Chapter 1.

Introduction
1.1. TUBERCULOSIS: A Problem with Persistence

Tuberculosis – A global problem

Approximately eight million people develop active tuberculosis (TB) every year, with two million dying from the disease. In addition to this already huge burden of disease, it is estimated that up to two billion people have been infected with the causative agent, *Mycobacterium tuberculosis* (34). Most people control the initial infection by mounting a cell-mediated immune response that prevents disease but can leave a residual population of viable mycobacteria. Between 5–10% of individuals who become infected subsequently develop clinical disease (7). Primary TB develops within 1 or 2 years after an initial infection and, particularly in children, is often associated with disseminated disease. Post-primary TB develops later in life, and can be caused either by reactivation of bacteria remaining from the initial infection or by failure to control a subsequent reinfection. Post-primary TB is predominantly a pulmonary disease, involving extensive damage to the lungs and efficient aerosol transmission of bacteria (Fig. 1). The risk of disease is highly dependent on the immune status of the host; coinfection with HIV markedly increases the incidence of both forms of disease (102).

Establishing infection

To establish an infection, *M. tuberculosis* has to penetrate deep into the alveoli of the lung where they are engulfed by macrophages (Fig. 1). The thick mycobacterial cell wall provides a degree of resistance against the microbicidal mechanisms of macrophages, which is complemented by the expression of a range of enzymes that can detoxify oxidative radicals. In addition, in common with other slow-growing mycobacteria, live *M. tuberculosis* interfere with intracellular trafficking events after uptake by phagocytes, allowing the bacteria to occupy an immature phagosomal compartment that is screened from the most potent antimicrobial armoury (93). These properties allow *M. tuberculosis* to replicate inside macrophages during the early stages of infection. The ability of mycobacteria to interfere with phagosome maturation is reduced in macrophages that have been activated by interferon-γ (IFN-γ) (128).
Mycobacteria, or mycobacterial fragments, are also taken up by local dendritic cells that migrate to the lymph nodes, where they prime antigen-specific T cells, which return to the lung, and strengthen antibacterial responses by the recruitment and activation of additional macrophages (56). In most individuals, this interplay between the innate and adaptive arms of the immune response serves to control the infection with no obvious clinical symptoms. The site of the infection can often be recognized as a granulomatous lesion or tubercle, which comprises a central area of mycobacterial-infected cells surrounded by other, non-infected phagocytic cells and foamy giant cells, with lymphocytes found at the periphery. The lesion is sealed off from surrounding tissue by a fibrotic capsule.

**Evidence for persistent infection**

Infection leaves an imprint on the T-cell repertoire, which can be detected by a positive skin-test response to tuberculin (a crude mixture of *M. tuberculosis* antigens), and which provides the experimental basis for estimates of global infection rates. More recent assays measuring the production of IFN-γ by peripheral blood lymphocytes that have been exposed to specific *M. tuberculosis* antigens in culture provide a more accurate test—allowing priming by *M. tuberculosis* to be distinguished from *Mycobacterium bovis* bacille Calmette–Guérin (BCG) vaccination, for example — but indicate a similarly high rate of infection in endemic populations (35). Immunological tests are a useful indication of the history of infection, but additional evidence is required to determine whether the immune response has resulted in elimination of the pathogen or in establishment of a persistent infection. There has been prolonged debate about the extent to which reactivation disease contributes to the overall
incidence of TB. In settings where there is a low risk of reinfection, an interval between tuberculin conversion and clinical disease shows that an initial infection can leave a population of viable bacteria capable of acting as a source of reactivation disease years or decades in the future (Fig. 2).

**Figure 2.** Compared with many other diseases, the timescales involved in TB are long and there is large variation between different individuals. Adopted from (137). Bacteria that reach the alveoli of the lung are ingested by macrophages, where they can initiate rounds of intracellular replication and cell lysis. Macrophages are key effector cells in mycobacterial killing, but can also provide a niche for bacterial multiplication. Dendritic cells engulf bacteria, or bacterial components, circulate to the draining lymph nodes and prime T cells, which then return to the lungs to orchestrate control of the infection. T cells enhance the antibacterial activity of macrophages by releasing cytokines, such as IFN-γ, which generally results in arrest or clearance of the infection. If the T-cell response is insufficient to control the initial infection, clinical symptoms will develop within 1 year in the form of primary progressive disease. Prior vaccination with BCG establishes a primed population of T cells and reduces severe primary disease in children. Most individuals develop a T-cell response in the absence of any clinical symptoms, which is defined as a latent infection and carries a risk of secondary disease owing to subsequent reinfection or reactivation of the initial infection. Autopsy studies show that latent infection is often associated with persistence of viable bacteria. Bacteria can persist within granulomas that function to contain bacterial spread. In adult pulmonary TB, breakdown of granulomas in the lung promotes mycobacterial replication, release of bacteria into the airways and effective aerosol transmission. Transmission is enhanced by the destruction of lung tissue, which is mediated by the same immune cells that are crucial for protection during the earlier stages of infection.

**Persistent bacteria**

The question of how bacteria survive for decades in immunologically educated hosts without causing disease has puzzled microbiologists for a century. One attractive solution envisages the formation of an inert spore-like form that is capable of surviving in harsh environments. Small, spherical non-acid-fast forms of the tubercle bacillus have been reported by several researchers, most famously Hans Much and Albert Calmette (15). However, this hypothesis has not been supported by molecular genetic analysis. Analysis of the *M. tuberculosis* genome (22) has not shown acquisition of a dedicated developmental pathway that is comparable to spore formation, and a less-distinctive form of adaptive response seems a more likely option. Two important observations have been made on the replication status of persistent bacteria in the mouse plateau model. Rees and D'Arcy Hart compared total microscopic counts of
bacteria in infected lungs with counts of viable organisms (87), reasoning that if the constant bacterial load reflected active replication in steady-state equilibrium with immune killing, total counts would increase over viable counts as dead bacteria accumulated. In fact, the total count was similar to the viable count throughout the experiment, indicating that bacteria were replicating either very slowly or not at all. Reduction in bacterial replication is also associated with a reduction in drug susceptibility. Drugs that target processes involved in cell division and cell-wall synthesis have a reduced effect on non-replicating cultures, and the induction of stress-response defences and a decrease in metabolic rate tend to increase tolerance to drug action. Transcriptional regulation in response to environmental changes encountered during infection is a common theme in bacterial pathogenesis, and similar concepts can be applied to persistent *M. tuberculosis*. The work of Larry Wayne has provided an important stimulus in this area (135). He reasoned that survival in poorly oxygenated lesions was likely to be an important element of persistence, encouraging several groups to investigate non-replicating persistence of mycobacteria in a low-oxygen environment. An important metabolic adaptation to hypoxic survival involves a switch to metabolism of lipid substrates using the enzymes isocitrate lyase and malate dehydrogenase to bypass the complete tricarboxylic-acid cycle by combining acetyl CoA with oxaloacetate by the glyoxylate shunt (134). A similar adaptation is observed in activated macrophages, and a mutant strain of *M. tuberculosis* H37Rv lacking isocitrate lyase was characterized by a defect in the persistent phase of infection in the C57BL/6 model (71). This indicates that the persistent bacteria use the abundant fatty acids of the host cells as a carbon source. Detailed analysis of the response of *M. tuberculosis* to Hypoxia using microarray-based whole-genome profiling has identified an important response regulon (80, 104). This includes a two-component regulator (DosR/DosS) that is responsible for the control of a set of genes, which include a prominent protein antigen, the α-crystallin chaperone. Increased expression of α-crystallin is seen under microaerophilic conditions, in stationary-phase cultures and in the C57BL/6 model (138). An important feature of the hypoxia regulon is that it can also be activated by exposure to nitric oxide (130). Hypoxia and nitric oxide are key elements of the inflammatory response, and an ability to respond to changes in these signals could allow *M. tuberculosis* to tune changes in phenotypic properties to changes in host immunity.
1.2. Stress Response in Mycobacteria

Hypoxic stress

Persistence phase of *M. tuberculosis* growth is particularly a major problem since it is difficult to cure. Most of the drugs are effective against actively growing Mycobacteria but not the persistent ones. The mechanism by which the persistent state is induced is unclear at present although it is generally accepted, that a variety of stress conditions trigger persistence. One such stress condition that has been the subject of intense study is Hypoxia or lack of Oxygen. Most of our knowledge regarding the molecular biology of the persistent state is derived from studies using the *in vitro* model developed by Wayne (132, 133). In this model a persistent or non replicating state is induced by the gradual deprivation of oxygen which is achieved by growing the cells in a stoppered tube, with little or no head space.

