The University of Calgary

Regulation of *groE* Expression in *Bacillus subtilis*

by

Gang Yuan

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF BIOLOGICAL SCIENCES

CALGARY, ALBERTA

October, 1995

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Regulation of *groE* Expression in *Bacillus subtilis*" submitted by Gang Yuan in partial fulfillment of the requirements for the degree of Master of Science.

Supervisor, Dr. S.- L. Wong, Department of Biological Sciences

Dr. P. W.K. Lee, Department of Microbiology and Infectious Diseases

Dr. K. E. Sanderson, Department of Biological Sciences

Dr. G. Voordouw, Department of Biological Sciences

Jan. 15. 1996

Date

ABSTRACT

The regulatory mechanism controlling *groE* expression in *Bacillus subtilis* is characterized. This operon is shown to be transcribed by the σ^A containing RNA polymerase. An inverted repeat sequence located downstream from the transcription start site plays two key functional roles in regulating *groE* expression. It serves as a signal ensuring the fast turnover of the *groE* transcript under the non-heat shock condition and also functions as an operator. Isolation and characterization of seven *groE* regulatory mutants demonstrate that *orf39*, the first gene in the *dnaK* operon, encodes a repressor that binds specifically to the inverted repeat sequence. All these findings allow the establishment of the regulatory mechanism controlling the expression of heat shock genes encoding molecular chaperones in *B. subtilis*. A similar mechanism is likely to exist in at least twenty eight different eubacteria.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor Dr. Sui-Lam Wong for his support and careful supervision.

I would to like to thank the members of my supervisory committee, Dr. Kenneth E. Sanderson and Dr. Gerrit Voordouw, for their helpful suggestions about this project. I would also like to thank Dr. Patrick W.K. Lee for reading the thesis.

I would like to express my many thanks to everybody in Dr. Wong's lab, particularly to Ruiqiong Ye and Xu-Chu Wu who help me a lot throughout my study.

I would like to acknowledge Mr. Rongdian Fu, for always making his time to help me solve problems, and giving me encouragement when I was depressed.

A special thanks to Sau-Ching Wu and Anita Arora for sequencing some of the *groE* mutants. The proof-reading of this thesis by Fiona Cowman is appreciated very much.

Finally, I am very grateful to my wife Chunli Zhang's patience during my study and all she did for me over these years. Without her help, these works cannot be finished.

Dedicated to my parents and my family

Table of Contents

List of Tables

List of Figures

ABBREVIATIONS

amyE	the structural gene of α-amylase
bgaB	the structural gene of ß-galactosidase gene from *Bacillus stearothermophilus*
bla	the structural gene of ß-lactamase
bp	base pair
cat	chloramphenicol acetyl transferase gene
CIRCE	controlling inverted repeat of chaperone expression
DTT	dithiothreitol
hsp	heat shock protein
IPTG	isopropylthio-ß-D-thiogalactopyranoside
kb	kilobase pairs
LacZ	*E. coli* ß-galactosidase
lacZ	*E. coli* ß-galactosidase gene
NTG	N-methyl-N'-nitro-N-nitrosoguanidine
ONPG	o-nitrophenyl-D-galactoside
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
SDS	sodium dodecyl sulfate
Tris	tris (hydroxymathyl) aminomethane
Xgal	5-bromo-4-chloro-3-indoly-ß-galactoside

I Literature Review and Objectives

I.1 Literature Review

I.1.1 The Heat Shock Response and Molecular Chaperones

When cells are shifted from low to high temperature, a set of proteins designated heat shock proteins (hsps) is induced. Synthesis of hsps increases upon temperature upshift and decreases upon the temperature downshift. The heat shock response exists in most organisms ranging from bacteria, lower eukaryotes to higher eukaryoties including mammals and plants. Not only the response to the temperature upshift is conserved, but also the hsps are conserved. Some of the hsps are described as Molecular Chaperones, which are defined as a family of unrelated proteins that mediate the correct folding and assembly of other proteins but do not form part of the functional assembled structures (Ellis, 1991). These proteins include the GroE, DnaK, DnaJ, and GrpE, etc. Their functions involve assisting protein folding, assembly, export and preventing aggregation.

Among these molecular chaperones, *E. coli* GroEL and GroES are well characterized. They are identified by mutations affecting the morphogenesis of several bacteriophages (Georgopoulos *et al.*, 1973; Hohn *et al.*, 1979; Tilly and Richardson, 1985). GroEL exists as a complex of a 14-subunit oligomer which forms two stacked rings with seven subunits per ring. Each ring has a central hole (Hendrix, *et al.* 1979, Hohn *et al.*, 1979), whereas GroES is a single ring of seven subunits (Chandrasekhar *et al*, 1986). Under physiological conditions, GroEL transiently exists as a complex with GroES. A single GroES heptamer binds to one end of the GroEL complex conferring

the asymmetry to the chaperone complex. In *E. coli*, GroEL and GroES are also required for viability under normal growth conditions (Fayet *et al.*, 1989; Georgopoulos *et al.*, 1973; Sternberg, 1973; Tilly *et al*, 1981). These proteins mediate the folding of many intracellular proteins (Horwich *et al.*, 1993). GroEL has been shown *in vitro* to form a complex with several secretory proteins, including pro-OmpA, pre-PhoE and pre-ß-lactamase (Bochkaneva *et al.*, 1988; Kumamoto, 1991; Lelker *et al.*, 1989). Requirement of GroEL for the export of ß-lactamase is established *in vivo* (Kusukawa *et al.*, 1989). Besides interacting with intracellular proteins, overproduction of GroEL in *E. coli* can facilitate the export of *lacZ* hybrid proteins (Phillips *et al.*, 1991). It is not surprised to find the homologs of GroES and GroEL in many other organisms including bacteria and yeast, plant and human (Ellis and van der Vies, 1991).

In *E. coli*, *groEL* and *groES* are organized in an operon (Hemmingsen *et al.*, 1988). Regulation of *groE* and other hsps genes has been well studied. Although several *B. subtilis* hsps genes have been characterized, the regulation of hsps expression in *B. subtilis* is still not clear. To study the regulation of *groE* expression in *B. subtilis*, a summary of the current information in the regulation of the heat shock genes expression in *E. coli* and *B. subtilis* is presented.

I.1.2 Regulation of the Heat Shock Response in *E. coli*

I.1.2.1 Heat shock genes form a regulon in *E. coli*. A set of proteins including GroEL, GroES and DnaK has been identified as heat shock proteins (hsps). Each of these proteins exhibits a characteristic increase in its rate of

synthesis upon the temperature upshift. The increase of hsps synthesis happens immediately after the temperature shift, and reaches its maximum at about 5 minutes, then goes down to a new steady-state rate of synthesis somewhat greater than that at the low temperature (Yamamori *et al.,* 1978; Yamamori *et al.,* 1980). Heat shock genes form a regulon and express at the same time. Sequence alignment of the promoter region of heat shock genes generated consensus -35 and -10 sequences (Figure 1). These conserved promoters were confirmed to be recognized by the σ^{32} containing RNA polymerase (Cowing *et al.*, 1985).

I.1.2.2 σ^{32} is the regulatory factor for the expression of heat shock genes. Genetic experiments found that mutations in *rpoH* (the structural gene for σ^{32}) prevented the transient increase in hsp synthesis, upon the temperature upshift from 30°C to 42°C (Neidherdt and Vanbogelen, 1981). The level of hsps produced depended on the amount of the *rpoH* gene product (Grossman *et al.*, 1984; Skelly *et al.*, 1987). *rpoH* was cloned and characterized (Landick *et al.*, 1984). The amino acid sequence of RpoH shows striking similarity to that of σ^{70}, the major sigma factor in *E. coli.* Therefore, it could be a new sigma factor. *In vitro* transcription study showed that the core RNA polymerase with the purified 32-kDa *rpoH* product could initiate transcription from the promoter of heat shock genes (Bloom *et al.*, 1986; Cowing *et al.*, 1985). These findings indicate that the RpoH or σ^{32} is a sigma factor which positively regulates the heat shock response in *E. coli.*

I.1.2.3 The cellular level of σ^{32} controls the regulation of heat shock response. Several lines of experimental evidence suggested that the

Figure 1. Consensus -35 and -10 sequences in the promoter region of heat shock operons in *E. coli.*

-35 -10

TCTC-CCCTTGAA---13-17 bp----CCCAT-TA

heat shock response of *E. coli* was regulated by changes in the cellular concentration of σ^{32}. 1) Overexpression of *rpoH* gene product without a temperature upshift can increase the synthesis of heat shock proteins(Grossman *et al.*, 1987); 2) Western blot analysis of σ^{32} showed that the σ^{32} level rapidly increased about 20 fold following a temperature upshift for the first 5 minutes, then decreased to a new steady-state level, which is still several fold higher than the level before heat shock (Straus *et al.*, 1987); 3) By using the S-30 extract (*E. coli* extract prepared by centrifuging the cell lyzate at 30,000 g. This extract contains all the components for *in vitro* transcription and translation) for the *in vitro* transcription, the extract prepared from the heat-shocked cells can transcribe heat shock genes about 8 times higher than that from the extract prepared from the non-heat-shocked cells (Skelly *et al.*, 1987). However, what is the control mechanism to modulate the level of σ^{32} in *E. coli* ? It has been demonstrated that σ^{32} is regulated by different mechanisms. Two regulatory mechanisms, stability and synthesis, have been well studied.

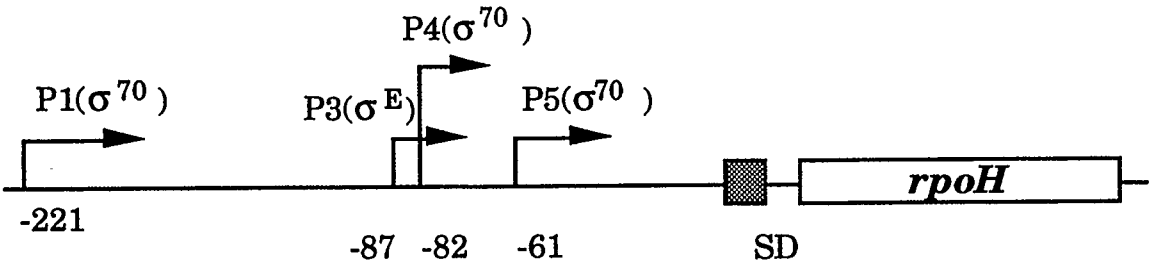
I.1.2.4 Stability of σ^{32}. σ^{32} is very unstable. Its half life is about 1 minute at both 37°C and 42°C (Straus *et al.*, 1987). However, following the temperature shift from 37°C to 42°C, σ^{32} is transiently stabilized in the first 4 minutes ($T_{1/2}$ = 8 mins). With the accumulation of σ^{32} in sufficient amount, this leads to the elevated synthesis of hsps (Straus *et al.*, 1987). Then, σ^{32} is unstable again. Mechanisms controlling the stability of σ^{32} are still not very clear. It was observed that mutations in *dnaK*, *dnaJ* and *groE* resulted in increasing the level of σ^{32} by increasing the σ^{32} stability (Tilly *et al.*, 1983, Tilly *et al.*, 1989). Since none of the above hsps are proteases, it is possible

that DnaK, DnaJ and GroE exert negative regulatory roles by binding to σ^{32} and presenting it to intracellular proteases. Alternatively, Hsps may work indirectly through their potential effects on protease activation. Recently, an essential protein HflB (FtsH), known to control proteolysis of the phage λcII protein, was found to mediate the degradation of σ^{32}. The half life of σ^{32} increased 12 fold in the *hflB* mutant (Herman *et al.*, 1995). *In vitro* experiment also showed that HflB specifically degraded the σ^{32} (Tomoyasu *et al.*, 1995; Herman *et al.*, 1995).

I.1.2.5 Regulation of σ^{32} synthesis

I.1.2.5.1 Transcriptional regulation. Analysis of the promoter region of *rpoH* by S1 nuclease mapping indicates that there are five different transcription start sites. They are transcribed from promoters P1, P2, P3, P4 and P5. Transcription from P2 was confirmed to be an artifact. Thus, there are four promoters in *rpoH* (Erickson *et al.*, 1987, Nagai *et al.*, 1990) (Figure 2). The most distal (upstream) promoter, P1, is the strongest one, and plays a major role under the physiological conditions, whereas others play subsidiary but specific roles. The activity of P4 is enhanced about 2 fold upon temperature upshift from 30° to 42°C. P5 is a weak promoter that is induced by the presence of ethanol or the absence of glucose. P1, P4 and P5 are σ^{70} type promoters. P3 is recognized by another minor sigma factor, σ^{E}. σ^{E} is a 24-kDa sigma subunit of RNA polymerase that is essential for bacterial viability above 42°C. At 50°C, almost all the transcripts of *rpoH* are initiated from P3 promoter. However, only 15% of the *rpoH* transcription is carried out by σ^{E} at 42°C (Erickson *et al.*, 1987). This indicated that σ^{E} was required at much higher growth temperatures. Besides *rpoH*, overproduction of σ^{E}

Figure 2. Organization of promoters in the regulatory region of *rpoH*. SD is the ribosomal binding site. +1 is the transcription start site of *rpoH*.


P4(σ70)
P1(σ70)
P3(σE)
P5(σ70)
rpoH
-221
-87
-82
-61
SD

resulted in the induction of at least 10 proteins. These results indicated that the σ^E dependent genes constitute a second heat shock regulon. Their function may strengthen the effects of the σ^{32} regulon in coping with heat shock or other stresses (Rouviere *et al.*, 1995, Raina *et al.*, 1995). The arrangement of the *rpoH* promoters can allow the cells to maintain thelevel of *rpoH* mRNA under different growth conditions, as well as to maintain the critical levels of *rpoH* transcripts to tolerate some extremely stressful conditions.

I.1.2.5.2 Translational regulation. Several results suggested that the regulation of σ^{32} synthesis may be at the translational level, rather than at the transcriptional level. 1) Heat induced synthesis of *rpoH-lacZ* fusions relies on the specific translation initiation region of *rpoH*, not the promoter region of *rpoH* (Tilly *et al.* 1986, Nagai *et al.*, 1991). 2) Synthesis of σ^{32} after heat shock precedes the accumulation of the *rpoH* mRNA accumulation (Erickson *et al.*, 1987). 3) When RNA synthesis is inhibited by rifampicin, the heat induction of *rpoH-lacZ* fusion proteins still increases 5-fold (Nagai *et al.*, 1991). By extensive analysis of 3'- and 5'- deletion derivatives of *rpoH-lacZ* translational fusions (364 nucleotides of 5' end of *rpoH* fused to *lacZ*), it was found that two regions in the *rpoH* mRNA were involved in the synthesis of σ^{32}. Region A, corresponding to nucleotides 6-20 (The first nucleotide is 'A' at the first codon AUG) functioned as a positive element for *rpoH-lacZ* translation. This sequence shows similarity to the downstream box in gene 0.3 of phage T7. The downstream box can enhances the expression of gene 0.3 and is complementary to nucleotides 1469-1483 of 16S rRNA. It

stimulates translation by stabilizing the binding of the message to the 30S ribosomal subunit (Sprengart *et al* ., 1990). Deletion of this region resulted in the reduction of the expression of *rpoH-lacZ* up to 15 fold. Region B (nucleotides 153-247) functioned as a negative regulatory element. Deletion of this region resulted in the constitutive expression of *rpoH-lacZ* fusion proteins (Nagai *et al.*, 1991). Based on the secondary structure prediction for the *rpoH* mRNA predicted by Nagai *et al*, it was suggested that part of the sequence from region A can form a stem structure with part of the sequence from region B (Nagai *et al.*, 1991). This kind of structure may lead to a sequestration of the initiation codon. The secondary structure of mRNA may be transiently disrupted upon the temperature upshift to allow the release of the translational repression. This model was supported by the observation that under the non-heat shock condition the synthesis of *rpoH-lacZ* increased when deletion and base substitution of regions A and B resulted in preventing the formation of the stem structure.

Another regulation occurs at the posttranslational level. It was found that the *rpoH-lacZ* translational fusions, which carried the region from 364 to 433 (amino acids 122 to 145) could shut off the *rpoH-lacZ* expression after heat induction. A frame shift mutation in this region resulted in constitutive expression of *rpoH-lacZ* fusions following heat induction. When the *rpoH-lacZ* fusion (carrying the fragment covering the region from 364 to 443) was transformed into *dnaK* or *dnaJ* mutant strains, the heat induction of *rpoH-lacZ* occurred normally, but failed to shut off the *rpoH-lacZ* fusion expression. Hence, the peptide (Amino acids 122 to 145 of RpoH) is involved in the interaction with DnaK and DnaJ which may interfere with the expression of *rpoH*.

I.1.2.6 A working model for the heat shock regulation in *E. coli* by σ^{32}. A set of heat shock operons forms a regulon in *E. coli*. The promoters of these heat shock genes are recognized by the σ^{32}-containing RNA polymerase. When the cells are under the heat shock condition, a transient increase in the cellular concentration of σ^{32} leads to a transient increase in the rate of hsps synthesis. σ^{32} is transcribed by σ^{70} containing RNA polymerase under heat shock and non-heat shock conditions. Under the non-heat shock condition, expression of *rpoH* is repressed at the translational level, σ^{32} is very unstable through binding to DnaK, DnaJ that may deliver σ^{32} to the protease, HflB, to be degraded. Following the temperature upshift: 1) Synthesis of σ^{32} is induced through the transient disruption of the σ^{32} mRNA secondary structure; 2) In the first several minutes, the amount of DnaK and DnaJ decreases due to the binding to the unfolded/denatured proteins. σ^{32} will become more stable. Thus, the cellular level of σ^{32} is increased. An increase in the concentration of σ^{32} is followed by a proportional increase in the transcription of heat shock genes, resulting in a transient increase in the rate of hsps synthesis.

I.1.3 Regulation of Heat Shock Response in *B. subtilis*

I.1.3.1 The heat shock response in *B. subtilis*. *Bacillus subtilis* can induce as many as 66 proteins with a temperature upshift from 37 to 48°C (Brian *et al.*, 1991). According to the induction condition, heat shock proteins can be classified into two main groups: specific heat shock proteins which are specifically induced by heat shock and general stress proteins that can be induced by heat shock and other stresses such as salt stress, glucose and

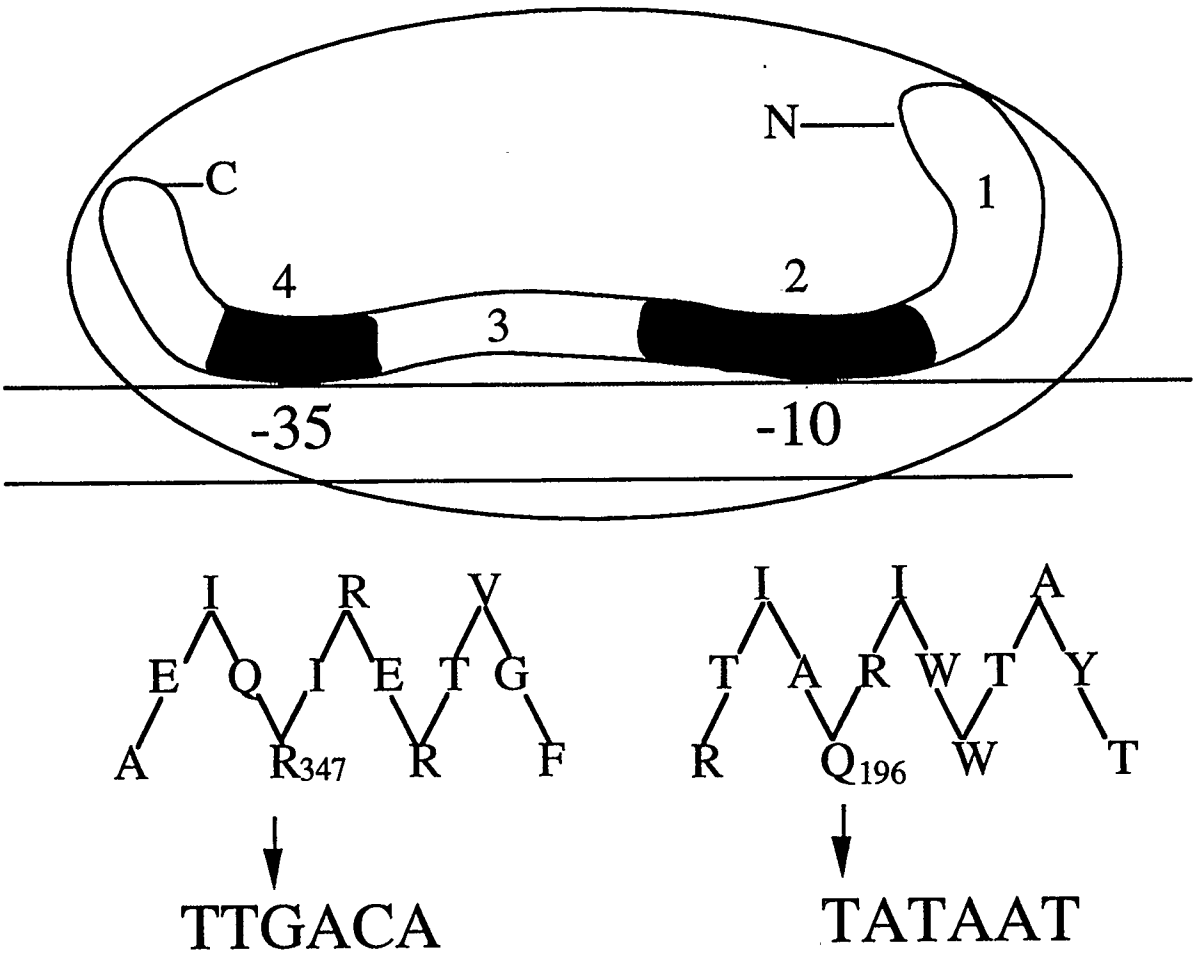
oxygen limitation or oxidative stress, *etc.*(Volker *et al.*, 1994). Many of these heat shock proteins exhibit functional and structural similarity to their homologs in *E. coli*. As discussed early, the heat shock response in *E. coli* is regulated by a minor sigma factor, σ^{32}. Early studies on the regulation of heat shock response in *B. subtilis* were focused on the following questions: Does *B. subtilis* use a similar mechanism to regulate the heat shock response? Which sigma factor mediates the regulation of the expression of the heat shock genes in *B. subtilis* ?

I.1.3.2 **σ factor in *B. subtilis*.** Regulation of gene expression in *B. subtilis* occurs primarily at the level of transcription. The specificity of transcription relies on interactions between RNA polymerase and the promoter sequence. RNA polymerases are isolated in two forms: the core RNA polymerase (which contains several subunits, α, β, and β') and the holoenzyme. The core RNA polymerase can catalyze the polymerization of ribonucleotides into RNA, whereas the σ factor recognizes and interacts with the specific promoter sequences. The holoenzyme is the complex of the core polymerase with the σ factor. There are at least 10 σ factors (σ^A to σ^L) in *B. subtilis*. Each σ factor recognizes different promoter sequence. The specific promoter sequences are usually located at the -35 and -10 regions of the promoter with +1 as the transcription start site. σ^A, the major sigma factor which is functionally equivalent to σ^{70} in *E. coli*, is responsible for the expression of housekeeping genes. σ^B is responsible for transcribing some of the heat shock genes. The function of σ^C is not clear. σ^D is involved in flagellar synthesis (Helmann, 1988). σ^E, σ^F, σ^G , σ^H, and σ^K are involved in the sporulation process. σ^L regulates the expression of the levanase operon (Haldenwang, 1995).

I.1.3.3 The Structure of σ factors. Sequence comparison of the known sigma factors at the protein level reveals that four regions are conserved among these proteins. These regions are numbered from 1 to 4, beginning at the N-terminal end of the protein. The function of region 1 and 3 is not clear. Region 2 is responsible for the interaction with the -10 region of the promoter, and region 4 interacts with the -35 region. (Helmann and Chamberlin, 1988). Among the σ factors in *B. subtilis*, σ^A has been well characterized. Figure 3 illustrates the critical amino acid residues in σ^A that can interact with the -35 and -10 regions of the σ^A - dependent promoter.Amino acid conversions at region 2 and region 4 can change the specificity of σ^A to recognize the mutated promoter sequences. Amino acid substitutions at 347 (R to H) and 196 (Q to R) can restore the reduced σ^A-type promoter activity caused by mutations at the third nucleotide (G to A) in the -35 region and the first nucleotide (T to C) in the -10 region, respectively (Kenney and Moran, 1991).

I.1.3.4 σ^D (σ^{28}) does not transcribe the heat shock genes in *B. subtilis*. Three pieces of evidence indicated that the regulatory mechanism of heat shock gene induction in *B. subtilis* was different from that in *E. coli* : 1) A minor factor σ^D (σ^{28}) (Gilman *et al.*, 1981) in *B. subtilis* can initiate transcription for the *E. coli* σ^{32} specific promoters *in vitro* (Briat *et al.*, 1985). However, the cell with disrupted σ^D gene (*rpoD*) showed a normal heat shock response (Helmann *et al.*, 1988); 2) Promoters of the major *E. coli* heat shock genes are not functional in *B. subtilis* (Wetzstein and Schumann, 1990); 3) Several *B. subtilis* heat shock operons have been cloned and characterized. None of the characterized promoters showed a consensus sequence recognized by σ^{32} (Li and Wong, 1992; Schmidt *et al*, 1992; Wetzstein *et al*, 1992).

Figure 3. Interaction between σ^A and the σ^A dependent promoter. The sigma factor may lie between RNA polymerase and the promoter, where the sigma factor specifically contacts with the promoter. Sigma A factor have four domains, indicated as regions 1 to 4. Amino acids in region 2 and region 4 contact the - 10 and -35 regions of the promoter, respectively. A mutation at region 2 (Q-196-R) can suppress the effect of a single base-pair transition (T to G or C) at the first position of the -10 hexamer of the promoter. A mutation at region 4 (R-347-H) can suppress the effect of a substitution (G to A or C) at the third base pair of the -35 hexamer. This diagram is prepared to the illustrate of the contacts between SigA and the SigA dependent promoter based on the genetic data currently available (see text, Kenney and Moran, 1991).


N
C
1
4
2
3
-35
-10
I R V
E Q I E T G
A R347 R F
TTGACA
I I A
T A R W T Y
R Q196 W T
TATAAT

I.1.3.5 σ^B **controls the expression of the general stress heat shock genes in *B. subtilis*.** Although σ^B has been found more than ten years ago, its physiological role remains obscure. Null mutations in the σ^B structural gene (*sigB*) have no obvious effect on vegetative growth or sporulation. In 1993, Benson & Haldenwang reported that the cellular level of σ^B increased 5 to 10 fold when the cell culture was shifted from 37°C to 48°C (Benson and Haldenwang, 1993). At least 5 known operons were found to have the σ^B-dependent promoter preceding the genes which encoding the general stressproteins (Figure 4). The major group of general stress proteins (about 40 hsps) failed to be induced following heat shock in a *sigB* mutant (Volker *et al*, 1994). These results suggest that σ^B is responsible for the induction of a major group of general stress proteins by heat stress as well as by other stimuli. Besides this mechanism, another regulatory system may exist for the expression of some general stress protein operons since: 1) ClpC and LonA that are the heat inducible ATP-dependent proteases that show a normal heat induction in the null σ^B mutant; 2) Both *clpC* and *lonA* have a σ^A-type promoter in their regulatory region (Kruger *et al.*, 1994; Riethdorf *et al.*, 1994).

Figure 4. Consensus sequences of the SigB-dependent promoters.

-35 -10

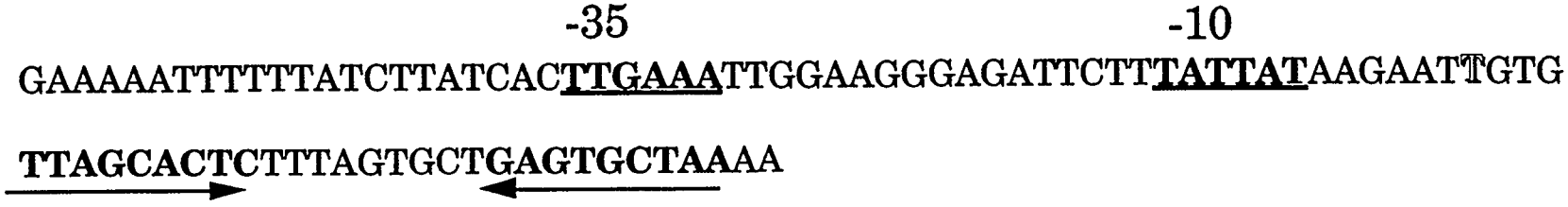
-----GTTTAA-------12-14 bp-----GGGGTAT---

I.1.3.6 σ^A may transcribe *groE* and *dnaK*. The *groE* operon was isolated and characterized in our lab and Schumann's lab independently (Li and Wong, 1992; Schmidt *et al.*, 1992). Two genes, *groES* and *groEL*, are located in the operon. Analysis of the promoter region in the *groE* operon showed that a typical σ^A-dependent promoter sequence could be found in *groE*. The spacer region is 17 bp long between the putative -35 and -10 sequences (Figure 5). The transcription start site of *groE* was mapped by primer extension analysis under both heat shock and non-heat shock conditions. The same transcription start site was observed under these conditions. A σ^A-type promoter was also found in the *dnaK* operon (Figure 5) (Wetzstein *et al.*, 1992). This heat inducible operon consists of at least four genes (*orf* 39, *grpE*, *dnaK*, *dnaJ*). The products encoded by *grpE, dnaK and dnaJ* are molecular chaperones. The function of ORF39 is unknown. No homolog of ORF39 has been found in *E. coli*.

