LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

A STUDY OF THE EXPRESSION AND ROLES OF *BACILLUS SUBTILIS* HtrA-TYPE SERINE PROTEASES

David Noone

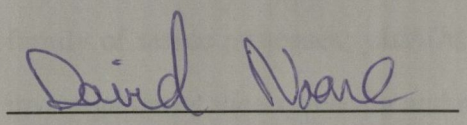
Thesis Submitted to the University of Dublin, Trinity College, for the Degree of Doctor of Philosophy

Department of Genetics
University of Dublin,
Trinity College.

December 2002.

DECLARATION

This thesis has not been previously submitted to this or any other university for examination for a higher degree. The work presented here is entirely my own except where noted. This thesis may be made available for consultation within the university library. It may be photocopied or lent to other libraries for the purposes of consultation.



David Noone
December 2002.

SUMMARY

Upon sequencing a region of the *Bacillus subtilis* genome assigned to the author of this thesis, an open reading frame encoding a homologue of an HtrA-type serine protease, YkdA, was identified. The HtrA serine protease family are common to nearly all eubacteria which can possess one or more paralogues. The HtrA proteases generally respond to stress conditions, typically heat or oxidative insult, and in many pathogens contribute to the virulence of the bacterium. *B. subtilis* contains three genes encoding members of the HtrA family of serine proteases; *ykdA*(*htrA*), *yvtA* and *yycK*(*yyxA*). This work describes an investigation of the expression and roles of the *Bacillus subtilis* HtrA-type serine proteases.

Sequence and topological analysis predict that all three HtrA-type proteases are anchored to the cell membrane by a single transmembrane domain and that the catalytic domains are exposed on the outside of the cell. Additionally, the control regions of both *ykdA* and *yvtA* contain a highly homologous (91% identical) 56-bp sequence not found upstream of *yycK*. This suggested that *ykdA* and *yvtA* might be controlled by a similar mechanism. Using a combination of transcriptional fusions and direct transcriptional analysis employing Northern blotting and primer extension analysis, it was shown that both *ykdA* and *yvtA* were heat shock inducible. In contrast no heat shock induction of *yycK* was observed. Heat shock induction of *ykdA* was not affected by mutations in any of the three genes encoding the well characterised regulators of Class I (HrcA), Class II (σ^B) or Class III (CtsR) regulons. This implied that *ykdA*, and presumably *yvtA*, were members of a novel heat shock regulon. A deletion analysis of the *ykdA* control region identified a 60bp sequence necessary for normal heat shock induction. This fragment contains four copies of an octomer repeat motif upstream of a well conserved SigA-type -10 box appropriately positioned relative to the transcriptional start site. Given their proposed cellular localization, both YkdA and YvtA were shown to be greatly upregulated upon overexpression of heterologous secretable amylases, in comparison to the level of induction observed during heat shock. It has been shown previously that production of large amounts of secretable proteins causes saturation of the secretion apparatus and leads to extracytoplasmic misfolding. The induction observed occurred at the transcriptional level and from the same promoters, as determined by primer

extension analysis, that effected heat shock induction. No increase in the expression of *yycK* was observed in strains overproducing amylase.

A novel aspect of the regulation of *ykdA* and *yvtA* was the negative and reciprocal autoregulation between the genes. In either a *ykdA* or *yvtA* mutant strain the basal temporal expression of either gene was clearly increased. Furthermore upon heat stress these increased levels of expression were dramatically amplified compared to the wild type situation under identical stress conditions. The importance of YkdA or YvtA to the physiology of *B. subtilis* is implied from the phenotypes of strains mutated in both genes. These strains display a dramatic growth phenotype producing small round mucoid colonies rapidly accumulating one or more suppressor mutations which permit normal (wild type) growth and colony formation at 37˚C. However the suppressor strains have an inability to grow at elevated temperatures where rapid lysis is evident. This temperature sensitive phenotype is common to many bacteria bearing mutations in their *htrA* genes. One further surprising and unexpected result was the finding that strains mutated for *ykdA* or *yvtA* or expressing an elevated level of a protease negative variant of YkdA[N289H S290M] displayed increased oxidative stress resistance to lethal levels of hydrogen peroxide. This may indicate a non-proteolytic , and by inference to the activities of DegP (a HtrA-type protease) from *E. coli*, chaperone role for YkdA.

In conclusion, the work presented in this thesis would indicate that both YkdA and YvtA perform a stress responsive role to degrade or chaperone misfolded proteins found on the extracytoplasmic surface of the *B. subtilis* cell envelope. This conclusion is supported by the non activation of either gene in response to the production of truncated cytoplasmic proteins resulting from puromycin exposure.

DEDICATION

To Regina, Rowan and Dáire.

ACKNOWLEDGEMENTS

I would firstly like to express my thanks to my supervisor Prof. Kevin Devine for his enthusiastic support and diligence during the course of the work leading to this thesis. I also wish to acknowledge all the people in the KD lab whose advice, support and friendship have always been welcome and appreciated. Special thanks to Libby for well reasoned questions and criticisms, to Kasper likewise and for his expertise and advice with RNA experiments. To the many students who have been laboured with various HtrA spin-off projects I am also grateful for your contributions. This applies especially to Alistair and to Ross for constructing important strains. I also wish to thank the newer crew of the KD lab, Brian, Lorna and Paola. To Mary also, I owe much thanks for advice, friendship and chats.

I wish also to say a big thank you to all members of the department but especially to the prep room guys, Dave, Paul, Louis and Brenda for their assistance whenever needed. Not all days in the lab went smoothly and to anyone who had to endure or listen to my whinging your time and patience is acknowledged. This especially applies to colleagues past and present particularly those in the TK lab, thanks Alan, Sinead, Jannet and Nga.

I wish also to thank the National Pharmaceutical Biotechnology Centre, BioResearch Ireland and especially Prof. Clive Williams for financial support and for making it possible to allow me undertake this thesis project. I also wish to thank David McConnell for the two opportunities to work in the Genetics Department.

Lastly and most importantly I wish to thank my family for their wonderful support and patience! To Regina my wife I guess if you're reading this you know its done at last - I couldn't have made it without you. Thanks also to the Lads, Rowan and Dáire, for your many visits and cheerfulness. Finally to my mother and father, Frances and Michael and my aunt Joan your encouragement was truly appreciated.

TABLE OF CONTENTS

	Page No.
Declaration	I
Summary	II
Dedication	IV
Acknowledgements	V
Table of Contents	VI
List of Tables	XI
List of Figures	XII
Abbreviations	XV
Publications	XVII

Page No.

CHAPTER 1. GENERAL INTRODUCTION.

1.1. *B. subtilis* as a model organism	1
1.1.1. Natural Competence of *B. subtilis*	1
1.1.2. Exoenzyme production by *B. subtilis*	2
1.1.3. Developmental programme of Sporulation	2
1.1.4. Genome Sequencing and Functional Analysis of *B. subtilis*	3
1.2. The Heat Shock Response	4
1.2.1. The *E. coli* Cytoplasmic Heat Shock response	4
1.2.2. DnaK/DnaJ chaperone system	5
1.2.3. GroEL/GroES chaperone system	6
1.2.4. Heat Shock Proteases	6
1.3. Regulation of the Cytosolic Heat Shock Response in *E. coli*.	8
1.3.1. Increased synthesis of σ^{32}	8
1.3.2. Increased Stability and Activity of σ^{32}	9
1.4. The Extracytoplasmic Heat Shock and Unfolded Protein Response in *E. coli*	10
1.4.1. Chaperones and protein folding agents	11
1.4.2. Proteases	11
1.4.2.1. OmpT protease	11
1.4.2.2. Tsp (Prc)	12
1.5. HtrA proteases	13
1.5.1. Structure of HtrA proteases	13
1.5.2. N-termini	13
1.5.2.1. *E. coli* DegP	13
1.5.2.2. Other Bacterial HtrA proteases	14

1.5.2.3 Eukaryotic HtrA proteases 15
1.5.3. Catalytic domains of HtrA proteases 16
1.5.4. HtrA proteases contain PDZ substrate recognition domains at their C-termini 17
1.5.4.1. PDZ domains 17
1.5.4.2 Bacterial PDZ domains 17
1.5.4.3. PDZ domain of DegP 18
1.5.4.4. Human HtrA PDZ domains 19
1.5.5. Multimeric Structure of DegP and HtrA2(Omi) 19
1.5.6. Functions of HtrA proteases 20
1.5.6.1. HtrA proteases respond to varied stress stimuli 20
1.5.6.2. The role of DegP paralogues 21
1.5.6.3. Pathogenicity Functions of other Bacterial HtrA proteases 22
1.5.6.4. Functions of Eukaryotic HtrA proteases 23
1.5.6.5. Biotechnological and Biomedical Applications of HtrA proteases 24
1.5.7. Regulation of the expression of HtrA proteases 25
1.5.7.1. Regulation of DegP by the σ^E alternate ECF-type sigma factor 25
1.5.7.2. Stress induction of σ^E 26
1.5.7.3. Signal transduction mechanism of the σ^E stress response 27
1.5.7.4. σ^E regulon members and inducing signals 30
1.5.8. Identification of the Cpx system in controlling an envelope stress response 32
1.5.8.1. The CpxRA two-component system 34
1.5.8.2. Regulation of the Cpx response 36
1.5.8.3. Cpx regulon members and inducing signals 37
1.5.8.4. A common mechanism for signal transduction to activate the Cpx and σ^E responses ? 38
1.5.9. Regulation of HtrA proteases in other bacteria 40
1.6. The Heat Shock Response of *Bacillus subtilis* 42
1.6.1. Class I heat shock genes 42
1.6.2. Class II heat shock genes 45
1.6.3. Class III heat shock genes 47
1.7. Extracytoplasmic stress response and HtrA proteases in *B. subtilis* 49

CHAPTER 2. MATERIALS AND METHODS.

2.1. Growth and Maintenance of Bacterial Strains. 52
2.1.1. Antibiotic Concentrations. 52
2.1.2. Composition of Growth Media. 52
2.2. Stress Induction Protocols 53

2.2.1. Stress Inductions 53
2.2.2. Phenotypic Analyses 53
2.2.2.1 Thermosensitivity 53
2.2.2.2. Oxidative Stress Resistance 53
2.3. DNA Manipulations 54
2.3.1. Cloning Procedures 54
2.3.2. Polymerase Chain Reaction 54
2.3.3. Plasmid DNA Isolation 54
2.3.4. DNA Fragment Isolation 54
2.3.5. *B. subtilis* Chromosomal DNA Isolation 55
2.4. Transformation Procedures 55
2.4.1. *E. coli* Transformation 55
2.5.2. *B. subtilis* Transformation 55
2.5. Sequencing and Sequence Assembly of *B. subtilis* Genomic Fragments. 56
2.5.1. Generation of "mini-banks" of *B. subtilis* Clones in *E. coli* 56
2.5.2. Sequencing Reactions 56
2.5.3. Sequence Analysis 57
2.5.4. Sequence Assembly 57
2.6. Plasmid and Strain Constructions 57
2.7. Transcriptional Analyses 65
2.7.1. RNA Isolation from *B. subtilis* 66
2.7.2. Northern Blot Analysis 66
2.7.3. Primer Extension Analysis 68
2.7.3.1. Radiolabelling Oligonucleotide Probes 68
2.7.3.2. Primer Extension Reactions 68
2.8. Measurement of β-Galactosidase Activity 68
2.9. Routine and General Sequence Analysis 70

CHAPTER 3. SEQUENCING AND SEQUENCE ANALYSIS OF THE *ykcA - ykfD* REGION OF THE *B. SUBTILIS* CHROMOSOME

3.1. Introduction 81
3.2. Results and Discussion 83
3.2.1. Sequencing of pSal16 83
3.2.1.1. Operon structure and open reading frames (ORFs) 83
3.2.1.2. *ykcABC* operon 84
3.2.1.3. *ykdA* operon 85
3.2.1.4. *ykeA* (*proG*) operon 85
3.2.1.5. *dppABCDE* operon 85

3.2.1.6. *ykfABCD* operon 86
3.2.2. The *Bacillus subtilis* Functional Analysis Project 87
3.2.2.1. Expression analysis of the *ykfABCD* operon 88
3.2.2.2. Other phenotypes 89
3.2.3. HtrA-Family of Serine Proteases in *B. subtilis* 89
3.2.3.1. *B. subtilis* contains two additional members of the HtrA family of proteases 90
3.2.3.2. Predicted topology of *B. subtilis* HtrA proteases 91
3.2.3.3. PDZ domains of *B. subtilis* HtrA proteases 92
3.2.3.4. Expression signals for the *ykdA, yvtA* and *yycK* genes in *B. subtilis* 93

CHAPTER 4. A STUDY OF *ykdA* ENCODING A HtrA-TYPE SERINE PROTEASE

4.1. Introduction 110
4.2. Results 113
4.2.1. Sequence analysis of *ykdA* 113
4.2.2. Induction Of *ykdA* Expression 114
4.2.2.1. Heat Shock Induction of *ykdA* Expression 114
4.2.2.2. Effect of other stressors on *ykdA* expression 114
4.2.3. Expression of *ykdA* is Not Affected in Strains Containing Mutations in the Regulators of Class I, Class II or Class III Heat Shock Genes 115
4.2.4. Investigation of the effects of mutations in genes encoding ECF-type sigma factors on the expression of *ykdA* 116
4.2.5. Control sequences involved in thermoinduction of *ykdA* 118
4.2.6. YkdA negatively autoregulates its own expression 119
4.2.6.1. Increased levels of YkdA reduce the heat shock induced expression of *ykdA* 120
4.2.6.2. The protease activity of YkdA is not required for negative autoregulation 121
4.2.7. Transcriptional analysis of *ykdA* expression 121
4.2.8. Phenotype of *ykdA* mutants 122
4.2.8.1. Increased hydrogen peroxide resistance of strain DN26 is not mediated through PerR inactivation 122
4.3. Discussion 124

CHAPTER 5. CO-ORDINATED EXPRESSION AND FUNCTIONAL RELATIONSHIP BETWEEN YkdA AND YvtA.

5.1. Introduction 142

5.2. Results 145
5.2.1. Temporal expression of the genes encoding the HtrA-like serine proteases 145
5.2.2. Heat shock induction of the genes encoding the HtrA-like serine proteases 146
5.2.3. Induction of *ykdA* and *yvtA* expression in response to heterologous amylase production 148
5.2.3.1. Investigation of the effect of amylase production on the expression of *yycK* 150
5.2.4. Transcriptional analysis of *yvtA* 150
5.2.4.1. AmyL induction of both *ykdA* and *yvtA* is at the level of transcription and utilises the heat shock induced promoters 150
5.2.5. Phenotypic analysis of *ykdA* and *yvtA* mutants 151
5.3. Discussion 153

CHAPTER 6. GENERAL DISCUSSION

6.1. Summary of Results 168
6.2. Structure and localisation of *B. subtilis* HtrA-like proteases 169
6.3. Expression Analysis of *B. subtilis* HtrA-like Proteases 171
6.3.1. Stress induced expression of *ykdA* and *yvtA* 171
6.3.2. Secretion stress induction of YkdA and YvtA 172
6.4. Regulation of Expression of HtrA-like Proteases of *B. subtilis* 173
6.4.1. Autoregulatory nature of *ykdA* and *yvtA* expression 174
6.4.2. The CssR/S two component system regulates expression of *ykdA*(*htrA*) and *yvtA* 176
6.4.3. Another regulator(s) for HtrA-like proteases in *B. subtilis* ? 177
6.5. Roles of *B. subtilis* HtrA-like proteases 178
6.6. Concluding Remarks 182

REFERENCES 185

LIST OF TABLES

		Page No.
Table 2.1.	Bacterial Strains and Plasmids Used in This Study	71
Table 3.1.	Features and Homologies of Open Reading Frames Found in Proximity to *ykdA*	95
Table 3.2.	*In silico* Promoter Analysis of Opening Reading Frames	96
Table 3.3.	Probable Transcriptional Terminators Defining Likely Operon Structures	97
Table 4.1.	Levels of β–Galactosidase Activity from Strains DN2 and PB198 Exposed to Different Stressors	127
Table 4.2.	Expression from the P*ykdA-bgaB* Fusion in Strains Containing Mutations in Genes Encoding Putative ECF-type Sigma Factors	128
Table 4.3.	Expression of *bgaB* Fused to Deleted Derivatives of the *ykdA* Promoter in Wild-type and *ykdA* Mutant Strains	129
Table 5.1.	Expression of Transcriptional Fusions between the *ykdA* and *yvtA* Promoters and the *bgaB* Reporter gene in Various Genetic Backgrounds at 37°C and 48°C	156

LIST OF FIGURES

Page No.

Figure 1.1. Structure of DegP 19

Figure 1.2. Regulation of the σ^E response in *E. coli.* 25

Figure 1.3. Regulation of the CpxRA two-component extracytoplasmic stress response 36

Figure 1.4. Regulation of Class I heat shock genes in *B. subtilis* 44

Figure 1.5. Regulation of the Class II σ^B-dependent general stress response in *B. subtilis* 45

Figure 1.6. Regulation by the CtsR repressor of Class III heat shock genes in *B. subtilis* 49

Figure 2.1. Generation of transcriptional fusions to *bgaB* using the integration vector pDL 80

Figure 3.1. Physical map of the *B. subtilis* chromosome region covered in pSAL16 98

Figure 3.2. Generation of mutant/fusion strains using the integrating vector pMUTIN 99

Figure 3.3. Expression analysis of the *ykfABCD* operon 100

Figure 3.4. Sequence alignment of the *E. coli* DegP and *B. subtilis* YkdA proteins 101

Figure 3.5. Chromatograms displaying identification and correction of sequencing errors in the *yvtAB* nucleotide sequence 102

Figure 3.6. Translation of the *yvtA* sequence after correction of detected frameshifts 103

Figure 3.7. Multiple sequence alignment and features of HtrA proteins from *B. subtilis* 105

Figure 3.8. PDZ domains of HtrA proteases from *B. subtilis* and DegP of *E. coli* 106

Figure 3.9. Secondary structure analysis of the PDZ domains of *B. subtilis* HtrA-like proteases and *E. coli* DegP 107

Figure 3.10. Comparison of the controlling regions of *ykdA* and *yvtA* genes 108

Figure 3.11. ECF-type promoter motifs in the promoter regions of *ykdA* and *yvtA* compared to ECF-type promoters from other bacterial genes 109

Figure 4.1. A comparison of the promoter regions of *ykdA* and *yvtA* 130

Figure 4.2. Profile of growth and LacZ accumulation in strain DN2 during growth at 37°C and after heat shock at 48°C 131

Figure 4.3. Expression of *ykdA* transcriptional fusions in *B. subtilis* strains mutated in the regulators of Class I, II and III heat shock regulons 132

Figure 4.4. The end-points of the 5'- and 3'- deletions of the *ykdA* promoter region generated by *Bal* 31 exonuclease 133

Figure 4.5. Negative autoregulation of *ykdA* 134

Figure 4.6. Expression and heat shock induction of P*ykdA-bgaB* transcription fusion in a strain overproducing YkdA 135

Figure 4.7. Construction of strain AH24, containing an *ykdA* allele mutated in the catalytic serine codon 136

Figure 4.8. Role of YkdA protease activity in negative autoregulation of P*ykdA-bgaB* activity 137

Figure 4.9. Transcriptional analysis of *ykdA* 138

Figure 4.10. Survival of wild-type and DN26 (*ykdA*Δ439) strains after exposure to heat and hydrogen peroxide 139

Figure 4.11. Expression of an *ahpCF-lacZ* transcriptional fusion in wild type

and *ykdA* mutant backgrounds 140

Figure 4.12. Phylogenetic analysis of the HtrA proteases from *B. subtilis* and *E. coli*. 141

Figure 5.1. Profiles of growth and β–galactosidase accumulation from strains bearing P*ykdA-bgaB* or P*yvtA-bgaB* fusions in various genetic backgrounds 157

Figure 5.2. Temporal expression of *yycK* in different media. 158

Figure 5.3. Expression of *yycK* following heat shock 159

Figure 5.4. Profiles of growth and β–galactosidase accumulation from P*ykdA* transcriptional fusions in the presence and absence of heterologous amylase production during the growth cycle in LB broth. 160

Figure 5.5. Expression and secretion of AmyL in strain KS408 in the presence or absence of xylose inducer 161

Figure 5.6. Expression of a *yycK'-bgaB* transcriptional fusion in response to heterologous amylase (AmyL) production 162

Figure 5.7. Primer extension analysis of *yvtA* following heat shock 163

Figure 5.8. Primer extension analysis of *ykdA* and *yvtA* in response to production of the heterologous protein AmyL 164

Figure 5.9. Growth of strains DN2 (WT), DN27 (*ykdA*Δ439) and DN112 (*yvtA*::kan) at 37°C and 48°C 165

Figure 5.10. Growth and survival profiles of *ykdA* and *yvtA* single and combinatorial mutant strains grown at elevated temperatures 166

Figure 5.11. Survival of *ykdA* and *yvtA* single and combinatorial mutant strains exposed to 10mM hydrogen peroxide 167

Figure 6.1. Model for the regulation of *ykdA* and *yvtA* by the CssRS two-component system 184

ABBREVIATIONS

aa	Amino acids
Ap^r	Ampicillin resistance
ATP	Adenosine triphosphate
BFA	*Bacillus subtilis* Functional Analysis Project
BgaB	β-galactosidase (*B. stearothermophilus* derived)
BLAST	Basic local alignment search tool
bp	Base pair
C	Carboxyl
Cm^r	Chloramphenicol resistance
Da	Dalton
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
E	Core RNA polymerase
ECF	Extracytoplasmic function
EDTA	Ethylenediaminetetraacetic acid
Erm^r	Erythromycin resistance
GCG	Genetics Computer Group
h	hour
HK	Histidine kinase
Hsp	Heat shock protein
IAP	Inhibitor of Apoptosis Protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Inverted repeat
Kbp	Kilobase pairs
kDa	kilo-Dalton
Km^r	Kanamycin resistance
LacZ	β-galactosidase
LB	Luria Bertani
MCS	Multiple cloning site
min	minute
mJ	Millijoules
N	Amino
~P	Phosphorylated
OD	Optical density
OMP	Outer membrane protein
ONPG	*o*-Nitrophenyl β-D-galactopyranoside
ORF	Open reading frame
PCR	Polymerase chain reaction

ABBREVIATIONS (cont.)

PDZ	Post synaptic density, Disc large and Zo-1 proteins
RBS	Ribosome binding site
RNA	Ribonucleic acid
Rpm	Revolutions per minute
RR	Response regulator
SSC	Salt sodium citrate
SDS	Sodium dodecyl sulfate
SM	Schaeffer's sporulation medium
Sp^r	Spectinomycin resistance
XBAP	X11 version, Sequence assembly program (Staden package)
X-gal	5-Bromo-4-chloro-3-indolyl-b-D-galactopyranoside
XNIP	X11 version, Nucleotide interpretation program (Staden package)

Publications Arising From This Work

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., ...Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., (and 138 other authors) Danchin, A., (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**: 249-256.

Noone, D., Howell, A., and Devine, K.M. (2000) Expression of *ykdA*, encoding a *Bacillus subtilis* homologue of HtrA, is heat shock inducible and negatively autoregulated. *J Bacteriol* **182**: 1592-1599.

Noone, D., Howell, A., Collery, R., and Devine, K.M. (2001) YkdA and YvtA, HtrA-like serine proteases in *Bacillus subtilis*, engage in negative autoregulation and reciprocal cross-regulation of *ykdA* and *yvtA* gene expression. *J Bacteriol* **183**: 654-663.

Scanlan, E., O'Reilly, M., Krogh Andersen, K., Noone, D., and Devine, K.M. (2001) Screening for Increased Sensitivity or Resistance and Level of LacZ Reporter Gene Expression in Response to the Oxidising Agents Hydrogen Peroxide and Methyl Viologen. In *Functional Analysis of Bacterial Genes: A Practical Manual*. Schumann, W., Ehrlich, S.D. and Ogasawara, N. (eds). Chichester UK.: Wiley, pp. 227-230.

CHAPTER ONE

GENERAL INTRODUCTION

1.1. *B. subtilis* as a model organism

Bacteria must adapt rapidly to changes in their environment and this is especially important in bacteria which do not inhabit relatively stable protected niche environments. As the soil environment can change rapidly bacteria must therefore be able to mount rapid changes to allow them to adapt and survive. One such bacterium is *Bacillus subtilis* which is a natural soil dweller. *B. subtilis* is a Gram-positive bacterium which has been extensively studied as a model organism particularly in relation to its responses to environmental and nutritional stresses (Msadek, 1999). A range of different adaptational responses are possessed by *B. subtilis* to combat nutrient starvation, such as competence development, degradative exoenzyme synthesis, motility and antibiotic production. If these responses fail to alleviate the impoverished nutritional stress of the cell then *B. subtilis* has the ability to enter into a dormant state via the developmental pathway of sporulation (Stragier and Losick, 1996). These developmental and stress responses have been the focus of many studies in *B. subtilis*.

1.1.1. Natural Competence of *B. subtilis*

One of the main factors in the development of *B. subtilis* as a model organism was the ability of *B. subtilis* to take up and recombine exogenous DNA into its genome as part of its response to nutritional stress. This response was termed competence and was recognised early in the pioneering work of Anagnostopoulos and Spizizen as providing a valuable tool for genetic manipulation of the bacterium (Anagnostopoulos and Spizizen, 1961). There now exists an array of specialised plasmid and phage vectors which are routinely used to introduce mutations, transcriptional or translational reporter fusions or complementing genes to the *B. subtilis* genome (Dowds *et al.*, 1988; Perego, 1993) The regulation of competence development and mechanism of DNA uptake in *B. subtilis* have been studied primarily by the Dutch laboratories of Venema and Bron and also by Dubnau, Zuber and Grossman. The regulation of competence development is now well understood and greater insights into the mechanisms of DNA uptake are also being elucidated (Dubnau, 1999; Dubnau and Provvedi, 2000).

1.1.2. Exoenzyme production by *B. subtilis*

Release of exoenzymes for the breakdown of other less favourable nutrients is also a response to nutritional stress in *B. subtilis*. The system controlling these events are mediated by the DegU-DegS two component system. The enzymes synthesised as a result of the activation of DegU~P are secreted into the medium/environment and hydrolyse large complex sugars and proteins. The DegU protein is also able to act as a repressor of competence thus acting as dual regulator of different stationary phase responses (Msadek, 1999).

1.1.3. Developmental programme of Sporulation

If the development of competence and the synthesis of degradative exoenzymes cannot prolong the survival of *B. subtilis* then the cells undergo the developmental programme of sporulation as a last resort. This involves the production of an endospore within the once vegetative cell, referred to as the mother cell. The spore can lie dormant for long periods until more favourable conditions exist for germination and resumption of growth. The process of sporulation is complex and fascinating, the study of which is viewed as a paradigm of intercompartmental signal transduction in prokaryotes (Stragier and Losick, 1996). The key regulators of commitment to the sporulation pathway are the levels of phosphorylated SpoOA protein and other mechanisms for sensing the nutritional and growth status of the cell. A key sensor of this latter point is the intracellular level of GTP nucleotide which determines the activity of the repressor CodY (Ratnayake-Lecamwasam *et al.*, 2001). When low levels of GTP are present the activity of CodY is reduced allowing an increase in the transcription of many of the important early stationary phase genes necessary for the commitment to sporulation. A parallel system of other sensors also contribute to the increase in the levels of SpoOA~P, collectively termed the phosphorelay (Dworkin and Losick, 2001 and references therein).

Once initiated, sporulation proceeds by a well defined series of morphological events. In concert with these steps a series of compartmental specific transcriptional programmes is undertaken. The basis for these co-ordinated transcriptional events is the production and

activation of mother cell and forespore specific sigma factors. The σ^F and σ^G factors are active in the forespore while the σ^E and σ^K factors are active in the mother cell. Initially all the sigma factors are produced in an inactive form, as pro-sigma factors in the mother cell and as anti-sigma factor bound proteins in the forespore. Activation of the sigma factors requires the action of proteins or protein complexes bound to the membrane dividing the mother cell and forespore and necessitating interdepartmental signalling termed crosstalk. This results in a temporally and spatially defined series of events culminating in the release of the endospore from the mother cell.

1.1.4. Genome Sequencing and Functional Analysis of *B. subtilis*.

The genesis of this thesis emanated from a collaborative European and Japanese project to undertake the sequencing of the genome of *B. subtilis* (Devine, 1995). As part of that project approximately 15-kbp of the genome between the *ykcA* and *ykfD* genes were sequenced by this author (Kunst *et al.*, 1997). During sequencing of that region an open reading frame (ORF), *ykdA*, was found to encode a homologue of the *degP*(*htrA*) gene from *E. coli*. (see section 3.2). The analysis of the other ORFs found in the region sequenced is presented in chapter three. Genes whose encoded putative proteins did not show any homology to those in the databases or where the function of homologues in other organisms were not known were designated 'genes of unknown function' and given the prefix 'y'. A second international collaborative project, involving many of the same laboratories that participated in the genome sequencing project, was undertaken to analyse the function of the '*y*' genes (see section 3.1).

One of the genes sequenced as part of the work for this thesis, *ykdA*, was homologous to a gene from *E. coli degP*(*htrA*) encoding a heat shock inducible serine protease. The expression and regulation of *degP* was novel as it was independent of the sigma 32 heat shock response which regulates many of the important cytoplasmic chaperones and proteases. Subsequently, a large amount of work resulted in the elucidation of the regulatory mechanisms controlling *degP* expression (see section 1.5.7.). Furthermore, most eubacteria were found to encode one or more HtrA-like serine proteases in their genomes. Despite this

finding, the regulation of their expression remained unknown for the most part. The focus of my research was the investigation of the expression and roles of HtrA-like proteases in *B. subtilis* which is presented in chapters four and five. The remaining part of this introduction presents reviews of (i) the heat shock response, with particular emphasis on the extracytoplasmic heat shock response in *E. coli*; (ii) the knowledge to date on the roles and regulation of expression of HtrA-like serine protease in other bacteria and eukaryotic organisms.

1.2. The Heat Shock Response.

The process of correctly folding proteins is vital to all organisms, to produce functional enzymatic and structural polypeptides and prevent deleterious protein aggregation. Some proteins can fold spontaneously to their native state but many larger proteins must be aided in this process by families of highly conserved proteins termed molecular chaperones. Proteins which cannot attain their native structure are degraded to ensure their accumulation or deleterious interaction with other proteins does not become lethal to the cell. This degradation is carried out by proteases, usually in an energy dependent manner. The proteases responsible for this detoxification are highly conserved. Chaperones and proteases are required during both normal growth and under stress conditions when folding defects and aggregation increase. Our current understanding of the roles, mechanisms and regulation of the chaperones and proteases derives from the studies of protein folding and degradation under conditions of heat stress. This universally conserved biological event is termed the heat shock response. Most of the structural and mechanistic studies of the heat shock proteins (hsps) have been undertaken in the model organism *Escherichia coli* or on heat shock proteins isolated from *E. coli* . A brief overview of the two main classes of heat shock proteins, the molecular chaperones and proteases, is presented here with emphasis on referral to many in depth reviews and books dedicated to this subject.

1.2.1 The *E. coli* Cytoplasmic Heat Shock response.

The study of the heat shock response in *E. coli* has focused primarily on the induction of molecular chaperones and proteases localised to the cytoplasm. The vast majority of these proteins, identified by genetic and biochemical analyses, are regulated by the heat shock sigma factor, σ^{32} (Gross, 1996; Yura *et al.*, 2000). An overview of their structure and roles is presented below.

1.2.2 DnaK/DnaJ chaperone system.

DnaK is the archetypal protein for the HSP70 class of heat shock proteins (see reviews by Buchberger and Bukau, 1997; Frydman, 2001) Homologues are found throughout all bacteria and in most eukaryotes. The DnaK protein functions in conjunction with its co-chaperone DnaJ and a nucleotide exchange factor GrpE. The DnaK chaperone acts both co- and post-translationally and has been shown to bind to nascent polypeptides displaying hydrophobic residues in a linear conformation (Bukau *et al.*, 2000). Substrates sizes for the DnaK system are extensive unlike that of the GroEL system. There may be redundancy in co-translational chaperoning activities within the cell since another small chaperone termed trigger factor (TF), has also been shown to act on polypeptides emerging from the ribosome (Bukau *et al.*, 2000). The DnaK protein is composed of 2 domains; an N-terminal domain required for ATP binding and hydrolysis, and a C-terminal possessing the substrate binding domain. The mode of folding by DnaK involves cycles of binding and release of its substrate molecule concomitant with hydrolysis of ATP which is mediated by its co-chaperone DnaJ and the nucleotide exchange factor GrpE. The DnaK chaperone machine has roles in both folding newly synthesised proteins and those denatured or aggregated by stress conditions. DnaK has recently been shown to co-operate with the ClpB chaperone in preventing and/or reversing protein aggregation (Tomoyasu *et al.*, 2001). *E. coli* requires DnaK at heat shock temperatures but not under normal growth conditions where an overlapping role with TF seems to occur. The DnaK chaperone machine also plays an important role in autoregulating the cytoplasmic heat shock response by controlling the activity and stability of σ^{32}, the main cytosolic heat shock transcription factor (see section 1.3).

1.2.3. GroEL/GroES chaperone system

The GroEL/ES chaperone machine is the second main chaperone system in *E. coli* and all other eubacteria (see Buchberger and Bukau, 1997; Frydman, 2001). Homologous systems are also found in plant and eukaryote organelles. Similar to the DnaK system, GroEL and its associated co-chaperone GroES use the binding and hydrolysis of ATP for cycles of binding and release to achieve correct substrate folding. However, the structure and mechanisms used by the GroEL/ES are different. GroEL forms two stacked heptameric rings of identical subunits and is capped by the GroES subunit when folding is in progress. This arrangement provides a protected cylindrical environment where substrate folding can take place. The substrates bound by GroEL are also different to those bound by DnaK. GroEL binds polypeptides of 60 kDa or less and does not act co-translationally. Many of GroEL's substrates are proteins that are unable to attain their correct conformation through association with DnaK. Each subunit of GroEL has three domains; an outer domain for binding to exposed hydrophobic residues of its substrate molecules which is also used to contact GroES, an intervening hinge like domain and an inner ATP binding and hydrolysing domain (see Frydman, 2001). The GroEL/ES system is essential in *E. coli* and in many other bacteria for viability. GroEL represents one of the most studied of all chaperone proteins and detailed reviews on the structural and functional aspects of the GroEL machine have been presented by Bukau and Horwich (1998) and Horwich *et al.* (1999).

1.2.4. Heat Shock Proteases

The second major class of heat shock proteins are proteases. In *E. coli* these proteins are involved in degrading proteins which have not achieved their functional folded confirmation or in controlling the action of regulators by proteolysis. A comprehensive review detailing the major cytosolic proteases, their structures, roles and cellular targets has been presented by Gottesman (1996).

In common with the cytosolic chaperones, the main heat shock proteases, ClpAP, ClpXP, Lon and FtsH, also require ATP for substrate binding and hydrolysis (Hoskins *et al.*, 1998).

In two of these proteases, Lon and FtsH, the proteolytic and ATP binding domains are present in the same polypeptide. The structure of the third major family of proteases, the Clp systems, differ in that the substrate and ATP binding domains reside in separate proteins which combine with the proteolytic component ClpP or ClpQ (HslV) to form a heteroligomeric complex (see review by Porankiewicz *et al.*, 1999)

Two of the cytosolic proteases of *E. coli,* Clp and FtsH, share a common feature with that of the GroEL/ES chaperonins in forming oligomeric ring structures to undertake their roles (Shotland *et al.*, 1997; Horwich *et al.*, 1999). The substrate recognition and binding mechanisms used by both Lon and Clp proteases would also now appear to be similar (Smith *et al.*, 1999). Protease substrates can in general be targeted by exposed hydrophobic stretches of amino acids or by tagging at their N- or C- termini.

The heat shock proteases in the *E. coli* cytosol are now known to have two roles. The first which predominates under heat stress is to degrade unfolded proteins. In this role they may be co-operation between the classical chaperones in presenting the substrates to the proteases. However, it is now know that the substrate binding components of the Clp proteases may aid in disaggregation of substrates before their presentation to the proteolytic complex. Furthermore, independent chaperone activity has been shown for ClpA and ClpX regulatory Clp subunits and for Lon and FtsH proteases (Smith *et al.*, 1999; Wickner and Maurizi, 1999). Which of these fates awaits the substrates may be determined by " triage and kinetic partitioning of non-native proteins" whereby some substrates may be competent for refolding or alternatively be degraded due to the extent of damage or aggregation (Gottesman *et al.*, 1997). In addition to the vital role of cytosolic proteases in degrading stress damaged proteins each of the three main proteases perform regulatory functions by proteolysis of global regulators to effect precise control over their transcriptional activities (Gottesman, 1999).

1.3. Regulation of the Cytosolic Heat Shock Response in *E. coli*.

Heat shock is a universally conserved response to an increase in the amount of misfolded or damaged proteins. It is characterised by a swift and transient induction of heat shock proteins (hsps). The induced proteins, primarily chaperones and proteases function to decrease the accumulation of these potentially lethal misfolded products. Hsps are present in cells growing at physiological temperatures reflecting their importance in the general process of folding and degradation. Homologues of the ATP-dependent cytosolic proteases described above have been identified in many different bacteria (Koonin *et al.*, 2000). However, given its importance as a paradigm for the detailed investigation of biological systems it is not surprising that the understanding of the regulation of the heat shock response is at an advanced stage in *E. coli*. The regulation of the heat shock response in *E. coli* and some other bacteria has been recently reviewed and an overview is presented here (Gross, 1996; Yura *et al.*, 2000) . Expression of members of the cytosolic heat shock response are transcribed from both σ^A and σ^{32} type sigma factors. σ^A is the vegetative sigma factor and σ^{32} the heat shock sigma factor. Investigation has shown that the increase in heat shock protein production results from an increase in the synthesis, stability and activity of σ^{32}.

1.3.1. Increased synthesis of σ^{32}.

Transcriptional analysis of the *rpoH* gene encoding σ^{32} had identified four promoters, three of which are σ^A -type and the fourth is a σ^E -type, an alternate sigma factor induced at near-lethal high temperatures and by extra-cytoplasmic stress signals (Erickson *et al.*, 1987; Erickson and Gross, 1989) (see section 1.5.7). Studies of these promoters showed that after heat shock from 30°C to 42°C there is only a small increase in *rpoH* transcription. This increase in transcript could not account for the rapid increase in the level of σ^{32} which was observed suggesting that the response was regulated post-transcriptionally.

As a starting point for the investigation into the post-transcriptional model of regulation translational fusions between the *rpoH* gene and the *lacZ* gene were constructed. Use of nested sets of deletion fusions with differing amounts of the *rpoH* gene showed that there were two regions at the 5' end of the *rpoH* gene necessary for correct thermoregulation

(Yura *et al.*, 1993). Structural analysis of these regions showed that they could potentially form a stable secondary structure at normal temperatures and that this would prevent access of the translational machinery. Very recent detailed RNA structural data, from the laboratory of Yura, has now confirmed the existence of stable secondary structure (Morita *et al.*, 1999). The RNA secondary structure is directly responsive to the growth temperature of the cell; at heat shock temperatures the duplexes sequestering the ribosome binding site are melted, translational repression is relieved and synthesis of σ^{32} from a pool of pre-existing mRNAs can take place. The significance of this finding has prompted the model of *rpoH* RNA acting as a cellular thermometer (Storz, 1999).

1.3.2. Increased Stability and Activity of σ^{32}.

Earlier studies of the heat shock response in *E. coli* showed that mutations in the *dnaK*, *dnaJ* or *grpE* genes increased the activity and amount of σ^{32} under non heat shock conditions (Straus *et al.*, 1990). These results led to the discovery that a third region of the *rpoH* gene determined stability of the β-galactosidase fusion proteins (Nagai *et al.*, 1994). The fusion proteins were seen to be stabilised in an DnaK-DnaJ-GrpE dependent manner. Subsequently a role for the membrane bound FtsH heat shock protease in regulating σ^{32} was elucidated and confirmed by *in vitro* analysis with purified proteins (Herman *et al.*, 1995). The role of proteolysis in controlling the stability and activity of σ^{32} was not unique however as cells mutated in *ftsH* did not show a large increase in the level of hsps. The third regulatory region of σ^{32}, region C, demonstrated that it was a binding region for the DnaK chaperone machine proteins and that in the *ftsH* mutant cells the increased amount of σ^{32} was held in a largely inactive form (Tatsuta *et al.*, 1998; Blaszczak *et al.*, 1999). In recent work it has been shown that modulating the levels of the DnaK chaperone machine to small extents has significant effects on σ^{32} activity. In particular, the adaptation phase of the heat shock response controlling reduction of σ^{32} activity after the initial induction is very much reliant on DnaK chaperone function (Tomoyasu *et al.*, 1998). The current model suggests that there is competition between the DnaK chaperones and RNA polymerase for binding of σ^{32}. When bound by chaperones σ^{32} may be presented to FtsH or other ATP-dependent heat shock proteases such as Lon or ClpQY for degradation. This system provides for an autoregulatory

circuit in which the level and activity of σ^{32} is rapidly increased following heat shock by the titration of chaperones and proteases to damaged proteins. The rapid increase in Hsps including the DnaK chaperone machine proteins and proteases allows the sequestration of σ^{32}, reducing its activity and subsequent stability by degradation. This results in the level of heat shock proteins decreasing after the initial burst but it is still above the pre-heat shock levels because increased synthesis is maintained by the increased translation from the melted RNA secondary structures of the *rpoH* transcript. Further elements in the control mechanisms are provided by the increased proteolysis of σ^{32} by ClpQY and FtsH at higher temperatures (Kanemori *et al.*, 1999). This checks the accumulation of σ^{32} resulting from increased synthesis. Lastly a newly discovered protease EcfE has been implicated in the control of σ^{32} levels (Dartigalongue *et al.*, 2001a). EcfE is an essential inner membrane protease that has been shown to degrade σ^{32} both *in vivo* and *in vitro*, although this function of EcfE has recently been disputed (see section 1.5.7.4).

1.4. The Extracytoplasmic Heat Shock and Unfolded Protein Response in *E. coli*

The cytoplasmic heat shock response is controlled by σ^{32}. However, *E. coli* has three further distinct compartments; the inner membrane, the periplasm and the outer membrane, collectively termed the cell envelope. It is now know that the cell envelope can respond to localised protein unfolding and aggregation in a similar manner to the cytoplasm by the production of its own specialised chaperones and proteases. Furthermore, it now appears that there are at least four mechanisms by which a response(s) can be mounted. Two of these systems, the σ^{E} response and the Cpx response have been recently reviewed (Raivio and Silhavy, 2000), while the UptR mediated system appears novel (Guigueno *et al.*, 2001). A fourth system is implied from studies demonstrating suppression of envelope toxicity caused by a processing-defective OMP, which appears to be independent of the other three systems (Cosma *et al.*, 1998).

1.4.1. Chaperones and protein folding agents

A major function of the Cpx and σ^E controlled responses is to ensure that proteins under normal physiological and stress conditions can achieve their proper folded state and that proteins are prevented from forming unproductive and potentially lethal aggregates. Unsurprisingly, members of Cpx and σ^E regulons include proteins with the ability to aid the folding of inner membrane, outer membrane and periplasmic proteins. An increasing number of these so-called "foldases" have been identified (Raivio and Silhavy, 2000; Dartigalongue *et al.*, 2001b). These include the disulphide bond proteins (Dsb) which facilitate the formation and rearrangement of correct disulphide bonds (Rietsch and Beckwith, 1998). Another important class of folding agents are the peptidyl-prolyl isomerases (PPI's) which are involved in isomerisation of X-Pro bonds. A group of general (Skp) and specific chaperones (Pap) are also important. The total number of proteins and their known or putative functions have been significantly increased by the recent studies of Dartigalongue and co-workers (Dartigalongue *et al.*, 2001b). This work identified addition proteins involved in lipopolysaccharide synthesis, sensory proteins, lipoproteins and a number of key regulatory transcription factors, e.g. RpoD as being regulated by σ^E and/or CpxRA.

1.4.2. Proteases

In addition to foldases, a number of proteases have been identified in this cellular compartment.

1.4.2.1. OmpT protease

OmpT is localised to the outer membrane of *E. coli*. It has been shown to be an unusual serine-type protease (Vandeputte-Rutten *et al.*, 2001). A large number of different substrates have been identified for the OmpT protein reflecting an unusual proteolytic site specificity. A recurring theme that is evident from the literature on OmpT is its induction and subsequent interference in the production of recombinant proteins when these are induced in *E. coli* overproducing strains (Baneyx and Georgiou, 1990; Henderson *et al.*, 1994; Meerman and Georgiou, 1994). To combat this unwanted effect many of the commercially available strains contain mutations in the *ompT* gene. A second OmpT-like protease (OmpP) exists in *E. coli*

which is plasmid encoded (Matsuo *et al.*, 1999). The OmpT protease has been shown to be induced by heat shock but the mechanism for its regulation is not currently known (Gill *et al.*, 2000).

1.4.2.2. Tsp (Prc)

Tail specific protease (Tsp) is a second extracytoplasmic protease of *E. coli*. This protease is located in the periplasm and the polypeptide contains separate catalytic and substrate binding domains (Keiler and Sauer, 1996). The substrate binding region has been shown to be a PDZ domain, specified by an internal sequence of ~100 amino acid residues (Beebe *et al.*, 2000). The substrate catalytic domain is located at the carboxyl terminus. As with OmpT the catalytic residues are somewhat unusual being composed of a serine-lysine dyad. *E. coli* mutated in the *tsp* gene show a minor heat sensitivity when grown in low salt media and also increased sensitivity to antibiotics. However the most studied aspect of Tsp function is its role in degradation of SsrA mediated carboxyl terminally tagged proteins. This system adds a run of 10 amino-acid residues to proteins stalled on ribosomes which then targets the tagged protein for degradation by Tsp in the periplasm or Clp proteases in the cytosol (Gottesman *et al.*, 1998). The types of amino acids added are important and invariably are composed of a terminal triplet of non-polar small residues, this class being most important for the C-terminal residue (Keiler and Sauer, 1996). The specificity of recognition and binding to these tagged proteins by Tsp is due to the PDZ domain (Beebe *et al.*, 2000). The regulation of expression of Tsp has not been uncovered as of yet.

The third type of extracytoplasmic proteases in *E. coli* are members of the HtrA family of serine proteases. Because the work contained within this thesis is directly concerned with the study of HtrA-like proteases in *Bacillus subtilis* a detailed review of HtrA proteases both from *E. coli* and from other prokaryotic and eukaryotic organisms will be presented in the next section.

1.5. HtrA proteases

One of the emerging aspects from the continually increasing number of microbial genomes being sequenced is the ubiquity of certain classes of proteases contained in most if not all bacteria (Koonin *et al.*, 2000). These include homologues of the cytosolic heat shock proteases, Lon, Clp and FtsH described above. In regard to extracytosolic bacterial proteases the most prevalent would appear to be HtrA family proteases a feature noted in a previous review (Pallen and Wren, 1997). Since the review by Pallen and Wren (1997) a substantial body of literature has accumulated regarding the structure, distribution, function, applications and regulation of HtrA proteases in both prokaryotes and eukaryotes. I will review this protease family under these headings.

With regard to nomenclature, the Deg acronym, relating to the original phenotype of reduced ***deg***radation of certain fusion proteins used for screening purposes, will be used to refer exclusively to the *E. coli* prototype enzyme and encoding gene (Strauch and Beckwith, 1988; Strauch *et al.*, 1989). Homologues of DegP from other organisms will be referred to using the name given in the first reported publication or called HtrA-like, as this general collection of proteases are now termed the HtrA family.

1.5.1. Structure of HtrA proteases

Proteases of the HtrA family have in general three distinct domains; (i) an NH_2 terminus which generally contains a localisation or signal sequence, (ii) a central region containing the catalytic core and (iii) a C terminus containing the protein substrate recognition domain mediated by one or more PDZ domains.

1.5.2. N-termini

1.5.2.1. *E. coli* DegP

Initial identification of DegP came from the observation that fusion proteins secreted to the periplasm and previously degraded were stabilised in the *degP* background (Strauch and Beckwith, 1988). This was the first indication that DegP protein might be periplasmic. Subsequent cloning and sequencing of the gene revealed that the N-terminus of the encoded

protein had a canonical signal peptide composed of hydrophobic residues followed by a recognised cleavage site for leader peptidases (Lipinska *et al.*, 1988). Experimental evidence for the periplasmic localisation of DegP came from the use of DegP-PhoA fusions which showed that strains harbouring such fusions showed alkaline phosphatase activity (Strauch and Beckwith, 1988). Since the phosphatase activity is only active in extracytoplasmic environments this indicated that the DegP part of the fusion protein directed transport of alkaline phosphatase to the periplasm. Further evidence was provided by overexpression of the DegP protein and characterisation of its size and location in the overexpressing strains. The full length protein was seen as a 51-kDa band but most of the DegP polypeptide was observed to run at 48-kDa (Strauch *et al.*, 1989). This indicated that the putative signal peptide was processed after export to the periplasm. However, even in the overexpressing strain, the DegP protein was found in some membrane fractions indicating it may be membrane associated. Definitive data obtained through sphaeroplasting techniques has shown that the DegP protein is in fact truly periplasmic in nature and that the previous membrane association was probably artifactual. This likely occurred because of the native structure of DegP, as a hexamer (see below), may have prevented its escape through the peptidoglycan layer from osmotically shocked cells, which was not the case in the sphaeroplasted cells (Sassoon *et al.*, 1999).

1.5.2.2. Other Bacterial HtrA proteases

Similar to DegP nearly all other bacterial DegP homologues show amino-termini bearing subcellular localisation signals. In the majority of Gram-negative bacteria this is usually a canonical export signal sequence for export of the HtrA-like protein to the periplasm as is the case for HtrA-like proteins from *H. influenza* (Loosmore *et al.*, 1998), *Yersinia enterocolitica* (Li *et al.*, 1996), *Y. pestis* (Williams *et al.*, 2000), *Salmonella typhimurium* (Johnson *et al.*, 1991), *Brucella abortus* (Elzer *et al.*, 1994), *Legionella pneumoniae* (Pedersen *et al.*, 2001), *Bartonella henselae* (Resto-Ruiz *et al.*, 2000) and *Pseudomonas aeruginosa* (Boucher *et al.*, 1996). This latter organism contains two HtrA-like proteins, AlgW and MucD, both of which contain runs of hydrophobic residues at their N termini. Only the MucD protein contains a putative signal peptide cleavage site.

