</tr>
<tr>
<td>pJE4</td>
<td>seq#2</td>
<td>seq#5</td>
<td>much weaker than BCSP31</td>
</tr>
<tr>
<td>pJE5</td>
<td>seq#3</td>
<td>seq#4</td>
<td>still weaker</td>
</tr>
<tr>
<td>pJE6</td>
<td>seq#3</td>
<td>seq#5</td>
<td>about 1% of BCSP31</td>
</tr>
</tbody>
</table>
Fig. 21: HindIII digests of the various expression vectors with the BCSP31 sequence inserted at the polycloning site. Lane 1: the plasmid based on pJE7. Lane 2-7: the plasmids based on pJE1-6, respectively. The correct sized fragment is released from each plasmid.
Fig. 22: Expression of *Brucella abortus* 31K protein. (22A) Western blot probing for the expression of the 31K from the regulated plasmid pJE7 with antiserum R217. Lane A represents expression in the presence IPTG and in lane B, IPTG is absent. Despite the absence of IPTG, there is a low level of expression indicating the "tac" promoter is not very tightly regulated. Lane C is the control lane representing *E. coli* without the 31K gene. (22B) Western blot probing for the expression of the 31K protein with antiserum R217. LaneA through F represents progressively increased expression from the unregulated plasmid pJE6 through pJE1 respectively.
DISCUSSION

Studies of the Brucella abortus HSP70 Gene in E. coli

The heat shock response has been intensively investigated by many laboratories around the world for more than a decade. This continuous interest reflects the importance of these protein families, their universality, and the conservation of their function and structure. These studies have provided insight to pathways of protein secretion, folding, stabilization and degradation. At the present time, heat shock proteins represent an exciting opportunity for investigating the bacterial response to the stresses of infection, and they also may represent useful antigens for development of subunit vaccines. Investigation of the molecular biology of the dnaK protein from various bacteria may provide new insight into the protein's function, may lead to new and unexpected strategies for combating bacterial infection, and will certainly increase our knowledge of the biology of the protein.

The cloning and expression of the Brucella dnaK gene in E. coli provides us the opportunity to study the biological properties of the protein independent of other contaminating Brucella macromolecules, and provides a convenient source for the further characterization of the protein, a convenient means for studying it's regulation, and makes possible future testing of its immunogenicity in animals. This report describes the sequencing and expression of the gene encoding Brucella abortus Hsp70 (dnaK). The nucleotide sequence is given in Fig. 4 along with the predicted amino acid sequence.

In E. coli, DnaK exists together with DnaJ in an operon. The dnaJ protein is highly conserved amongst widely different organisms. The most conserved sequences characteristically include four tandem repeats of a motif consisting of Cys-x-x-Cys-x-Gly-x-Gly (67). Partial sequencing of the DNA downstream of the
**Brucella** dnaK gene demonstrated an open reading frame that shows high homology to the *E. coli* DnaJ gene and includes one copy of this highly conserved repeat motif. Because of this, we conclude that the **Brucella** dnaJ gene homologue exists also in an operon with dnaK gene.

In *E. coli* although the DnaK gene product is produced abundantly, the amount of the DnaJ protein produced is substantially less. Until very recently, information on the DnaJ protein has been very limited, but recent publications (21, 44) show, that in *E. coli*, DnaJ gene interacts with DnaK gene and is involved with both the expression and regulation of DnaK. The sequence conservation and the similarity of the organization of the DnaK-J operon in *E. coli* and *B. abortus* suggests that there may be similar interactions between the two gene products in *B. abortus* as well making an understanding of the expression of **Brucella** dnaJ necessary for a complete understanding of the expression and regulation of dnaK.

Initially, study of the protein expressed from the 11-8, and other similar plasmid clones in recombinant *E. coli* cells seemed to be a very difficult undertaking since the protein was expressed at too low a level to be visible on Laemmli gels. To facilitate characterization of the protein, we have replaced the natural promoter to achieve a higher level of expression. Since we did not know whether or not constitutive high expression of the protein would be deleterious, we juxtaposed a regulated high expression promoter to the coding sequence. The artificial "tac" promoter in the plasmid pKK223-3 seemed well suited for this (100). The main drawback was that pKK223-3 must be transformed into a host that carries the lacΩ gene to prevent over-expression from the 'tac' promoter.