Global transcriptional profiling performed using Miroarray based techniques as well as 2D gel electrophoresis methods revealed that the gene expression profile changed remarkably as the bacteria was exposed to oxygen limiting conditions. One of the genes that is found to be reproducibly induced at a high level under these conditions is *acr*. This gene encodes a α-crystallin homolog (Acr) (127). α-crystallin is a member of the small molecular weight heat shock protein family, members of which are ubiquitously present. The most well studied members of this family are the eye lense proteins (125). However counterparts of the eye lense proteins are also found in bacteria. The primary function of these classes of proteins is to prevent irreversible denaturation of proteins under stress conditions. The *M. tuberculosis* Acr protein also known as Hsp16.3 or HspX has been characterized in details using biophysical methods. Such studies have clearly shown that Hsp16.3 protects against irreversible denaturation of proteins (96). It performs this protective function by exposing the hydrophobic interior, which interacts with the exposed hydrophobic surfaces of denatured proteins thereby preventing its aggregation (19). However unlike the major chaperones such as DnaK and GroEL, it does not aid the folding of the protein. It is likely that, *in vivo* too Hsp16.3 functions by protecting proteins against irreversible denaturation. However there may be other functions which are not yet known. *In vivo*
it has been demonstrated that over-expression of this protein resulted in slower decline in viability following the log phase. It appears therefore that the major function of this protein is to stabilize the stationary phase of growth. Deletion of the gene for this protein resulted in loss of virulence properties and inability to survive in the persistent phase (139). It has been proposed that an hspX deleted strain may be used as a vaccine since it appears to be attenuated (96).

Stationary phase and starvation

The stationary phase arises when a bacteria is allowed to grow in liquid medium to saturation. In the stationary phase, stress can arise due to variety of reasons such as accumulation of toxic products, acidity, high osmolarity etc (106). However the most important form of stress under these conditions arises due to nutritional deficiency. Bacteria have evolved different ways of adapting to limiting nutritional conditions such as formation of spores, fruiting bodies etc. However not all bacteria form such specialized structures. Instead they adapt to the stationary phase by changing their size and shape and also by altering their metabolic activities. In case of M. tuberculosis it has been found that the bacteria isolated from the lungs are morphologically distinct from those that are actively growing. The cells become spherical and they lose their acid fast staining property (they become chromophobic) (77). This phenotype could be reproduced under in vitro conditions. By starving Mycobacterial cells in distilled water it could be demonstrated that the cells develop morphological characters that are typical of persistent bacteria (77). Interestingly these cells could survive for prolonged periods of time, as long as 2 years. When after this long period the cells were re-introduced into nutritionally rich medium, normal growth was encountered.

Signal transduction under stress.

Stress response in bacteria is known to be involved in elaborate signalling network. Perhaps the best example of how stress signals are sensed and converted into appropriate responses is the phenomenon of sporulation observed in B. subtilis. Multiple stress signals are channelized in such a way that B. subtilis can survive under stress conditions either by sporulating or by adapting other means which
includes development of competence, production of proteases and surfactants and by killing other cells to obtain nutrients (cannibalism) (42).

Two component signalling systems constitute an integral part of the stress response pathway in *B. subtilis*. From whatever is known so far the *M. tuberculosis* stress response is also highly complex involving multiple signalling systems. As in case of other bacteria, two component systems play a strategic role in the signal transduction pathways leading to an appropriate stress response. There are 11 complete two component systems encoded by the *M. tuberculosis* genome (68). Two such two component systems which have been relatively well studied and which play crucial roles during response to either Hypoxia or nutrient starvation or both are discussed below.

**DevR – DevS**

Two component signalling systems comprise a sensory component which is usually membrane associated and a Response Regulator (RR) component which in many cases, though not all functions as a transcriptional regulator (81). The signals are perceived by the sensory component through sensory domains such as GAF, PAS etc (98). The leads to auto-phosphorylation at a specific Histidine residue. The phosphate is then transferred to the RR component which becomes active and performs its action (115).

The DevR-DevS system was discovered long before the *M. tuberculosis* genome was sequenced. This two component system was found to be differentially expressed in the virulent Rv strain of *M. tuberculosis* but not in Ra, the avirulent one (58). Hence the *devR-devS* system was considered to be potentially a virulence gene cassette. The realization that DevR-DevS plays a key role in controlling the dormancy regulon became clear when microarray experiments revealed that operon encoding this two component system is specifically induced under hypoxia along with 26 other genes (74). In this two component system DevR (Rv3133c) functions as the Response Regulator. Deletion of the corresponding gene resulted in abrogation of the hypoxia response (68). While DevR functions as response regulator, DevS (DosS) is the sensory component. How DevS senses hypoxia is not clearly understood. It appears that DevS senses the lowering of redox potential, whereas another sensor DosT expressed from
the same operon is responsible for sensing the absence of O$_2$ (80). Alternatively it may be that DosS and DosT both sense the depletion of O$_2$, but threshold concentration of O$_2$ required to activate these sensors (121).

**MprA-MprB**

Apart from DosR-DosS, the MprAB two-component system, also involved in the development of persistence has been well characterized. Initial studies revealed that deletion of the gene encoding MprA resulted in inability of the mutant to survive in the persistent phase (79). Subsequent studies revealed that MprA, the response regulator which binds upstream of the genes encoding stress responsive Sigma factors SigE and SigB, and activates their synthesis (116, 117). MprAB based signalling systems respond to a variety of stress conditions which include SDS, low pH and nutritional stress. The sensory component, in this case is the activity of the response regulator MprA which is controlled by the sensory kinase MprB. MprAB is also involved in the expression of the virulence factor Acr2. Acr2, like Hsp16.3 (for the sake of convenience Hsp16.3 is referred to as Acr1) is an $\alpha$-crystallin family protein, but its synthesis is induced under heat stress and not Hypoxia (78). The protein becomes associated with ribosomes and it has been suggested that it might stabilize ribosomes at higher temperatures. Deletion of the Acr2 does not lead to reduced bacterial load but reduced pathology in the host was observed. This indicates that Acr2 may be involved in the host-pathogen interaction mechanism. Regulation of the Acr2 gene is multifactorial and very complex. There are binding sites for the alternative sigma factors SigE and SigH, MprAB and finally the repressor HspR which is the subject matter of this study. How the multiple factors cooperate with each other to bring about an orchestrated response is unclear. Further insight may be obtained only after the role of the individual elements comes to light.