HtrA homologues have also been identified in Gram-positive bacteria. Gram-positive bacteria do not possess a double membrane structure enclosing a distinct periplasmic compartment. However in these bacteria signal peptides and membrane anchoring motifs are found within secreted or exported proteins (Tjalsma *et al.*, 2000; van Wely *et al.*, 2001). HtrA-like proteases have been identified in *Lactobacillus helveticus* (Smeds *et al.*, 1998), *Lactococcus lactis spp. lactis* strain IL1403 (Poquet *et al.*, 2000), *Streptococcus pneumoniae, S. pyogenes, S. gordonii, S. mutans, Enterococcus faecalis, Staphylococcus aureus* (Jones *et al.*, 2001) and *Bacillus subtilis* (Noone *et al.*, 2000) (see section 3.2). All of the HtrA-like proteins possess stretches of hydrophobic amino acid residues at their N-termini predicted to form transmembrane helices capable of anchoring these proteins to the cell surface. The catalytic domains positioned C-terminal to these membrane anchors are inferred from topological analysis to be extracytoplasmically located.

1.5.2.3 Eukaryotic HtrA proteases

At least two human homologues of DegP, L56 (humHtrA1) (Zumbrunn and Trueb, 1996; Hu *et al.*, 1998), and Omi (humHtrA2) have been identified and characterised (Faccio *et al.*, 2000; Gray *et al.*, 2000). Both proteins contain localisation signals at their amino termini, a cleavable signal sequence in the L56 protein and a transmembrane domain in the Omi sequence. The L56 protein has been shown to be secreted but the subcellular localisation of Omi was disputed until very recently. A surge of literature concerning Omi has now emerged related to its involvement in apoptosis (Suzuki *et al.*, 2001; Martins *et al.*, 2002; Verhagen *et al.*, 2002) (see section 1.5.6.4.). These studies have now definitively shown that Omi has two N-terminal localisation signals, an extreme N-terminal mitochondrial targeting sequence followed by a transmembrane domain motif. In normal cells Omi is present in the mitochondria, specifically localised to the mitochondrial inter-membrane space. However, upon receiving apoptotic stimuli Omi is released to the cytoplasm. Isolation of cDNAs from other animals has also indicated the presence of conserved HtrA-like proteins among mammals (Hu *et al.*, 1998).

Two HtrA-like proteins have been characterised in plants. DegP1 was characterised in pea and *Arabidopsis thaliana* and DegP2 only from *A. thaliana* (Itzhaki *et al.*, 1998; Haussuhl *et al.*, 2001). Both proteins contain localisation signals in the form of transit peptides, the equivalent terminology used for signal peptides in plants, at their N-termini. Both DegP1 and DegP2 are localised to the chloroplast thylakoid subcellular organelle where photosynthesis occurs. DegP1 is membrane bound with its proteolytic domain exposed to the lumen while the DegP2 protein is proposed to be peripherally membrane associated with its catalytic domain facing the stroma.

1.5.3. Catalytic domains of HtrA proteases

The highly conserved catalytic domain is immediately juxtaposed to the N-termini of the HtrA family . The *E. coli* DegP protein was shown to be an endopeptidase following its overexpression and ability to degrade casein *in vitro* (Lipinska *et al.*, 1990). This proteolytic activity was independent of ATP but was sensitive to the presence of diisopropylflourophosphate, a know inhibitor of serine proteases. In agreement with these results the primary sequence of DegP was noted to contain the sequence GNSGGAL which is a highly conserved motif surrounding the active site serine of trypsin-like serine proteases (Brenner, 1988). In addition to the serine at the catalytic site two other residues were know to be important for serine protease activity: a histidine and an asparagine both located N-terminal of the serine residue. Mutagenesis of either the Ser or His resulted in the abolition of protease activity from DegP confirming its activity as a serine protease (Skorko-Glonek *et al.*, 1995).

The very highly conserved central core is a theme of HtrA proteases found in eubacteria and eukaryaea. This sequence conservation is primarily present in the immediate areas of the residues constituting the catalytic triad. The highly conserved motifs are readily apparent when multiple alignments were performed on HtrA protease sequences as diverse as bacteria and mammals (Boucher *et al.*, 1996; Pallen and Wren, 1997; Gray *et al.*, 2000; Jones *et al.*, 2001). These conserved catalytic residues together with their surrounding motifs are the defining features used to identify new members of the family of HtrA proteases.

1.5.4. HtrA proteases contain PDZ substrate recognition domains at their C-termini.

1.5.4.1. PDZ domains

The primary feature of HtrA proteases is the conserved central core. However members of the HtrA proteases also contain a second conserved element at their carboxyl termini (Pallen and Ponting, 1997; Pallen and Wren, 1997). This element is a substrate recognition and binding element termed a PDZ domain. The name PDZ derives from the three proteins where this element was first recognised; Post synaptic density protein, Disc large protein, and Zo-1 protein all of which are eukaryotic proteins (Kennedy, 1995). The significance and functional characterisation of PDZ domains had previously been investigated prior to their discovery in bacterial proteins (for reviews see Fanning and Anderson, 1996; and Ponting *et al.*, 1997). PDZ domains are in general composed of eighty to one hundred amino acid residues. Their secondary structure contains five or six beta sheets and two alpha helices. They have been shown through crystallography studies to adapt a structure which allows the formation of a "carboxylate binding pocket". The primary sequence of PDZ domains are not well conserved, containing only a few strongly conserved residues. In contrast, the positional conservation of certain types of amino acids allows for a conserved overall tertiary structure. The PDZ domains of eukaryotic proteins have been shown to mediate binding to both relatively well conserved carboxyl terminal peptides and to motifs further removed from the extreme carboxyl termini of their target proteins. PDZ domains can also bind to the PDZ domains of other proteins allowing for the formation of homo- and hetero-oligomers. This feature together with their localisation at membrane sites permits them to act as scaffolding structures for the formation of signalling complexes between cells. The specificity for their target proteins is proposed to derive from the diverse amino acids present in the primary sequence that can partake in the formation of the binding pocket in PDZ domains.

1.5.4.2 Bacterial PDZ domains

The presence of PDZ domains in bacterial proteins is a relatively recent finding (Ponting, 1997). The laboratory of Sauer had identified the location of the domain conferring target specificity to tail specific proteases (Tsp) in *E. coli* (Keiler and Sauer, 1996). This domain

was subsequently shown to be a PDZ domain (Beebe *et al.*, 2000). Sauer's group also demonstrated that the recognition peptide sequences of Tsp could be more diverse than those from PDZ domain containing eukaryotic proteins, which in many cases consisted of the (T/S)XV carboxyl terminal motif. The SpoIVB protein of *B. subtilis* was one of the initial bacterial proteins shown to contain a PDZ domain (Ponting, 1997). This protein is a peptidase with a crucial role in the correct temporal induction of the activity of the mother cell specific sporulation sigma factor σ^K (Oke *et al.*, 1997; Wakeley *et al.*, 2000). Similar to both Tsp and eukaryotic PDZ domain containing proteins SpoIVB is initially localised to a membrane structure, the internal forespore membrane of the *B. subtilis* sporangium. Recent studies have demonstrated the importance of the PDZ domain in the functioning of SpoIVB. Site directed mutagenesis of some of the well conserved residues within the PDZ domain of SpoIVB caused delayed induction of σ^K controlled expression and reduced autoproteolysis of SpoIVB. The exact mechanism by which the PDZ domain contributes to the role of SpoIVB is still not known, but is speculated to involve sequential interactions with an inhibitor followed by self-interactions which may then invoke proteolytic activity and correct pro-σ^K processing and functioning (Hoa *et al.*, 2001).

1.5.4.3. PDZ domain of DegP

PDZ domains in Tsp and SpoIVB are located internally but are positioned at the carboxyl terminus of HtrA proteases. Additionally, the DegP protease of *E. coli* contains tandem PDZ domains a feature associated with most HtrA proteases of Gram-negative bacteria. The *E. coli* DegP protein is localised to the periplasmic face of the inner membrane which correlates with the membrane location of eukaryotic PDZ containing proteins. The importance of the PDZ domain for the activity of DegP was shown by both Sassoon *et al.* (1999) and Speiss *et al.* (1999). In the former, the production of deletion derivative of DegP containing the full proteolytic domain but truncated at the start of the PDZ domains conferred a temperature sensitive phenotype mimicking a DegP null strain. The production of both wild type and deleted forms together also produced temperature sensitive strains. This dominant mutant phenotype resulted from the assembly of non-functional heterodimers, contrasting with DegP which was shown to form homo-oligomeric hexamers. Likewise Speiss and co-

workers (1999) also showed that deletion of each PDZ domain individually or together had a drastic effect on thermosensitivity. Interestingly, these same authors were able to show that DegP had an intrinsic chaperone activity at lower temperatures and that this activity was not affected by removal of the PDZ domains of DegP. In contrast protease activity which had a dependence on higher temperatures was severely affected by deletion of the PDZ domains.

1.5.4.4. Human HtrA PDZ domains

The specificity of HtrA proteases for the carboxyl termini of target peptides has also been shown for Omi, a human HtrA protease. When the last amino acid of Mxi2, a known target protein for Omi, was changed the two proteins no longer interacted (Faccio *et al.*, 2000). Furthermore, the specificity of interactions between PDZ domains and target sequences was also highlighted by showing that the second human HtrA protease, L56, could not interact with Mxi2. With the determination of the crystal structure of HtrA2/Omi it has been shown that the PDZ domains of the monomers play an important role in regulating the activity of HtrA2/Omi (Li *et al.*, 2002).

1.5.5. Multimeric Structure of DegP and HtrA2(Omi).

A number of earlier studies with gel filtration assays suggested that native DegP formed an oligomeric structure (Kolmar *et al.*, 1996; Sassoon *et al.*, 1999). These studies disagreed as to the final number of subunits. Recently the crystal structures of both DegP and human HtrA2(Omi) have been solved. The DegP monomers form a hexameric molecule composed of two "tight" trimers (Krojer *et al.*, 2002) while HtrA2 monomers exist in a trimeric state (Li *et al.*, 2002) (Figure 1.1). In the DegP hexamer the PDZ domains are positioned equatorially and regulate access to the inner cavity formed by the protease domains. The structure of the PDZ domains are proposed to be flexible which would allow binding of substrate to be co-ordinated with translocation to the proteolytic cavity. This proposed mechanism is similar to the "anemone" model originally suggested by Pallen and Wren (1997). The overall proteolytic mechanism employed by DegP is similar to many of the cytosolic systems (e.g.,. Clp and FtsH) whereby the substrates must enter a central chamber for degradation to occur. However, in contrast to the multimeric cytosolic proteases whose

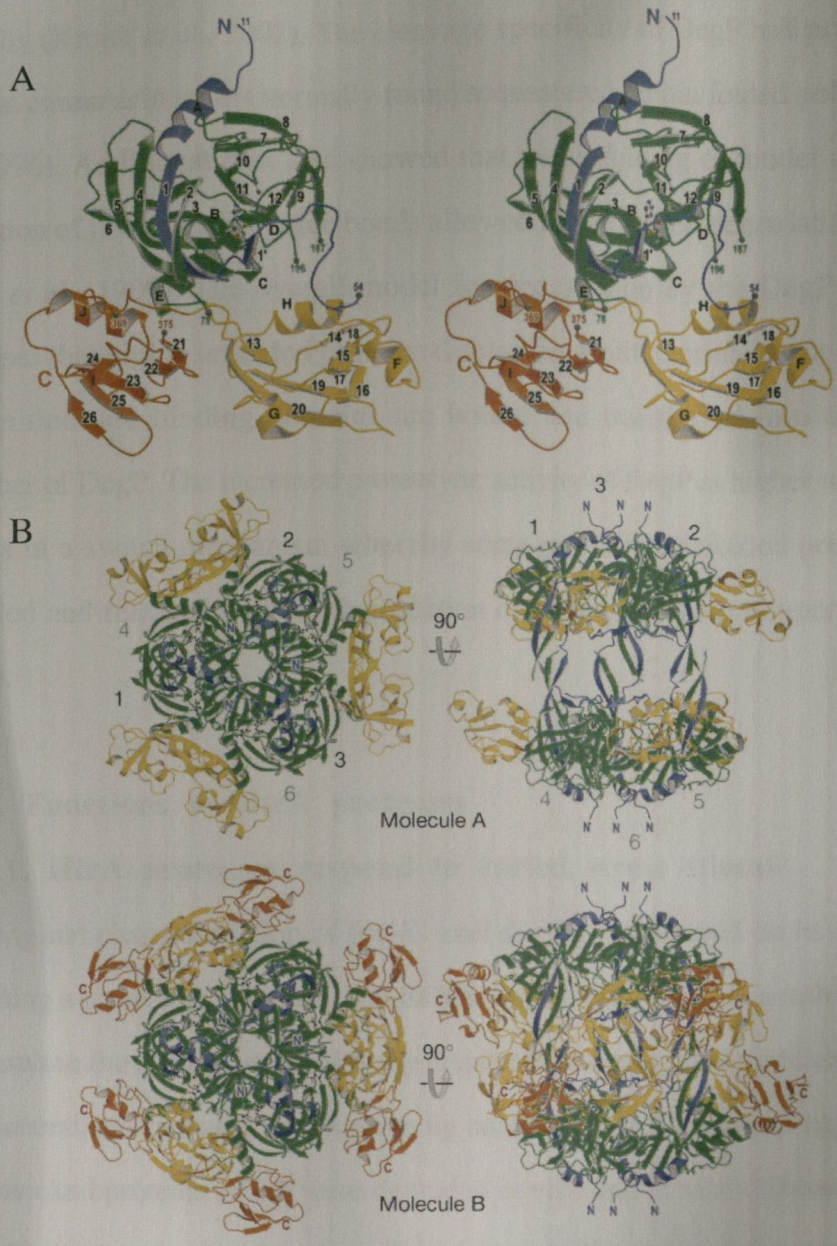


Figure 1.1. The structure of DegP from *E. coli* revealed by crystallography. (A) The DegP monomer is composed of a protease domain, green; and two PDZ domains, PDZ1 yellow; PDZ2, red. The N-terminal domain, purple, of the mature protein forms part of the protease domain. Secondary structural features are indicated by letters for α–helices and numbers for β–strands. (B) The DegP hexamer. Two different forms of the hexamer were observed. An open confirmation molecule "A", middle panel, in which the PDZ2 domain was not defined and a closed confirmation molecule"B" defining both PDZ domains, lower panels. The structures on the right are rotated by 90 degrees around the horizontal axis. This figure is from Krojer *et al.* (2002).

entrances to the core are positioned apically, the gateway to the DegP chamber is positioned laterally (Krojer *et al.*, 2002). The cleavage specificity of DegP had previously been shown to be at amino acid motifs normally found sequestered within folded polypeptides (Kolmar *et al.*, 1996). Additional data also showed that the unfolding of model or native proteins by reduction of intrinsic disulphide bonds allowed for increased degradation of these substrates (Kim *et al.*, 1999). The overall model for degradation by the DegP protease of *E. coli* proposes that highly unfolded proteins displaying motifs normally not available to the PDZ recognition and binding domains are bound and transferred into the inner proteolytic chamber of DegP. The increased proteolytic activity of DegP at higher temperatures probably results in a switch mechanism whereby some partially misfolded proteins can be bound, refolded and released without degradation occurring at lower temperatures (Spiess *et al.*, 1999).

1.5.6. Functions of HtrA proteases

1.5.6.1. HtrA proteases respond to varied stress stimuli

The original characterisation of the *E. coli degP* gene focused on its induction and role in providing a protective response against heat stress. Heat stress is a rather broad stimulus and can result in the accumulation of many denatured or aggregated proteins. The role of DegP in this generalised response was shown by analysing the amount of aggregated endogenous heat shocked proteins which were degraded *in vivo* and *in vitro*. These experiments showed a prominent role for DegP in degrading aggregated proteins and suggested a mechanism of recognition and proteolysis involving generally exposed residues rather than specific recognition motifs (Laskowska *et al.*, 1996).

Following on from the general role of DegP in degrading misfolded proteins, attention focused on the search for specific substrates. To date only a limited number of DegP substrates have been identified, the first of which was Colicin A lysis protein (Cavard *et al.*, 1989). Other substrates have since been identified, e.g. MalS and MalF when overproduced without its natural protein partner (Mourez *et al.*, 1997) and most recently the PapA subunit of P-pili (Jones *et al.*, 2002). Given its cellular location in the periplasm, many workers

utilised *degP-lacZ* fusions as indicators of extracytoplasmic stress. It is safe to say it has been one of the most exploited fusions for this purpose, mainly because of its control by dual regulatory systems (see sections 1.5.7. and 1.5.8) . Many different stimuli have now been shown to induced the transcription of *degP*, such as heat shock (Lipinska *et al.*, 1988), ethanol addition, alkali shock (Danese and Silhavy, 1998), oxidative stress (Skorko-Glonek *et al.*, 1999), overexpression of outer membrane proteins (Mecsas *et al.*, 1993), perturbation of membrane lipid synthesis (Danese *et al.*, 1998), imbalanced synthesis of receptor proteins (Jones *et al.*, 1997) and the mutation of most genes encoding periplasmic or outer membrane folding agents (see review by Raivio and Silhavy, 2000) The obvious result of many of these stimuli and mutations is to increase the amount of periplasmic misfolding and aggregation and fits with the role of DegP to degrade or refold the damaged protein(s). The role of DegP in chaperoning is a recently discovered one and has been shown both *in vitro* and *in vivo* (Spiess *et al.*, 1999).

1.5.6.2. The role of DegP paralogues

Despite its role in combating extracytoplasmic stress, DegP is only essential when *E. coli* are grown at high temperatures (> 42°C). In contrast DegS a paralogue of DegP, is essential. DegS is also a periplasmic protease but may be bound to the periplasmic face of the inner membrane as it does not have a consensus signal peptide cleavage site and co-fractionates with inner membranes. The role of DegS was recently elucidated by the laboratory of Carol Gross (Alba *et al.*, 2001). DegS enables σ^E to be released from its physical interaction with RseA, an anti-sigma factor binding to σ^E at the inner membrane. DegS performs this role through proteolysis since serine active site mutations of DegS prevent proper induction of σ^E controlled genes. The essential function of both DegS and σ^E can be overcome by accumulation of suppressor(s) although the nature or identity of these has not been identified (De Las Penas *et al.*, 1997a; Ades *et al.*, 1999).

DegQ, a third HtrA-type *E. coli* protease, can suppress the temperature sensitive growth phenotype of *degP* mutants (Waller and Sauer, 1996). The DegQ protease is very similar to DegP as it localises to the periplasmic space and also contains a second tandem PDZ domain

at its C-terminus unlike DegS which atypically for a Gram-negative HtrA protease, contains only one PDZ domain.

1.5.6.3. Pathogenicity Functions of other Bacterial HtrA proteases

The HtrA proteases are a rapidly expanding family of proteins, due in part to the increasing number of bacterial genomes that are being sequenced. In many cases the characterisation of HtrA proteases has begun with an investigation of their expression and in particular their response to stress stimuli. The majority of HtrA proteases are responsive to stress stimuli, in particular to heat shock and oxidising agents although there are exceptions. One of the interesting phenotypes originally associated with mutation of HtrA encoding genes was the reduced ability of *Salmonella typhimurium* to survive within macrophages (Johnson *et al.*, 1991). This phenotype also correlated with reduced resistance to oxidising agents and may indicate that intracellular *S. typhimurium* are more prone to the oxidative burst utilised by macrophages as a means of killing engulfed bacteria. Since this initial finding linking reduced virulence or intracellular survival to HtrA mutants, the same phenotype of reduced virulence has been observed in other bacterial pathogens bearing mutations in *htrA* encoding genes such as *Legionella pneumoniae* (Pedersen *et al.*, 2001), *Yersinia enterocolitica* (Li *et al.*, 1996), *Streptococcus pyogenes* (Jones *et al.*, 2001)., *Klebsiella pneumoniae* (Cortes *et al.*, 2002) and *Streptococcus pneumoniae* (Sebert *et al.*, 2002). Reduced virulence was originally reported for *htrA* mutants of *Brucella abortus* and *Brucella melitensis* in calves and goats respectively (Phillips *et al.*, 1995; Elzer *et al.*, 1996). The significance of these findings has now been downplayed as the original strains contained a second mutation which has been shown to have been primarily responsible for the reduced virulence (Phillips and Roop, 2001). Nevertheless the newly generated *htrA* mutants are still less resistant to cultured host defence cells and to oxidising agents *in vitro*.

The characterisation of an *htrA* mutant of the industrially important Gram-positive bacterium *Lactococcus lactis* has shown that a variety of model substrates were reliant on the single HtrA protease contained by this bacterium for degradation and/or processing (Poquet *et al.*, 2000). These same authors also predict that the HtrA protease may have a role in processing

of secreted virulence factors and this may account, along with reduced oxidative stress resistance, for the diminished virulence of some pathogenic bacteria bearing *htrA* mutations.

1.5.6.4. Functions of Eukaryotic HtrA proteases

In eukaryotes the same theme of a role in stress responses has also been demonstrated in both human and plant HtrA proteases. Up to thirteen HtrA-like proteases are encoded in the *A. thaliana* genome (Adam *et al.*, 2001). Two of these have been experimentally studied. The synthesis of DegP1 was shown to be elevated in response to heat shock (Itzhaki *et al.*, 1998) while the second HtrA protease, DegP2, was shown to be responsible for degradation of light induced damaged D1 reaction centres, a very important part of the photosynthetic complex in thylakoid membranes (Haussuhl *et al.*, 2001).

In humans at least two different HtrA proteases have been identified and studied. L56 (humHtrA1) has been shown to be upregulated in osteoarthritic cartilage (Hu *et al.*, 1998). A second human HtrA protease Omi (humHtrA2) has been shown to have increased expression when a stress stimulus was applied to a mouse kidney model system *in vivo* (Faccio *et al.*, 2000). Recently, Omi has been identified as providing an additional component in the regulation of apoptosis in mammalian cells. In various screens to isolate proteins interacting with inhibitors of apoptosis proteins (IAPs) Omi (HtrA2) has been identified as one such protein and its mode of action characterised (Suzuki *et al.*, 2001; Martins *et al.*, 2002; Verhagen *et al.*, 2002). When cells are stimulated to undergo apoptosis it has been shown that Omi is released from the mitochondria into the cytoplasm and interacts with X-linked inhibitor of apoptosis protein (XIAP). This interaction occurs through sequences at the extreme N-terminus of mature processed Omi. Furthermore, the pro-apoptotic role of Omi requires a functional serine protease activity. It has also been shown that overexpression of Omi outside the mitochondria can lead to caspase-independent atypical cell death. This finding has been questioned by other studies suggesting it may derive purely from the amounts of Omi produced as moderate levels of mitochondrially produced Omi did not lead to cell death but did sensitise cells to apoptotic stimuli (Martins *et al.*, 2002).

1.5.6.5. Biotechnological and Biomedical Applications of HtrA proteases

Some of the possible applications of HtrA proteases can be inferred from its dual role as both a protease and a chaperone. *E. coli* strains mutated for DegP in combination with other extracellular proteases have already been constructed (Meerman and Georgiou, 1994). These strains have shown improved yields of secreted overproduced proteins. This result confirms the stress responsive role whereby the degradative function of DegP is employed to remove increased amounts of heterologous proteins in the cell envelope. In contrast, co-expression of DegP in strains producing native envelope proteins has led to increased amounts of functional overexpressed protein by reducing the formation of inclusion bodies in the periplasm, demonstrating the advantageous chaperone activity of DegP (Lin *et al.*, 2001). As cited above, the surface localised HtrA of *L. lactis* has been shown to influence the production and degradation of a number of heterologous secreted proteins (see section 1.5.6.3. Poquet *et al.*, 2000). Further studies have confirmed this earlier finding and have shown dramatically improved production of other viral and bacterial proteins from a controlled promoter in a *htrA* mutant strain of *L. lactis* (Miyoshi *et al.*, 2002).

The use of live attenuated *Salmonella* strains mutated in *htrA* and *aroCD* (imparts auxothrophy for aromatic amino acids) has proved to be a useful and effective delivery vector for the production of multiple protective antigens (Levine *et al.*, 1996). The mutation of *htrA* encoding genes in other pathogenic bacteria and their use as vaccines has also provided varying degrees of protective immunity when the immunised animals are challenged with the respective virulent strains (see above). Given their strong association with pathogenicity, the HtrA proteases could also be targeted for rational antimicrobial drug design. The relatively high degree of non sequence conservation potentially confers upon PDZ domains a high degree of substrate specificity. This may allow for specific peptides to be designed and targeted at specific PDZ recognition domains without affecting host proteases or beneficial flora.

1.5.7. Regulation of the expression of HtrA proteases

The HtrA proteases are either localised to the periplasm, as in Gram-negative bacteria or are putatively membrane bound with their catalytic domains positioned extra-cytoplasmically as in Gram-positive bacteria. Therefore, they are positioned to respond to protein misfolding in these extracytoplasmic compartments. However, the presence of the cytoplasmic membrane requires a signal transduction pathway to sense the stress and convey this signal across the membrane to regulate increased HtrA expression appropriately. The elucidation of the regulatory mechanisms to control synthesis of HtrA proteases has focused largely in *E. coli*. It is now known that at least two regulatory systems control DegP production; the alternate heat shock sigma factor σ^E and the CpxR/CpxA two component system (see Figures 1.2 and 1.3.). There is tentative evidence for the involvement of a third possible system (Connolly *et al.*, 1997). While many other organisms have been shown to possess a stress responsive HtrA protease, the regulatory mechanisms responsible have not for the most part been identified.

1.5.7.1 Regulation of DegP by the σ^E alternate ECF-type sigma factor.

The identification of a second heat shock response came about from investigations into the transcriptional analysis of *rpoH*, encoding the cytoplasmic heat shock sigma factor σ^{32}. These studies showed that *rpoH* produced four transcripts, three of which were universal among different strains and the fourth was strain specific. Under normal growth conditions promoters P1 and P4 are responsible for the basal level of transcription. Expression from the third promoter P3 is activated upon heat shock to 42°C. The induction observed was weak as the primary mechanism for induction of the cytoplasmic heat shock response had previously been shown to result from post-transcriptional mechanisms to increase the levels and stability of σ^{32} (see section 1.3). However when *E. coli* cells were heat shocked to 50°C the level of the P3 transcript increased more dramatically while the P1 and P4 transcript levels decreased. *In vitro* analysis confirmed that the P1 and P4 transcripts depended upon the σ^{70} (σ^A) vegetative sigma factor for activation but this was not the case for the P3 transcript which was not activated by σ^{70} or σ^{32}. These experiments suggested the existence of a second heat shock factor (Erickson *et al.*, 1987).

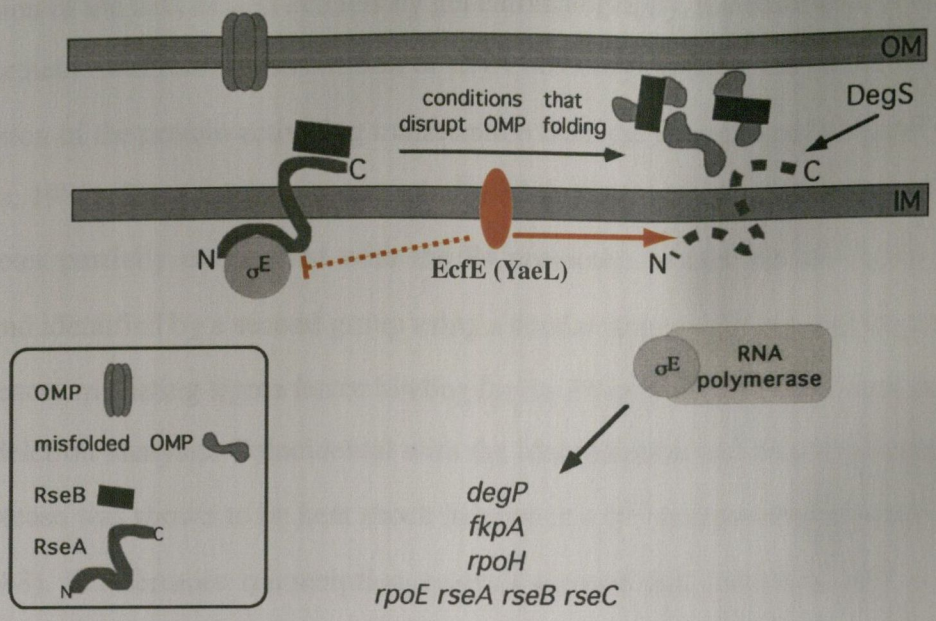


Figure 1.2. Regulation of the σ^E response in *E. coli.* During normal physiological conditions σ^E is prevented from directing transcription by its association with the anti-sigma factor RseA. During conditions which lead to increased amounts of protein unfolding or aggregation a signal is generated and sensed which leads to the sequential proteolytic degradation of RseA by DegS and EcfE(YaeL). DegS makes an initial cleavage upon the periplasmic domain of RseA followed by a second proteolytic attack by EcfE at cytoplasmic and/or integral membrane sites. The degradation of RseA frees σ^E to direct transcription of its regulon genes and effect an appropriate response to alleviate extracytoplasmic stress. The newest elements in the regulation of σ^E are indicated in red, the dashed line indicates a questionable negative role for EcfE in degrading σ^E (see section 1.5.7.4.). OMP, outer membrane protein; OM, outer membrane; IM, inner membrane. This diagram is adapted from Raivio and Silhavy (2001).

Using crude extracts from cells heat shocked to 50°C and extracts from a strain of *E. coli* containing a temperature sensitive allele of the *rpoD* gene, encoding σ^{70}, the laboratory of Gross confirmed the existence of the P3 activating factor (Erickson *et al.*, 1987). Subsequent fractionation of the heat treated extracts by gel chromatography, isolation and renaturation of specific protein bands and reconstitution of RNA holoenzyme ($E\sigma$) led to the isolation and identification of the protein activating transcription from the P3 promoter of *rpoH* (Erickson and Gross, 1989). The protein was named σ^E and further *in vitro* analysis showed that the P3 promoter partially overlapped with the P4 promoter. In fact the σ^E factor was also isolated and identified by a second group using a similar strategy (Wang and Kaguni, 1989). The sequences mediating sigma factor binding for the P4 and P3 promoters were determined through deletion analysis. Coincidental with the identification and characterisation of σ^E, DegP protease was shown to be heat shock induced in a σ^{32} independent manner (Lipinska *et al.*, 1988). Furthermore transcription analysis showed that induction still occurred at 50°C. A comparison of the P3 promoter regions of *rpoH* and the *degP* promoter showed both had almost identical motifs upstream of the respective transcriptional start points. *In vitro* analysis with reconstituted $E\sigma^E$ proved that σ^E was responsible for the activation of *degP* transcription (Erickson and Gross, 1989). It was also noted by the same authors that the *dagA* and *phl* genes, encoding an agarase and a phospholipase from *Streptococcus coelicolor* and *Serratia liquefaciens* respectively, had similar motifs within the promoter regions. This suggested that the genes controlled by σ^E or its homologues in other bacteria would be involved in the response to extracytoplasmic stress.

1.5.7.2. Stress induction of σ^E

A study by Mecsas *et al.* was undertaken to identify genes encoding proteins that affect expression of σ^E-dependent promoters (Mecsas *et al.*, 1993). Genetic screens employing two differing approaches were carried out to identify signals affecting activation of σ^E. Both strategies utilised the σ^E-dependent *rpoH* P3 promoter. The first involved the generation of a multicopy plasmid library which would increase expression of *rpoH* P3-*cat* fusions and thereby allow growth in the presence of increased levels of chloramphenicol. The second strategy used the same *rpoH* P3 promoter fused to *galK*, producing phenotypically distinct

red colonies in a wild type background when grown on indicator media. The multicopy approach showed that increased levels of outer membrane proteins (OMPs) increased expression of the σ^E-dependent *rpoH* P3 promoter. The transposon experiments indicated that mutation of the response regulator OmpR decreased expression of the same promoter. These experiments indicated that the level of OMPs influenced the activity of σ^E. The exact nature of the inducing signals was identified more precisely in further studies. Constructs which caused accumulation or misfolding of periplasmic or OMPs in different subcellular compartments were tested for analyse their effect on σ^E activity. This series of experiments demonstrated that the accumulation or misfolding of outer membrane proteins in the periplasm had the greatest effect, while misfolding or overexpression of periplasmic proteins in the periplasm had no affect. In addition a gene encoding a periplasmic disulphide bond catalyst, *dsbA,* had a minor affect. Overall the data showed that the activating signal for σ^E was specific both for outer membranes proteins and for a specific location (i.e. the periplasm).

1.5.7.3. Signal transduction mechanism of the σ^E stress response

Two laboratories simultaneously identified and cloned the gene encoding the σ^E sigma factor. The laboratory of Georgopoulos used two genetic screens utilising *rpoH*P3 and *degP* promoter fusions to *lacZ* to clone the *rpoE* gene (Raina *et al.*, 1995). The laboratory of Gross relied on a bioinformatical analysis of putative open reading frames from database entries to match amino acid sequences of tryptic fragments of the previously isolated σ^E polypeptide (Rouviere *et al.*, 1995). After overexpression and isolation of the *rpoE* gene product, the 24 kDa protein was able to direct transcription from both the *rpoH*P3 and *degP* promoters *in vitro*, confirming its identity and activity with the *in vivo* isolated protein (Raina *et al.*, 1995; Rouviere *et al.*, 1995).

Transcriptional analysis of the *rpoE* gene showed that two transcripts, P1 and P2 were synthesised from the *rpoE* promoter region *in vivo*. Mapping of the transcriptional start points showed that the sequences upstream of the P2 transcript matched closely those of *rpoH* and *degP* σ^E dependent promoters suggesting possible autoregulation. Using *rpoE*

promoter fragments *in vitro* analysis confirmed that reconstituted $E\sigma^E$ generated the P2 transcript but not the P1 transcript. Furthermore, the P2 transcript was missing from RNA isolated from a *rpoE* mutant strain. The *rpoE* gene was therefore shown to be autoregulated by its own gene product and became the third member of the σ^E regulon (Raina *et al.*, 1995; Rouviere *et al.*, 1995). The activity of the *rpoE* promoter was analysed to determine if the increased activity of σ^E dependent promoters, *rpoH*P3 and P*degP*, was reflected by increased transcriptional activation of *rpoE*. It was found that under heat shock conditions only a small increase in transcription of *rpoE* was observed. However under non heat shock conditions in the presence of overexpressed OMPs the level of σ^E activity was correlated to increased transcription. This indicated a possible post-transcriptional activation of σ^E occurring after heat shock (Rouviere *et al.*, 1995).

A search was undertaken for factors that function to limit the positive autoregulatory nature of σ^E expression. Sequence analysis of a AlgU, a σ^E homologue, from *Pseudomonas aeruginosa*, indicated that the gene encoding this sigma factor was cotranscribed with a gene encoding an anti-sigma factor MucA (Schurr *et al.*, 1996). This was also found to be the case for σ^E as *rpoE* was contained within a four gene operon (De Las Penas *et al.*, 1997b). Mutation of the *rpoE* operon downstream genes, termed *rseABC*, for regulators of sigma-E, showed that the mutation of *rseA* resulted in a large increase in σ^E activity under non stressed conditions as judged by induction of a *rpoH*P3-*lacZ* fusion (De Las Penas *et al.*, 1997b). In contrast, mutations in *rseB* had only a minor effect and mutations in *rseC* had no effect. When these same strains were assayed under conditions of OMP overexpression, the *rpoH*P3 fusions were inducible in the *rseB* and *rseC* backgrounds but no further induction of the existing high level of activity was observed in the *rseA* mutant background. This showed that *rseA* was a negative regulator of $E\sigma^E$ activity and was necessary for signal transduction to activate $E\sigma^E$ activity (De Las Penas *et al.*, 1997b). The RseA protein was predicted to be a membrane protein with a cytoplasmic N-terminal domain and a periplasmic C-terminal domain. *In vitro* analysis with purified truncated RseA containing only the cytoplasmic N terminal domain showed it bound σ^E preventing transcription from reconstituted $E\sigma^E$ but not from $E\sigma^{70}$. These results were also reproduced *in vivo* and demonstrated that RseA is an

anti-sigma factor. In addition, RseB was shown to bind to the periplasmic domain of RseA which fitted with RseB's predicted cellular location and suggested it may positively modulate affinity of RseA for σ^E (De Las Penas *et al.*, 1997b).

Since RseA was shown to be the major determining factor controlling $E\sigma^E$ activity following stress this implied that there must be a system by which the negative effect of the RseA-$E\sigma^E$ interaction should be relived to allow induction of the σ^E regulon. One of the possibilities explored was that of proteolysis of the RseA molecule following stress sensing. This idea was prompted by a similar role for proteolysis in controlling key cytoplasmic regulators such as σ^{32} and σ^S. The first indication that such a system might exist was indicated when the level of RseA was shown to be drastically reduced following overexpression of OMPs while the level of σ^E at the same time increased (Ades *et al.*, 1999). Since the rate of synthesis was found to be equal for both proteins under these conditions this implied the level of RseA was being decreased rapidly. Pulse chase experiments showed there was an eight to fourteen fold decrease in the half life of RseA depending on the stress used, while the stability of σ^E remained unchanged. A similar decrease in the half life of RseA was also seen in mutant backgrounds which increased $E\sigma^E$ activity and lastly the mutation of *rseB* produced a small decrease in RseA correlating with the minor increase in σ^E dependent activity (Ades *et al.*, 1999).

Candidate extracellular proteases responsible for the reduced amount of RseA included DegP, OmpT, DegQ and DegS. No effect on the activity of $E\sigma^E$ was observed in strains mutated for the first of these three proteases. However, the activity of $E\sigma^E$ was dramatically reduced in a *degS* mutant strain while wild type activity levels were restored in cells complemented by a plasmid carrying the *degS* gene (Alba *et al.*, 2001). Further evidence for the proteolytic role of DegS in degrading RseA was shown in cells mutated for *degS* but containing an active site mutation (S210A) in the complementing clone, in this case wild type activity was not restored. The requirement for DegS were also observed under stress conditions as no induction occurred without a wild type *degS* gene being present. The role of DegS in activating the σ^E dependent response was further strengthened by demonstrating the

topology of DegS as a membrane bound protease with its active site in the periplasm was crucial. When DegS devoid of its membrane anchor was overproduced in a *degS* mutant background only a fraction of wild type σ^E-dependent activity was regained (Alba *et al.*, 2001).

The importance of proteolysis in reducing the amounts of RseA binding to σ^E and inducing the σ^E dependent stress response has been disputed recently. The laboratory of Missiakas have produced data showing the level and half life of RseA following stress are not reduced to the extent reported by Alba and co-workers (Alba *et al.*, 2001). Instead these authors have shown that complexes are formed by σ^E-RseA-RseB and that the affinity of RseA for σ^E is greatly increased in the presence RseB. An alternative model involving a titration of RseB from the ternary complex by misfolded periplasmic proteins and hence a lessening of the affinity of RseA for σ^E is proposed to account for the activation of the Eσ^E dependent regulon (Collinet *et al.*, 2000).

1.5.7.4 σ^E regulon members and inducing signals

Initial characterisation of signals or stresses which induced Eσ^E activity led to the proposal that the regulon responded to misfolded proteins in the periplasm. Both heat shock and the overexpression of outer membrane proteins had shown this but some selectivity in the signal had also been suggested because overexpression of periplasmic proteins did not induce the σ^E response (Mecsas *et al.*, 1993). The use of *rpoH*P3 or *degP* promoter fusions to *lacZ* were used to screen for the isolation and characterisation of proteins involved in protein folding in the periplasm or outer membrane. These data showed that the mutation of genes encoding protein folding agents and catalysts resulted in significant increases in the activity of Eσ^E. The proteins identified included disulphide bond oxido-reductases DsbA and DsbC, peptidyl-prolyl isomerases SurA and FkpA, and an outer membrane protein OmpH (Skp) (Missiakas *et al.*, 1996). However, it was also noted that deletion of *degP* did not change Eσ^E activity (Mecsas *et al.*, 1993).

Overexpression of σ^E led to the induction of at least ten proteins (Raina *et al.*, 1995). However the known genes of the σ^E regulon were until very recently limited to those which had been characterised during the initial work on σ^E. In a twin approach to estimating the size and identifying the members of the σ^E regulon, 2-D gel electrophoresis and reporter fusion analysis was undertaken. These experiments identified at least 20 new promoters shown to respond to signals which activate σ^E (Dartigalongue *et al.*, 2001b). Transcriptional analysis by primer extension identified motifs upstream of the deduced start sites which were similar to the consensus sequence for σ^E-dependent activation. The function of the gene products identified are wide ranging, from transcription factors σ^{32}, σ^E and σ^{70}, to folding factors (see above) which were shown to be σ^E regulon members and proteins with roles in lipopolysaccharide synthesis. Some of the genes were contained within operons having both upstream and internal σ^E controlled dual promoters.

The *rpoE* gene has been described as essential, but much analysis of σ^E mutant strains and regulon had taken place prior to this proposal. This was possible because of suppressor mutations which allowed for apparently normal growth at physiological temperatures although the frequency and nature of these suppressors appears to be strain specific (De Las Penas *et al.*, 1997a). To try and identify which of the possible σ^E regulon genes might be the cause of the regulons essentiality, systematic mutation of the newly identified genes (*ecf* genes) was undertaken. All but two of the *ecf* genes were mutable, the *ecfE* and *ecfL* genes could not be mutated unless complementing plasmids carrying the respective genes were first introduced. The precise function of their gene products is not known, both are predicted to be inner membrane proteins with EcfE a possible protease (Dartigalongue *et al.*, 2001b). EcfE is suggested to be involved in regulating both the σ^{32} and σ^E regulons through direct proteolytic degradation of these sigma factors (Dartigalongue *et al.*, 2001a). Contrary to this finding two independent studies have very recently indicated that EcfE(YaeL) is required to undertake a sequential proteolytic processing step upon RseA (after the initial DegS-dependent step) in order to activate σ^E (see Figure 1.2.) (Alba *et al.*, 2002; Kanehara *et al.*, 2002).

1.5.8. Identification of the Cpx system in controlling an envelope stress response

Investigations in the laboratory of Silhavy into the mechanisms suppressing the toxicity conferred by model envelope proteins led to the identification of a second regulatory component controlling expression of *degP* (Cosma *et al.*, 1995; Danese *et al.*, 1995). Three different model proteins were used in these analyses; a tripartite protein LamB-LacZ-PhoA which forms large periplasmic aggregates, LamBA23D defective in signal sequence processing in the periplasm and LamB-LacZ a protein which generates toxicity at the secretion apparatus. Two different approaches were used in the screens for suppressors of theses toxic proteins. The first involved growth of strains expressing the proteins on medium known to mediate toxicity of the model proteins. Strains which were able to survive this increased toxicity were chosen and the loci conferring the suppression mapped to the Cpx locus. Similarly, a screen for multicopy suppressors of the toxicity of the LamB-LacZ-PhoA model protein identified the Cpx locus as the mechanism of suppression. It was shown that overexpression of a newly identified lipoprotein, NlpE, relieved toxicity and that this relief was dependent on the presence of an intact Cpx system (Snyder *et al.*, 1995).

Cpx mediated suppression was accompanied by a dramatic decrease in the levels of LamB-LacZ-PhoA, a first indication of the involvement of DegP (Cosma *et al.*, 1995). The levels of transcription from a *degP-lacZ* fusion were then measured in the different Cpx suppressors identified. Increased levels of transcription occurred and the relative amounts reflected the strength of the different Cpx suppressors (Danese *et al.*, 1995). Mutation of *degP* resulted in the stability of LamB-LacZ-PhoA protein. In addition, the presence of an intact *degP* gene was required for full suppression of the toxicity resulting from LamB-LacZ-PhoA. Therefore, it appeared that the mechanism of suppression relied largely upon DegP mediated proteolysis (Cosma *et al.*, 1995).

In contrast to the situation for LamB-LacZ-PhoA, the stability of LamBA23D was not affected by a *degP* mutation in the suppressor strains. Furthermore, overexpression of DegP resulted in stabilisation of LamBA23D. (Cosma *et al.*, 1995). This was possibly in

retrospect an early indication of DegP's chaperone role. However, the effectiveness of the suppression was greatest in strains containing a wild type *degP* locus.

An unexpected result came about from the use of the third model protein LamB-LacZ. The same Cpx suppressors also abolished the toxicity of this protein and were seen to result from a degradation of the fusion protein. The degree of suppression was also allied to the presence of an intact *degP* allele. Since the LamB-LacZ protein affected the secretion machinery this result suggested that the Cpx suppressors enabled the transport of the toxic protein to the periplasm and subsequent degradation by DegP. Overall, the results showed an important but not essential role for DegP in the mechanism of suppression and implied that the Cpx suppressors controlled other factors affecting protein folding and degradation.

The results with the mutant proteins showed that *degP* had two regulators. To analyse what relationship if any existed between them, experiments were undertaken to see if the overexpression of the lipoprotein NlpE also induced the σ^E regulon. This showed that the *rpoH*P3 σ^E -dependent promoter was not induced by NlpE overexpression and that expression of other lipoproteins did not induce *degP* through the Cpx system. Furthermore the transcriptional induction of *degP* by NlpE overexpression was greatly reduced in an *rpoE* mutant background showing that much of the induction of *degP* by the Cpx suppressors occurred through association with σ^E. However induction was still observed in the *rpoE* mutant strain implying other factors could promote Cpx mediated induction of *degP*. These results showed that separate signals from a common cellular location could effect transcriptional induction of a common target *degP* through two separate systems (Danese *et al.*, 1995).

A role for a third regulatory system in the expression of DegP has also been reported as residual expression of *degP* remained even in a *rpoE cpx* double mutant strain (Connolly *et al.*, 1997).

1.5.8.1. The CpxRA two-component system

The suppressor mutations allowing growth of the strains in the presence of the toxic model proteins mapped to the previously identified Cpx locus. This locus had previously been identified as the locus in which mutations affecting **c**onjugative **p**lasmid e**x**pression mapped. It was also noted that mutations in the two Cpx genes *cpxA* and *cpxR* affected both inner and outer membrane composition, discussed in (Danese *et al.*, 1995). Subsequent work had identified the two Cpx proteins as members of the two component signal transduction system family (Dong *et al.*, 1993).

Two component regulatory systems are used by bacteria to sense changes, mostly of a stressful nature, in their environment. This information is then transmitted through the cell membrane to effect rapid transcriptional induction of genes encoding proteins which allow the cell to cope with the emerging stress. The central components of bacterial two component systems have now been shown to be highly conserved amongst a wide variety of bacterial species (Hoch and Silhavy, 1995). As their name suggests two component systems are composed of two proteins, a membrane bound sensor histidine protein kinase, (HK), and a cytoplasmic protein acting as a transcriptional regulator called a response regulator, (RR). The sensor component binds or senses the signal through its sensor domain resulting in autophosphorylation of a histidine residue in its phosphotransferase domain. The phosphotransferase domain then transfers the phosphoryl group to an aspartic acid residue in the input domain of the response regulator. This event induces a conformational change in the RR, allowing the RR to bind to its target promoters resulting in altered transcription patterns. The HK phosphotransferase domain in some cases has phosphatase activity by which it can dephosphorylate the RR~P to reduce transcriptional activity following adaptation to the stress.

The Silhavy group hypothesised that the increase in the transcription of the *degP* gene most probably resulted from alterations of the sensor kinase CpxA (Danese *et al.*, 1995). The group of Jon Beckwith was able to demonstrate by DNase footprinting that the CpxR protein bound to the *degP* promoter fragment *in vitro* (Pogliano *et al.*, 1997). More importantly it

was also shown that the phosphorylated form of CpxR~P bound with greater efficiency (at higher DNA to protein ratios) to the *degP* promoter. The sequence of two regions to which CpxR~P protein bound were determined and found to contain a repeated GTAA motif (Pogliano *et al.*, 1997).

A similar and perhaps more comprehensive *in vivo* and *in vitro* biochemical analysis was also undertaken by Silhavy's group. These analyses were carried out with plasmids overexpressing fusion proteins between a cytoplasmic derivative of maltose binding protein to CpxA and CpxR. Overexpression of the MBP-CpxR fusion in a *cpxR* mutant restored transcription of *degP* and the purified MBP-CpxR protein also showed binding in gel shift assays to a *degP* promoter fragment *in vitro*. When phosphorylated the MBP-CpxR~P caused enhanced supershifting of the *degP* promoter fragment confirming the results of Pogliano and co-workers (Raivio and Silhavy, 1997).

The genes encoding the CpxA and CpxR proteins from the suppressor strains (designated Cpx*) were sequenced and found to contain various missense and deletion mutations (Raivio and Silhavy, 1997). The missense mutations clustered around (i) the second transmembrane domain (TM) preceding the phosphotranferase domain, (ii) the histidine autophosphorylatable residue and (iii) the periplasmic domain of CpxA. Further work with the MBP fusions to wild type, CpxA* and CpxR* proteins showed that MBP-CpxA* had the normal autokinase activity and also the ability to kinase CpxR *in vitro*. However, a MBP-CpxA* protein did not possess the correct phosphatase activity to dephosphorylate CpxR~P and thereby downregulate CpxR activity. This was the probable reason for increased amounts of DegP and other unidentified folding agents. Some of the Cpx* suppressors were shown to confer constitutive *degP* transcription. The genes encoding this class of suppressors were found to contain mutations in the periplasmic domains of CpxA*. When these alleles were tested for their ability to induce *degP* by overproduction of NlpE, the Cpx activator, no further induction occurred. It was hypothesised that the periplasmic loop of the CpxA HK was necessary for sensing and transmitting a signal to switch off the response by possibly binding a factor from the periplasm. Under normal conditions wild type CpxA can

do this but the CpxA* cannot. Thus in some Cpx* mutants CpxA* is constitutively active leading to a relief of the toxicity caused by the model proteins, mediated by increased levels of DegP and/or other Cpx regulon members.

1.5.8.2. Regulation of the Cpx response

Increasing interest in the Cpx regulon led to the identification of other members in addition to *degP*. Some of these investigations used a candidate gene approach (Pogliano *et al.*, 1997) to see if newly identified proteins with protein folding or isomerase activities were part of the Cpx regulon while others relied on overexpression of the regulators CpxR or σ^E (Danese and Silhavy, 1997). A more classical approach utilising transposon generated random *lacZ* fusions found the gene *cpxP* encoding a periplasmic protein with a canonical signal peptide sequence (Danese and Silhavy, 1998). *cpxP* was shown to be regulated exclusively by the Cpx system in contrast to the dual regulation of *degP* by Cpx and σ^E. Neither deletion of *rpoE* nor overexpression of OMPs affected expression of *cpxP*. Fusion of a signal sequenceless PhoA to CpxP directed export of PhoA to the periplasm. Deletion of *cpxP* led to increased sensitivity, relative to wild type, of strains producing a toxic protein even when the Cpx response was induced by NlpE overexpression. This implied a role for CpxP in dealing with extracytoplasmic stress, while the location of *cpxP*, divergently transcribed from the *cpxRA* operon, was a further indicator of its possible role in regulating Cpx function (Danese and Silhavy, 1998).