High expression of the gene was apparently achieved by this construction, but the protein was degraded, leading to the accumulation of a 23 kDa polypeptide. Amino acid sequencing showed that this peptide was derived from the N-terminus of
the predicted mature protein. DnaK proteins from a variety of other organisms have been overproduced in *E. coli* and no other similar problems with degradation have been reported.

The molecular principles that govern the regulation of expression of heat shock genes are only partially understood. The very fact that the source of induction of HSP synthesis can vary greatly (e.g., heat, foreign protein, toxic metals, macrophage ingestion, etc.) suggests that these genes may be under many different transcriptional and post-transcriptional controls. The unusual small open reading frame upstream of the *dnaK* coding sequence (see Fig. 4) and extensive secondary structure predicted for the mRNA in this region (see below) suggested to us that some unusual type of regulation of gene expression was occurring in *Brucella*. The initial question we addressed was: are there regulatory sequences upstream from the *dnaK* coding sequence which function in *E. coli*? Deletion experiments then showed that upstream DNA sequences play a critical role in determining whether or not the full length stable Hsp70 is produced by the artificial constructions. Immediately upstream of the initiator ATG is a small open reading frame (ORF) which is preceded by a good ribosome binding sequence (see Fig. 4). Translation of this ORF would lead to the expression of an eleven amino acid peptide which is clearly amphipathic if structured as an alpha helix (Fig. 23). Computer calculations predict that an RNA molecule including this sequence region would exhibit extensive secondary structure. In particular, predicted stem-loop structures would seem to tie up the ribosome binding sequence of the small ORF but not the ribosome binding sequence required for *dnaK* expression (Fig. 24).

The meaning of these observations is unclear. We have only looked at one aspect of the regulation of the gene expression, the presence of an upstream regulatory sequence of *Brucella dnaK*. The present work demonstrated that the
Fig. 23: Computer projection of an amphipathic helix formed by the eleven amino acid peptide coded by the ORF. Hydrophobic amino acids are boxed.

\( \square = \text{Hydrophobic} \)
Fig. 24: Secondary structure of *Brucella abortus* Hsp70 mRNA as drawn by the computer program SQUIGGLE
presence of this ORF is necessary for the production of the intact 70kDa protein. Our preferred interpretation of this data is that the small ORF is important for the stability of the protein and that subtle changes in RNA secondary structure have large effects on its expression.

The next step would be to understand how the presence of this ORF regulates gene expression. DnaK is a complicated and multifunctioned protein, to fully appreciate the regulation of expression of the protein will require further understanding of the transcriptional, translational and post translational controls with closer analysis of additional cis and trans acting elements. The rpoH gene product, sigma32, appears to be one of the main factors that positively regulates the transcriptional level of the heat shock response in \textit{E. coli}. So, to further understand the regulation of \textit{Brucella dnaK} expression, the demonstration of the possible presence and an understanding of the action of a \textit{Brucella rpoH} like gene would be necessary. Other important gene products may include DnaJ and a number of proteases.

The small upstream ORF seems to be unique, an extensive literature search has not revealed another similar phenomenon. The regulation of expression of the ORF itself could be rather complicated and the expression of this sequence in turn seems to affect the regulation of the \textit{Brucella abortus} HSP70 protein. The ORF could act in cis or trans, and control the expression at the DNA, RNA, or peptide level. One attractive possibility is that the small upstream ORF may code for a peptide that binds to the dnaK protein itself protecting it from cleavage and subsequent degradation by an existing protease. Alternatively, the peptide could be a protease inhibitor. The peptide might compete for a binding site on the protease, or some other macromolecule, inhibiting binding to the HSP70 protein and thus allowing the expression of the full length HSP70 protein.
The ORF may also function at the RNA level, binding or assisting in the formation of secondary structure of the dnaK coding sequence. It may also act as an antisense RNA by interacting with a region in the middle of the DnaK gene and modulating the expression of the full length 70 kDa or the 23 kDa fragment. Naturally occurring antisense regulation occurs in bacteria and it has been shown that antisense regulation often is not absolute. We have observed the expression of the 23 kDa peptide does not always preclude the expression of the 70 kDa protein. In fact, as the level of expression of the 23 kDa peptide increases, there is a slight corresponding increase of the expression of the 70 kDa protein. These are but a few of the possible mechanisms. The ORF sequence itself could directly cause premature termination of either transcription or translation of the dnaK coding sequence by an unknown mechanism. Further work is necessary to elucidate the regulation of expression of this ORF and the dnaK gene.