**The stringent response - the role of RelA**

As mentioned above starvation constitutes an important signal for the transition of *M. tuberculosis* from actively growing state to the persistent phase. Starvation conditions can induce the stringent response. This response is a broad transcriptional program encompassing 80 genes in *E. coli* that mediates prokaryotic adaptation under...
starvation conditions (82). Induction of the stringent response in *E. coli* stimulates polyphosphate synthesis (85), increases fatty acid cyclopropanation, inhibits fatty acid and phospholipid synthesis, upregulates glycogen synthesis, upregulates the stationary phase sigma factor $\sigma^S$ and inhibits stable RNA synthesis. The stringent response can be induced by amino acid, carbon, nitrogen or phosphorous starvation. The stringent response is mediated by the increased levels of hyperphosphorylated guanine nucleotides (ppGpp) (20). In *E. coli* and other gram negative bacteria two proteins have been identified that regulate the synthesis of ppGpp, RelA and spoT (Fig. 3). RelA is ribosome associated and is activated by the binding of uncharged tRNAs to the ribosomes, upon depletion of amino acids. *E. coli* strains lacking both RelA and SpoT cannot grow in minimal media (82). *M. tuberculosis* has only one SpoT/RelA homolog (23). The RelA in *M. tuberculosis* is responsible for producing ppGpp. Mutants of RelA were found to be incapable of long term survival. MprAB two component system indirectly controls the activity of the relA gene by regulating the expression of $\sigma^E$, the sigma factor that is essential for the expression of the *rel* gene. The activity of MprAB is controlled by Polyphosphate kinase (116). The ppk1 gene of *M. tuberculosis* has been demonstrated to be essential for long term survival under hypoxia conditions. Polyphosphate kinase is involved in the synthesis of inorganic polyphosphates (PolyP), which are polymers of hundreds of phosphate residues. MprB the sensory component is phosphorylated in the presence of PolyP (Fig. 4). The phospho-residues are then transferred to MprA (116). MprA-P then acts as an activator of transcription of $\sigma^E$, which in turn activates expression of RelA. The chain of events beginning with the synthesis of polyphosphates and ending with the production of ppGpp, the alarmone, is crucial for the survival of the bacterium under stress conditions. The stringent

![Image](https://example.com/figure3.png)
response is bistable in nature. Positive feedback and noise can contribute to bimodal expression of the \( \text{relA} \) gene resulting in subpopulations which either express or do not express \( \text{relA} \) resulting in some cells that are more competent in resisting stress than others (117).

**Alternative Sigma factors**

Sigma factors are subunits of bacterial RNA polymerase and confer promoter specificity to the holoenzyme complex. The unique affinity of each sigma factor for its promoter consensus sequence is an essential component in many gene regulation systems (33). The \( \sigma \) factors have been divided into two main families namely the \( \sigma^{70} \) and \( \sigma^{54} \). All bacteria possess at least one \( \sigma^{70} \) family sigma factor. However not all bacteria possess \( \sigma^{54} \) class of sigma factors. *M. tuberculosis* does not possess \( \sigma^{54} \) family sigma factor (57). The \( \sigma^{70} \) family has been divided into three groups (47, 53). Group 1 comprises the essential housekeeping \( \sigma \) factors sometimes referred to as primary sigma factors. Group 2 are closely related to Group 1 except that they are not essential in nature. Group 3 comprises what are known as alternative sigma factors many of which are involved in the transcription of stress responsive genes. Some of the alternative ECF family \( \sigma \) factors or Extra Cytoplasmic Function \( \sigma \) factors as their functions are manifested following the perception of Extracellular Signals (53). All bacteria possess at least one essential \( \sigma \) factor that transcribes the genes required for cell viability and most bacteria harbour alternative \( \sigma \) factors that transcribe regulons in response to specific stimuli. The number of alternative \( \sigma \) factors correlates generally with the variability of the environments encountered by a given bacterial species. *M. tuberculosis* encodes a repertoire of 13 \( \sigma \) factors of which \( \sigma^{A} \) and \( \sigma^{B} \) are representatives of groups 1, 2 respectively while the remaining 11 sigma factors belong to Group 3 (95). It is interesting that *M. tuberculosis* has the highest \( \sigma \) factor to genome ratio amongst obligate pathogens, suggesting a highly complex regulatory mechanism for transcription of MprA/MprB (95). Perhaps the best
characterized alternative Mycobacterial sigma factor is $\sigma^E$. This sigma factor, which is a part of the ECF subgroup, is essential for growth in macrophages, and for causing virulence in mice. $\sigma^E$ is subject to an extremely complex regulation transcriptional, translational and post-transcriptional. As mentioned above the $\sigma^E$ gene expression is positively controlled by phosphorylated MprA (123). Additionally $\sigma^E$ is controlled at the post translational level by RseA an anti-$\sigma$ factor (Fig. 4). RseA may be phosphorylated by the eukaryotic type protein kinase PknB (95). The phosphorylated form of RseA can be degraded proteolytically resulting in the release of $\sigma^E$. Apart from $\sigma^E$ several other alternative $\sigma$ factors such as $\sigma^H$, $\sigma^F$ and $\sigma^C$ have been found to be involved in virulence in some form or the other.

**Heat shock response**

The classical heat shock proteins (HSPs) are stress-inducible molecular chaperones which represent the most conserved proteins in cellular life. In prokaryotes and eukaryotes their main role is to maintain a correctly folded and assembled protein component of the cell (50).

What makes the expression of HSPs so interesting to the pathogen biologist is that pathogen HSPs are also recognized by specific receptors on host immune cells, triggering an inflammatory immune response (60, 64). In addition to this, the conservation between host and pathogen HSPs means that pathogen HSPs can be utilized by the host to shuttle peptides into the HSP-mediated antigen presentation pathway (118). Thus, while the pathogen needs to increase expression of its HSPs in response to the stresses induced by host defences it must temper this need so as not to alert the host immune response to its presence (13).

Of course, heat shock is a complex response which varies with both time and temperature, but this snapshot of the response provides a reference with which to compare transcriptomic changes in defined regulatory mutants. One general observation on the heat shock transcriptional profile is that the response is not simply the elevated transcription of the known HSPs but encompasses genome-wide changes in gene expression.
To dissect this response, two likely heat shock repressor proteins in *M. tuberculosis*, HspR and HrcA, which were identified by homology to regulators in *Streptomyces* (10) and *Bacillus* (51) were knocked out. Mutants of both of these regulators using gene replacement system were made. Comparison of the expression profiles of strains lacking these regulators with wild-type, combined with identification of repressor binding sites in promoter regions, established the HspR and HrcA regulons (114). All members of these regulons were also found to be upregulated during heat shock and amongst them were many of the classical heat shock chaperones. Comparative studies examining the regulation of other arms of the mycobacterial stress response reveals a high degree of crosstalk and overlap between the different stress regulons. For example, the Hsp70 regulon forms a central element of the heat shock response and is under negative regulation by HspR in complex with Hsp70 itself. However, the Hsp70 operon is also under control of the heat-inducible alternative sigma factor σH (36), which in addition promotes transcription at the stress response sigma factors σB and σE (84). Further to this, the functional activity of these sigma factors is under the control of anti-sigma factor pathways (52).

The infection profile of *M. tuberculosis* is critically governed by a dynamic relationship with the host immune response. The vital importance of HSP regulation in *M. tuberculosis* has been demonstrated by generating a mutant strain lacking the HspR repressor protein, thus effecting dysregulation of the Hsp70 response (112). The mutant strain constitutively overexpressed Hsp70 and associated HSPs, and during murine infection its survival was emphatically reduced. The underlying cause for this attenuation was enhanced immune recognition of the bacterium.