A Cpx binding site had been deduced from DNase footprinting analyses of *degP* and other Cpx controlled promoters. It was also noted that the upstream region of the *cpxRA* operon also contained a possible Cpx binding site. This suggested a possible autoregulatory loop for the Cpx system to induce transcription of *cpxRA* (Pogliano *et al.*, 1997). Support for this proposal came from western blotting experiments which showed that in Cpx* suppressor strains increased levels of CpxA and CpxR protein were observed. This model was tested by assaying levels of LacZ from a *cpxR-lacZ* promoter fusions in wild type, *cpxR* mutant and cpx* backgrounds. Reduced levels of *cpxRA* were seen in the mutant backgrounds while increased levels were seen in the Cpx* strains. The activation of the Cpx response by NlpE

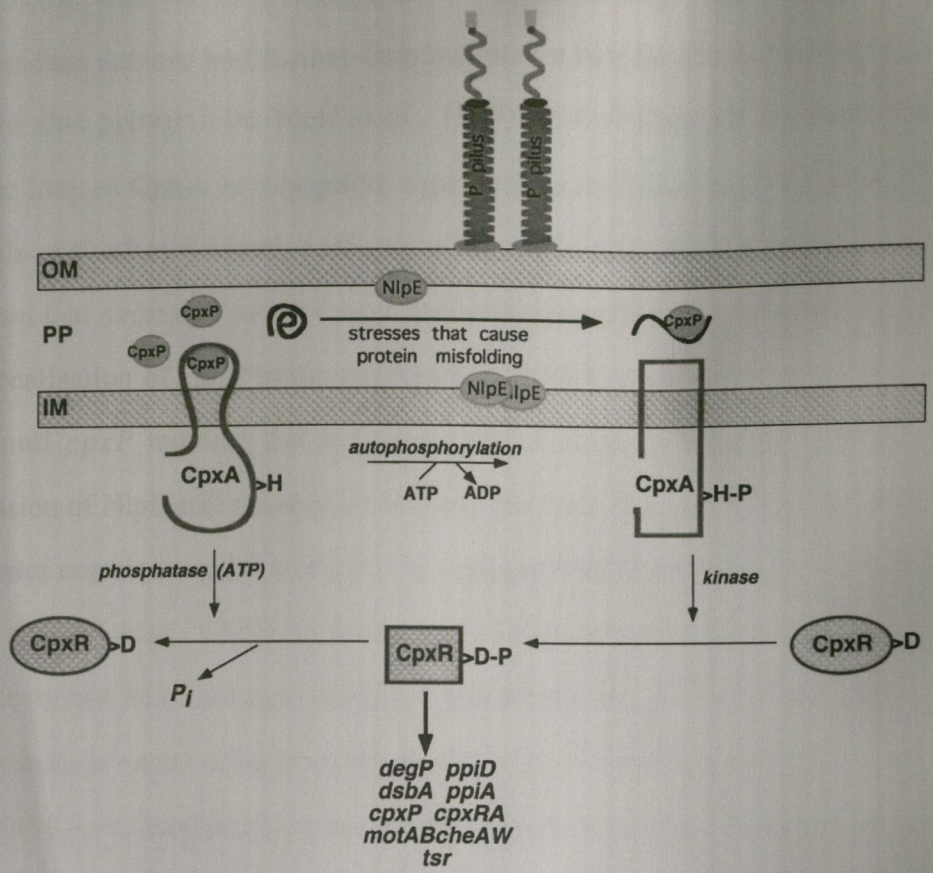


Figure 1.3. Regulation of the CpxRA two-component extracytoplasmic stress response. During normal growth of *E. coli* the CpxRA system is maintained at a basal level of activity by the negative affect of CpxP and the phosphatase activity of CpxA. An increase in the level of periplasmic protein misfolding or cell adhesion mediated via P-pili generates a signal to activate the Cpx system. This activation may involve titration of the CpxP protein away from its interaction with the periplasmic loop of CpxA, causing autophosphorylation of CpxA~P. The phosphate group is transferred to the cytoplasmic response regulator CpxR~P which increases transcription of the Cpx regulon genes. The response is negatively and positively autoregulated by the increased expression of both CpxP and CpxRA respectively. OM, outer membrane; IM, inner membrane; PP, periplasm. This figure is from Raivio and Silhavy (2001).

overexpression also demonstrated increased *cpxRA* transcription (Raivio *et al.*, 1999). In parallel work De Wulf and co-workers also showed the autoregulation of *cpxRA* in a growth phase dependent manner and further demonstrated a role for the stationary phase sigma factor σ^S in this process (De Wulf *et al.*, 1999). In revisiting the proposed role of the periplasmic loop of CpxA as a negative signalling factor Silhavy's group investigated the possible role of CpxP as a negative effector of Cpx activity by binding to this CpxA domain. It was shown that overexpression of *cpxP* had a negative effect on Cpx activity and that this required localisation of *cpxP* to the periplasm and a wild type CpxA periplasmic domain. However null *cpxP* mutants did not show full activation of the Cpx response and the overexpression of NlpE still produced a normal Cpx stress response. This implied that CpxP is not a master negative regulator of the Cpx response and that other factors are necessary for activation (Raivio *et al.*, 1999). In recent work the involvement of *cpxP* in regulating the Cpx response has been further studied. Sphaeroplasting involves the removal of the peptidoglycan layer under controlled osmotically balanced conditions to prevent cell lysis and results in the loss of periplasmic contents. This procedure has been shown to induce the Cpx response. In keeping with its role in downregulating the Cpx response, binding of CpxP to the outer face of the inner membrane through a non cleavable MBP-CpxP fusion derivative severely reduces Cpx induction following sphaeroplasting treatment (Raivio *et al.*, 2000).

1.5.8.3. Cpx regulon members and inducing signals

As expected for a regulon that responds to extracytoplasmic stress, the members of the Cpx regulon identified to date consist of proteins involved with refolding, attainment of correct conformation and degradation of aggregated proteins. Work by the Silhavy laboratory identified the *dsbA* gene as a member of the Cpx regulon, the same gene was identified by Beckwith's group using a candidate gene approach (Danese and Silhavy, 1997; Pogliano *et al.*, 1997). With the identification of the putative Cpx binding site a further member of the Cpx regulon was shown to be the peptidyl-prolyl isomerase protein PpiA (RotA) (Pogliano *et al.*, 1997). A second peptidyl-prolyl, PpiD, was also shown to be a member of the Cpx regulon and also to be regulated by the cytoplasmic heat shock regulator σ^{32} (Dartigalongue and Raina, 1998). Together with *degP* and *cpxP* these genes were the sole members of the

Cpx regulon until very recently. A newly enlarged σ^E regulon has been described and as part of the analysis of this regulon an investigation of possible twin control mechanisms by σ^E and Cpx was undertaken (Dartigalongue *et al.*, 2001b). This study revealed that three members of the σ^E regulon, *dsbC*, *skp*, and *ecfI* were also subject to additional transcriptional control by the Cpx system. In agreement with this experimental observation all three genes had motifs matching the Cpx binding site.

The signals inducing the Cpx regulon are varied. As detailed above the initial stimuli identified were the production of a range of model proteins which cause extracytoplasmic toxicity. The presence of the Cpx system in other bacteria was also noted and this led to the finding that the Cpx system was induced by alkali pH and that Cpx mutants were sensitive to growth at high pH (Raivio *et al.*, 1999). The expression of specific endogenous proteins can also induce the Cpx system as exemplified by the lipoprotein NlpE. It has also been shown that while heat shock induces the σ^E response it does not activate the Cpx system. While the Cpx system has primarily been shown to respond to stress stimuli a further role in monitoring the pathway that leads to production of adhesive organelles has been shown (Jones *et al.*, 1997). In uropathogenic *E. coli* strains the presence of P pili mediates adhesion to host cell surfaces. The production of P pili from cytoplasmic subunits requires their export from the inner membrane, chaperoning, transport and assembly to the outer membrane. The process involves the use of a specific chaperone PapD to prevent unwanted "off pathway" intermediates accumulating from the pilus subunits. Studies in strains lacking the chaperone PapD showed that the Cpx system is induced by the overexpression of two different subunits, PapG and PapE. The σ^E response was also shown to be induced but only by overexpression of PapG (Jones *et al.*, 1997).

1.5.8.4. A common mechanism for signal transduction to activate the Cpx and σ^E responses ?

The two systems that monitor protein folding in the extracytoplasmic compartment of *E. coli* have been shown to overlap at the regulation of *degP* and recently at three other genes; *dsbC*, *skp* and *ecfI* (Dartigalongue *et al.*, 2001b). Both regulons respond to unique and common

sets of stimuli and cross regulation has also been observed where mutation of Cpx regulon members (*ppiD* and *spy*) induces the σ^E response (Dartigalongue and Raina, 1998; Raivio *et al.*, 2000). The current data show that the Cpx regulon is upregulated by increasing the levels of *cpxA* kinase activity relative to its phosphatase activity resulting in more of the phosphorylated form of CpxR~P. The current hypothesis backed by experimental data is that the periplasmic sensing domain of CpxA is important in modulating the relative ratio of kinase to phosphatase activity. A negative role for CpxP in modulating this activity has also been shown. This implies that under non-stress conditions the response is in an "off mode". A similar mechanism has also been proposed for activating the σ^E response, whereby the interaction of the negative regulator RseA with σ^E keeps the response in an "off mode". How then is the presence of misfolded proteins sensed and transmitted to the negative regulators CpxP/CpxA and RseAB. The current models favour a titration mechanism which allows the sensing of misfolded proteins directly or indirectly by CpxP and/or the periplasmic loop of CpxA resulting in increased kinase activity and upregulation of the Cpx system. The system can then be shut off as the stress is alleviated resulting in increased amounts of CpxP binding to the CpxA sensor domain.

In the case of the σ^E response the induction relies on the degradation of the negative regulator RseA by the DegS protease. The mechanism of sensing the presence of the misfolded proteins to activate this degradation has also been proposed to involve titration. Gross has proposed that the levels of free protein folding factors are sensed by binding to RseA-RseB and/or DegS (Ades *et al.*, 1999). An increase in misfolded proteins probably titrates these folding factors relieving the negative effect of this binding and resulting in degradation of RseA and induction of the σ^E response. The response will also increase the levels of RseAB and when the stress has been removed these proteins can again mediate their negative effect in conjunction with the increased pool of folding factors. This mechanism mirrors that found to control the activity of the cytoplasmic heat shock regulator σ^{32} by the DnaK/J chaperone system (Yura *et al.*, 2000).

Using a model toxic hybrid protein, DsbA-PhoA, a new mechanism of relieving extracytoplasmic toxicity has recently been reported. This system does not require or induce the Cpx system or invoke the σ^E response and is mediated by an untranslated RNA termed UptR for unfolded protein toxicity relief factor (Guigueno *et al.*, 2001).

1.5.9. Regulation of HtrA proteases in other bacteria.

It has now been established that many organisms possess one or more HtrA proteases, as occurs in *E. coli* (Pallen and Wren, 1997). In contrast while the regulation of *degP* expression in *E. coli* has been well studied very little is known regarding the mechanisms of regulation of the HtrA proteases in other organisms apart from stress induction analyses. This even applies to the regulation of the two *E. coli* DegP homologues DegQ (HhoA) and DegS (HhoB) (Bass *et al.*, 1996; Waller and Sauer, 1996). Neither *degQ* nor *degS* are heat shock inducible and appear to be individually transcribed from σ^{70} type promoters even though they are tandemly arranged on the chromosome (Waller and Sauer, 1996).

One exception to the general lack of information regarding the regulation of HtrA-like proteases applies to *Pseudomonas aeruginosa*. Alginate is an extracellular polysaccharide which is expressed in *P. aeruginosa* isolates of patients suffering from cystic fibrosis. During studies on the complex regulation of alginate biosynthesis the role of a sigma factor AlgU (AlgT) belonging to the extracytoplasmic function (ECF) family of σ factors was demonstrated (Martin *et al.*, 1993; Schurr *et al.*, 1996). AlgU is the *P. aeruginosa* equivalent of σ^E and was shown to reside in an operon of similar structural organisation encoding the negative regulators MucA and MucB, homologues of RseA and RseB. In efforts to identify other regulators of alginate synthesis two genes, *algW* and *mucD*, were identified as negative regulators of alginate synthesis (Boucher *et al.*, 1996). Both MucD and AlgW are HtrA-like serine proteases based on homology analysis. *mucD* is the fifth gene of the *algU* operon and its mutation leads to the production of alginate and the associated mucoid phenotype. *mucD* is very probably transcribed as part of the *algU* operon and since this operon is positively autoregulated by AlgU *mucD* is under the control of the σ^E-like sigma factor AlgU. Mutation of *algW* also leads to induction of mucoidy in strains subjected

to sublethal levels of the redox cycling compound paraquat. The regulation of *algW*, which is not linked to the *algU* operon, is not known and does not have any recognisable σ^E-type ECF promoter motifs (Boucher *et al.*, 1996).

In *Mycobacterium tuberculosis* a gene specifying an HtrA protease was found to be part of an operon encoding an ECF sigma factor, SigE, similar to the situation of *mucD* in *P. aeruginosa*. (Wu *et al.*, 1997). *In vitro* analysis with reconstituted mycobacterial holoenzyme $E\sigma^E$ showed that transcription initiated from the regulatory region of the σ^E operon indicating positive autoregulation. However the determined *in vivo* start sites were different and not fully dependent on the presence of a *sigE* gene. Lastly, there was only limited homology to ECF-type promoter motifs in the regions upstream of the *in vivo* transcription start points. Therefore, the regulation of at least one *htrA* gene from *M. tuberculosis* by an ECF-type sigma factor is possible.

The regulation of stress inducible *htrA* genes from *Bartonella henselae* and *Legionella pneumophilia* by ECF-type sigma factors has also been indicated as the promoter regions of both genes contain ECF-type promoter motifs. Transcriptional analysis has shown an *in vivo* derived transcriptional start point downstream of the putative ECF-type promoter of *htrA* in *B. henselae* (Resto-Ruiz *et al.*, 2000). Deletion of the ECF-type promoter region of the *L. pneumophilia htrA* gene on a complementing plasmid failed to rescue the reduced pathogenic phenotype of a *L. pneumophilia htrA* mutant *in vitro*. This indicates that the ECF-type promoter is necessary for expression of *htrA* in *L. pneumoniae* (Pedersen *et al.*, 2001). The regulation of the htrA gene from *S. pneumoniae* has been elucidated through microarray transcriptome analysis. This recent finding has demonstrated that the *htrA* gene is regulated by the CiaRH two component system. Furthermore, an important role for HtrA in the pathogenesis of *S. pneumoniae* is indicated.(Sebert *et al.*, 2002).

1.6. The Heat Shock Response of *Bacillus subtilis*

Homologues of many of the heat shock response proteins identified in *E. coli* have now been identified and characterised in many other bacteria (Koonin *et al.*, 2000). The laboratories of Hecker, Schumann and Wong isolated and cloned the genes encoding the classical heat shock chaperones DnaK-DnaJ and GroES-GroEL (Li and Wong, 1992; Schmidt *et al.*, 1992; Wetzstein *et al.*, 1992). In parallel with these studies the technique of 2D gel electrophoresis was used successfully by the Hecker laboratory to characterise a stress regulon controlled by the alternative sigma factor σ^B (σ^{37}) (Volker *et al.*, 1994). During these investigations it also became clear that a subset of proteins from the regulon controlled by σ^B was controlled by additional mechanisms. Currently the various proteins responsive to heat shock and other stress stimuli are grouped into at least four different classes based on the regulatory mechanisms governing their expression (Hecker and Volker, 1998). The class I and class III heat shock responses are induced at the transcriptional level by relieving the effects of negative regulators, while the class II response is induced by release of a sigma factor from its interaction with an anti-sigma factor. The regulatory systems governing control of other heat stress inducible genes, grouped into class IV, have not been yet been defined (Schumann *et al.*, 2002).

1.6.1. Class I heat shock genes.

The genes encoding the *B. subtilis* DnaK-DnaJ chaperones were isolated and cloned by the Schumann and Hecker groups using a cross species probe from *Bacillus megaterium* (Wetzstein *et al.*, 1992). The *dnaK* and *dnaJ* genes were shown to be the third and fourth genes of a four gene operon. The first ORF encoded a gene of unknown function while the second gene encodes a homologue of the *E. coli* GrpE protein which functions as part of the *E. coli* DnaK-DnaJ chaperone machine. Transcriptional analysis showed that after heat shock induction increased amounts of mRNA were produced. It also demonstrated that the increase was transient and that at least three different sized transcripts were present (Wetzstein *et al.*, 1992).

A similar strategy, again by the same two laboratories, of using cross species probe was also successful in cloning the *groE* locus (Schmidt *et al.*, 1992). This locus was simultaneously isolated by Wong's laboratory using a serum cross reacting against *E. coli* GroEL to isolate the *B. subtilis* GroEL protein from heat shocked cells after SDS\PAGE analysis (Li and Wong, 1992). N-terminal protein sequencing of the isolated band was used to design degenerate oligonucleotide probes and the complete *groE* locus was isolated from a *B. subtilis* plasmid library. The operon was shown to comprised two genes *groES* and *groEL*. Transcriptional analysis by primer extension and Northern blotting showed that increased amount of a single transcript accumulated after heat shock. The GroE proteins were shown to be essential by placing the operon under the transcriptional control of a sucrose inducible promoter P*sacB* and demonstrating that viability relied upon the presence of the inducer (Li and Wong, 1992).

A feature of both the *dnaK* and *groE* operons was the presence of an 9-bp inverted repeat at the 5' nontranslated region situated between the transcriptional and translational start points (Li and Wong, 1992; Wetzstein *et al.*, 1992). It was proposed that this could be used as part of a regulatory mechanism for controlling the induction of the two operons. Mutational analysis of the inverted repeat (IR) preceding the *dnaK* operon showed that disruption of the dyad symmetry caused increased amount of transcription from the *dnaK* operon but not that of the *groE* operon. This implied the IR acted as a *cis* acting negative element and furthermore the IR could function to downregulate other genes if placed appropriately within their control regions (Zuber and Schumann, 1994). Additional mutational analysis of individual genes of the *dnaK* operon showed that insertional mutation of the *grpE*, *dnaK* or *dnaJ* did not affect basal or heat shock regulation of the *dnaK* or *groE* operons. However mutation of the first gene ORF39 caused high level of transcription from the *groE* operon (Schulz *et al.*, 1995). This implicated ORF39 as encoding a possible repressor acting at the IR (now termed CIRCE) for controlling inverted repeat of chaperone expression. This was subsequently demonstrated by showing that regulatory mutants of the *groE* operon contained point mutations in the ORF39 coding sequence and that these mutations could be complemented *in trans* by plasmid encoded ORF39. This study by Yuan (Yuan and Wong,

1995a) also indicated the ability of the CIRCE element to bind the ORF39 protein produced from crude lysates in overproducing *E. coli* strains. The regulation of expression by ORF39 at CIRCE was also shown to be functional in *E. coli* using a binary plasmid system. The product of ORF39 was called HrcA for heat regulation at CIRCE (Schulz and Schumann, 1996).

Further analysis of the mechanism controlling the relief of repression by HrcA at CIRCE demonstrated that the GroE chaperone machine was the main effector of this repression (Mogk *et al.*, 1997). Deletion of the genes encoding the components of the DnaK chaperone system did not produce any effect on repression. However, changing the levels of GroES-EL from a xylose inducible promoter produced dramatic effects. Increasing the level reduced the basal and induced levels of the RNA from both the *groE* and *dnaK* loci, while decreasing the amounts of GroES-EL increased the basal level of these transcripts. The levels of GroES-EL also affected the cells thermosensitivity further substantiating the essential role of the GroE system. To analyse the role of GroEL in interacting with HrcA overexpression and isolation of *B. subtilis* HrcA was attempted. This failed due to insolubility of the protein but was successful when the *hrcA* gene of *B. stearothermophilus* was used. *In vitro* analysis with the purified HrcA protein in the presence of GroEL and ATP demonstrated binding to and retardation of the *groE* probe (Mogk *et al.*, 1997).

The current model for regulation of the *groE* and *dnaK* operons comprising the class I genes of *B. subtilis* involves the chaperoning role of GroEL (and probably GroES) in maintaining HrcA in a nonaggregated and function state under non stress conditions (see Figure 1.4.). Upon heat shock the positive effect of GroEL on HrcA's function is likely to be removed by titration of GroEL to misfolded proteins. This reduces the negative repression by inactive HrcA leading to induction of the class I genes. This in turn produces more GroEL and HrcA creating a negative feedback loop to downregulate transcription of the class I genes after the initial transient increase (Mogk *et al.*, 1997; Schumann *et al.*, 2002).

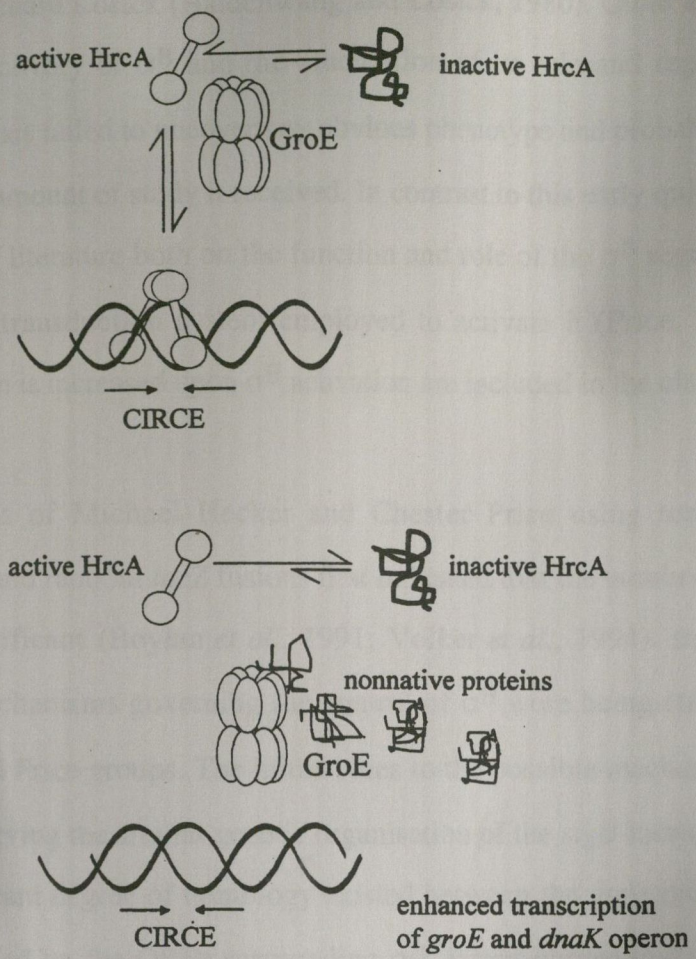


Figure 1.4. Regulation of Class I heat shock genes in *B. subtilis*. At normal growth temperatures there is enough GroE chaperone protein to maintain sufficient amounts of HrcA in an active confirmation. HrcA bound to the inverted repeat CIRCE located in the promoter regions of the *dnaK* and *groE* operons maintains a basal level of transcription of both operons. During heat shock or when protein misfolding occurs in the cytoplasm GroE chaperones are titrated away by the increased levels of misfolded proteins. This results in greater amounts of inactive HrcA unable to bind to the CIRCE operator and leads to a rapid increase in the levels of *dnaK* and *groE* operon transcripts. The gene products act to reduce the levels of misfolded proteins and also increase the level of HrcA, whose gene is part of the *dnaK* operon. The greater availability of GroE acts to again increase the amount of active HrcA and downregulate the transcriptional response. This diagram is from Schumann *et al.* (2002).

1.6.2. Class II heat shock genes

The first alternative sigma factor to be discovered in *B. subtilis* was σ^B isolated in the laboratory of Richard Losick (Haldenwang and Losick, 1980). Quite a time period elapsed between the discovery of σ^B and the elucidation of its role and regulatory mechanism. Mutational analysis failed to uncover any obvious phenotype and probably contributed to the relatively small amount of study it received. In contrast to this early quiescence there is now an abundance of literature both on the function and role of the σ^B regulon and on the very complex signal transduction system employed to activate it (Price, 2002). Those genes whose expression is increased upon σ^B activation are included in the class II group.

The laboratories of Michael Hecker and Chester Price using techniques of 2D gel electrophoresis and random *lacZ* fusions first indicated that the number of genes controlled by σ^B was significant (Boylan *et al.*, 1991; Volker *et al.*, 1994). In parallel with these analyses the mechanisms governing the control of σ^B were being studied by the Losick, Haldenwang and Price groups. The initial clues to the possible mechanisms involved were gleaned by observing the similar genetic organisation of the *sigB* locus to that of the *spoIIA* locus. A significant degree of homology existed between the gene products regulating σ^F and those encoded by the ORFs surrounding σ^B. It was shown that the *sigB* operon was autoregulated similar to the spoIIA locus (Kalman *et al.*, 1990). Since this primary discovery a great many different and varied stress stimuli have been shown to induce the expression of the σ^B regulon which may contain upwards of two hundred genes (Helmann *et al.*, 2001; Petersohn *et al.*, 2001; Price *et al.*, 2001)

The signal transduction mechanism which operates to induce the activation of σ^B is now very well understood and has proved to be exceptionally complex as detailed in the latest model offered by Haldenwang (Zhang *et al.*, 2001) and shown in Figure 1.5. There are two types of stress which induce the activation of the σ^B protein: (i) environmental stimuli, such as heat shock, exposure to ethanol, salt and acid stress and (ii) energy stress signals resulting from reduced amount of carbon, oxygen and phosphate. Each of these stresses produce signals processed by two separate signal transduction systems in the early part of the signal

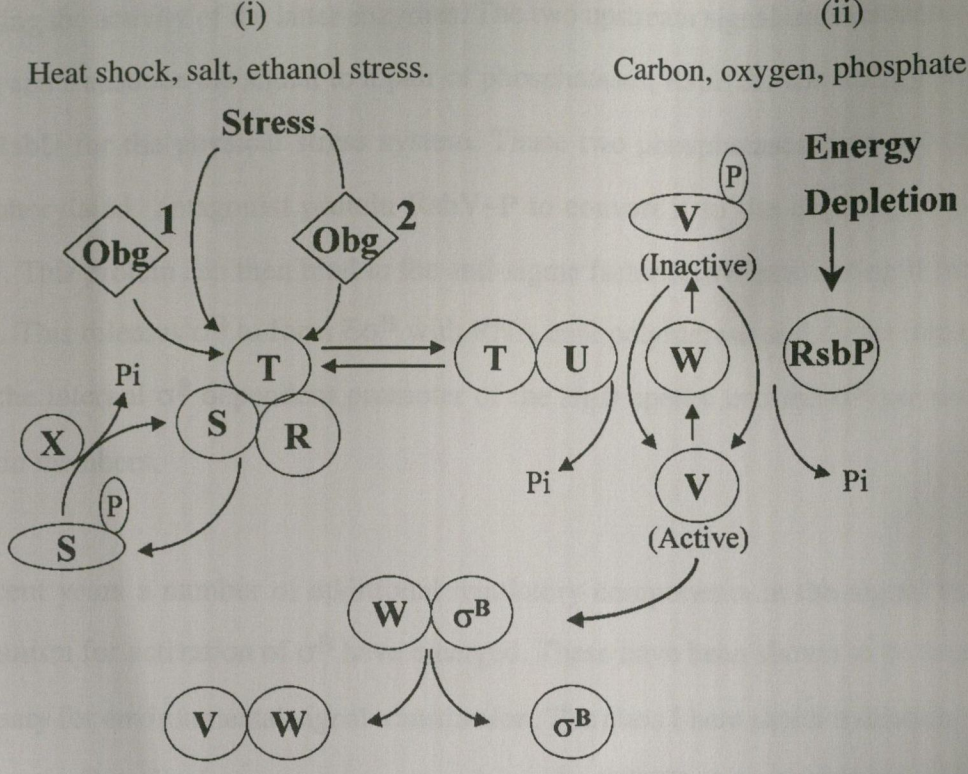


Figure 1.5. Regulation of the general stress response of *B. subtilis*. This response controls the heat shock induction of the σ^B-dependent class II heat shock genes. Two types of stress can induce the σ^B regulon; (i) environmental or physical stresses such as heat, salt, ethanol and acid, and (ii) energy depletion due to lack of carbon, oxygen or phosphate. Each stress type is sensed by separate "modules" composed of the Rsb regulatory proteins (RsbP/R/S/T/U/V/W/X), which ultimately result in the release of σ^B from its anti-sigma factor RsbW and enabling transcription of the σ^B regulon genes. The illustration shows the complex interaction of the members of each module comprised of phosphatases, kinases and activators. The final step by each module involves the phosphatase activity of both RsbU and RsbP to accomplish activation of RsbV. P and Pi indicate phosphorylated proteins and inorganic phosphate respectively. This diagram is adapted from Zhang *et al.* (2001).

transduction cascade. The two signalling modules coalesce to relieve the negative regulation of σ^B by the anti-sigma factor RsbW (regulator of sigma-B). The key proteins of both signal transduction systems are comprised of kinases, phosphatases and modulatory proteins affecting the activity of the latter enzymes. The two upstream signal transduction components sense and transduce the signal to a pair of phosphatases, RsbP for the energy stress system and RsbU for the physical stress system. These two phosphatases then act on the same phosphorylated antagonist protein RsbV~P to convert it to the dephosphorylated form RsbV. This protein can then bind to the anti-sigma factor RsbW preventing it from binding to σ^B. This releases σ^B to form $E\sigma^B$ with RNA core polymerase and direct transcription of both the internal σ^B dependent promoter of the *sigB* operon and the σ^B promoters of the regulon members.

In recent years a number of additional regulatory components in the signal transduction mechanism for activation of σ^B have emerged. These have been shown to be important and necessary for environmental signal transduction. The class I heat shock induction system can function in *E. coli* when a promoter containing the CIRCE element and the gene encoding the HrcA protein was present (Mogk *et al.*, 1997). However, when the components of the σ^B system are present in *E. coli* no induction by physical stress was observed. This indicated that there was a requirement for a *B. subtilis* specific factor (Scott *et al.*, 1999). This same study also observed that varying the levels of the DnaK or GroE chaperone machines in *B subtilis* did not affect the level of σ^B activity which corroborates studies showing that truncated proteins generated by use of puromycin did not induce the response and furthermore that heat shock produces the weakest σ^B-dependent response (Akbar *et al.*, 1997; Mogk *et al.*, 1997). Therefore, the presence of misfolded proteins is not a major factor in the σ^B signal transduction activating pathway in contrast to that of the *B. subtilis* class I genes and of the σ^{32} and σ^E responses in *E. coli*.

In recent studies the necessity of an essential GTP-binding protein for σ^B activation by environmental stresses has been shown (Scott and Haldenwang, 1999). The protein Obg was found to interact with various members of the signal transduction system (Rsb proteins)

in screening yeast two hybrid libraries. Additionally, both the Obg and Rsb proteins were found to co-fractionate with ribosomes from *B. subtilis* (Scott *et al.*, 2000). Lastly, it has also been demonstrated that the L11 ribosomal protein is required for signal transduction of environmental stress to activate σ^B (Zhang *et al.*, 2001). These findings have prompted the suggestion that the ribosome of *B. subtilis* is one of the first sensors of certain types of stress. This sensor mechanism produces a signal which is then communicated to the Rsb proteins through or in conjunction with Obg to activate σ^B. The requirement for Obg and L11 proteins is specific to σ^B activation by environmental stresses as energy stress signalling to activate σ^B is unaffected by loss of L11 or depletion of Obg proteins.

1.6.3. Class III heat shock genes

Analysis of the class II heat shock genes by 2D gel electrophoresis showed that a number of proteins were still induced in a *sigB* mutant background (Volker *et al.*, 1994; Bernhardt *et al.*, 1997). Furthermore these proteins were not members of the class I heat shock proteins described above. This indicated the presence of a third class of heat shock proteins in *B. subtilis*. Two groups subsequently isolated and cloned the gene encoding the ClpC ATPase which was identified as being a member of this new class of heat shock genes (Kruger *et al.*, 1994; Msadek *et al.*, 1994). The phenotypes associated with mutations in the *clpC* gene were pleiotrophic in nature affecting sporulation, secretion and competence. Following on from the isolation and mutation of the *clpC* gene and its associated effects, the *clpP* gene encoding the proteolytic subunit of the Clp protease was cloned and analysed. This gene and its encoded protein showed a similar *sigB*-independent stress induction pattern and mutation of the gene produced similar phenotypes to that of a *clpC* mutant (Gerth *et al.*, 1998; Msadek *et al.*, 1998).

Transcriptional and nucleotide analyses of the *clpC* gene indicated it was the fourth gene of a six gene operon while the *clpP* gene is monocistronic. Primer extension analyses showed that both genes were transcribed from two promoters and in each case one is σ^B-dependent. In a *sigB* mutant strain the alternate promoters became significantly more stress inducible to stimuli such as heat and ethanol stress but also to oxidative stress and by exposure to the

antibiotic puromycin (Kruger *et al.*, 1996; Gerth *et al.*, 1998). These later two stresses were shown not to induce σ^B-dependent promoters (see above). The sequences upstream of the σ^B-independent inducible promoters showed similarity to SigA-like -35 and -10 motifs.

Sequence analysis suggested that the first gene of the *clpC* operon encoded a protein which contained a helix-turn-helix motif similar to many DNA binding proteins with regulatory functions. This regulatory role was confirmed by in frame mutations in the first gene ORF1 which led to derepression of both *clpP* and *clpC* operon expression. ORF1 was renamed *ctsR* for class three stress gene regulator (Kruger and Hecker, 1998; Derre *et al.*, 1999b). Following overexpression and purification of the CtsR protein, the ability of CtsR to bind to sites in the promoter region of the *clpP* and *clpC* operons was demonstrated. DNase footprinting determined the recognition sequence for binding by CtsR to be a repeated heptameric motif (Derre *et al.*, 1999b). A third member of the CtsR controlled heat shock genes *clpE*, encoding another Clp ATPase, was also shown to contain the heptameric motif (Derre *et al.*, 1999a). However, even in the strains mutated for CtsR resulting in greatly increased expression of *clpC* and *clpP* operons, there remains a residual 2-3 fold induction of expression under stress induction. This indicates a second or parallel system is used in addition to the negative effect of the CtsR repressor to increase expression after stress induction (Derre *et al.*, 1999a).

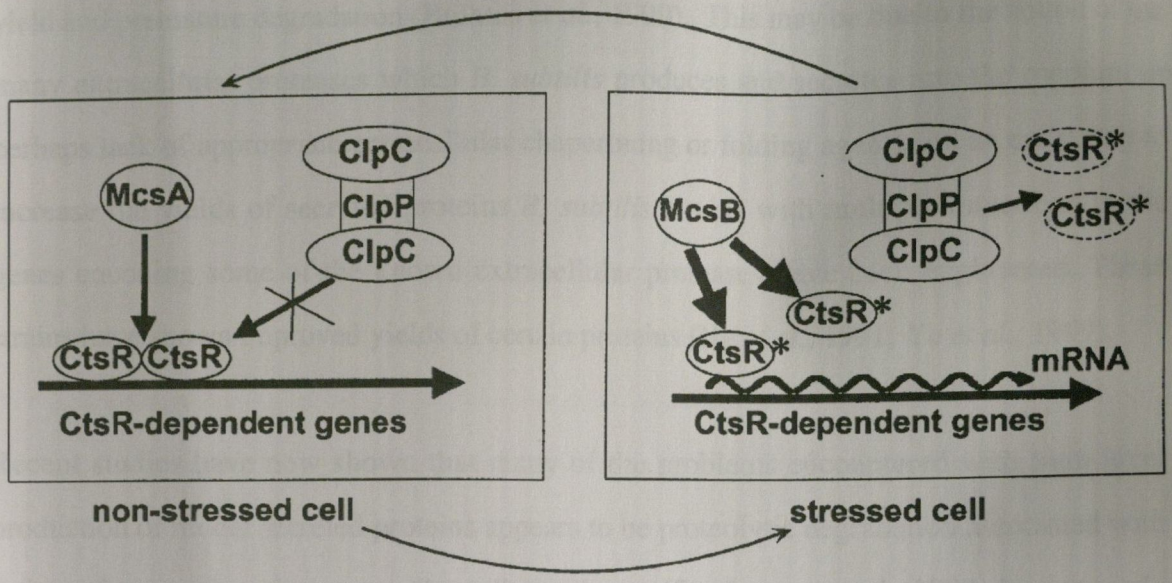
Further investigation of regulation by CtsR has demonstrated a role for the ClpP protease and the ClpC ATPase subunit requiring the chaperone functions of the latter protein. Investigations also showed that CtsR functioned as a dimer and that the CtsR protein has a number of distinct functional domains. The amino terminal end was shown to be responsible for dimerisation and in combination with the central region may confer heat sensitivity (Derre *et al.*, 2000). Further work on the roles of the second and third proteins encoded in the *clpC* operon showed that they functioned as modulators of CtsR repression (McsA, McsB). Cells with separate non polar in frame mutations of *mcsA* and *mcsB* were created and shown to have opposing effects on the ability of CtsR to function (Kruger *et al.*, 2001). *mcsA* mutations showed depression under non stress conditions but still retained stress induction

ability. *mcsB* mutations were only slightly affected in the basal level of transcription but displayed a much reduced induction ability under stress conditions. Biochemical analysis has shown that McsB possesses kinase activity and is responsible for phosphorylating CtsR *in vivo*. This activity correlates to the reduced stability of CtsR under stress conditions. McsA may play a role in promoting the dimerisation and binding of CtsR to DNA. The current model now postulates that McsA aids the ability of CtsR to dimerise and bind to its recognition sites under nonstress conditions. After stress is imposed McsB phosphorylates CtsR reducing its ability to dimerise and bind DNA and makes the protein accessible to the ClpCP protease for degradation (Kruger *et al.*, 2001 and see Figure 1.6.).

1.7. Extracytoplasmic stress response and HtrA proteases in *B. subtilis*.

The three classes of heat shock genes described above primarily encode cytosolic proteins. The class I and III genes encode the classical DnaK and GroE chaperone systems and components of the Clp proteolytic systems respectively. The class II response is comprised of perhaps up to two hundred genes which regulates the induction of mainly cytosolic proteins, although some secreted proteins have recently been shown to be regulated by σ^B (Petersohn *et al.*, 2001; Price *et al.*, 2001).

Genome and proteome analysis has indicated that *B. subtilis* contains about 300 secretory proteins (Tjalsma *et al.*, 2000; Antelmann *et al.*, 2001). The secretory apparatus of *B. subtilis* has been extensively studied both as a model system for Gram positive bacteria and from a biotechnological viewpoint (Tjalsma *et al.*, 2000; van Wely *et al.*, 2001). *B. subtilis* is an industrially important bacterium reflected in its leading role as a major source of both commercial enzymes and pharmaceutical vitamins (Sonenshein *et al.*, 1993). These applications came about from the observations that the wild type bacterium has a natural ability to secrete large amounts of exoproteins during growth in complex media. Many of these proteins are degradative enzymes utilised to break down complex macromolecules extending the growth cycle and delaying the advent of sporulation (Msadek, 1999).



igure 1.6. Regulation by the CtsR repressor of class III heat shock genes in *B.subtilis*. 'nder non-stressed conditions the CtsR repressor is able to bind as a dimer to its operator tes upstream of the *clpC* operon and *clpP* and *clpE* genes. The binding of CtsR is ositively modulated by the activity of McsA. When cells are subjected to heat shock or ccumulate increased amounts of unfolded proteins the binding of CtsR protein to its perator sequence is diminished. The McsB protein through its kinase activity upon CtsR egatively modulates CtsR binding. The unbound phosphorylated CtsR is then accessible to ıe ClpCP protease. The *ctsR*, *mcsA*, *mcsB* and *clpC* genes constitute the clpC operon. ıcreased levels of these proteins coupled with titration of ClpCP by unfolded proteins establish the pre-stressed levels of the Clp regulon proteins to downregulate the system. his figure is from Kruger *et al.* (2001).

However despite the success of utilising *B. subtilis* as a 'cell factory' for the production of a number of commercial enzymes and vitamins, the potential for greater exploitation remains. Many of the problems encountered with the production of other compounds relate to poor yield and premature degradation (Bolhuis *et al.*, 1999). This may be due to the action of the many extracellular proteases which *B. subtilis* produces and secretes into the medium or perhaps lack of appropriate extracellular chaperoning or folding agents. As an early step to increase the yields of secreted proteins *B. subtilis* strains with multiple mutations in the genes encoding some of the known extracellular proteases have been engineered. These strains have shown improved yields of certain proteins (He *et al.*, 1991; Ye *et al.*, 1999).

Recent studies have now shown that many of the problems encountered with high level production of model secreted proteins appears to be proteolytic degradation associated with or just after translocation across the cell membrane (Stephenson *et al.*, 1998; Jensen *et al.*, 2000). This is in contrast to the fate of secreted proteins in the medium. Expression of the model *B. licheniformis* secretory protein AmyL in *B. subtilis* is stable in the medium and most degradation of AmyL occurs in a cell associated context. Improved yields of AmyL were found in cells containing a mutation in the gene encoding a cell wall protease WprA (Stephenson and Harwood, 1998). Other limiting factors in the efficient high level production of other model proteins seem to point to an important role for the lipoprotein PrsA (Kontinen and Sarvas, 1993; Vitikainen *et al.*, 2001). This peripheral membrane protein is a homologue of SurA, the periplasmically localised *E. coli* chaperone/peptidyl-prolyl *cis-trans* isomerase (Lazar and Kolter, 1996; Behrens *et al.*, 2001). A third problem associated with increasing yields of secreted proteins is the large anionic charge density of the thick cell wall of *B. subtilis* which may impede release of correctly translocated proteins leading to their degradation by cell membrane-cell wall proteases (Stephenson and Harwood, 1998; Hyyrylainen *et al.*, 2000).

In contrast to the well studied σ^E and Cpx systems operating to maintain homeostasis in the periplasm of *E. coli*, no defined equivalent system has yet been identified in *B. subtilis*. This may reflect the major differences in the cell envelope structure between Gram negative

and Gram positive organisms. The periplasm of *E. coli* is a self contained double membrane bounded space through which all cell exported and environmentally imported molecules must traverse. The σ^E and Cpx systems are vital in maintaining this space free of misfolded or denatured proteins. Gram-positive bacteria such as *B. subtilis* have in contrast no defined periplasmic space. These bacteria contain a single cell membrane and a very thick cell wall composed of many layers of proteoglycan and associated anionic polymers such as teichoic acids (Archibald *et al.*, 1993; Foster and Popham, 2002). It has been proposed that the space immediately external to the outer leaflet of the cell membrane and before the beginning of the peptidoglycan layer constitutes an equivalent "periplasm" in *B. subtilis* (Merchante *et al.*, 1995). This space may be the environment where post-translocation folding/chaperoning of secreted and exported proteins can take place. In the absence of correct folding or perhaps due to overloading of the folding systems proteolytic enzymes would be required to remove the potentially toxic accumulation of these misfolded proteins.

During the course of the *B. subtilis* genome sequencing project three HtrA-like proteases were found to be encoded by *B. subtilis* (Kunst *et al.*, 1997). One of these, YkdA, was found in the region of the genome sequenced by the author of this work. An analysis of the expression of *ykdA* and of the other genes *yvtA* and *yyxA*(*yycK*) encoding the remaining HtrA proteases is the theme of this thesis. The work addresses the functional and regulatory relationships of the three proteases particularly in comparison to the situation in *E. coli*.

CHAPTER TWO

MATERIALS AND METHODS

2.1. GROWTH AND MAINTENANCE OF BACTERIAL STRAINS

Bacterial strains and plasmids used in this study are listed in Table 2.1. *E. coli* and *B. subtilis* were maintained on Luria-Bertaini (LB) agar for short term storage, for long term storage glycerol was added to a final concentration of 30% (vol/vol) to overnight cultures grown in LB broth and stored at -70°C. *E. coli* and *B. subtilis* were grown in liquid media with ample aeration (200-250 rpm) in a Gallenkamp rotary incubator. 5-Bromo-4-chloro-3-indolyl-D-galactoside (X-gal) was included in solid media when required at a concentration of 40 μg/ml. Isopropyl-β-D-thiogalactopyranoside (IPTG) was included in liquid or solid media at the concentrations indicated in the text.

2.1.1. Antibiotic concentrations

Antibiotics were added to media for selection and/or maintenance of constructs at the following concentrations: for *E. coli* , ampicillin 100 μg/ml; for *B. subtilis* , chloramphenicol 3 μg/ml, erythromycin 1 μg and 150 μg/ml, kanamycin 10 μg/ml and spectinomycin 100 μg/ml.

2.1.2. Composition of Growth Media

Luria-Bertani (LB) broth :

10 g tryptone, 10 g NaCl and 5 g yeast extract per litre.

Buffered LB broth :

as for LB broth but with 21 mM K_2HPO_4 and 11 mM KH_2PO_4 (pH6.9) replacing the NaCl (Boylan *et al.*, 1993)

Schaeffers medium (SM) (Schaeffer *et al.*, 1965):

Nutrient broth 8g, KCl 1g and $MgSO_4$ 0.25g per litre. After autoclaving the following were added from sterile stocks: $CaCl_2$ to 1mM, $MnCl_2$ to 0.01mM and $FeSO_4$ to 0.001mM.

Spizizen minimal salts (SMS); KH_2PO_4 6 g, K_2HPO_4 14 g, $(NH_4)_2SO_4$ 2 g, $MgSO_4.7H_2O$ 0.2 g, Na_3citrate 1 g per litre.

Solid media were made by the addition of agar to a final concentration of 1.5% (wt/vol).

2.2. STRESS INDUCTION AND PHENOTYPIC ANALYSES

2.2.1. Stress inductions

Strains of *B. subtilis* grown at 37°C in LB broth to an optical density of 0.3 measured at 550nm ($OD_{550} \approx 0.3$), were subjected to stresses by the following procedures; for heat stress one half of the culture was transferred to a prewarmed flask in a water bath and growth continued at 48°C.

Other stresses were imposed using the same procedure except that both halves of the culture were maintained at 37°C with one half exposed to stressors at the following concentrations: NaCl, 0.3M; ethanol, 4% (vol/vol); hydrogen peroxide, 0.1mM and puromycin 15 μg/ml.

2.2.2. Phenotypic analyses

2.2.2.1. Thermosensitivity

A qualitative measurement of the thermosensitivity of *B. subtilis* strains was obtained by splitting exponentially growing cultures. One half was maintained at 37°C while the other was transferred to the sub-lethal temperature of 52°C (Volker *et al.*, 1999). The optical density at 550 nm (OD_{550}) of both cultures were then recorded at the indicated time points.

Quantitative measurements of the thermosensitivity of *B. subtilis* strains were gained by shifting 10 ml of an exponentially growing culture to the lethal heat shock temperature of 54°C (Volker *et al.*, 1999) and the remainder maintained at 37°C. Aliquots of both cultures were removed at the indicated time points, serially diluted in sterile LB, spread onto LB agar plates and incubated for 18 hrs at 37°C. Percentage survival was determined from the number of viable cells from each culture at the indicated time points.

2.2.2.2. Oxidative stress resistance

Resistance to lethal concentrations of hydrogen peroxide was determined by splitting an exponentially growing culture (in SM broth) and exposing one half to broth containing freshly diluted hydrogen peroxide to a final concentration of 10mM. Aliquots of the stressed and control cultures were serially diluted in sterile SM broth and plated as above. The

percentage survival was determined from the number of cells in the stressed and unstressed cultures at each time point.

2.3. DNA MANIPULATIONS.

2.3.1. Cloning procedures

Routine cloning procedures such as DNA restriction digestion, ligation reactions, agarose gel electrophoresis and Southern blotting, caesium chloride density gradient centrifugation of plasmid DNA, were undertaken using standard methods (Sambrook *et al.*, 1989).

Restriction enzymes, DNA polymerase I (Klenow), T4 DNA polymerase and *Bal*31 nuclease were purchased from New England Biolabs, T4 DNA ligase was purchased from Boeringher Mannheim.

2.3.2. Polymerase chain reaction (PCR)

PCR reactions were performed essentially as described (Sambrook *et al.*, 1989) based on single volume reactions of 25 μl containing; 1X PCR buffer (1.5mM $MgCl_2$ final conc.), 10 pmoles each primer (0.4 μM), 200 μM each dNTP and template DNA (1-5 ng) and either recombinant *Taq* DNA polymerase (Invitrogen, Paisley, UK) or *Pfu* polymerase (Stratagene, La Jolla, CA, USA)

PCR amplified promoter fragments used to create *lacZ* or *bgaB* fusions were sequenced to check the fidelity of the amplification process.

2.3.3. Plasmid DNA isolation

Plasmid DNA for sequencing and subcloning was prepared by the boiling lysis method (Holmes and Quigley, 1981). For *de novo* cloning of *B. subtilis* fragments using "minibanks" (see section 3.2.1) plasmid DNA was purified from CsCl gradients.

2.3.4. DNA fragment isolation

DNA fragments obtained by PCR or from restriction digestion of plasmids were isolated and purified from agarose gels. The desired DNA fragments were isolated from gel slices by

elution into dialysis tubing (Sambrook *et al.*, 1989) or using the CONCERT PCR purification kit (Life Technologies, Paisley, UK) following the instructions of the manufacturer.

2.3.5. *B. subtilis* chromosomal DNA isolation

B. subtilis chromosomal DNA was isolated from 1-2 ml volumes of overnight cultures. Cells were pelleted from the medium by a brief spin at 12,500 rpm in an eppendorf centrifuge and all traces of medium removed. The cell pellet was resuspended in 300 μl of solution I (0.05M Tris pH8, 0.05M EDTA, 25% [wt/vol] sucrose) and 100 μl of a 20 mg/ml solution of lysozyme freshly prepared in solution I was added. The mixture was incubated at 37°C for 30 minutes followed by addition of 400 μl of solution II (0.01M Tris pH8, 0.005M EDTA, 1% [wt/vol] SDS), 40 μl RNase solution (10 mg/ml), mixed and incubated as above for a further 15 min. The mixture was extracted by thorough mixing with 500 μl phenol and spun for 5 min at 12,5000rpm. The aqueous phase was removed to a fresh tube and re-extracted with an equal volume of a 50% (vol/vol) solution of phenol/chloroform, spun as above, the aqueous phase removed to a fresh tube and the DNA precipitated by addition of an equal volume of isopropanol. The DNA was pelleted by centrifugation, dissolved in 100 μl 0.1M potassium acetate and re-precipitated with 3 volumes of 96% (vol/vol) ethanol. The final DNA pellet was washed in 70% (vol/vol) ethanol, air dried and dissolved in 50-100 μl sterile water.