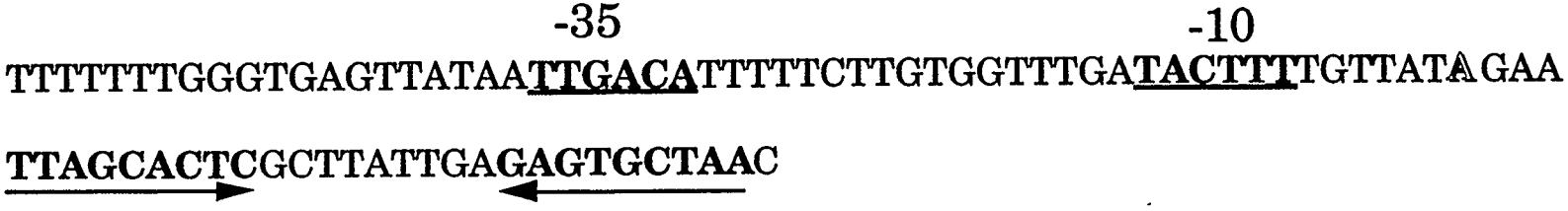
The following findings indicate that σ^A may transcribe *groE* and *dnaK*. 1) Expression of *groE* and *dnaK* was drastically reduced in a temperature-

Figure 5. Organization of the promoter regions from *groE* and *dnaK*. The arrow indicated the inverted repeat sequence. The -35 and -10 regions of promters are underlined. The transcription start sites of *groE* and *dnaK* are outlined.

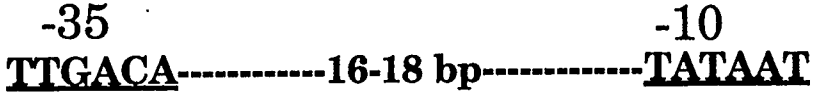
groE promoter sequence



dnaK promoter sequence



Consensus sequences of SigA dependent promoter



sensitive *sigA* (the structural gene for σ^A) mutant after temperature upshift; 2) σ^A with the core RNA polymerase can initiate transcription from the *groE* promoter *in vitro* (Chang *et al.*, 1994). Since σ^A is the major sigma factor in *B. subtilis* and its expression shows no obvious increase following the temperature upshift, some other regulatory factors should exist to mediate the heat inducible expression of both *groE* and *dnaK*.

I.1.3.7 An inverted repeat sequence (CIRCE) may be the key element to regulate the expression of both *groE* and *dnaK* in *B. subtilis* and many other eubacteria. With further analysis of the promoter of *groE*, a inverted repeat sequence (IR) was found downstream of the transcription start site. A similar IR sequence was found in the *dnaK* operon, which also located downstream of the transcription start site of the *dnaK* operon (Figure 4). It suggests that these two operons are likely to be regulated by the same mechanism. Besides *B. subtilis*, the conserved inverted repeats can also be found in *groE*, *dnaK*, and *dnaJ* operons in many othe gram-positive and gram-negative bacteria. Table 1 lists the sequences of all the published IR elements. By comparing all these sequences, a consensus sequence of the inverted repeat can be generated as follows: TTAGCACTC-N9-GAGTGCTAA. In most cases, this element is usually located downstream of the transcription start site. These inverted repeats may serve as a common regulatory element in most eubacteria to control the expression of these genes under the heat shock condition and is named "CIRCE" (controlling inverted repeat of chaperone expression) (Zuber and Schumann, 1994).

Table1: Consensus sequences of the IR sequences

Bacterial Species	Gene	Position [a]	Sequences
Agrobacterium tumefaciens	*groESL*	+2	CTGGCACTC-N9-GAGTGCTAA
Bacillus stearothermophilus	*groESL*	+3	TTAGCACTC-N9-GAGTGCTAA
Bacillus subtilis	*groESL*	+3	TTAGCACTC-N9-GAGTGCTAA
Brucella abortus	*groESL*	ND [b]	TTAGCACTC-N9-GAGTGCTAA
Clostridium acetobutylicum	*groESL*	+2	TTAGCACTC-N9-GAGTGCTAA
Clostridium acetobutylicum	*dnaK*	+6	TTAGCACTC-N9-GAGTGCTAA
Clostridium perfringens	*groESL*	ND	TTAGCACTC-N9-GAGTGCTAA
Chlamydia pneumoniae	*groESL*	ND	TTAGCACTT-N9-GAGTGCTAA
Chlamydia psittaci	*groESL*	ND	GTAGCACTT-N9-AAGTGCTAA
Chlamydia trachomatis	*dnaK*	ND	CTAGCACTC-N9-GAGCGCTAA
Chromatium vinosum	*groESL*	ND	TTAGCACTC-N9-GAGTGCTAA
Erysipelothrix rhusiopathiae	*dnaK*	ND	TTAGCACTC-N9-TATTGCTAA
Lactococcis lactis	*groESL*	ND	TTAGCACTC-N9-GAGTGCTAA
Lactococcis lactis	*dnaK*	ND	TTAGCACTT-N9-GAGTGCTAA
Lactococcis lactis	*dnaJ*	-38	TTAGCACTC-N9-GAGTGCTAA
Leptospira interrogans	*groESL*	+6	TAAGCACTC-N9-TAGTGCTAA
Mycobacterium bovis BGC	*groES*	ND	CTAGCACTC-N9-GAGTGCTAG
Mycobacterium bovis BGC	*groEL*	ND	CTTGCACTC-N9-GAGTGCTAA
Mycobacterium leprae	*groESL*1	ND	CTAGCACTC-N9-GAGTGCTAG
Mycobacterium leprae	*groESL*2	ND	ATTGCACTC-N9-GAGTGCTAA
Mycobacterium tuberculosis	*groESL*	ND	CTTGCACTC-N9-GAGTGCTAG
Mycobacterium pneumoniae	*dnaK*	ND	TTAGCACTC-N9-AAGTGATAA

Mycobacterium pneumoniae	*dnaJ*	ND	TTAACACTC-N9-AAGTGCTAA
Rhizobium meliloti	*groESL*	ND	TTAGCACTC-N9-GAGTGCTAA
Rhizobium leguminosarum	*groESL*1	ND	CTAGCACTC-N9-GAGTGCTAA
Rhizobium leguminosarum	*groESL*2	ND	ATGGCACTC-N9-AAGTGCGGC
Rhizobium leguminosarum	*groESL*3	ND	TTGGCACTC-N9-GAGTGCCAC
Staphylococcus aureus	*groESL*	ND	TTAGCACTC-N9-AAGTGCTAA
Staphylococcus aureus	*dnaK*	ND	TTAGCACTT-N9-GAGTGCTAA
Streptomyces albus	*groESL*1	-48	TTGGCACTC-N9-GAGTGCTAA
Streptomyces albus	*groESL*1	-2	CTGGCACTC-N9-GAGTGCCAA
Streptomyces albus	*groESL*2	-35	CTTGCACTC-N9-GAGTGCTAA
Streptomyces albus	*groESL*2	+2	TTAGCACTC-N9-GAGTGACAG
Streptomyces coelicolor A3	*groESL*1	-38	TTGGCACTC-N9-GAGTGCTAA
Streptomyces coelicolor A3	*groESL*1	+1	CTGGCACTC-N9-GAGTGCCAA
Streptomyces coelicolor A3	*groEL*2	-33	CTGGCACTC-N9-GAGTGCCAA
Streptomyces coelicolor A3	*groEL*2	+1	TTAGCACTC-N9-GAGTGCTAA
Synechococcus sp. 6391	*groESL*	ND	TTAGCACTC-N9-GAGTGCTAA
Synechococcus sp. PCC7942	*groESL*	-8	TTAGCACTC-N9-GAGTGCTAA
Synechococcus sp. PCC6803	*groESL*1	ND	TTAGCACTC-N9-GAGTGCTAA
Synechococcus sp. PCC6803	*groESL*2	ND	TTAGCACTC-N9-GAGTGCTAA
Zymomonas mobilis	*groESL*	ND	TTGGCACTC-N9-GAGTGCCAG
Consensus sequences			**TTAGCACTC-N9-GAGTGCTAA**

[a] +1 is the transcription start site.

[b] ND: the transcription start site is not determined.

I.1.3.8 The putative functions of CIRCE. Site-directed mutagenesis of CIRCE in *B. subtilis dnaK* operon resultes in a higher level of *dnaK* expression at 37°C, indicating that this inverted repeat sequence acts as a negative regulatory element (Zuber and Schumann, 1994). It is possible that CIRCE can exert its negative regulatory effect at the DNA level as a repressor binding site or at the RNA level by affecting the mRNA stability or acting as a transcription terminator. In addition, CIRCE can possibly function at both levels. Besides the negative regulation, it is not certain whether other mechanisms (such as positive factors) are involved in regulating the heat induction of these operons.

I.1.3.9 A model for the regulation of *groE* in *B. subtilis*. From the above mentioned information, a regulatory model for *groE* expression was proposed. Transcription of *groE* is directed by the σ^A containing RNA polymerase under both heat shock and non-heat shock conditions. CIRCE serves as a negative regulatory element that is responsible for the low level expression of *groE* under the non-heat shock condition. This mechanism may also regulate the expression of *dnaK*. After the temperature upshift, the putative negative factor (possibly a repressor) is somehow inactivated, therefore, the expression of *groE* and *dnaK* is increased.

I.2 Objectives

To determine the regulatory mechanism controlling *groE* and *dnaK* expression in *B. subtilis*. I have the following objectives.

I.2.1 To determine whether *groE* is transcribed by the σ^A containing RNA polymerase

As has already been discussed, the σ^A containing RNA polymerase may transcribe *groE*. However, there is no direct genetic evidence to demonstrate that *groE* transcription is indeed transcribed by the σ^A containing RNA polymerase. To address this problem, the putative *groE* promoter was mutated at both the -10 and -35 regions. The effect of these mutations on the promoter activity and the use of specific *sigA* mutants to suppress these mutations in an allele-specific manner are examined. The details of the studies is presented in Chapter IV.

I.2.2 To determine the functional roles of the IR element (CIRCE) in the regulation of *groE* expression.

Since CIRCE is located downstream of the *groE* transcription start site, this sequence is present in the *groE* mRNA. It is possible that CIRCE can exert its function at either the RNA level or the DNA level. To study the function of CIRCE, half-lives of the *groE* messenger RNA under both 37°C and 42°C were determined. Effects for the presence and absence of CIRCE in *groE* on the stability of the *groE* transcripts were also examined. The effect of varying the distance between the *groE* transcription start site and CIRCE was systematically studied. The results are presented in the Chapter V.

I.2.3 To develop a genetic system to study the structural gene encoding a negative regulatory factor controlling *groE* expression

A negative factor which bind to the CIRCE sequence may regulate the *groE* and *dnaK* expression. Therefore, one of my objectives is to identify this

factor. To identify this gene, I would like to isolate regulatory mutants which can constitutively express *groE* and *dnaK*. Once isolated, these regulatory mutants can be used to clone and characterize the gene for the negative factor. Results of these studies are described in Chapter VI.

II Materials and Methods

II.1 Biochemical reagents

Restriction endonucleases were purchased from Bethesda Research laboratories (BRL), Pharmacia, and New England Biolabs Inc. Klenow fragment, T4 Kinase, Calf Intestinal Alkaline Phosphatase, T4 DNA Ligase were obtained from BRL. DNA sequencing Kits, λ DNA, dNTP, Taq DNA polymerase and AMV Reverse Transcriptase were purchased from Pharmacia Inc. [α-^{35}S]-dATP (400 Ci mmol^{-1}; 10 mCi ml^{-1}), [γ-^{32}P]-dATP (3000 Ci mmol^{-1}; 10 mCi ml^{-1}) and [α-^{32}P]-dATP (3000 Ci mmol^{-1}; 10 mCi ml^{-1}) were obtained from either Amersham Corp or ICN. Bacterial growth media were purchased from Difco. The DNA primers for PCR and sequencing were synthesized at the University Core DNA Service, the University of Calgary. Agarose, acrylamide, and bis-acrylamide were purchased from BRL Inc. Nitrocellulose and Hybond-N hybridization transfer membranes were acquired from Amersham Corp. Low-molecular-weight protein electrophoresis markers were obtained from Bio-Rad Laboratories. All other chemical reagents were obtained from Sigma, ICN and Fisher Scientific Company.

II.2 Bacterial strains and plasmids

B. subtilis strains used in this study are listed in Table 2. Plasmids used in this study are described in Table 3. EU9001 *tms*, EU9002 *tms*-13C, EU9004 *tms*-35A (Kenney & Moran, 1991) are the *B. subtilis* strains used for the *in vivo* study to demonstrate that the expression of *groE* is directed by SigA. They carry either an extra copy of the wild type or mutated *sigA* under

the control of an IPTG inducible promoter, P*spac*. These strains also carry either the wild type *tms* promoter or the mutated (-13C or -35A) *tms* promoter fused with *lacZ*. *Bacillus subtilis* 168 (*trpC*2) and *E.coli* DH5α [(ϕ80*lacZ*dΔM15) *endA*1 *recA*1 *hsd*17 (r^-m^-) *supE*44 *thi*-1 λ^- *gyrA relA*1 F^- Δ (*lacZYA-argF*) *U* 169] served as hosts for routine transformations. *B. stearothermophilus* ATCC 7954 (strain code: 9A1) was obtained from the Bacillus Genetic Stock Center for the amplification of *bgaB* (ß-galactosidase gene). The *E. coli* Bluescribe plasmid (pBS, from Stratagene) and *B. subtilis* plasmid pUB18 were used for routine subcloning of DNA fragments and sequencing.

II.2.1 Construction of pDL and pDK

To have a heat stable reporter to monitor gene expression under the heat shock condition, the *E. coli lacZ* gene in pDH32M (Kraus, 1994, see Table 3) was replaced by *bgaB*, a thermostable ß-galactosidase gene from *B. stearothermophilus*. The promoterless *bgaB* with the sequence corresponding to nucleotides 395-2478 of the reported sequence (Hirata, 1986) was generated through PCR amplification. A *Sna*BI site and a *Sst*I site were introduced at the 5' and the 3' end of the fragment, respectively. The *Sna*BI and *Sst*I double-digested *bgaB* fragment was inserted to the *Sna*BI-*Sst*I digested, 7.6-kb pDH32M to generate pDL. To construct pDK, the *Eco*RI-*Sal*I fragment containing the *cat* gene in pDL was replaced by an *Eco*RI-*Sal*I fragment carrying the P43-Kanamycin cassette. The *kan* gene was amplified through PCR using pUB18 (Wang and Doi, 1987) as the template.

Table 2: *B. subtilis* and *E. coli* strains used in this study

Strain	Genotype and derivation, comments and reference.
B. subtilis 168	*trpC2* (Spizizen,1958).
B. subtilis PY852	*chr*:: Tn 917 Ω HU 160 *glyC*, containing a silent Tn917 transposon insertion (Kenney & Moran, 1991).
B. subtilis EU9000	PY852 *ermC::cat,* a derivative of PY852 which the ermC gene contained within the transposon was replaced with the *cat* (Kenney & Moran, 1991).
B. subtilis EU9001	EU9000 *Pspac-sigAwt,* a derivative of EU9000 which contains a wild type *sigA* gene controlled by an IPTG inducible promoter, P*spac,* (Kenney & Moran, 1991).
B. subtilis EU9002	EU9000 *Pspac-sigA* Q-196-R, a derivative of EU9000 which contains a mutated *sigA* allele controlled by P*spac* (Kenney & Moran, 1991).
B. subtilis EU9004	EU9000 *Pspac-SigA* R-347-H, a derivative of EU9000 which contains a mutated *sigA* allele controlled by P*spac* (Kenney & Moran, 1991).
B. subtilis EU9001*tms*	EU9001 SPß::*tms-lacZ* , a derivative of EU9001 which contains a *lacZ* controlled by the *tms* promoter (Kenney & Moran, 1991). SPß is a specialized transducing phage.
B. subtilis EU9002*tms*-13C	EU9002 SPß::*tms-13C-lacZ*, a derivative of EU9002 which contains *lacZ* controlled by the mutated *tms* promoter (Kenney & Moran, 1991).
B. subtilis EU9004*tms*-35A	EU9004 SPß::*tms-35A-lacZ,* a derivative of EU9004 which contains a *lacZ* controlled by the mutated *tms* promoter (Kenney & Moran, 1991).
B. subtilis WBG1	*B. subtilis* 168 containing a pDL plasmid integration in the *amyE* locus (This work).
B. subtilis WBG2	*B. subtilis* 168 containing a pDL2 plasmid integration in the *amyE* locus (This work).
B. subtilis WBG3	*B. subtilis* 168 containing a pDL3 plasmid integration in the *amyE* locus (This work).

B. subtilis WBG4	*B. subtilis* 168 containing a pDL4 plasmid integration in the *amyE* locus (This work).
B. subtilis WBG5	*B. subtilis* 168 containing a pDL5 plasmid integration in the *amyE* locus (This work).
B. subtilis WBG6	*B. subtilis* 168 containing a pDL6 plasmid integration in the *amyE* locus (This work).
B. subtilis WBG7	*B. subtilis* 168 containing a pDL7 plasmid integration in the *amyE* locus (This work).
B. subtilis WBG8	*B. subtilis* 168 containing a second copy of *groE*. The IR sequence in this copy of *groE* is deleted. It was integrated in the *amyE* locus (This work).
B. subtilis WBG15	*B. subtilis* 168 containing a pDL15 plasmid integration in the *amyE* locus (This work).
B. subtilis WBG16	*B. subtilis* 168 containing a pDL16 plasmid integration in the *amyE* locus (This work).
B. subtilis WBG17	*B. subtilis* 168 containing a pDL17 plasmid integration in the *amyE* locus (This work).
B. subtilis WBG101	Same as WBG2, except a point mutation in *orf39,* the first gene in *dnaK* operon (This work).
E.coli DH5α	ϕ80*lacZ*dΔM15 *endA*1 *recA*1 *hsd*17 (r^-m^-) *supE*44 *thi*-1 λ^- *gyrA relA*1 F^- Δ (*lacZYA-argF*) *U*169].
E.coli BL21 (DE 3)	F^- *ompT lonA* r_B^- m_B^-, this strain has an integrated structural gene encoding T7 RNA polymerase controlled by an IPTG inducible promoter (Novagen, Inc).

II.2.2 Construction of plasmids pDL2 and pDL3

DNA fragments carrying the *groE* promoter with (-238 to +36, +1 is the transcription start site) or without the inverted repeat sequence (-238 to +2) were generated through PCR amplification using *B. subtilis* chromosomal DNA as the template. An *Eco*RI site and a *Bam*HI site were introduced at the 5' and 3' ends, respectively, so that these fragments can be inserted into *Eco*RI and *Bam*HI digested pDL to generate pDL2 and pDL3. The resulting

plasmids were linearized by digesting with *Pst*I and integrated into *amyE* locus in *B. subtilis* to generate WBG2 and WBG3.

Table 3: Plasmids used in this study

Plasmid	Comment and reference
pBS	*E. coli* cloning vector, bluescribe plasmid from Strategene.
pBS-GROEP	containing a 272-bp *E.co*RI-*Bam*HI *groE* promoter (-238 to +36) fragment in pBS (This work).
pBS-MTL	Insertion of a 1.5-kb *Hind*III-*Bgl*II fragment carrying part of the *B. subtilis* mannitol operon (This work).
pBS-P43ORF39	Insertion of a 1.4-kb *Eco*RI-*Hind*III fragment carrying the P43-ORF39 expression cassette in pBS. This insert is from pUBP43ORF39 (This work).
pDH32M	A *B. subtilis* integration vector with the promoterless *lacZ* as the reporter gene. This *lacZ* cassette with upstream *cat* gene is flanked by the front and back portions of *amyE* which allows the integration of the transcriptional fusion at the *amyE* locus. This plasmid also carries the *cat* gene as the selection marker after integration. The pDH32M carries the replication origin of pBR322 and *bla* gene (Kraus et al., 1994).
pDL	A derivative of pDH32M. It is a *B. subtilis* integration vector with *bgaB* as the reporter gene (This work).
pDL2	Insertion of a 274-bp *Eco*RI-*Bam*HI *groE* promoter fragment (with the IR sequence) in pDL. This fragment was generated by PCR (This work).
pDL3	Same as pDL2 except that no inverted repeat sequence in the *groE* promoter region (This work).
pDL4	Same as pDL 2 except that a 5-bp fragment was inserted between the *groE* transcription start site and the inverted repeat sequence (This work).

pDL5	Same as pDL 2 except that a 9-bp fragment was inserted between the *groE* transcription start site and the inverted repeat sequence (This work).
pDL6	Same as pDL 2 except that a 17-bp fragment was inserted between the *groE* transcription start site and the inverted repeat sequence (This work).
pDL7	Same as pDL 2 except that a 21-bp fragment was inserted between the *groE* transcription start site and the inverted repeat sequence (This work).
pDL15	Insertion of a 46-bp *Eco*RI-*Bam*HI *groE* promoter fragment (-44 to +2) in pDL (This work).
pDL16	Insertion of a 114-bp *Eco*RI-*Bam*HI *groE* promoter fragment (-112 to +2) in pDL (This work).
pDL17	Insertion of a 185-bp *Eco*RI-*Bam*HI *xyl* (xylose operon) promoter fragment (-180to +5) in pDL (This work).
pET29b	An overexpression *E. coli* plasmid with T7 promoter (Novagen)
pET29b (ΔXB)	A derivative of pET29b except that the sequence from *Xba*I to *Bgl*II was deleted through restriction digestion and end reparing (filling in)(This work)
pK-GRO(+)	Insertion of the 2.3-kb *Sst*I fragment carrying the *groE-bgaB* transcription fusion from pDL2. The inserted *groE* regulatory region contains both the *groE* promoter and the inverted repeat sequence (This work).
pK-GRO(-)	Insertion of the 2.3-kb *Sst*I fragment carrying the *groE-bgaB* transcription fusion from pDL3. The inserted *groE* regulatory region contains only the *groE* promoter (This work).
pK184	An *E. coli* vector with the p15a replicon and a kanamycin resistance marker (Jobling, 1990).
pUB-P43	Insertion of a 300-bp *Eco*RI-*Kpn*I fragment carrying the *B. subtilis* P43 promoter in pUB18 (Li and Wong, 1992).

pUB-P43ORF39	Insertion of a 1.1-kb *Kpn*I-*Xba*I fragment carrying the promoterless *orf39* gene (nucleotide 221-1319 of the reported sequences) in pUB-P43. This insert was generated through PCR (This work)
pUB18	*B. subtilis* cloning vector; Km^r (Wang, 1987).

II.2.3 Construction of pDL4 and pDL7

To study the effects of varying the distance between the *groE* transcription start site and the inverted repeat sequence (CIRCE) on gene expression, a *Bgl*II site was introduced between nucleotides +2 and +3 of the *groE* promoter region through the PCR based site-directed mutagenesis using pBS-GROEP as the template. The primers for this study were designated ESL *Bgl*II and ESL10L, respectively. The mutated *groE* promoter was inserted into pDL to generate pDL4. This plasmid was digested by *Bgl*II, followed by filling in and ligation to generate pDL5. pDL6 was generated by inserting an 8-bp *BclI* linker to the *Bgl*II digested, end-repaired (by the fill-in reaction) pDL4. pDL6 was cut by *Bcl*I, followed by filling in and ligation to generate pDL7. These plasmids resulted in inserting 5, 9, 17 and 21 bp between nucleotides +2 and +3 in the *groE* regulatory region, respectively. These plasmids were linearized by *Pst*I and integrated to the *amyE* locus in *B. subtilis* 168 to generate WBG4 to WBG7.

II.2.4 Construction of pDL15 and pDL16

Different versions of the *groE* promoter were generated either by annealing the synthetic oligonucleotides (-44 to +2) or by PCR (-112 to +2). These promoter sequences were inserted between *EcoRI* and *BamHI* sites in

the integrated vector pDL to generate *groE-bgaB* transcription fusions. The resulting vectors were pDL15 and pDL16. Integration of these plasmids to *B. subtilis* 168 generated in *B. subtilis* strains, WBG15 and WBG16, respectively.

II.2.5 Construction of pDL17

The DNA fragment covering *B. subtilis xyl* (Wilhelm and Hollenberg, 1985) promoter (xylose operon from -180 to +5, +1 is the transcription start site) was generated through PCR amplification using the *B. subtilis* chromosome DNA as the template. The amplified fragment was inserted to pDL at *Eco*RI and *Bam*HI sites. The resulting plasmid was designated pDL17. Integration of *Pst*I linearized pDL17 to *B. subtilis* 168 generated WBG17.

II. 2.6 Introduction of a copy of the modified *groE* at the *amyE* locus

To study whether the IR sequences has any effects on the stability of *groE* transcript, *B. subtilis* carrying a modified *groE* in the *amyE* locus was created. This modified *groE* does not carry the IR sequence in the *groE* regulatory region. To construct this strain, a 2.0-kb *Bam*HI-*Sst*I fragment carrying *bgaB* in pDL16 was replaced by a 2.1-kb *Bam*HI-*Sst*I fragment carrying *groESL* (+33 to +2165, generated by PCR). The ligation mixture was digested by *Sca*I to linearize the ligated plasmid and transformed to *B. subtilis* 168. The resulting strain is designated WBG8.

II.2.7 Construction of pUB-P43ORF39, and pBS-P43 ORF39

To study the function of *orf39*, the first gene in the *dnaK* operon, pUB-P43ORF39 and pBS-P43ORF39 were constructed. The DNA fragment containing the promoterless *orf39* was generated by PCR amplification with the introduction of *KpnI* and *XbaI* sites at the 5' end and 3' end, respectively. The *Kpn*I and *Xba*I digested DNA fragment was then ligated to pUB-P43 digested with the same restriction enzymes. The resulting plasmid was designated pUB-P43ORF39. In this plasmid, the expression of *orf39* was controlled by the *B. subtilis* P43 promoter, a constitutively expressed promoter (Li and Wong, 1992). To generate pBS-P43ORF39, the 1.5-kb *Eco*RI and *Hind*III digested pUB-P43ORF39 carrying the P43-*orf39* cassette was inserted into pBS digested by *Eco*RI and *Hind*III.

II.2.8 Construction of pET-ORF39

To overproduce ORF39, pET-ORF39 was constructed. pBS-P43ORF39 was digested with *Kpn*I and *Xba*I. The 1.2-kb fragment was then inserted into pET29b(ΔXB) digested by the same enzymes. The resulting plasmid was designated pET-ORF39. The expression of *orf39* is controlled by an IPTG inducible T7 promoter. This plasmid was used for the overexpression of ORF39.

II.2.9 Construction of pK-BGAB, pK-GRO(+) and pK-GRO(-)

In order to study the function of *orf39*, a binary plasmid system was developed. The first plasmid was pBS-P43ORF39, which is a high copy number plasmid in *E. coli*. The second plasmid was the derivatives of pK184 (Jobling *et al.*, 1990, see Table 2). pK184 is a low copy number (5-15)

plasmid. To generate pK-GRO(+), the 2.4-kb *Sst*I digested pDL2, which carries the *groE* promoter with IR and *bgaB*, was inserted into the *Sst*I digested pK184. pK-GRO(+) was then cut by *Kpn*I to remove the *groE* promoter and the IR sequence to produce plasmid pK-BGAB. To generate pK-GRO(-), pDL3 was digested by *Sst*I. The 2.4-kb fragment which contains the *groE* promoter without the IR sequence was insert to pK184 at the *Sst*I site.

Table 4 : DNA primers used in this study

Primer	Gene	Location	Sequence
ESL 10U	*groE*	-22 to +2 [a]	GGAGATTCTTCATTATAAGAA TTG
ESL 10L	*groE*	-23 to -43 [a]	CTTCCAATTTCAAGTGATAAG
ESL 35U	*groE*	-22 to +2 [a]	GGAGATTCTTTATTATAAGAAT TG
ESL 35L	*groE*	-23 to -43 [a]	CTTCCAATTTTAAGTGATAAG
ESL BglII	*groE*	-22 to +13 [a]	GGAGATTCTTTATTATAAGAAT TGAGATCTGTTAGCACTC
GROEP	*groE*	-115 to -95 [a]	GGGGGAGCTCTTATGCATATT ATGTTGCCAACTGTCGG
GROEP1	*groE*	+2 to -22 [a]	GGAGATCTGGATCCAATTCTT ATAATAAAGAATCTC
GROEU	*groE*	+32 to +55 [a]	GGGGATCCAATTACATATTCA TACTATTGAGG
GROEB	*groE*	+2166 to +2150 [a]	GGGGTACCGAGCTCATATGAC CGCTAATATGG
GROPE	*groE*	+99 to +81 [a]	CAATGACAACGCGATCACC
GROESX	*groE*	+350 to +333 [a]	GAATTAGCCGATAACAGC

BTBF	*bgaB*	465 to 502	GGGGTACGTAGGTACCTAAAT TTTAACTTAATTTATAATTAAA CGAAAATTAG
BTBB	*bgaB*	2506 to 2479	GGGGGAGCTCTTATGCATATT ATGTTGCCAACTGTCGG
BSXYF	*xyl*	-180 to -156	GGGGAATTCATGAAAAACTAA AAAAAAATATTG
BSXYB	*xyl*	+5 to -17	GGGGGATCCATTTTATGTCAT ATTGTAAG
BGABPE	*bgaB*	492 to 475	GGGTTATAATCTCCTCCG
ORF39F	*orf39*	221 to 244	GGGGTACCGTGCTAACAGAGG TGATGATG
ORF39B	*orf39*	1319 to 1299	GCTCTAGAGGAGAAGGACAGG TGCTGCCC

[a] +1 is the transcription start site of *groE*. The other nucleotide sequences are numbered according to the reported sequences. The order of the number indicates the direction of the primer. The direction of the primer is from 5' end to 3' end.