Understanding the regulatory mechanisms behind mycobacterial HSP expression may allow the development of novel strategies for the treatment of tuberculosis. The dysregulation of the *M. tuberculosis* Hsp70 response allows the host to mount a more effective immunological response against the bacterium (110). Thus, drugs that disrupt HSP regulation by interfering with specific regulators make an attractive mechanism by which to enhance host immunity. The regulation of expression of heat shock proteins in bacteria will be discussed in the next chapter.
1.3. Regulation of Heat shock response

Role of heat shock proteins

What began as a molecular curiosity in fruit flies over 30 years ago, the so-called heat shock or stress response, now constitutes an active area of research in cell biology and biochemistry. The Heat Shock Proteins (HSPs), one of the most highly conserved groups of proteins so far characterized, are being implicated as essential components in a number of diverse biological processes. Although referred to as heat shock proteins, most of these proteins in fact are expressed at rather significant levels in all cells maintained under normal growth conditions and are essential for cellular growth at all physiologically relevant temperatures. Much of the current interest in the HSPs follows from recent studies demonstrating their role as molecular "chaperones", being intimately involved in various steps of protein maturation. Members of the Hsp70 (DnaK) and Hsp60 (GroEL) families, for example, participate in protein folding, protein translocation, and perhaps higher ordered protein assembly. Yet other members of the heat shock protein family, such as Hsp90, play important roles in the regulation of certain transcription factors and protein kinases. Many of the HSPs are constitutively present in the cell even under no-stress situation and perform important housekeeping functions (24). The major HSPs assist in correct folding and assembly of proteins and are involved in diverse cellular processes including DNA replication, RNA transcription, flagella synthesis and UV mutagenesis (140) (Fig. 5). GroEL together with GroES facilitates protein translocation across membrane barriers and possibly also secretion. The fundamental functions of these HSPs are to prevent protein denaturation and to reactivate partially denatured proteins. Non-repairable denatured proteins are degraded by another class of HSPs which represent either an ATP dependent protease (Lon, La) or a catalytic (C1pP, Ti) or regulatory (ATPase) subunit (C1pB) of another protease C1p (24). The pattern of HSP synthesis is dependent on the severity of the heat stress with two distinct phases being discernible in the heat shock response. Although the induction of HSPs is a universal response, the regulatory mechanisms controlling HSP synthesis differ widely between organisms. Bacteria
regulate the transcription of heat shock genes by means of both positive and negative mechanisms.

**Positive regulation**

*Escherichia coli* has long served as the paradigm for bacterial heat shock regulation. Regulation in this species is based on the use of alternative sigma factors that direct the RNA polymerase to heat shock gene promoters that differ from housekeeping promoters (Table 1, adopted from (75)). Transcription of some 30 genes constituting the σ^{32} (*rpoH*) regulon is transiently increased after a temperature upshift as a consequence of an elevated cellular level of σ^{32}, which is mainly caused by enhanced translation of *rpoH* mRNA and by increased stability of σ^{32}. The response is feedback-controlled by the DnaK machinery that sequesters σ^{32} under physiological conditions (activity control) and may also deliver it to FtsH, a protease that degrades σ^{32} (stability control). The key concept of this homeostatic control is that, after heat shock, denatured proteins titrate the DnaK chaperones away from σ^{32} leaving it stable and active (Fig. 6) (75). A second heat shock regulon in *E. coli* is controlled by σ^{6}, which belongs to a class of sigma factors that respond to extracytoplasmic stimuli, for example to unfolded proteins in the periplasm. The expression and activity of σ^{6} itself is controlled by a complex regulatory system (72). Finally, transcription of a third set of heat-inducible *E. coli* genes, the phage shock operon, is promoted by σ^{54}, an alternative sigma factor that always requires an activator (PspF, in this case) for transcriptional activation (73).

![Figure 6](image)

*Figure 6.* At low temperatures, σ^{32} interacts with DnaKJE and the complex is cleaved by FtsH, making σ^{32} unstable. At high temperatures, the DnaKJE chaperone machinery binds to unfolded polypeptides, and is involved in the folding process, and σ^{32} remains free. The free σ^{32} interacts with RNA-polymerase and the cleavage by FtsH is hindered. The half life of σ^{32} gets increased at 42 °C. Adopted from (75).
In *Bacillus subtilis*, more than 40 general stress genes (class II genes) are under the control of $\sigma^B$ (51). The activity of this sigma factor is modulated in a stress and starvation (energy limitation) responsive way. Evidence for additional regulatory mechanisms transpired from the finding that several stress genes in *B. subtilis*, among them the *dnaK* and *groE* chaperone operons, remained heat inducible in a $\sigma^B$ mutant and from an inspection of the corresponding promoter sequences that revealed typical housekeeping promoters. To confirm the postulated $\sigma^A$ (SigA) dependence (18) used a temperature-sensitive $\sigma^A$ mutant and demonstrated that *groESL* transcription is SigA-dependent. Other examples of heat shock genes transcribed from vegetative promoters were subsequently discovered in a number of bacteria, and regulatory DNA elements located in the promoter regions have been identified or proposed as repressor binding sites responsible for transcriptional repression under standard growth conditions. Overall, it now appears as if negative heat shock regulation is more prominent in eubacteria than positive regulation by alternative sigma factors.

### Negative regulation

**CIRCE/HrcA regulon**

CIRCE elements have been identified in more than 60 eubacteria, including gram-positive bacteria, proteobacteria, and cyanobacteria. The established consensus is TTAGCACTC N9 GAGTGCTAA. CIRCE has been reported to be associated mainly with the *dnaK* and/or *groEL* genes or their cohorts, *grpE*, *dnaJ* and *groES*. In *Streptomyces*, the *groES-groEL1* operon and the *groEL2* gene present tandem copies of the CIRCE element (Fig. 7). A new heat shock operon containing *hrcA* and *dnaJ* paralogs has been

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Representative organism</th>
<th>Regulated gene(s) or operon(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive control</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^{32}$ (RpoH)</td>
<td><em>Escherichia coli</em></td>
<td><em>dnaK</em>, <em>grpE</em>, <em>groESL</em> and more</td>
</tr>
<tr>
<td>$\sigma^{24}$ (RpoE)</td>
<td><em>Escherichia coli</em></td>
<td><em>htrA</em>, <em>rpoE</em>, <em>rpoH</em> operon and more</td>
</tr>
<tr>
<td>$\sigma^B$ (SigB)</td>
<td><em>Bacillus subtilis</em></td>
<td><em>gspA</em>, <em>csbA</em>, <em>katE</em> and more (class II)</td>
</tr>
<tr>
<td>$\sigma^N$ (RpoN), PspF</td>
<td><em>Escherichia coli</em></td>
<td><em>pspABCDE</em></td>
</tr>
<tr>
<td><strong>Negative control</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIRCE/HrcA</td>
<td><em>Bacillus subtilis</em></td>
<td><em>groESL</em>, <em>dnaK</em> operon (class I)</td>
</tr>
<tr>
<td>HAIR/HspR</td>
<td><em>Streptomyces albus</em></td>
<td><em>dnaK</em> operon, <em>clpB</em></td>
</tr>
<tr>
<td>RheA</td>
<td><em>Streptomyces albus</em></td>
<td><em>hsp18</em></td>
</tr>
<tr>
<td>ROSE</td>
<td><em>Bradyrhizobium japonicum</em></td>
<td><em>hspArpoH</em> and more</td>
</tr>
<tr>
<td>CstR</td>
<td><em>Bacillus subtilis</em></td>
<td><em>clpC</em>, <em>clpP</em> and more (class III)</td>
</tr>
</tbody>
</table>
identified in *Streptomyces albus* \(^{(45)}\). The hrcA-dnaJ2 operon is preceded by a canonical CIRCE sequence and the transcription of this operon is heat-induced. CIRCE operator sequences are highly conserved among bacterial species whereas the HrcA repressors of different species share few similarities. Disruption of the hrcA gene increases transcription of the groES-groEL1 operon and of the groEL2 gene. Despite considerable accumulation of the GroE chaperones in the hrcA mutant, formation of the aerial mycelium and sporulation were not affected, indicating that neither hrcA nor the level of groE gene expression are directly involved in the regulation of morphological differentiation in *Streptomyces* \(^{(103)}\). CIRCE motif is also known to present in other actinomycetes, i.e. *Mycobacterium tuberculosis* and *Mycobacterium leprae*. It is noteworthy that a CIRCE motif also precedes those paralogous genes \(^{(103)}\).