2.4. TRANSFORMATION of *E. coli* and *B. subtilis*

2.4.1. *E. coli* transformations.

E.coli were made competent and transformed by the calcium chloride method as described in Sambrook (Sambrook *et al.*, 1989).

2.4.2. *B. subtilis* transformations.

B. subtilis was transformed by the a modified two step procedure based upon the method of Anagnostopoulos and Spizizen (Anagnostopoulos and Spizizen, 1961). Briefly, an overnight culture grown in LB broth was diluted 10-fold into Broth I comprising; Spizizen's minimal salts containing 0.5% (wt/vol) glucose, 5 mM $MgSO_4$, 0.02% (wt/vol) yeast

extract, 0.02% (wt/vol) casamino acids and 0.05 mg/ml tryptophan and grown for four and a half hours. An aliquot of this culture was diluted five fold into Broth II comprising Spizizen's minimal salts containing 0.5% (wt/vol) glucose, 5 mM $MgSO_4$, 0.01% (wt/vol) yeast extract, 0.01% (wt/vol) casamino acids and 0.005 mg/ml tryptophan. Plasmid DNA (0.5 μg) or chromosomal DNA (12.5 ng) was added to 0.5 ml of the cells diluted in Broth II and incubation continued for a further 90 minutes. Transformants were selected on LB agar plates with appropriate antibiotic selection.

2.5. SEQUENCING AND SEQUENCE ASSEMBLY OF LARGE *B. SUBTILIS* GENOMIC FRAGMENTS

2.5.1. Creation of "mini-banks" of *B. subtilis* clones in *E. coli.*

A plasmid, pSAL16, containing a fragment of the *B. subtilis* genome was obtained from Mary O'Reilly (pers. comm.). To sequence the insert a shotgun sequencing strategy was undertaken (Sambrook *et al.*, 1989). CsCl purified pSAL16 plasmid was subjected to limited DNase I digestion in a buffer containing $MnCl_2$. Fragments of approximately 2-kbp were gel purified, blunt ended with Klenow and T4 DNA polymerases and dephosphorylated with bacterial alkaline phosphatase. The prepared fragments were ligated to *Hinc*II digested pUC19 cloning vector and aliquots of the reaction transformed into competent *E. coli* TG-1. White and pale blue transformants were picked from LB plates containing ampicillin and X-gal. PCR screening of colony lysates was carried out with M13 universal forward and reverse primers to determine the authenticity and size of the putative clone. Recombinant plasmids were sequenced with M13 universal forward and reverse primers.

2.5.2. Sequencing reactions

Sequencing reactions employed the Sanger dideoxy chain-terminating chemistry (Sanger *et al.*, 1977). Flouresently labelled universal forward or reverse M13 primers were used with either the Genpak PyroSeq(Genetix, New Milton, UK) or Thermosequenase (Amersham Biosciences, Uppsala, Sweden) kits following the instructions of the manufacturer. Alternatively, the Prism DyeTermintor kit from Perkin Elmer (Cambridge, UK) was used with custom primers. Completed reactions were purified and concentrated by ethanol

precipitation or Sephadex G-50 gel spin chromatography and electrophoresised through a 6% denaturing polyacrylamide gel on a ABI 373A automated sequencer (Applied Biosystems, Warrington, UK).

2.5.3. Sequence Analysis

Identification of open reading frames (ORFs) in the completed single contig generated from the sequencing of the pSAL16 insert was undertaken with the XNIP programme from the Staden package (Staden *et al.*, 2000) The nucleotide sequence was translated in all six reading frames for ORFs containing putative proteins of 100 amino acid or greater. Putative proteins identified by this *in silico* analysis were further analysed for possible frameshift errors by homology to proteins in the GenBank™ non-redundant database using the BlastP server (http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul *et al.*, 1997). For example, if two sequential putative proteins from the pSAL16 sequence matched a single database protein, the overlapping nucleotide sequence encoding the end and beginning of each protein was re-examined thoroughly. This generally identified the frameshift or otherwise the region in question was re-sequenced from a clone or PCR amplified product.

2.5.4. Sequence assembly

Sequence assembly was performed by imputing the ABI generated sequence data files into the XBAP programme from the Staden package (Staden, 1986). The XBAP programme created contigs from overlapping clones which were visually inspected to detect possible errors. Gap closure between the contigs was achieved by sequencing with custom designed oligonucleotide primers orientated outwards from the end of each contig using pSAL16 as the template.

2.6. PLASMID AND STRAIN CONSTRUCTION

Transcriptional fusions to the *bgaB* reporter gene for ectopic insertion at the *amyE* locus were generated using the plasmid pDL (Yuan and Wong, 1995b). PCR generated fragments contained *Eco*R1 and *Bam*H1 ends introduced during PCR were cloned into *Ecc* R1/*Bam* H1 digested pDL. Correct integration at the *amyE* locus by double cross-over using *Pst* I

linearised pDL constructs were verified by monitoring the Amy$^-$ phenotype on LB-agar containing 0.2% starch flooded with iodine solution (0.1% [wt/vol] iodine, 0.1% [wt/vol] KI_2) and by PCR. Double cross over integrations at the *amyE* locus were confirmed by PCR with primers AMY1 5'-GCTCGGGCTGTATGACTGG and AMY2 5'-CGTATCATGCGACTCTACCC in which the wild type *amyE* gene produces a 368-bp fragment but strains containing double cross over events at the *amyE* fail to produce an amplicon. Figure 2.1 shows a map of plasmid pDL and the integration of a pDL derived construct, pDN2, at the *amyE* locus.

To verify that plasmid integrations into the *B. subtilis* chromosome had occurred at the correct locus, Southern blot analysis, PCR or a combination of both methods were undertaken. For PCR analysis a set(s) of primers were chosen to generate a fragment(s) of a unique size(s) which confirmed the expected rearrangement at the integration site (Volker *et al.*, 1998; Pragai and Harwood, 2000).

Strain DN1. Strain DN1 contains plasmid pDL, without a promoter fragment inserted upstream of the promoterless *bgaB* reporter gene, integrated at the *amyE* locus of stain 168. This strain was used to measure the background level of BgaB and this level subtracted from all subsequent levels determined for other strains (see Table 4.3).

Strain DN2. Strain DN2 was constructed by cloning a 341-bp PCR-generated fragment (synthesised using primers YKDAP1 5'-GCGGATCCGATGATGAATGACATTGC-3' and YKDAP2 5'-GCGAATTCAGCTGTCTAGCGATCATATC-3' (the underlined sequences represent introduced restriction sites to facilitate cloning) into pDL to generate plasmid pDN2. Plasmid pDN2 was linearised and transformed into *B. subtilis* strain 168 with selection for chloramphenicol resistance yielding strain DN2.

Deletion derivatives of the *ykdA* promoter region were generated by *Bal*31 deletion of the insert in pDN2 from the 5' and 3' directions to generate plasmids pDN8 to pDN11, and pDN12 to pDN14 respectively. pDN2 linearised by *Eco*R1 digestion served as the template for 5' deletions while pDN2 digested with *Bam*H1 served as the template for 3' deletions. Deleted promoter fragments were recloned into pDL, after subcloning through pUC19, the

deletion end-points determined by sequencing and the appropriate deletion constructs were then integrated into the *amyE* locus of selected *B. subtilis* strains.

Strain DN3. Strain DN3 was constructed by cloning a 177-bp PCR amplified *ykdA* fragment synthesised using primers YKDA6 5'-GCGAATTCTAAACTCAAGTCATAAACCT-3' and YKDAP1, (see above strain DN2) containing the ribosome binding site and 5' part of the gene into *Eco*R1-*Bam*H1 digested pMUTIN4 to generate plasmid pDN3. Plasmid pDN3 was then integrated into the chromosome of *B. subtilis* strain 168 by a Campbell-type single cross over event with selection for erythromycin resistance to generate strain DN3. In this strain the *ykdA* promoter directs expression of the *lacZ* reporter gene and the *ykdA* structural gene is under the control of the P*spac* inducible promoter (see Figure 4.5)

Strains DN8 to -11. Strains DN8 to -11 were generated by transformation of strain DN2 with plasmids pDN8, pDN9, pDN10 and pDN11 respectively with selection for chloramphenicol resistance. Figure 4.4 shows the *ykdA* promoter deletion end points contained in plasmids pDN8 to pDN11.

Strains DN12 to -14. Strains DN12 to -14 were generated by transformation of strain 168 with plasmids pDN12, pDN13 and pDN14 respectively with selection for chloramphenicol resistance. Figure 4.4 shows the *ykdA* promoter deletion end points contained in plasmids pDN12 to pDN14.

Strains DN15 to -18. Strains DN15 to -18 were generated by transformation of strain DN25 with plasmids pDN8, pDN9, pDN10 and pDN11 respectively with selection for chloramphenicol resistance.

Strains DN19 to -21. Strains DN19 to -21 were generated by transformation of strain DN26 with plasmids pDN12, pDN13 and pDN14 respectively with selection for chloramphenicol resistance.

Strain DN25. Strain DN25 was constructed by cloning a 265-bp internal *ykdA* fragment amplified with primers YKDADEL2F 5'-AGAAGGGGCATCATCAC and YKDADEL2R 5'-TTGAAACCGTTCTGTCCAC into the *Eco*RV site of pMOR60 to generate plasmid pDN25. Plasmid pDN25 was transformed into *B. subtilis* strain 168 with selection for erythromycin resistance to generate strain DN25.

Strain DN26. Strain DN26 has a deletion in the *ykdA* gene that was generated using the method of Biswas *et al.* (Biswas *et al.*, 1993). Plasmid pDN26 was constructed by sequential cloning of two *ykdA* fragments synthesised using the primers YKDA6 and YKDAP1 (see above strain DN2) [fragment 1] and primers YKDADEL2R and YKDADEL2F (see above strain DN25) [fragment 2] into pGhost4^{+}. The fragments were juxtaposed in this plasmid and in the same orientation as they existed on the *B. subtilis* chromosome. *B. subtilis* strain 168 was transformed with pDN26 and erythromycin resistant transformants selected. Excision of the chromosomal DNA between the two homologous fragments resulted in strain DN26 (which has a 439-bp deletion within the *ykdA* gene). The chromosomal rearrangement was confirmed by PCR and Southern blot analysis.

Strain DN27. Strain DN27 was made by transforming strain DN26 with plasmid pDN2 and selecting for chloramphenicol resistance.

Strain DN28 Strain DN28 was created by transforming strain DN3 with chromosomal DNA from strain BT01 with selection for chloramphenicol and erythromycin resistance.

Strain DN29 Strain PB105 was transformed with plasmid pDN2 with selection for chloramphenicol resistance to create strain DN29.

Strain DN30 To create strain DN30 plasmid pDN29 was constructed in two steps. First a 466-bp fragment containing the *ctsR* gene was PCR amplified using primers CTSRF 5'-CGGGATCCGTCAAAGTCAGTAAAGGAGG and CTSRR 5'-CGGGATCCTTGCCTTCATCATTCTCGC digested with *Bam* HI and cloned into *Bam*H1 digested pUC19 to create plasmid pDN28. Next a *Bcl* I fragment from pDG646 containing the Erm cassette was inserted into the unique *Bcl* I site of plasmid pDN28 to construct plasmid pDN29. Strain DN2 was transformed with plasmid pDN29 linearised by digestion with *Eco* RI and *Pst* I, with selection for erythromycin resistance to generate strain DN30. The correct double crossover event at the *ctsR* locus in strain DN30 was confirmed by PCR.

Strain DN32 Strain DN32 was created in a number of steps. Plasmid pDN31 was constructed by subcloning the deleted ykdA promoter fragment [5'Δ79] from pDN10 into EcoR1/BamH1 digested pUC19. The spectinomycin cassette from plasmid pDIA5324 obtained as a *Sal* I fragment was cloned into pDN31 to create pDN32. Finally the Pspac promoter element from pDE was cloned as a *Kpn* I-*Sac* I fragment, after subcloning through

pBluescript II, into pDN32 to create pDN33. Transformation of B. subtilis 168 with plasmid pDN33 generated strain DN32. The correct integration of plasmid pDN33 by a Campbell-type single cross over event at the *ykdA* locus of strain DN2 was confirmed by PCR.

Strain DN33 Strain DN33 was generated by transforming strain DN32 with chromosomal DNA from strain DN2 and selecting for chloroamphenicol and erythromycin resistance.

Strain DN34 Strain DN34 was generated by transforming strain DN26 with chromosomal DNA from strain ES993 with selection for chloramphenicol resistance. Retention of the markerless deletion (*ykdA*Δ439) at the *ykdA* locus was confirmed by PCR.

Strain DN40 Strain DN40 was generated by transforming wild type strain 168 with plasmid pDN40 and selecting for erythromycin resistance. Plasmid pDN40 was created by PCR amplifying a 326-bp internal fragment of *yvtA* using primers YVTAB1 5'-ACGCGTCGACCATGTCGTTGAAAGGCGCG and YVTAB2 5'-ACGCGTCGACCGGATTGATCGCTGCATC (see above) and ligating into the *Sal* I site of pDE.

STRAIN DN41. Transformation of strain 168 with plasmid pDN41 with selection for erythromycin created strain DN41. Plasmid pDN41 was constructed by cloning a 191-bp PCR amplified internal fragment of *yycK* using primers YYCK1 5'-AAGGGTCGACTGTTTGCAGGACTTCAGCG and YYCK2 5'-TGGGTCGACAAGCAGAGTCGCTGATTTCGG into the *Sal* I site of pDE.

Strain DN50 Strain DN50 was generated by transforming strain BFSA59 with plasmid pDN2 with selection for erythromycin and chloramphenicol. resistance.

Strain DN51 Strain DN51 was generated by transforming strain DN2 with plasmid pDN51 with selection for erythromycin resistance. Plasmid pDN51 was constructed by cloning a 161-bp internal fragment of *sigW*, PCR amplified with primers YBBLR3 5'-CCGGAAATTTTCAACTTGGC and YBBLR3 5'-GCATCTTCAGGCAAAACACC, into *Eco*RV digested pMOR60.

Strain DN52 Strain DN52 was generated by transforming strain DN2 with plasmid pDN52 with selection for erythromycin.. Plasmid pDN52 was constructed by cloning a 313-bp internal fragment of *sigX*, PCR amplified with primers SIGXF 5'-

GAAGACCTTCTTCAAGAGG and SIGXR 5'-GAATATCCCTGAATAAACCG, into *Eco*RV digested pMOR60.

Strain DN53 Strain DN53 was generated by transforming strain HB4245 (*sigyY*::Erm) with plasmid pDN2 and selecting for chloramphenicol resistance.

Strain DN54 Strain DN54 was constructed by transformation of strain DN2 with plasmid pDN54. Plasmid pDN54 was constructed as follows; primers YHDMDEL1 5'-TCCGACGCGTAGTCTGTTTGAAGCCGC and YHDMDEL2 5'-TCCGACGCGTCTGTTTCGATTGCGTTCC were used to amplify a 1773-bp fragment containing the *yhdM* coding sequence. The amplicon was digested with *Hpa*II, the extreme 5' and 3' fragments gel purified, ligated and reamplified with the YHDM primers. The desired Δ*yhdM* fragment was then double digested with *Hpa* II and *Mlu* I and ligated to *Cla* I digested pMOR60 to generate pDN54. This construct was sequenced to verify its integrity, an aliquot linearised with *Mlu* I and transformed into strain DN2. Transformants were selected for resistance to erythromycin and the correct double crossover event at the *yhdM* locus verified by Southern blot analysis.

Strain DN55 Strain DN55 was constructed by transformation of strain DN2 with plasmid pDN55. Plasmid pDN55 was constructed as follows; primers YLACDEL1 5'-TCCGACGCGTATGGAGAAACACAAGAGG and YLACDEL3 5'-TCCGACGCGTGATAAGCCTCCTCAACG were used to amplify a 994-bp fragment containing the *ylaC* coding sequence. The amplicon was digested with *Hpa* II, the extreme 5' and 3' fragments gel purified, ligated and reamplified with the YLAC primers. The desired Δ*ylaC* fragment was then double digested with *Hpa* II and *Mlu* I and ligated to *Cla* I digested pMOR60 to generate pDN55. This construct was sequenced to verify its integrity, an aliquot linearised with *Mlu* I and transformed into strain DN2. Transformants were selected for resistance to erythromycin and the correct double crossover event at the *ylaC* locus verified by Southern blot analysis.

STRAIN DN110. Strain DN110 was generated by transforming strain 168 with plasmid pDN110 and selecting for chloramphenicol resistance. Plasmid pDN110 was constructed by cloning a 263-bp PCR-generated fragment, synthesised with primers YVTAPF 5'-

GGAATTCGGCTCTTCACATCCTTTCAACG-3' and YVTAPR 5'-CGGGATCCTAACGGTTATTCATTTATCG-3', into pDL.

STRAIN DN111. To generate strain DN111, a 326-bp fragment of the *yvtA* gene was amplified by PCR, synthesised with primers YVTAB1 5'-ACGCGTCGACCATGTCGTTGAAAGGCGCG-3' and YVTAB2 5'-ACGCGTCGACCGGATTGATCGCTGCATC-3' and gel purified. The purified fragment was digested with *Hin*dIII, yielding two fragments that were subsequently ligated to a *Hin*dIII fragment from pDG780 that encoded the kanamycin resistance gene. An aliquot of the ligation mixture was transformed into strain 168 with selection for kanamycin resistance. That integration had occurred by a double-crossover into the *yvtA* gene was confirmed by PCR.

STRAINS DN112 AND DN113. Strains DN112 and DN113 were generated by transforming strains DN111 and DN26 respectively with plasmid pDN110 and selecting for chloramphenicol resistance.

Strain DN114. Strain DN114 was created by transforming strain DN111 with plasmid pDN2 and selecting for chloramphenicol resistance.

Strain DN115. The double mutant strain DN115 was generated by transforming strain DN26 with chromosomal DNA of strain DN111 and selecting for kanamycin resistance. Confirmation that strain DN115 contained a *ykdA* null allele (*ykdA*Δ439) was confirmed by PCR amplification using primers YKDA6 and YKDADEL2F.

Strain DN116. Strain DN116 was generated by transformation of strain DN115 with plasmid pDN2 and selecting for chloramphenicol resistance.

Strain DN200. Strain DN200 was generated by transforming strain AH24 with chromosomal DNA from strain DN111, selecting for kanamycin, chloramphenicol and erythromycin resistance.

Strain DN201 and DN203. Strains DN201 and DN203 were generated by transformation of strain KS408 with chromosomal DNA of strains DN40 and DN41 respectively and selecting for erythromycin and chloramphenicol resistance in each transformation.

Strain AH22. Strains AH22-AH24 were constructed by Alistair Howell as an undergraduate student under my supervision. Strain AH22 has a 570-bp deletion at the *ykdA* locus, extending from upstream of the active site serine residue to beyond the transcriptional terminator. Strain AH22 was generated by the method of Biswas *et al.* (Biswas *et al.*, 1993) using plasmid pAH22. Plasmid pAH22 was generated by cloning two fragments into pGhost4$^+$: fragment 1 was synthesised using primers AL3 5'-CGGGAATTCGAGGTTACTGCAAAGCTG-3' and AL4 5'-AAAACTGCAGGCTGCGTCTG TCTGAATG-3' and fragment 2 was generated using primers AL5 5'-AAAACTGCAGCTTTCTGACGAGATATCCG-3' with AL6 5'-CCCAAGCTTGCGTTATAT TGGGGGCG-3'. Transformation of pAH22 into *B. subtilis* strain 168, followed by excision according to the method of Biswas *et al.* (Biswas *et al.*, 1993) resulted in the generation of strain AH22.

Strain AH23. Strain AH23 was constructed by transforming strain AH22 with plasmid pDN2 and selecting for chloramphenicol resistance.

Strain AH24. Strain AH24 contains a *ykdA* allele that has the active site serine residue of the protease mutated to a methionine. This was achieved by integration of plasmid pAH24 into the chromosome of AH23. Plasmid pAH24 was constructed as follows: a DNA fragment was amplified using primers AL3 (see above) and AL2 5'-GGAATTCGCCcaTaTgACCTGGATTAATTGCTG-3' and was cloned into *Eco*R1 digested pMOR60 to generate pAH24'. A second 819-bp fragment was synthesised using primers AL6 (see above) and AL1 5'-GGAATTCGGTcAtAtgGGCGGTCCTTTGTTAA-3' and was cloned into *Nde* I *Hin* dIII digested pAH24' to generate pAH24 (the lower case letters within primer sequences are the introduced site directed mismatches that create an *Nde* I site and alter the active site). The complete insert in pAH24 was sequenced to verify that the construct contained only the desired site directed mutations. The plasmid pAH24 was then transformed into strain AH23 by a Campbell-type event to generate strain AH24, thereby reconstituting an intact *ykdA* gene containing the desired point mutations. The correct integration was confirmed by PCR analysis. The construction of strain AH24 is schematically summarised in figure 4.7.

Strains RC010 and RC011. Strains RC010-RC011 were constructed by Ross Collery as an undergraduate student under my supervision. Strains RC010 and RC011were generated by transforming strains KS408 and KS405b with chromosomal DNA of strain DN3 with selection for erythromycin and chloramphenicol resistance in each transformation.

Strain BFS1807. An 251-bp internal fragment of the *ykfA* gene, PCR amplified using primers YKFAF 5'-TCAGCACGGATATTGGC-3' and YKFAR 5'-TTTGATAAGGCTCCTCG-3', was blunt-end cloned into the *Hinc*II site of pUC19. A 211-bp subfragment was released by digestion with *Hin*dIII and cloned into *Hin*dIII digested pMUTIN1 to generate plasmid pDU1807. Transformation of strain 168 with plasmid pDU1807 applying selection for erythromycin created strain BFS1807.

Strain BFS1808. An 432-bp internal fragment of the *ykfB* gene, PCR amplified using primers YKFBF 5'-ACCATGTGTTGAAGCCG-3' and YKFBR 5'-GCATCCTCCATTTTCCG-3', was blunt-end cloned into the *Hinc*II site of pUC9. This same fragment was released by digestion with *Hin*dIII and *Bam*HI and cloned into *Hin*dIII-*Bam*HI digested pMUTIN1 to generate plasmid pDU1808. Transformation of strain 168 with plasmid pDU1808 applying selection for erythromycin created strain BFS1808.

Strain BFS1809. An 278-bp internal fragment of the *ykfC* gene, PCR amplified using primers YKFCF 5'-AAGGTACTTGTGACGGC-3' and YKFCR 5'-TTCCCCAAAAACGGTCG-3', was blunt-end cloned into the *Hinc*II site of pUC19. A 138-bp subfragment was released by digestion with *Hin*dIII and cloned into *Hin*dIII digested pMUTIN1 to generate plasmid pDU1809. Transformation of strain 168 with plasmid pDU1809 applying selection for erythromycin created strain BFS1809.

Strain BFS1810. An 347-bp internal fragment of the *ykfD* gene, PCR amplified using primers YKFDF 5'-CCTTCAACCGCAAACTGC-3' and YKFDR 5'-TCTCTTTTTGAAGCCGC-3', was blunt-end cloned into the *Hinc*II site of pUC9. This same fragment was released by digestion with *Hin*dIII and *Bam*HIand cloned into *Hin*dIII-*Bam*HI digested pMUTIN1 to generate plasmid pDU1810. Transformation of strain 168 with plasmid pDU1810 applying selection for erythromycin created strain BFS1810.

2.7.TRANSCRIPTIONAL ANALYSES

2.7.1. RNA Isolation.

Total RNA was prepared from *B. subtilis* cells during normal growth and during heat shock as follows: 10 ml aliquots were harvested at designated times, centrifuged for 1 minute (8,000 rpm) at 4°C and cell pellets snap-frozen in a dry ice/ethanol bath. Pellets were either stored at -80°C or processed immediately. The cell pellet was resuspended in 0.2 ml sterile water and transferred to a 2 ml screw cap tube containing 1 ml of a guanidine thiocyanate/acid phenol mixture (TRI reagent, Sigma, St Louis MO), 80 μl hexadecyltrimethylammonium bromide (CTAB) and 0.5 mg of zirconium/silica beads (0.1 mm, Biospec, Bartlesville, OK) . The contents were shaken with three 1-minute pulses in a minibead beater (Biospec) or three 45s pulses in a Ribolyser (Bio101). The tube was cooled for 1 minute on ice between each pulse. Chloroform (0.2 ml) was then added, mixed and the tubes were centrifuged. The upper aqueous layer was removed and extracted with 0.5 ml chloroform followed by a second extraction with an equal volume of acid phenol/chloroform pH 4.7 (Sigma). The RNA was then precipitated with an equal volume of isopropanol and 1/10 volume 3M sodium acetate (pH 4.7). The RNA concentration was determined by measuring the A_{260} and aliquots stored at -70°C.

2.7.2. Northern Blot Analysis.

Twenty five microgram aliquots of total RNA was electrophoresed through an agarose gel (1.2% w/v) containing 2.2M formaldehyde according to standard methods (Sambrook *et al.*, 1989). Separated RNA was transferred to a positively charged nylon membrane (Pall Gelman, Ann Arbor MI) by capillary blotting overnight. The membrane was UV treated (150mJ) to fix the RNA and prehybridised for 2 hrs in high SDS buffer containing blocking reagent (Boehringer, Mannheim, Germany). Hybridisation was performed overnight in the same solution containing a heat denatured 1046-bp digoxigenin labelled *ykdA* DNA probe generated using the synthetic primers YKDA6 5'-CGGAATTCTAAACTCAAGTCATAAACCT-3' and RE1 5'-CCCAAGCTTTTTGACGTCATTGCTTGG-3' using a PCR DIG probe synthesis kit (Boerhinger, Mannheim, Germany) according to the manufacturer's instructions. Final stringency washes were performed in 0.5X SSC, 0.1% SDS at 68°C, twice for 15 minutes

each. Transcripts were visualised with the digoxigenin luminescent detection kit (Boehringer, Mannheim, Germany).

2.7.3. Primer Extension Analysis.

2.7.3.1. Radiolabelling of oligonucleotides

Oligonucleotide primers were 5'- end labelled in a final reaction volume of 20μl containing 20pmoles probe, 2μl γ ^{32}P ATP (6,000 Ci/mmol), 1x T4 polynucleotide kinase buffer and 2μl T4 polynucleotide kinase (Promega Corporation, Madison, WI, USA). The reaction mixture was incubated at 37°C for 30 minutes and terminated by heating to 70°C for 15 minutes.

2.7.3.2. Primer Extension.

Primer extension analysis was performed using 25μg total RNA isolated from cells harvested at appropriate times. The RNA was added to 2pmol of radioactively labelled primer and heated to 90°C for 2 minutes. For *ykdA*, the primer YKDART3 (5'-CCTTTCGTTCTGTTTTCATCACG-3') was used; for *yvtA* the primer YVTART2 (5'-CGCATGTACAAGTTCAATTGTCC-3') or primer YVTART1 (5'-GAATGTCCAATCAGCTTCTG G- 3') was used. Following denaturing the mixture was annealed for 30 minutes at 55°C in 15μl 1X Superscript II buffer (Life Technologies, Paisley, UK). The reaction was then cooled to 48°C and 5μl of a prewarmed solution containing 2mM dNTP's, 4X Superscript II buffer, 20mM DTT and 200 units Superscript II was added. This reaction was incubated at 48°C for 45mins. The reactions were ethanol precipitated overnight at -20°C, the pellet resuspended in 50% 10mM Tris-HCl, 1mM EDTA pH8.0, 50% stop solution (Amersham Biosciences, Uppsala, Sweden), denatured and electrophoresed through a 6% denaturing polyacrylamide gel containing a sequencing reaction generated with the same primers using plasmid pDN27 for *ykdA* or a PCR product containing the *yvtA* promoter region as templates.

2.8. MEASUREMENT OF β-GALACTOSIDASE ACTIVITY.

The BgaB reporter enzyme is suited for the measurement of heat shock gene induction because of its thermostability (Hirata *et al.*, 1984; Hirata *et al.*, 1985). It is also ideal for the measurement of low amounts of target gene transcriptional activity as endogenous β–galactosidase can may destroyed by heat treatment (70°C) (Schrogel and Allmansberger,

1997). The heat treatment of lysates will also destroy LacZ enzymatic activity so that a *bgaB* fusion construct can be introduced into a strain already containing a *lacZ* fusion, hence only BgaB activity will remain measurable (Yuan and Wong, 1995b).

BgaB activity was measured as previously described (Hirata *et al.*, 1985) with the following modifications: cells were lysed for 25 minutes at 37°C in Z buffer (Miller, 1972) containing 25mM β-mercaptoethanol, 100 μg/ml lysozyme, and 10 μg/ml DNase. Lysates were heat treated at 70°C for 15 minutes and spun at 12,500 rpm for 5 minutes. Aliquots of the supernatants were reserved for protein concentration determination. Aliquots (typically ranging from 15 to 730 μl depending on activity) of the supernatant were added to reactions containing o-nitrophenyl-β-D-galactopyranoside (ONPG) as substrate and incubation was carried out at 55°C. The reaction was stopped by addition of 1.2M $NaCO_3$. The optical density of the reaction was read at 420 nm. Measurement of LacZ activity was performed as outlined above except that dithiothreitol (1mM) replaced β-mercaptoethanol in the Z buffer and the incubation at 70°C was omitted. Incubation temperature for the reaction was 28°C. Both BgaB and LacZ assays were performed with duplicate samples from each time point. Protein concentration was determined using the Bio-Rad microassay (Bio-Rad, Hercules, CA) following the instructions of the manufacturer. One activity unit is defined as 1 nanomole of ONPG hydrolysed per minute per milligram of protein.

BgaB and LacZ specific activities were expressed in two ways using the following formulae

a) units/mg protein /min

$$\frac{A_{420nm} \times 1,5}{\text{mg/ml of proteins in sample} \times \text{ml of sample (0.2)} \times t \times 0.00486}$$

(http://locus.jouy.inra.fr/genmic/madbase/mutant/galang.html.)

For LacZ assays undertaken in microtitre format, an additional multiplication factor of 2.2 was added to the top line of the equation. 2.2 is the correction factor for reading samples in a microtitre plate rather than a cuvette. It was determined from calibration curves of samples measured in microtitre plates compared to those measured in cuvettes. This accounts for the difference in pathlength between absorbance measurements in microtitre plates and

absorbance measurements in cuvettes (Kasper Andersen, pers. comm). 1.5 is the correction factor for the dilution with Na_2CO_3. The molar extinction coefficient of ONPG is 4860.

b)

$$\text{Activity in Miller units} = \frac{A420 \times 222.2}{\text{mg protein} \times \text{volume sampled in ml} \times \text{mins}}$$

mg protein calculated from OD_{550};

μg protein / ml = OD_{550} x 83

$$\text{mg protein} = \frac{\mu\text{g protein}}{\text{ml}} \times \frac{\text{ml used in assay}}{1000}$$

this formula was obtained from the laboratory of Dr. Jim Hoch.

2.9. ROUTINE SEQUENCE ANALYSIS.

Routine analysis of DNA and protein sequences for restriction mapping, identification of ORFs, RNA structural analysis or for primer design employed the commercially available sequence analysis software programmes Genejockey II (Biosoft, Cambridge, UK), the GCG package (ALIGN, MAP, RNAFOLD, FINDPATTERNS) (Accelrys, Cambridge, UK) (Womble, 2000), MacVector (Accelrys, Cambridge, UK), PsiPred (http://bioinf.cs.ucl.ac.uk/psipred/) (McGuffin *et al.*, 2000); Mfold (http://bioinfo.math.rpi.edu/~mfold/rna/form1.cgi) (Mathews *et al.*, 1999), SMART (Schultz *et al.*, 1998), TREEVIEW (Page, 1996).

TABLE 2.1. Bacterial strains and plasmids used in this study.

Strain or Plasmid	Relevant characteristics	Source, or reference[a]
E.coli strains		
TG-1	*supE hsd*Δ5 *thi* Δ(*lac-proAB*) F'[*traD*36 *proAB*$^+$ *lacI* ^{q}*lacZ*ΔM15]	(Sambrook *et al.*, 1989)
B.subtilis strains		
168	*trpC* 2	(Anagnostopoulos and Spizizen, 1961)
BT01	1012 *hrcA*::cat	(Schulz *et al.*, 1995)
PB105	*sigB*Δ1	(Kalman *et al.*, 1990)
ES993	*trpC2 amyE*::P*ahpCF lacZ* Cm^r	E. Scanlan unpublished data
HB4245	*pheA*1 *trpC* 2 *sigY::*erm	(Huang and Helmann, 1998)
PB198	*trpC* 2 *amyE* ::(*ctc-lacZ*)	(Boylan *et al.*, 1992)
BFS59	*trpC* 2 *sigV* ::pHV30227 (*sigV-lacZ Pspac-sigV*)	Ehrlich S.D., unpublished data
BFS1807	*trpC* 2 *ykfA*::pDU1807 (*ykfA-lacZ*) Erm^r	pDU1807 → 168
BFS1808	*trpC* 2 *ykfB*::pDU1808 (*ykfB-lacZ*) Erm^r	pDU1808 → 168
BFS1809	*trpC* 2 *ykfC*::pDU1809 (*ykfC-lacZ*) Erm^r	pDU1809 → 168
BFS1810	*trpC* 2 *ykfD*::pDU1810 (*ykfD-lacZ*) Erm^r	pDU1810 → 168
DN1	*trpC2 amyE: :* pDL Cm^r	pDL → 168
DN2	*trpC2 amyE*::P$_{ykdA}$-*bgaB* Cm^r	pDN2 → 168

Strain or Plasmid	Relevant characteristics	Source, or reference[a]
DN3	*trpC2 ykdA*::pDN3 (P_{ykdA}-*lacZ* P_{spac}-*ykdA*) Erm^r	pDN3 → 168
DN8 to DN11	*trpC2 amyE*::pDN8 to 11(5'ΔP_{ykdA}-*bgaB*) Cm^r	pDN8,9,10,11 → 168
DN12 to DN14	*trpC2 amyE*::pDN12 to14 (3'ΔP_{ykdA}-*bgaB*) Cm^r	pDN12,13,14 → 168
DN15 to 18	*trpC2 ykdA*::pMOR60 *amyE*::pDN8 to 11(5'ΔP_{ykdA}-*bgaB*) Erm^r Cm^r	pDN8,9,10,11 → DN25
DN19 to 21	*trpC2 ykdA*Δ439 *amyE*::pDN12 to 14 (3'ΔP_{ykdA}-*bgaB*) Cm^r	pDN12,13 14 → DN26
DN25	*trpC2 ykdA*::pMOR60 Cm^r	pDN25 → 168
DN26	*trpC2 ykdA*Δ439	pDN26 → 168
DN27	*trpC2 ykdA*Δ439 *amyE*::P_{ykdA}-*bgaB* Cm^r	pDN2 → DN26
DN28	*trpC2 ykdA*::pDN3 (P_{ykdA}-*lacZ* P_{spac}-*ykdA*) *hrcA*::cat Erm^r Cm^r	BT01 → DN3
DN29	*trpC2 sigB*Δ2*ykdA*::pDN3 (P_{ykdA}-*lacZ* P_{spac}-*ykdA*) Erm^r	pDN3 → PB105
DN30	*trpC2 amyE*::P_{ykdA}-*bgaB ctsR*::*erm* Cm^r Erm^r	pDN29 → DN2
DN32	*trpC2 ykdA*::pDN23 (Pspac-*ykdA* [*lacI* -]) Sp^r	pDN32 → 168
DN33	*trpC2 ykdA*::pDN23 (Pspac-*ykdA* [*lacI* -]) *amyE*::P_{ykdA}-*bgaB* Sp^r Cm^r	pDN2 → DN31
DN34	*trpC2 ykdA*Δ439 *amyE*::P*ahpCF lacZ* Cm^r	ES993 → DN26
DN40	*trpC2 yvtA*::pDN40 *(yvtA'-bgaB)* Erm^r	pDN40 → 168
DN41	*trpC2 yycK*::pDN41 (*yycK'-bgaB)* Erm^r	pDN41 → 168
DN50	*trpC2 sigV* ::pMutin(P*sigV-lacZ* P*spac-sigV*) *amyE*::P_{ykdA}-*bgaB* Cm^r Erm^r	pDN2 → BFA59

Strain or Plasmid	Relevant characteristics	Source, or reference[a]
DN51	*trpC2* *sigW*::pMOR60 *amyE*::P_{ykdA}-*bgaB* Cm^r Erm^r	pDN51 → DN2
DN52	*trpC2* *sigX*::pMOR60 *amyE*::P_{ykdA}-*bgaB* Cm^r Erm^r	pDN52 → DN2
DN53	*trpC2* *sigY* ::pMOR60 *amyE*::P_{ykdA}-*bgaB* Cm^r Erm^r	pDN2 → HB4245
DN54	*trpC2* Δ*yhdM*::pMOR60 *amyE*::P_{ykdA}-*bgaB* Cm^r Erm^r	pDN54 → DN2
DN55	*trpC2* Δ*ylaC*::pMOR60 *amyE*::P_{ykdA}-*bgaB* Cm^r Erm^r	pDN55 → DN2
DN110	*trpC* 2 *amyE*::P_{yvtA}-*bgaB* Cm^r	pDN110 → 168
DN111	*trpC2* *yvtA*::kan Km^r	'*yvtA*'::kan → 168
DN112	*trpC2* *yvtA*::kan *amyE*::P_{yvtA}-*bgaB* Km^r Cm^r	pDN110 → DN111
DN113	*trpC2* *ykdA*Δ439 *amyE*::P_{yvtA}-*bgaB* Km^r	pDN110 → DN26
DN114	*trpC2* *yvtA*::kan *amyE*::P_{ykdA}-*bgaB* Km^r Cm^r	pDN2 → DN111
DN115 *sup* [b]	*trpC2* *ykdA*Δ439 *yvtA*::kan *sup* Km^r	DN111 → DN26
DN116 [b]	*trpC2* *ykdA*Δ439 *yvtA*::kan *amyE*::P_{ykdA}-*bgaB sup* Km^r Cm^r	pDN2 → DN115 *sup*
DN200	*trpC2* *ykdA*1[N289H S290M] *yvtA*::kan *amyE*::P_{ykdA}-*bgaB* Km^r Cm^r Erm^r	DN111 → AH24
DN201	*trpC2* *amyE* *xylR*::P*xyl--amyL* *yvtA*::pDN40(*yvtA'-bgaB)* Cm^r Erm^r	DN40 → KS408
DN203	*trpC2* *amyE* *xylR*::P*xyl--amyL* *yycK*::pDN41(*yycK'-bgaB*) Cm^r Erm^r	DN41 → KS408
AH22	*trpC2* *ykdA*Δ570	pAH22 → 168
AH23	*trpC2* *ykdA*Δ570 *amyE*::(P_{ykdA}-*bgaB*) Cm^r	pDN2 → AH22

Strain or Plasmid	Relevant characteristics	Source, or reference[a]
AH24	*trpC2 ykdA*::pAH24(*ykdA*1 [N289H S290M]) *amyE*::P_{ykdA}-*bgaB* Erm^r Cm^r	pAH24 → AH23
KS408	*trpC2 amyE xylR*::P*xyl--amyL* Cm^r	(Stephenson and Harwood, 1998)
KS405b	*trpC2 amyE xylR*::P*xyl-amyLQS* Cm^r	(Stephenson *et al.*, 1998)
RC010	*trpC2 amyE xylR*::P*xyl--amyL* P*spac-ykdA* P_{ykdA}-*lacZ* Cm^r Erm^r	DN3 → KS408
RC011	*trpC2 amyE xylR*::P*xyl-amyLQS* P*spac-ykdA* P_{ykdA}-*lacZ* Cm^r Erm^r	DN3 → KS405b
Plasmids		
pDL	integration vector for the introduction of single copy transcriptional fusions to *bgaB* by double cross over at the *amyE* locus (Ap^r Cm^r)	(Yuan and Wong, 1995)
pMUTin4	integration vector containing the promotorless *E. coli lacZ* gene and the inducible Pspac promotor (Ap^r Erm^r)	(Vagner *et al.*, 1998)
pMOR60	pBluescript SKwith the erythromycin cassette from pGhost4+ inserted into *SacI*/*Sma*I allowing Campbell-type integration into the *B. sutbilis* chromosome (Ap^r Erm^r)	M O'Reilly unpublished

Strain or Plasmid	Relevant characteristics	Source, or reference[a]
pGhost4+	vector containing Ts replicon for conditional integration and excision allowing construction of markerless deletions (Ap^r Erm^r)	(Biswas *et al.*, 1993)
pUC9	cloning vector (Ap^r)	(Vieira and Messing, 1982)
pUC19	cloning vector (Ap^r)	(Yanisch-Perron *et al.*, 1985)
pDG646	General purpose *B. subtilis* integration vector containing an erythromycin resistance cassette	(Guerout-Fleury *et al.*, 1995)
pDIA 5324	pUC18 containing a spectinomycin cassette in a modified polylinker	Philippe Glaser (Institute Pasteur, Paris), unpublished data
pDG780	general purpose *B. subtilis* integration vector containing a kanamycin resistance cassette	(Guerout-Fleury *et al.*, 1995)
pDIA5304	pBluescript II containing a *cat* gene inserted at the unique *Nae* I site	(Glaser *et al.*, 1993)
pSAL16	pDIA5304 containing a fragment of the *B. subtilis* chromosome	Mary O'Reilly, unpublished data
pDU1807	pMUTIN containing an 211-bp internal fragment of the *ykfA* gene (Ap^r Erm^r)	This work
pDU1808	pMUTIN containing an 432-bp internal fragment of the *ykfB* gene (Ap^r Erm^r)	This work
pDU1809	pMUTIN containing an 138-bp internal fragment of the *ykfC* gene (Ap^r Erm^r)	This work

Strain or Plasmid	Relevant characteristics	Source, or reference[a]
pDU1810	pMUTIN containing an 347-bp internal fragment of the *ykfD* gene (Ap^r Erm^r)	This work
pDE	pMUTIN based integrating plasmid with the *lacZ* gene replaced with *bgaB* (Ap^r Erm^r)	E. Deuerling unpublished data
pDN2	pDL containing the full *ykdA* control region on a PCR amplified fragment (Ap^r Cm^r)	This work
pDN3	pMUTin4 with a PCR amplified fragment containing an inactive part of the *ykdA* promoter, ribosome binding site and the first 112-bp of the *ykdA* structural gene cloned into *Eco*R1-*Bam*H1 (Ap^r Erm^r)	This work
pDN8	pDL containing the *ykdA* control region with 47-bp deleted from the 5' end [5'Δ47] (Ap^r Cm^r)	This work
pDN9	pDL containing the *ykdA* control region with 55-bp deleted from the 5' end [5'Δ55] (Ap^r Cm^r)	This work
pDN10	pDL containing the *ykdA* control region with 79-bp deleted from the 5' end [5'Δ79] (Ap^r Cm^r)	This work
pDN11	pDL containing the *ykdA* control region with 103-bp deleted from the 5' end [5'Δ103] (Ap^r Cm^r)	This work
pDN12	pDL containing the *ykdA* control region with 33-bp deleted from the 3' end [3'Δ33] (Ap^r Cm^r)	This work

Strain or Plasmid	Relevant characteristics	Source, or reference[a]
pDN13	pDL containing the *ykdA* control region with 61-bp deleted from the 3' end [3'Δ61] (Ap^r Cm^r)	This work
pDN14	pDL containing the *ykdA* control region with 74bp deleted from the 3' end [3'Δ74] (Ap^r Cm^r)	This work
pDN25	pMOR60 containing a PCR amplified internal fragment from *ykdA* cloned into *Eco*R V (Ap^r Erm^r)	This work
pDN26	pGhost4+ containing juxtaposed 5'flanking and internal fragments fragments from *ykdA* cloned into *Eco*R1 /*Hin*dIII (Ap^r Erm^r)	This work
pDN27	pUC19 containing the *ykdA* control region cloned into the *Eco*R1+ *Bam*H1 sites (Ap^r)	This work
pDN28	pUC19 containing a 466bp PCR amplified fragment of *ctsR* cloned into *Bam*H1 site	This work
pDN29	*Bam*H1 fragment containing the *erm* cassette of pDG646 inserted into the unique *Bcl*1 site of pDN28	This work
pDN31	pUC19 containing the deletion fragment of the *ykdA* promoter from pDN10	This work
pDN32	pUC19 containing the deletion fragment of the *ykdA* promoter from pDN10 and the spectinomycin resistance cassette from pDIA5324	This work

Strain or Plasmid	Relevant characteristics	Source, or reference[a]
pDN33	pUC19 containing a deletion fragment of the *ykdA* promoter [5'Δ79] including the RBS, Pspac promoter and the spectinomycin cassette from pDIA5324 (Ap^r Sp^r)	This work
pDN40	pDE containing a 326bp PCR amplified internal fragment of *yvtA*	This work
pDN41	pDE containing a 191bp PCR amplified internal fragment of *yycK*	This work
pDN51	pMOR60 containing PCR amplified internal fragment of *sigW* (Ap^r Cm^r)	This work
pDN52	pMOR60 containing PCR amplified internal fragment of *sigX* (Ap^r Cm^r)	This work
pDN54	pMOR60 containing juxtaposed 5' and 3' *Hpa* II fragments flanking *yhdM* (Ap^r Cm^r)	This work
pDN55	pMOR60 containing juxtaposed 5' and 3' *Hpa* II fragments flanking*ylaC* (Ap^r Cm^r)	This work
pDN110	pDL containing the *yvtA* control region on a PCR amplified fragment (Ap^r Cm^r)	This work
pAH22	pGhost4+ containing PCR amplified juxtaposed fragments spanning the 3' end of *ykdA*, inserted into *Eco*R1 /*Hin*dIII sites.(Ap^r Erm^r)	This work

Strain or Plasmid	Relevant characteristics	Source, or reference[a]
pAH24	PCR amplified fragment of the *ykdA* gene incorporating the site directed mutations (N289H S290M) cloned into *Eco*R1/*Hin*dIII sites of pMOR60 (Ap^r Erm^r)	This work

[a] → indicates construction of the strain on the right of the arrow with the plasmid or chromosomal DNA on the left of the arrow.

[b] The suppressor mutation in these strains is uncharacterized.

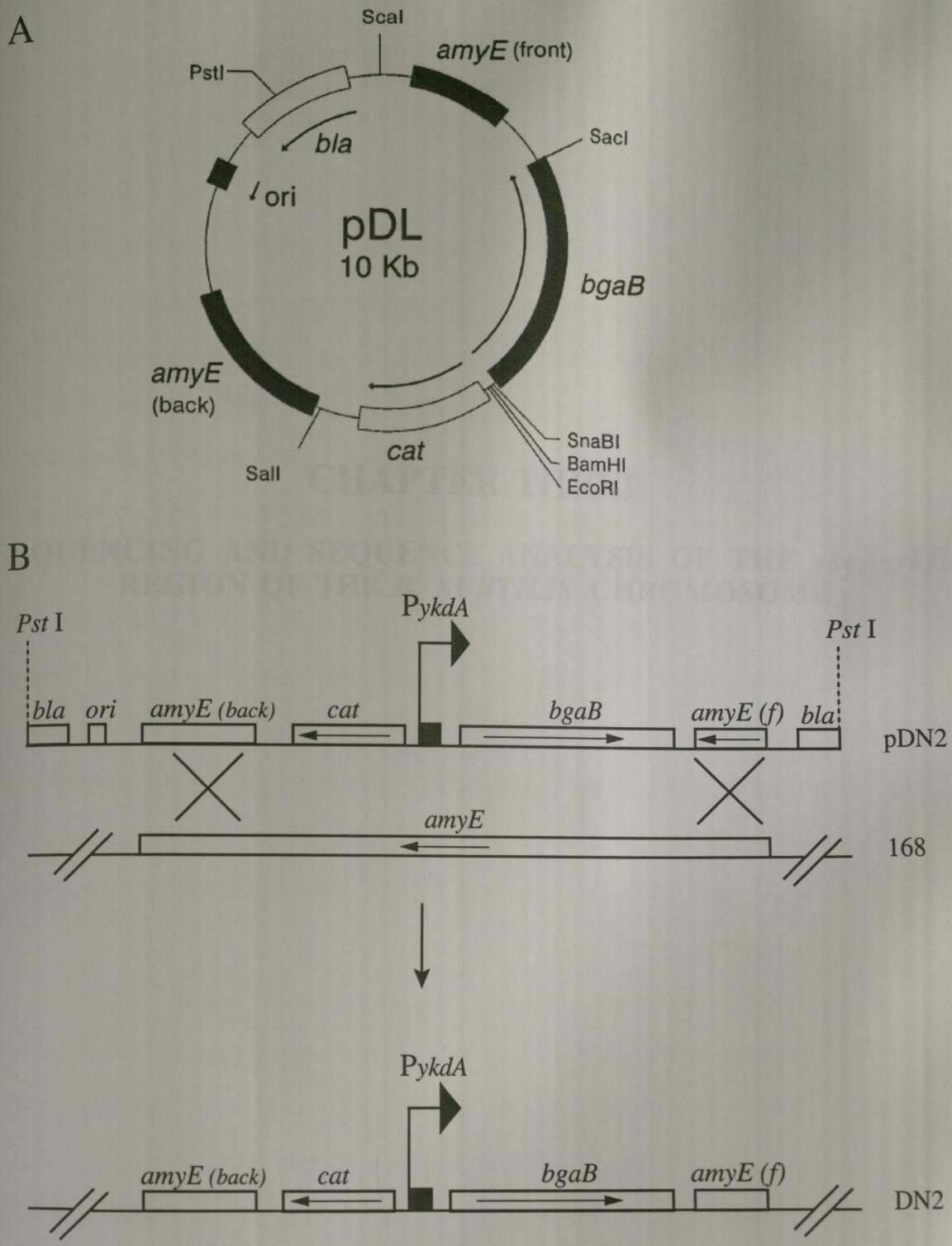


Figure 2.1. Generation of transcriptional fusions to *bgaB* using the integration vector pDL. (A) Map of pDL showing the promoterless *bgaB* gene, the *cat* gene for selection in *B. subtilis*, the front and back regions of the *amyE* gene and the *Eco*RI/*Bam*HI cloning sites. Arrows indicate direction of transcription. (B) Schematic representation of a double crossover event between a *Pst*I linearised pDL derived construct, pDN2, carrying the *ykdA* promoter (black box/bent arrow), and the *amyE* gene on the chromosome of *B. subtilis* 168 to generate strain DN2. Gene sizes are not to scale and the figure in (A) is from Yuan *et al.*, (1995b).