A possibility which deserves consideration is that the 23 kDa polypeptide has a biological function. It has been shown that sequences in the N-terminal two-thirds of the mammalian Hsp70 are necessary for ATP binding and dispensable for localization (41). The three dimensional X-ray structure of an N-terminal fragment of bovine HSC70 has recently been solved at a resolution of 2.2 angstrom (18), revealing that the ATPase domain consists of two domains with the nucleotide bound at the base of a deep cleft between them. The size of the 23 kDa polypeptide matches exactly the size of the first lobe if the structure of the human HSC 70 protein is extrapolated to *Brucella* HSP70. Sequences in the C-terminal third of the human protein have been shown necessary for localization and dispensable for ATP binding. In one study, it was shown that cultured cells expressing a mutant human HSP70 from which the ATP binding region has been deleted remain fully heat resistant (52). This surprising finding is potentially important because it contradicts
the widely accepted notion that ATP binding and hydrolysis modulate interactions between HSP70 and its targets, including nascent polypeptides and denatured proteins (64). Furthermore, Li et al. (43) removed amino acids 120-428 from human HSP70. This deletes nearly the middle third of the protein including the ATP binding site. This also does not affect the protection of the cell from heat shock implying that ATP binding is unnecessary for its thermal protective function. The binding of ATP to HSP70 has been suggested to affect its interaction with cellular proteins, which may, in turn prevent detrimental aggregation of these proteins. ATP hydrolysis presumably allows recycling of HSP70 by dissociating it from its targets. Alternatively, ATP binding and/or hydrolysis by HSP70 may facilitate dissociation of protein aggregates already formed. Perhaps defective HSP70s lacking the ATP-binding domain can still bind to cellular proteins and prevent their aggregation at elevated temperatures (43). The results described in these previous studies do not give a clear picture of the actual function of the ATP binding domain of the protein. It is possible that the 23 kDa polypeptide retains ATP binding capability and has some sort of biological activity.

Work has been done which shows that antibodies to foreign HSP70 proteins are preferentially directed towards the variable carboxyl terminal regions of the protein (15). Monoclonal antibodies raised against either *M. bovis* or *M. leprae* were observed to react with the C-terminal fragments but not the N-terminal fragments. Sera from humans with Lepromatous leprosy sera also reacted with the C-terminal fragments rather than with the N-terminal fragments of HSP70 in immunoblots (15). In addition, murine polyclonal anti-HSP70 antisera showed preferential binding to the C-terminal fragments. Therefore the C-terminal portion of *M. leprae* HSP70, which includes the region of maximum divergence from human HSP70 (50), is the major target for the humoral immune response to the protein (15). The divergence is
greatest in the C-terminal 100 amino acids. The sequence conservation in the N-terminal portions of HSP70s from various species may be necessary to preserve the ATPase function of the molecule (18). All these studies suggest that the C-terminal region of the protein is the major target of the human and murine immune responses and the N-terminal, ATP binding, region is highly conserved and therefore must have some very important cellular functions.