Unfortunately, *in vitro* analyses have been hampered by the insolubility of HrcA, which tends to form inclusion bodies. The addition of GroEL retarded the aggregation of HrcA as measured by light scattering and improved the DNA binding activity in a band shift assay (in the presence of 0.5 M urea) \(^{(75)}\). The same report provides *in vivo* evidence that GroEL is a specific modulator of the CIRCE regulon. HrcA lost its repressor function in the absence of GroEL, leading to high levels of CIRCE-controlled DnaK protein. Conversely, dnaK expression decreased when the GroEL level was elevated. This knowledge strongly supports a GroE titration model (Fig. 8) analogous to the DnaK/DnaJ titration model for \(\sigma^{32}\)-dependent genes in *E. coli* \(^{(75)}\). The key prediction is that HrcA depends on GroEL to acquire an active conformation. The steady-state pool of GroEL under normal growth conditions maintains HrcA in an inhibitory state, providing feedback repression of the *groESL* operon (and repression of other CIRCE-dependent operons, if present). After a heat shock, the chaperonins

![Figure 7. HrcA binds to a well-conserved 9-bp inverted repeat called CIRCE and negatively regulates the expression of groESL operon. Adopted from \(^{(103)}\).](image)

![Figure 8. Titration models for homeostatic control of chaperone expression. A. GroEL as cellular thermometer in HrcA-controlled systems. Dotted arrows indicate protein–protein interactions; (+) symbolizes activation, and (−) repression. Adopted from \(^{(75)}\).](image)
become engaged in refolding denatured proteins which depletes the GroEL pool. HrcA is rendered inactive allowing for continuous production of GroEL (Fig. 8). The response is self-limiting when cells have recovered and GroEL becomes available again. This titration model regarding GroEL as cellular thermometer is very appealing for its simplicity and striking analogy to the σ32-mediated heat shock control of E. coli in which DnaK functions as the thermosensor (75).

HAIR/HspR regulon

The group of W. Schumann showed, in 1996, that the first gene of the dnaK operon of B. subtilis encodes HrcA (100). In S. coelicolor, the simultaneous determination of the nucleotide sequence of the dnaK operon by the teams of C. Smith (12) demonstrated the absence of CIRCE and hrcA sequences associated with this operon. A sequence similar to that of the transcriptional regulator gene glnR was identified at the 3' end of the operon. Bucca et al. (1993) showed that the product of this gene bound the dnaK promoter region, and the gene was named hspR for heat shock protein regulator. A hspR mutant of S. albus was generated. In this mutant, transcription of the dnaK operon is constitutively derepressed at all temperatures, with the DnaK protein highly overproduced. These results, and those demonstrating the binding of HspR to the promoter of the dnaK operon, showed that HspR is the repressor of the dnaK operon (46) (Fig. 9). The clpB gene is also constitutively overexpressed in the hspR mutant. This observation led to a study of the regulation of this gene. The ClpB protein belongs to the Clp ATPase family. Some members of this family associate with the ClpP proteolytic subunit to form ATP-dependent proteases. However, they may also act alone, as bona fide molecular chaperones. This is the case for ClpB, which has never been reported to be associated with a proteolytic complex. The clpB has been cloned and a clpB mutant has been generated (44). This mutant strain, like the hspR mutant, shows no major impairment of the developmental cycle. Primer extension identified a transcription start site preceded by typical vegetative -10 and -35 hexamer sequences. This site is associated with an inverted repeat that is also present in the promoter

18
region of the *dnaK* operon (IR3). DNA binding experiments have shown that HspR regulates *clpB* transcription by interacting directly with this motif. The HspR binding site was called HAIR (HspR-associated inverted repeat). The consensus sequence of this motif is CTTGAGT N7 ACTCAAG.

Expression of the *groEL* and *hsp18* genes is normal in the *hspR* mutant. Thus, in contrast to the situation in *B. subtilis* and other gram-positive bacteria, the transcription of *dnaK* and *groEL* is not controlled by the same repressor in *Streptomyces*. Genes similar to *hspR* have also been identified in *M. leprae, M. tuberculosis* and in bacteria related to *Campylobacter jejune* (54) and *Helicobacter pylori*, and *Aquifex aeolicus* (31). HspR binding sites have been identified in these bacteria upstream from various heat shock genes, suggesting that these genes are regulated by HspR, and HAIR/HspR has been reported to be involved in heat shock regulation in *Helicobacter pylori* (108). The HspRs of Eubacteria belonging to the phylum *Thermus-Deinococcus* have also been identified (99). The DNA sequence of *S. coelicolor* was searched for the HAIR motif. In addition to the HAIR motifs associated with *dnaK* and *clpB* (Fig. 10), a HAIR sequence was found upstream from a gene closely related to the *lon* gene, which encodes a ubiquitous ATP-dependent protease. Analysis of the transcription of the *S. lividans* *lon* gene confirmed that it was regulated by HspR (107). There are no genes related to *lon* in the pathogens *M. tuberculosi*s and *M. leprae*, but there is an unreported HAIR motif upstream from *lon* in *Mycobacterium smegmatis*, suggesting that HspR/HAIR regulation of *lon* may be characteristic of all actinomycetes in which this gene is present.

**Mycobacterium tuberculosis** HspR

In *M.tuberculosis*, open reading frame (ORF) Rv2374c encodes a homolog of the HrcA repressor, whereas Rv0353 encodes a protein similar to HspR, a repressor identified in *hsp70* regulation in *Streptomyces* and in *Helicobacter pylori*. In *M. tuberculosi*s *hspR* is the fourth gene in the *dnaKJE* operon, comprising of the *hsp70* or *dnaK*, followed by
genes encoding \textit{grpE} and \textit{dnaJ}, HSPs that have functional interactions with Hsp70 and are known as their co-chaperones (65) (Fig. 11).