CHAPTER THREE

SEQUENCING AND SEQUENCE ANALYSIS OF THE *ykcA-ykfD* REGION OF THE *B. SUBTILIS* CHROMOSOME

3.1. INTRODUCTION

The advent of whole genome sequencing has truly impacted upon molecular biology. This is evidenced by the debate and degree of public awareness (and apprehension) that has resulted from the completion of the sequencing of the human genome. However, it was bacterial genome sequencing that paved the way for much of the technological and computational advances that facilitated the sequencing of much larger eukaryotic genomes. The first genome sequenced was that of the Gram-negative bacterial pathogen *Haemophilus influenzae* (Fleischmann *et al.*, 1995). Since then a plethora of microbial genomes, from both pathogens and model organisms, have been sequenced among them *B. subtilis*. Reasons for sequencing the genome of *B. subtilis* can be ascribed to its position as the *de facto* Gram-positive model bacterium and its importance as a producer of industrially useful compounds (Sonenshein *et al.*, 1993; Sonenshein *et al.*, 2002). Two major aspects of the physiology of *B. subtilis* have been responsible for its role as a model organism; (i) the ability to take up exogenous DNA (natural competence) with concomitant recombination into the host genome and (ii) the study of the developmental pathway of sporulation and other adaptational stress responses such as secretion of exoenzymes and antibiotics (see section 1.1). Many of the genes that are involved in these adaptational responses had already been cloned, characterised and mapped which presented a good frame of reference for the genome sequencing project. The pre-existing body of work generated an expertise and an array of genetic techniques, vectors and methodologies which would enable rapid analysis and exploitation of newly acquired sequence data (Harwood and Cutting, 1990). However, despite the knowledge that had accrued, many important aspects of the physiology of *B. subtilis* was still unknown. Genome sequencing of *B. subtilis* could provide many answers which, in addition, would be directly applicable to its role as an important industrial bacterium. Additionally, functional analysis of the *B. subtilis* gene products would be of benefit in assigning functions to homologues found in other closely related Gram-positive bacteria especially those in which genetic study was not as amenable.

Once the genome sequence of an organism is known it allows the conceptual translation of the open reading frames (ORF). This generates a list of putatively encoded proteins which

make up the proteome of the organism. The function of the encoded proteins may be known from previous genetic analysis or inferred from the high degree of similarity they show to proteins whose function is known in other bacterial species. However, for those proteins which exhibit only limited homology an experimental approach is required to analyse their functional activity. This was the case for about 2000 genes of *B subtilis* (Schumann *et al.*, 2001). To undertake the systematic analysis of these genes of unknown function (y-genes) a consortia of European and Japanese laboratories collaborated in the *Bacillus* Functional Analysis (BFA) project. The strategy followed was to generate mutants in each of the *y*-genes using an integration vector, pMUTIN. These integrations also generate a transcriptional *lacZ* fusion allowing a measurement of the expression level and profile of the gene of interest. Each of the participating laboratories generated mutants which were sent to a central depository at Jouy-en-Josas, France, from where all the mutant strains were sent to all laboratories for testing of certain phenotypic traits. This was to allow a tentative assignment of gene function to five broad groups: (i) metabolism of small molecules; (ii) macromolecule metabolism; (iii) cell structure and mobility; (iv) stress and stationary phase; (v) cell processes (e.g., cell cycle , competence, sporulation) (Ehrlich, 2001).

My involvement in the genome sequencing project resulted in the sequencing of an approximately 15-kbp fragment of the region between *ykcA* and *ykfD*. An analysis of this region is presented below. One of the ORFs, *ykdA*, was found to encode a protein which was homologous to the DegP(HtrA) protein of *E. coli.* A more in-depth analysis of this protein and its paralogues, *yvtA* and *yycK* (*yyxA*) is also presented. I also generated insertion mutants as a contribution to the functional analysis project. Finally, I participated in analysing the levels of expression of all BFA mutant strains in response to oxidative stress and in monitoring the oxidative stress phenotypes of these strains (Scanlan *et al.*, 2001).

3.2. RESULTS and DISCUSSION

3.2.1. Sequencing of pSal16

As part of our contribution to the *Bacillus subtilis* genome sequencing project the sequence of the region between *xre* and *fruAB* genes was determined. Various members of the laboratory undertook the sequencing of sub-regions contained within the larger section. I undertook the sequencing of a plasmid clone pSal16, containing a fragment of the *B. subtilis* genome. pSal16 (provided by Dr. Mary O'Reilly) was generated using standard plasmid rescue techniques (Dowds *et al.*, 1988) using the integrating plasmid pDIA5304 (Glaser *et al.*, 1993). A small segment of *ykcA* gene positioned at the end of the previously cloned fragment was inserted into pDIA5304 and integrated into the chromosome. Total chromosomal DNA from this strain was restricted with *Sal* I, ligated and cloned into *E. coli* strain TP611 (Glaser *et al.*, 1993). The chromosomal insert was ≈ 15-kbp. The whole of plasmid pSal16 was fragmented by partial DNase treatment and the resulting fragments were cloned into pUC19 as detailed in Materials and Methods. The sequence of inserts in the subsequent mini bank were then determined, assembled using the XBAP programme from the STADEN package, annotated and submitted to GenBank as part of entry NC000964. A physical map of the open reading frames determined from the sequencing of the *B. subtilis* chromosomal insert contained within the pSAL16 plasmid is shown in Figure 3.1.

3.2.1.1. Operon structure and open reading frames (ORFs)

The chromosomal insert of pSAL16 was searched for putative open reading frames (ORF) encoding proteins of ≥ 100 amino acids, and for rho-independent transcriptional terminators. These analyses lead to the putative operon structures shown in Figure. 3.1. An analysis of the sequences is presented in this section. In Table 3.1 the co-ordinates of the ORFs, their size in amino acids and the putative RBS preceding each ORF are shown. Table 3.1 also lists the putative function or similarity of the ORFs to other proteins. Table 3.2 presents an analysis of promoter structure using the FINDPATTERNS programme of the GCG sequence analysis package using the consensus sequences listed in the review by Haldenwang (1995). It should be noted that this promoter analysis is rather speculative and would require considerable experimental data to validate the *in silico* analysis. Table 3.3 lists sequences

located at the end of putative operons which are capable of forming stem loop structures and functioning as rho-independent terminators.

3.2.1.2. *ykcABC* operon. This operon is composed of three ORFs and ends with a putative strong rho-independent terminator (ΔG= -24.30kcal/mol) containing a 13-bp perfect inverted repeat including a run of six thymidine nucleotides (Table 3.3). Each gene in the operon contains sequences resembling ribosome binding sites (Shine-Delgarno boxes) (Table 3.2). There are four putative promoters preceding the operon, matching consensus sequences of the σ^D, σ^H, σ^L, σ^K controlled promoters. This spectrum of promoters, if real would suggest that the proteins play a role in motility and chemotaxis(σ^D), stationary phase(σ^H), nitrogen metabolism(σ^L) and sporulation(σ^K). Primer extension analysis would be required to verify these promoters.

ykcA encodes a protein of 316aa (amino acids). A similarity search using the BLAST programme indicate that the YkcA protein is homologous to ABC transporter binding proteins.

ykbB encodes a putative protein of 716 amino acids and is most similar to another *B. subtilis* ORF, YycA, and to a protein from the Gram-positive bacterium *Lactobacillus fermentum*. When analysed by the TMHMM and SIGNALP programmes YkcB is suggested to possess a cleavable signal peptide and 13 transmembrane domains. In keeping with this proposed localisation YkcB shows homology to the family of Dolichyl-phosphate-mannose-protein mannosyltransferase proteins. Enzymes of this family are generally membrane bound and play a role in the glycosylation of exported proteins or in maintenance of cell wall rigidity (Gentzsch and Tanner, 1996).

ykcC encodes a protein that is similar to proteins from *Synechocystis* sp. (strain PCC 6803) and other bacteria including *E. coli* which encode dolichol-phosphate mannose synthases. These enzymes are involved in sugar transfer associated with cell wall/peptidoglycan biosynthesis (Foster and Popham, 2002). In support of this putative function, the YkcC protein contains two putative C-terminal transmembrane domains which may anchor it to the cell membrane. The YkcC protein could there act in a preceding step to that of YkcB in the transfer of sugar residues to a lipid carrier intermediate.

3.2.1.3. *ykdA* operon. A more detailed analysis of this protein and its *B. subtilis* paralogues, *yvtA* and *yycK*(*yyxA*) is presented in section 3.2.3 below.

3.2.1.4. *ykeA* (*proG*) operon.

The *ykeA* operon consists of a single ORF *ykeA* which encodes a protein of 272 amino acids. The ORF is preceded by a sequence similar to a ribosome binding site (Table 3.1) and is followed by a structure capable of forming a rho-independent terminator, including a run of seven thymidine nucleotides, with a calculated free energy of -11.9 kcal/mole (Table 3.3). The upstream region contains sequences with similarity to motifs from σ^A, σ^B and σ^K controlled promoters. This spectrum of promoters would suggest expression during vegetative growth, under some stress conditions and during sporulation.

The encoded YkeA protein is similar to members of the pyrroline-5-carboxylate reductase family of enzymes. These enzymes perform the terminal step in the synthesis of proline from glutamate or arginine. *B. subtilis* has recently been shown to contain four potential pyrroline-5-carboxylate reductases, encoded by the *proH*, *proI*, *comER* and *ykeA*(*proG*) genes. Experimental genetic analysis has demonstrated that *proG*, *proH* and *proI* can individually allow the growth of *B. subtilis* in minimal media lacking proline in contrast to that of the *comER* gene product, implying a different function for ComER (Belitsky *et al.*, 2001). Two other genes involved in upstream steps of proline biosynthesis, *proAB* (co-ordinates 1377808- 1380155) are located about 18-kbp downstream of *ykeA*. This spatial separation of genes encoding pyrroline-5-carboxylate reductases and genes involved in the proline biosynthetic pathway also occurs in other bacteria (Kenklies *et al.*, 1999).

3.2.1.5. *dppABCDE* operon. This operon has previously been the subject of detailed expression and functional analysis. The operon encodes a dipeptide transport system utilised during nutritional stress conditions and/or sporulation (Mathiopoulos *et al.*, 1991; Slack *et al.*, 1991). Two differences between the sequence obtained in this study and that previously deposited in GenBank were noted (DppE L4→V, L31→K). The *dppA* gene has recently

been shown to encode a D-aminopeptidase often associated with cell envelope reconstruction (Cheggour *et al.*, 2000). In addition the *dpp* operon has been identified by global transcriptional analysis to be subject to catabolite repression (Schumann *et al.*, 2001). This mode of regulation was previously established by more classical methods (Slack *et al.*, 1991). The *dpp* operon is negatively regulated by the global negative regulators AbrB and a novel GTP-binding repressor, CodY (Ratnayake-Lecamwasam *et al.*, 2001).

3.2.1.6. *ykfABCD* operon.

The *ykf* operon contains four ORFs, *ykfA* (234aa), *ykfB* (366aa), *ykfC* (296aa) and *ykfD* (327aa). Each of the ORFs is preceded by a ribosome binding site (Table 3.1). There are overlaps between the translational stop site of *ykfA* and start site of the *ykfB* ORFs and similarly between the *ykfB* and *ykfC* ORFs. The *ykf* operon contains putative sequences for σ^H, σ^L and σ^G type promoters (Table 3.2). These putative promoters may indicate roles for the *ykf* operon in stationary phase (σ^H), degradative enzyme production (σ^L) and sporulation (σ^G). As mentioned above (section 3.2.1.1.) validation of the promoter sequences would require experimental confirmataion by primer extension and/or fusion construct methodology. Downstream from the *ykfD* ORF there is a strong putative rho-independent terminator with a calculated free energy of -19.7 kcal/mole (Table 3.3)

ykfA encodes a protein of 234 amino acids with homology to conserved proteins in a number of other bacteria including *Streptomyces coelicolor, Neisseria meningitidis*, *Synechocystis* sp., and *E. coli*. The homologous protein in *E. coli* is a microcin C7 immunity protein, MccF, endowing host cells with protection against microcin translational inhibitors (Gonzalez-Pastor *et al.*, 1995)

ykfB encodes a protein of 366 amino acids displaying greatest similarity to an ORF in *Bacillus halodurans*. Furthermore both putative proteins are homologous to the family of chloromuconate cycloisomerases. These enzymes, from *Pseudomonas sp.* strain B13 and *Alcaligenes eutrophus* JMP134, participate in degradation of chloroaromatic componds where they catalyse the conversion and dehalogention of the compound chloro-*cis*,*cis*-muconate to *trans*-dienelactone (Vollmer and Schlomann, 1995).

ykfC The YkfC protein has been proposed to encode an endopeptidase and to form part of the large family of proteins termed autolysins in *B. subtilis* through detailed sequence analysis by Smith *et al.* (Smith *et al.*, 2000). Many of these enzymes play a role in cell wall turnover, in particular to processes involved in peptidogylcan remodelling. Experimental evidence for this role was also provided by the above authors through a mutant construct obtained from the author of this thesis (personnel communication from T. Smith and S. Foster to K. Devine and D. Noone).

ykfD shows a high degree of homology to many proteins encoding oligopeptide transport ATP-binding proteins both from *B. subtilis* and other bacteria. Interestingly another *B. subtilis* strain, ATCC 6633, also possesses an operon encoding a self immunity protein (see YkfA above) and an oligopeptide ATP-binding protein transporter (Klein and Entian, 1994).

Overall these analyses indicate that many of the proteins encoded in this region of the *B. subtilis* chromosome are potentially localised to, and/or have roles in the cell envelope of *B. subtilis*.

3.2.2. The *Bacillus subtilis* Functional Analysis Project.

In spite of the explosion in sequencing microbial genomes, many of the resulting predicted genes (up to 50%) have not yet been assigned a function. To systematically investigate the roles of genes of unknown function (*y* genes) a project, comprising two laboratory consortia (European and Japanese), was undertaken. The strategy chosen was to inactivate the *y* gene by integration of the vector pMUTIN using a DNA fragment internal to the ORF (Vagner *et al.*, 1998). The structure of this vector allows the inactivation of the target gene while placing the downstream gene(s) under inducible control. This creates a null mutant strain while simultaneously generating a fusion between the controlling elements for expression of the targeted gene and the reporter gene *lacZ*. To prevent polar mutations arising from the integration event, the plasmid contains an IPTG-inducible promoter Pspac. The consequence of an integration event by a pMUTin derived plasmid construct is shown in Figure 3.2 for the case of the *ykfABCD* operon. Each of the strains generated by the consortia laboratories

was first tested for the expression profile of the targeted gene by growing the strain in rich and minimal media and measuring LacZ activity during a growth cycle. In addition, all the mutants were subjected to a series of broad phenotypic tests to determine the function (Schumann *et al.*, 2001) of the inactivated gene(s). Details of all tests undertaken and results from both expression and phenotypic analyses are available from the Micado database (http://locus.jouy.inra.fr/cgi-bin/genmic/madbase_home.pl).

3.2.2.1. Expression analysis of the *ykfABCD* operon

I used the pMUTIN strategy to inactivate the four *ykf* genes generating BFA strains 1807-1810. Plasmid constructs to disrupt the genes were made by cloning a PCR-amplified internal part of each gene into plasmid pMUTIN1, this was then transformed into the wild type parental strain 168. Mutagenic insertions were obtained for each of the four genes (see Figure 3.2) and integration at the correct loci was verified by PCR. Cells grew normally in the absence of IPTG indicating that the insertion did not have a detrimental polar effect at least under standard laboratory conditions .

The expression profile for each gene was determined by growing strains BFA1807-1810 in Schaeffer's medium, a nutrient based broth used for growth and sporulation of *B. subtilis*. The profile of β–galactosidase expression and growth patterns for the four *ykf* fusion strains is presented in Figure 3.3. The expression profile is broadly similar for the four strains with induction of expression occurring in late exponential/transition phase of the growth cycle. After this point in the growth cycle expression decreases for the four genes as the cells begin to sporulate. There is however a difference in the levels of activity measured in the four strains. The levels of β–galactosidase are similar in strains BFA1807 and 1810. In contrast β–galactosidase activity is approximately two-fold higher in strains BFA 1808 and 1809 compared to strains BFA1807 and 1810 at the end of the exponential growth phase. These differences could be due to internal processing sites and/or differential stabilities for transcripts as a result of any processing events. However, the overall similarity in expression profiles of the fusions is in agreement with the structure of the *ykfABCD* operon.

3.2.2.2. Other phenotypes

The focus of the laboratory within the BFA project was to establish phenotypes in response to the oxidising agents hydrogen peroxide and paraquat. None of the *ykf* mutants differed in their resistance to either of these agents when compared to the wild type parental strain 168. There was no visual change in expression of the *lacZ* fusions in any of the four strains in the presence of either compound (Schumann *et al.*, 2001). These data indicate that the *ykf* operon does not respond to oxidative stress. The expression of *ykfB* in strain BFA1808 has been shown to be responsive to glucose levels in minimal medium by the laboratory of Dr. George Rapport (http://locus.jouy.inra.fr/cgi-bin/dev/chiapell/consult-old.operl). Expression of *ykfB* was increased in medium with low levels of glucose. The *ykfC* mutant strain BFS1809 has a minor mother cell lysis phenotype when combined with other mutations known to affect autolytic processes in *B. subtilis* (personnal communication from T. Smith and S. Foster to Kevin Devine and D. Noone).

3.2.3. HtrA-Family of Serine Proteases in *B. subtilis*.

One of the ORFs found in the region of the genome sequenced as part of this work (see Figure 3.1) was found to encode a protein, YkdA(HtrA), with a high level of similarity to the HtrA/DegP heat shock protease of *E. coli* (Lipinska *et al.*, 1988; Strauch *et al.*, 1989). This family of proteases is the major focus of this thesis. When this gene was sequenced in 1995 there were only a few other putative HtrA-like proteins in the public databases, the most characterised being the DegP protease from *E .coli.* In *E. coli* DegP is located in the compartment created by the inner and outer membranes termed the periplasm, a feature unique to Gram-negative bacteria. In contrast Gram-positive bacteria do not possess an outer membrane, instead having a thick peptidoglycan layer often called the cell wall. The number of putative HtrA-like proteins in the databases has increased vastly. The HtrA family of proteins are now recognised as a distinct and important family of proteases/chaperones necessary for a variety of cellular functions (Pallen and Wren, 1997). They are present in virtually all bacteria sequenced with only a few exceptions, *e.g.* the genomes of two eubacteria *Mycoplasma pneumoniae* and*Mycoplasma genitalium* do not appear to encode HtrA-like serine proteases as revealed by a BLAST search of completed genomes. HtrA

proteases are also absent from genomes of many archae prokaryotes including *Pyrococcus abyssi, Aeropyrum pernix, Thermoplasma volcanium* and *Methanococcus jannaschii.*

All bacterial HtrA-like proteins possess a conserved central domain with homology to the trypsin-like serine proteases which contain a triad of His, Asp and Ser residues responsible for the proteolytic activity of these enzymes. The *B. subtilis* YkdA protein encodes a protein of 449 amino acids including the appropriately positioned catalytic triad of residues referred to above. An alignment of YkdA and the *E. coli* DegP(HtrA) protein is shown in Figure 3.4. The highest of homology is observed between the central portion of both proteins, from the amino acids GSGVI to GIGFAIPS. This region contains the conserved catalytic domain common to all HtrA family members. In contrast, the N-termini of both proteins are quite dissimilar, YkdA having considerably more residues prior to the catalytic core region. YkdA has a single PDZ domain while the DegP protein has a longer C-terminal extension due to the occurrence of a second tandem PDZ domain (see below and Figure 3.8).

3.2.3.1. *B. subtilis* contains two additional members of the HtrA family of proteases.

During the course of this work a second probable HtrA-like protein was found as part of the *B. subtilis* genome sequencing project (Kunst *et al.*, 1997). The *yycK* (*yyxA*) gene was postulated to encode a protein of 400 amino-acids with the conserved catalytic residues of HtrA-like proteins present in the central part of the YycK protein. The *yycK* gene was shown to be in a hexacistronic operon, *yycFGHIJK* that includes the only essential two component signalling transduction system in *B. subtilis, yycFG* (Fabret and Hoch, 1998). YycK expression is controlled by the regulatory system controlling expression of the operon and by an intergenic sporulation induced promoter, positioned immediately upstream of the *yycK* structural gene, containing a σ^G-type consensus sequence (Fabret and Hoch, 1998; Fukuchi *et al.*, 2000)

The *yvtB* gene also encoded a protein with homology to HtrA-like proteases. However, the *yvtB* gene encoded a putative protein of only 224 amino acids, containing only the Asp and

Ser residues of the catalytic triad. The upstream cistron *yvtA* encoded a 122 amino acid protein with some homology to YkdA also but did not contain the conserved His residue. We therefore suspected that the nucleotide sequence was incorrect and undertook to resequence selected parts of the *yvtAB* ORFs. Two regions of the *yvtAB* ORFs were amplified and resequenced with custom primers. As suspected, two changes were identified compared to the sequence deposited in GenBank: the first change involved insertion of an additional cytosine (FS1) and the second change involved deletion of an adenine (FS2) as shown in Figure 3.5. The corrected sequence was conceptually retranslated and shown to encode a single ORF, *yvtA*, capable of encoding a 458 amino acid protein (Figure 3.6). There is strong homology between YvtA and both YkdA 52% identity, (88% similarity) and YycK 36% identity, (72% similarity) (Figure 3.7).

3.2.3.2. Predicted topology of *B. subtilis* HtrA proteases.

Sequence analysis of the three HtrA paralogues in *B. subtilis* showed that they contain the catalytic triad of amino acids in a conserved central domain (red shading in Figure 3.7). The DegP protein of *E. coli* is secreted to the periplasm and contains a typical signal peptide which is removed from the pre-protein following translocation across the cytoplasmic membrane (Lipinska *et al.*, 1990; Skorko-Glonek *et al.*, 1997). This extra-cytoplasmic sub-cellular location is also postulated or has been demonstrated for a number of other HtrA-like proteins (Pallen and Wren, 1997; Jones *et al.*, 2001). The three *B .subtilis* proteins were analysed using the SIGNALP programme to determine if any of these proteins possessed signal peptides which might direct their export across the cytoplasmic membrane. None of the three proteins contained a signal peptide nor a putative cleavage site for signal peptidases. However, the TMHMM programme predicted that all three proteins contained a single stretch of hydrophobic amino acids at their N-termini predicted to form a transmembrane (TM) helical region, boxed in Figure 3.7. Furthermore this programme also predicts that the region of the protein located on the C-terminal side of the TM, which contains the catalytic core of the proteins, is located extracytoplasmically, with the N-terminal region preceding the TM motif having a "cytoplasmic-in" location (Figure 3.7). In agreement with the prediction that the catalytic domain of YkdA is on the "outside" of the cell, it is unique among the three

proteins in containing two serine rich regions (overlined in Figure 3.7). It is know that such serine rich regions occur in proteins localised to the cell envelope in *B. subtilis* and other Gram-positive bacteria (Arnold *et al.*, 1999; Ohnishi *et al.*, 1999). A FASTA search using the twin polyserine region of YkdA against the *B. subtilis* proteome revealed that 10 of the top 25 protein matches have a putative envelope localisation.

3.2.3.3. PDZ domains of *B. subtilis* HtrA proteases.

The HtrA-family of bacterial proteases all contain at least one PDZ domain at their C-termini which are predicted to mediate substrate recognition and binding (Pallen and Ponting, 1997; Ponting, 1997). HtrA proteins in Gram-negative bacteria possess in the main two tandemly arranged PDZ domains while the proteases of Gram-positive bacteria contain single PDZ domains. An alignment of the of three *B. subtilis* proteins with the DegP protein of *E. coli* shows that the longer C-terminal extension of DegP is due to a second PDZ domain as previously reported, (Figure. 3.8) (Pallen and Ponting 1997). The alignment also shows that the PDZ domains contain the conserved residues as previously noted (Pallen and Ponting, 1997; Ponting, 1997). PDZ domains have a conserved secondary structure architecture consisting of five or six beta-strands and two alpha-helices (Doyle *et al.*, 1996; Fanning and Anderson, 1996). The PDZ domains of the three *B. subtilis* HtrA proteases show a high level of secondary structural similarity. Additionally they conform to the general structure for PDZ domains as shown in Figure 3.9. Recently the crystal structure of *E. coli* DegP has been determined and this shows an enlarged PDZ domain with extra structural elements (Krojer *et al.*, 2002). The additional elements are accounted for by the insertion of approximately thirty residues between the conserved βB and βC strands, as noted previously by Fanning and Anderson (1996). However, for the analysis of the *B. subtilis* HtrA PDZ domains, the boundaries identified in the alignment created by Pallen and Ponting (1997) are used. The roles of PDZ domains are primarily in mediating protein-protein interaction (Fanning and Anderson, 1996). This type of domain could therefore be used by proteases for substrate recognition and binding as has been shown for the Tsp protease of *E. coli* (Beebe *et al.*, 2000). Furthermore, the binding action could be coupled with concomitant signal transmission as proposed for the SpoIVB protein of *B. subtilis* (Hoa *et al.*, 2001).

3.2.3.4. Expression signals for the *ykdA*, *yvtA* and *yycK* genes in *B. subtilis*.

Given that *B. subtilis* contains three HtrA-like serine proteases, a search for common controlling sequences in the promoter regions of the three genes *ykdA*, *yvtA* and *yycK* (*yyxA*) was undertaken. The *ykdA* and *yvtA* genes are both contained within monocistronic operons. Inspection of the sequences upstream of*ykdA* and *yvtA* revealed a region of 60-bp with a high degree of similarity. The *yycK* gene is located in a hexacistronic operon and neither the operon promoter area, nor the region immediately upstream of*yycK* contain sequences matching those of the *ykdA* or *yvtA* promoters. An alignment of the promoter regions of *ykdA* and *yvtA* is presented in Figure 3.10. which shows that there is 91% identity over a stretch of 60-bp between the *ykdA* and *yvtA* controlling sequences. However, despite the fact that both genes possess this common sequence there are differences between their controlling regions. There are 64 nucleotides between the 3' end of the common motif to the translational start site of *ykdA*, while this same region of the *yvtA* promoter is extended by an additional 101 bases to a total of 165 nucleotides (see Figure 3.10). This extended region may contain additional sequences for controlling expression of *yvtA*.

The *E. coli degP* gene is regulated by a σ^E controlled promoter (Erickson and Gross, 1989). There is a high degree of conservation in the promoters of other genes transcribed by RNA polymerase holoenzymes directed by ECF-type sigma factors to which σ^E belongs (Lonetto *et al.*, 1994; Missiakas and Raina, 1998). An inspection of the *ykdA*, *yvtA* and *yycK* controlling sequences found that both upstream regions of *ykdA* and *yvtA* contained possible ECF-type promoter sequences. These potential promoter sequences are compared in Figure. 3.11 to ECF-type promoters from a number of other bacterial species and genera. The spacing between the -35 and -10 motifs is within specified limits while the putative *ykdA* ECF-type controlling sequence also contains a run of adenine nucleotides a feature noted in some other ECF-type promoters (Raina *et al.*, 1995; Rouviere *et al.*, 1995). In contrast no ECF-type promoter sequences were noted in the upstream regions of *yycK*.

The *B. subtilis* genome sequencing project revealed that there are 7 sigma factors of the ECF family encoded in the chromosome of this bacterium. To date experimental data is available for three of these sigma factors, σ^X (Huang *et al.*, 1997), σ^W (Huang *et al.*, 1998) and σ^M (Horsburgh and Moir, 1999). None of the consensus sequences derived from the promoters of genes known to be controlled by these sigma factors are present in the upstream regions of *ykdA*, *yvtA* or *yycK*. Furthermore, none of the three genes encoding the HtrA-like proteins of *B. subtilis* were identified in a consensus driven approach as being potential members of the regulons controlled by the *B. subtilis* σ^X (Huang and Helmann, 1998) or σ^W sigma factors (Huang *et al.*, 1999). Nevertheless there is identity between some of the conserved nucleotides of these generic ECF-type promoter elements and the putative ECF-type promoters of *ykdA* and *yvtA*. Therefore a role for one or more of the ECF-type sigma factors in the expression of *ykdA* and *yvtA* seemed possible. Experiments to investigate if expression of *ykdA* was under the control of ECF-type sigma factors were undertaken (see section 4.2.4).

Table 3.1. Open Reading Frame features and homology with other proteins

ORF	Size (aa)	Coordinates[a]	Ribosome binding site[b] /Translation Start codon	Description/Function
ykcA	316	1352391 - 1353338	tGGAGGgacagcat**ATG**	Similar to ABC transporter binding proteins
ykcB	716	1353751 - 1355898	AAAGGAGtgAaaaataa**GTG**	Similar to YycA (*B. subtilis*) and Dolichyl-phosphate-mannose-protein mannosyltransferases
ykcC	323	1355913 - 1356881	AtAGGAGGcaAaaac**ATG**	Similar to glycosyl transferases
ykdA	449	1357405 - 1358751	AGAAAcGgaGTGATCatag**ATG**	Similar to HtrA serine proteases and YvtA and YyxA paralogues in *B. subtilis*
ykeA	272	1358920 - 1359735	GAAAGGActgagTCg**ATG**	Pyrroline carboxylate reductase
dppA	274	1359708 - 1360529	AAAGGAGGaGcgtt**ATG**	D-aminopeptidase
dppB	308	1360549 - 1361472	AAAGGgGGTtTtaggct**TTG**	Dipeptide permease
dppC	320	1361481 - 1362440	AAAGGgtGaagcgt**GTG**	Dipeptide permease
dppD	335	1362488 - 1363452	AGGAGGTaAgCtgt**ATG**	Dipeptide ATP binding protein
dppE	549	1363458 - 1365104	AAAGGgGGaagagg**ATG**	Dipeptide binding protein
ykfA	234	1365607 - 1366308	AtGAGGTaAaagcagtgtt**GTG**	Similar to Microcin C7 immunity protien
ykfB	366	1366308 - 1367405	AAGGgGcgctgaagac**ATG**	Possible muconate cycloisomerase
ykfC	296	1367405 - 1368292	GAAAGGgGaaaAagagca**ATG**	Autolysin of the DL-endopeptidase II family
ykfD	327	1368308 - 1369288	AAGGAGGgagaCaa**ATG**	Peptide transport ATP-binding proteins

[a] nucleotide bases from the origin of replication (see http://genolist.pasteur.fr/SubtiList/genome.cgi)

[b] bases complementary to the 3' end of the 16S RNA of *B. subtilis* (3' -UCUUUCCUCCACUAG) (McLaughlin *et al.*, 1981)

Table 3.2. Motifs upstream of ORFs with similarity to consensus *B. subtilis* sigma factor binding sites

ORF	Putative vegatative sigma factor binding sites detected[a]	Sporulation sigma factor binding sites detected[a]
ykcBC	sigmaD taaa→15nt←gccgatat gAAA→15nt←GCaaATAT	sigmaK ac→17nt←catannnta cC→18nt←CATAATTaA
	sigmaH rvaggawwt→14nt←mgaat GGAGGAAAT→14nt←gGgAg	
	sigmaL tggcac→4nt←ttgcannn TGttAC←4nt←TTtCAAAG	
ykdA	sigmaD taaa→15nt←gccgatat TAAA→16nt←atCaATAT	sigmaG gmatr→18nt←catwmta Gataa→19nt←CAataTA
		sigmaK ac→17nt←catannnta AC→17nt←CAcAATTTc
ykeA	sigmaA ttgaca→17nt←tataat TTGAtA→17nt←TATcAT	sigmaK ac→17nt←catannnta AC→16nt←CAaAAACTg
	sigmaB rggwttra→14nt←gggtat GGGATaAA→14nt←GGGaAa	
ykfABCD	sigmaH rvaggawwt→14nt←mgaat GCgGtATTT→14nt←CGtAT	sigmaG gmatr→18nt←catwmta GgATG→19nt←CATgAgA
	sigmaL tggcac→4nt←ttgcannn TtcCtC→4nt←TTGCACGT	

[a] nucleotide matches to those of sigma factor consensus binding sites compiled by Haldenwang (1995); the first line of each match represents the input search motif including some spacer flexibility and the second line is the sequence found in the upstream regions of the indicated genes.

Table 3.3. Sequence elements capable of forming stem loop structures and likely to function as rho-independent transcriptional terminators.

ORF	Sequence of putative rho-independent terminator[a]	ΔG values
	>>>>>>>>>>>>>>>> <<<<<<<<<<<<<<	
ykcC	**taa**cctcAAACCCCCTGTCCgtaatGGACAGGGGGTTTttat	-24.30 kcal/mol
	>>>>>>>>>>>>> <<<<<<<<<<<<	
ykdA	**taa...**GCCTCAGgCCGtaaaagCGGtCTGAGGCtttttatt	-23.7 kcal/mole
	>>>>>>>> <<<<<<<<	
ykeA	**tg**GAATCCTtgaaagAGGATTCtttttttat	-11.9 kcal/mol
	>>>>>>>>>>>>>> <<<<<<<<<<<<<<	
dppE	**tga...**ggAAATACtGCTtCTTtatgcgAAGgAGCgGTATTTttcctctttctt	-19.4 kcal/mole
	>>>>>>>>>>>>>>>>>>>> <<<<<<<<<<<<<<<<<<<<	
ykfD	**taa**gATAAAtAAAaaGAGTGGCTCctttttactaGAGCCACTCcaTTTgTTTAT	-19.7 kcal/mol

[a] nucleotide in bold lower case represent the stop codons, some bases between the stop codon and the terminator structure are omitted for formatting purposes; bases capable of pairing are in capitals.

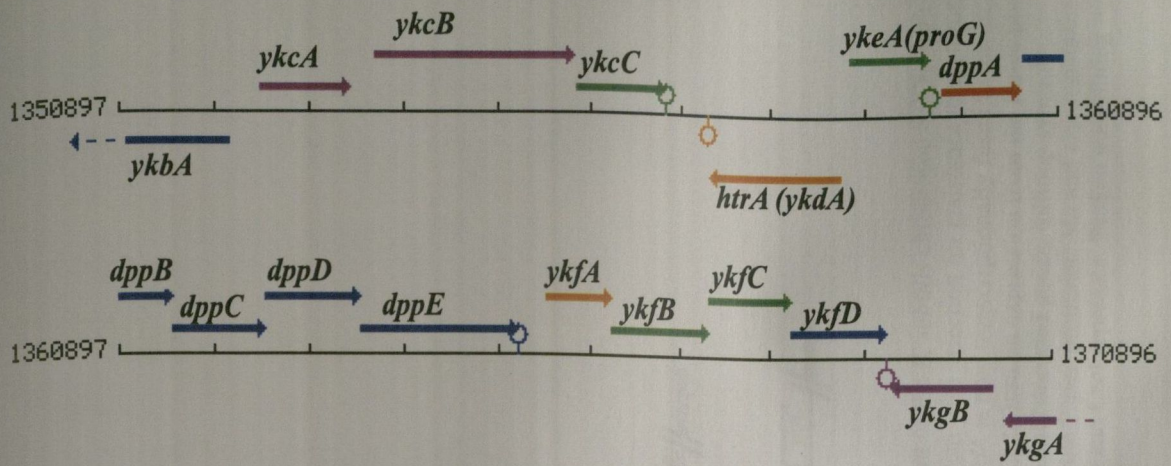


Figure 3.1. Physical map of the *B. subtilis* chromosome region covered in pSAL16. The thin line represents the chromosome and tick marks indicate 1 kbp intervals. Genes are indicated by arrows. Genes below the line are encoded on the lagging strand. Lollipops represent transcriptional terminators and numbers represent bp from the origin. Genes are colour coded to indicate known or putative function: ■, cellular processes; ■, information pathways; ■, similar to unknown proteins; ■, intermediary metabolism; ■, other functions. (figure modified from SubtiList, http://genolist.pasteur.fr/SubtiList/).

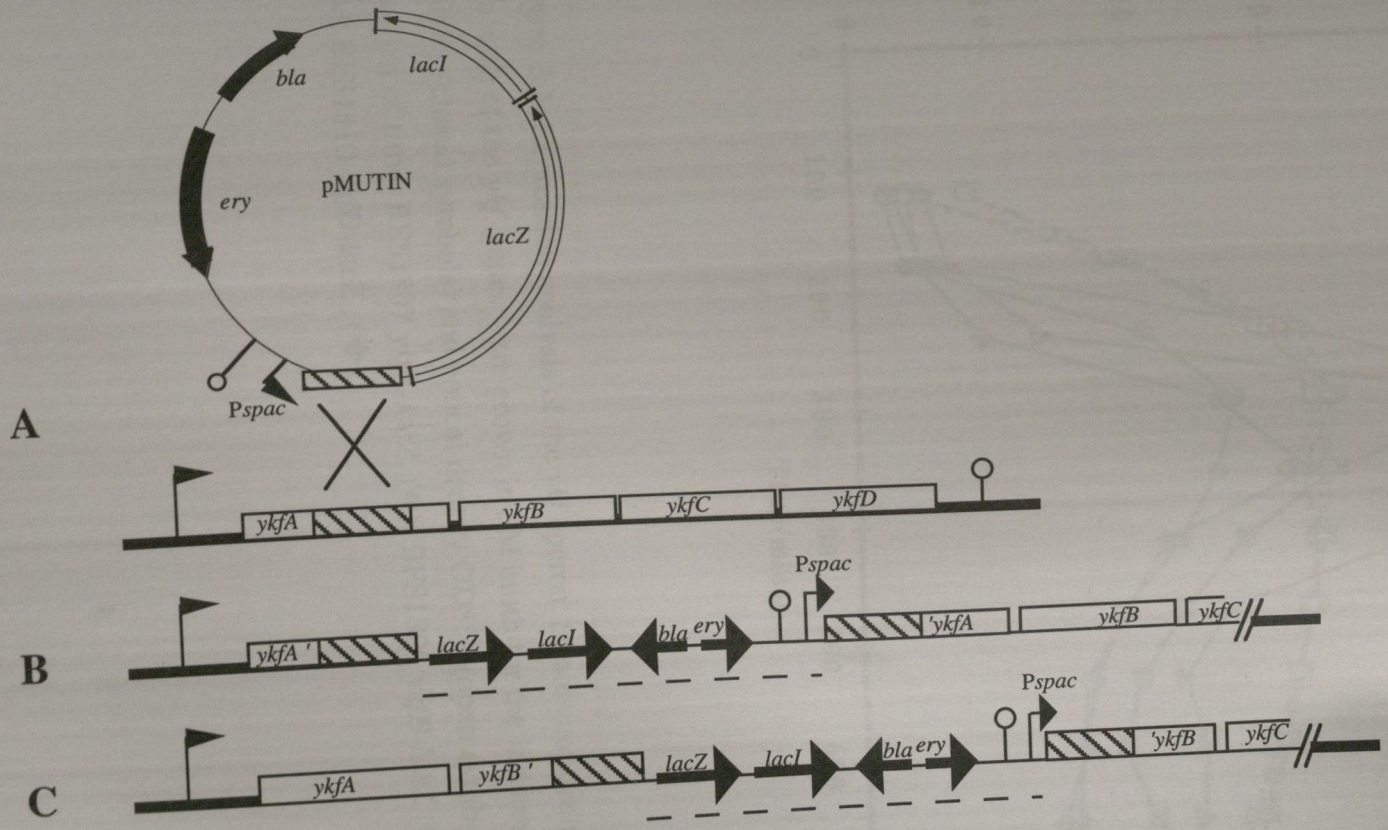


Figure 3.2. Generation of mutant /fusion strains using the pMUTIN integrating vector. (A) pMUTIN containing a fragment of the *B. subtilis* chromosome (▭) from the middle of the targeted gene when introduced into *B. subtilis* recombines into the homologous locus by single cross over event with selection for resistance to erythromycin (*ery*). (B) This creates an integration mutant, of *ykfA* in the case of shown here. The promoter controlling expression of *ykfA* is now transcriptionally fused to the *lacZ* gene. In the event of polar effects on downstream genes *ykfBCD* their expression can be induced by addition of IPTG to the media to turn on expression from the P*spac* promoter. (C) The case of integration into *ykfB* is demonstrated here. Open boxes, chromosomally encoded genes; heavy arrows, plasmid encoded genes; lollipop symbol, transcriptional terminator to prevent transcriptional readthrough from plasmid to chromosome; broken arrow, IPTG inducible P*spac* promoter; *bla* and *ery*, resistance determinants for ampicillin and erythromycin respectively; *lacZ* and *lacI*, genes encoding β-galactosidase reporter enzyme and LacI repressor. Broken underlining represents bounderies of plasmid encoded elements.

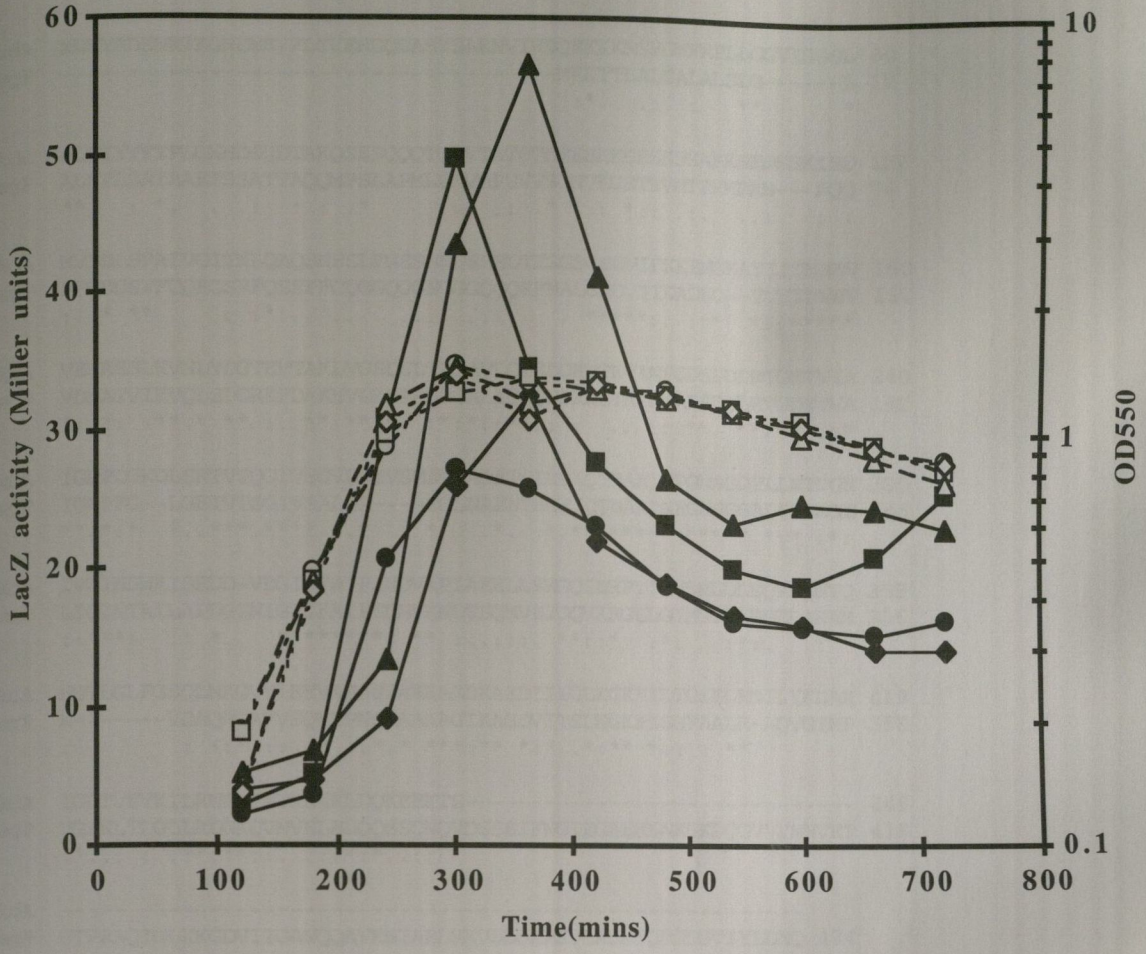


Figure 3.3. Expression analysis of the *ykf* operon. Strains containing pMUTIN1 insertions into each of the *ykf* genes were grown in SM broth for the indicated time. ß-galactosidase activity (closed symbols) and optical density OD550 (open symbols) was determined as a function of time. BFS1807 *ykfA-lacZ*, (●); BFS1808 *ykfB-lacZ* (■); BFS1809 *ykfC-lacZ*, (▲); BFS1810 *ykfD-lacZ* , (◆).

```
YkdA  MDNYRDENRTKGNENEVFLTKENDQSASYSARNVIHDQEKKKRGFGWFRPLLGGVIGGSL 60
DegP  -----------------------------------MKKTTLALSALALSLG------L 17
                                          :*.. .:. :   **      *

YkdA  ALGIYTFTPLGNHDSQDTAKQSSSQQQTQSVTATSTSSESKKSSSSSSAFKSEDSSKISD 120
DegP  ALSPLSATAAETSSATTAQQMPSLAPMLEKVMPSVVSINVEGSTTVNTPRMPRN---FQQ 74
      **.  : *.  . .:  : : .*     :.* .: .* : : *:: .:.  ..:   :.:

YkdA  MVEDLSPAIVGITNLQAQSNSSLFGSSSSDSSEDTESGSGSGVIFKKENGKAYIITNNHV 180
DegP  FFGDDSPFCQEGSPFQSSPFCQGGQGGNGGGQQQKFMALGSGVIIDADKG--YVVTNNHV 132
      :. * **     : :*:.. ..   .......::.  . *****:. ::*  *:*****

YkdA  VEGASSLKVSLYDGTEVTAKLVGSDSLTDLAVLQISDDHVTKVANFGDSSDLRTGETVIA 240
DegP  VDNATVIKVQLSDGRKFDAKMVGKDPRSDIALIQIQNPKNLTAIKMADSDALRVGDYTVA 192
      *:.*: :**.* ** :. **:**.*. :*:*::**.: :  .. ::.**. **.*: .:*

YkdA  IGDPLGKDLSRTVTQGIVSGVDRTVSMSTSAGETSINVIQTDAAINPGNSGGPLLNTDGK 300
DegP  IGNPFG--LGETVTSGIVSALGR----SGLNAENYENFIQTDAAINRGNSGGALVNLNGE 246
      **:*:*  *..***.****.:.*    *   .*.  *.******** *****.*:* :*:

YkdA  IVGINSMKISEDD-VEGIGFAIPSNDVKPIAEELLSKGQIERPYIGVSMLDLEQVPQNYQ 359
DegP  LIGINTAILAPDGGNIGIGFAIPSNMVKNLTSQMVEYGQVKRGELGIMGTELNSELAKAM 306
      ::***:  :: *.   ********* ** ::.:::. **::*  :*:   :*:.   :

YkdA  EGTLGLFGSQLNKGVYIREVASGSPAEKAGLKAEDIIIGLKGKEIDTGSELRNILYKDAK 419
DegP  K-------VDAQRGAFVSQVLPNSSAAKAGIKAGDVITSLNGKPISSFAALR-AQVGTMP 358
      :        : ::*.:: :* ..*.* ***:** *:* .*:** *.: : **

YkdA  IGDTVEVKILRNGKEMTKKIKLDQKEEKTS---------------------------- 449
DegP  VGSKLTLGLLRDGKQVNVNLELQQSSQNQVDSSSIFNGIEGAEMSNKGKDQGVVVNNVKT 418
      :*..: : :**:**::. :::*:*..::

YkdA  ------------------------------------------------------
DegP  GTPAAQIGLKKGDVIIGANQQAVKNIAELRKVLDSKPSVLALNIQRGDSTIYLLMQ 474
```

Figure 3.4. Alignment of the *E. coli* DegP and *B. subtilis* YkdA proteins. Identical, conserved and semi-conserved residues are represented by (*), (:) and (.) respectively, while gaps introduced for maximum fit are indicated by a dash (-). Amino acid residues forming the catalytic triad of HtrA-type serine proteases are shaded red. The region of greatest homology within the catalytic domain is indicated starting and ending with the residues shaded blue. The alignment was generated using the CLUSTALW programme.