II.3 DNA Primers

Table 4 lists all the primers used in this study. ESL 10U and ESL 10L were used to generate the T (-12) to C mutation in the *groE* promoter. ESL 35U and ESL 35L were used to generate G (-33) to A mutation in *groE* promoter. ESLBglII and ESL10U were used to introduce the *Bgl*II site between the transcription start site and the IR sequence. GROEP and GROEPl were used to amplify the *groE* promoter (-115 to +2). GROEU and GROEB were used to generate the *groESL* fragment by PCR. GROEPE was used for the primer extension study to determine the half life of *groE* mRNA. BTBB and BTBF were used to produce the thermostable ß-galactosidase gene (*bgaB*) from *B. stearothermophilus*. BgaBPE was used for primer extension to detect the half life of *groE-bgaB* mRNA. BSXYF and BSXYB were used to

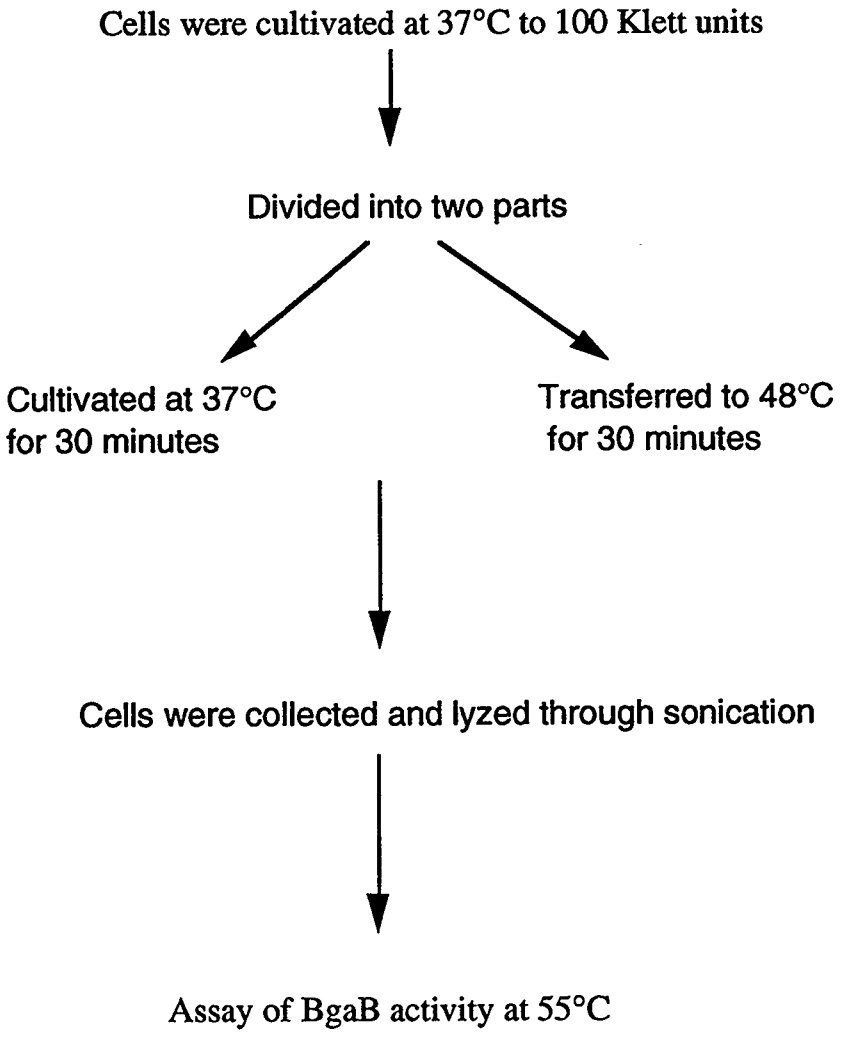
produce the promoter fragment from the *B. subtilis* xylose operon. ORF39F and ORF39B were used to generate the *orf39* through PCR.

II.4 Culture conditions

L-broth was used routinely to culture both *B. subtilis* and *E. coli*. Figure 6 illustrates cell culture conditions for both the heat shock and non-heat shock treatments and the temperature for BgaB (ß-galactosidase from *B. stearothermophilus*) assay. For determining the BgaB specific activity under heat shock and nonheat shock conditions, cells were cultured up to 100 Klett units and divided into two sets. One set was kept at 37°C, the other set was at 48°C. After a 30-minute cultivation, cells were collected for enzyme (BgaB) assays. To determine the effect of the mutated *sigA* on the expresion of *groE-bgaB,* cells were cultured under the same condition as mentioned above except that IPTG at a final concentration of 1 mM was added at the beginning of the heat shock and the cells were cultured for one hour. To prepare RNA for the half-life determination, *B. subtilis* was cultured at 37°C until the cell density reached 100 klett units. Cells were then splited into two sets. One was kept at 37°C, the other one was transferred to 48°C. These cells were cultured for 10 minutes. Rifampicin at the final concentration of 10 μg/ml was added. Cells were collected at different time points for RNA isolation. For the thermotolerant study, *B. subtilis* cells were grown in minimal medium for overnight, and then were transferred to fresh minimal medium. The initial cell density was 0.06 OD at A_{540}. The cells were divided into four sets. Three of them were cultured at 37°C, 48°C, 52°C respectively. The fourth one was cultivated at 48°C for thirty minutes then transferred to 52°C. To overproduce ORF39, *Ecoli* BL21 (DE3) carrying pET-ORF39 was

cultured in L-broth containing kanamycin (30 μg per ml) at 37°C to 100 Klett units. IPTG was added at the final concentration of 0.4 mM. Cells were collected after a 2-hour induction. To study the functional role of ORF39 in *E.coli*, cells with the appropriate plasmids were cultured in L-broth at 37°C up to 100 Klett units. The cells were divided into two sets. One set was first transferred to 42°C for 15 minutes and then transferred to 48°C for 30 minutes. The other set was kept at 37°C as the control.

Figure 6. Culture conditions for the heat shock and non-heat shock treatment and the assay temperature for BgaB activtiy (ß-galactosidase from *B. stearothermophilus*).


Cells were cultivated at 37°C to 100 Klett units
Divided into two parts
Cultivated at 37°C
for 30 minutes
Transferred to 48°C
for 30 minutes
Cells were collected and lyzed through sonication
Assay of BgaB activity at 55°C

II.5 Detection of integration at the *amyE* locus in *B. subtilis*.

After the pDL plasmid was transformed into *B. subtilis*, the cells were plated on TBAB agar plates with 5 microgram per ml of chloramophenicol The plates were kept at 37°C for overnight. The transformants were then spotted on the TBAB plates containing 0.1% starch. The plates were incubated at 37°C for 12 hours. The iodine solution (24 mM I_2, 24 mM KI) was then added to the plate. The colonies without a halo around them indicated that the *amyE* operon was disrupted through recombination.

II.6 Nucleic acid manipulations

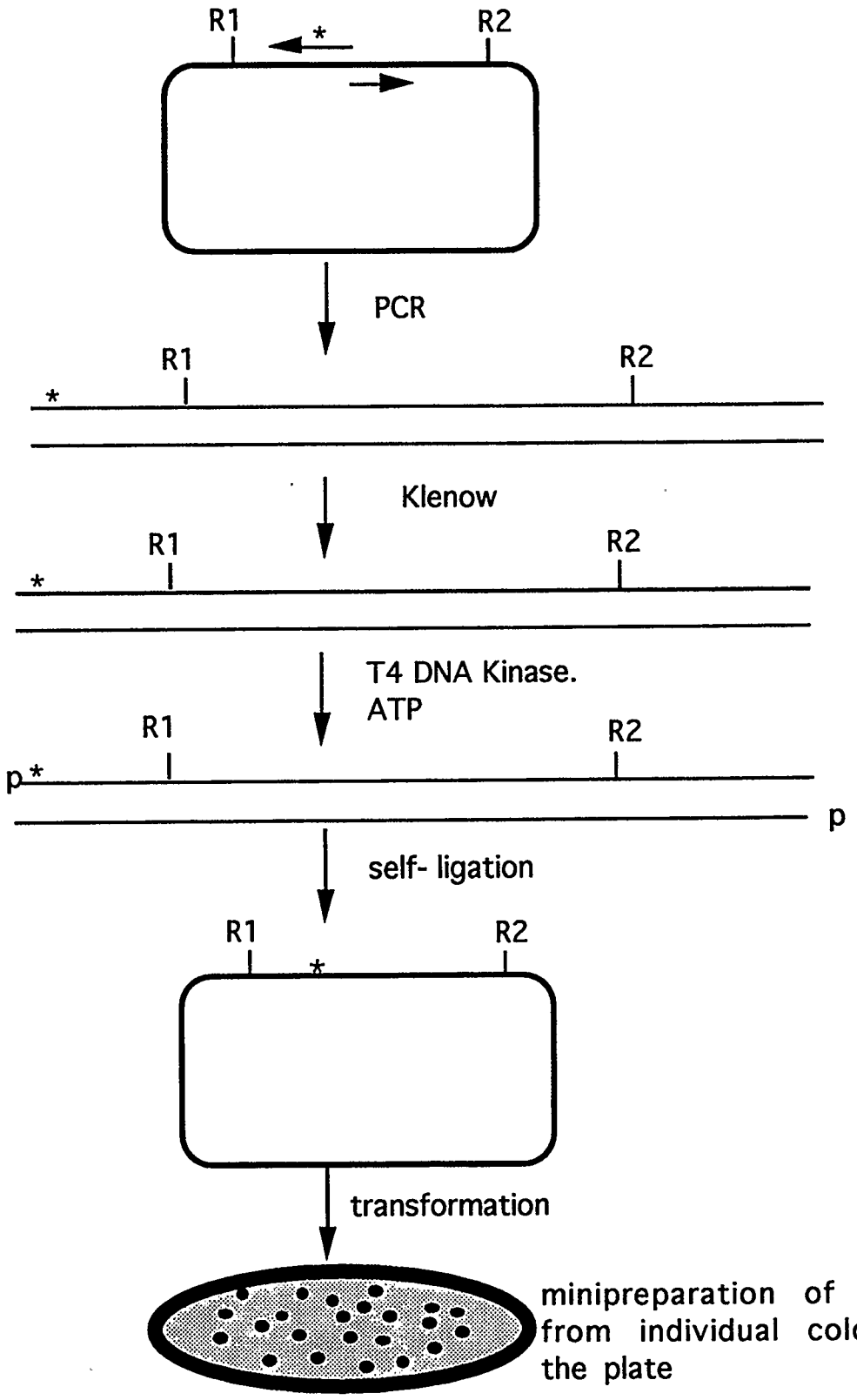
II.6.1 DNA isolation. Isolation of plasmid from *E. coli* was carried out according to Maniatis *et al.*, (1989). Preparation of plasmid from *B. subtilis* was performed according to Wu *et al.*, (1991). Chromosomal DNA from *B. subtilis* was prepared as described by Wu *et al.*, (1991).

II.6.2 DNA manipulations. Restriction endonucleases and other modification enzymes were used according to the suggestions from the manufacturers. PCR amplification and DNA sequencing were performed by methods described previously (Li & Wong, 1992). For genomic sequencing, sequences of interest were generated through PCR amplification by using Vent DNA polymerase (New England Biolabs Ltd., Canada) with the condition suggested by the manufacturer. The amplified DNA samples were used directly as templates for sequencing by using the fmol DNA sequencing system (Promega, Wisconsin).

Transformation of *E. coli* cells was carried out according to Maniatis *et al.*, (1989). Transformation of *B. subtilis* was performed according to Dubnau and Davidoff-Abelson, (1971).

II.6.3 Site-mutagenesis of the *groE* promoter. Figure 7 illustrates a method to generate site-directed mutagenesis by PCR (Hemsley *et al.*, 1989). Two primers are used to generate mutations in the method. The target DNA to be mutated was inserted to a vector. The resulting plasmid was used as the template for PCR. The first primer carrying the mutated sequence is complementary to one strand of the template, the second primer is complementary to the other strand of the template. There is no gap between the two primers. After PCR, linear DNA fragments with mutation were generated. The fragments were treated with Klenow and phosphylated with T4 polynucleotide kinase in the presence of the ATP. The resulting DNA fragments were self-ligated and transformed to an appropriate host strain. The resulting plasmid can be sequenced to confirm for the presence of the site-directed mutation. By using this method, *groE* promoters carrying -12C (T to C) or -33A (G to A) mutation were generated. The mutated promoters were inserted into pDK between *EcoR*I and *BamH*I sites. The resulting plasmids were integrated at the a*myE* locus in three *B. subtilis* strains (EU9001*tms*, EU9002*tms*-13C, and EU9004*tms*-35A, see Table 2) developed by Kenney and Moran.

Figure 7: A PCR based site-mutagenesis method. The arrows represent the primer used for PCR. " * " indicates the mutation in the primer. R1 and R2 represent the unique restriction sites. The fragment to be mutated is inserted between restriction sites R1 and R2.


R1
*
R2
PCR
R1
R2
*
Klenow
R1
R2
*
T4 DNA Kinase.
ATP
R1
R2
p*
p
self- ligation
R1
R2
*
transformation
minipreparation of plasmids from individual colonies on the plate

II.6.4 Genomic sequencing. The protocol for genomic sequencing was described as follows: The target DNA for sequencing was amplified through PCR by using chromosome DNA as the template and then purified through electroelution to get rid of primers and other non-specific DNA fragments. For each set of the sequencing reaction, four 0.5 ml microcentrifuge tubes (G, A, T, C) were labeled. Two microliter of the appropriate dNTP/ddNTP mix were added to each tube. Then, the sequencing reaction mix was prepared as follows: 500 fmol DNA (from PCR) was mixed with 3.0 pmol Primer and 0.5 μl [α–^{35}S] dATP (> 1000Ci/mmol) and the final volume was adjusted to 16 μl with sterilized distilled water. One microliter of sequencing grade *Taq* DNA polymerase (5 u/μl) was added to the mixture. Four microliter of this enzyme/primer/template mix were added to the inside wall of each tube containing dNTP/ddNTP mix. One drop (approx. 20 μl) of mineral oil was added to each tube and mixed by briefly spin. The reaction tubes were put into a thermal cycler that has been preheated to 95°C. The conditions for sequencing depended on the length and GC content of a primer. If a primer is over 24 nucleotides or is a short primer with a GC content more than 50%, using the following condition: 95°C for 2 minutes, then 95°C for 30 seconds, 70°C for 0.5 minute. It need 30 cycles for the reaction. If a primer is less 24 bases or with a GC-content less than 50%, the following conditions should be adopted: 95°C for 2 minutes, then 95°C for 30 seconds, 42°C for 30 seconds and 70°C for 1 minutes. The total cycle number is 30. After the reaction, three microliter of sequencing stop solution were added to each tube. Sequencing samples can be stored at -20°C. The reactions were heated at 70°C for 2 minutes before loading to a sequencing gel.

II.6.5 RNA isolation. The method to purify the RNA from *B.subtilis* was described by Li and Wong, (1992). Five milliliters of the cell culture were harvested by centrifugation (SS 34 rotor, 10,000 rpm for 5 minutes). The cell pellet was washed by the disruption buffer (30 mM Tris-HCl, pH 7.4, 100 mM NaCl, 5 mM EDTA) and resuspended in 0.5 ml of disruption buffer. 25 µl of SDS (20%) and 2.5 µl of proteinase K (20 mg/ml in disruption buffer) were added. The cells were sonicated three times (10 seconds pulse interval, 50 watts output) with a Bronson sonicator equipped with a microtip. Another 2.5 µl of proteinase K (20 mg/ml) was added immediately after sonication. The sample was kept at 37°C for 1 hr and extracted twice by phenol/chloroform (1:1). RNA in the sample was precipitated by adding 50 µl of 3M sodium acetate (pH 4.8) and 2 volumes 95% ethanol. After drying by vacuum, the precipitated nucleic acid was dissolved in 300 µl of 20 mM Tris-HCl, pH 7.5, 10 mM $CaCl_2$, 100 µg/ml RNA free DNAase. The digested sample was incubated 1 hr at 37°C and extracted 3 times by phenol/chloroform (1:1) and precipitated by ethanol. This methods generates DNA-free RNA preparations.

II.6.6 Determination of the half life of the mRNA for *groE* and *groE-bgaB*. The *groE* mRNA was detected by primer extension and Northern blot described by Li and Wong (1992). The primer used for primer extension is GROEPE (Table 4) which has the sequence complementary to the *groES* coding sequence from position +81 to +99 (Li and Wong, 1992), +1 corresponds to the transcription start site of *groE*. For Northern blot hybridization, two DNA probes were used. One was specific to *groES* with

the sequence covering nucleotides +20 to +326 while the other was specific to *groEL* with sequence covering nucleotides +617 to +1640. These probes were generated through PCR amplification. The probes were radiolabeled with [α-^{32}P]-dATP by using the random primer labeling kit (Bethesda Research Laboratories) with the condition specified by the manufacturer. To determine the half life of *groE-bgaB* transcripts by primer extension, a *bgaB* specific primer (BgaBPE) with the sequence complementary to nucleotides 475 to 492 of the reported *bgaB* sequence was designed. The band intensity in the autoradiograms was quantified with an LKB Ultrascan XL enhanced laser densitometer.

II.7 Chemical Mutagenesis

B. subtilis WBG2 was inoculated into L-broth (25 ml) and cultured at 37°C. When the cell density reached 0.8 OD_{600}, the cells were collected and washed once with SC (0.15 M NaCl, 0.01 M Sodium Citrate, pH 7.0). The cell pellet was then resuspended in 10 ml of SC with the addition of N-methyl-N'-nitro-N-nitrosoguanidine (NTG) at a final concentration of 100 μg/ml. After culturing at 37°C for 30 minutes. The cell pellet was resuspended in 1 ml L-broth and kept at 37°C for an hour. The culture was collected and resuspended in 10 ml of SC. Cells were diluted and plated on TBAB agar medium (with X-gal 30 μg/ml, chloramphenicol al 5 μg/ml) and kept at 37°C for overnight.

II.8 Cellular protein manipulation

B. subtilis and *E. coli* cells were lyzed by sonication for three times (10 seconds pulse interval, 50 watts output) in SET buffer (20% surcose, 50 mM

Tris-HCl, 50 mM EDTA, pH 7.6). Cell debris was removed by centrifugation at 12,000 g for 5 minutes. The supernatant was kept at 4°C. Protein concentration was determined by the Bradford method with the coomassie brilliant blue reagant from Bio-Rad Laboratories. Recommended procedure for the protein assay was followed.

II.9 Band shift assays

Crude extracts were prepared from *E. coli* BL21 (DE3) carrying either pET29 bΔ(XB) or pET-ORF39. Cells were lyzed by sonication in the breakage buffer (10 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 0.1 mM EDTA, 1% glycerol, 1 mM DTT). Cell debris and insoluble proteins were removed by centrifugation. Supernatant was saved for the mobility shift assay. The DNA probe was a 70-bp fragment containing the *B. subtilis groE* regulatory sequence from -22 to +34. This fragment was generated by PCR amplification and end-labeled by T4 DNA kinase with γ-[^{32}P]-ATP. The band shift assay was performed under following conditions. The total volume of the assay was 30 µl. The assay mixture contained 10 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 100 mM NaCl, 2mM DTT and 10% glycerol, 1 µg sonicated salmon sperm DNA. 0.2 pmol of the DNA probe and 15 µg of total proteins from the crude extract were added to the mixture for the assay. The sample was kept at 25°C for 15 minutes and loaded to an 8% polyacrylamide gel, which was then run at 180 Volt in 0.5X TBE buffer. After runing, the gel was dried and autoradiographed.

II.10 Determination of the specific activity of BgaB

The method was described by Hirata *et al.*, (1985). The cells were lyzed by sonication in Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4 \cdot 7H_2O$, 50mM mercaptoethanol). Cell debris was removed by centrifugation at 10,000 X g for 5 minutes. The cell extract was diluted with Z buffer to 0.8 ml, which contains 50 to 100 microgram protein. Two hundred microliter of ONPG (o-nitrophenyl-ß-D-galactoside, 10 mg/ml in Z buffer) was added. The reaction was kept at 55°C for 2 to 10 minutes. The absorbance at 420 nm was determined. One unit of BgaB activity is defined as the hydrolysis of 1 nmol of substrate per min. The specific activity of BgaB is expressed as units per µg cellular protein.

III. Development of a Vector to Study the Expression of Heat Shock Genes in *B. subtilis*

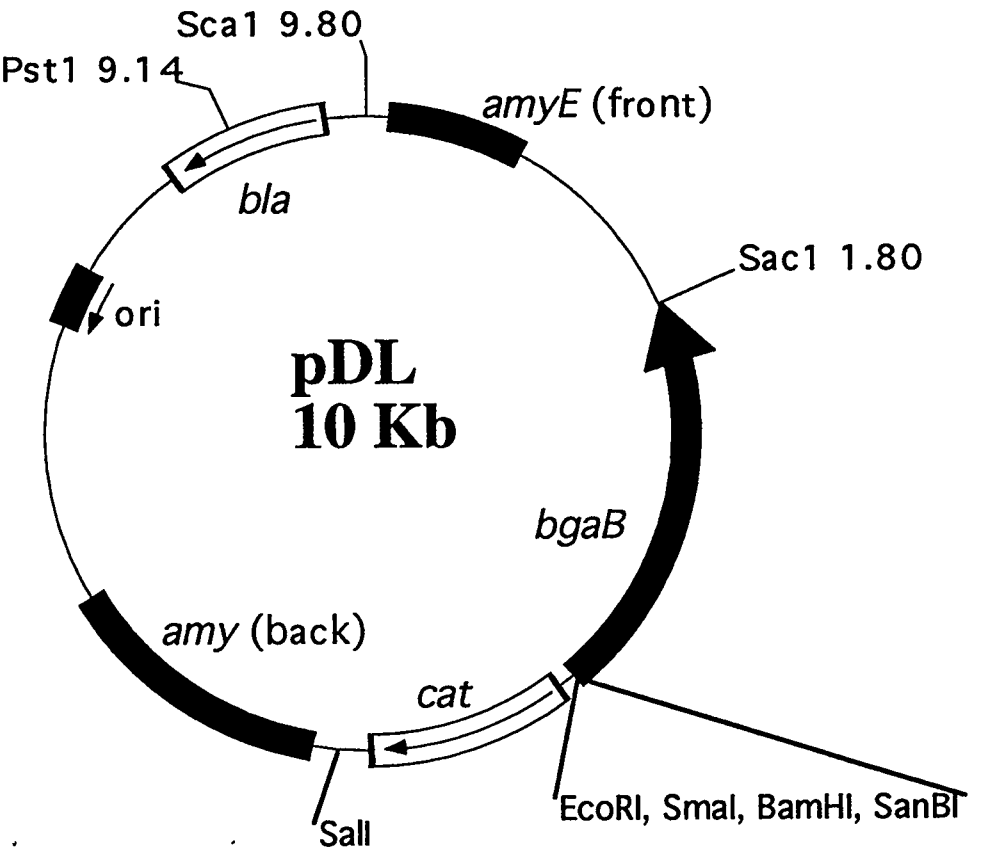
III.1 Introduction

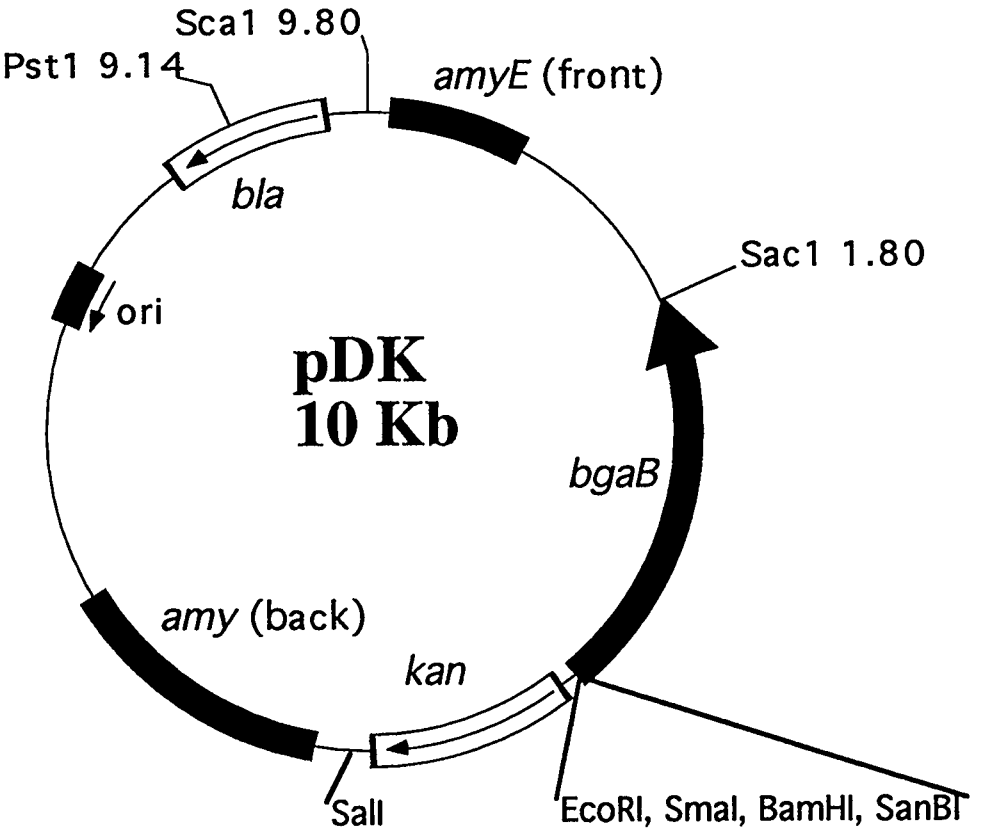
Gene expression in certain cases is difficult to study since the gene product is not easily measured quantitatively. For these cases, a reporter gene is applied to monitor the gene expression. The most commonly used reporter gene is *lacZ*. The gene product of *lacZ*, ß-galactosidase, can be assayed easily and accurately. However, ß-galactosidase was not a suitable reporter to monitor the expression of the heat inducible genes in *B. subtilis* due to the loss of activity under the heat shock condition (Zuber & Schumann, 1994). Therefore, it is important to have a suitable reporter similar to ß-galactosidase to study the gene expression under heat shock conditions. In this study, a ß-galactosidase gene, *bgaB*, was selected as a reporter gene to monitor the expression of *groE* in *B. subtilis*.

III.2 Construction of pDL and pDK to study the gene expression under heat shock and non-heat shock conditions

bgaB, a ß-galactosidase gene, was amplified through PCR using the *B. stearothermophilus* chromosomal DNA as the template. Two primers, BTBF and BTBB, were used for PCR amplification. The first primer BTBF (GGGGTACGTAGGTACC**TAA**ATTT**TAA**CTTAATTTATAAT**TAA**ACGAAAA TTAG) with the sequence corresponding to nucleotides 465 to 502 of the reported *bgaB* sequence (Upstream of the open reading frame) contains the *Sna*BI site and *Kpn*I site, it also contains translational stop codons (highlighted

Figure 8: Features of the integration vector pDL and pDK. Unique restriction enzyme sites in pDL and pDK were indicated in the figure. *bgaB*, *amyE*, *cat* and *bla* are structure genes for ß-galactosidase, α-amylase, chloramphenicol acetyl transferase, and ß-lactamase, respectively. *kan* is kanamycin resistant marker.


Sca1 9.80
Pst1 9.14
amyE (front)
bla
Sac1 1.80
ori
pDL
10 Kb
bgaB
amy (back)
cat
SalI
EcoRI, SmaI, BamHI, SanBI


Sca1 9.80
Pst1 9.14
amyE (front)
bla
Sac1 1.80
ori
pDK
10 Kb
bgaB
amy (back)
kan
SalI
EcoRI, SmaI, BamHI, SanBI

in primer sequences). The presence of translational stop codons in all three reading frames in front of *bgaB* will prevent the formation of any potential translational fusions. The primer BTBB corresponding to the 3' end of *bgaB* contains a *Sst*I site. The amplified promoterless *bgaB* (465 to 2506 of the reported sequence, Hirata *et al.*, 1986) fragment was digested by *Sna*BI and *Sst*I and inserted into the *Sna*BI-*Sst*I digested pDH32M (Kraus *et al.*, 1994, see Table 3). The resulting plasmid was designated pDL. pDK was the same as pDL except that the selection marker, *cat,* was replaced by *kan* (kanamycin resistant marker, which was generated through PCR with pUB18 as the template). Figure 8 shows the features of both plasmids, pDL and pDK. The pDL has several features: 1) A ColE origin allows the replication of the plasmid in *E. coli*; 2) A *bla* (The ß-lactamase gene) selection marker for selecting the transformed *E. coli* cells; 3) a *cat* (chloramphenicol acetyl transferase gene) gene for the selection of *B. subtilis* transformants; 4) A promoterless *bgaB* serves as the reporter gene; 5) In front of *bgaB*, there is a polylinker with several unique restriction sites which allow the insertion of the promoter of interest; 6) Sequences from 5' and 3' ends of the *amyE* operon allow the integration of the transcription fusions to the *amyE* locus in *B. subtilis*; 7) Two unique restriction sites *Pst*I and *Sca*I can be used to linearize pDL. The linearized plasmid can be integrated to the *amyE* locus through the double reciprocal recombination.

III.3 **Detection of the integration of pDL plasmid at the *amyE* locus in *B. subtilis***

After the pDL plasmid integrated to the *amyE* locus, *amyE* was disrupted. The cells with the integration can be easily detected by spotting the

transformed cells on TBAB plates containing 0.1% starch. The α-amylase activity can be monitored by adding an iodine solution to the plate. A halo surrounding the colonies indicates the production of an active α-amylase. Integration of pDL should disrupt the α-amylase gene. Consequently, no α-amylase activity would be expected (Figure 9).