\textbf{Proposed mechanism of interaction between HspR & DnaK}

The mechanism by which DnaK functions as a co-repressor has been investigated in a previous study using HspR of \textit{Streptomyces coelicolor} which like \textit{M. tuberculosis} belongs to the phylum Actinobacteria. The HspR protein interacts tightly with Hsp70 \textit{in vitro} (112) and DnaK functions as a transcriptional co-repressor by binding to HspR at its operator sites in \textit{S. coelicolor} (Fig. 13). A system where this heterodimer forms the functional repressor unit with feedback achieved by titration of Hsp70 away from the HspR complex in the presence of unfolded polypeptides represents an attractive
model for regulation (9). DnaK activates HspR repressor, rather than inactivating an activator (such as σ^32). Despite this reversal of function, the end-result is the same, the molecular chaperones negatively regulate their own synthesis by modulating the activity of their respective transcription factors. Thus, although mechanisms controlling molecular chaperone gene expression vary widely, homeostatic feedback control by a key chaperone is emerging as a common theme. In DnaK-mediated regulation of the *S. coelicolor* dnaK operon, under normal growth conditions, native HspR binds DnaK, and this complex binds avidly to the dnaK promoter region to repress transcription of the operon efficiently. In the (transient) absence of DnaK, HspR does not bind with high affinity to its DNA target. Thus, during heat shock, when DnaK is sequestered by denatured or partially unfolded proteins, the HspR protein is unable efficiently to repress transcription and the operon is induced. It was also demonstrated that DnaK can enhance the binding activity of HspR, in an ATP independent manner. Hence it was proposed that the chaperone function of DnaK is not involved in the process. Contrary to this observation, investigations done in this laboratory (27) with the mycobacterial counterpart revealed that the DnaK dependent DNA binding activity of HspR can

![Figure 13](image-url)

**Figure 13.** A model for DnaK mediated feedback regulation of *Streptomyces coelicolor* dnaK operon. (A) represents normal growth condition, (B) scenario after heat shock. Adopted from (9).
indeed be stimulated in the presence ATP, and hence the chaperone function of DnaK may have a role to play. It was also shown that the mycobacterial GroELs, GroEL1 and GroEL2 play co-stimulatory role in the process of reactivation of HspR, when the amount of HspR used is limiting. The role of DnaJ and GrpE, the co-chaperons of DnaK, was also established. The presence of GrpE inhibited the operator of binding activity (Fig. 14).

These studies have established the fact that the interaction between HspR and DnaK is necessary for DNA binding activity of HspR. However the mechanism by which DnaK activates HspR is not clear. There are many questions yet to be answered regarding this interaction. What is the molecular basis of this interaction? Does DnaK have any influence on the structure of HspR? What type of protection DnaK provides against denaturation of HspR? What is the role of DnaK in folding of HspR? How does DnaK recognize HspR? To address these questions it becomes necessary to discuss about DnaK in the next chapter.
1.4. Protein folding and role of DnaK

Protein folding problem

To be biologically active, all proteins must adopt specific folded three-dimensional structures. Yet the genetic information for the protein specifies only the primary structure, the linear sequence of amino acids in the polypeptide backbone. Most purified proteins can spontaneously refold in vitro after being completely unfolded, so the three-dimensional structure must be determined by the primary structure. The pioneering work of Anfinsen (2) with the *in vitro* refolding purified ribonuclease A left the long-lasting impression that the folding of a newly synthesized polypeptide was an intrinsic feature of its primary structure, independent of other factors. Most of the *in vitro* protein refolding experiments are usually carried out by first denaturing a given purified polypeptide and then removing the denaturant. Under these conditions, most polypeptides quickly collapse into a compact structure, usually called "molten globule", thought to possess extensive secondary structure, but still exposing hydrophobic groups, which may lead to aggregation (25) (Fig. 15). The probability that a given unfolded polypeptide will fold properly increases at relatively low protein concentrations (which limit inter-polypeptide aggregation) and low temperature (which attenuates hydrophobic interactions). However, the relatively high protein concentrations in the cytosol [or the specialized organelles, i.e. endoplasmic reticulum (ER), mitochondria, chloroplasts] may subject the growing polypeptide chains, as they emerge from the ribosomes (or as they enter various cellular compartments), to premature interactions with other intra- or inter polypeptide domains, thereby leading to misfolding and aggregation (41). To deal with these sorts of problems, a set of proteins, collectively called *chaperones*, has

![Protein Folding Pathway](image)

*Figure 15.* Protein folding pathway. During refolding protein experiences three conformational states, stable at equilibrium. The ideal unfolded protein is the random coil. The native state is properly folded and/or assembled form, which is operative and functional. In this hydrophobic side chains comprises the folded interior and hydrophilic parts are exposed at surface. Molten globule is a collapsed molecule with native-like secondary structure and liquid like interior. The interconversions with the fully unfolded state are rapid and nonco-operative, but slow and co-operative with the fully folded state.
been identified and whose primary function is to ensure that polypeptides will fold or assemble properly in the cell. These chaperone proteins act primarily by binding to the reactive surfaces of polypeptides (such as the hydrophobic surfaces exposed by the molten globule intermediates of various proteins). In doing so, chaperones sequester these reactive sites from the rest of the reactive surfaces present in their vicinity, thus effectively preventing aggregation and favouring the proper folding pathway. The chaperone proteins act without covalently modifying their polypeptide substrates and without being part of the finished product. Because high temperatures tend to favour both protein unfolding on the one hand and hydrophobic interactions on the other, there is an extra need for chaperone protein function to prevent protein aggregation in vivo. This is most likely the reason why many protein chaperones are expressed at higher levels in cells after heat shock (41).

**Figure 16.** Schematic of protein synthesis, folding and misfolding and aggregation in an E. coli cell. Recent genetic and biochemical experiments have revealed that DnaK/DnaJ/GrpE function co-translationally in cooperation with ribosome-bound trigger factor, which is a prolyl isomerase. An essential post-translational function of DnaK is to break up large protein aggregates, and DnaK accomplishes this in concert with the Hsp104 chaperone ClpB. TF, K, J and E denote trigger factor, DnaK, DnaJ and GrpE respectively. I denotes a productive folding intermediate. The concentration of the free unfolded protein in a cell is vanishingly small. Adopted from (105).

**DnaK**, the 70 kDa family of heat shock protein chaperone (Hsp70), is the focus of this chapter. Figure 16 illustrates the effect of kinetic partitioning on the distribution of protein states within a cell. Within the text of this chapter, nascent and folded states are designated by $U_n$ and $F$ respectively; the free unfolded, misfolded and aggregated states are designated $U$, $M$ and $A$ respectively. Spontaneous protein folding occurs on the timescale of milliseconds. However, depending on the size of the protein and its energy landscape, a fraction of the unfolded molecules often misfolds and aggregates (41). Misfolded and aggregated proteins may be trapped in these states for minutes or even hours because the half-times for the return to folding-competent intermediates are very long. Under a variety of stressful conditions, or as a consequence of deleterious amino acid changes to a protein's sequence, misfolded and aggregated proteins may accumulate, causing a cell to be deprived of its essential proteins, which leads to cell death. A wealth of data indicate that Hsp70 molecular chaperones use free energy from ATP binding and/or
hydrolysis to minimize the concentrations of M/A states in the cell. The question is, how do they do this?

**Holding versus unfolding**

One widely held view is that Hsp70 chaperones selectively bind short hydrophobic segments of nascent proteins or heat-denatured proteins and, by sequestering these segments, misfolding and aggregation is inhibited. Another way to state this is that Hsp70 chaperones possess antifolding or holding activity, in that they stabilize unfolded states of other proteins. Rothman (1989) (90) formulated four essential properties that a chaperone must possess to be a holdase: (i) chaperones should bind unfolded segments of polypeptide chains; (ii) binding should persist through the time required to synthesize a protein or translocate it through a membrane (the duration of protein synthesis is ≈ 20 s in bacterial cells and 300 s in eukaryotic cells); (iii) dissociation must occur at a useful rate in order to keep pace with synthesis and translocation; and (iv) chaperone function is energy dependent. Hsp70 molecular chaperones possess the essential properties (i), (iii) and (iv). However complexes between DnaK and its substrate proteins are so short-lived in the presence of its co-chaperones that DnaK does not fulfil (ii); thus, it should seriously be considered that DnaK (and probably other Hsp70s) does not function via holdase activity (105).