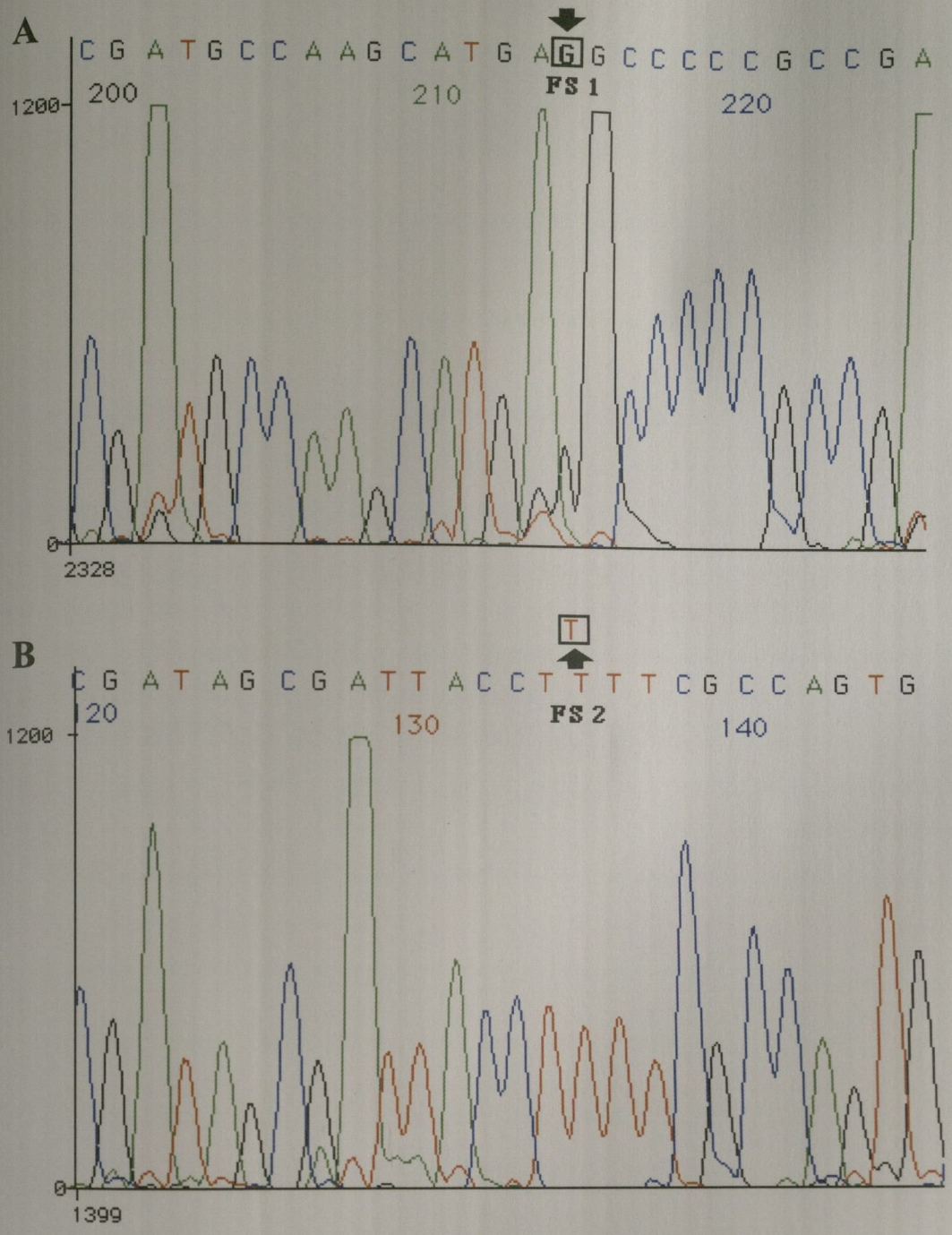
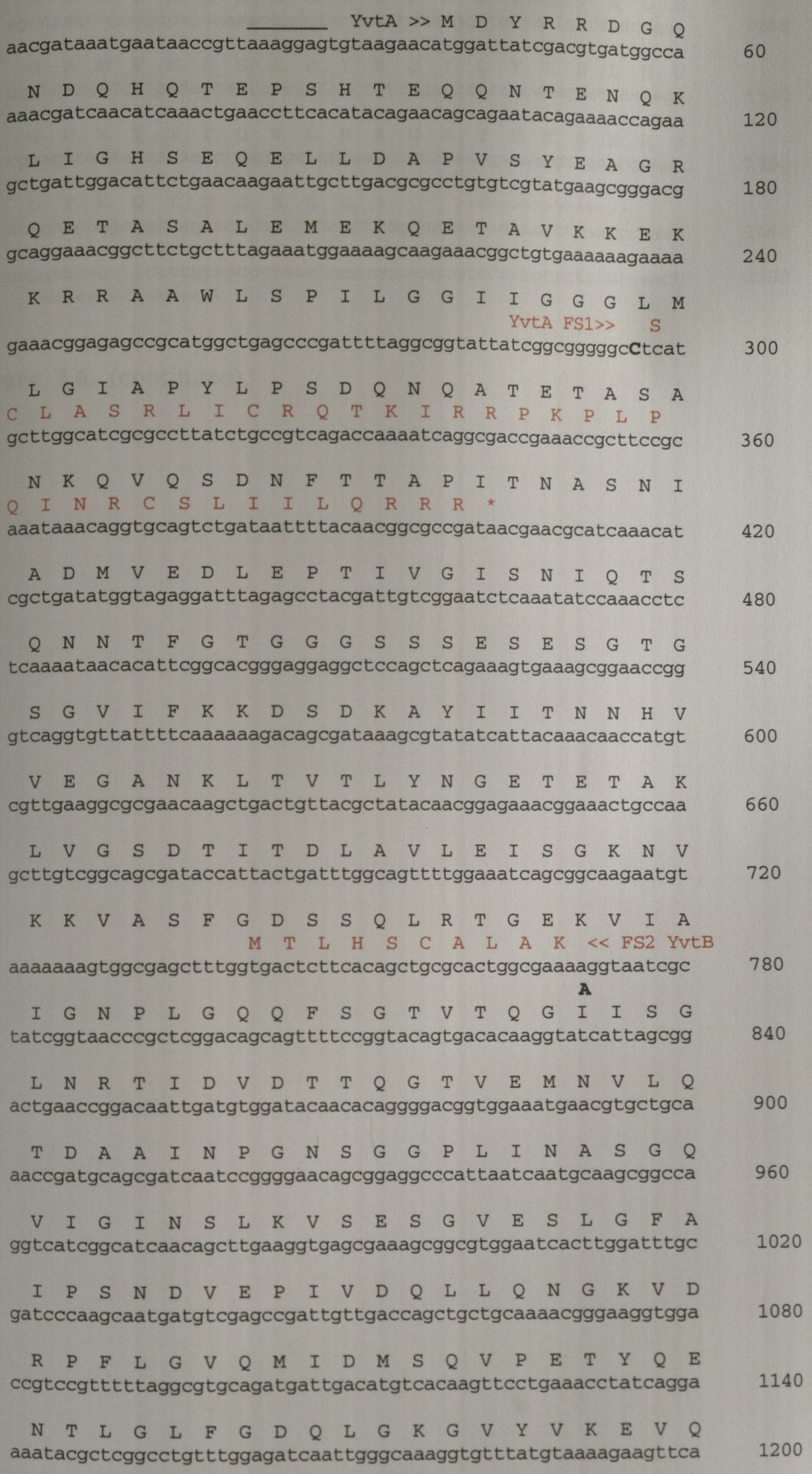


Figure 3.5. Identification and correction of sequencing errors in the *yvtAB* nucleotide sequence. Selected regions of the *yvtAB* ORFs were PCR amplified and subjected to direct sequencing to identify suspected errors, chromatographs from the ABI 373A output are shown. (A) The upstream frameshift (FS1) was identified as omission of a cytosine nucleotide at the indicated position. (B)The downstream frameshift (FS2) was identifed as the incorrect inclusion of an additional thymidine at the indicated position. Both sequences were determined for the non-coding strand and the complementary changes are shown in Figure 3.6.

Figure 3.6. Translation of the *yvtA* sequence after correction of detected frameshifts. The *yvtA* ORF encodes a 458 amino acid protein. Deduced protein sequence (black capitals) from the corrected *yvtA* ORF is shown above the nucleotide sequence. Previous incorrect amino acid sequences and stop codon are shown in red text for YvtA (GGG→SCLAS...*) and YvtB (ALAKK→VIAIG). Inclusion of an additional cytosine and deletion of the extraneous adenine are highlighted in **bold.** * indicates stop codons, >>> represents nucleotides forming a putative transcriptional terminator, and the overlined sequence shows the putative ribosome binding site for *yvtA*.


YvtA >> M D Y R R D G Q
aacgataaatgaataaccgttaaaggagtgtaagaacatggattatcgacgtgatggcca 60
N D Q H Q T E P S H T E Q Q N T E N Q K
aaacgatcaacatcaaactgaaccttcacatacagaacagcagaatacagaaaaccagaa 120
L I G H S E Q E L L D A P V S Y E A G R
gctgattggacattctgaacaagaattgcttgacgcgcctgtgtcgtatgaagcgggacg 180
Q E T A S A L E M E K Q E T A V K K E K
gcaggaaacggcttctgctttagaaatggaaaagcaagaaacggctgtgaaaaaagaaaa 240
K R R A A W L S P I L G G I I G G G L M
YvtA FS1>> S
gaaacggagagccgcatggctgagcccgattttaggcggtattatcggcgggggcCtcat 300
L G I A P Y L P S D Q N Q A T E T A S A
C L A S R L I C R Q T K I R R P K P L P
gcttggcatcgcgccttatctgccgtcagaccaaaatcaggcgaccgaaaccgcttccgc 360
N K Q V Q S D N F T T A P I T N A S N I
Q I N R C S L I I L Q R R R *
aaataaacaggtgcagtctgataatttacaacggcgccgataacgaacgcatcaaacat 420
A D M V E D L E P T I V G I S N I Q T S
cgctgatatggtagaggatttagagcctacgattgtcggaatctcaaatatccaaacctc 480
Q N N T F G T G G G S S S E S E S G T G
tcaaaataacacattcggcacgggaggaggctccagctcagaaagtgaaagcggaaccgg 540
S G V I F K K D S D K A Y I I T N N H V
gtcaggtgttattttcaaaaaagacagcgataaagcgtatatcattacaaacaaccatgt 600
V E G A N K L T V T L Y N G E T E T A K
cgttgaaggcgcgaacaagctgactgttacgctatacaacggagaaacggaaactgccaa 660
L V G S D T I T D L A V L E I S G K N V
gcttgtcggcagcgataccattactgatttggcagttttggaaatcagcggcaagaatgt 720
K K V A S F G D S S Q L R T G E K V I A
M T L H S C A L A K << FS2 YvtB
aaaaaaagtggcgagctttggtgactcttcacagctgcgcactggcgaaaaggtaatcgc 780
A
I G N P L G Q Q F S G T V T Q G I I S G
tatcggtaacccgctcggacagcagttttccggtacagtgacacaaggtatcattagcgg 840
L N R T I D V D T T Q G T V E M N V L Q
actgaaccggacaattgatgtggatacaacacaggggacggtggaaatgaacgtgctgca 900
T D A A I N P G N S G G P L I N A S G Q
aaccgatgcagcgatcaatccggggaacagcggaggcccattaatcaatgcaagcggcca 960
V I G I N S L K V S E S G V E S L G F A
ggtcatcggcatcaacagcttgaaggtgagcgaaagcggcgtggaatcacttggatttgc 1020
I P S N D V E P I V D Q L L Q N G K V D
gatcccaagcaatgatgtcgagccgattgttgaccagctgctgcaaaacgggaaggtgga 1080
R P F L G V Q M I D M S Q V P E T Y Q E
ccgtccgtttttaggcgtgcagatgattgacatgtcacaagttcctgaaacctatcagga 1140
N T L G L F G D Q L G K G V Y V K E V Q
aaatacgctcggcctgtttggagatcaattgggcaaaggtgtttatgtaaaagaagttca 1200

```
  A  N  S  P  A  E  K  A  G  I  K  S  E  D  V  I  V  K  L  N
agcaaattcgccggcagaaaaagcaggtattaaatctgaggatgtcattgtcaaactaaa      1260

  G  K  D  V  E  S  S  A  D  I  R  Q  I  L  Y  K  D  L  K  V
cggtaaagacgtggaaagcagtgcggatatccgtcagatcctttacaaagacttaaaagt      1320

  G  D  K  T  T  I  Q  V  L  R  K  G  K  T  K  T  L  N  A  T
cggagataaaacaacaatccaagtgctgcgaaagggaaaaacgaaaacgttgaacgcaac      1380

  L  T  K  Q  T  E  S  S  S  S  *        >>>>>>>>>>>>>>>>
actgaccaagcagacagaaagcagttcaagctaagaaaaaacaaaaagctgaacccgatt      1440

aaggttcagctttttgttaccctaaataagaggaaagcatccacacttgtttttcaact      1500
  <<<<<<<<<<<<<<<<<
```

Figure 3.6. (continued).

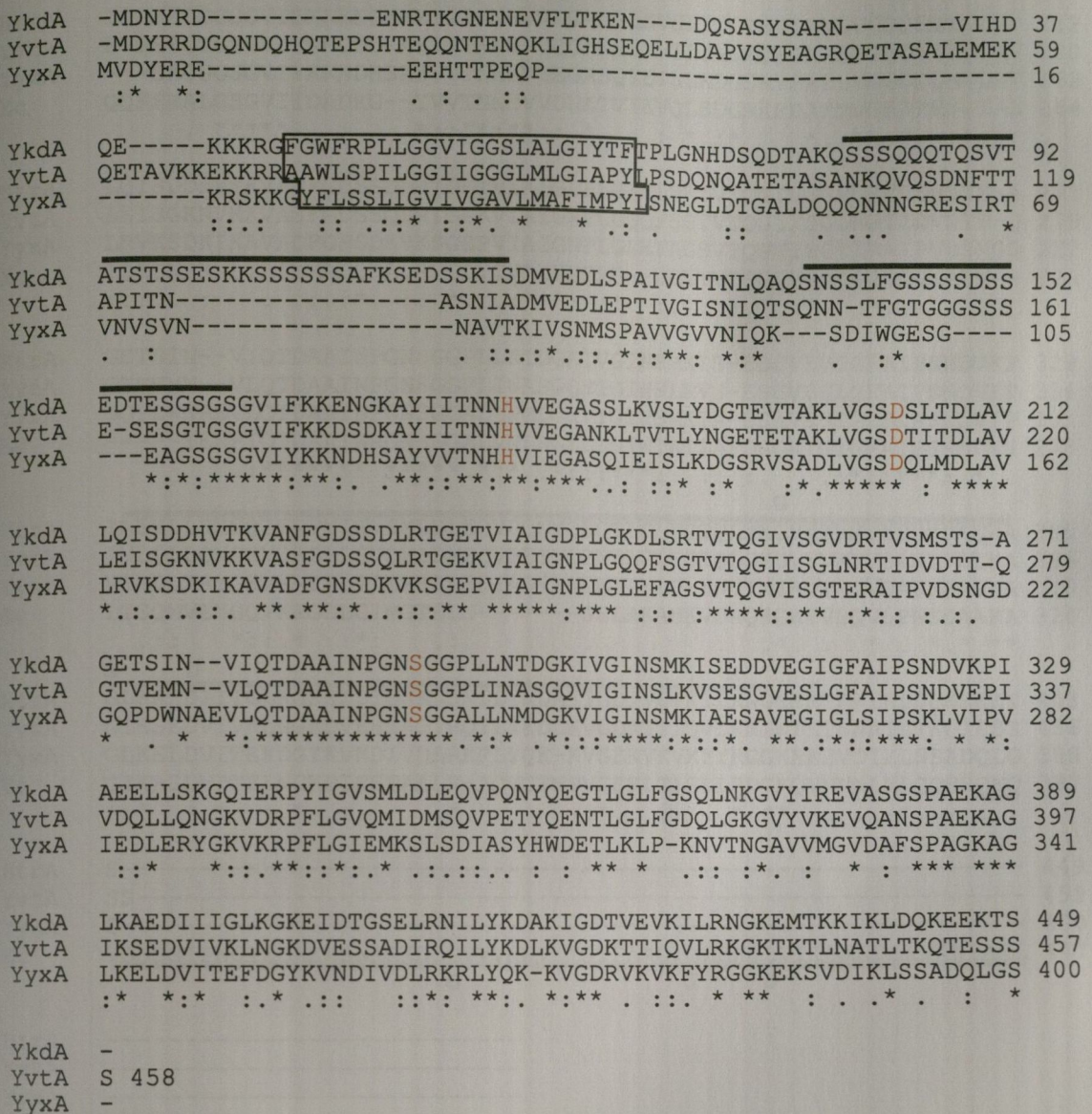


Figure 3.7. Multiple alignment and features of HtrA proteases from *B. subtilis*. The three HtrA-like proteases YkdA, YvtA and YyxA from *B. subtilis* are shown. Stretches of hydrophobic residues predicted to specify transmembrane domains in all three proteins are boxed. The heavy overlines above the YkdA sequence shows the two serine rich motifs with possible involvement in cell wall attachment. Red shading denotes the conserved catalytic residues. Identical, conserved and semi-conserved residues are represented by (*), (:) and (.) respectively, while gaps introduced for maximum fit are indicated by a dash (-).

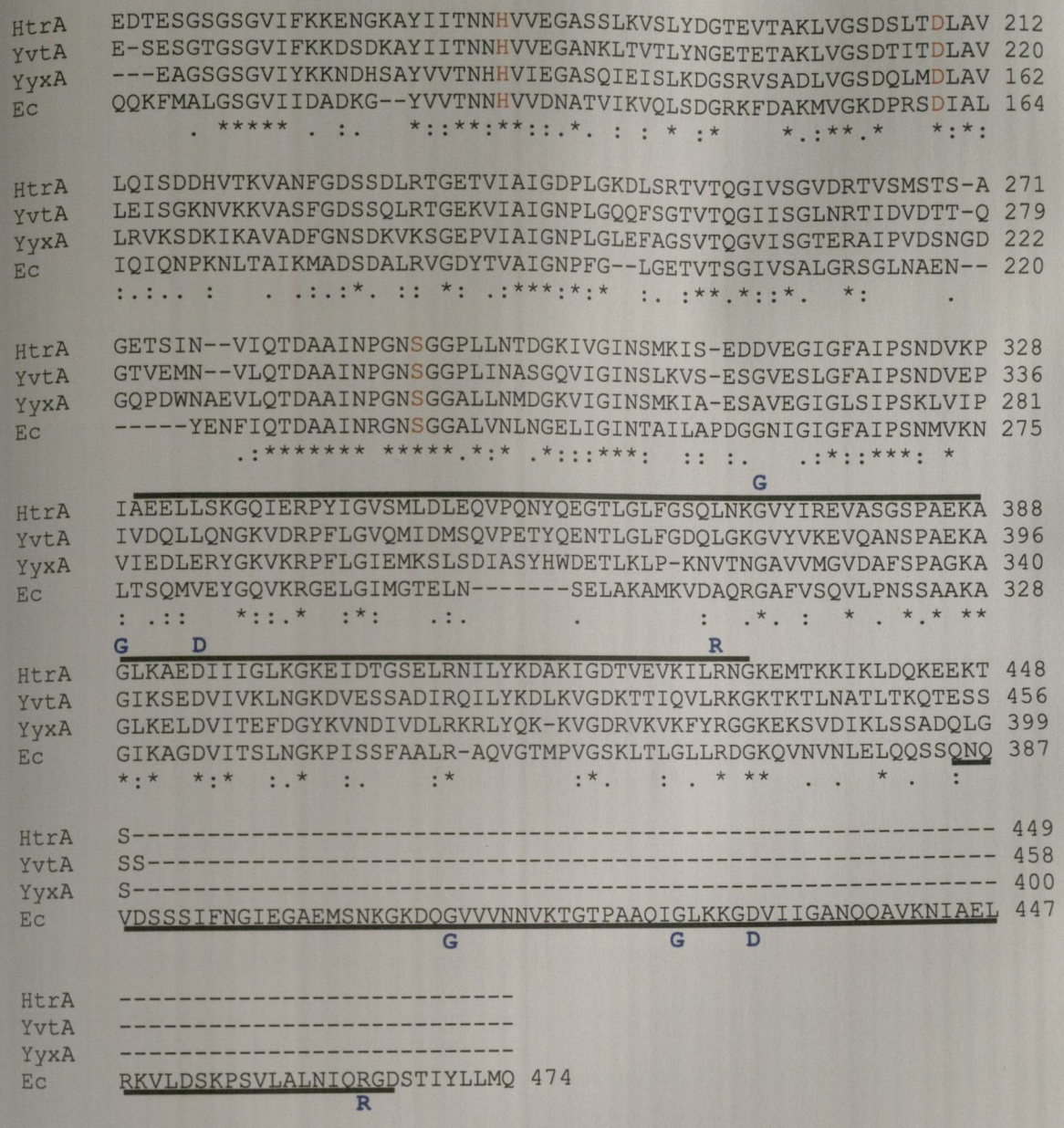
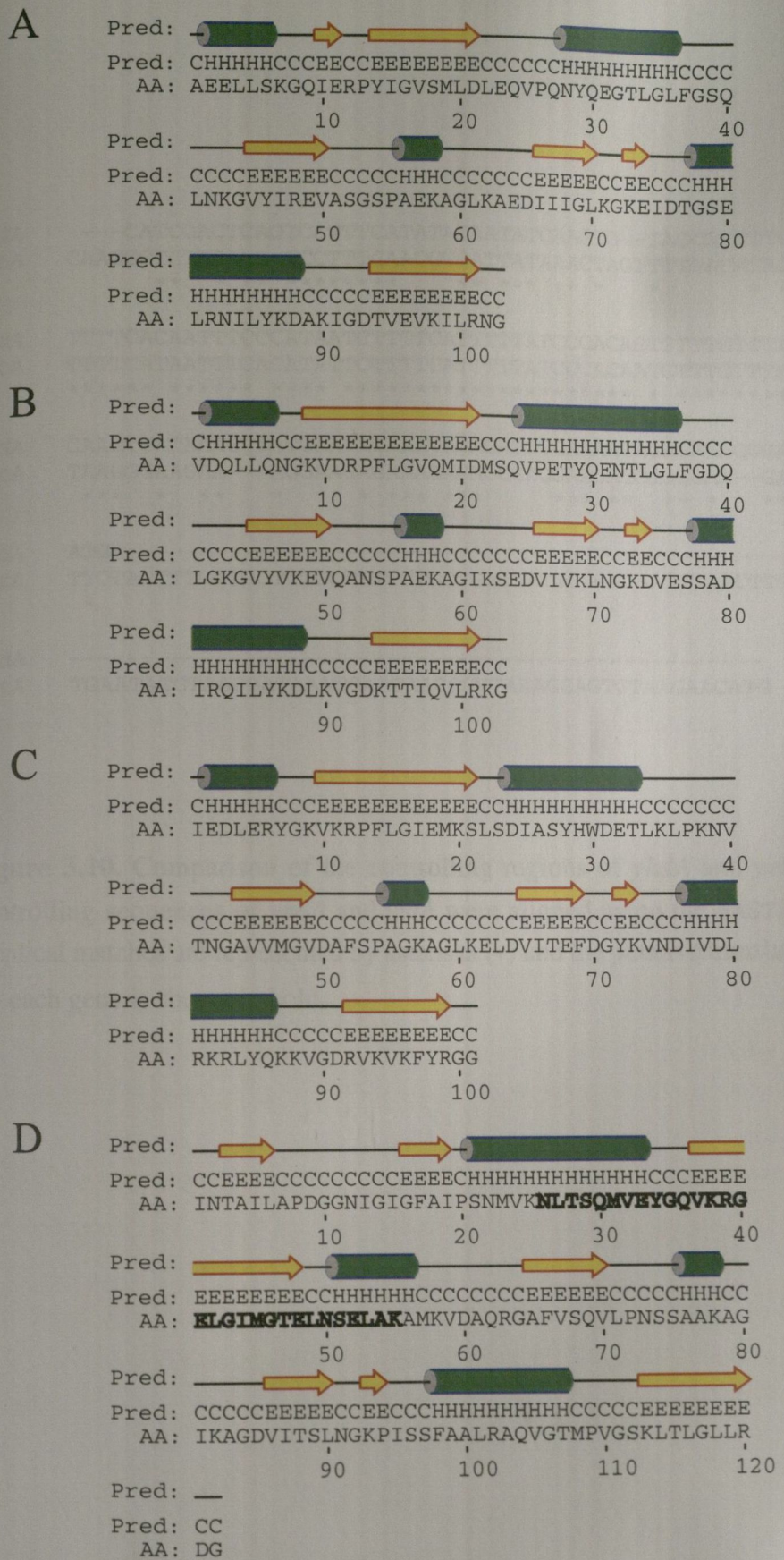


Figure 3.8. PDZ domains of HtrA proteases from *B. subtilis* and DegP(HtrA) of *E. coli* [Ec]. PDZ domains were identified using the SMART database and by visual comparison to the alignment of Pallen and Ponting (1997). Conserved residues of the catalytic core are shaded red. The single carboxyl- terminal PDZ domains of the three *B. subtilis* proteins and the first PDZ domain of DegP are indicated by the overline above the sequences. The second PDZ domain of DegP is shown by the underline. The highly conserved residues common to most PDZ domains are emphasized in blue. Identical, conserved and semi-conserved residues are represented by (*), (:) and (.) respectively, while gaps introduced for maximum fit are indicated by a dash (-). Numbers represent residues from the N terminus.

Figure 3.9. Secondary structure of PDZ domains from *B. subtilis* HtrA-like proteases and *E. coli* DegP. (A) YkdA. (B) YvtA. (C) YyxA (YycK). (D) DegP. The sequences used to predict the secondary structure elements of the *B. subtilis* PDZ domains are the same as in figure 3.8. The sequence analysed for the structure of the DegP PDZ1 domain was that previously used containing an additional 30 N-terminal residues, highlighted in bold, (Fanning and Anderson, 1996), which accounts for the additional motifs found in the crystal structure of the PDZ-1 domain of DegP (Krojer *et al.*, 2002). The secondary structure predictions were generated by the PSIPRED programme and the figure was generated by modifying the output postscript files. Green cylinders and yellow arrows denote α-helices (H) and β-strands (E) respectively, coiled regions are indicated by plain line and (C).


A
Pred:
Pred: CHHHHHCCCEECCEEEEEEEECCCCCCHHHHHHHHHCCCC
AA: AEELLSKGQIERPYIGVSMLDLEQVPQNYQEGTLGLFGSQ
10 20 30 40
Pred:
Pred: CCCCEEEEEECCCCCHHHCCCCCCCEEEEECCEECCCHHH
AA: LNKGVYIREVASGSPAEKAGLKAEDIIIGLKGKEIDTGSE
50 60 70 80
Pred:
Pred: HHHHHHHHCCCCCEEEEEEEECC
AA: LRNILYKDAKIGDTVEVKILRNG
90 100
B
Pred:
Pred: CHHHHHCCEEEEEEEEEEEECCCHHHHHHHHHHHHCCCC
AA: VDQLLQNGKVDRPFLGVQMIDMSQVPETYQENTLGLFGDQ
10 20 30 40
Pred:
Pred: CCCCEEEEEECCCCCHHHCCCCCCCEEEEECCEECCCHHH
AA: LGKGVYVKEVQANSPAEKAGIKSEDVIVKLNGKDVESSAD
50 60 70 80
Pred:
Pred: HHHHHHHHCCCCCEEEEEEEECC
AA: IRQILYKDLKVGDKTTIQVLRKG
90 100
C
Pred:
Pred: CHHHHHCCCEEEEEEEEEEEECCHHHHHHHHHHHCCCCCCC
AA: IEDLERYGKVKRPFLGIEMKSLSDIASYHWDETLKLPKNV
10 20 30 40
Pred:
Pred: CCCEEEEEECCCCCHHHCCCCCCCEEEEECCEECCCHHHH
AA: TNGAVVMGVDAFSPAGKAGLKELDVITEFDGYKVNDIVDL
50 60 70 80
Pred:
Pred: HHHHHHCCCCCEEEEEEEECC
AA: RKRLYQKKVGDRVKVKFYRGG
90 100
D
Pred:
Pred: CCEEEECCCCCCCCCEEEECHHHHHHHHHHHHHCCCEEEE
AA: INTAILAPDGGNIGIGFAIPSNMVKNLTSQMVEYGQVKRG
10 20 30 40
Pred:
Pred: EEEEEEEECCHHHHHHCCCCCCCCEEEEEECCCCCHHHCC
AA: KLGIMGTELNSELAKAMKVDAQRGAFVSQVLPNSSAAKAG
50 60 70 80
Pred:
Pred: CCCCCEEEEECCEECCCHHHHHHHHHHCCCCCEEEEEEEE
AA: IKAGDVITSLNGKPISSFAALRAQVGTMPVGSKLTLGLLR
90 100 110 120
Pred: —
Pred: CC
AA: DG

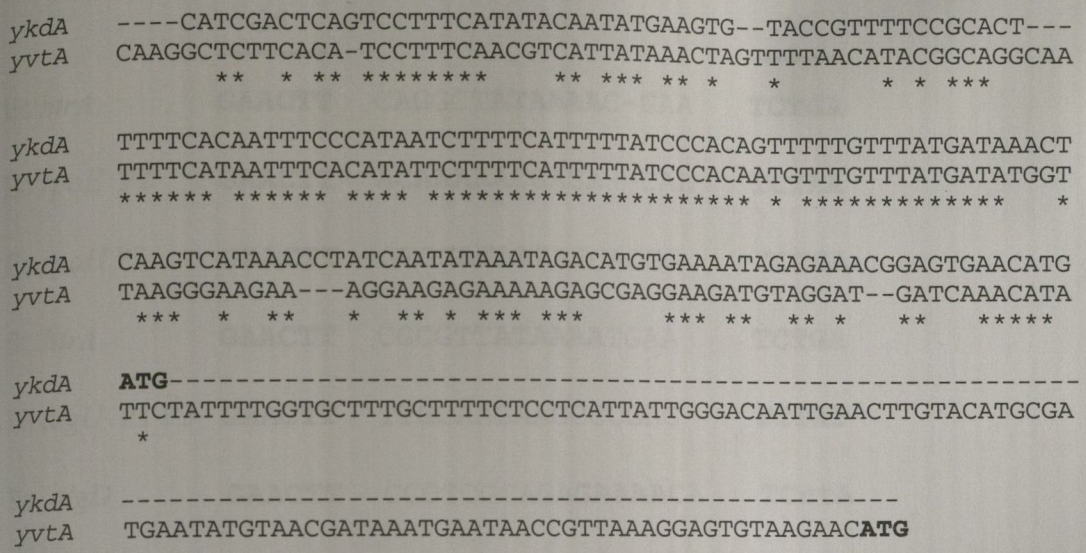


Figure 3.10. Comparison of the controlling regions of *ykdA* and *yvtA* genes. The controlling sequences of *ykdA* and *yvtA* were aligned using the FASTA programme. Identical matches are represented by asterisks (*) and the putative translational start site for each gene is shown in bold.

	-35		-10
Ec *htrA*	**GAACTT**	CAGGCT**ATAAAA**C-GAA	**TCTGA**
Ec *rpoE* P2	**GAACTT**	AC**AAAAA**CGAGA--CAA	**TCTAA**
Ec *rpoH* P3	**GAACTT**	GTGG**ATAAAA**TCACGG	**TCTGA**
St *htrA*	**GAACTT**	CGCGTT**ATAAAA**TGAA	**TCTGA**
Pa *algU* P1 P3	**GAACTT**	TTGCAAGAAGCCCGAG	**TCTAT**
Pa *algD*	**GAACTT**	CCCTCGCAG**AGAAAA**CA	**TCctA**
Sc *dagA* P2	**GAACTT**	TTTGCACGCACGCGAGC	**TCTCG**
Bs *yvtA*	**GAACTT**	GTACATGCGATGAATA	**TgTAA**
Bs *htrA*	**aAACcT**	ATCAAT**ATAAA**TAGACA	**TgTGA**

Figure 3.11. ECF-type promoter consensus sequences upstream of the *ykdA* and *yvtA* genes from *B. subtilis* and in genes from other bacterial species encoding extra-cytoplasmic function proteins. -35 and -10 elements of each promoter are indicated in bold as are adenine rich stretches present in some of these motifs. Ec *htrA*, Ec *rpoE*P2 and Ec *rpoH*P3 are σ^E controlled genes from *E. coli*; Pa *algU*P1P3 and Pa *algD* are AlgU sigma factor controlled genes from *Pseudomonas aeruginosa*; St *htrA* is the σ^E controlled *htrA* gene from *Salmonella typhimurium*; Sc *dagA*P2 is the σ^E controlled gene from *Streptomyces coelicolor*. Lower case bold indicates atypical nucleotides found at the indicated positions, dashes represent gaps introduced for formating purposes.

CHAPTER FOUR

A STUDY OF YkdA ENCODING A HtrA-TYPE SERINE PROTEASE

4.1. INTRODUCTION

Members of the HtrA family of serine proteases are widely distributed among bacteria and have also been found in yeast, *Arabidopsis* and humans (for review see Pallen and Wren, 1997). While some bacterial genomes encode more than one HtrA-like protease, (*e.g. Escherichia coli* and *Bacillus subtilis* each encode three genes), no representative of this family has been identified in the completely sequenced genomes of *Mycoplasma genitalium* and the archaebacteria *Methanococcus janaschii*, *Pyrococcus horikoshii* OT3, *Archaeoglobus fulgidis* and *Methanobacterium thermoautotrophicum*. The three HtrA-like serine proteases in *E. coil*, HtrA (DegP), HhoA (DegQ) and HhoB (DegS), can be divided into three structural regions: (i) an amino terminal region that determines subcellular localisation; (ii) a core domain containing the catalytic triad of amino acids (H, D, S) required for enzymatic function and (iii) one or more PDZ domains positioned in the carboxyl terminus. All three proteins in *E. coli* have signal sequences with HtrA itself being localised to the periplasmic face of the cytoplasmic membrane (Skorko-Glonek *et al.*, 1997; Sassoon *et al.*, 1999). PDZ domains function to recognise peptide motifs usually located at the carboxyl terminus of proteins. The PDZ domain of HtrA functions to assemble protein monomers into the functional hexameric complex (Sassoon *et al.*, 1999). It may also function in targeting the protease to its natural substrate *in vivo* (Fanning and Anderson, 1996; Pallen and Ponting, 1997; Pallen and Wren, 1997). Recent work has shown that *E. coli* DegP can function both as a molecular chaperone and as a protease (Spiess *et al.*, 1999). The switch between these activities is temperature dependent with the chaperone activity predominating at lower temperatures and the protease activity predominating at high temperature (Spiess *et al.*, 1999). Null *htrA* mutants of *E. coli* are thermosensitive and are deficient in degrading abnormal periplasmic proteins (Strauch and Beckwith, 1988; Lipinska *et al.*, 1989). An additional interesting feature of HtrA proteases is that they appear to play an important but as yet uncharacterised role in pathogenesis of some bacteria (Pallen and Wren, 1997). Strains of *Salmonella typhimurium, Brucella abortus and Yersinia enterocolitica* with null mutations in *htrA* genes show attenuated virulence (Johnson *et al.*, 1991; Elzer *et al.*, 1994; Li *et al.*, 1996).

The *htrA* gene of *E. coli* is a member of the SigE stress regulon (Erickson and Gross, 1989). This regulon functions to maintain the periplasm free of misfolded proteins. The induction signals of this regulon include increased levels of non-native periplasmic proteins and unbalanced levels of periplasmic proteins (Mecsas *et al.*, 1993; Rouviere *et al.*, 1995; Connolly *et al.*, 1997). Induction can also be effected by mutation of genes encoding enzymes that participate in periplasmic protein folding (for example DsbA, DsbB, DsbC, and FkpA). Constituent genes of the SigE regulon include *sigE, rpoH, fkpA* (encoding a peptidyl-prolyl isomerase) and *ompK* (Pallen and Wren, 1997; Missiakas and Raina, 1998). The activity of SigE is controlled by a sigma factor/anti-sigma factor partner switching mechanism. RseA is an anti-sigma factor located in the inner membrane that can signal stress in the cell envelope and whose activity can be modulated by RseB and RseC (for reviews see (Pallen and Wren, 1997; Missiakas and Raina, 1998). The SigE / RseA ratio determines the level of SigE activity with the level of RseA in the cell being regulated by DegS, a HtrA paralogue (Ades *et al.*, 1999). The *htrA* gene is also a member of the CpxRA regulon (Danese *et al.*, 1995; Connolly *et al.*, 1997). CpxR/CpxA is a two component system that participates in the response to cell envelope stresses. It regulates expression of genes that have Sig70 (e.g. *ppiA*), Sig32 (e.g. *ppiD*) and SigE (e.g. *htrA*) promoters (Raina *et al.*, 1995; Pogliano *et al.*, 1997; Dartigalongue and Raina, 1998). Constituent genes of the CpxRA regulon include a periplasmically located disulphide oxido-reductase (DsbA), two peptidyl-prolyl-isomerases (PpiA and PpiD) and HtrA, indicating that one role of this regulon is to maintain protein folding homeostasis within the cell envelope. However the functional repertoire of the CpxR / CpxA regulon is likely to be more extensive with recent reports showing that positive autoregulation is effected in conjunction with RpoS and that expression of some chemotaxis and motility genes is also CpxRA dependant (De Wulf *et al.*, 1999; Raivio *et al.*, 1999).

The *B. subtilis* heat shock response can be resolved into four classes of genes (Schumann *et al.*, 2002). There are three well-characterised regulons: HrcA/CIRCE (Hecker *et al.*, 1996), SigB (Hecker and Volker, 1998) and CtsR (Kruger and Hecker, 1998; Derre *et al.*, 1999b). The fourth class comprises a grouping of genes whose expression is responsive to heat

stress, but the mechanism of induction is not effected by HrcA, SigB or CtsR. It is likely that there are heterogeneous heat shock induction mechanisms within this group. Examination of the constituent genes indicates that none of these four gene classes correspond to the SigE or CpxRA regulons identified in *E.coli*. We therefore chose to analyse how expression of *ykdA* (a *htrA* homologue) is regulated in order to ascertain if *B. subtilis* has an extracytoplasmic heat shock response similar to that of *E. coli*. In this study we show that *ykdA* expression is heat-inducible, that increased expression occurs at the transcriptional level and that YkdA negatively regulates its own expression during exponential growth and during heat shock.

4.2. RESULTS

4.2.1. Sequence analysis of *ykdA*

There are three HtrA-like proteases encoded in the *B. subtilis* chromosome, and listed in SubtiList as *htrA*, *yvtB* and *yyxA* (Kunst *et al.*, 1997). Because a functional correlation between the three *B. subtilis* members of this family and the three *E. coli* members (HtrA, HhoA and HhoB) has not yet been established, we give the *B. subtilis htrA* gene its original designation of *ykdA* in this thesis. The *yvtB* gene listed in SubtiList is a truncated HtrA-like protease. We re-sequenced this region of the chromosome and have shown that the full-length serine protease encoding gene (designated *yvtA*, GenBank accession number AF188296, see section 3.2.3) comprises both the *yvtA* and *yvtB* open reading frames listed in SubtiList. The *yyxA* gene has been renamed *yycK* (Fabret and Hoch, 1998). YkdA contains the catalytically important triad of histidine, aspartate and serine residues found in this family of proteases and one PDZ domain. The *E. coli* and *B. subtilis* members of the family are approximately 40% identical (60% similar) within a core region comprising the catalytic and PDZ domains. However, there is little similarity at the amino terminal regions. The amino terminus region of YkdA is longer than the *E. coli* proteases and does not have a recognisable signal peptide. Instead the amino terminal 44 hydrophilic amino acids are followed by a 23 amino acid segment with the potential to form a transmembrane helix suggesting that the protein has a membrane association.

A comparison of the promoter regions of *ykdA* and *yvtA* is shown in Figure 4.1. There is a 56 bp sequence found in both control regions that displays 91% identity (51/56 bases). This level of identity is higher than that observed between the two structural genes confirming the importance of this sequence. Within this region are four copies of an octameric repeat similarly arranged (consensus TTTTCACA). There is a conserved 13 base pair motif (TTTGTTTATGATA) positioned a half-turn of the helix downstream of octamer repeats III and IV and a putative ECF-type promoter located in the control region of *ykdA* that is very similar to the SigE promoter of *htrA* from *E. coli* (Figure 4.1, and also see Figure 3.11).

4.2.2. Induction Of *ykdA* Expression

4.2.2.1. Heat Shock Induction of *ykdA* Expression

To investigate *ykdA* expression in *B. subtilis*, a 341 bp fragment containing the entire intergenic region 5' to the gene and 112 bp of the *ykdA* coding sequence was directionally cloned into the integrating plasmid pDL generating a transcriptional fusion with the *bgaB* (thermostable β-galactosidase) reporter gene. This fusion was then positioned in single copy at the *amyE* locus of the *B. subtilis* chromosome generating strain DN2. The expression profile of this fusion was examined for 60 minutes at 37°C and after temperature upshift to 48°C (Figure 4.2). Only a very low (approximately 2 units) and constant level of activity is observed when cells are grown at 37°C. However, the β-galactosidase activity level increases biphasically when cells are shifted to 48°C with an accumulation of approximately 35 units of activity after 60 minutes growth at this temperature. The biphasic nature of activity accumulation is very reproducible. These data show that expression of *ykdA* is thermoinducible in *B. subtilis*.

4.2.2.2. Effect of other stressors on *ykdA* expression

To establish if *ykdA* expression is induced by stressors other than heat, strain DN2 was grown in media containing ethanol (4% v/v), NaCl (0.3M), H_2O_2 (0.1mM) and puromycin (15μg/ml). The results are shown in Table 4.1. It is clear the range of activities in the unstressed cultures (2-6 U[Miller]) is practically identical to those in the stressed cultures (2-8 U[Miller]). Therefore it is concluded that none of these stressors function to induce *ykdA*. To ensure that induction conditions were operative strain PB198 (*amyE*::*ctc-lacZ*) was exposed to 0.3M NaCl. This gene has a SigB promoter and is therefore inducible by NaCl (Boylan *et al.*, 1993). It can be seen that it is expressed lowly in the absence of NaCl. Ten minutes after NaCl addition there is an increase to 259 U and a drop to 127 U at 30 minutes post induction. This is the expected profile and level of expression for this SigB dependent stress inducible gene.

4.2.3. Expression of *ykdA* is Not Affected in Strains Containing Mutations in the Regulators of Class I (CIRCE/HrcA controlled), Class II (SigB dependent) or Class III (CtsR controlled) Heat Shock Genes.

To investigate whether thermoinduction was mediated through one of the three established heat-shock regulons in *B. subtilis*, the kinetics of *ykdA* thermoinduction was investigated in *hrcA*, *sigB* and *ctsR* null mutant strains. To determine if a mutation in the regulator of Class I genes, HrcA, affected expression of *ykdA*, the *hrcA*::*cat* mutation from strain BT01 (Schulz *et al.*, 1995) was introduced into strain DN3, containing a P*ykdA-lacZ* fusion at the *ykdA* locus, to generate strain DN28. Expression from the P*ykdA-lacZ* fusion in strains DN3 and DN28 was measured throughout the growth cycle in SM media and is presented in Figure 4.3(A). It is clear that the basal level of expression of both strains is similar. Since HrcA is a repressor, these data indicate that *ykdA* is not a member of the CIRCE/HrcA regulon.

To investigate if a mutation in the *sigB* gene, controlling the expression of Class II heat shock and general stress response genes affected the induction of *ykdA*, plasmid pDN2 containing the P*ykdA-bgaB* fusion was transformed into the sigma B mutant strain PB105 to generate strain DN29. Strain DN29 was subjected to heat shock and levels of BgaB activity measured. A comparison of the basal and induced expression profiles for *ykdA* in strains DN2 and DN29 is presented in Figure 4.3(B). The basal levels of expression in both strains are almost identical. The heat shock induced expression of *ykdA* in strain DN29 is more pronounced than that seen in the wild type background of strain DN2. However at 30 and 60 minutes post induction the induced levels in both strains are very similar. These data indicate that induction of *ykdA* is not dependent on the alternate sigma factor SigB. Interestingly the initial heat shock induction of *ykdA* is hastened in a *sigB* mutant background.

The most recently discovered regulator of a heat shock response in *B. subtilis* is the CtsR repressor protein (Kruger and Hecker, 1998) (Derre *et al.*, 1999b) . CtsR negatively

regulates the induction of the *clpC* operon, *clpP* and *clpE* genes in addition to negatively autoregulating its own expression. To determine if the CtsR protein regulated expression of *ykdA*, strain DN30 containing a *ctsR* null allele was constructed. The levels of BgaB activity from the P*ykdA-bgaB* in strain DN30 was measured before and after heat shock and is presented in Figure 4.3(C) together with the levels determined in the wild type background of strain DN2. The basal levels of expression are almost identical in each strain. Heat shock induced expression of *ykdA* is more linear in strain DN30 without the plateau observed in the wild type background. Therefore the absence of the negative regulator CtsR does not increase the expression of *ykdA* to any significant degree. Therefore taken together these data show that thermoinduction of *ykdA* is neither directly nor indirectly controlled by HrcA, SigB or CtsR and must be part of a separate heat-shock regulon in *B. subtilis*.

4.2.4. Investigation of the effects of mutations in genes encoding ECF-type sigma factors on the expression of *ykdA*

The *htrA* gene in *E. coli* is controlled by two different, but partially overlapping regulatory systems (Connolly *et al.*, 1997; Raivio and Silhavy, 1999). One of these systems utilises an alternative essential sigma factor σ^E (De Las Penas *et al.*, 1997a). Sequence analysis shows that the *B. subtilis ykdA* gene possesses a sequence in its control region that displays a high degree of similarity to the σ^E -dependent promoter of *E. coli htrA* (see Figure 3.11). In the sequence of the complete genome of *B. subtilis* there are seven genes encoding putative ECF-type sigma factors (Huang *et al.*, 1998). To determine if any of the putative ECF-type sigma factors contributed to the expression of *ykdA*, a series of strains DN50 - DN55, were generated each harbouring a P*ykdA-bgaB* fusion at the *amyE* locus and carrying a mutated copy of one of six from seven genes encoding putative ECF-type sigma factors. The heat shock induction of *ykdA* was investigated in each strain and the results are shown in Table 4.2. Results for induction of P$_{ykdA}$-*bgaB* in a wild type background show that levels increase to ~9 U(Miller units) at 30 minutes and 14 U at 60 minutes after temperature upshift respectively. It is evident that P$_{ykdA}$-*bgaB* induction in DN50 (*sigV* $^-$), DN51 (*sigW* $^-$), DN53 (*sigY* $^-$), DN54 (*yhdM* $^-$) and DN55 (*ylaC* $^-$) are very similar to those observed in the wild type indicating that heat shock induction of *ykdA* is not mediated by these sigma

factors. It is also evident that there is a modest reduction in *ykdA* induction in a *sigX* mutant background, suggesting partial involvement of this sigma factor in heat shock induction.

4.2.5. Control sequences involved in thermoinduction of *ykdA*

To delineate the promoter elements involved in thermoinduction, the complete *ykdA* control region (cloned in pDN2) was sequentially deleted from the 5'- and 3' ends and the deletion end-points are shown in Figure 4.4. The promoter activity of each fragment was established by generating a *bgaB* transcriptional fusion, which was then placed in single copy at the *amyE* locus. The level of β-galactosidase activity in each strain at 30 and 60 minutes post-thermoinduction in wild-type (strains DN8-DN14) cells is shown in Table 4.3. The biphasic profile of β-galactosidase activity after thermoinduction of strain DN8 (5'Δ47) is similar to that of strain DN2 harbouring the full-length control region. However there is a two-fold increase in the activity level after thermoinduction at T_{30} whereas the level at T_{60} is approximately the same as that observed for the full-length promoter. There is also a two-fold difference observed with this promoter deletion in a *ykdA*⁻ background (compare levels in strains DN15 with DN27) which is observed at both time points. These data suggest the existence of a negative regulatory element positioned upstream of deletion end-point 5'Δ47. Deletion of a further 8 base pairs (strain DN9, 5'Δ55), where most of octamer repeat I is removed, leads to attenuated thermoinduction of *bgaB*, especially at 30 minutes. Thermoinduction is further reduced in strains DN10 (5'Δ79) and DN11 (5'Δ103) with activity levels being virtually undetectable after 60 exposure to heat. Results from the 3' deletions show that *ykdA* expression is heat inducible until a sequence between the 3'Δ61 (DN13) and 3'Δ74 (DN14) is deleted, which results in very low levels of *ykdA-bgaB* expression. A further nuance of the regulation of *ykdA* expression is revealed by strain DN13 (3'Δ61). Deletion of one arm of the putative stem-loop structure is accompanied by a two-fold increase in expression after thermoinduction suggesting the existence of a second negative regulatory element between deletion end-points 3'Δ61 and 3'Δ33. These data show that the sequences required for *ykdA* thermoinduction are located on a 60 bp fragment (between deletion end-points 5'Δ47 and 3'Δ61) containing octamer repeats I-IV and the 13 bp conserved motif and suggest the existence of two negative regulatory elements within the promoter region.

4.2.6. YkdA negatively autoregulates its own expression.

To investigate the possibility that YkdA participates in the regulation of its own expression, strain DN3 was constructed with the following salient features shown in Figure 4.5(A): (i) the full *ykdA* promoter region is fused to the *lacZ* reporter gene and (ii) an intact copy of *ykdA* is placed under the control of the IPTG-inducible SPAC promoter (Yansura and Henner, 1984). The profiles of β-galactosidase accumulation during the growth cycle in the presence of differing levels of IPTG inducer are presented in Figure 4.5(B). In cultures exposed to either 0.1 mM or 1 mM IPTG, β-galactosidase activity accumulates to less than 10 units, the level normally observed during growth at 37°C. However when cells are grown either in the absence of IPTG or in the presence of 10 μM IPTG, β-galactosidase activity increases during exponential growth and reaches a maximum level of approximately 80 units at the transition phase of the growth cycle. This is followed by a sharp decrease in activity levels during the stationary phase of the growth cycle. This 8-fold increase in β-galactosidase expression levels when YkdA levels are low or absent implies that the basal level of *ykdA* expression at 37°C during the growth cycle is negatively regulated by YkdA.

To further investigate negative autoregulation of *ykdA* expression during the heat shock response, the full length promoter construct was transformed into a *ykdA*$^-$ background generating strain DN27. The strain was heat shocked as previously described and the results are shown in Table 4.3. The level of expression is approximately 2-fold higher in the *ykdA*$^-$ strain than in the *ykdA*$^+$ strain when cells are grown at 37°C. However after heat shock at 48°C, expression levels are up to 20-fold higher in the *ykdA*$^-$ strain than in the *ykdA*$^+$ strain at both time points. To delineate the promoter elements responsible for the elevated levels of expression during heat shock, the 5'- and 3'-deleted promoter constructs fused to *bgaB* were transferred into a *ykdA*$^-$ background generating strains DN15 to DN21. Expression levels are tabulated in Table 4.3. The profiles of expression after successive promoter deletions are similar to those obtained in the *ykdA*$^+$ background indicating that the same elements are involved in the elevated expression levels observed in the *ykdA*$^-$ strains. However it is apparent that in cells grown at 48°C the level of expression observed for each deletion

construct is up to 20-fold higher (depending on the deleted construct) than in cells grown at 37°C.

4.2.6.1. Increased levels of YkdA reduce the heat shock induced expression of *ykdA*.

The above data demonstrates that depletion or absence of the YkdA protein increases both the basal and induced levels of expression from the *ykdA* promoter. Further evidence that the level of YkdA protein influences heat induced expression from the *ykdA* promoter was obtained from strain DN33. In this strain the *ykdA* gene is under the control of an non-repressible P_{spac} promoter, i.e. there is no *lacI* gene in this strain, see Figure 4.6(A). Therefore in strain DN33 the *ykdA* gene is expressed constitutively and expression is monitored through a P*ykdA-bgaB* fusion located at the *amyE* locus. The results of heat shock induction of the P*ykdA-bgaB* fusions in strain DN33 compared with strain DN2 are presented in Figure 4.6(B). Both strains show similar levels of BgaB activity when strains were grown at 37°C. However the level of BgaB activity in strain DN33 did not show an immediate increase in response to temperature upshift and remained close to that of the non induced culture. At later time points there is a small increase in BgaB levels in the induced culture of strain DN33 relative to the control culture but the level is 3 to 5 fold reduced compared to the level seen in the wild type background of strain DN2. Although there is no direct evidence of increased YkdA protein levels in strain DN33 presented here, the expression levels of genes placed under Pspac control were comprehensively studied by Vagner (Vagner *et al.*, 1998) In that study the *lacZ* gene was placed under control of the Pspac promoter and levels of LacZ activity were determined in the absence or presence of differing amounts of the inducer IPTG which antagonises the repressor properties of the LacI protein. The results demonstrated that levels of LacZ activity were low and repressed without induction by IPTG. As increasing amounts of IPTG were added to the cultures LacZ activity increased in a linear fashion such that a 200-fold increase in activity was demonstrated at IPTG levels of 300 μM. Therefore in strain DN33, containing no LacI repressor, it is assumed that expression of the *ykdA* gene is induced to an degree comparable to a fully IPTG induced Pspac promoter.