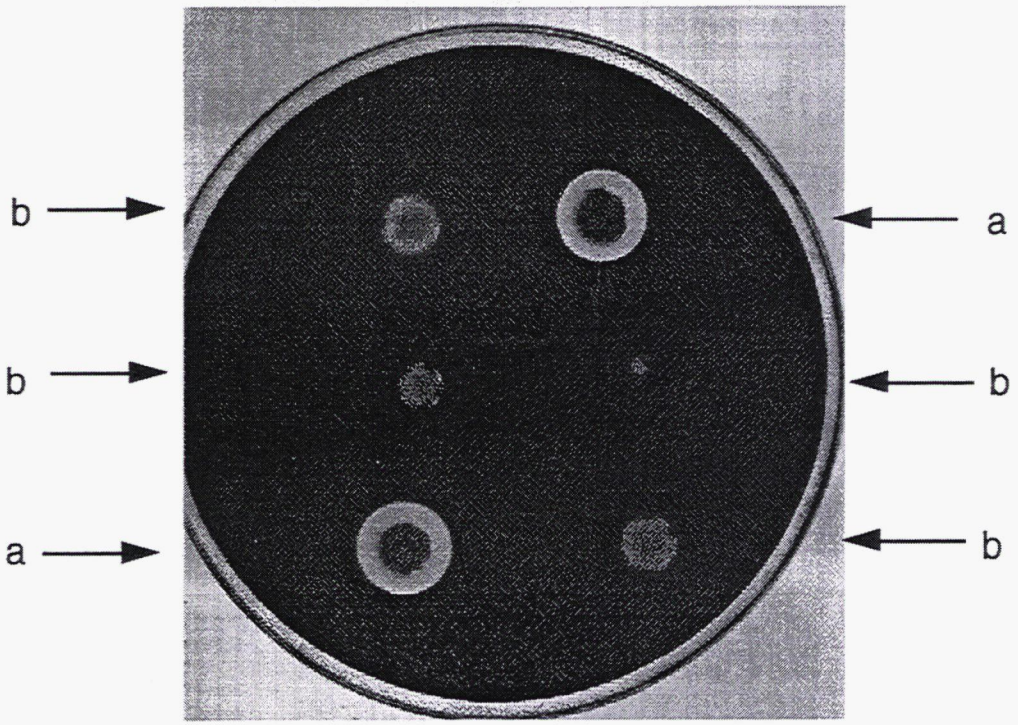


Figure 9: Detection of integration of pDL at the *amyE* locus in *B. subtilis*. The cells were spotted on the starch containing agar plate. Iodine solution was added for detecting the integration. a: *B. subtilis* 168 with intact *amyE* operon. The halo formed around the colony. b: *B. subtilis* with the integration of pDL. The *amyE* operon was disrupted, and no halo could be observed.

III.4 *groE* expression studied by using the promoter-probe plasmid, pDL

Figure 10 illustrates the process to use pDL to generate the transcriptional fusion with *bgaB* and integrate the pDL plasmids to the *amyE* locus. The DNA fragment covering the *groE* regulatory region (-238 to +36) including the inverted repeat sequence was inserted into pDL. The resulting plasmid was integrated at the *amyE* locus of *B. subtilis* 168 to generate strain WBG2. WBG1 is the control strain (*B. subtilis* 168 integrated with pDL). As shown in Table 5, the background BgaB specific activity in WBG1 is about 23 units per microgram protein. The low level of the constitutive expression of *groE* at 37°C was reflected by a 3.5-fold increase in BgaB specific activity in WBG2. Under the heat shock condition, an 8.5-fold increase in the BgaB specific activity in WBG2 was observed. This reflects the typical heat-induced expression of *groE*. These results indicate that the expression of *groE* can be monitored by detecting the BgaB activity from *groE-bgaB*.

Figure 10: Integration of the linearized pDL at the *amyE* locus of *B. subtilis*. Arrows indicate the direction of gene transcription. The rectangle with shadow indicates the inserted promoter fragment.

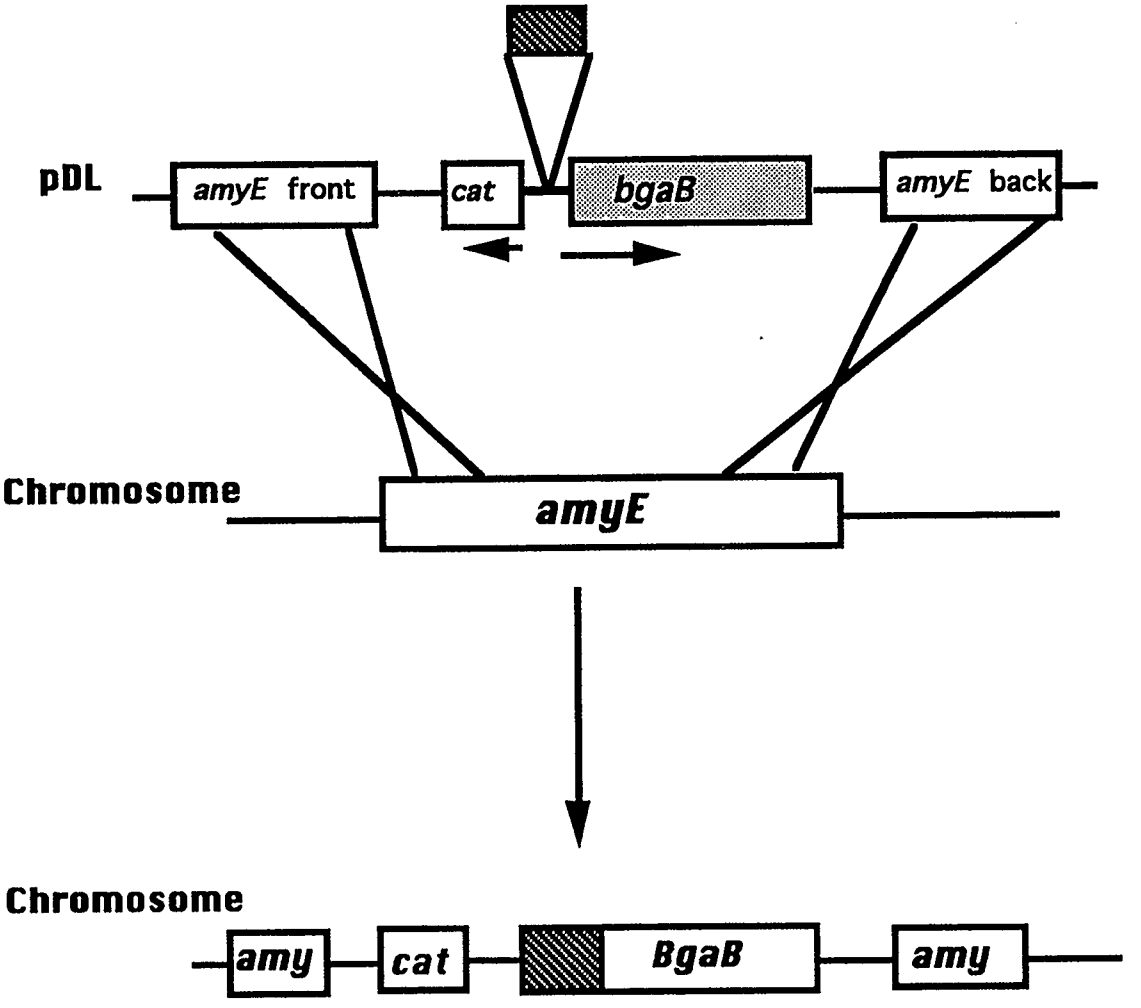

pDL
amyE front
cat
bgaB
amyE back
Chromosome
amyE
Chromosome
amy
cat
BgaB
amy

Table 5: BagB specific activities from *groE-bgaB* (with CIRCE) transcription fusion

Strains	Constructions	BgaB activity [b]		Ratio [a]
		37°C [c]	48°C [c]	
WBG1		23±4	25±9	—
WBG2		81±12	692±62	8.5

[a] Ratio is calculated as follows: BgaB activity from WBG2 under the heat shock condition (48°) to that from WBG2 under the non-heat shock condition (37°).

[b] Data presented represent the average of three experiments, BgaB units is defined as the hydrolysis of 1 nmol OPNG per minute. The BgaB activity represents the sepectific activity of BgaB (units/μg protein). All the assays were performed at 55°C.

[c] Cultivation temperature.

The vertical rectangle in WBG1 indicates the promoter insertion site.

III.5 Discussion

Previous studies to use *E. coli* ß-galactosidase (LacZ) as a reporter to monitor the heat-inducible expression of *dnaK* in *B. subtilis* was not successful (Zuber & Schumann, 1994). Under the heat shock condition, the cells carrying *dnaK-lacZ* showed little activity of ß-galactosidase. The loss of activity for ß-galactosidase may be due to the degradation of ß-galactosidase by the heat-inducible proteinases. Recently, a *lonA* and *lonB* (encoding ATP-dependent proteases) null mutant of *B. subtilis* with the *groE-lacZ* transcriptional fusion at the *amyE* locus was shown to have a ten-fold higher ß-galactosidase activity relative to that from the control (*groE-lacZ* in wild type *B. subtilis* strain) under the heat shock condition (Ye and Wong, unpublished result). This result supported the hypothesis that the heat inducible proteases could affect the stability of LacZ under the heat shock condition.

For the first time, a thermostable ß-galactosidase, was introduced to monitor the *groE* expression in *B. subtilis* in this study. Our results showed that expression of *groE* can be monitored by determining the expression of *bgaB* under both the heat shock and non-heat shock conditions. This system provides a simple and accurate method to monitor gene expression under these conditions.

IV. *B. subtilis groE* is Transcribed by the σ^A Containing RNA Polymerase under both Heat Shock and non-Heat Shock Conditions

IV.1 Introduction

In this chapter, the role of σ^A in the expression of *groE* is studied. A set of specific *B. subtilis* strains was used in this study. The principle and detail strategy to demonstrate that the σ^A containing RNA polymerase can transcribe *groE* is described in Chapter I (I.1.3.3) and the following text.

IV.2 A short synthetic DNA fragment carrying the putative *groE* promoter is transcriptionally active.

The σ^A-like promoter of *groE* was initially mapped by primer extension with RNA prepared under heat shock and non-heat shock conditions (Li and Wong, 1992). The same transcription start site was observed under these conditions. To confirm this putative *groE* promoter is active, a DNA fragment, which only covered the minimum elements (-10 and -35 regions) of the σ^A-like *groE* promoter (-44 to +2, +1 is the transcription start site), was synthesized. To confirm that this region indeed functions as a promoter, this DNA fragment was fused with the promoterless *bgaB* and integrated to *B. subtilis* 168 to generate WBG15. Compared to the control strain WBG1, expression of *bgaB* can be observed under heat shock and non-heat shock conditions in WBG15 (Table 6). This indicated that this sequence indeed functioned as a promoter. The *bgaB* expression at 37°C was found to be 5-fold higher in another *groE-bgaB* fusion strain (WBG16) which carried a longer *groE* promoter (-112 to +2). Extending the upstream sequence of the *groE* promoter to -238 did not result in

any significant increase of *bgaB* activity in WBG3. This result suggested that the sequence between -44 and -112 contained additional regulatory elements that could enhance the transcription efficiency of the *groE* promoter.

Table 6:BgaB activity in various *groE-bgaB* transcriptional fusions

Strain	DNA inserted sequence from the *groE*	BgaB activity[a]	
		37°C[b]	48°C[b]
WBG1	none	23±4	25±9
WBG15	-44 to +2	204±21	562±54
WBG16	-112 to +2	1127±168	2108±240
WBG3	-238 to +2	974±120	1870±344

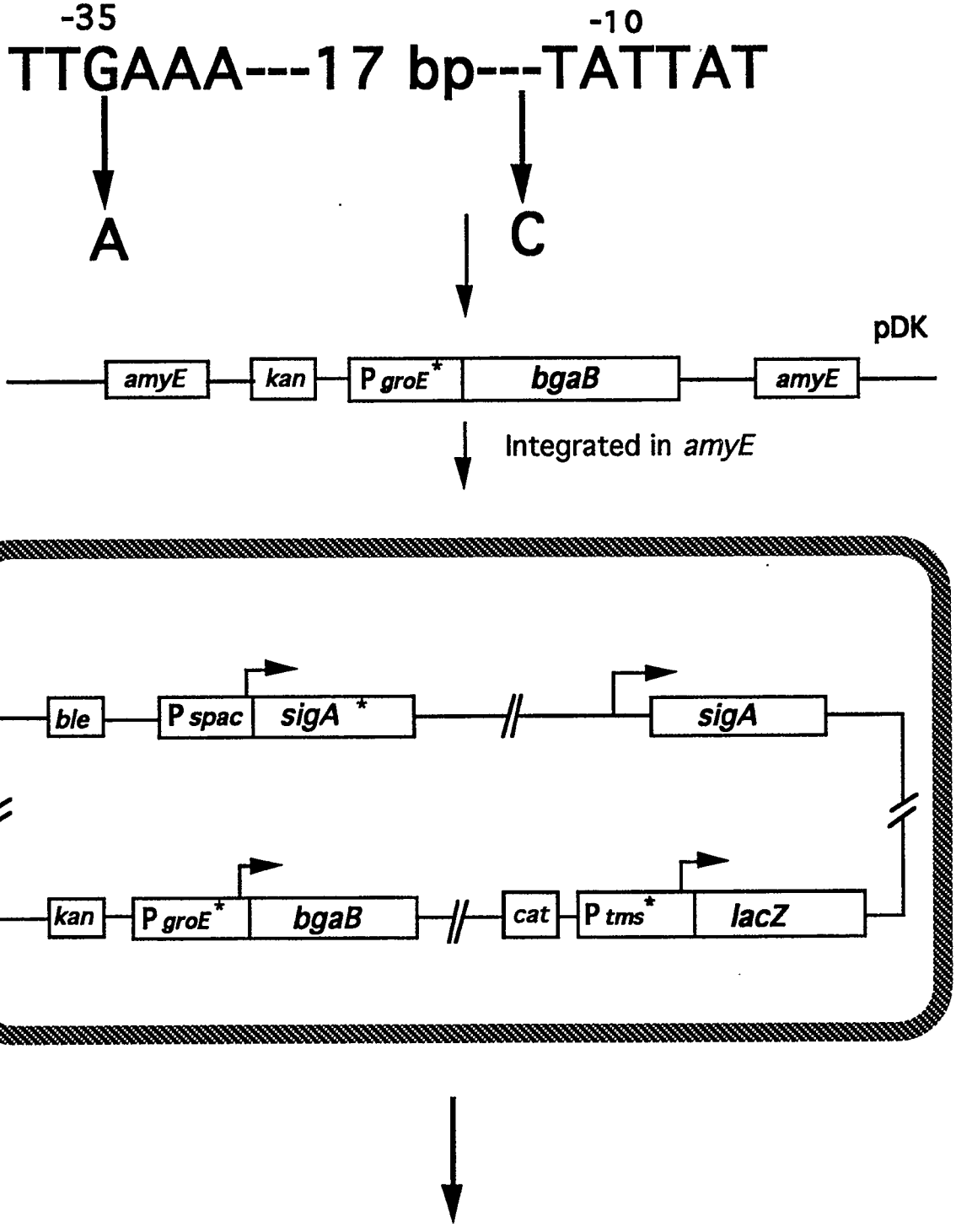
[a] Data presented represent the average of three experiments.

[b] Cultivation temperature.

IV.3 Effect of mutations in the -10 and -35 regions of the *groE* promoter.

From the above study, we demonstrated that the *groE* promoter was active under heat shock and non-heat shock conditions. The *groE* promoter contains a 5 out of 6-bp match with each hexameric consensus sequence found in the -10 and -35 regions of the σ^A - dependent promoters. The spacer between -10 and -35 regions is 17 bp in length (Figure 11). To confirm the

Figure 11: The strategy to study the interaction between the *groE* promoter and σ^A (see text for details). Expression of the second *sigA* (wild type or mutated) was controlled by the IPTG induced promoter P*spac*. *ble* is the phleomycin resistance marker. *kan* is kanamycin resistance marker. *cat* is chloramphenicol resistance marker. *bgaB* encodes the ß-galactosidase gene from *B. stearothermophilus*. *lacZ* encodes the ß-galactosidase gene from *E. coli*. P*tms* is a σ^A-dependent promoter from *B. subtilis*. P*groE* is the *groE* promoter (-238 to +2) from *B. subtilis*. The asterisk indicates either mutated or wild type *groE* promoter or *sigA*.



Add IPTG to induce the expression of mutated SigA gene. Then assay BgaB activity.

any significant increase of *bgaB* activity in WBG3. This result suggested that the sequence between -44 and -112 contained additional regulatory elements that could enhance the transcription efficiency of the *groE* promoter.

Table 6:BgaB activity in various *groE-bgaB* transcriptional fusions

Strain	DNA inserted sequence from the *groE*	BgaB activity[a]	
		37°C[b]	48°C[b]
WBG1	none	23±4	25±9
WBG15	-44 to +2	204±21	562±54
WBG16	-112 to +2	1127±168	2108±240
WBG3	-238 to +2	974±120	1870±344

[a] Data presented represent the average of three experiments.

[b] Cultivation temperature.

IV.3 Effect of mutations in the -10 and -35 regions of the *groE* promoter.

From the above study, we demonstrated that the *groE* promoter was active under heat shock and non-heat shock conditions. The *groE* promoter contains a 5 out of 6-bp match with each hexameric consensus sequence found in the -10 and -35 regions of the σ^A - dependent promoters. The spacer between -10 and -35 regions is 17 bp in length (Figure 11). To confirm the

putative -10 and -35 region is important for the function of the *groE* promoter, we used site-directed mutagenesis to generate two single base-pair substitutions at the highly conserved -10 (-12T to C) and -35 (-33G to A) regions in the *groE* promoter, respectively (Figure 11). To determine the effects of these mutations on the *groE* promoter activity, the mutated promoters were fused with promoterless *bgaB* in pDK and integrated into three different of strains developed by Kenney *et al.*, (Table 8). In these strains, a second copy of σ^A gene was integrated into chromosome. Transcription from the second *sigA* was directed by an IPTG inducible promoter, P*spac*. The second σ^A gene encodes either the wild type σ^A or one of the mutated σ^A (Q-196-R or R-347-H). Q-196-R represents glutamine at position 196 was replaced by arginine and R-347-H is the replacement of arginine at position 347 by histidine. These cells also carry an integrated *lacZ* transcription fusion directed by either the wild type or mutated *tms* promoter (-13T→C or -34G→A, Figure 12). The *tms* promoter is a σ^A-dependent promoter from *B. subtilis* (Kenney and Moran, 1991). The activities of the *tms* promoter with mutations in the -10 region (-13T→C) and the -35 region (-34G→A) can be restored selectively by cells producing mutated SigA factors with the Q-196-R and R-347-H conversion, respectively. The presence of the *tms-lacZ* fusion in these strains allows the verification of the properties of these strains before integration of the *groE-bgaB* fusion to the chromosome. After integration, all nine strains carry two reporter genes, *E. coli lacZ* and *B stearothermophilus bgaB*. Since only BgaB is thermostable, it is possible to eliminate the activity of *E. coli* ß-galactosidase (LacZ) by a brief heat treatment when BgaB and LacZ exist in the same host. As shown in Table 7, the strain E9001tms-lacZ (carrying *lacZ* under the control of the *B. subtilis tms* promoter) can express *E.*

coli ß-galactosidase at a high level. *E. coli* ß-galactosidase from the crude extract of E9001tms-lacZ was inactivated after a heat treatment at 70°C for 15 minutes. However, WBG2 retained 90% of the BgaB activity assayed at 55°C after the heat treatment. BgaB shows weak activity at 37°C as the optimal assay temperature for this thermostable enzyme is 55°C. Thus, the activity contributed by BgaB can be determined, even in the presence of LacZ. As shown in Table 8, the cells (WtsigA) with the mutated *groE* promoters carrying either the -12T→ C mutation or the -33G → A mutation showed a 5 to 6 fold reduction in BgaB activity compared to that from the strain carrying the wild type *groE* promoter (P*groE*). These data demonstrated that nucleotides at positions -12 and -33 are important for *groE* expression.

Table 7. Heat stability of BgaB and *E.coli* β–galactosidase

Strain	Assay temperature	Activity[a] non-heat treatment	heat treatment
WBG2[b]	37°C	53±3.2	68±4.6
	55°C	1202±146	972±123
WtsigA tms[c]	37°C	4227±790	1.7±1.5
	55°C	5417±954	1.9±1.7

[a] β–galactosidase specific activity for the cell crude extract with and without the 70°C pre-treatment for 15 minutes.

[b] WBG2 is *B. subtilis* 168 with the *groE-bgaB* fusion. The *groE* promoter (-238 to +2) inserted does not carry the CIRCE sequence.

Table 8: Restoration of the mutated *groE* promoters by *sigA* mutants under the non-heat shock condition

Strain	Promoter	Activity[a]		Induction ratio[b]
		Uninduced	Induced	
wtsigA	P*groE*	980±149	960±214	1
	-12T→C	200±25	255±19	1.3
	-33G→A	157±24	185±40	1.2
Q-196-R	P*groE*	861±42	765±65	1
	-12T→C	208±21	726±32	3.9
	-33G→A	63±8	32±3	0.6
R-347-H	P*groE*	690±62	542±29	1
	-12T→C	132±9	84±10	0.8
	-33G→A	146±20	603±67.5	5.3

[a] BgaB specific activity 60 minutes after induction with 1mM IPTG. Data presented represent the average of the three experiments

[b] Induction ratio is calculated as follows: (mutant promoter +IPTG/-IPTG/ wild-type promoter +IPTG/-IPTG)

IV.4 Amino acid substitutions in σ^A suppress the effects of mutation in the *groE* promoter

To confirm that the σ^A containing RNA polymerase is responsible for transcribing *groE,* the ability of various *sigA* mutants to restore the promoter activity from the mutated *groE* promoters in an allele-specific manner was studied at 37°C and 48°C. As shown in Tables 8 and 9, no change in the level of expression of the wild-type (P*groE*) or mutant *groE* promoters was detected upon induction of the wild-type *sigA* gene. In contrast, induction of the *sigA* gene with the Q-196-R mutation resulted in a 3 to 4 fold increasing in the activity from the *groE* promoter with a mutation at -12 (T→ C) . With the induction of R-347-H allele of *sig,* a 3-5 fold increase of BgaB activity from the *groE* promoter with a mutation at -33 (G →A) was also observed. The induction of the *sigA* allele with the Q-196-R mutation cannot restore the T to C mutation at position -12 of the *groE* promoter. *sigA* mutant (R-347-H) also failed to suppress the G to A mutation at position -33 of the *groE* promoter. These data illustrate the high selectivity of the *sigA* alleles in restoring the promoter mutants with specific mutations. Since the mutated σ^A factors can compete with the wild-type σ^A to bind to the core RNA polymerase, this may provide an explaination for the observed slight reduction of the BgaB activity from the strain with the wild type *groE* promoter upon induction.

Table 9: Restoration of mutated the *groE* promoters by *sigA* mutants under the heat shock condition

Strain	Promoter	Activity[a]		Induction ratio[b]
		Uninduced	Induced	
wtsigA	P*groE*	1254±141	1685±277	1
	-12T→C	288±63	486±53	1.2
	-33G→A	213±50	286±54	1
Q-196-R	P*groE*	1654±249	1336±113	1
	-12T→C	453±115	1073±232	2.9
	-33G→A	133±31	94±18	0.6
R-347-H	P*groE*	1048±102	742±107	1
	-12T→C	227±72	212±59	1.3
	-33G→A	477±94	1207±249	3.6

[a] BgaB specific activity 60 minutes after induction with 1mM IPTG. Data presented represent the average of three experiments

[b] Induction ratio is calculated as follows: (mutant promoter +IPTG/-IPTG/ wild-type promoter +IPTG/-IPTG)

IV.5 Discussion

Chang *et al.* reported that: 1) *groE* promoter can be transcribed by the σ^A containing RNA polymerase *in vitro*. 2) Heat induction of the *groE* gene *in vivo* was much lower in a temperature sensitive *sigA* mutant than that in the wild type *sigA* strain (Chang *et al.*, 1994). These findings suggest that σ^A plays a vital role in expressing *groE*. However, certain complications from these findings cannot be ignored. 1) It is possible for the existence of trace amounts of minor σ factor in the RNA polymerase preparation used for the *in vitro* transcription. The RNA polymerase associated with the minor σ factor may be the one that actually initiates the transcription from the *groE* promoter; 2) Expression of *groE* may be regulated by a possible positive regulatory factor that is transcribed by the σ^A containing RNA polymerase; 3) The ts *sigA* mutant showed pleiotropic effects on gene expression (e.g. an abrupt production of two proteins, alkyl hydroperoxide reductase and glucose starvation-inducible protein B, which are not normally expressed at such a high level under heat stress).

In this study, different approaches were used to determine whether the *groE* promoter is recognized by σ^A. Our results have confirmed that: 1) The short 46-bp putative *groE* promoter fragment is functional *in vivo*; 2) Mutations in the putative -35 and -10 regions of the *groE* promoter severely reduced the function of the *groE* promoter in *B. subtilis*. 3) These mutations in the *groE* promoter can be suppressed in an allele-specific manner by the appropriate *sigA* mutants at both 37°C and 48°C. These findings clearly demonstrate that the σ^A containing RNA polymerase is directly responsible for the transcription of *groE* that under these conditions.

We also found a positive regulatory element located upstream of the -35 sequence which could enhance the transcription of the *groE* promoter. A third recognition element in many *E. coli* promoters was found to interact with the α subunit in the *E. coli* RNA polymerase (Ross *et al.*, 1993). These elements are not conserved in sequence but they are AT rich. This element can increase the transcription efficiency. It is usually located between -40 and -60 region. This element is called "upstream activation sequence" (UAS). In *B. subtilis*, the promoter of flagellin gene can be activated over 20-fold by a UAS sequence. The α subunit of the *B. subtilis* RNA polymerase can interact with this upstream activation sequence (Fredrick *et al.*, 1995). Alignment of 236 promoters recognized by the *B. subtilis* σ^A containing RNA polymerase revealed that weakly conserved AT rich regions are presented upstream of the -35 region (Helmann, 1995). These data suggested that the sequence upstream of the -35 region can enhance the transcription. In the *B. subtilis* *groE* promoter, sequence between -40 to -56 is highly AT rich (Figure 5). Therefore, this region may function as the upstream activation sequence (UAS) to further stimulate the expression of *groE*.

Since σ^A is the major σ factor that is produced constitutively in the cell and its production level does not increase further under heat shock (Chang *et al.*, 1994), it is of interest to determine what is the induction mechanism of *groE* at elevated temperatures. An inverted repeat sequence was found downstream of the *groE* promoter. This element and σ^A factor may regulate the expression of *groE* and *dnaK*. Chapter V summaries the study of the functional roles of the IR sequence in the *groE* regulatory region.