An opposing view is that Hsp70 chaperones are unfoldases that use free energy from ATP binding and/or hydrolysis to unfold or pull apart misfolded and aggregated proteins to yield productive folding intermediates (M → I and A → I) (90). Long-lived Hsp70–substrate complexes are not required for this mode of chaperone action. The productive folding intermediate then folds spontaneously in milliseconds to the native state (I → F). In this mechanism, Hsp70 proteins do not change the microscopic rate constant for the folding reaction (I → F); instead, they lower the activation energy barrier for the M → I and A → I reactions, and thus increase the microscopic rate constants for these reactions (105). Note that, during protein synthesis, a nascent chain can misfold (Un→M). Acting as an unfoldase, a Hsp70 chaperone could reverse this reaction, thus keeping the polypeptide chain in a folding-competent state until synthesis is completed.
DnaK’s Structure and functions

Composed of two functional domains, the ATPase domain (residues 1–385) and the polypeptide-binding domain (residues 389–508), which is capped by a lid domain (residues 509–638) (Fig. 17), DnaK binds ATP tightly but hydrolyses it incredibly slowly (94). The hallmark of DnaK and other Hsp70 chaperones is that nucleotide modulates peptide binding and release: in the absence of co-chaperones, ADP-bound DnaK binds and releases peptides over a timescale of minutes or even hours, whereas ATP-bound DnaK binds and releases peptides over a timescale of seconds or even milliseconds. The structure of the ATPase domain is defined by a central cavity for nucleotide binding at the base of two lobes (49). The C-terminal peptide-binding domain of DnaK is composed of a unique β-sandwich domain arranged in two sheets of four antiparallel β-strands, and this domain is followed by an α-helical lid-like domain (αA–αE) (Fig. 17) (141). The helical subdomain of the Substrate Binding domain D acts like a lid over the β-sandwich subdomain. The α-helices A and B are packed onto the β-sandwich, and helix B covers the substrate-binding cavity. The distal part of helix B, together with helices C, D, and E, builds up a hydrophobic helical core that acts as a “lid” (37). Although interdomain coupling occurs in the absence of the lid and the bound peptide contacts the β-structure but not the α-helical subdomain, its presence has been confirmed to be essential in stabilizing Hsp70-substrate complexes, especially at stress temperatures. A network of hydrogen bonds and a salt bridge links the lid noncovalently to the β-sandwich. The salt bridge Asp526–Asp445 (Fig. 18) plays a pivotal role in the proper positioning of the lid. The bound peptide interacts with the β-
sandwich but not the lid. The role of the lid in the chaperone activity cycle is still not precisely understood. Interdomain coupling occurs in the absence of the lid (69), although ATP-induced peptide dissociation is significantly accelerated compared to the wild-type protein (14).

**Substrate Binding specificity of DnaK**

Molecular chaperones form a class of proteins that selectively bind to nascent, unfolded, or aggregated polypeptides. It has been suggested that the recognition of non-native proteins by chaperones is mediated by their binding to exposed hydrophobic areas, to the polypeptide backbone, or to specific secondary structures. DnaK, member of hsp70 family protein, can distinguish native proteins from their non-native forms (66), owing to the specificity of its peptide binding site. Richarme et al. (89) showed that the peptide dependent ATPase activity of DnaK is specifically stimulated by several hydrophobic amino acids (Ile, Phe, Leu and Val). Since these amino acids are buried inside the hydrophobic core of native proteins and tend to be exposed in non-native forms, it was proposed that their specific interaction with DnaK is the basis of the specific interaction of chaperone with non-native proteins during protein folding, protein targeting to membrane, or protein denaturation. The sequence specificity of DnaK-peptide binding was then studied using f1 phage peptide-display library (43). The phage carries random six amino acid insertions near the N-terminus of the pIII surface protein (101). Peptides with affinity for DnaK were first screened out and then, heptapeptides were designed according to the sequences of peptide epitopes of individual selected phage. Results showed that peptides containing internal hydrophobic residues and terminal polar residues are preferential substrates for DnaK. Peptides with net negative charge bind poorly. This study also showed that amongst these peptides NRLLLTG binds best to DnaK. Further studies showed that DnaK interacts specifically with the positively charged amino acids arginine and lysine (30). The identification of crystal structure of the COOH-terminal substrate binding domain of *E. coli* DnaK complexed with NRLLLTG (Fig. 19) (141) revealed a hydrophobic substrate binding cleft with a central pocket tailored to bind leucine residue. From the structure it was observed that the hydrophobic side-chain of NRLLLTG contacts between the loops and the three central leucines in the peptide.
hydrophobic pocket in the floor of the channel. It is surmounted by a hydrophobic “arch” that also interacts with the hydrophobic side chains of the neighbouring leucine residues in the peptide. The interactive surface of DnaK at the ends of the hydrophobic channel is negatively charged. This negatively charged interactive surface favours the presence of basic residues at the end positions of the peptide (e.g., arginine at position 2). However this structural information and other previous studies could not be able to provide information regarding DnaK binding sites in biologically relevant protein substrates. A study was performed by Rüdiger et al. to identify the binding sites within protein sequences and the substrate binding motif of the DnaK chaperone (91). For this purpose they screened cellulose-bound peptide scans (88) representing complete protein sequences for DnaK binding. From this study the consensus substrate binding motif of DnaK was identified. The consensus motif recognized by DnaK consists of a central hydrophobic core of four to five residues and two flanking regions, of approximately four residues each that are enriched in basic residues.

\[
\text{Hy} \text{Hy} \text{Hy} \text{Hy} \text{Hy} + + B + B IM B
\]

(Fig. 19, L3-5) makes the most extensive contacts. The central-most leucine side chain dwells in a hydrophobic pocket in the floor of the channel. It is surmounted by a hydrophobic “arch” that also interacts with the hydrophobic side chains of the neighbouring leucine residues in the peptide. The interactive surface of DnaK at the ends of the hydrophobic channel is negatively charged. This negatively charged interactive surface favours the presence of basic residues at the end positions of the peptide (e.g., arginine at position 2). However this structural information and other previous studies could not be able to provide information regarding DnaK binding sites in biologically relevant protein substrates. A study was performed by Rüdiger et al. to identify the binding sites within protein sequences and the substrate binding motif of the DnaK chaperone (91). For this purpose they screened cellulose-bound peptide scans (88) representing complete protein sequences for DnaK binding. From this study the consensus substrate binding motif of DnaK was identified. The consensus motif recognized by DnaK consists of a central hydrophobic core of four to five residues and two flanking regions, of approximately four residues each that are enriched in basic residues.

The features of this motif agree well with those of the structure of the substrate-binding domain of DnaK (141). The binding cavity is suited to interact with approximately five consecutive residues. A central hydrophobic pocket, which makes the major energy contribution to binding of the co-crystallized heptapeptide, is tailored to bind Leu, but also Ile and Val, although probably yielding lower binding energies. However, the crystal structure of the substrate binding domain in complex with the peptide substrate did not elucidate the entire amino acid spectrum capable of associating with the substrate binding cavity, in particular with the four positions

**Figure 19.** The crystal structure of DnaK (cyan in colour) in complex with NRLLTGT (blue in colour). PDB code: 1DKZ.

**Figure 20.** Prediction of DnaK binding sites in protein sequences. (A) The DnaK binding motif used to establish an algorithm consists of a five residue hydrophobic core (Hy) flanked by four residue segments (+, with weights for the algorithm decreasing with increasing distance from the core) enriched in Arg and Lys. (B) Energy score distribution as a percentage of total DnaK binding (-) or non-binding (--) peptides. (C) Score prediction validity as percentage for total DnaK binding (--), good binding (---), binding (—) and non-binding (...) peptides. Adopted from (91).
outside the central hydrophobic pocket, and the contributions of negatively charged residues surrounding the binding cavity (141). This study defined (91) (i) the consensus binding motif including the entire amino acid spectrum capable of associating with DnaK (Fig. 20), (ii) the residues disfavored in binding sites (in particular Glu and Asp), (iii) the number of hydrophobic residues present in individual good DnaK binding sites (two to four on average), (iv) the extraordinary importance of Leu for DnaK binding (present in 87% of the tested hydrophobic cores of DnaK binding sites) and (v) a role for basic residues adjacent to the hydrophobic core, most likely allowing electrostatic interactions with the negatively charged surface surrounding the substrate binding cavity. This information on the consensus binding motif as well as the sequence identity of many individual DnaK binding regions provided a basis for further dissection of structural features of DnaK-substrate complexes.