4.2.6.2. The protease activity of YkdA is not required for negative autoregulation

To investigate if YkdA protease activity is required for negative autoregulation, strain AH24 was constructed in which the active site serine residue was changed to a methionine (Figure. 4.7). Mutation of the catalytic serine residue has been shown to inactivate the activity of this protease family (Skorko-Glonek *et al.*, 1995) The expression profiles of a P*ykdA-bgaB* fusion in a *ykdA* null background and in the site specific mutant strain AH24 (YkdA N289H S290M) are shown in Figure 4.8(A) and 4.8(B) respectively. The profile and levels of P*ykdA-bgaB* expression before and after heat shock of AH24 is very similar to that observed for wild type, compare Figure 4.8(B) and Figure 4.2 and differs from the parental *ykdA* null mutant of strain AH23 (Figure 4.7B). It is evident therefore that the protease activity is not required for negative autoregulation of *ykdA* expression.

4.2.7. Transcriptional analysis of *ykdA* expression.

Expression of *ykdA* was analysed by Northern blotting (Figure 4.9A) and by primer extension (Figure 4.9B) analysis to establish transcript size and transcription initiation point. Total RNA samples were prepared from exponentially growing *B. subtilis* 168 cells before (0 minutes) and at 5, 9, 13 and 17 minutes after heat shock. The results of the Northern blot show that there is a single transcript in all samples that migrates at approximately the same position as the 1383 base RNA marker. The level of this transcript is significantly increased in the 50°C samples at 5, 9 and 13 minutes but decreases to non-heat shocked levels at 17 minutes after heat shock.

Primer extension analysis was performed on RNA samples prepared from strain DN26 (*ykdA*Δ439) since expression levels are higher a strain mutated in *ykdA* (see section 4.2.6). A doublet of reverse transcripts was observed in all samples that mapped to the adjacent CA bases marked by an asterisk in Figure 4.9. This demonstrates that thermoinduction is effected from a single promoter. While there is a good match to the –10 consensus (TATGAT) of a SigA promoter corresponding to this initiation point of transcription, there is no apparent SigA –35 region (Figure 4.4). The level of reverse transcript in the 50°C

samples was significantly higher than that observed in the 37°C samples consistent with the observation that YkdA is a negative regulator of its own expression. These results indicate that heat shock induction of *ykdA* occurs at the transcriptional level from a single promoter that has a good match to the –10 region of a SigA promoter but no apparent match to the –35 region. The size of the single transcript is consistent with the monocistronic operon predicted from the sequence.

4.2.8. Phenotype of *ykdA* mutants

To establish if mutation of *ykdA* leads to altered sensitivity to stressors, strain DN26 (*ykdA*Δ439) was exposed to increased temperature (54°C) and hydrogen peroxide (10mM) and cell survival profiles were compared with the parental *B. subtilis* strain 168. Cell survival profiles (the values are the average of at least three independent experiments) after heat (Figure 4.10A) and hydrogen peroxide (Figure 4.10B) exposure are shown. It is evident that inactivation of *ykdA* leads to increased thermotolerance at all the times sampled after temperature upshift. Similarly, inactivation of *ykdA* leads to an increased tolerance to hydrogen peroxide. It is evident the survival levels of wild-type and mutant cells are similar (2-3 logs killing) after 5 minutes exposure to hydrogen peroxide. However whereas exposure of wild-type cells for a further 30 minutes results in an additional 2 logs killing, the survival of the *ykdA* mutant is unaffected by this extended treatment. These experiments show that mutation of *ykdA* leads to increased tolerance to heat and hydrogen peroxide indicating that YkdA expression and the peroxide stress response are somehow linked in *B. subtilis*.

4.2.8.1. Increased hydrogen peroxide resistance of strain DN26 is not mediated through PerR inactivation.

Resistance to hydrogen peroxide in *B. subtilis* is primarily mediated by the action of catalases. *B. subtilis* contains three catalases, encoded by the *katA*, *katE* (*katB*) and *katX* genes. KatA is the primary catalase in exponentially growing cells and is a member of the PerR hydrogen peroxide inducible regulon (Bol and Yasbin, 1991; Herbig and Helmann, 2002). Expression of the *katE* and *katX* genes are regulated by the general stress regulatory

sigma factor σ^B and inducible by a variety of stresses but not by hydrogen peroxide (Engelmann *et al.*, 1995; Petersohn *et al.*, 1999). The *katX* gene is in addition under the control of the sporulation specific sigma factor σ^F (Petersohn *et al.*, 1999). To further investigate the increased resistance of strain DN26 to hydrogen peroxide it was decided to determine if mutation of *ykdA* lead to derepression of the PerR regulon. Such derepression would lead to increased expression of *katA* and *ahpCF* that would decrease the susceptibility of cells to oxidising agents. Therefore strain DN34 was constructed in which an *ahpCF-lacZ* fusion was placed in the *ykdA* null background of strain DN26. Profiles of *ahpCF-lacZ* expression before and after hydrogen peroxide induction in strain DN34 were compared with those for strain ES993 under similar conditions. The results presented in Figure 4.11 show that the profiles and levels of *ahpCF-lacZ* expression both before and after induction by hydrogen peroxide are very similar. These data show that there is no general derepression of PerR regulated genes in a *ykdA* mutant background in the absence of hydrogen peroxide. In addition the kinetics of *ahpCF* induction in response to hydrogen peroxide are very similar in wild type and *ykdA* mutant cells. These data indicate that the increased resistance of *ykdA* mutant cells to hydrogen peroxide is not mediated through interference with PerR controlled expression of enzymes that protect against peroxides.

4.3. DISCUSSION

There are three members of the HtrA serine protease family encoded in the *B. subtilis* genome (Kunst *et al.*, 1997). An analysis of one of these genes *ykdA* is presented in this thesis. The core domain (catalytic region and one PDZ domain) of YkdA is 38% identical (61% similar) to the other *B. subtilis* members of this family. This core region was used to generate a phylogenetic tree of each of the three HtrA members from *B. subtilis* and *E. coli* using HtrA from *Synechocystis* as an outgroup. The tree shows that the three *B. subtilis* members are more closely related to each other than to any of the *E. coli* members and that *ykdA* is more closely related to *yvtA* than to *yycK* (Figure 4.12). This grouping is supported by the conservation of a regulatory motif in the promoter regions of *ykdA* and *yvtA*. Therefore duplication of the genes encoding HtrA-like serine proteases most likely occurred after the divergence of *E. coli* and *B. subtilis*.

Expression of *ykdA* is thermoinducible, which is effected at the transcriptional level. Primer extension analysis shows that thermoinducible expression is driven by a promoter with a SigA-type –10 region but the –35 region shows no similarity to SigA-type promoters. There is a series of octamer repeats positioned in this region of the promoter, at least one of which is necessary for normal thermoinduction. The importance of these repeats is indicated by the fact that both their number and positioning are conserved in the promoter region of *yvtA*, encoding a second HtrA-like serine protease in *B. subtilis*. Therefore we predict that *yvtA* is also thermoinducible by a similar mechanism to *ykdA*. Our data also suggest the existence of two negative regulatory elements within the promoter region. These elements are worthy of further investigation, with the caveat that they may be due to novel juxtaposition of promoter and plasmid sequences generated through the deletion process. However each region has a particular feature, (octamer repeats and a stem-loop in the upstream and downstream regions respectively) suggesting that the negative regulatory elements are physiologically relevant. Although both *ykdA* and *yvtA* have potential ECF-type promoter sequences (showing high homology to SigE-dependent promoters from *E. coli*; indicated in Figure 3.11) positioned in their control regions, we were unable to activate this putative *ykdA* promoter by a variety of

stressors including heat. Therefore it is likely that *ykdA* expression responds to additional stimuli that can induce an ECF-type sigma factor-controlled regulon.

Expression of *ykdA* is also negatively autoregulated both during exponential growth at 37°C and during heat shock at 48°C. The level of β-galactosidase steadily increases in *ykdA*$^-$ cells throughout exponential growth in contrast to *ykdA*$^+$ cells where expression levels are low and constant. This negative regulation is manifest most clearly however in heat shocked cells: at 60 minutes post-thermoinduction the level of β-galactosidase accumulation is 20-fold higher in *ykdA*$^-$ cells than in *ykdA*$^+$ cells. Primer extension and Northern analysis show that the increased expression occurs at the level of transcription. We have evidence to show that the YkdA protease activity is not required to mediate negative autoregulation. The observed profiles of thermoinduction suggest that negative autoregulation operates by YkdA controlling the level of the inducing signal. Therefore in *ykdA* null mutants grown at 37°C, loss of YkdA results in only a small increase in the inducing signal. However growth at 48°C results in a high level of inducing signal that remains high and persistent due to the absence of YkdA.

Counterintuitively, null *ykdA* mutants are more resistant than wild-type cells to heat and to hydrogen peroxide exposure. Mutant *ykdA* cells are approximately 10-fold more resistant to heat exposure at 54°C than wild-type cells. This result differs from those obtained with some other bacteria where mutation of *htrA* leads to a thermosensitive phenotype (Pallen and Wren, 1997). In addition whereas null *ykdA* mutants of *S. typhimurium* (Johnson *et al.*, 1991), *Yersinia enterocolitica* (Yamamoto *et al.*, 1996) and *Pseudomonas aeruginosa* (Boucher *et al.*, 1996) are more sensitive than wild-type strains to oxidative stress, the *B. subtilis ykdA* null mutant is up to 80-fold more resistant to hydrogen peroxide than wild-type cells. It has been established in *E. coli* that the relationship between HtrA and oxidative stress involves the cell envelope. Cumene hydroperoxide (which partitions to the membrane) induces *htrA* and *htrA* mutants are more sensitive to this oxidising agent than are wild-type cells. However mutant *htrA* cells are not more sensitive to hydrogen peroxide than wild-type cells nor does hydrogen peroxide induce *htrA* expression (Skorko-Glonek *et al.*, 1999). In

addition ferrous ions lead to oxidative damage of membrane proteins that can be alleviated by membrane associated anti-oxidants but not by cytosolic antioxidants (Skorko-Glonek *et al.*, 1999). Mutation of *ykdA* in *B. subtilis* obviously results in a stimulus that leads to increased expression of a gene(s) that protect cells against both heat and hydrogen peroxide. One such candidate gene might be *yvtA*, encoding the closely related YkdA paralogue. Our evidence shows that the *yvtA* expression level increases when *ykdA* is mutated (see section 5.2.2). Therefore the increased thermo- and perhaps oxidative-tolerance of *ykdA* mutants may be due to the compensatory patterns of expression of these two proteases.

Table 4.1. Levels of β-galactosidase activity from strains DN2 (P*ykdA-bgaB*) and PB198 (P*ctc-lacZ*) exposed to different stressors.

Strain	Stressor	β-galactosidase Activity (Miller units)				
		Control culture			Stressed culture	
		T_0	T_{10}	T_{30}	T_{10}	T_{30}
DN2	No stressor	3	4	2	N/A[b]	N/A
DN2	Ethanol (4% vol/vol)	4	3	6	6	8
DN2	H_2O_2 (0.1mM)	4	3	5	4	8
DN2	Puromycin (15μg/ml)	2	2	5	2	4
DN2	NaCl (0.3M)[a]	5	6	6	5	6
PB198	NaCl (0.3M)[a]	3	9	10	259	127

[a] Salt stress was imposed by adding an appropriate amount of a 5M NaCl solution to cultures growing in buffered LB, all other cultures were grown in regular LB.

[b] N/A, not applicable.

Table 4.2. Expression from a P*ykdA-bgaB* fusion in strains containing mutations in genes encoding putative ECF-type sigma factors

Strain (ECF-σ mutation)	BgaB activity (Miller units)			
	37°C		48°C	
	T 30	T60	T30	T60
DN50 (*sigV*) [a]	1	1	9	16
DN52 (*sigX*)	1	1	4	10
DN51 (*sigW*)	1	2	7	14
DN53 (*sigY*)	1	2	6	13
DN54 (*yhdM* [*sigM*])	1	1	8	16
DN55 (*ylaC*)	1	1	9	16
DN2 (wild type)	2	3	9	14

[a]Strain DN50 is a *sigV* conditional mutant; BgaB activities were obtained in the absence of IPTG.

Table 4.3. Expression of *bgaB* fused to deleted derivatives of the *ykdA* promoter in *ykdA*$^+$ and *ykdA* mutant strains.

Strain	BgaB [a]sp. act. in *ykdA*$^+$ strains				Strain	BgaB [a]sp. act. in *ykdA* mutant strains			
	37°C		48°C			37°C		48°C	
	T_{30}[b]	T_{60}	T_{30}	T_{60}		T_{30}	T_{60}	T_{30}	T_{60}
DN2 (full)	2	4	10	39	DN27 (full)	6	12	230	723
DN8 (5' Δ47)	3	4	17	39	DN15 (5' Δ47)	12	25	275	825
DN9 (5' Δ55)	2	1	4	24	DN16 (5' Δ55)	1	1	11	55
DN10 (5' Δ79)	0	0	1	9	DN17 (5' Δ79)	0	0	0	5
DN11 (5' Δ103)	0	0	0	2	DN18 (5' Δ103)	0	0	0	3
DN12 (3' Δ33)	2	3	18	27	DN19 (3' Δ33)	16	24	308	729
DN13 (3' Δ61)	5	7	40	58	DN20 (3' Δ61)	78	108	479	940
DN14 (3' Δ74)	2	2	5	5	DN21 (3' Δ74)	0	0	0	0

[a] specific activity, one activity unit is defined as 1nmol of *o*-nitrophenyl-β-D-galactopyranoside hydrolysed per min per mg of protein. The background activity level found with a promoterless *bgaB* fusion has been subtracted from each value.

[b] Time in miutes after splitting the culture.

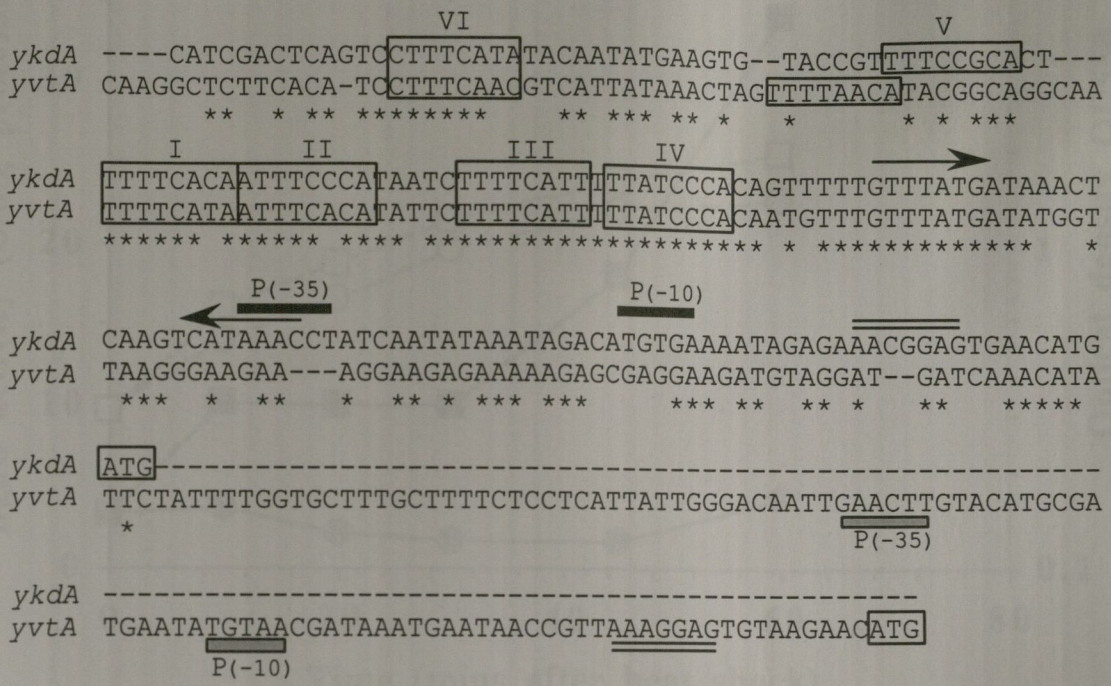


Figure 4.1. A comparison of the promoter regions of *ykdA* (upper sequence) and *yvtA* (lower sequence). Nucleotide identity is indicated by an asterisk below the sequences and gaps introduced to maximize the alignment are signified by dashes (-). The octameric repeats are boxed (labelled I through VI). The ECF-type promoters are indicated by solid bars over the sequence (*ykdA*) and hatched bars under the sequence (*yvtA*) signifying the —35 and —10 regions (P-35 and P-10) respectively. The arrows represent inverted repeats, the putative ribosome binding sites have double lines over (*ykdA*) or under (*yvtA*) the sequence and the boxed ATG is the putative start codon of each gene.

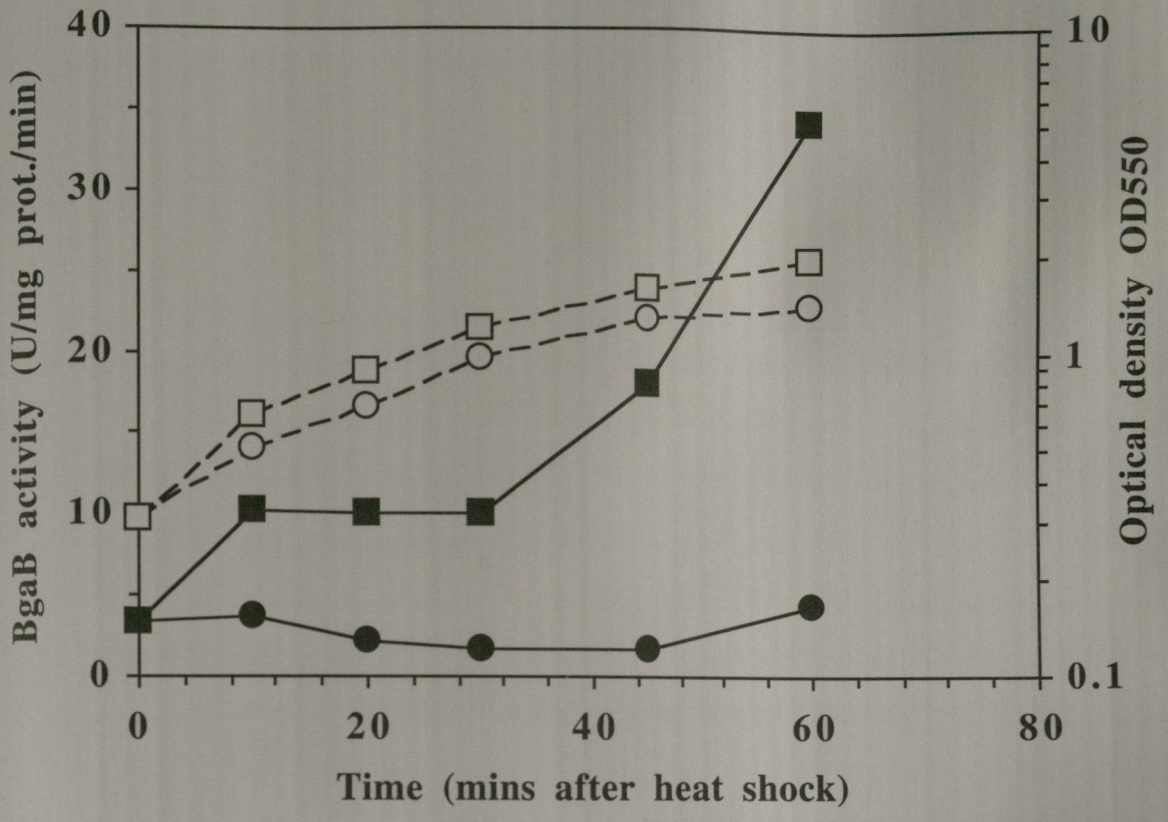


Figure. 4. 2. Profile of growth and β-galactosidase accumulation in strain DN2 during growth at 37°C and after heat-shock at 48°C. Growth (open circles) and β-galactosidase accumulation (closed circles) at 37°C: growth and β-galactosidase accumulation at 48°C are indicated by open and closed squares, respectively.

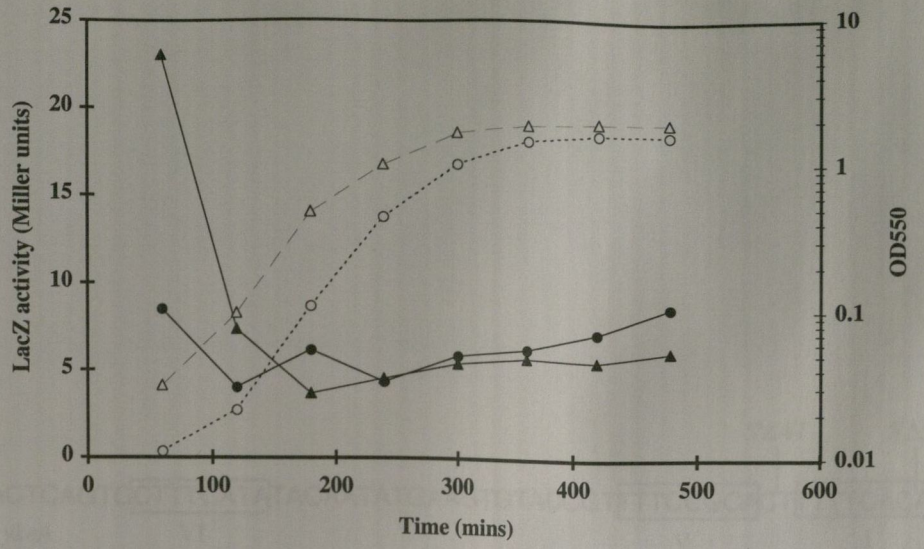
Figure. 4.3. Expression of *ykdA* transcriptional fusions in strains mutated in the regulators of class I (A), class II (B) and class III (C) heat shock regulons.

(A) Level of P*ykdA*-*lacZ* expression in strain DN3 (P*ykdA*-*lacZ* P*spac*-*ykdA*) (●) and strain DN28 (P*ykdA*-*lacZ* P*spac*-*ykdA* *hrcA*::cat) (▲). Both strains were grown in SM broth in the presence of 1mM IPTG. The growth of the strains measured by monitoring the optical density at 550nm (OD550) is shown for the duration of the experiment by open circles (O) for DN3 and by open triangles (Δ) for DN28.

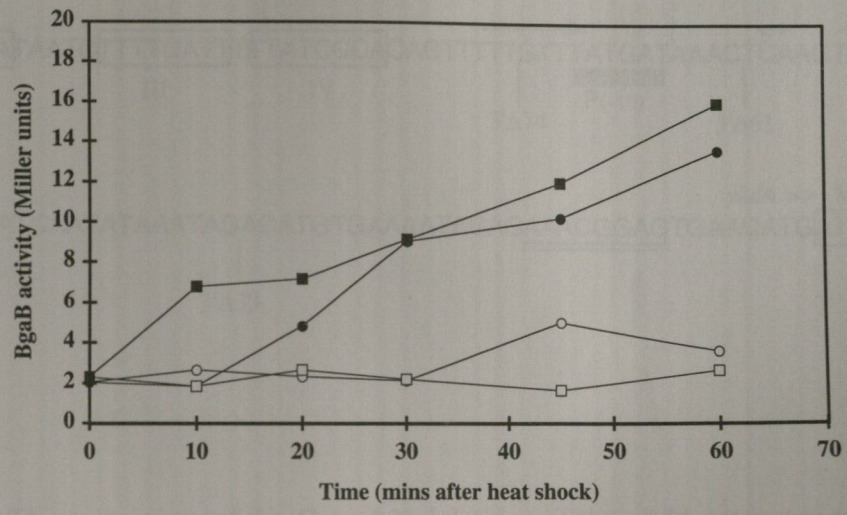
(B) Levels of BgaB activity in strain DN2 (*amyE*::P*ykdA*-*bgaB*) grown at 37˚C (O) and at 48˚C (●); and in strain DN29 (*amyE*::P*ykdA*-*bgaB sigB*Δ2) grown at 37˚C (□) and at 48˚C (■). Strains were grown in LB broth.

(C) Levels of BgaB activity in strain DN2 (*amyE*::P*ykdA*-*bgaB*) grown at 37˚C (O) or 48˚C (●); and in strain DN30 (*amyE*::P*ykdA*-*bgaB ctsR::erm*) grown at 37˚C(◊) or 48˚C (♦). Strains were grown in LB broth.

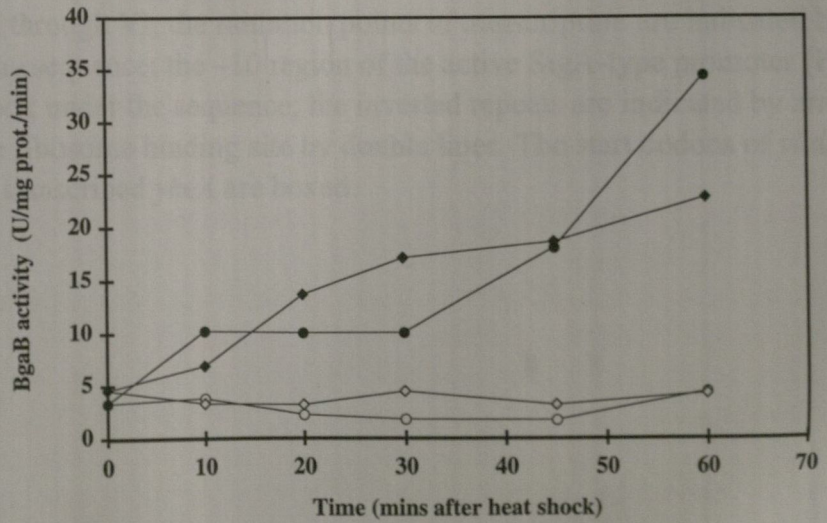
A


LacZ activity (Miller units)
OD550
Time (mins)
0
5
10
15
20
25
0.01
0.1
1
10
0
100
200
300
400
500
600

B


BgaB activity (Miller units)
Time (mins after heat shock)
0
2
4
6
8
10
12
14
16
18
20
0
10
20
30
40
50
60
70

C


BgaB activity (U/mg prot./min)
Time (mins after heat shock)
0
5
10
15
20
25
30
35
40
0
10
20
30
40
50
60
70

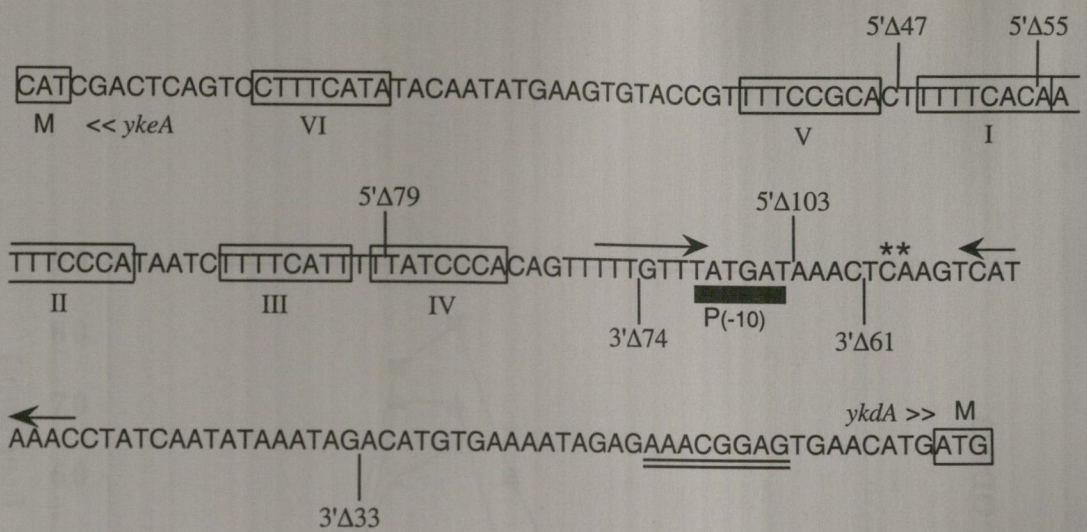
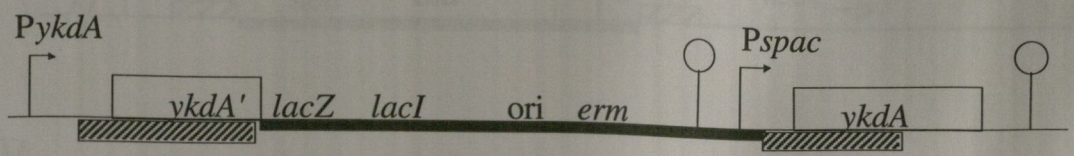


Figure. 4.4. The end-points of the 5'- and 3'-deletions of the *ykdA* promoter region generated by Bal31 exonuclease. The nomenclature for each deletion indicates the end from which the deletion was made (5'-end and 3'-ends above and below the sequence respectively) and the number of bases deleted from the intergenic region. Some features of the promoter region are indicated: the octameric repeats are boxed and numbered I through VI; the initiation points of transcription are indicated by asterisks over the sequence; the –10 region of the active SigA-type promoter (P-10) is indicated by a box under the sequence; the inverted repeats are indicated by arrows and the putative ribosome binding site by double lines. The start codons of *ykdA* and the divergently transcribed *ykeA* are boxed.

A



B

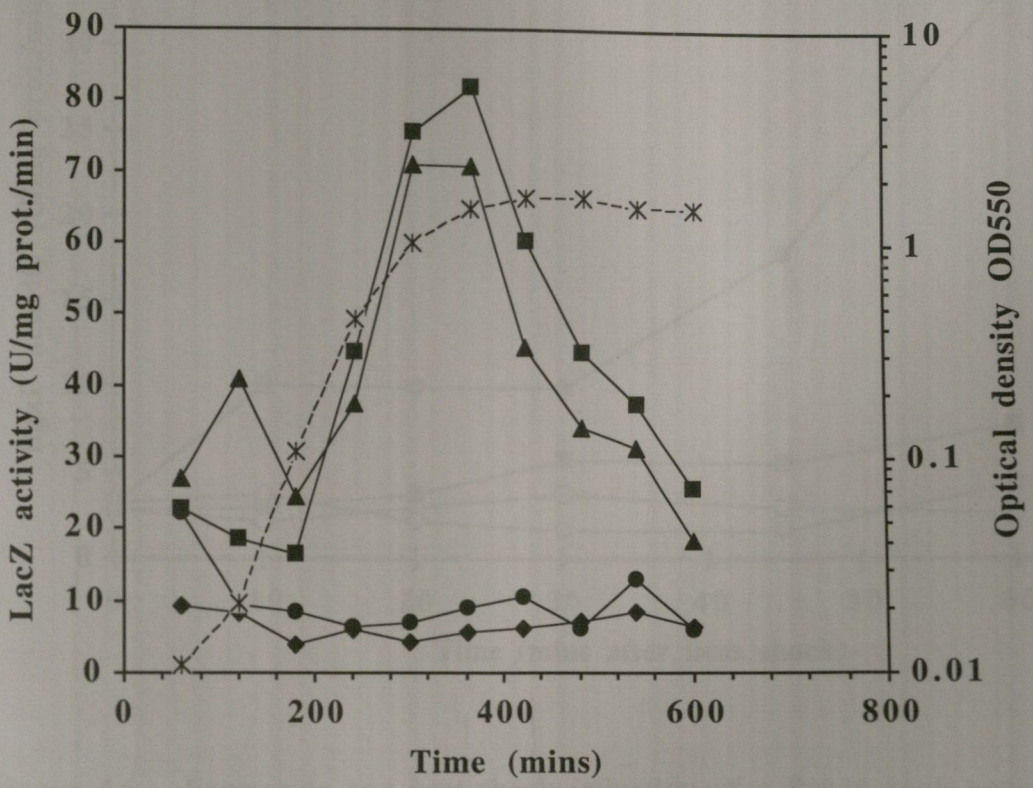
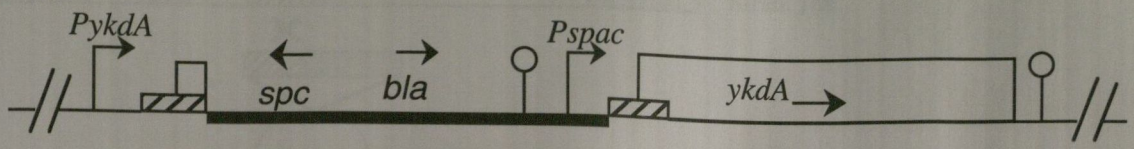


Figure. 4. 5. Negative autoregulation of *ykdA*. (A) A schematic of the construct at the *ykdA* locus in strain DN3. Promoters (P_{ykdA} and P_{SPAC}) are indicated by bent arrows. In this strain, the full intergenic *ykdA* control region is transcriptionally fused to the *lacZ* gene and the intact *ykdA* structural gene (open box) is under the control of the inducible P_{SPAC} promoter. Bold line, plasmid DNA; hatched box, fragment from upstream region of *ykdA* used to integrate pDN3 by Campbell-type integration; *ykdA*', 3' truncated *ykdA* ; lollipops represent transcriptional terminators. The figure is not to scale. (B) Expression of the P_{ykdA}-*lacZ* transcriptional fusion during the growth cycle in SM medium containing different levels of IPTG: no IPTG (▲), 10μM IPTG (■), 100μM IPTG (◆) and 1mM IPTG (●). The growth cycle is indicated by the dashed line and is representative for the four growth curves.

A



B

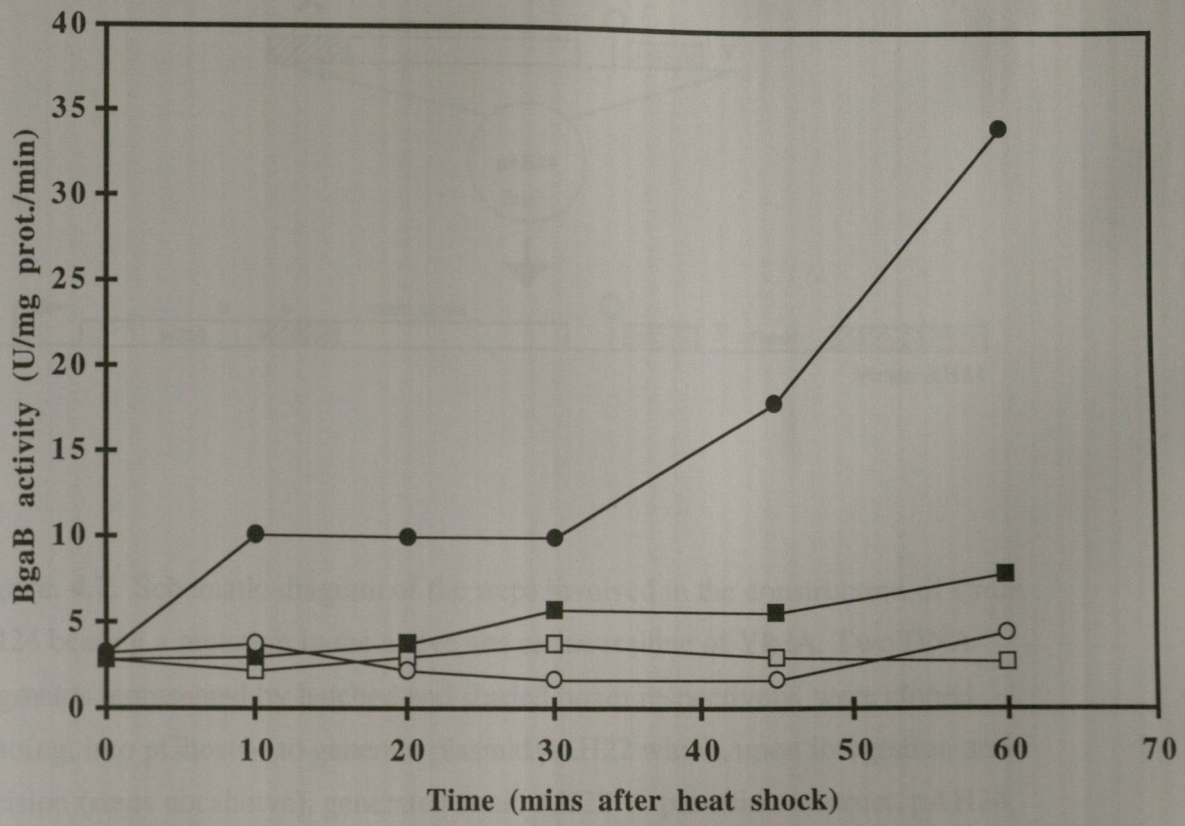


Figure. 4.6. Expression and heat shock induction of a P*ykdA-bgaB* transcriptional fusion in a strain overproducing YkdA. (A) Schematic diagram of the integration of plasmid pDN33 at the *ykdA* locus. The deleted *ykdA* promoter fragment 5'Δ103 mediating integration is indicated by the hatched box, the bent arrows represent promoters, the open box represents the *ykdA* structural gene, arrows represent direction of transcription, lollipops indicate transcriptional terminators, heavy and thin lines represents plasmid and chromosomal DNA respectively, *spc* and *bla* are genes encoding resistance to spectinomycin and ampicillin.

(B) Exponentially growing cultures of strains DN33 (P_{spac} *-ykdA lacI*$^-$ *amyE*::P*ykdA-bgaB*) (■, □) and DN2 (*amyE*::P*ykdA-bgaB*) (O, ●) were subjected to heat shock and levels of BgaB measured from cultures growing at 37°C (open symbols) and 48°C (closed symbols).

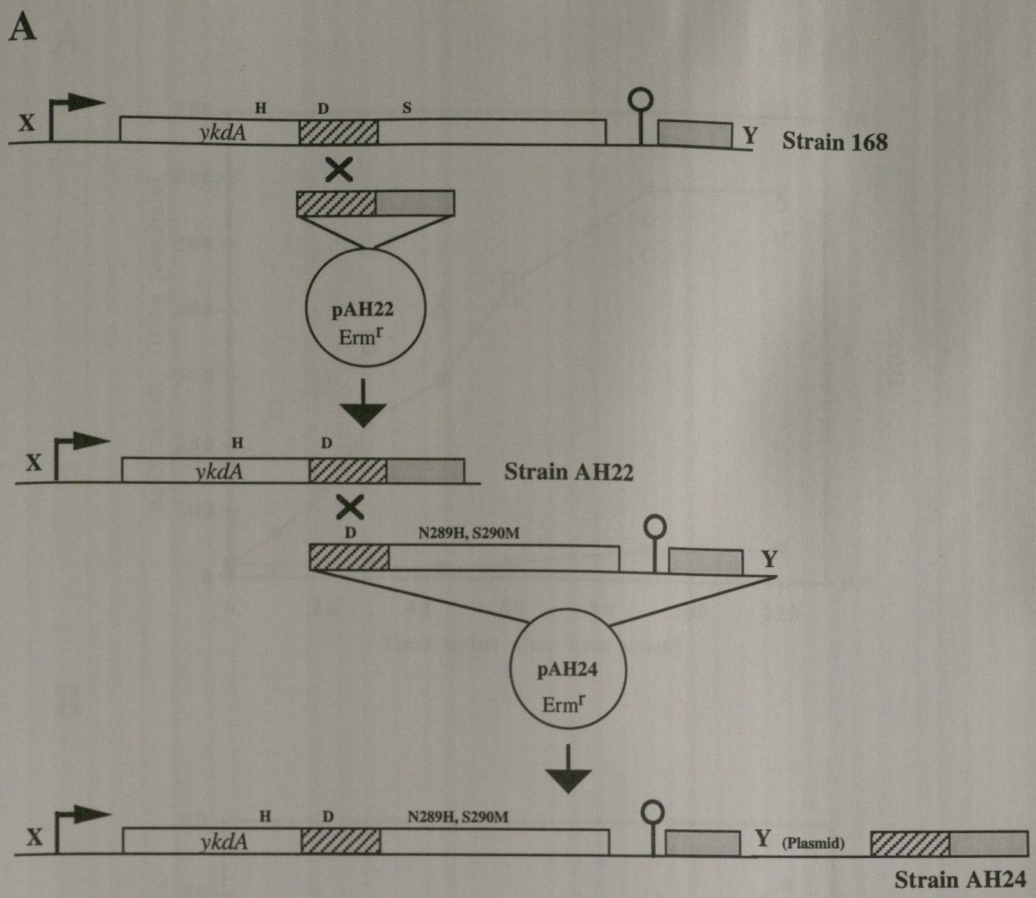
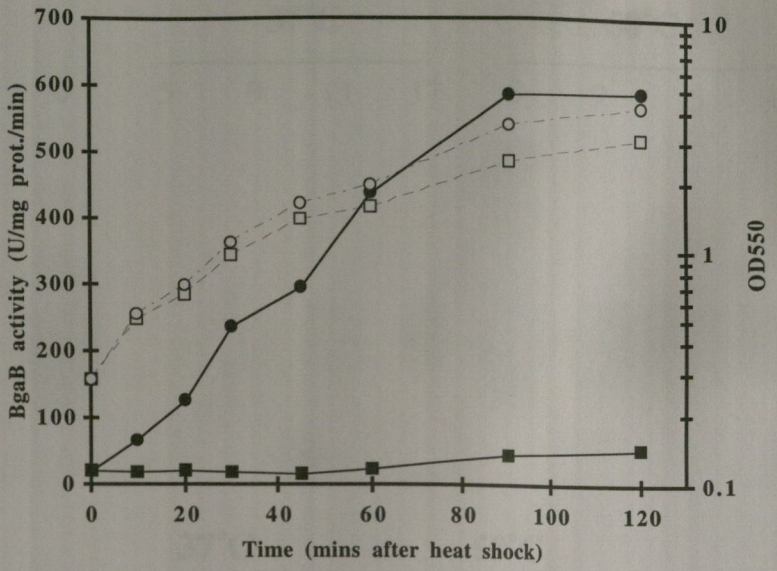


Figure. 4.7. Schematic diagram of the steps involved in the construction of strain AH24 bearing a mutation in the active site serine residue of YkdA. Two DNA fragments represented by hatched and shaded boxes respectively, were cloned abutting, into pGhost4+ to generate plasmid pAH22 which, upon integration and excision (steps not shown), generated strain AH22. A plasmid construct, pAH24, containing the desired site specific mutations was integrated into the deleted *ykdA* allele to regenerate an intact *ykdA* gene in strain AH24. H, D and S represent the catalytic triad of active site residues; X and Y are included for orientation purposes.

A



B

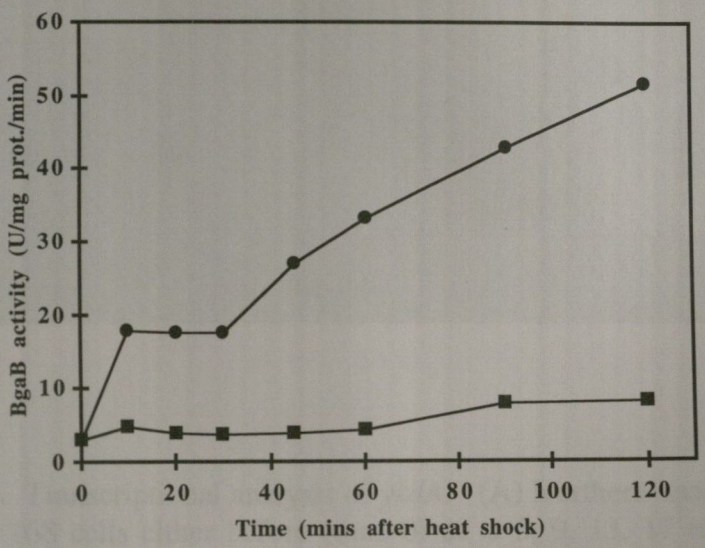


Figure 4.8. Role of YkdA protease activity in negative autoregulation of P*ykdA-bgaB* activity. (A) Expression of the P*ykdA-bgaB* transcriptional fusion in strain AH23 (*ykdA*Δ570 *amyE*::P*ykdA-bgaB*) at 37°C (■) and 48°C (●). The growth profiles of the strains at 37°C (□) and at 48°C (O) during the course of the experiment are also shown. (B) Expression of the P*ykdA-bgaB* transcriptional fusion in strain AH24 (*ykdA*1 [N289H S290M] *amyE*::P*ykdA-bgaB*) containing an intact *ykdA* gene mutated at the active site (N289H S290M). BgaB activity is shown for cells grown at 37°C (■) and 48°C (●) for the duration of the experiment

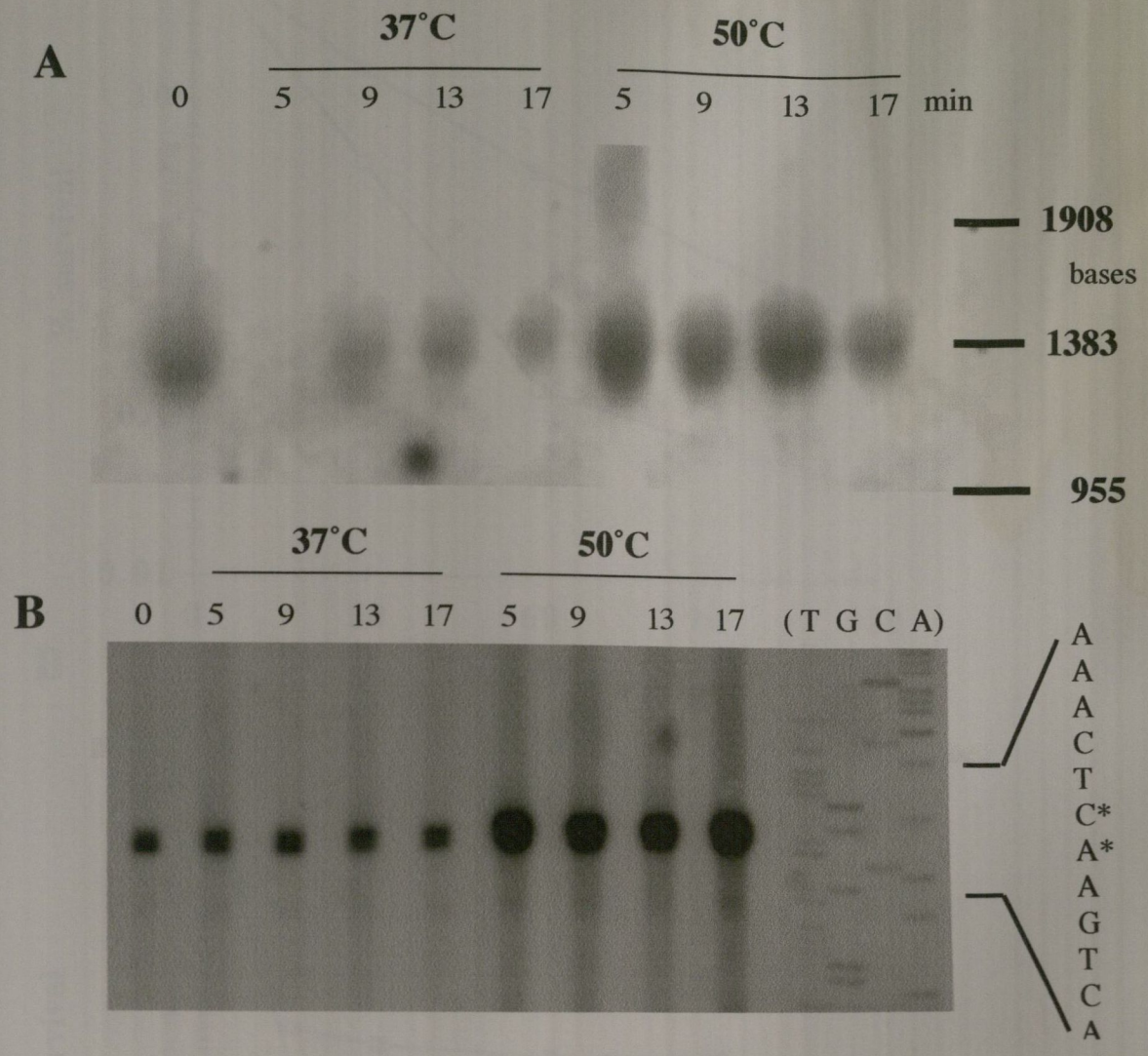


Figure 4. 9. Transcriptional analysis of *ykdA*. (A) Northern analysis: RNA was prepared from strain 168 cells either before (time 0) or at 5, 9, 13, 17 minutes after splitting the culture. Twenty five μg of total RNA was loaded onto each lane (the sample in the 5 minute lane at 37°C was lost). The positions to which the RNA size markers (1908, 1383 and 955 bases) migrated are indicated. (B) Primer extension analysis: RNA was prepared from strain DN26 (*ykdA*Δ439) before (0 minutes) and at 5, 9, 13 and 17 minutes after splitting the culture. One half of the culture was maintained at 37°C while the other was heat shocked at 50°C. The amount of transcript in each lane is that obtained from 12.5μg of total RNA. A sequencing ladder to size the reverse transcript is shown at the right of the figure and the complement of a portion of this sequence is indicated beside it. The two bases at which transcription initates are beside it.