The significance of the accumulation of the 23 kDa HSP70 fragment is unclear. The production of the fragment could simply be the natural and accidental consequence of the degradation of a highly abundant protein that has outlived its function. On the other hand, its accumulation in *E. coli* suggests that this 23 kDa polypeptide is very stable in the cell. Since *Brucella* is an intracellular parasite with the capability of survival in the hostile environment of the macrophage, expression of its own HSPs is probably critical. At the same time, a high expression of *dnaK* could possibly induce a strong immune response from its host. By producing or retaining only the highly conserved N-terminal domain of the protein which may retain some biological activity, it might be able to evade or slow down the host's immune response to this important and abundant protein. As mentioned previously others have shown that monoclonal antibodies raised against *Mycobacterium* react with the C-terminal but not the N-terminal portion of the *Mycobacterium* HSP70 protein (13). It is interesting that *Mycobacterium* are also intracellular parasites that are capable of colonizing and multiplying in the macrophage in a similar manner as *Brucella*. Construction of deletion mutants of the *dnaK* gene and the study of the pathogenicity of the mutant bacteria might shed some light on these host/parasite relationships. While heat shock proteins are not likely to be the unique key to understanding all aspects of virulence, it is probable that further study of the function and regulation of these ubiquitous macromolecules will provide a means to dissect some of the
molecular mechanisms involved in the complex interactions between microbial parasites and their mammalian hosts.

Construction of the Regulated Expression Vector

The usefulness of pKK223-3 is severely limited because the "tac" promoter becomes active any time the plasmid is introduced into a strain which does not overproduce the lac repressor. This was demonstrated by the poor survival of the strains Y1090 and GM161 which lack the lacI gene when transformed with the high expression plasmids. We have also observed that genes cloned downstream from the tac promoter may be expressed at undesirable and significant levels even in lacI hosts in the absence of the inducer IPTG. In pKK223-3, it seemed therefore, that the only way to effectively obtain a high level of expression of Brucella dnaK outside a lacI host was to have the plasmid carry its own lacI gene to assure that expression from the "tac" promoter is regulated. The lacI gene from the plasmid pMJR1560 constitutively produces lac repressor at a rate about 10 times higher than its normal counterpart in E. coli (note the difference in expression of the 23 kDa peptide in fig.18). Other problems with pKK223-3 are that it is not useful for production of single strand DNA for sequencing, and cloning into pKK223-3 is quite restricted since it only has a small polycloning site and two of the restriction sites, BamH1 and Sal1, are also present outside the polycloning site rendering them almost useless for cloning. All of these problems were solved by re-engineering the plasmid. We substituted a longer polycloning sequence, destroyed the existing BamH1 and Sal1 sites outside the polycloning sequence, and introduced both the lacI gene and the M13 origin of replication into the plasmid. The new plasmid, pJE7, should aid in the in vitro production and characterization of many proteins.
Construction of a Series of Unregulated Expression Vectors

For the past few years our lab has been studying the pathogenicity of Brucella abortus, its ability to survive in the macrophage, and in vivo expression of Brucella proteins in live carrier vaccine experiments. To serve as a carrier for immunological and vaccine experiments it would be convenient if the bacterium constitutively express the experimental protein over days or even weeks in the host animal. The foreign gene must be stably expressed and not be deleterious to the bacterium. A major problem which has been encountered in a number of laboratories is establishment of the proper level of expression of the heterologous antigens. Experience thus far indicates that this problem must be handled differently for each antigen investigated, and possibly for each carrier strain being tested, a laborious and time consuming undertaking. For most proteins, it may be necessary to try different levels of expression to find the optimal balance between plasmid stability (including carrier survival in the host) and a strong immune stimulation. Over expression is often deleterious and under expression may lead to insufficient immunity.