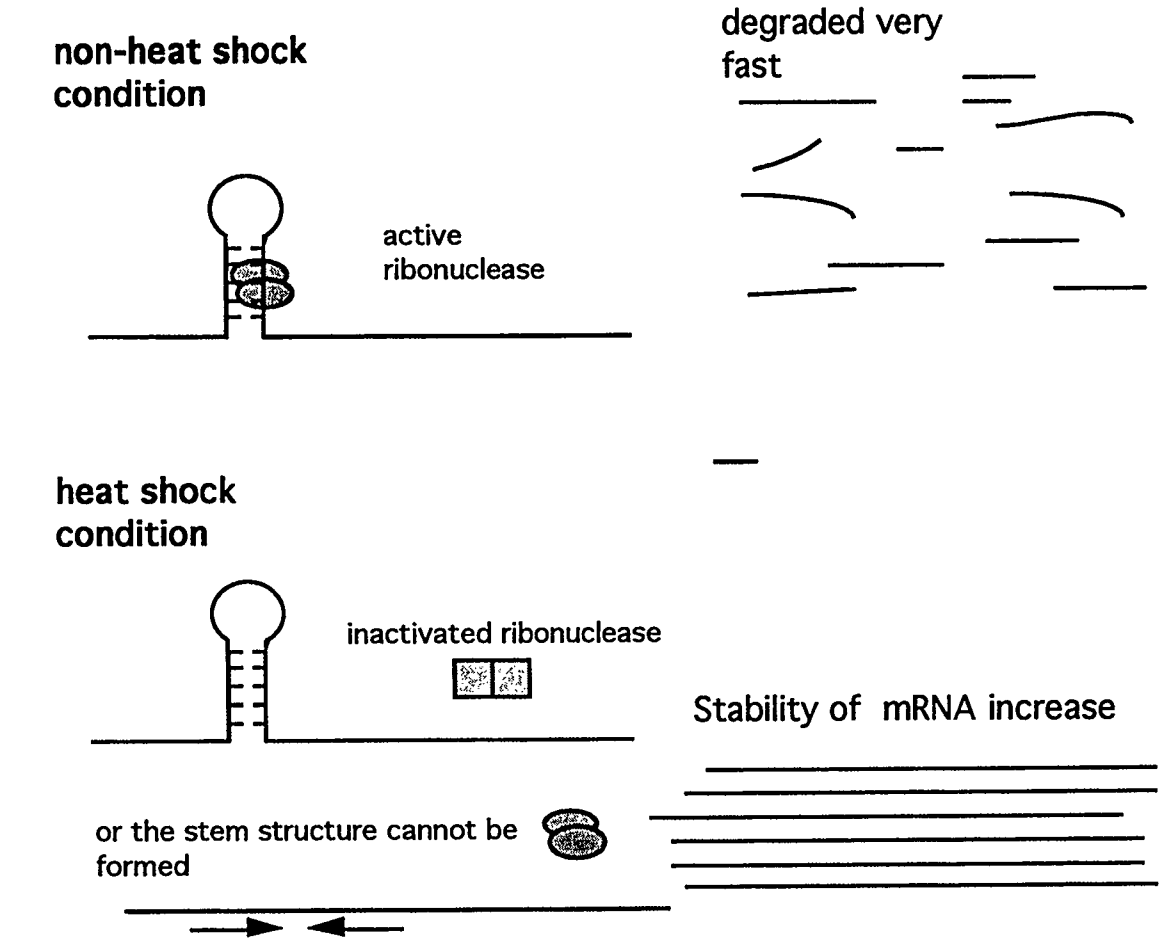
V Regulatory Functions of the Inverted Repeat Sequence (CIRCE) in the *B. subtilis groE*

V.1 Introduction

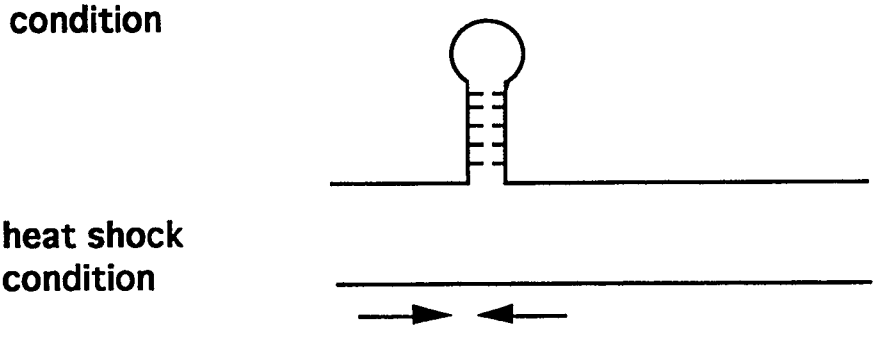
In this chapter, the functional role of CIRCE has been investigated. Point mutations in CIRCE of the *dnaK* operon resulted in the expression of *dnaK* at high levels (Zuber and Shumann, 1994). This suggests that CIRCE functions as a negative element. However, these point mutations may not be able to completely abolish the negative effects of CIRCE. It would be difficult to tell whether other regulatory elements or factors exist to regulate the expression of *groE*. In order to determine the functional role of CIRCE, CIRCE was deleted from the regulatory region of *groE*. The expression of *groE-bgaB* transcription fusion with the deletion of CIRCE was studied and compared with the expression of *groE-bgaB* transcription fusion. Our deletion results showed that CIRCE negatively regulated the *groE* expression (Table 10). As a negative element, CIRCE may exert its regulatory effect at either the RNA level or the DNA level (Figure 12). Since CIRCE is located downstream of the *groE* transcription start site, the *groE* mRNA should carry CIRCE at the 5' end. At the RNA level, two models are proposed: 1) CIRCE could be a target for some heat-sensitive ribonucleases. Under the non-heat shock condition, the *groE* transcripts are degraded by the putative ribonuclease. Following the temperature upshift, either the ribonuclease is inactivated or the recognition site of the ribonuclease cannot form. This leads to the increase of the half life of *groE* transcript. Consequently, the amount of *groE* mRNA and GroE proteins would increase; 2) CIRCE can be a transcription terminator. Under the non-heat shock condition, transcription of *groE* by the σ^A containing RNA

Figure 12: Models for the function of the inverted repeat sequence (CIRCE) in regulating *groE* and *dnaK* expression. 1. Regulation at the RNA level as a recognition site for ribonucleases or as a transcription terminator. 2. Regulation at the DNA level as a repressor binding site. The arrows indicate the IR sequence (CIRCE) in *groE*.

a. The ribonuclease model

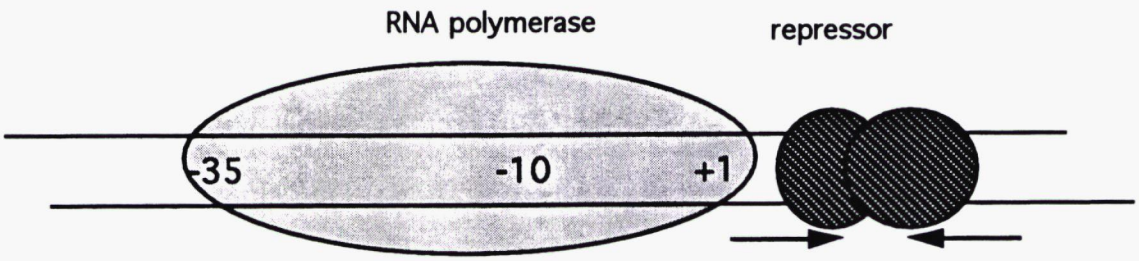


b. IR as a transcriptional terminator

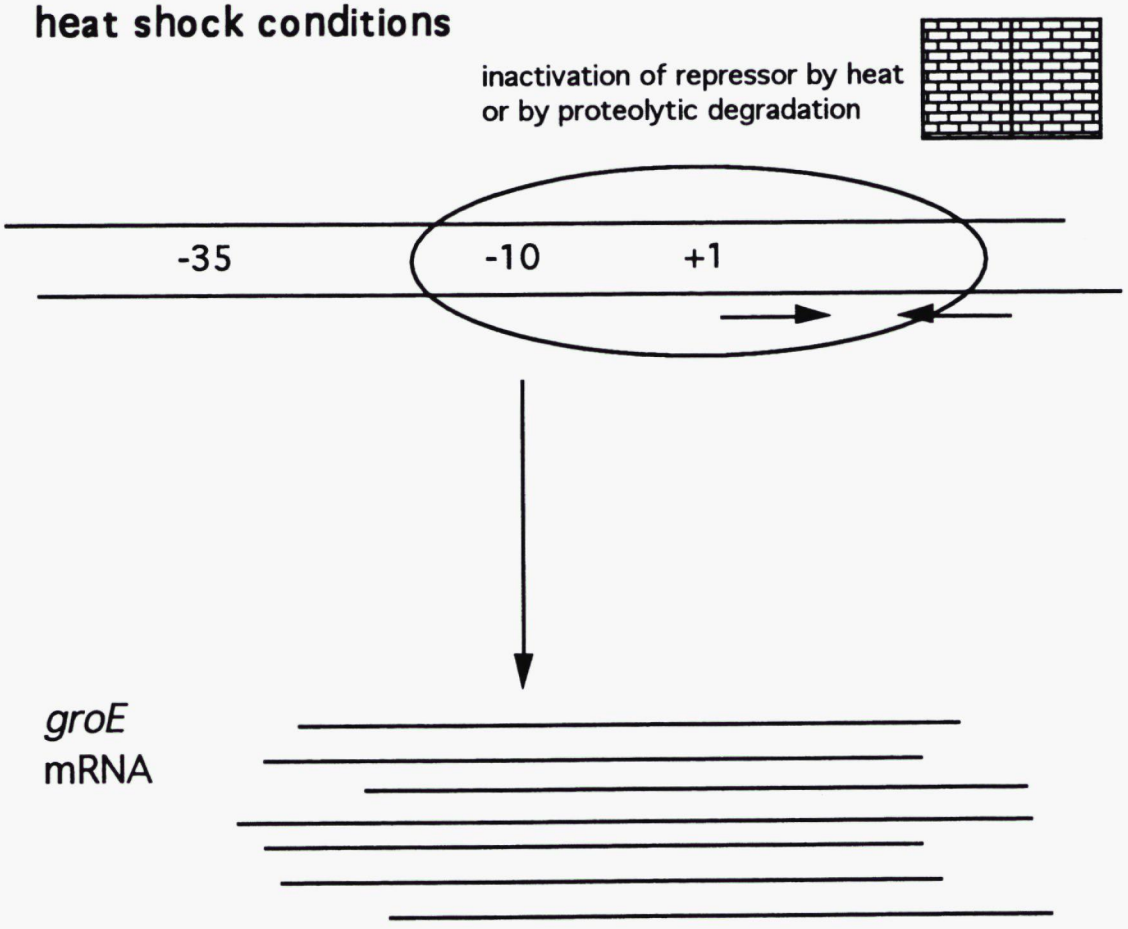


2. DNA level

non-heat shock condition



heat shock conditions



polymerase will be terminated. This terminator structure may be disrupted at high temperatures. The σ^A containing RNA polymerase can pass through this region. The expression of *groE* was then increased. For the DNA model, a repressor can bind to CIRCE. Transcription of *groE* was shut off by the repressor under the non-heat shock condition. When the cells are cultured at high temperatures, the putative repressor can be inactivated by either heat denaturation, chemical modifications, or proteolytic degradation. Then, the σ^A containing RNA polymerase can initiate the transcription of *groE*.

The ribonuclease model can be examined by determining the half life of the *groE* transcripts under heat shock and non-heat shock conditions. If this model is correct, the half life of the *groE* transcripts is expected to increase under the heat shock condition. The model of transcription terminator can be examined by the insertion of a small DNA fragment between the transcription start site and CIRCE. If this model is right, no matter where CIRCE locates the expression of *groE* will not be affected. For the operator model, the short distance between the operator and promoter is usually important for the repressor to shut off the gene expression efficiently. This model can also be examined by the insertion of the short DNA fragment between the transcription start site and CIRCE. The repressor may not be able to efficiently abolish transcription with the insertion of a long sequence between the *groE* promoter and CIRCE. With these approaches, we should be able to obtain vital information to see how CIRCE exerts its negative effects.

V.2 CIRCE is a negative regulatory element controlling the expression of *groE*

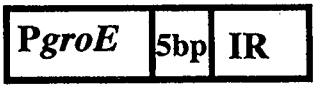
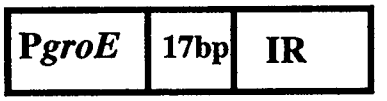
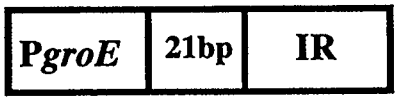
To determine the functional roles of CIRCE, the *groE* promoter (-238 to +2) without the inverted repeat sequence was inserted into pDL. The resulting plasmid was integrated at the *amyE* locus of *B. subtilis* 168 to generate WBG3. The BgaB specific activity was determined under both heat shock and non-heat shock conditions. As shown in Table 10, at the non-heat shock condition (37°C), deletion of CIRCE resulted in the constitutive expression of *groE-bgaB*. Compared to WBG2, a 12-fold increase in BgaB specific activity was observed. These data indicated that CIRCE functioned as a negative regulatory element in *groE* expression under the non-heat shock condition.

V.3 Introduction of a second copy of the modified *groE* gene at the *amyE* locus

To investigate whether the presence of CIRCE affects the stability of the *groE* transcripts, WBG8 was generated. WBG8 carries another copy of the modified *groE* operon in which the regulatory region, CIRCE, was deleted. This *groE* gene was integrated at the *amyE* locus. Figure 13a illustrates the process to create WBG8. Since we have hard time to generate the integration plasmid pDL-*groE* (ΔIR) in *E. coli*, the ligation mixture containing the *Bam*HI and *SstI* digested pDL16 (pDL with the *groE* promoter from -115 to +2) and the promoterless *groE* fragment digested with the same restriction enzymes was transformed directly to *B. subtilis*. The strain with the proper insert integrated at the *amyE* locus was detected by the loss of the AmyE activity and confirmed by PCR. Figure 13b shows the PCR results using the WBG8 chromosomal DNA as the template. The primers used for PCR are illustrated

in the Figure 14a and their sequences are listed in Table 4. The first PCR reaction was performed by using a pair of primers from the *groE* operon, one was GROEPE that was complementary to the *groE* sequence from +99 to +81, the other one was GROEP that corresponds the *groE* sequence from -115 to -95. +1 represents the transcription start site of *groE*. Two bands were observed by this PCR amplification, one band was expected to be 214 bp (Figure 13b,band b in lane 1) from the wild type *groE,* the other band was expected to be 198 bp from the integrated *groE* with the deletion of the IR sequence (Figure 13b, band c in lane 1). This result confirmed that there were two copies of *groE* in the WBG8 chromosomal DNA. The second PCR reaction was performed by using the GROEPE primer (+98 to +81) and the pH1 primer which was 75 bp upstream of the inserted *groE* promoter and located in *cat*. A 290-bp fragment was amplified. This result indicated that the second copy of *groE* was in the pDL plasmid integrated at the *amyE* locus (Figure 13b, lane 2). Since, it was demonstrated that CIRCE (the inverted repeat sequence) functioned as a negative element in the expression of *groE*. Deletion of the inverted repeat sequence (CIRCE) in *groE* operon should result in the constitutive expression of *groE* under the non-heat-shock condition. As shown in Figure 14, a protein was found to give higher levels of expression in WBG8. The size of this protein is identical to that of GroEL. This result confirmed again that CIRCE was a negative element for *groE* expression. WBG8 is used to study the effect of IR on the stability of the *groE* transcript.

Table 10: Specific activities of BgaB from various *groE-bgaB* transcription fusions.

Strains	Constructions	BgaB activity[b] 37°C [c]	BgaB activity[b] 48°C [c]	Ratio[a] 37°C
WBG1		23±4	25±9	0.3
WBG2		81±12	692±62	1
WBG3		974±120	1870±344	12.0
WBG4		310±52	1095±94	3.8
WBG5		660±38	1708±73	8.1
WBG6		1002±102	2080±240	12.3
WBG7		983±56	2130±248	12.1
WBG17		729±49	790±71	-

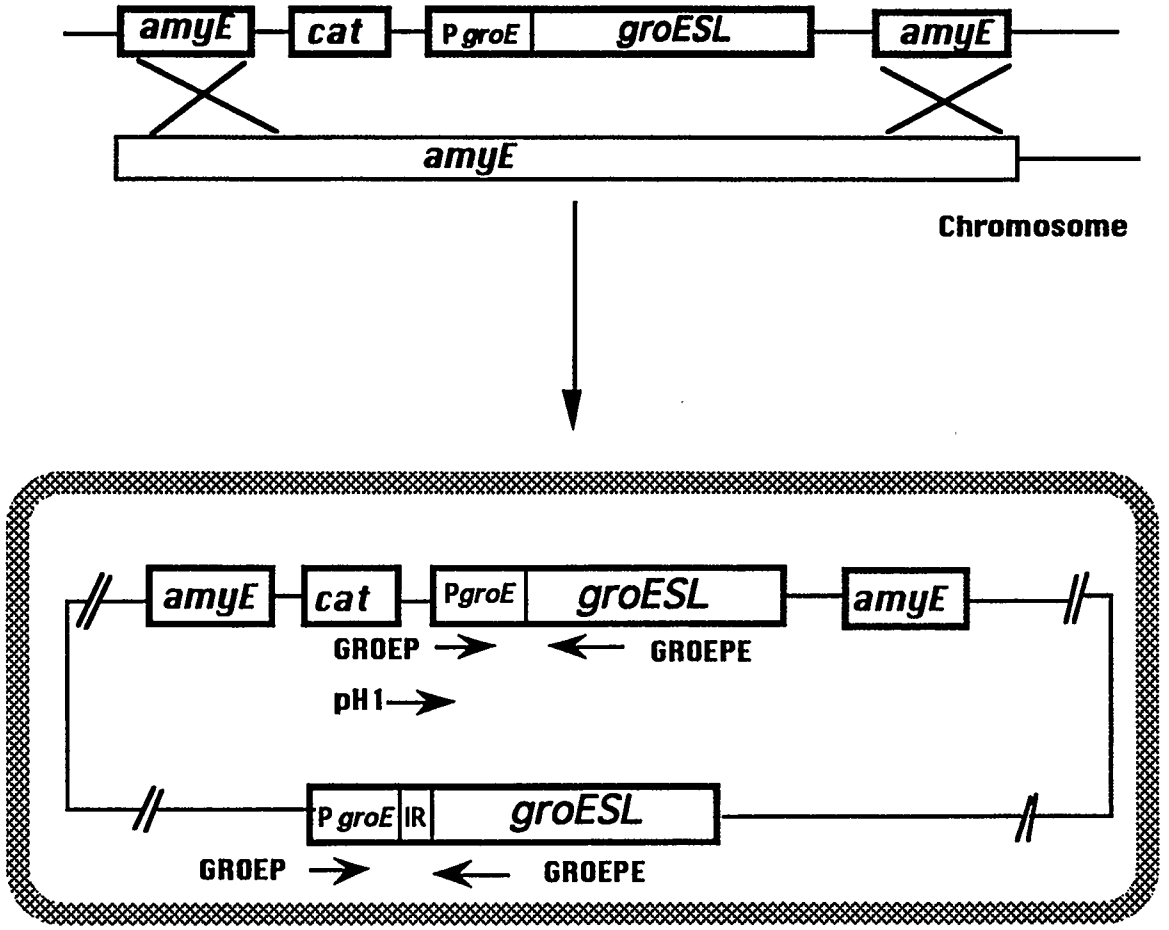
[a] Ratio is calculated as follows: specific activity (37°C) of BgaB from WBG1 toWBG7/ specific activity of BgaB from WBG2. The 5' end of the inserted promoters in WBG2 to WBG7 is at position -238.

[b] Data presented represent the average of three experiments.

[c] Cultivation conditions.

Figure 13: a. Construction of WBG8. WBG8 carries two copies of the *groE* operon. The Lower one is wild-type gene at the *groE* locus, the upper one is the modified *groE* gene without carrying CIRCE. The modified *groE* locates at the *amyE* locus. b. PCR confirmation of the integration of the *groE*(ΔIR) on the chromosome. Lane 1: PCR using primer GROEP and GROEPE, the upper band is from the wild type *groE* gene (band b) with the 27 bp inverted repeat sequence, the lower band is modified *groE* with IR deleted (band c). Lane 2: PCR product using GROEPE and upstream pDL primer, pH1 (band a). This band is specific for the modified *groE* [*groE*(ΔIR)].

a


amyE
cat
P groE
groESL
amyE
amyE
Chromosome
amyE
cat
PgroE
groESL
amyE
GROEP
GROEPE
pH1
P groE
IR
groESL
GROEP
GROEPE

b

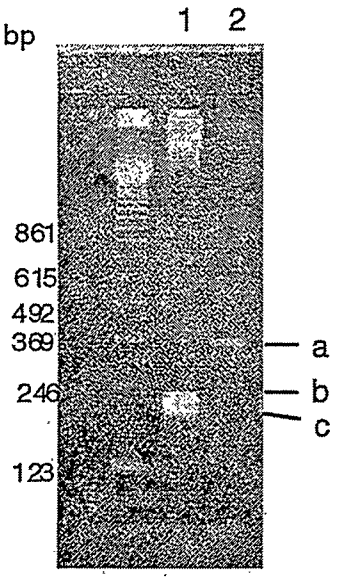
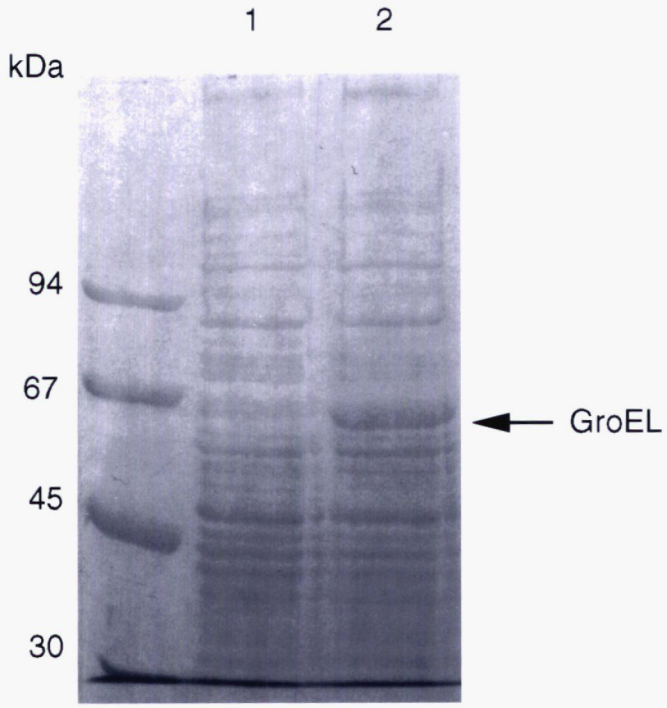

bp
1
2
861
615
492
369
246
123
a
b
c

Figure. 14 : SDS-PAGE gel profile of the cellular proteins in WBG8. Lane 1 is the control strain WBG2, which has pDL2 integrated at the *amyE* locus. Lane 2 is the sample from WBG8. An overproduced protein band with the molecular mass of 66 kDa is indicated. This overproduced protein is expected to be GroEL.


1
2
kDa
94
67
45
30
GroEL

V.4 Half-lives of *groE* mRNA under heat-shock and non-heat shock conditions

Half lives of the *groE* transcripts were determined by northern blot hybridization (Figure 15a and b) and primer extension (Figure 15c), the half-life of the *groE* transcript was about 5 minutes when the cells were cultured at 37°C or 48°C (Figure 15a, 15b and 15c, Table 11). Figure 16 is decay curve of the *groE* mRNA by using the result shown in Figure 15c. The plots of other data are not shown here. Similar results were observed for the *groE-bgaB* transcript from WBG2, the half-lives of *groE-bgaB* mRNA were about the same (3.6 and 2.8 min.) at 37°C and 48°C (Figure 15d, Table 11). Therefore, mRNA stability is not a key factor for the rapid increase of the *groE* mRNA level during the early phase of the heat shock response.

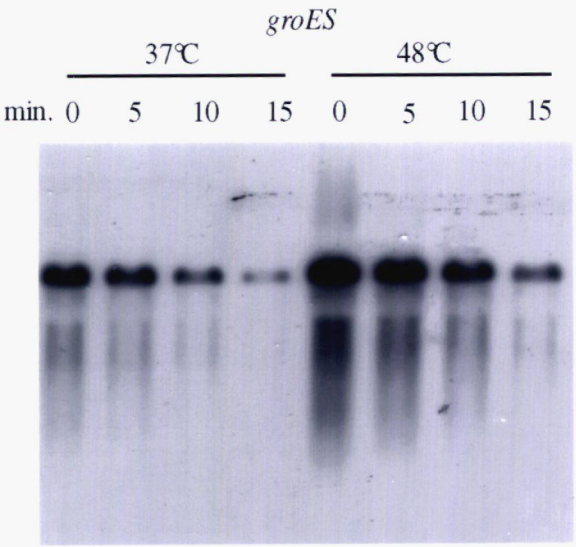
V.5 The functional role of CIRCE on the stability of the *groE* transcript

Although mRNA stability is not responsible for the increase of *groE* mRNA under heat-shock condition, it is still interesting to know whether CIRCE affects the stability of the *groE* mRNA. To address this issue, the half-life of the *groE* transcript from WBG8 was determined. In WBG8, there are two different *groE* transcripts. The longer one corresponds to the transcript from the wild type *groE* and the shorter one is from the modified *groE* in which CIRCE was deleted. When WBG8 was cultured at 37°C, the half life of the *groE* mRNA (without CIRCE) was 17 minutes (Figure 17a, Table 11), which increased about 3 times in comparison with that of the wild-type *groE* mRNA. The *groE* mRNA without CIRCE (shorter one) showed a half life of 7.5 minutes at 48°C (Figure 17a, Table 11). This is comparable to that (5 minutes) of the wild-type *groE* mRNA under the same condition. The same experiment was

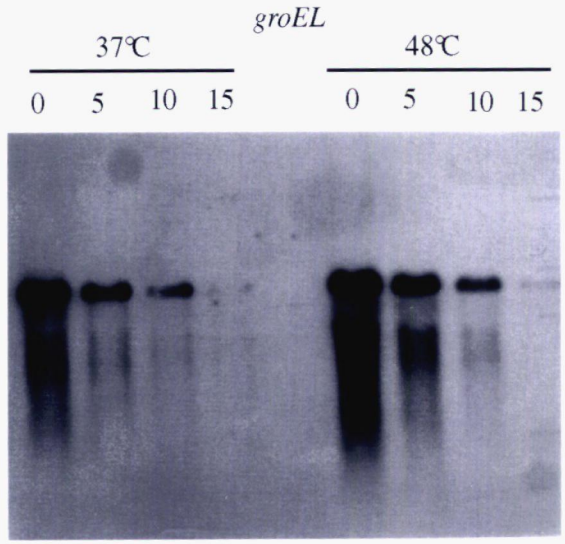
performed in WBG2, which carried the *groE-bgaB* fusions without the CIRCE sequence. The *groE-bgaB* fusion transcripts without CIRCE showed the half-life of 20 minutes (Figure 17b), which increased about 6 times more than that of *groE-bgaB* (with the IR) at 37°C. (Figure 17b, Table 11). These results indicated that the presence of CIRCE at 5' end of mRNA could reduce the half-life of the transcript under the non-heat-shock condition.

Figure 15: Half life of *groE* mRNA determined by Northern blot hybridization with RNA isolated from *B. subtilis* 168 probed with the *groES* probe (a) and the *groEL* probe (b). c) Half life of the *groE* mRNA determined by primer extension. d) Half life of the *groE-bgaB* transcript determined by primer extension.

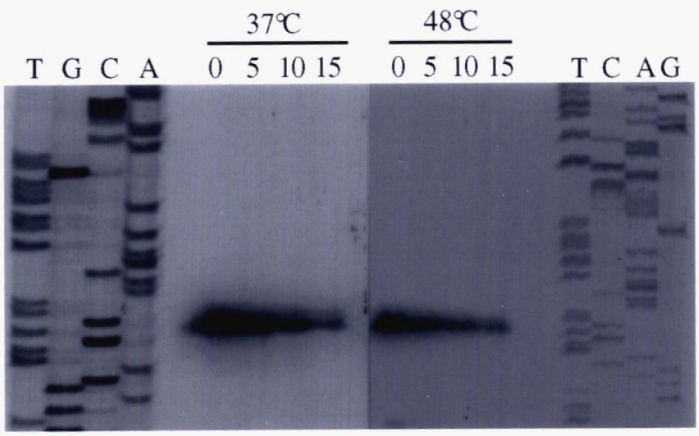
a


groES
37℃
48℃
min. 0 5 10 15 0 5 10 15

b


groEL
37℃
48℃
0 5 10 15 0 5 10 15

c


37℃
48℃
T G C A 0 5 10 15 0 5 10 15 T C A G

d

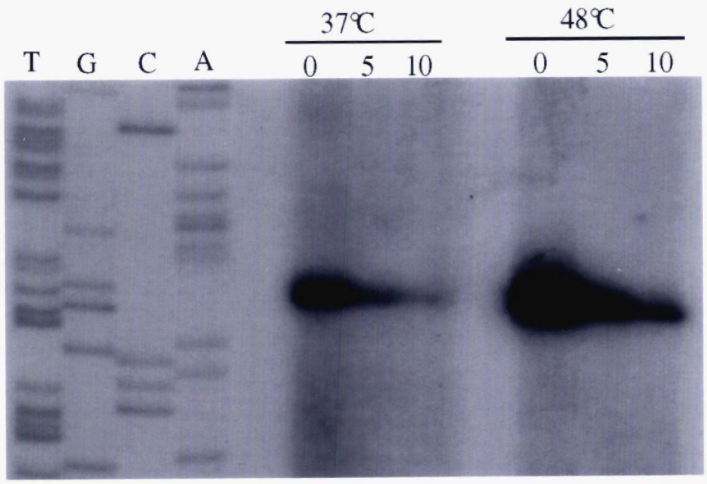

37℃
48℃
T G C A 0 5 10 0 5 10

Figure 16. Rate of the *groE* mRNA decay under heat shock (diamond) and non-heat shock (square) conditions by using the data from Figure 15c.

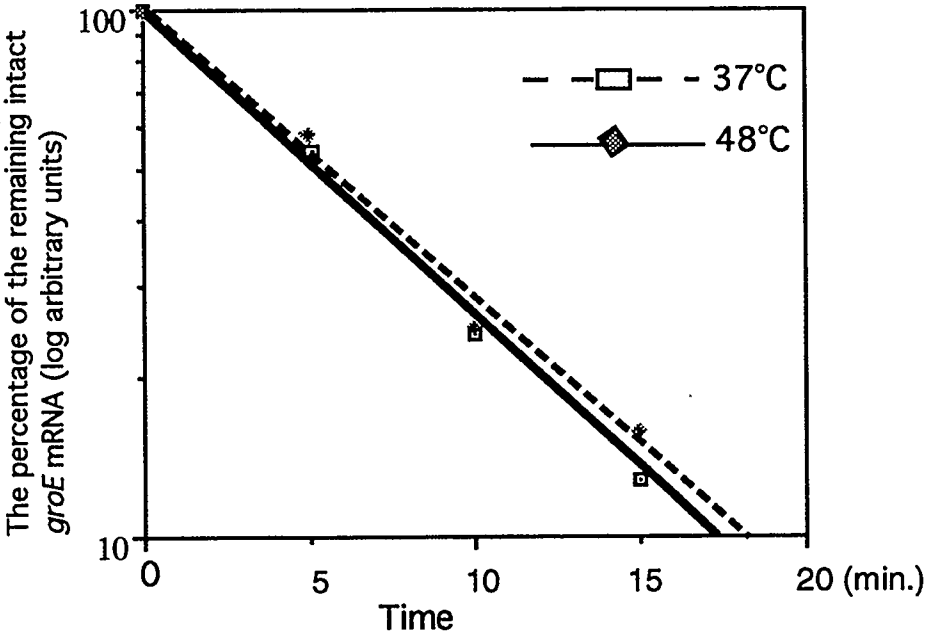
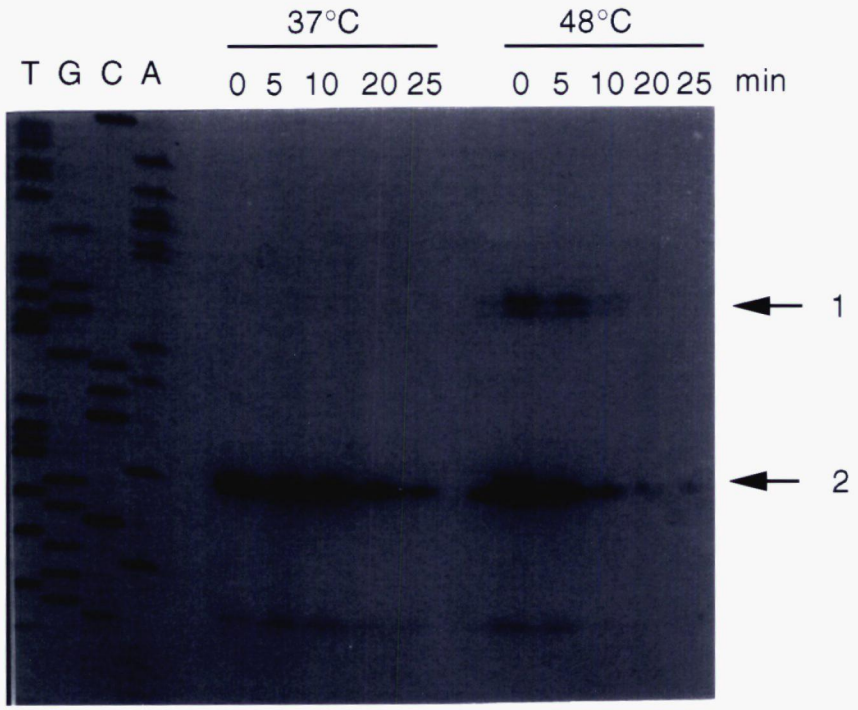

100
10
The percentage of the remaining intact groE mRNA (log arbitrary units)
37°C
48°C
0
5
10
15
20 (min.)
Time

Figure 17: Half life of *groE*(ΔIR) mRNA (with the deletion of CIRCE) (a) and *groE*(ΔIR)-*bgaB* mRNA (with the deletion of CIRCE) (b) determined by primer extension. Transcript 1 is from the native *groE*. Transcript 2 comes from *groE*(ΔIR).

a


37°C
48°C
T G C A
0 5 10 20 25
0 5 10 20 25
min
1
2

b

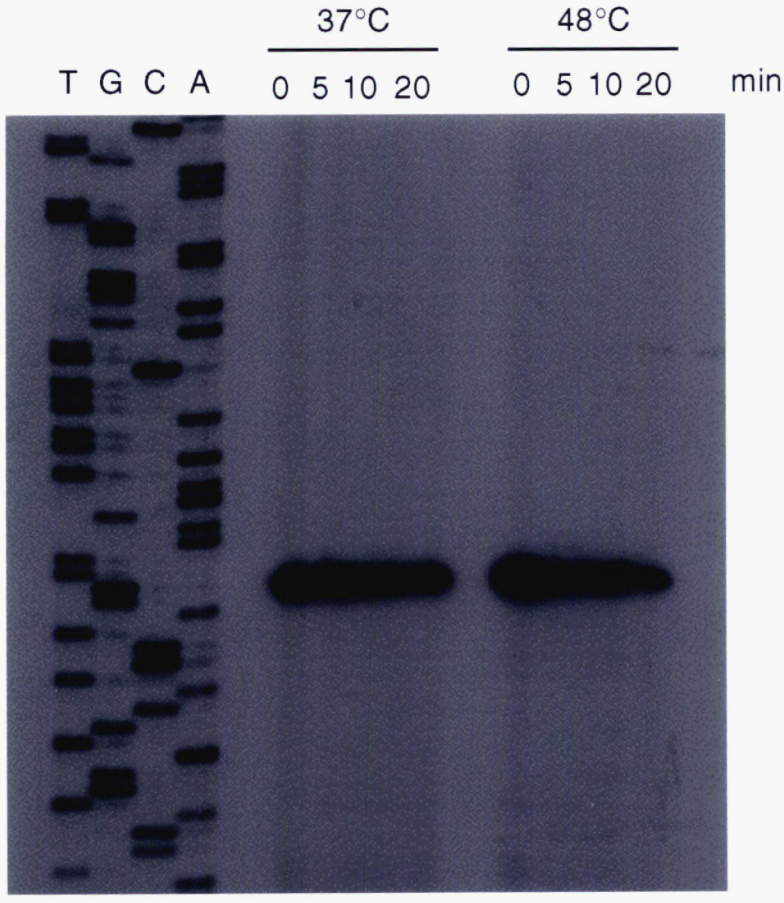

37°C
48°C
T G C A
0 5 10 20
0 5 10 20
min

Table 11. mRNA stability of *groE* and its derivatives under heat shock and non-heat shock conditions.