Chaperone Cycle of DnaK

The key to the entire functional cycle of Hsp70 proteins lies in the mechanism by which the action of DnaJ and GrpE proteins couples the regulated ATPase cycle of Hsp70 with productive substrate binding. Such coupling prevents stimulating futile ATPase cycles in the absence of substrate. In the case of the DnaK-DnaJ system, it has been observed that in addition to binding to DnaK, DnaJ itself associates with substrates of the DnaK system with kinetics fast enough to prevent their aggregation (62). This is achieved possibly by binding to a sequence motif similar to that recognized by DnaK. Furthermore, the efficiency with which DnaJ stimulates the ATPase activity of DnaK is strongly increased by the presence of polypeptide substrates. These data suggest a model of the functional cycle.

Figure 21. Model of the Chaperone Cycle of the DnaK System. The cycle starts with the association of DnaJ with a substrate, followed by transfer of the substrate to the ATP form of DnaK (K-ATP). This transfer is coupled to the locking-in of the substrate in the substrate-binding pocket of DnaK by ATP hydrolysis. Following substrate transfer, DnaJ leaves the complex, and GrpE associates with the DnaK-substrate complex to trigger ADP release from DnaK. This allows binding of ATP and subsequent release of GrpE and substrate from DnaK.
of the DnaK system: (Fig. 21) (1) The cycle starts with the transient and rapid association of DnaJ with substrates, although in some cases the cycle may start with the association of DnaK-ATP with a substrate. (2) DnaK-ATP accepts polypeptide from the DnaJ-substrate complex in a process requiring two steps. The first being the transient interaction of DnaK-ATP with the J domain of DnaJ through an undetermined DnaK binding site and the second is the transfer of the substrate protein from DnaJ to the open substrate binding pocket of DnaK-ATP. Both these steps together are required to stimulate ATP hydrolysis by DnaK, resulting in stabilization of the DnaK substrate complex and tightly coupling ATP hydrolysis to substrate binding by DnaK. (3) Upon substrate transfer to DnaK and conversion of DnaK to the ADP state, the affinity of the DnaK-substrate complex for DnaJ gets reduced. This leads to the dissociation of DnaJ from the complex. This step is reflected in observations that ternary DnaK-DnaJ-substrate complexes are unstable (40) and that DnaJ acts catalytically in targeting DnaK to substrates (67) (4) GrpE is the nucleotide exchange factor and hence it binds to the DnaK-ADP-substrate complex, thereby triggering the release of ADP. Consequently, (5) ATP binds rapidly to DnaK, which releases the bound substrate and GrpE and returns DnaK to its initial state.
1.5. Objective

Tuberculosis remains a major health problem worldwide. The problem is particularly acute in third world countries like India. Despite the availability of various control measures and chemotherapeutic agents, it has not been possible to eradicate TB. This is partly due to the fact that none of the existing drugs can target the persistent state of the bacterium. Persistent bacteria may be generated spontaneously, but stress conditions are known to increase the probability of formation of persistent bacteria. In the granuloma, where persistent Mycobacteria reside, the bacteria are subjected to various kinds of stress- anaerobic, acid, superoxide, NO etc. Pathogenic bacteria such as *M. tuberculosis* have to overcome these various types of stress as they colonize their hosts. Understanding how the stress response operates in such bacteria is important from the overall perspective of bacterial pathogenesis.

In recent years most of the work in the area of stress response has been performed under either hypoxic or nutritional deficiency conditions. Significant information has been obtained regarding how *M. tuberculosis* reacts under these stress conditions. While in case of hypoxia stress related signals are transduced through the two component system known as DevR-DevS, in case of nutritional stress signal transduction involves the alarmone ppGpp, the production of which is indirectly controlled by the two component system MprAB. Stationary phase of growth is another well accepted model for studying stress response. In the stationary phase conditions prevail that are similar to both hypoxia as well as nutrient starvation. It is likely that in the stationary phase both DevR-DevS and ppGpp pathways are active.

In contrast to hypoxia or nutritional stress very little information is available regarding how heat shock response is regulated in *M. tuberculosis*. This is so despite the fact the fact that studies into how heat shock response in *M. tuberculosis* was initiated long time back. Most of these early studies were focused on the qualitative aspects of the response rather than the mechanistic ones. However investigation into the mechanism of action is important as heat shock proteins play a crucial role in pathogenesis. On the one hand these proteins protect the pathogens under stress condition, whereas on the other they serve as signaling modules which alert the host’s immune system.
The advent of the genomics era has remarkably changed our views about how stress responses operate. Global transcriptional profiling experiments have revealed that stress activates wider arrays of genes and operons than what was earlier believed to be. Many of these genes encode proteins that contribute to pathogenesis. Although different types of stress conditions are likely to activate specific subsets of stress responsive genes, there is apparently no sharp demarcation. One type of stress response may affect genes that are expected to be induced/repressed by some other type of stress conditions. An example of such intricate networking is observed in the case of the expression of the gene acr2. The regulatory region of this gene has a binding site for the heat sensitive repressor HspR, but at the same time there are multiple sites for the regulator MprA which is activated under nutritional deficiency and stationary phase conditions through the Polyphosphate Kinase pathway. This indicates possible cross talk between Heat Shock and Nutritional Stress Response pathways.

Previous studies in Streptomyces and Mycobacterium have established that DnaK physically interacts with HspR and that DnaK-HspR interaction is necessary for HspR’s activity. How this interaction aids HspR’s DNA binding activity is unclear. A variety of questions arise regarding how DnaK activates HspR. Does DnaK change HspR structure? Does it protect against thermal denaturation of HspR? How does DnaK recognize HspR? Is there any specific sequence involved? These are essentially the issues that this study seeks to address.

**Objective 1.** Although previous attempts have been made to understand HspRs DNA binding activity, the studies were carried out in the presence of DnaK. The native protein cannot be isolated in the pure form free from DnaK contamination (9). Hence a method had to be devised which would allow renaturation of HspR following its isolation in the denatured state, The renatured protein can then be used in various assays to study how HspR interacts with DnaK. **Devising a methodology to renature the protein into a soluble and biologically active form, without the aid of DnaK is therefore one of the goals of this investigation.**

**Objective 2.** HspR is a conserved protein, which belongs to the MerR family of winged Helix-turn-Helix DNA binding proteins. Given that a large number of HspR sequences are available in the databases therefore in silico investigations should give valuable
clues regarding the sequences that are important for either DNA binding or DnaK-binding or both. Hence bio-informatic analysis of these sequences to identify potentially important sequence motifs is another issue that has been taken up in this study.

Objective 3. Understanding protein-protein interaction mechanisms require the application of recombinant DNA methods, particularly mutational analysis. Combined with direct binding experiments in vitro and indirect functional assays in vivo such studies can potentially lead to valuable insights into mechanism of protein-protein interaction. Hence mutational analysis of HspR followed by binding studies could give us the necessary molecular insight into the nature of DnaK-HspR interaction. Thus the final and the most important objective of this study is to understand the structural basis of the HspR-DnaK interaction phenomenon.