It is clear from the work of many investigators, that the rate of transcription of a gene in E. coli is determined by a pair of short DNA sequences known collectively as the promoter. These sequences are always located upstream from the gene. The relative strength of the promoter is frequently but not always regulated by the binding of either, or both, negative and positive regulatory proteins. Even in the absence of regulatory protein function, promoters vary in strength approximately 10,000 fold depending on their sequences (29). Analysis of more than 100 native promoter sequences and more than fifty promoter mutations has lead to a rather clear picture of promoter requirements in E. coli (29, 73). The strongest possible E. coli promoter probably consists of the sequence TTGACA (the -35 sequence) followed in 17 base
pairs by the sequence TATAAT (the -10 sequence). Variations in the surrounding nucleotide sequence seem to have only a limited influence on this activity. The separation of the two sequences can vary from 15 to 21 base pairs, but each deviation of one base pair from the ideal of 17 reduces the activity of the promoter by about ten fold (29). Likewise, substitution of any of the ideal promoter base pairs by another can also reduce the strength of the promoter by as much as one hundred fold, although the effect is usually less. Using in vitro oligonucleotide synthesis it is possible to synthesize a promoter of any desired strength (up to a maximum). Indeed, a number of cloning vectors have been constructed which contain artificial promoters. Unfortunately, the emphasis has been on the construction of vectors which utilize very strong regulated promoters. Such vectors are very useful for producing proteins in vitro, but not for the constitutive long term synthesis of proteins inside an animal which will be required for live carrier vaccines.

We have constructed six expression plasmids which contain unregulated promoters of different strengths followed by a polycloning site, allowing insertion of cloned DNA sequences downstream of each unregulated artificial promoter. Each of these plasmid constructions has been tested for deleterious effect on E. coli with BSCP31, and plasmid stability measured when the gene is expressed at a high level for an extended period of time. The promoters were based on the B. abortus BCSP31 promoter, and therefore are likely to function in B. abortus also. Since they contain no flanking Brucella sequences, they are unlikely to be regulated even in Brucella. Each differs from one of the others by a single nucleotide substitution, and based on the known rules of E. coli promoters, have been designed to differ in strength, in E. coli, by a maximum of about 1000 fold. Each promoter was constructed in two parts. This allowed for the synthesis of smaller oligonucleotides. It also permitted some savings by allowing mixed matching of oligonucleotides, and
provided greater flexibility in case additional sequences needed to be constructed in the future. We chose not to simply copy known promoters from nature (except for the BCSP31 promoter) in order to contribute new data to what is known about the effect of individual base pair changes on promoter strength. Note, that the design of these plasmids requires that each gene to be expressed includes a ribosome binding site (Shine-Dalgarno) sequence. This design allows much greater flexibility in juxtaposing various genes upstream from the promoter since hundreds of base pairs may separate the initiation codon from the promoter rather than the five or six required if the SD sequence is provided with the vector.

We, have therefore, created a series of unregulated expression vectors which will constitutively express heterologous proteins in strains of *E. coli* and *Salmonella* over a broad range of levels in an effort to facilitate the identification of optimal *in vivo* expression for particular experiments. These plasmids allow us to examine the amount of antigen necessary to achieve a desired level of immunity. Present procedures for doing this are time consuming and must be tailored to each protein antigen being investigated. These graded expression plasmids may serve as tools needed to streamline this procedure such that future testing would be accelerated and simplified.

It has been demonstrated in *Mycobacterium*, that both HSP70 and HSP60 are dominant antigens which can induce strong immune responses (94, 96). It should be feasible, especially when cloned and expressed from the graded expression vectors that we have constructed, to study the use of these proteins in the live carrier vaccine concept.
LITERATURE CITED


40. Lesley, S., N. Thompson and R. Burgess. 1987. Studies of the role of the *E. coli* heat shock regulatory proteins sigma 32 by the use if monoclonal


of mutations induced in transfected DNA by mammalian cells. EMBO J. 3:3117-3121.


Gene and Develop. 3:2003-2010.


100. The complete nucleotide sequence is available from Pharmacia, Inc., Milwaukee, WI.
ACKNOWLEDGEMENTS

This dissertation is dedicated to my family in Hong Kong and especially to my late brother Dr. Simon Chin.

The author wishes to acknowledge Dr. John E. Mayfield for his teaching and guidance during the course of this work. I also want to thank my friend Dr. Stephen Smith for all his kind encouragement and assistance. Finally, my warmest appreciation to my sons Vachel and Emile as well as my special friend Dr. Howie Smith for all their patience, perseverance, encouragement and good-natured acceptance of my time away from them.