Strain	mRNA	Methods	Half life of mRNA (min)	
			37°C	48°
B. subtilis 168	*groE*	Primer Extension	5.5	5.0
B. subtilis 168	*groE*	Northern blot (*groES*)[a]	5.8	5.0
B. subtilis 168	*groE*	Northern blot (*groEL*)[b]	5.0	5.5
WBG8	*groE* (ΔIR)	Primer Extension	17.0	7.5
WBG2	*groE*-(+IR)-*bgaB*	Primer Extension	3.6	2.8
WBG4	*groE*-(5 bp+IR)-*bgaB*	Primer Extension	3.4	2.8
WBG7	*groE*-(21bp+IR)-*bgaB*	Primer Extension	5.2	4.2
WBG3	*groE*-(ΔIR)-*bgaB*	Primer Extension	20.5	7.2

[a] Using *groES* as the probe.
[b] Using *groEL* as the probe.

V.6 Effects of varying the distance between the transcription start site and CIRCE on *groE* expression

To further understand how CIRCE exerts its negative role in controlling *groE* expression, the distance between the *groE* transcription start site and CIRCE was increased by inserting DNA fragment of 5 to 21 bp. Table 12 shows the inserted sequences between the *groE* transcription start site and CIRCE. These *groE* promoter derivatives were inserted into the *Eco*RI and *Bam*HI digested pDL to generate pDL4 to PDL7. These pDL plasmids were integrated to the chromosome in *B. subtilis* 168. The resulting strains were designated WBG4, WBG5, WBG6, and WBG7. The BgaB activity was gradually increased as the distance between the *groE* transcription start site and CIRCE increased. Increase of the distance up to 17 bp can totally abolish the negative effects of CIRCE in regulating the *groE* expression at 37°C (Table 10). This increase is not due to changes in the transcription start site or the changes in the stability of *groE-bgaB* transcripts, since the primer extension results showed that the same transcription start site was used and the half lives of *groE-bgaB* with 5 and 21 bp inserts were similar to that of *groE-bgaB* (WBG2) at both 37°C and 48°C (Figure 18a and 18b, Table 11). Thus, the short distance between CIRCE and transcription start site is essential to CIRCE to exert its negative regulatory effects.

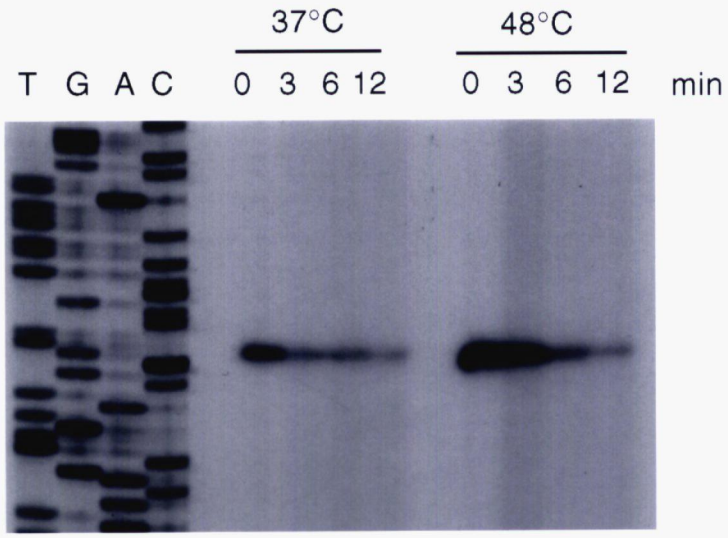
Table 12: Sequence inserted between the *groE* transcription start start and the inverted repeat sequence

plasmid	insertion	Sequence
pDL2	0	*▼ TTGTGTT
pDL4	5	TTG*AGATC*TGTT
pDL5	9	TTG*AGATCGATC*TGTT
pDL6	17	TTG*AGATCCTGATCAGGATC*TGTT
pDL7	21	TTG*AGATCCTGATCGATCAGGATC*TGTT

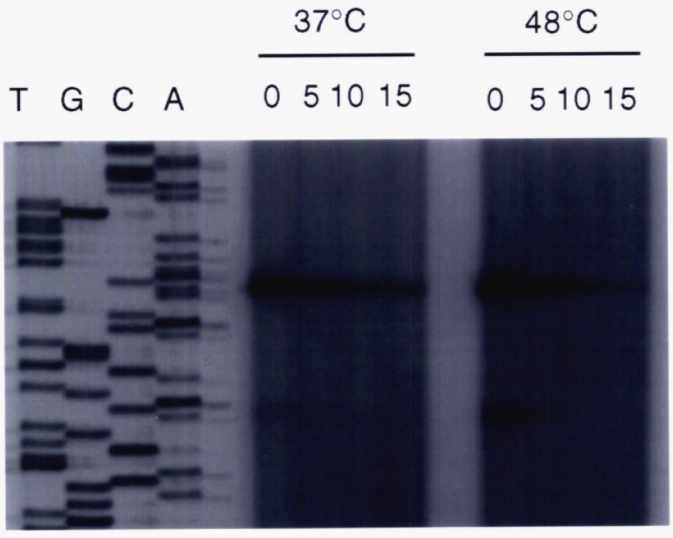
An asterisk indicates the *groE* transcription start site and the insertion site is marked by an arrow. Sequences inserted are shown in italic

Figure 18. Half life of the *groE*(+5 bp)-*bgaB* mRNA (a) and the *groE* (+21bp)-*bgaB* mRNA (b) determined by primer extension. pDL6, which containing a 17-bp insert between the *groE* transcription start site and CIRCE was used as the sequence ladder in (b). Thus the transcripts of *groE* (+21 bp) shows a 4-bp upshift relative to the transcription start site "T".

a



b



V.7 Other regulatory mechanisms may mediate the expression of *groE* under the heat shock condition

In the case of WBG3, WBG6 and WBG7, the negative role of CIRCE was abolished by deletion of CIRCE or by insertion of short sequences between CIRCE and transcription start site. However, a 2-fold increase in BgaB specific activity was observed in WBG3, WBG6 and WBG7 following the shift of the culture temperature from 37°C to 48°C (Table 10). This enhancement was *groE* specific. No increase in BgaB activity could be detected under the same condition by using the *B. subtilis xyl* (xylose operon) promoter as the control (WBG17, Table 10). Therefore, some other regulatory mechanisms may exist to increase the expression of *groE* under heat shock.

V.8 Discussion

V.8.1 CIRCE is a negative regulatory element for *groE* expression. Deletion of the entire CIRCE resulted in a high level expression of *groE-bgaB* constitutively. It clearly demonstrated that CIRCE was a negative element controlling the *groE* expression.

V.8.2 The stability of the *groE* mRNA is not changed under both heat shock and non-heat shock conditions. For the ribonuclease model, either the specific ribonuclease that recognized CIRCE sequence was heat-inactivated or the secondary structure of CIRCE which serves as the target site was changed at high temperatures. The *groE* mRNA of *groE* becomes more stable under heat shock. Our data (the mRNA half lives) did not support this model, since the half lives of the *groE* transcript show no changes under

both heat-shock and non-heat shock conditions. The *groE* mRNA has a short half life (3 to 5 minutes) under both heat shock and non-heat shock conditions. Our results are also consistent with the transcriptional analysis of the two *groE* genes in *S. colicolor* A3 (Duchene *et al.*, 1994). The conserved CIRCE sequence can be found in the two *groE* genes in this bacterium. The stability of the *groE* mRNA in *S. colicolor* A3 was reported to have no drastic changes under heat shock and non-heat shock conditions.

V.8.3 The presence of CIRCE can destabilize the *groE* transcripts under the non-heat shock condition. Although the stability of the *groE* transcript is not important for the heat induction of *groE,* our result indicated that deletion of CIRCE can prolong the half life of the *groE* transcripts at 37°C (Figure17a, Table 11, WBG8). This suggests that CIRCE may serve as a mechanism to ensure the rapid turnover of the *groE* transcripts in *B. subtilis* under non-heat shock conditions. This mechanism may allow fine-tuning of the GroE level in the cell.

V.8.4 CIRCE does not functions as a transcription terminator. CIRCE (IR) could be a transcription terminator. At low temperature, the transcription of *groE* would be terminated, and at high temperature, this effect would be abolished. The transcription of *groE* was then increased. If this assumption is right, insertion of short sequences between the transcription start site and CIRCE should not affect the terminator function of CIRCE, on the condition that the inserted sequences show no homology to CIRCE sequence. Our results also did not support this hypothesis, since the insertion of 5, 9, 17, 21 bp DNA fragment gradually abolished the negative regulatory

function of CIRCE (Table 10). Other pieces of evidence also do not support this hypothesis: 1) Similar amount of transcripts were observed when *groE* was used as the template for *in vitro* transcription at 37°C and 48°C (Chang *et al.*, 1994); 2) CIRCE can be found in the *groE* operon ina thermophilic bacterium, *B. stearothermophilus* which grows at 55°C (Schumann, 1993). In this case, *B. stearothermophilus groE* is heat-inducible when the cultivated temperature is elevated to a higher temperature. At 55°C, CIRCE is unlikely to form any stable secondary structure.

V.8.5 CIRCE may be a binding site for a putative repressor. Our data (Table 12) favor the operator model for CIRCE to exert its negative regulatory effects. At low temperature, a repressor binds to the CIRCE sequence. Following the temperature upshift, the repressor is inactivated directly by heat or indirectly by proteolytic degradation or chemical modifications. The *groE* transcription was increased due to derepression. In order to exert its function efficiently, the binding site for repressor usually overlaps with the promoter or is located closely to the promoter (Lanzer and Bujard, 1988). The further away is the binding site, the less is the repression effect. This is well demonstrated in *E. coli* through systematic changes of the distance between the operator and the promoter (Morita *et al.*, 1988). Other pieces of evidence are also consistent with the operator model which are listed as follows: 1) CIECE can be located upstream of the transcription start site in the *groE* operon of *S. albus* and the *dnaJ* operon of *L. lactis* (Guglielmi *et al.*, 1993; Van Asseldonk *et al.*, 1993); 2) By using the crude extracts and the partial purified fractions from *S. coelicolor* A3, the mobility of the CIRCE

sequence from the *groE* operon of this strain was retarded (Duchene *et al.*, 1994).

V.8.6 Existence of other positive factors in regulating the *groE* expression. Our data also suggest that positive factor(s) may be involved in regulating the *groE* expression under the heat shock condition. This was supported by the fact that the activities of BgaB were consistently increased about 2 times in WBG3, WBG6, WBG7, WBG15, WBG16 (Table 6 and Table 10). In these strains, the negative effect from CIRCE was abolished either by deletion of the entire CIRCE or by increasing the distance between CIRCE and the transcription start site. This increase may be due to the modification of σ^A to increase its affinity to the promoter or by some other positive factors that can enhance the transcription efficiency of *groE*. In WBG15, the *groE-bgaB* contains only the 46-bp *groE* promoter, it is unlikely to have other cis-acting sequence in this region for other regulating factors. However, the presence of the positive regulatory factor was also supported by the isolation of a *B. subtilis* mutant that shows a low level of *groE* induction under the heat shock condition. This mutation was mapped between *leu* and *phe* locus that is not in the *groE* locus on the chromosome (Miller *et al*, 1993). It is possible that this factor may increase the activity of the σ^A containing RNA polymerase to the *groE* promoter.

V.8.7 mRNA processing was not observed in the *groE* transcript of *B. subtilis*. Although our study shows that the stability of *groE* mRNA does not change under heat shock and non-heat-shock conditions, it is still possible that RNA processing can be a mechanism in regulating the expression

of *groEL* and *groES*. In *A. tumefaciens*, a heat-shock dependent processing of *groE* transcript was observed (Segal and Ron, 1995). It generates a *groES* transcript with short half life and a *groEL* transcript with long half life. This processing is mediated by an inverted repeat which shows no homology with CIRCE. RNA processing is suggested to serve as a mechanism to maintain the ratio of GroES and GroEL *in vivo*. In the case of *B. subtilis*, our Northern blot analysis indicated that this phenomenon did not exist in *B. subtilis* (Figure 15a and 15b).

In conclusion, 1) CIRCE may serve as a repressor binding site; 2) The presence of CIRCE can destabilize the *groE* transcripts at 37°C; 3) Positive regulatory factor(s) may exist which can enhance the expression of the *groE* transcription.

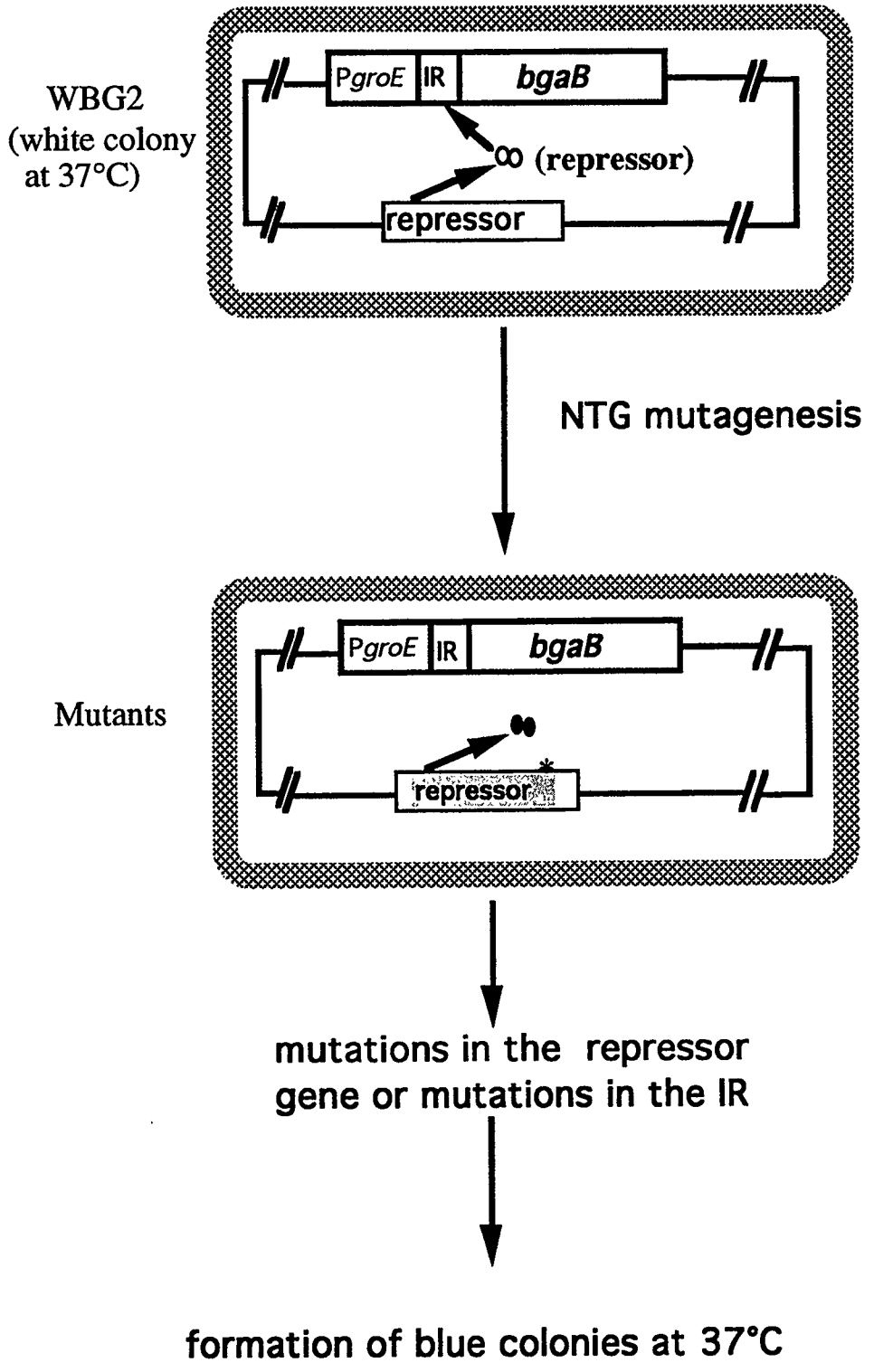
VI Isolation and characterization of the *B. subtilis groE* regulatory mutants: evidence for *orf39* at the *dnaK* locus encoding the repressor controlling the expression of *dnaK* and *groE*

VI.1 Introduction

Our previous results (Chapter IV & V) demonstrate that the σ^A RNA polymerase transcribes the *B. subtilis groE* under both the heat shock and non-heat shock conditions. CIRCE acts at the DNA level to negatively regulate the expression of *groE* and is likely to function as an operator. If this is the case, a repressor should exist in the cell to interact with CIRCE and reduce the transcription of *groE* under non-heat shock conditions. Since the *dnaK* operon also contains a similar CIRCE sequence, this repressor is likely to regulate the expression of both *groE* and *dnaK*. To determine the existence of such a repressor gene in *B. subtilis*, the regulatory mutants which exhibit constitutive expression of both *groE* and *dnaK* at high levels under the non-heat shock condition should be isolated.

Figure 19 illustrates the strategy to isolate regulatory mutants for *groE* and *dnaK*. *B. subtilis* WBG2 was mutagenized by N-methy-N'-nitro-N-nitrosoguanidine and plated on X-gal containing plate. Since the expression of *bgaB* was regulated by the *groE* promoter with the downstream of the IR sequence (CIRCE), the expression of *bgaB* is at low levels under the non-heat shock condition (37°C). WBG2 colonies were white in color on X-gal containing TBAB plate at 37°C. These colonies will turn blue at 48°C. If the putative repressor was completely inactivated by the NTG mutagenesis, *bgaB* in WBG2 should be expressed at a high level under normal culture conditions

Figure 19: Isolation of *groE* regulatory mutants. NTG is N-methy-N'-nitro-N-nitrosoguanidine. P*groE* is the promoter of *groE*. *bgaB* is the ß-galactosidase gene from *B. stearothermophilus*. IR is the inverted repeat sequence (CIRCE) downstream of the *groE* promoter. The open circle is the active repressor. The closed circle represents the inactivated repressor. The asterisk indicates the mutation in the repressor gene.


WBG2
(white colony
at 37°C)
PgroE
IR
bgaB
(repressor)
repressor
NTG mutagenesis
Mutants
PgroE
IR
bgaB
repressor
mutations in the repressor
gene or mutations in the IR
formation of blue colonies at 37°C

(37°C). Colonies of the regulatory mutants should show dark blue color. The expression of *groE* and *dnaK* should be at high levels as well.

In this study, we have isolated the regulatory mutants. We also found that the first gene in the *dnaK* operon, *orf39*, could restore the phenotype of these mutants back to normal. Furthermore, the nature of the mutations in these mutants was determined and the binding of ORF39 to CIRCE sequence in *groE* was demonstrated.

VI.2 Isolation of the *groE* regulatroy mutants

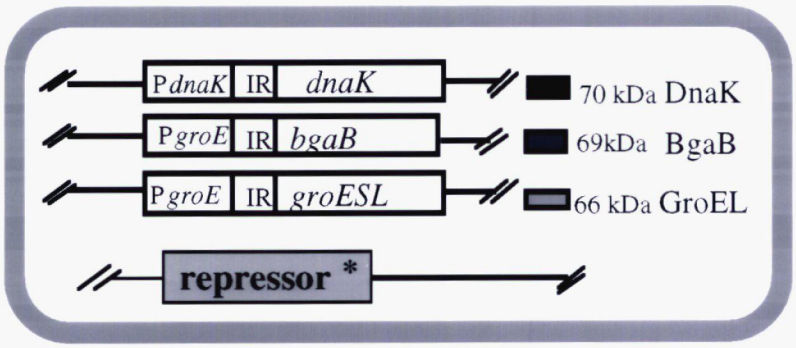
WBG2 was mutagenized with N-methy-N'-nitro-N-nitrosoguanidine. The cells were plated on the X-gal containing TBAB plate and incubated at 37°C. Seventeen dark blue colonies were observed from a total of about 5×10^5 NTG-treated WBG2. Many colonies with a frequency of 2 per 1,000 plated cells also showed light blue color. Seven (WBG101 to WBG107) of the 17 dark blue colonies were selected for further characterization.

VI.3 Characterization of the *groE* regulatory mutants

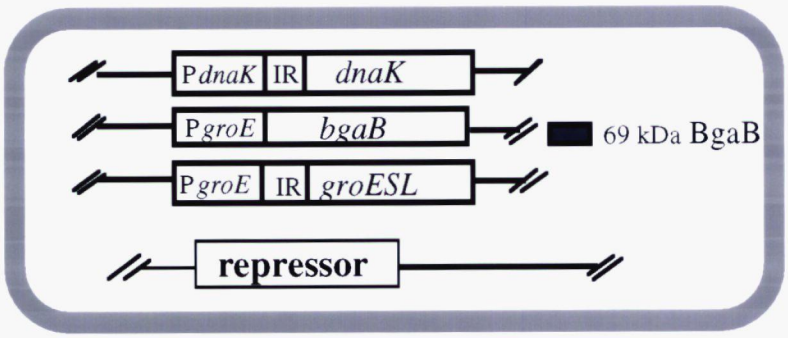
Figure 20a shows the expected features of the *groE* regulatory mutants. At least three genes *dnaK*, *groE* and *bgaB* were regulated by CIRCE in WBG2. Inactivation of the repressor should result in the constitutive expression of *groE*, *dnaK* and *bgaB* under the non-heat shock condition (37°C). WBG3 was used as the control strain, which is identical to WBG2 except CIRCE was deleted from the *groE-bgaB* transcription fusion. Figure 20b shows the phenotype of WBG3, *bgaB* is expressed constitutively at a high level under both non-heat shock and heat shock conditions, whereas the expression of *groE*

Figure 20: a) The expected phenotype of the *groE* regulatory mutants. b) The phenotype of the control strain, WBG3. c) SDS-PAGE of the total proteins from the regulatory mutants. The gel was stained with coomassie blue. Lane1: the control strain WBG2, Lanes 2 to 4: mutants WBG101 to WBG103; Lane 5: the control strain WBG3; Lanes 6 to 9: mutants WBG104 to WBG107.

a


PdnaK
IR
dnaK
70 kDa DnaK
PgroE
IR
bgaB
69kDa BgaB
PgroE
IR
groESL
66 kDa GroEL
repressor *

b


PdnaK
IR
dnaK
PgroE
bgaB
69 kDa BgaB
PgroE
IR
groESL
repressor

c

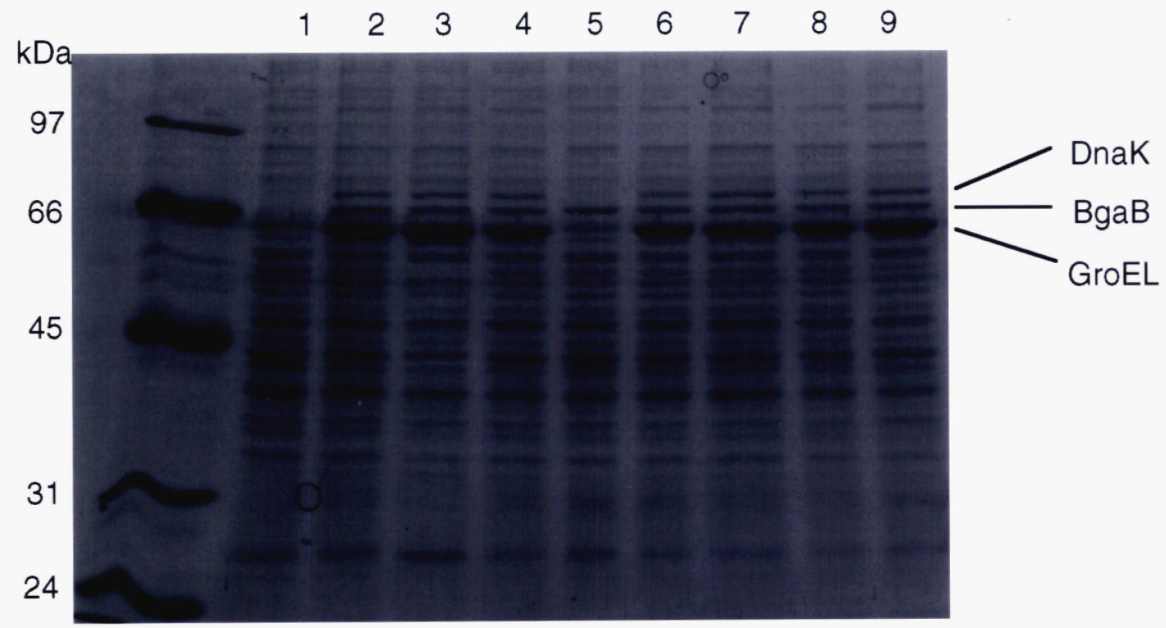
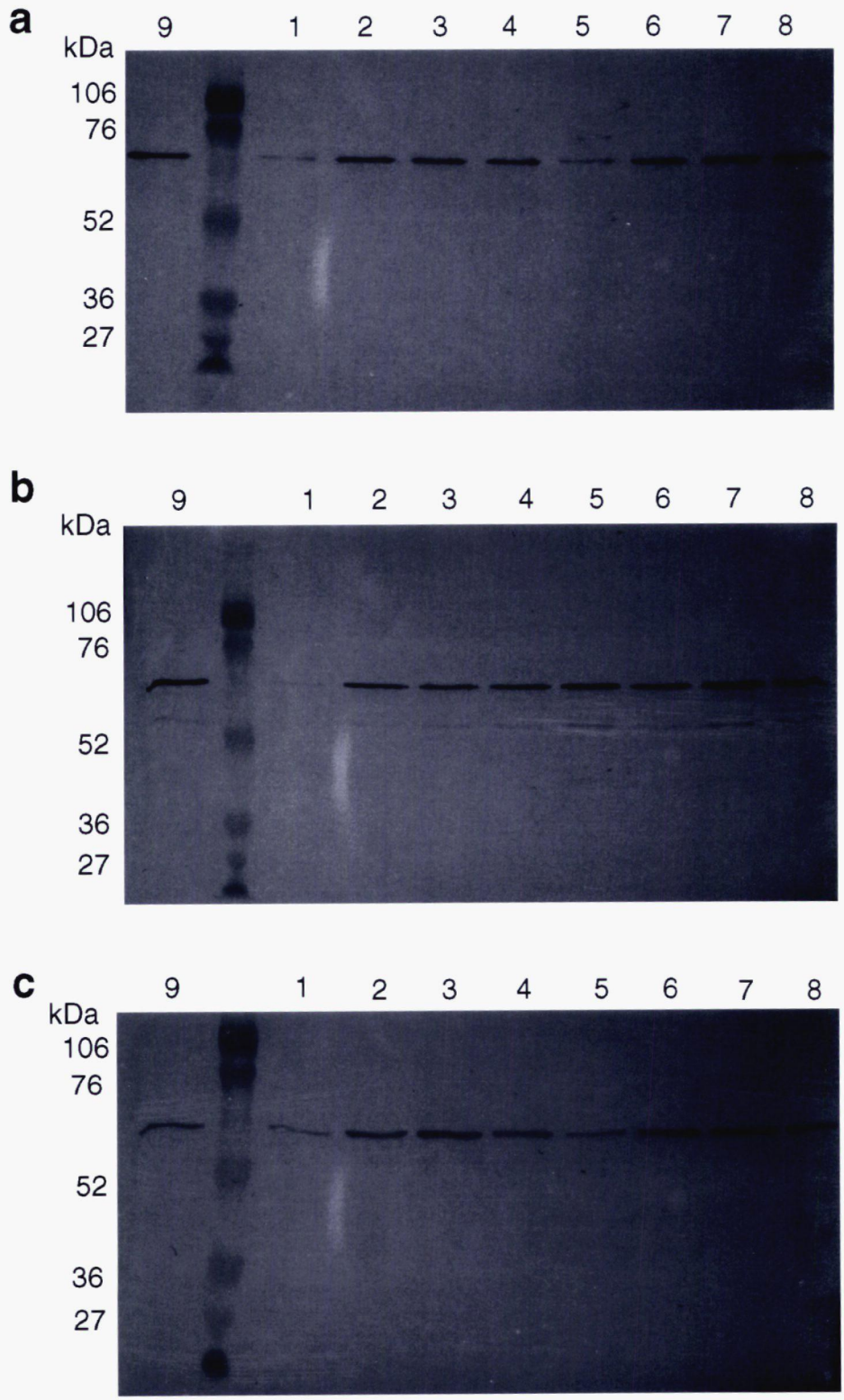

1 2 3 4 5 6 7 8 9
kDa
97
66
45
31
24
DnaK
BgaB
GroEL

Figure 21:. Western blot analysis of cellular proteins from different *groE* regulatory mutants. a) with DnaK specific antiserum; b) with BgaB specific antiserum. c) with GroEL specific antiserum. Lane 1: *B. subtilis* WBG2, the control strain; Lanes 2 to 4 mutants WBG101 to WBG103; Lane 5: the control strain WBG3; Lanes 6 to 9: mutants WBG104 to WBG107


a
9 1 2 3 4 5 6 7 8
kDa
106
76
52
36
27
b
9 1 2 3 4 5 6 7 8
kDa
106
76
52
36
27
c
9 1 2 3 4 5 6 7 8
kDa
106
76
52
36
27

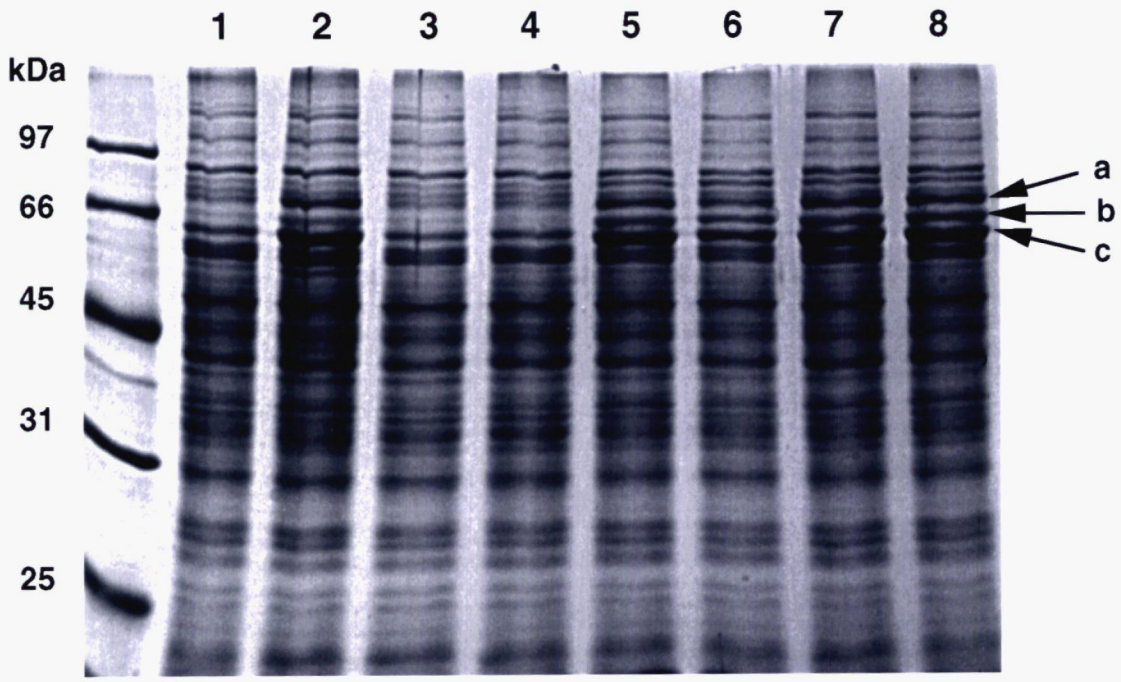
and *dnaK* is at low levels under the non-heat shock conditions. The phenotype of the mutants was studied by culturing the mutants at 37°C. The cells were collected and lyzed by sonication. The protein profile of mutants was analyzed by SDS-PAGE and Western blot with DnaK, BgaB, and GroEL specific antibodies. Relative to WBG2, three overproduced protein bands were found in these mutants (Figure 20c). The 70-kDa and 66-kDa protein were correspondent to DnaK and GroEL, respectively. The 69-kDa protein showed exactly the same position as BgaB in WBG3. Western blot results indicated that they were indeed DnaK, BgaB and GroEL respectively (Figure 21a, b, c). There are two possibilities to overproduce DnaK, BgaB and GroE. (i) Mutation is in the putative repressor gene, and (ii) Mutations are in the CIRCE sequences of *groE-bgaB*, *dnaK* and *groE*. To eliminate the second possibility, one of the mutants, WBG101, was used for genomic sequencing. No mutations could be found in the *groE* promoter and the CIRCE sequence in *groE-bgaB* locus. Therefore, WBG101 was postulated to carry mutation(s) in the putative regulatory repressor gene.

VI.4 *orf39* in the *dnaK* operon can restore the *groE* regulatory mutants to normal

Recent study by Schultz *et al.* suggests that either *orf39* or *grpE* can be the negative factor to regulate the expression of *groE*. Disruption of the entire *dnaK* operon, which includes four genes *orf39*, *grpE*, *dnaK* and *dnaJ*, increases the level of the *groE* mRNA at 37°C. Inactivation of both *dnaK* and *dnaJ* does not affect the expression of *groE* (Schultz *et al.*, 1995). To investigate whether *orf39* is the regulatory gene controlling *groE* and *dnaK* expression, a 1.1-kb DNA fragment containing *orf39* (nucleotides 221 to 1319 of the reported

sequence, Wetzstein *et al.*, 1992) was amplified through PCR by using the *B. subtilis* chromosome DNA as templates. This promoterless *orf39* was inserted into the vector pUB-P43. Expression of *orf39* was controlled by the constitutively expressed promoter P43 (Wang & Doi, 1987). The resulting plasmid was designated pUB-P43ORF39. When the pUB-P43ORF39 was introduced to the seven mutants (WBG101 to WBG107) by transformation, white colonies on X-gal containing TBAB plate resulted. The protein profile from one of the mutants WBG101 carrying the pUB-P43ORF39 was analyzed by SDS polyacrylamide gel. Figure 22 shows the SDS-PAGE analysis by using the crude extract from cells cultivated under heat shock and non-heat shock conditions. At 37°C, overproduction of DnaK, GroEL and BgaB could be observed only in WBG101 carrying pUB-P43 (Lane 2), low level expression of these genes was observed in WBG101 carrying pUB-P43ORF39 (Lane 4). WBG2 carrying pUB-P43 (Lane 1) and pUB-P43ORF39 (Lane 3) express *groE*, *dnaK* and *bgaB* at low levels at 37°C. Expression of these genes (*dnaK, groE* and *bgaB*) at high levels could be observed for all of these strains under the heat shock condition (Lanes 5 to 8). These results suggest that expression of the functional *orf39* from the plasmid can restore the *groE* and *dnaK* regulatory mutants to the wild-type phenotype. Co-transformation of *orf39* containing DNA fragment and pUB-18 to the mutants can restore these mutants back to normal too.

Figure 22: SDS polyacrylamide gel analysis of cellular proteins from WBG101 transformed with pUB-P43ORF39. Lanes 1 to 4: WBG2 with pUB-P43, WBG101 with pUB-P43, WBG2 with pUB-P43ORF39, and WBG101 with pUB-P43ORF39 at 37°C. Lanes 5 to 8: the same order as lanes 1 to 4 except the cell were cultivated at 48°C. a) DnaK. b) BgaB. c) GroEL


1
2
3
4
5
6
7
8
kDa
97
66
45
31
25
a
b
c

VI.5 Sequencing analysis of the mutations in WBG101 to WBG107

To determine whether there is any mutation(s) in *orf39* in these mutants, genomic sequencing was performed. As shown in Table 13, three groups of mutants were found based on the location of the mutations in *orf39*. Therefore, mutation in *orf39* resulted in the constitutive expression of *groE-bgaB*, *groE*, and *dnaK*. Sequences of ORF39 homologs from other microorganisms are reported and aligned. Two of the amino acid residues changed, Gly63→Glu and Gly307→Asp, were found to locate in the highly conserved region in ORF39 (Figure 23). Besides the single mutation found in the reading frame of *orf39* in each mutant, there are two discrepancies in the *orf39* sequence relative to the sequence reported previously in the seven mutants (Weztstein *et al*, 1992). An extra G should be added between nucleotides 112-113 (GenBank accession number M84964) and a T at nucleotide 148 should be deleted. Both nucleotides were located upstream of the *dnaK* promoter. Genomic sequencing of this region in *B. subtilis* 168 also exhibited the same sequences as observed in these mutants at these two locations. Therefore, discrepancies at these locations were not considered as mutations.

Table 13. Mutations in *orf39* in the *groE* regulatory mutants (WBG101 to WBG107)

Mutant	Location of the mutation[a]	Base substitution	Amino acid replacement
Group 1 (WBG101, WBG104, WBG105)	1161	GGC → GAC	Gly-307 → Asp-307
Group 2 (WBG102, WBG106, WBG107)	606	TCC → TTC	Ser-122 → Phe-122
Group 3 (WBG103)	429	GGA → GAA	Gly-63 → Glu-63

[a] Location is numbered according to the reported DNA sequence (Wetzstein *et al,* 1992)

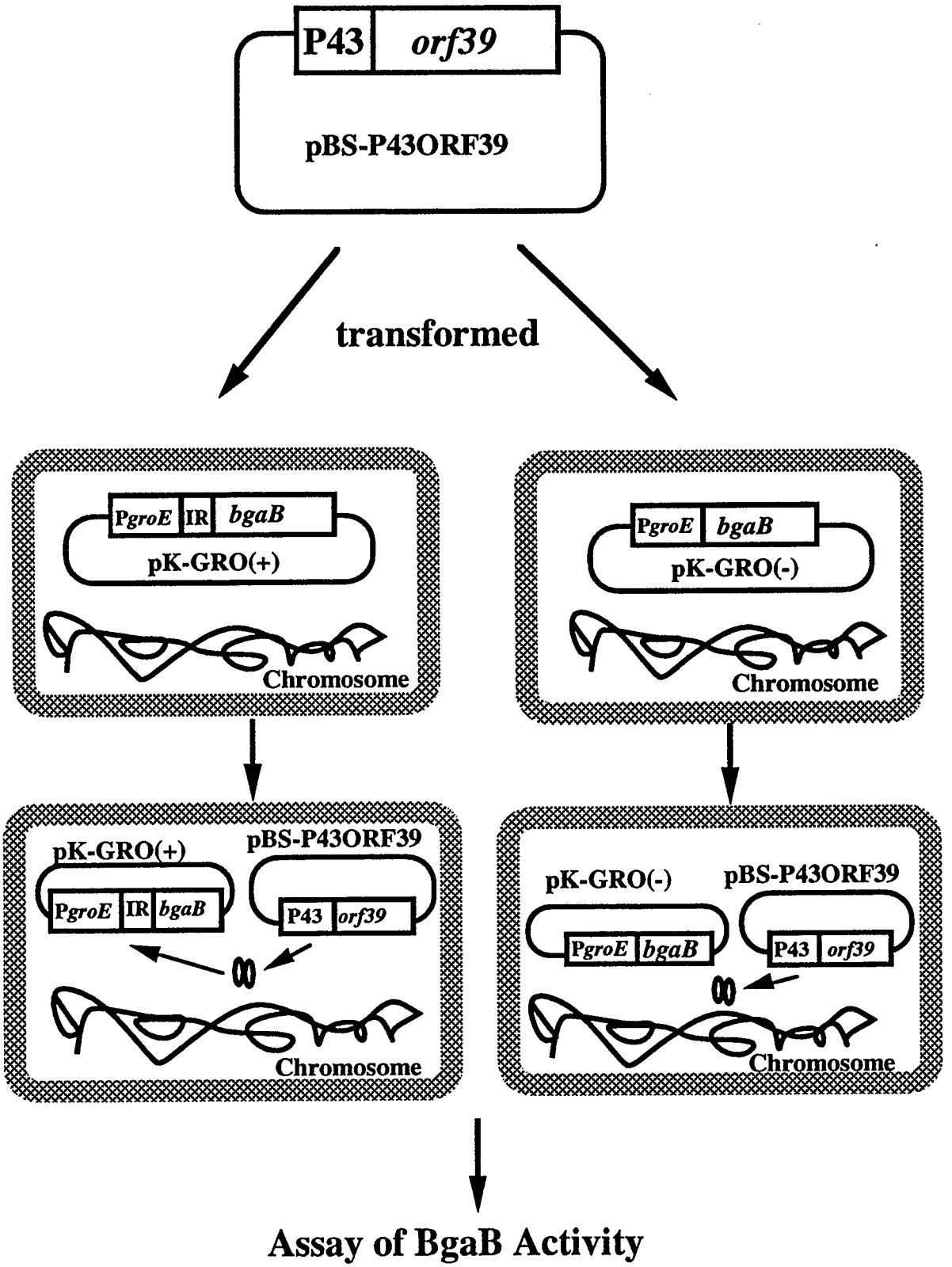
Figure 23: Sequence alignment of *B. subtilis* ORF39 with ORF39 homologs from other eubacteria. Identical and conserved residues are marked with an asterisk and a dot, respectively. Sequences in the highly conserved regions are shown in bold. Amino acid residues that were replaced in the mutants are shown in italic. BS: *B. subtilis*; CA: *Clostridium acetobutylicum;* SA: *Strphylococcus aureus*; LL: *Lactococcus lactis* and CT: *Chlamydia trachomatis*

```
ORF39BS     M---LTNRQLL----------ILQVIINDFIKSAQPVGSRTLSKDEITF    37
ORF39CA     ME--MEERHLK----------ILQAIINDYINNGEPVGSRTIAKYNLGI    38
ORF37SA     M---ITDRQLS----------ILNAIVEDYVDFGQPVGSKTLIERHNLNV    37
ORF39LL     M---ITERQRQ----------ILNLIVSLYAKDHTPIGSKSLLD--SIQA    35
ORF45CT     MENRIEMSQLRASKKDSKISYVLLMATKLYLESGQPVGSKLLKETYCSDL    50
            *   .. ..           .*   .. . .  *.**. . .

ORF39BS     SSATIRNEMADLEELGFIEKTHSSSGRVPSEKGYRYYVDHLLSPYKLTKS    87
ORF39CA     SSATIRNEMADLEEMGYIEQLHTSSGRKPSDKGYRLYVDRLMEIPSMSVE    88
ORF37SA     SPATIRNEMKQLEDLNYIEKTHSSSGRSPSQLGFRYYVNRLLEQTSHQKT    87
ORF39LL     SSATIRNDMKALERLGLIQKEHTSSGRIPSVSGYKYFVENVIQLEEFSQN    85
ORF45CT     SSATIRNYFAQLETDGFLRKQHISGGRIPTDLAFRYYADH--NAPFLEQE    98
            *.***** . .**  . . . * *.** *.  ... ....         .

ORF39BS     DLDQIHS-IFKEKIF--ELEKTVQKSAQILSDLTNYTSIVLGPKLSENYL   134
ORF39CA     EEMLINAIIDSALY--EIDKLVKQAMSLVSEMTKLTCVVKSLSARKSYI   136
ORF37SA     NKLRRLNQLLVENQY--DVSSALTYFADELSNISQYTTLVVHPNHKQDII   135
ORF39LL     DLFKYM-KAFDGDFY--RLSDLFKTAAKSLSELTGLTSFVLNAPQRDQQL   132
ORF45CT     EILAIQQKLTELPEYSKNIVKDLQKASEVLSDILQLPVCFSSPRFESDSV   148
            .           .  . .        .*..  .          .

ORF39BS     KQIQIIPIQPDMAVAILVTNTGHVENKTINFPTKMDLSDIEKLVNILNDR   184
ORF39CA     KSISLNIEPNMILCVFITDSGMIKNSIIRVKSNIENSSLERIANILNSK   186
ORF37SA     NNVHLIPANPNLVIMVIVFSSGHVEHVHLASDIPFSNDKLNTISNFVTNK   185
ORF39LL     VSFEAACLDNHSVLSVITLGTGEVRTNQFILPKSMTEADLAVFSNLVKER   182
ORF45CT     INIQLVAIDDQRVVFVLSTEFGQVFTDVLWLPEQLPENSLKRIEGFLQNY   198
               .. . .  . ..  . * .    .     .  ...  . ... .

ORF39BS     LSGVPMDELMERIFKEVVMYLRQHIKNYDNILDALRSTF--HSTNHVEKL   232
ORF39CA     LKGLTIEQINLEVINNIKKDLREYGHIFDCIMPNLYDIL--READSTE-V   233
ORF37SA     LT-----EFNQNLQDDIVSFVQSEQE--EIFINKLINTMNNHISNQSNSI   228
ORF39LL     LVGNKVIDIHYTLRTEIPQIVQRYFKVTSEVLQLFESIFDDLFKEH---L   229
ORF45CT     LRKQPSDSLLSQKEEDLGMVL------YNEVVVRYLTRYCHFSEED---L   239
            *       .    ...   .        .  .   .        .   .

ORF39BS     FFGGKINMLNQPEFHDITRVRSLLSLIEKEQ--DVLKLVQSPHTGISIKI   280
ORF39CA     YKEGTDNLFNYPEFKDIEKAKEFLSVID-DR--RILDTLFNASGGVTVNI   280
ORF37SA     YMGGKVTLIDALNESNVSSIQPILQYIESNRIAELLQDISSPN--INVKI   276
ORF39LL     TVAGHEMFDYAT-DNLAELYKLFS--DDERMLHEIREITNNDEMRAVKF   276
ORF45CT     YQTGLSLLKYETFKDPETLAQGLAFFENRKHMCQLLNTYLHKETPTAFI   289
              .*  ....  .  . .     .  .  .  .          .  .

ORF39BS     GKENDY---EEMENCSLITASYSVDQKQIGSIAIIGPTRMNYSRVVSLLQ   327
ORF39CA     GNENSI---KEARDFSVVSSVYKYNGRPLGTIGIIGPTRIPYSKVIKVIM   327
ORF37SA     GNEID----DSLSDISIVTSQYHFDETLKGQIAVIGPTAMHYQNVIQLLN   322
ORF39LL     DNDENF-----MKNLTIISQKFVIPYRGFGTLTVVGPVEMDYQRTLSVLD   321
ORF45CT     GRELTDIVGNTDPSCAVITIPYYMDRTPLGTFGVLGPMNLPYQQVFGTLS   339
            ... .       . ....  .        * ....**  . * ..   .

ORF39BS     HVTSDLSKALTS-LY----------------------DE----------   343
ORF39CA     EVVDQINNNLDK-MN----------------------NS----------   343
ORF37SA     RI----------------------------------W-----------   325
ORF39LL     LVAKVLTDSLSD-YYRYL-------------DGNHYEISK----------   347
ORF45CT     LFTERLNVILTQSFYKFKLSFRRPCPTDPRCSQRPAELTRSSSIKLLPAK   389

ORF39BS     ---   343
ORF39CA     ---   343
ORF37SA     ---   325
ORF39LL     ---   347
ORF45CT     ELS   393
```

VI.6 ORF39 can exert its negative function in *E. coli*

Our studies demonstrated that ORF39 was a negative regulatory factor for *groE* and *dnaK* expression. However, how this ORF39 exerts its negative role in regulating the *groE* and *dnaK* expression is still not clear. It is still possible that expression of *groE* and *dnaK* is mediated by several regulatory factors and ORF39 may be just one of them. Since expression of heat shock genes in *E. coli* is not regulated by the CIRCE mediated mechanism, CIRCE specific repressor should not exist in this host. In order to study the function of *orf39*, a binary plasmid system in *E. coli* was established as shown in Figure 24. The first plasmid was pBS-P43ORF39 that is a high copy number plasmid and contains *orf39* driven by the P43 promoter. The second plasmid was pK-GRO(+), which is a low copy number plasmid with the p15a replicon. This pK-GRO(+) carried the *bgaB* operon under the control of the *groE* promoter with CIRCE. If ORF39 functions as a repressor, the *bgaB* expression should be shut off by overexpressing *orf39* from the high copy number plasmid in the same host. WGE106 serves as a control. This strain carries plasmids pKGRO(+) and pBM-MTL (pBS with a fragment of mannitol dehydrogenase operon). Constitutive expression of *bgaB* was observed in WGE106 (Table 14). When the pBS-ORF39 was introduced into the *E. coli* DH5α strain containing pK-GRO(+) (pK184 with *bgaB* controlled by P*groE* with CIRCE), the BgaB activity is reduced (135 units, WGE103) at 37°C relative to that of the control strain (1904 units, WGE106). This indicates that the expression of *orf39* can reduce the expression of *bgaB* controlled by the *groE* promoter with CIRCE. This reduction of *bgaB* activity is CIRCE specific, since the expression of *orf39* cannot shut off the expression of *bgaB* controlled by the *groE* promoter without CIRCE (Table 14, WGE102). WGE101 and WGE104 were the negitive control

Figure 24: A binary plasmid system to study the function of *orf39* in *E. coli.* IR is the inverted repeat sequence (CIRCE). P*groE* is the *groE* promoter (-238 to +2). P43 is a SigA dependent promoter in *B. subtilis*. *orf39* is the first gene in *dnaK* operon in *B. subtilis*. *bgaB* is the galactosidase gene from *B. stearothermophilus.*


P43
orf39
pBS-P43ORF39
transformed
PgroE
IR
bgaB
pK-GRO(+)
Chromosome
PgroE
bgaB
pK-GRO(-)
Chromosome
pK-GRO(+)
pBS-P43ORF39
PgroE
IR
bgaB
P43
orf39
Chromosome
pK-GRO(-)
pBS-P43ORF39
PgroE
bgaB
P43
orf39
Chromosome
Assay of BgaB Activity

strains, which contained pK-BGAB (pK184 with promoterless *bgaB* insert) with pBS-ORF39 or pBS-MTL. Regardless of the conditions, they showed low BgaB activities. It is interesting to find that the BgaB activity in WGE103 increased 5.4 fold when the cell was cultured at 48°C (WGE103). This observation suggested that ORF39 could be heat-inactivated either directly (e.g. temperature sensitive repressor) or indirectly (degraded by heat-inducible protease or inactivated through chemical modification).

Table 14. A binary plasmid system to study the regulation of *groE-bgaB* expression by ORF39 in *E. coli*

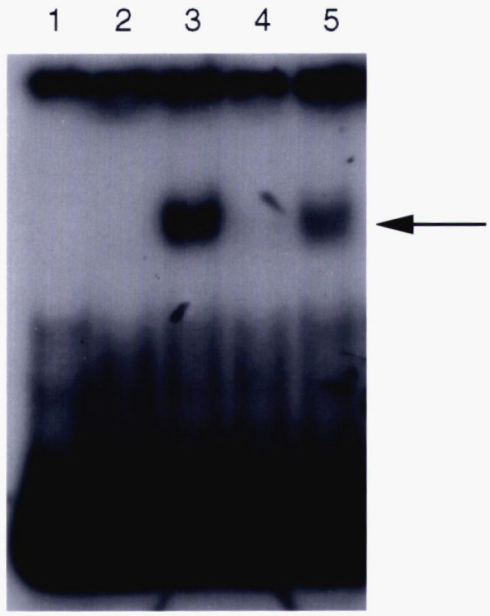
Strain	Plasmid 1	Plasmid 2	BgaB activity[a]	
			37°C[b]	48°C[b]
WGE 101	pBS-ORF39	pK-BGAB	7.9 ± 2	7.5 ± 1
WGE 102	pBS-ORF39	pK-GRO(-)	2041 ± 151	1884 ± 112
WGE 103	pBS-ORF39	pK-GRO(+)	135 ± 10	728 ± 26
WGE 104	pBS-MTL	pK-BGAB	3.1 ± 1	4.2 ± 0.3
WGE 105	pBS-MTL	pK-GRO(-)	1862 ± 182	1992 ± 63
WGE 106	pBS-MTL	pK-GRO(+)	1904 ± 178	2006 ± 172

[a] Data presented represent the average of the three independent experiments.
[b] Cultivation temperature.

VI.7 Gel mobility shift assay

Based on the above study, it is suggested that ORF39 could be a CIRCE specific repressor. In order to support this hypothesis, crude cell extracts were prepared from both the control strain [*E. coli* BL23 carrying pET29b(ΔXB)] and the ORF39 production strain [*E. coli* BL23 carrying pET-ORF39] for the band shift assay. Although ORF39 was produced at a high level, the majority of this protein was insoluble. However, there was a low percentage of ORF39 in the soluble form. The CIRCE-containing DNA probe was retarded by the cell extract containing ORF39 (Lane 3) (Figure 25). This retardation is CIRCE specific, since prebinding the cell extract with cold IR containing DNA probe can eliminate the gel shift (Lane 4), and prebinding with the sonicated salmon DNA does not significantly affect the interaction of the extract with the labeled CIRCE DNA probe (Lane 5). No band shift was observed when the extract was prepared from the strain that does not produce ORF39 (Lane 2). We conclude from these studies that ORF39 is the repressor for *groE* and *dnaK* expression. ORF39 exerts its function by binding to the inverted repeat sequences (CIRCE) located downstream of the *groE* and *dnaK* transcription start sites.

Figure 25: Gel retardation study of the CIRCE probe with crude extracts prepared from various *E. coli* strains. The strong band at the bottom of the gel is the free probe. The protein-probe complex is marked by an arrow. Lane 1: the labeled CIRCE probe alone. Lane 2: Crude extract from the control strain *E. coli* (DE3) [pET-29bΔ(XB)] was added. Lane 3: Extract from *E. coli* BL21(DE3)[pET-ORF39] was added. Lane 4: Same as lane 3 except 2 pmole of the unlabeled-IR probe was added for prebinding. Lane 5: Same as lane 3 except 4 μg of sonicated salmon sperm DNA was added for prebinding.

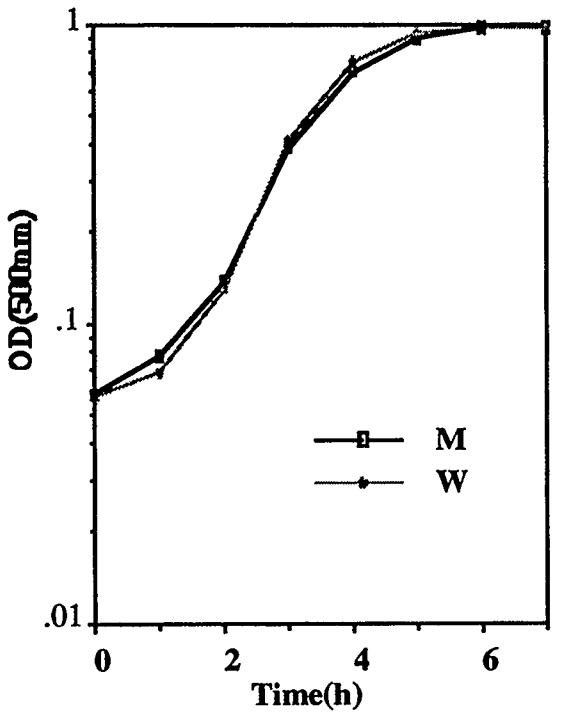

1
2
3
4
5

VI.8 Growth of WBG101 and WBG2 cultivated under different temperatures

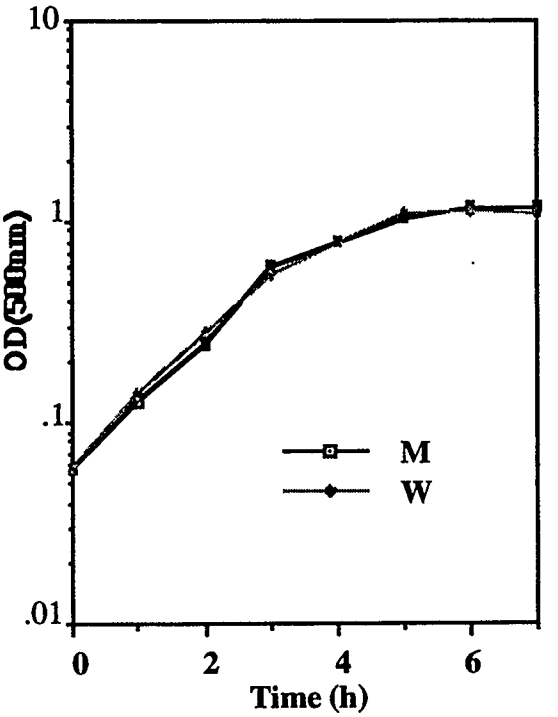
Since ORF39 is the negative regulatory factor controlling the expression of the *groE* and *dnaK*, inactivation of ORF39 resulted in the constitutive expression of at least *groE, dnaK, dnaJ, grpE*. The proteins encoded by these genes are GroEL, GroES, DnaK, DnaJ and GrpE. These proteins are known to function as molecular chaperones. It would be interesting to know whether the *orf39* mutant (WBG101) can survive better than the wild type *Bacillus subtilis* strain under heat shock conditions. Figure 26 shows the growth curve for WBG101 and WBG2 under different culture temperatures. At 37°C and 48°C, both the mutant (WBG101) and the wild type strain (WBG2) could grow well in the same manner (Figure 26a, b). If both cell strains were cultured at 48°C for 30 minutes, then moved to 52°C which is a lethal temperature for *B. subtilis*, they both grew well. The mutant clearly grew better than the wild type strain (Figure 26c). If WBG2 is cultivated at 52°C directly, the cells cannot grow at all (Figure 26d). However, with a seven hours lag, WBG101 can grow (Figure 26d). This result indicates that the mutant can survive better when the cells are cultured under the lethal temperature.

Figure 26: Growth of both the control strain (WBG2) and the mutant strain (WBG101) under various temperatures. W and M represents WBG2 and WBG101, respectively.

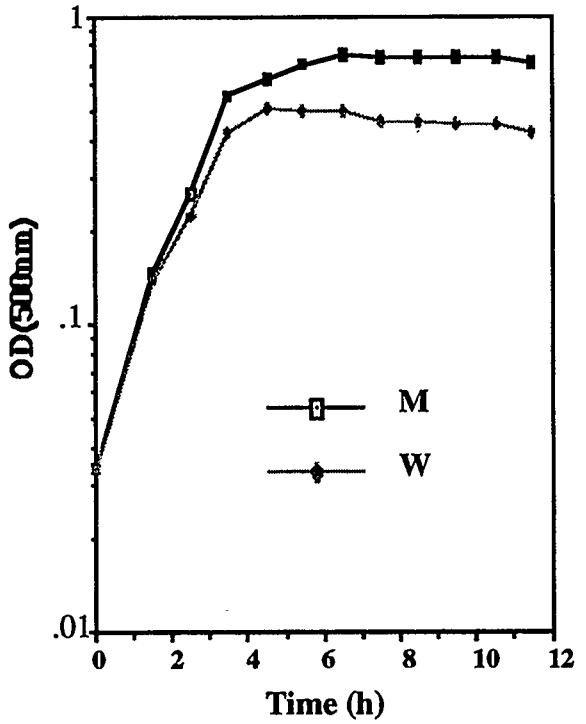
(a). Growth at 37°C



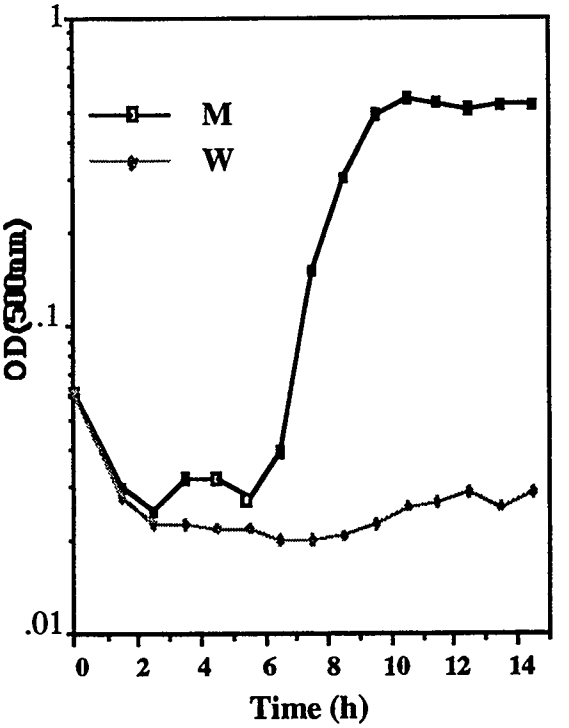
(b). Growth at 48°C



(c). Growth at 48°C(30min) then at 52°C



(d). Growth at 52°C



VI. 9 Discussion

The preliminary screening was based on the color of colony. WBG3, was used as a positive control. WBG3 has the same genetic background as WBG2 except that CIRCE in the *groE-bgaB* fusion was deleted. The putative repressor cannot shut off the expression of the *groE-bgaB* fusion. Therefore, *bgaB* is constitutively expressed at a high level at 37°C, and the colony of WBG3 is dark blue in color. If the mutants generated from WBG2 carried the mutations in the repressor gene, inactivation of the repressor should allow the expression of *groE-bgaB* to the same level as that in WBG3. By selecting the colony with a similar color intensity that observed in WBG3, it is possible to identify the potential *groE* regulatory mutants. Since *groE* and *dnaK* were controlled by the same regulatory element (CIRCE), mutants carrying the defective repressor gene should overproduce DnaK and GroE simultaneously at 37°C. This type of mutants should show a high level expression of *bgaB* as well. In addition, genomic sequencing of regulatory regions in *groE-bgaB* or *groE* and *dnaK* should be carried out to eliminate the possible mutations existed in CIRCE and the promoter of all three operons.

During the process of screening, many light blue colonies were observed. None of them showed a simultaneous overproduction of all three proteins (GroE, DnaK and BgaB). Several reasons may explain this observation :1) Mutations in either the *B. subtilis* endogenous ß-galactosidase gene which expressed at a low level or its regulatory gene encoding a repressor (Errington and Vogt, 1990); 2) Partial inactivation of the repressor of *groE* and *dnaK*; 3) Mutations in the CIRCE sequence of the *groE-bgaB* fusion. In these cases, the

mutations may weakly reduce the negative function of CIRCE; 4) Improper folding of proteins caused by chemical mutagenesis, which lead to an elevated expression of *groE* and *dnaK* and *groE-bgaB*.

NTG (N-methy-N'-nitro-N-nitrosoguanidine) was used to generate the *groE* regulatory mutant in this study. Seven mutants were isolated out of about 5×10^5 cells. each of the seven mutants carries a single mutation in *orf39*. The size of *orf39* is 1.029 kb, and the size of *B. subtilis* genome is about 4,000 kb. If each cell in 500,000 cells treated with NTG carried a single mutation, the expected number of mutants that have mutations in *orf39* is about 125. Since not every mutation in *orf39* can generate defective ORF39, only those mutants carrying mutations affecting the critical amino acid residues in ORF39 will show dark blue colony. We have aligned the sequence of ORF39 homologs. Two regions are highly conserved and these regions covered about 100 bp. Therefore, the number of mutants show dark blue colony should be around 12. Because, NTG prefers to generate mutations at G or C and the GC content of *B. subtilis* is about 40%, the expected number of mutants should be less than 10. Thus, seven mutants, the number of the *orf39* mutants obtained is in close agreement with the theoratical calculation. The dark blue colony may also come from the mutations in the CIRCE sequence in *groE-bgaB*. There are 27-bp nucleotides in CIRCE and only 18 nucleotides involved in the formation of the stem structure are importmant. Therefore, less one mutant carrying mutations in CIRCE is generated out of 500,000 cells treated with NTG. Of the seven mutants, three mutants have the same mutation (606C to T) and three mutants have another identical mutation (1161G to A). The possible reasons to generate the same mutation in these mutants are as follows: 1) These mutants are from a single parental mutant which propagated

twice during NTG mutagensis; 2) This mutated nucleotide is critical to the function of ORF39.

Isolation of these mutants (WBG101 to WBG107) indicates that a common negative regulatory factor controls the expression of both *groE* and *dnaK* in *B. subtilis*. The fact that *orf39* on a multiple copy plasmid can restore all seven mutants to normal and the existance of mutations in *orf39* in these mutants suggest that ORF39 should be the negative regulatory factor controlling the expression of *groE* and *dnaK*. Our results from the binary plasmid system in *E. coli* and the gel mobility shift assay also provide strong evidence to support the idea that ORF39 acts directly on the *groE* promoter by binding specifically to the CIRCE sequence.

Since the conserved CIRCE sequences can be found in many bacteria, it is interesting to know whether *orf39* homologs exists in these organisms. By search the known sequences in GenBank, four microorganisms including *Clostridium acetobutylicum*, *Staphylococcus aureus*, *Lactococcus lactis* and *Chlamydia trachomatis* with CIRCE sequences in their *dnaK* operons were found to carry a homologous *orf39* gene. Alignment of the amino acid sequences of these ORF39 homologs shows two highly conserved regions (Figure 23). These conserved regions are located at positions 9-77 (the N-terminal region) and positions 307-322 (the C-terminal region), respectively. The change of the highly conserved Gly-307 to Asp in the group 1 mutants and Gly-63 to Glu in the group 3 mutants resulted in the inactivation of ORF39. It suggests that these residues may play a critical role in the repressor function.

Inactivation of *orf39* results in the constitutive expression of *groE* and *dnaK* operon. Since these genes code for molecular chaperones which can help the cells survive the stress conditions. Our results show that WBG101 with

mutations in *orf39* can survive at 52°C, which is a lethal culture condition to the wild-type strain. Overproduction of these molecular chaperons indeed can help the cells survive better under harsh conditions. Over 60 proteins can be induced under the heat shock condition in *B. subtilis*. Therefore, besides GroEL, GroES, DnaK, DnaJ and GrpE, other heat shock proteins are also required to help the cells survive under the harsh conditions.

VII Concluding Remarks and Perspectives

VII.1 Regulation of the *groE* and *dnaK* expression in *B. subtilis*

In this thesis, the regulatory mechanism of *groE* was investigated extensively. On the basis of our results, a working model for the regulation of *groE* and *dnaK* is proposed as follows: The σ^A containing RNA polymerase transcribes *groE* and *dnaK* operons under both heat shock and non-heat shock conditions. Under the non-heat shock condition, *groE* and *dnaK* are expressed at low levels due to two mechanisms mediated by CIRCE. 1) The transcription of *groE* and *dnaK* was repressed through ORF39 by binding to CIRCE (the IR sequence): 2) The presence of CIRCE at the 5' end of the transcript can increase the degradation rate of the *groE* mRNA. The autoregulation of *orf39* allows cells to maintain the basal level of GroEL, GroES, GrpE, DnaK and DnaJ. These proteins are required to carry out the normal physiological functions at this temperature. *groE* is shown to be essential for cell growth in both *B. subtilis* and *E. coli*. Following the temperature upshift, ORF39 is inactivated by either heat denaturation, proteolytic degradation or chemical modifications, *groE* and *dnaK* can then be transcribed by the σ^A containing RNA polymerase. The increased transcription results in elevating the level of GroEL, GroES, GrpE, DnaK and DnaJ. Higher levels of these molecular chaperones can protect the nascent polypeptides from improper folding, allow refolding of denatured proteins and help degrading certain denatured proteins so that the cells can survive better under the heat shock condition.

Since CIRCE-like sequences can be found in *groE* and *dnaK* operons in 28 different bacteria, the regulatory mechanism deduced here can potentially be applied to these organisms.

VII.2 Perspectives

VII.2.1 Possible mechanisms for the inactivation of ORF39. As described in the model, ORF39 will lose its repressive function under the heat shock condition. However, whether it is inactivated by either heat denaturation, chemical modification (such as methylation) or proteolytic degradation is still unclear. To address this question, the purified ORF39 can be applied to study the ORF39-CIRCE interaction under non-heat shock (37°C) and heat shock conditions (48°C). If the repressor can be inactivated by heat denaturation, no retardation of the CIRCE containing DNA fragment would be expected. Otherwise, other mechanisms may involve in the inactivation of ORF39. Furthermore, *in vitro* transcription can also be performed by using the σ^A containing RNA polymerase in the presence or absence of ORF39, if the repressor were inactivated by heat denaturation, more transcripts would be observed under the heat shock condition.

VII.2.2 Regulation of *orf39*. It is particularly interesting to study how the expression of *orf39* is regulated. Since *orf39* is the first gene in the *dnaK* operon, expression of *orf39* is subject to autoregulation. Thus, the expression of *orf39* should also increase under the heat shock condition. In such a case, it is hard to explain why the cells would like to make more ORF39 which is useless under the heat shock condition. Two possible explanations can be provided: 1) ORF39 may have a dual functional roles. At low temperatures, it serves as a repressor. At high temperatures, it may have other physiological roles (e.g. molecular chaperone); 2) Some other mechanisms may control the expression of *orf39*. The ORF39 level may not actually increase. To address these questions, it is important to determine the expression of *orf39* at both the transcriptional and translational levels under different conditions. At present,

limited results were obtained. They tend to support the second assumption. 1) The transcript of *orf39* does not increase as high as the transcripts of other genes (*dnaK*, *dnaJ* and *grpE*) in the same operon under heat shock condition (Zuber and Schumann, 1995). 2) Under the heat shock condition, the level of ORF39 is not increased as analyzed by SDS-PAGE. Further confirmation is needed by Western blot.

Several studies can be performed to investigate the regulation mechanism for *orf39* expression. 1) Stability of the *orf39* mRNA may regulate the expression of *orf39*. 2) It is important to determine whether the transcript of the *dnaK* operon is processed. This mechanism was observed in the regulation of the *groE* expression in *A. tumefaciens*, which can reduce the expression of *groES* by generate an unstable *groES* transcript. 3) ORF39 may be very unstable due to proteolytic degradation. The half life of ORF39 should be determined in different protease deficient strains to investigate which proteolytic pathway may determine the stability of ORF39. 4) As in the case of sigma 32, molecular chaperones may mediate the stability of ORF39. This can be confirmed by studying the half life of ORF39 in different null mutants carrying defective genes for *dnaK* and *groE*.

REFERENCES

Asseldonk, V., A. Sinmons, H. Visser, W. M. De Vos, and G. Simons. 1993. Cloning, nucleotide sequence, and regulatory analysis of the *Lactococcus lactis dnaJ* gene. *J. Bacteriol.* 175:1637-1644.

Bloom, M., S. Skelly, R. VanBogelen, F. Neidhardt, N. Brot, H. Weissbach. 1986. *In vitro* effect of the *Escherichia coli* heat shock regulatory protein on expression of heat shock genes. *J. Bacteriol.* 166:380-84.

Bochkareva, E. S., N. M. Lissin, and A. S. Girshovich. 1988. Transient association of newly synthesized unfolded proteins with the heat-shock GroEL protein. Nature (London) 336:254-257.

Brian, S. M., T.E. Kennedy, and U.N. Streips. 1991. Molecular characterization of specific heat shock proteins in *Bacillus subtilis*. Current Microbiology 22:231-236

Briat, J. F., M. J. Chamberlin. 1985. *Bacillus subtilis* σ^{28} and *Escherichia coli* σ^{32} (hptR) are minor σ factors that display an overlapping promoter specificity. *J. Biol. Chem.* 260:2038-2041.

Chandrasekhar, G. N., K. Tilly, C. Woolford, R. Hendrix, and C. Georgopoulos. 1986. Purification and properties of the GroES morphogenetic protein of *Escherichia coli. J. Biol. Chem.* 261:12414-12419.

Chang, B.-Y., K.-Y. Chen, Y.-D. Wen, and C.-T. Liao. 1994. The response of a *Bacillus subtilis* temperature-sensitive *sigA* mutant to heat stress. *J. Bacteriol.* 176:3102-3110.

Cowing, D. W., J. C. A. Bardwell, E. A. Craig, C. Woolford, R. W. Hendrix, C. A Gross. 1985. Consensus sequence for *Escherichia coli* heat shock gene promoters. *Proc. Natl. Acad. Sci.* USA. 82:2679-83.

Cutting, S.M., and P.B. Vander Horn. 1990. Genetic Analysis, p. 62-63. In C.R. Harwood and S. M. Cutting (eds.). *Molecular Biological Methods for Bacillus.* John Wiley and Sons, Chichester, U. K.

Dubnau, D., and R. Davidoff-Abelson. 1971. Fate of transformation DNA following uptake by competent *Bacillus subtilis. J. Bacteriol.* 88: 904-911.

Duchene, A.-M., C. J. Thompson, and P. Mazodier. 1994. Transcriptional analysis of *groEL* genes in *Streptomyces coelicolor* A3 (2). *Mol. Gen. Genet.* 245:61-68.

Ellis, R. J., and S. M. van der Vies. 1991. Molecular chaperones. *Annu. Rev. Biochem.* 60:321-347.

Erickson, J. W., V. Vaughn, W. A. Walter, F. C. Neidhart, C. A. Gross. 1987. Regulation of the promoters and transcripts of *rpoH*, the *Escherichia coli* heat shock regulatory gene. *Genes Dev.* 1:419-432.

Errington, J. and C. Vogt. 1990. Isolation and characterization of mutations in the gene encoding an endogenous *Bacillus subtilis* ß-galactosidase and its repressor. *J. Bacteriol.* 172: 488-490.

Fayet, O., T. Ziegelhoffer, and C. Georgopoulos. 1989. The *groES* and *groEL* heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J. Bacteriol.* 171:1379-1385.

Fredrick, K., T. Caramori, Y.-F. Chen, A. Galizzi, and J. D. Helmann. 1995. Promoter architecture in the flagellar regulon of *Bacillus subtilis*: high-level expression of flagellin by the σ^D RNA polymerase requires an upstream promoter element. *Proc. Natl. Acad. Sci. USA* 92: 2582-2586.

Georgopoulos C. P., R. W. Hendrix, S.R. Casjens, and A. D. Kaiser. 1973. Host participation in bacteriophage lambda head assembly. *J. Mol. Biol.* 76:45-60.

Gilman, M., J. Wiggs, and M. J. Chamberlin. 1981. Nucleotide sequences of two *Bacillus subtilis* promoters used by *Bacillus subtilis* sigma-28 RNA polymerase. *Nucleic Acids Res.* 9:5991-6000.

Grossman, A. D., D. B. Straus, W. A. Walter, C. A. Gross. 1987. σ^{32} synthesiscan regulate the synthesis of heat shock protiens in *Escherichia coli. Genes Dev.* 1:179-184.

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

Guglielmi, G., A.-M. Duchene, C. Thompson, and P. Madzodier. 1993. Transcriptional analysis of two different *Streptomyces albus groEL*-like genes, p. 17-24. In R. H. Baltz, G. D. Hegeman, and P. L. Skatrud (ed.), *Industrial microogranisms: Basic and applied molecular genetics*. American Society for Microbiology, Washington, D.C.

Haldenwang W. G. 1995. The sigma factors in *Bacillus subtilis*. *Microbiological Reviews*. 59:1-30.

Helmann, J. D., L. M. Marquez, and M. J. Chamberlin. 1988. Cloning, sequencing and disruption of the *Bacillus subtilis* σ^{28} gene. *J. Bacteriol.* 170:1568-1574.

Helmann, J.D. 1995. Compilation and analysis of *Bacillus subtilis* σ^{A}-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. *Nucleic Acids Res.* 23: 2351-2360.

Helmann, J.D.and M.J. Chamberlin. 1988. Structure and function of bacterial sigma factors. *Ann. Rev. Biochem.* 57: 839-872

Hemmingaen, S. M., C. Wollford, S. M. van der Vies, K. Tilly, D. T. Dennis, C. P. Georgopoulos, R. W. Hendrix, and R. J. Ellis. 1988. Homologous plant and baterial proteins chaperone oligomeric protein assembly. *Nature* (London) 333:330-334.

Hemsley, A., N. Arnheim, M. D. Toney, G. Cortopassi, and D. J. Galas. 1989. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res.* 17:6545-6551.

Hendrix, R. W. 1979. Purification and properties of GroE, a host protein involved in bacteriophage assembly. *J. Mol. Biol.* 129: 375-392.

Herman, C., D. Thevenet, R. D' Ari, and P. Bouloc. 1995. Degradation of σ^{32}, the heat shock regulator in *Escherichia coli*, is governed by HflB. *Proc. Natl. Acad. Sci. USA.* 92: 3516-3520.

Hirata, H., S. Negoro, H. Okada. 1984. Molecular Basis of Isozyme Formation of ß-galactosidase Genes, *bgaA* and *bgaB*. *J. Bacteriol.* 166:9-14.

Hirata, H., T. Fukazawa, S. Negoro, and H. Okada. 1986. Structure of a ß-galactosidase gene of *Bacillus staerothermophilus*. *J. Bacteriol.* 166:722-727.

Hiratsu, K., M. Amemura, H. Nashimoto, H. Shinagawa, and K. Makino. 1995. The *rpoE* gene of *Escherichia coli*, which encodes σ^{E}, is essential for bacterial growth at high tmperature. *J. Bacterial.* 177: 2918-2922.

Hohn, T., B. Hohn, A. Engel, M. Wurtz, and P. R. Smith. 1979. Isolation and characterizatiion of the host protein GroE involved in bacteriophage lambda assembly. *J. Mol. Biol.* 129:359-373.

Horwich, A.L., K. Brooks Low, W.A. Fenton, I.N. Hirshfield, and K. Furtak. 1993. Folding *in vivo* of bacterial cytoplasmic proteins: Role of GroEL. Cell. 74: 909-917

Jobling, M. G. and R. K. Holmes. 1990. Construction of vector with the p15a replicon, kanamycin resitance, inducible lacZa and pUC18 or pUB19 multiple cloning sites. *Nucleic Acids Res.* 18: 5315-5316.

Kenney, T.J. and C. P. Moran, JR. 1991. Genetic evidence for interaction of σ^A with two promoters in *B. subtilis*. *J. Bacteriol.* 173:3282-3290.

Kraus, A., C. Hueck, D. Gartner, and W. Hillen. 1994. Catabolite repression of the *Bacillus subtilis xyl* operon involves a cis element functional in the context of an unrelated sequence, and glucose exerts additional *xylR*-dependent repression. *J. Bacteriol.* 176: 1738-1745.

Kruger, E., U. Volker and M. Hecker. 1994. Stress induction of clpC in *Bacillus subtilis* and its involvement in stress tolerance. *J. Bacteriol.* 176:3360-3367.

Kumamoto, C. A. 1991. Molecular chaperones and protein translocation across the *Escherichia coli* inner membrane. *Mol. Microbiol* 5:19-22.

Kusukawa, N., T. Yura, C. Ueguchi, Y. Akijama, and K. Ito. 1 989. Effects of mutations in heat-shock genes *groES* and *groEL* on protein export in *Escherichia coli. EMBO J.* 8:3517-3521.

Landick, R., V. Vaughn, E. T. Lau, R. A. VanBogelen, J. W. Erickson, F. C. Neidhardt. 1984. Nucleotide sequence of the heat shock regulatory gene of *E. coli* suggests its protein product may be a transcription factor. *Cell* 38:175-82.

Lecker, S., R. Lill, T. Ziegelhoffer, C. Georgopoulos, P. J. Bassford, Jr., C. A. Kumamoto, and W. Wickner. 1989. Three pure chaperone proteins of *Escherichia coli* SecB, trigger factor and GroEL form soluble complexes with precursor proteins *in vitro*. *EMBO J.* 8:2703-2709.

Li, M. and S.-L. Wong. 1992. Cloning and characterization of the *groESL* operon from *Bacillus subtilis*. *J. Bacteriol.* 174: 3981-3992.

Maniantis, T., E. F. Fritach, and J. Sambrook. 1989 *Molecular Cloning: a laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.

Miller, B. S., E. Ferrsri, U. N. Steips, and W. D. Crabb. 1993. Characterization of a mutant altered in the induction of the GroEL protein upon heat shock in *Bacillus subtilis*, abstr. P44. In *Abstracts of Seventh International Conference on Bacillus* 1993. Paris, France.

Miller, B. S., T. E. Kennedy, and U.N. Streips. 1991. Molecular characterization of specific heat shock proteins in *Bacillus subtilis*. *Curr. Microbiol.* 22:231-236.

Morita, T., K. Shigesada, F. Kimizuka, and H. Aiba. 1988. Regulatory effect of a synthetic CRP recognition sequence placed downstream of a promoter. *Nucleic Acids Res.* 16: 7315-7332.

Nagai, H., H. uzawa, T. Yura. 1991. Interplay of two cis-acting mRNA regions in translational control of σ^{32} synthesis during the heat shock response of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 88:10515-19.

Nagai, H., R. Yano, J. W. Erickson, T. Yura. 1990. Transcriptional regulation of the heat shock regulatory gene *rpoH* in *Escherichia coli*: involvement of a novel catabolite-sensitive promoter. *J. Bacteriol.* 172:2710-15.

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

Phillips, G. J., and T. J. Silhavy. 1990. Heat-shock proteins DnaK and GroEL facilitate export of LacZ hybrid proteins in *E. coli. Nature* (London) 344:882-884.

Raina, S., D. missiakas, and C. Georgopoulos. 1995. The *rpoE* gene encoding the σ^{E} (σ^{24}) heat shock sigma factor of *Escherichia coli. EMBO J.* 14: 1043-1055.

Riethdorf, S., U. Volker, U. Gerth, A. Winkler, S. Engemann, and M. Hecker. 1994. Cloning, nucleotide sequence, and expression of the *Bacillus subtilis lon* gene. *J. Bacteriol.* 176:6518-6527.

Ross, W., K. K. Gosink, J. Salomon, K. Igarashi, C. Zou, A. Ishiama, K. Severinov, and R. L. Gourse. 1993. A third recognition element in bacterial promoters: DNA binding by the a subunit of RNA polymerase. *Science* 262: 1407-1413.

Rouviere, P.E., A. De Las Penas, J. Mecsas, C. Z. Lu, K. E. Rudd, and C. A. Gross. 1995. *rpoE,* the gene encoding the second heat-shock sigma factor, σ^E, in *Escherichia coli. EMBO J.* 14: 1032-1042.

Schmidt, A., M. Schiesswohl, U. Volker, M. Hecker, and W. Schumann. 1992. Cloning, sequenceing, mapping, and transcriptional analysis of the *groESL* operon from *Bacillus subtilis. J. Bacteriol.* 174: 3993-3999.

Schultz, A., B. Tzschaschel, and W. Schumann. 1995. Isolation and analysis of mutants of the *dnaK* operon of *Bacillus subtilis. Mol. Micribiol.* 15:421-429.

Segal, G. and E. Z. Ron. 1993. Heat shock transcription of the *groESL* operon of *Agrobacterium tumefaciens* may involve a hairpin-loop structure. *J. Bacteriol.* 175:3083-3088.

Segal, G. and E. Z. Ron. 1995. The *groESL* operon of *Agrobacterium tumefaciens* : Evidence for heat shock-dependent mRNA cleavage. *J. Bacteriol.* 177: 750-757.

Skelly, S., T. Coleman, C.-F. Fu, N. Brot, H. Weissbach. 1987. Coregulation between the 32-kDa σ factor levels and *in vitro* expression of *Escherichia coli* heat shock genes. *Proc. Natl. Acad. Sci. USA* 84:8365-69.

Spizizen, J. 1958. Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxribonucleate. *Proc. Natl Acad. Sci. USA* 44:407-8.

Sternberg, N. 1973. Properties of a mutant of *Escherichia coli* defective in bacteriophage λ head formation (*groE*). *J. Mol. Biol.* 76:25-44.

Straus, D. B., W. A. Walter, C. A. Gross. 1989. *Escherichia coli* heat shock gene mutant are defective in proteolysis. *Genes Dev.* 2:1851-58.

Straus, D. B., W. A.Walter, C. A. Gross. 1987. The heat shock response of *E. coli* is regulated by changes in the concentration of σ^{32}. *Nature* 329:348-51.

Tamamori, T., K. Ito, Y. Nakamura, T, Yura. 1978. Transient regulation of protein synthesis in *Escherichia coli* upon shift-up of growth temperature. *J. Bacteriol.* 134:1133-40.

Tilly, K., H. Murialdo, and C. Geogopoulos. 1981. Identification of a second *Escherichia coli groE* gene whose product is necessary for bacteriophage morphogenesis. *Proc. Natl. Acad. Sci. USA* 78:1629-1633.

Tilly, K., J. Erickson, S. Sharma, C. Georgopoulos. 1986. Heat shock regulatory gene *rpoH* mRNA level increase after heat shock in *Escherichia coli*. *J. Bacteriol.* 168:1155-58.

Tilly, K., N. McKittrick, M. Zylicz, C. Georgopoulos. 1983. The *dnaK* protein modulates the heat shock response of *Escherichia coli* heat shock regulatory factor σ^{32}. *J. Bacteriol.* 171:1585-89.

Tomoyasu, T., J. Gamer, B. Bukau, M. Kanemori, H. Mori, A. J. Rutman, A.B. Oppenheim, T. Yura, K. Yamanaka, H. Niki, S. Hiraga and T. Ogura. 1995. *Escherichia coli* FtsH is a membrane-bound, ATP-dependent protease which degrades the heat shock transcription factor σ^{32}. *EMBO J.* 14: 2551-2560.

Volker, U., H. Mach, R. Schmid and M. Hecker. 1992. Stress proteins and cross-protection by heat shock and salt stress in *Bacillus subtilis*. *J. Gen. Microbiol.* 138: 2125-2135.

Volker. U, S. Engelmann, B. Maul, S. Riethdorf, A. Volker, R. Schmid, H. Mach and M. Hecker. 1994. Analysis of the induction of general stress proteins of *Bacillus subtilis*. *Microbiology.* 140:741-752.

Wang, L.-F. and R. H. Doi. 1987. Promoter switch during development and the termination site of the σ^{43} operon of *Bacillus subtilis*. *Mol. Gen Genet.* 207:114-119.

Wang, P.-Z.. and R. H. Doi. 1984. Overlapping promoters transcribed by *Bacillus subtilis* σ^{55} and σ^{37} RNA polymerase holoenzymes during growth and stationary phases. *J. Biol. Chem.* 259: 8619-8625.

Wetzstein, M., U. Volker, J. Dedio, S. Lobau, U. Zuber, M. Schiesswohl, C. Herget, M. Hecker, and W. Schumann. 1992. Cloning, sequencing, and molecular analysis of the *dnaK* locus from *Bacillus subtilis. J. Bacteriol.* 174: 3300-3310.

Wetzstein. M. and W. Schumann. 1990. Promoters of major *Escherichia coli* heat shock genes seem non-functional in *Bacillus subtilis. FEMS Microbiol Lett.* 60:55-58.

Wilhelm, M. and C. P. Hollenberg. 1985. Nucleotide sequence of the *Bacillus subtilis* xylose isomerase gene: extensive homology between the *Bacillus* and *E. coli* enzyme. *Nucleic Acids Res.* 13:5717-5722.

Wu, X.-C., S. Nathoo, A. S.-H. Pang, T. Carne, and S.-L. Wong. 1990. Cloning, genetic organization of a structural gene encoding Bacillopeptidase F from *Bacillus subtilis. J. Biol. Chem.* 265:6845-6850.

Yamamori, T., T. Yura. 1980 Temperature-induced synthesis of specific proteins in *Escherichia coli*: evidence for transcriptional control. *J. Bacteriol.* 142:843-51.

Yansura, D. G. and D. J. Henner. 1984. Use of the *Escherichia coli lac* repressor and operator to control gene expression in *Bacillus subtilis*. *Proc. Natl. Acad. Sci.* USA. 81:439-443.

Ye, R. Q. and S.- L. Wong. 1995. unpublished result

Zuber, U. and W. Schumann. 1994. CIRCE, a novel heat shock element involved in regulation of heat shock operon *dnaK* of *B. subtilis*. *J. Bacteriol.* 176:1359-1363.