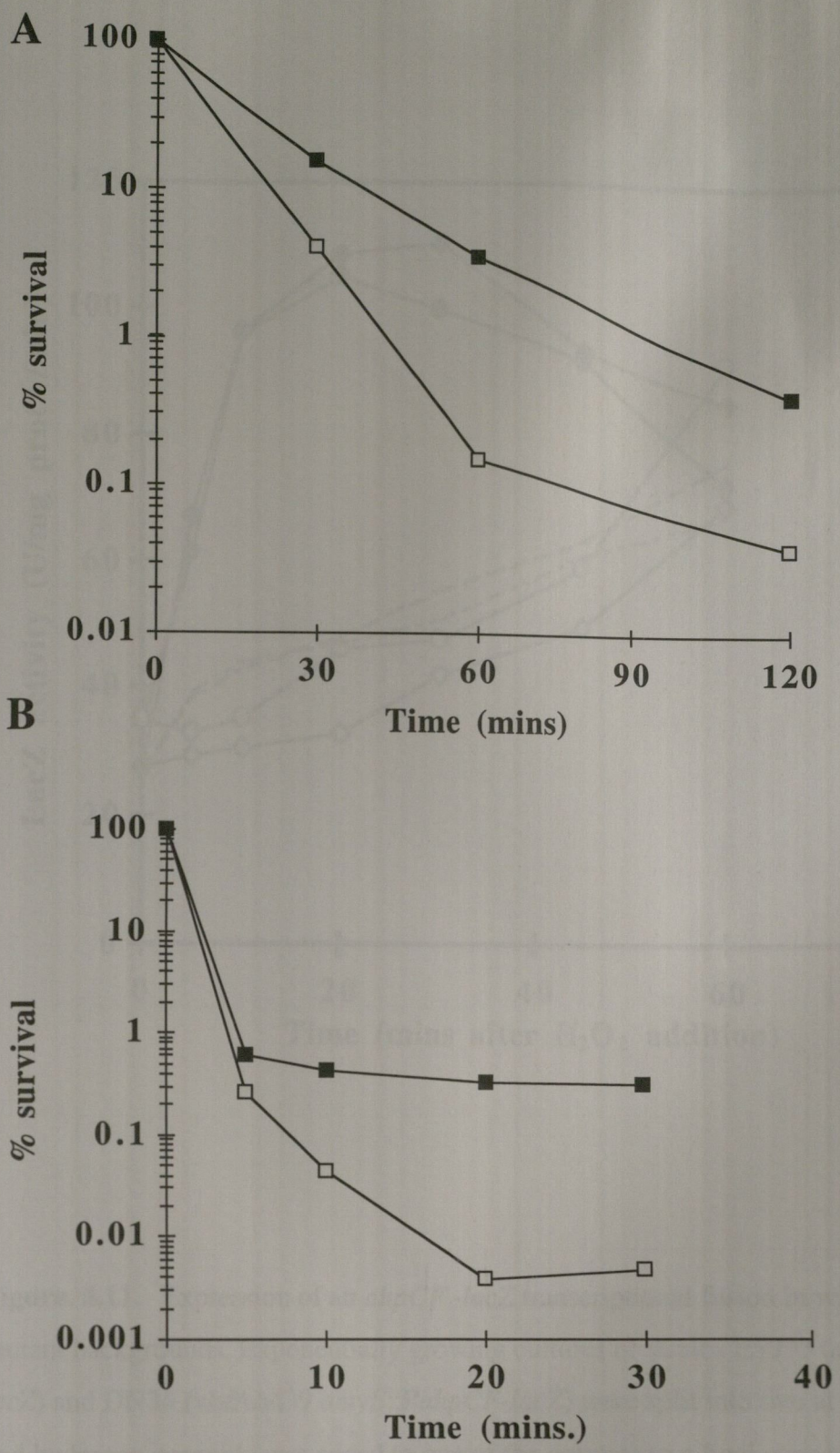


Figure. 4.10. Survival of wild-type and DN26 (*ykdA*Δ439) strains after exposure to heat (A) and hydrogen peroxide (B). Exponential phase cells were exposed to heat at 54°C and 10 mM hydrogen peroxide for the times indicated. Wild-type cells are indicated by open squares (□) and *ykdA* mutant cells are indicated by closed squares (■).

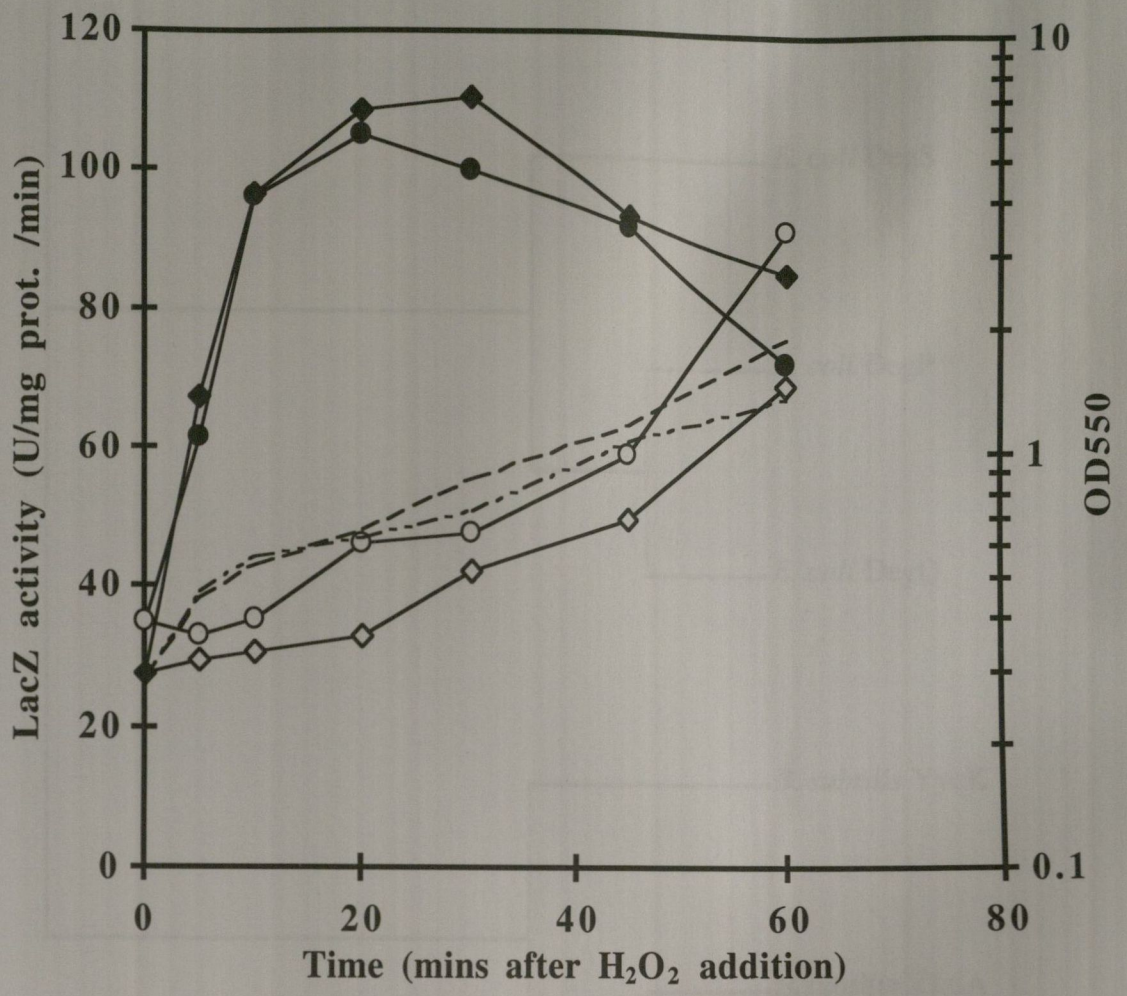


Figure. 4.11. Expression of an *ahpCF -lacZ* transcriptional fusion in wild type and *ykdA* mutant backgrounds. Exponentially growing cultures of strains ES993 (*amyE*::P*ahpCF-lacZ*) and DN34 (*ykdA*Δ439 *amyE*::P*ahpCF-lacZ*) were split into two at OD_{550} ≈0.3 (T0) and hydrogen peroxide was added to one of the cultures to a final concentration of 60 μM. The growth of strains in the presence or absence of hydrogen peroxide were very similar; the growth of strain ES993 is represented by dashed lines. LacZ activity is shown for strain ES993 (O, ●) and for strain DN34 (◊, ◆), open symbols represent activity from control cultures and closed symbols activity from induced cultures.

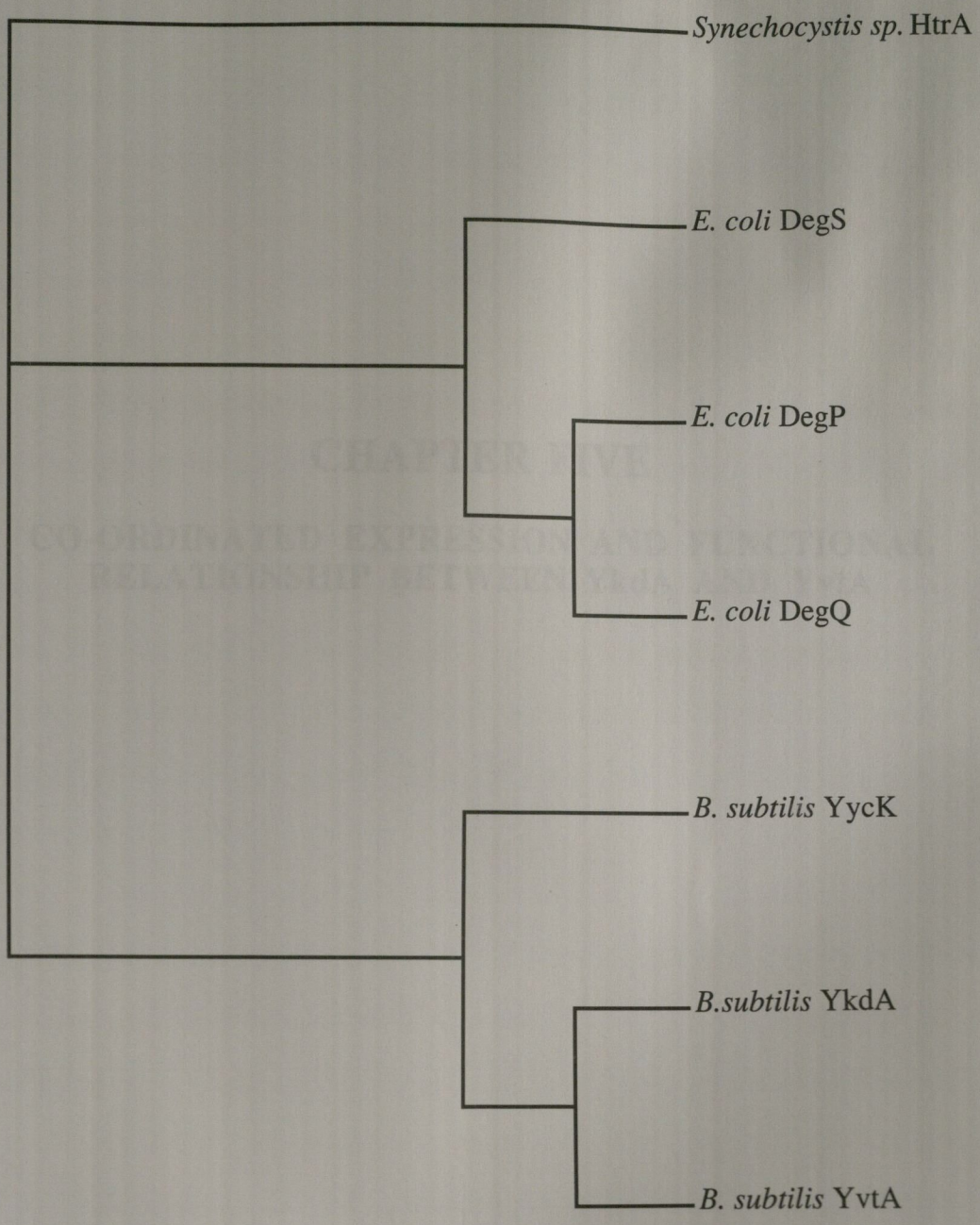


Figure 4.12. Phylogenetic tree of HtrA-like proteases from *E. coli* and *B. subtilis*. The core catalytic domains of the three proteases in *E. coli* and *B. subtilis* were aligned using CLUSTALW and the the resulting dendogram file from the output was used to draw the tree using TREEVIEW. The HtrA-like protein sequence of *Synechocystis sp.* PCC6803 was also used in the alignment as an outgroup. Accession numbers for the proteins analysed are DegP, P09376; DegQ, P39099; DegS, P31137; YkdA, AJ002571; YvtA, AF188296; YycK, Z99124 and *Snyechocystis* HtrA, NP440705.

CHAPTER FIVE

CO-ORDINATED EXPRESSION AND FUNCTIONAL RELATIONSHIP BETWEEN YkdA AND YvtA

5.1. Introduction

Members of the HtrA-family of serine proteases are widely distributed in nature from bacteria to humans (Pallen and Wren, 1997). The proteins are characterised by an amino terminal domain that participates in protein localisation, a catalytic domain containing an active serine residue and a PDZ domain that functions in multimerization of the protein into the active hexameric structure and perhaps also in identifying target proteins. Information derived from completely sequenced genomes shows that most eubacteria have a single HtrA-like serine protease. A recognisable member of the HtrA-protease family has been identified only in some archaebacteria while the very small genomes of *Mycoplasma pneumoniae* and *Mycoplasma genitalium* do not appears to encode such a protease. However, a significant number of bacterial genomes encode more than one HtrA-like serine protease. *Mycobacterium tuberculosis* has four genes; *Escherichia coli, Bacillus subtilis, Treponema pallidum, Deinococcus radiodurans* and *Synechocystis* each have three copies while *Haemophilus influenzae* and *Pseudomonas aeruginosa* each have two copies. These observations prompt questions about the physiological roles of each paralogue and of the extent to which their functions overlap.

A comparison of the amino terminal and PDZ domain regions of paralogues from each bacterium suggests significant functional divergence. *E. coli* and *T. pallidum* each have two HtrA-like proteases with two PDZ domains and one HtrA-like protease with a single PDZ domain. Similarly *H. influenzae* and *P. aeruginosa* each have one HtrA-like protease with two PDZ domains and one HtrA-like protease with a single PDZ domain. Divergence within the amino terminal regions suggests that paralogues may be located within different cellular compartments. Analysis of the N-terminal region for the three paralogues in *T. pallidum* and *D. radiodurans*, suggests that one protease is extracellular, having a cleavable signal sequence; a second has a transmembrane domain suggesting that it is anchored within the cytoplasmic membrane and the third has neither a signal sequence nor a transmembrane domain and its location is predicted to be cytoplasmic. A clear functional distinction has been experimentally established in *E. coli* between DegS and the other two HtrA-like proteases, DegP and DegQ. Sigma factor σ^E directs expression of an extracytoplasmic

stress regulon that includes DegP. It has been shown that mutation of *degS* (encoding the HtrA-like protease with only one PDZ domain) stabilises the anti-sigma factor RseA thereby reducing the basal and induced levels of σ^E activity, whereas mutation of *degP* and *degQ* have no effect on the levels of σ^E activity (Mecsas *et al.*, 1993; Ades *et al.*, 1999). Therefore DegS has a distinct regulatory role in controlling induction of the σ^E regulon that its paralogues DegP and DegQ do not. In *P. aeruginosa* Boucher *et al.* have shown that mutation of *mucD* (encoding a HtrA-like protease) results in conversion to mucoidy with a concomitant increase in sensitivity to heat and hydrogen peroxide (Boucher *et al.*, 1996). Mutation of *algW,* encoding the second *P. aeruginosa* HtrA-like protease, also leads to increased sensitivity to heat and hydrogen peroxide, but with the additional phenotypes of paraquat sensitivity and increased alginate production in the presence of sublethal paraquat levels. These analyses show that individual HtrA-like proteases can have both distinct and overlapping regulatory and housekeeping functions in the cell.

There are three members of the HtrA-like serine protease family encoded in the *B. subtilis* genome, YkdA, YvtA and YycK (Kunst *et al.*, 1997). The three proteins each have a single transmembrane domain positioned in the amino terminus and a single PDZ domain in the carboxyl terminus. All three proteins are predicted to span the membrane with the amino terminus located within the cytoplasm and the carboxyl terminal catalytic and PDZ domains located extracytoplasmically (see section 3.2.3.2 and Figure 3.7). The three proteases, and their genes also differ in important respects. YkdA alone has a poly-serine tract, that is characteristic of some cell wall associated proteins and each protein has an amino terminal peptide of different length (YkdA, 44 aa), (YvtA, 71 aa), (YycK, 22 aa) protruding into the cytoplasm (Figure 3.7). Promoter analysis suggests a partially overlapping expression profile for *ykdA* and *yvtA*, distinctive from that observed for *yycK*. The *yycK* encoding operon is expressed during exponential growth from the operon promoter and is also expressed from a σ^G-type promoter positioned immediately upstream of *yycK* that is active during sporulation (Fabret and Hoch, 1998). These analyses suggest that the three HtrA-like proteases encoded in *B. subtilis* may have distinctive but partially overlapping expression profiles and functions within the cell.

In this chapter we report on the induction stimuli and expression patterns of *ykdA*, *yvtA* and *yycK*, the genes encoding the three HtrA-like serine proteases in *B. subtilis*. Expression of *ykdA* and *yvtA* is induced both by heat shock and by secretion stress using a common mechanism whereas *yycK* is neither heat shock nor secretion stress inducible. We also show that inactivation of either *ykdA* or *yvtA* results in compensating over-expression of the other gene with a concomitant increase in resistance to heat and hydrogen peroxide stresses whereas inactivation of both genes leads to thermosensitivity.

5.2. Results

5.2.1. Temporal expression of the genes encoding the HtrA-like serine proteases

The expression profiles of *ykdA*, *yvtA* and *yycK* were examined throughout the growth cycle in LB. These analyses were undertaken with strains DN2 and DN27 for *ykdA* (Chapter. 4). To analyse expression of *yvtA*, a *yvtA-bgaB* fusion construct, pDN110 was generated (see section 2.6., pg.61). Similarly a *yycK-bgaB* construct, pDN41 was used to measure expression of *yycK* (see section 2.6., pg.60). In a wild-type background, both *ykdA* and *yvtA* are expressed at a low level during exponential growth and their synthesis continues during stationary phase, with ß-galactosidase accumulating to between 20 and 30 units at 8 hours after the transition phase (Figure 5.1A). The temporal pattern of *ykdA* and *yvtA* expression was also established in their respective null mutant backgrounds. During exponential growth, expression of both genes is slightly increased over wild-type levels. However, the level of expression increases significantly when the culture enters the transition phase (at approx. 200 minutes) of the growth cycle, and continues to increase for up to 6 hours, the duration of the experiment (Figure 5.1A). This profile of expression is observed both for *ykdA* and *yvtA* with the expression levels being slightly higher for *ykdA* than for *yvtA*. We also examined expression of *yvtA* in a *ykdA* null background during the growth cycle and found that the expression profile was similar to that observed in the *yvtA* null background (Figure 5.1B).

To investigate the temporal regulation of *yycK* strain DN41 (Δ*yycK* [*yycK'-bgaB*]) containing a *yycK-bgaB* fusion at the *yycK* locus was grown in SM media. Figure 5.2A shows that after the exponential phase of the growth, expression of *yycK* falls slightly followed by a slow increase in expression during sporulation. To assess expression of *yycK* in rich media strain DN203 (see section 2.6., pg.63), containing the *yycK'-bgaB* fusion, was grown in LB broth and BgaB activity was measured throughout the growth cycle. The results, presented in Figure 5.2B, show that in rich medium expression of *yycK* is highest in the exponential phase and decreases thereafter without any increase in BgaB activity in the post-exponential phase of growth. These data are in agreement with the results of Fabret and

Hoch (Fabret and Hoch, 1998). These data show (i) that the *yycK* gene can be distinguished from *ykdA* and *yvtA* on the basis of temporal expression and (ii) that inactivation of either *ykdA* or *yvtA* leads to a dramatic increase in their respective expression levels at the transition phase of the growth cycle.

5.2.2. Heat shock induction of the genes encoding the HtrA-like serine proteases

The *ykdA* gene is induced by heat shock at 48°C with the level of induction being greatly increased in a *ykdA* mutant background (see Tables 4.3 and 5.1). To determine if *yvtA* and *yycK* were heat shock inducible, transcriptional fusions were generated using the *bgaB* reporter gene. In strain DN110, which has a *yvtA-bgaB* transcriptional fusion positioned at the *amyE* locus, the level of ß-galactosidase accumulation is very low (1U) in cells grown at 37°C. There is a significant increase in ß-galactosidase activity accumulation (11 units) after growth at 48°C for 1 hour (Table 5.1). These activity levels are approximately one-third those observed with the *ykdA-bgaB* fusion in the wild-type background at each temperature (Table 5.1). To investigate whether the YvtA protease affected expression of the *yvtA* gene, as was observed for the YkdA protease, expression of a *yvtA-bgaB* transcriptional fusion was examined in a *yvtA* mutant background (strain DN112). The results for DN112 (Table 5.1) show that there is a 12-fold increase in accumulated ß-galactosidase activity when cells are grown at 37°C and a 16-fold increase in activity levels during growth at 48°C. When strain DN41, containing a *yycK-bgaB* fusion generated by a Campbell-type integration, was grown at 37°C and 48°C, expression of the fusion was low and no induction was observed for up to 1 hour after temperature upshift as presented in Figure 5.3. Therefore expression of *yvtA* is both heat shock inducible and negatively autoregulated, similar to that observed for its paralogue *ykdA*. In contrast, expression of *yycK* is not thermoinducible under these conditions.

Since mutation of *ykdA* and *yvtA* leads to increased expression of the *ykdA-bgaB* and *yvtA-bgaB* transcriptional fusions respectively both at 37°C and at 48°C, we examined whether mutation of *ykdA* would affect expression of the *yvtA-bgaB* fusion and similarly whether

mutation of *yvtA* would influence expression of the *ykdA-bgaB* fusion. The results (Table 5.1) for strain DN114 (*yvtA::kan ykdA-bgaB*) show that mutation of *yvtA* leads to a two-fold increase in ß-galactosidase levels (4U to 9U) at 37°C and an approximate four-fold increase in ß-galactosidase levels (34U to 127U) at 48°C. These increased levels of *ykdA-bgaB* expression in the *yvtA* null background are lower than those observed in the *ykdA* null background. The results (Table 5.1) for strain DN113 (*ykdAΔ439 yvtA-bgaB*) show that mutation of *ykdA* leads to an increase in *yvtA-bgaB* expression: ß-galactosidase levels increase 17-fold (from 1U to 17U) at 37°C and 24-fold (from 11U to 268U) at 48°C. These levels of *yvtA-bgaB* expression in a *ykdA* null background are higher than those observed in a *yvtA* null background. Therefore the level of expression of *ykdA* affects the level of expression both of *ykdA* and of *yvtA* at 37°C and 48°C. Similarly, the level of expression of *yvtA* influences the expression of *yvtA* and *ykdA* at both temperatures. It is also apparent, at least with regard to thermoinduction, that deletion of *ykdA* has a greater effect on the expression of both genes at both temperatures than has deletion of *yvtA*.

To examine expression of *ykdA* in a double mutant background, strain DN115 was constructed with a deletion in the *ykdA* gene and a kanamycin resistance cassette inserted into the *yvtA* gene by a double cross-over event. Strain DN115 grew extremely slowly and colonies were small, round and had a mucoid appearance and consistency. This colony morphology was not stable and rapidly growing cells with a normal colony appearance segregated from the small mucoid colonies. These cells have presumably accumulated (a) suppressor(s), compensating to some extent for the growth defect, mucoidy or protease deficiency of *ykdA yvtA* double mutants. Further experiments were performed on the fast growing stable double mutant strains designated DN115*sup*. A *ykdA-bgaB* transcriptional fusion was inserted into the *amyE* locus of DN115*sup* to generate strain DN116 and expression levels are shown in Table 5.1. Expression of *ykdA* at 37°C has risen in this strain, with accumulated ß-galactosidase activity being 20-fold higher than those observed in wild-type cells and 6- to 7-fold higher than those seen in a *ykdA* background at this temperature. This trend is also observed at 48°C where ß-galactosidase activity levels are 40-fold higher than those observed in wild type cells, and almost 2-fold higher than those

observed in cells carrying the single *ykdA* mutant. Expression of *ykdA* was also examined in strain DN200. In this strain the *yvtA* gene has been inactivated as described for strain DN116 and the *ykdA* gene is intact but has the active site serine codon changed to a methionine codon rendering the encoded protein protease negative (see section 4.2.6.2). While colonies of strain DN200 had a mucoid appearance similar to DN116, this morphology was stable and no faster growing segregants were observed. The level of *ykdA-bgaB* expression in DN200 at 48°C was similar to that observed in the *ykdA* single mutant at this temperature. However the level of *ykdA* expression at 37°C in strain DN200 is 50-fold higher than wild-type, 17-fold higher than the single *ykdA* mutant and 2-3 times higher than the double mutant strain DN116. These data show that both *ykdA* and *yvtA* are heat shock inducible. Reciprocal cross-regulation is also evident in that expression of each gene is negatively regulated both by its own gene product and by the product of the other protease gene. Expression levels are further increased in strains where both *ykdA* and *yvtA* are inactivated.

5.2.3. Induction of *ykdA* and *yvtA* expression in response to heterologous amylase production

HtrA in *E. coli* degrades aberrant and misfolded proteins in the periplasm (Pallen and Wren, 1997). We sought to establish if production of heterologous proteins in *B. subtilis* could play a role in induction of *ykdA* and/or *yvtA* expression. The heterologous proteins chosen were AmyL, the amylase from *B. licheniformis* and AmyLQS50.5, a modified amylase that has an increased net positive charge (Stephenson *et al.*, 1998). Strains RC010 and RC011 were constructed in which expression of *amyL* (RC010) and *amyLQS* (RC011) are under the control of a xylose inducible promoter and *ykdA* expression was monitored through the *ykdA-lacZ* transcriptional fusion resident on each chromosome. The results with both amylases were essentially the same and are shown in Figure 5.4A. In the absence of xylose there is very little *ykdA-lacZ* expression with approximately 4U ß-galactosidase activity present during exponential growth and rising to approximately 10-14U activity during the stationary phase of the growth cycle. In the presence of inducer there is still a very low level of expression during the exponential growth phase. However a dramatic increase in *ykdA-*

lacZ expression occurs at the transition phase of the growth cycle with ß-galactosidase levels rising to 500-700 units of activity (Fig. 5.4A). Activity levels drop slightly thereafter to between 350 and 500 units for the remainder of the growth cycle. A similar experiment was performed to assess induction of *yvtA* by AmyL. Strain DN201 was constructed in which *amyL* is under the control of a xylose inducible promoter and *yvtA* expression is monitored by the resident *yvtA-bgaB* transcriptional fusion. It is important to note that strain DN201 is a null mutant for *yvtA* owing to the fact that the *yvtA-bgaB* construct in this strain derives from pDN40 which integrates into the *yvtA* gene. The expression profile (Fig 5.4B) shows that, in the absence of inducer (*i.e.*. no AmyL production), there is a low level of *yvtA* expression during exponential growth but that expression rises sharply at the transition phase of the growth cycle, accumulating to a maximum of approximately 300 units over 6 hours. In the presence of inducer, expression of *yvtA-lacZ* increases even more sharply at the transition phase of the growth cycle, rising to 350 units within 2 hours and continuing to increase to more than 600 units during the remainder of the experiment (Fig. 5.4B). We also assayed for the presence of amylase activity in the culture supernatant and showed it is present extracellularly during both the exponential growth and stationary phases of the growth cycle (Figure 5.5), confirming previously published results (Stephenson *et al.*, 1998). To confirm that the observed induction of *ykdA* and *yvtA* is caused by amylase production and is not simply an effect of adding the xylose inducer, expression of a *ykdA-lacZ* fusion was examined throughout the growth cycle in strain DN3 in the presence and absence of xylose. The results (Fig. 5.4C) show that there is a low and constant level of ß-galactosidase throughout the growth cycle in the absence of xylose with an increase in activity levels to approximately 15 units at the late exponential /transition phase in the presence of xylose. Therefore while xylose leads to a small increase in *ykdA* expression perhaps caused by induction of xylosidase (Lindner *et al.*, 1994), we conclude that the dramatic increase in both *ykdA* and *yvtA* expression at the transition stage of the growth cycle is caused by the presence of the heterologous amylases.

5.2.3.1. Investigation of the effect of amylase production on the expression of *yycK*.

To ascertain if *yycK* encoding the third HtrA-type protease in *B. subtilis* responds to secretion stress strain DN203 was constructed by transformation of strain KS408 with plasmid pDN41. Expression of the *yycK-bgaB* transcriptional fusion present in strain DN203 was examined during the growth cycle in the presence and absence of xylose in strain DN203. Figure 5.6 shows that xylose addition to the medium has no effect on growth profiles. In addition the pattern and level of BgaB activity are very similar in the two cultures. Therefore it is concluded that *yycK* expression is not induced by AmyL production.

5.2.4. Transcriptional analysis of *yvtA*.

The control region of *yvtA* contains a sequence showing high homology to the region of the *ykdA* promoter that directs heat shock induction (see Figure. 4.2). Therefore we wished to ascertain if this region also participated in *yvtA* heat shock induction. Primer extension analysis was performed on RNA samples isolated from cultures of strain DN111 grown at 37°C and 50°C. The results show a low and constant level of transcript during the course of the experiment at 37°C (Fig. 5.7A). The level of transcript was greatly increased at 50°C, evident within 6 minutes of temperature upshift and remaining at this high level for up to 24 minutes. The initiation point of transcription is indicated by an asterisk (Fig. 5.7A). It is located at the same position relative to the conserved octamer repeats (boxed, Fig. 5.7B) and putative –10 region (overlined PA-10, Fig. 5.7B) as is the transcription initiation point of *ykdA* (arrows in the promoter alignment, Figure 5.7B).

5.2.4.1. AmyL induction of both *ykdA* and *yvtA* is at the level of transcription and utilises the heat shock induced promoters.

To determine if AmyL induction of *ykdA* and *yvtA* occurs at the transcriptional level and through the same promoter as heat shock induction, primer extension analysis was performed on total RNA extracts from cells of strain KS408 (*amyE xylR-amyL*) grown in the presence and absence of xylose, the AmyL inducer. Figure 5.8B and 5.8C show a low level of transcript for *ykdA* and *yvtA* respectively in the absence of xylose. However, there

is a large increase in transcript levels for both *ykdA* and *yvtA* genes at the 60 and 90 minute time points. These transcripts initiate at the same start points as those upon heat shock induction (see Figures 4.9 and 5.7). Therefore it is evident that induction of both genes by AmyL occurs at the transcriptional level and is mediated by the same promoter as heat shock induction.

5.2.5. Phenotypic analysis of *ykdA* and *yvtA* mutants.

A strain carrying a null mutation in *ykdA* has a growth profile similar to wild-type cells at both 37°C and 48°C. However when these cells are shifted to 54°C (the killing temperature) the *ykdA* mutant cells show a 10-fold increase in survival levels compared to wild-type cells (see Figure. 4.10). To further investigate these phenomena, we decided to examine the growth and survival profiles of strains with single and double mutations in *ykdA* and *yvtA* after exposure to heat and oxidising agents. The growth profiles at 37°C and 48°C of strains carrying mutations in either *ykdA* or *yvtA* are similar to those of wild-type cells grown at these temperatures (Figure 5.9). A comparison of the growth profiles of strains DN115*sup* and DN200 with that of the wild-type strain grown at 37°C and at the sub-lethal 52°C are shown (Figure. 5.10A). At 37°C the growth profile of strain DN115*sup* (mutated in both *ykdA* and *yvtA* but carrying an unspecified suppressor mutation), was essentially identical to the wild-type profile. However, growth of strain DN200 is significantly slower than wild-type cells at this temperature. When exponentially growing cultures of these strains are shifted from 37°C to 52°C, it is evident that the wild-type strain continues to grow, albeit at a slower rate. However when strains DN115*sup* and DN200 are shifted to 52°C, cell growth ceases immediately and cells slowly lyse during the remainder of the experiment. These experiments show that having a proteolytically functional copy of either *ykdA* or *yvtA* allows cells to grow essentially with wild-type characteristics at elevated temperature. A temperature sensitive growth phenotype is observed only when the proteolytic activites of both YkdA and YvtA are mutated. Furthermore, it also implies that any putative chaperone activity present in strain DN200 is unable to compensate for the proteolytic role at high temperatures.

Quantitation of thermosensitivity was scored by exposing exponentially growing cultures of each strain to 54°C (lethal temperature) for increasing time periods. The results, (shown are the average of three separate experiments) confirm that strains mutated in either *ykdA* or *yvtA* are more thermotolerant than are wild-type cells (Figure. 5.10B). It is also evident that DN115*sup* cells are up to 10-fold more sensitive than wild-type cells after exposure to 54°C for more than 60 minutes. Surprisingly, we found that the thermosensitivity of DN200 cultures was similar to wild-type for up to 2 hours exposure at 54°C. Since the *ykdA* expression level is greatly increased in this background (Table 5.1.), this result suggests that extreme overexpression of an intact YkdA protein devoid of protease activity may confer some thermoprotection on cells of this strain.

Previous results showed that mutation of *ykdA* also conferred resistance to lethal levels of hydrogen peroxide (see Figure. 4.10). To further investigate this phenomenon, we examined the resistance of strains both singly and doubly mutated in *ykdA* and *yvtA* to 10mM hydrogen peroxide for increasing time periods. The results (Fig. 5.11) show that strains singly mutated either in *ykdA* or *yvtA* are up to 100-fold more resistant to 10mM hydrogen peroxide than is the wild-type strain 168. Interestingly, strain DN115*sup* shows the same sensitivity as the wild-type strain. Strain DN200, while being slightly more sensitive to hydrogen peroxide than singly mutated strains, is still up to 100-fold more resistant than either wild-type or strain DN115*sup*. These data suggest that overproduction of either YkdA, YvtA or a protease negative form of YkdA confers some protection against lethal hydrogen peroxide levels during exposure periods greater than 10 minutes.

5.3. Discussion

In this study, we have investigated the relationship between the three HtrA-like serine proteases, YkdA, YvtA and YycK, encoded in the *B. subtilis* genome. Phylogenetic analysis shows that YkdA and YvtA are similar to each other with YycK being more distantly related (Figure 4.12). The results presented here support this classification. Expression profiles of the *ykdA* and *yvtA* genes are very similar. Both are inducible by heat shock during exponential growth and by heterologous amylases at the transition phase of the growth cycle. In cells with either *ykdA* or *yvtA* mutated, expression of the intact paralogue increases slightly during exponential growth, but rises dramatically at the transition phase and continues to increase throughout an extended stationary phase. The phenotypes observed in strains carrying null mutations in either *ykdA* or *yvtA* are also very similar showing increased tolerance to thermal stress and to lethal levels of hydrogen peroxide. The expression profile of *yycK* is quite different from that of *ykdA* or *yvtA* (Fabret and Hoch, 1998), our results). Expression is low during exponential growth, decreases at the transition phase of the growth cycle and is not inducible by heat shock. In addition, expression of *yycK* is directed by a σ^G-type promoter positioned immediately upstream of the gene under conditions conducive to sporulation. The negative auto- and reciprocal cross-regulation of *ykdA* and *yvtA* expression, and the similar phenotypes observed when either gene is mutated, suggests an overlap in the cellular roles of YkdA and YvtA. YycK in contrast has either a distinctive cellular role or alternatively functions similarly to YkdA / YvtA but under different environmental or developmental conditions.

There is a high level of identity between the promoters utilised for expression of *ykdA* and *yvtA* under the conditions examined in this study. A comparison of these promoters (Fig. 3B) shows that they have a consensus σ^A-type –10 region with a series of similarly spaced octamer repeats positioned at the –35 region. This suggests that both genes are under the control of the same regulator and that their expression is responsive to the same stimuli. Therefore it is likely that under some conditions YkdA and YvtA are both required to adequately deal with the cellular stress or condition that triggers their induction. The negative auto- and reciprocal cross-regulation observed in strains carrying either *ykdA* and

yvtA null mutants is likely to be a manifestation of compensatory overexpression of YkdA and YvtA. Therefore establishing the conditions under which such over-expression occurs gives an indication of the time and conditions when the inducing signal(s) are most prevalent. Two such conditions have been established in this study: (i) in single mutant backgrounds, even under normal growth conditions, expression of both genes rises dramatically at the transition phase and continues throughout stationary phase and (ii) production of heterologous amylases leads to an additional increase in *ykdA* and *yvtA* expression at the transition phase over that observed during normal growth conditions. Interestingly, both events occur at the transition and stationary phases of the growth cycle. In *B. subtilis*, the transition phase coincides with the production and secretion of many extracellular enzymes. In view of these results and of those reported by Poquet *et al.* (Poquet *et al.*, 2000) showing that HtrA in *L. lactis* functions both to degrade abnormal proteins and to process natural propeptides, it is tempting to speculate that YkdA and YvtA may have similar functions in *B. subtilis*. Perhaps the gene duplication was triggered by the level and/or number of enzymes produced and secreted during stationary phase by this bacterium.

Further insight into the mechanism of *ykdA* and *yvtA* induction can be obtained from our expression data in response to heterologous amylase production. We, and others (Stephenson *et al.*, 1998) have demonstrated that heterologous amylases are produced and secreted during the exponential and stationary phases of the growth cycle in the presence of the xylose inducer. However, *ykdA* and *yvtA* induction by such amylases was only triggered at the transition phase, continuing throughout the stationary phase of the growth cycle. Therefore despite amylase production and secretion during exponential growth, *ykdA* and *yvtA* expression was not significantly induced. These data suggest it is neither the heterologous nature of the amylases nor their secretion *per se* that is sufficient for *ykdA* and *yvtA* induction. Perhaps the induction stimulus is multifactorial, comprising (i) secretion load (*i.e.*. the total number of proteins and the amount of each protein being processed and/or secreted); (ii) the level of protein maturation and (iii) the level of aberrant protein formation. Further experiments will be required to dissect the nature of the inducing signal.

Redundancy in the function of YkdA and YvtA is also suggested by analysis of mutant phenotypes. In contrast to other bacteria, where mutation of *htrA* leads to thermosensitivity *e.g. htrA* of *E. coli*, *algW* and *mucD* from *P. aeruginosa* and *htrA* from *Yersinia enterocolitica* (Pallen and Wren, 1997), mutation of either *ykdA* or *yvtA* results in strains that are more resistant to both heat and hydrogen peroxide. It is necessary to mutate both genes in order to observe thermosensitivity. Strains carrying mutations in both genes are slow growing and form mucoid colonies but segregate fast-growing colonies very rapidly. These cells grow normally at 37°C, but still cannot grow at elevated temperatures showing that the suppressor mutation affects growth but does not compensate for the loss of both YkdA and YvtA. Such suppressor mutations do not accumulate in a strain carrying a null mutation in *yvtA* and a *ykdA* allele encoding a protease negative form of the YkdA protein (strain DN200). This strain has a severe growth defect at lower temperatures and is thermosensitive at elevated temperatures. Spiess *et al.*, has shown that HtrA from *E. coli* has a chaperone activity that predominates at lower temperatures and a protease activity that predominates at elevated temperatures (Spiess *et al.*, 1999). The difference in the phenotypes of strains DN115 and DN200 may suggest that the protease negative form of YkdA has such a chaperone activity that is sufficient to promote slow growth but that the protease activity is required for growth at higher temperatures. However, it is evident that the stimulus for *ykdA* induction is still extremely high in both these backgrounds with expression levels at 37°C being 20-40 fold higher than those observed in the wild-type background.

In conclusion our results show that two of the HtrA-like serine proteases in *B. subtilis* YkdA and YvtA, show significant similarities in terms of expression profiles and of the phenotypes of mutant strains, suggesting that they perform similar functions within the cell.

Table 5.1. Expression of transcriptional fusions between the *ykdA* and *yvtA* promoters and the *bgaB* reporter gene in various genetic backgrounds at 37°C and 48°C.

Construct expressed	Background (strain)	BgaB sp act at [a]:	
		37°C	48°C
P*ykdA-bgaB*	168 WT [b] (DN2)	4	34
	*ykdA*Δ439 (DN27)	12	723
	yvtA::*kan* (DN114)	9	127
	*ykdA*Δ439 *yvtA*::*kan sup* (DN116)	80	1200
	ykdA (N289H S290M) *yvtA*::*kan* (DN200)	215	880
P*yvtA-bgaB*	168 WT (DN110)	1	11
	*ykdA*Δ439 (DN113)	17	268
	yvtA::*kan* (DN112)	12	184

[a] All values presented are those found 60 min after the culture was split. One unit is defined as 1nmol of *o*-nitrophenyl-β-D-galactopyranoside hydrolysed per min per mg of protein.

[b] WT, wild type

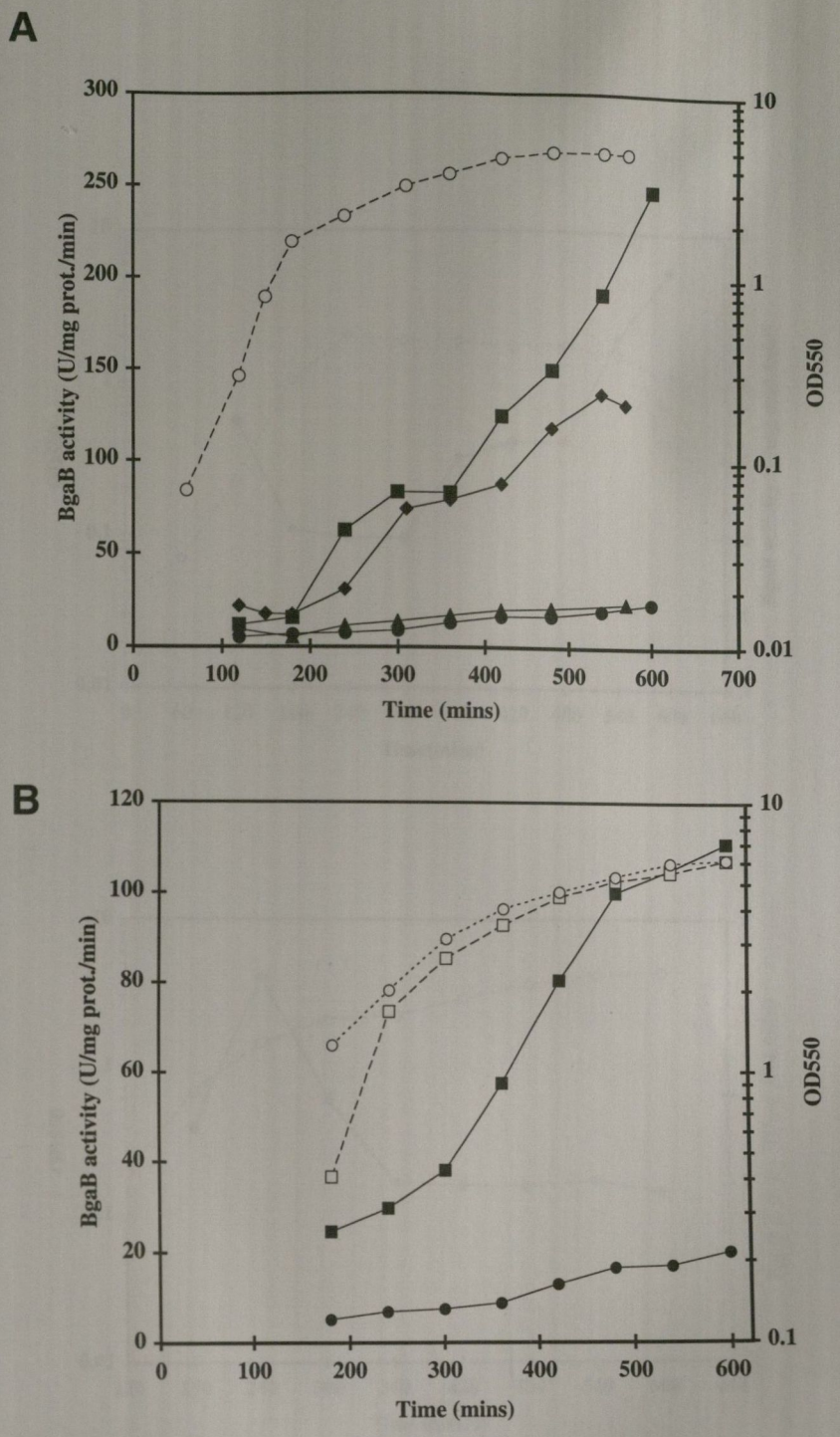
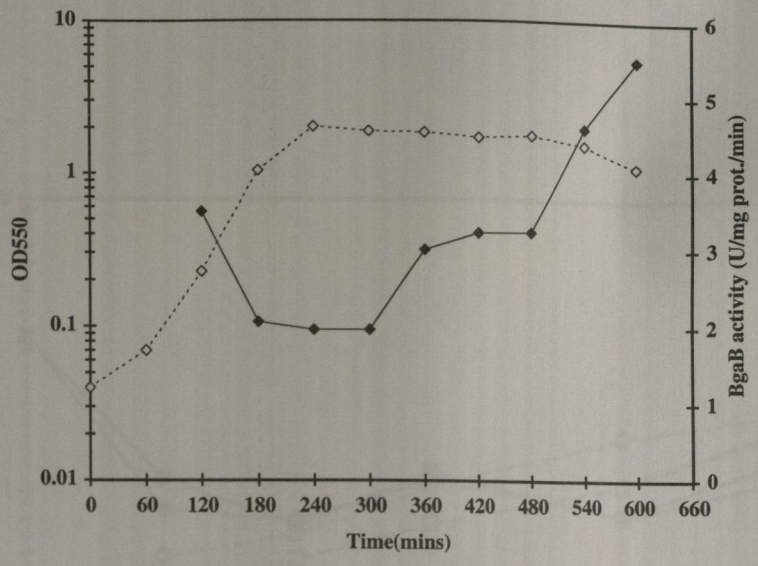


Figure 5.1. Profiles of growth and ß-galactosidase (BgaB) accumulation throughout the growth cycle in LB. Growth was measured at OD550 and is indicated by open circles. The single growth curve shown is representative for all four strains. (A) Beta-galactosidase accumulation is indicated by closed symbols for the following strains: strain DN2 (*amyE* ::P*ykdA -bgaB*)●; DN110 (*amyE* ::P*yvtA -bgaB*) ▲; DN27 (*ykdA* Δ439 *amyE* ::P*ykdA -bgaB*) ■; DN112 (*yvtA* ::kan *amyE* ::*PyvtA -bgaB*) ◆.
(B) Growth curve (□) and accumulation of BgaB (■) in strain DN113 (*ykdA* Δ439 *amyE* ::P*yvtA -bgaB*); the same data from growth of strain DN110 (*amyE* ::P*yvtA -bgaB*) are included for comparison purposes, indicated by open (O) and closed circles (●).

A



B

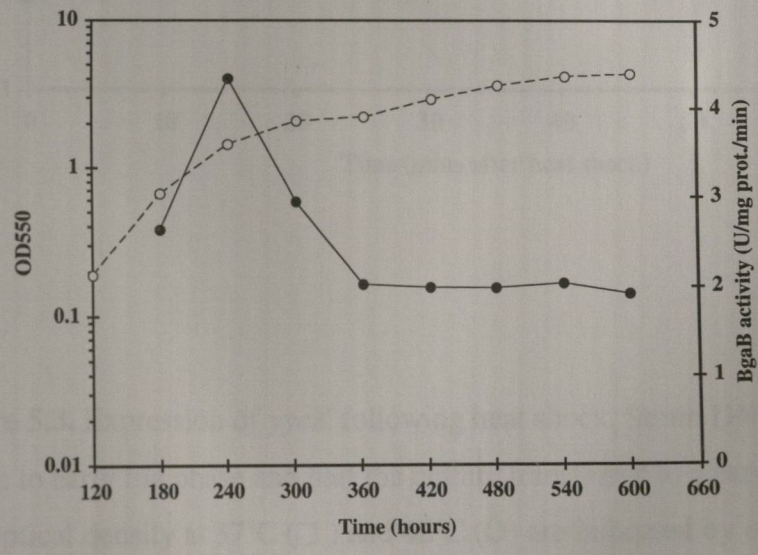


Figure 5.2. Temporal expression of *yycK* . (A) Expression of *yycK* in SM broth. Strain DN41 (*yycK* ::pDN41[*yycK' -bgaB*]) was grown in SM broth, growth (◊) and BgaB activity (◆) are shown.

(B) Expression of *yycK* in LB broth. Strain DN203 (*amyE xylR -amyL yycK* ::pDN41 [*yycK '- bgaB*]) was grown in LB broth, growth (O) and BgaB (●) activity are indicated.

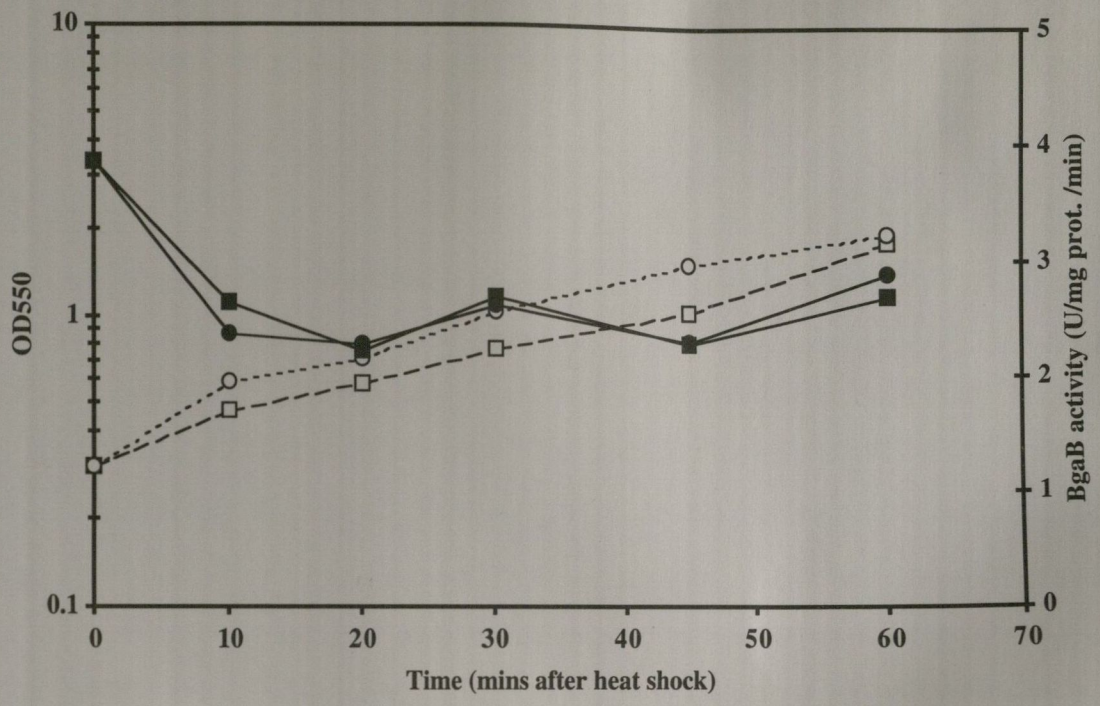
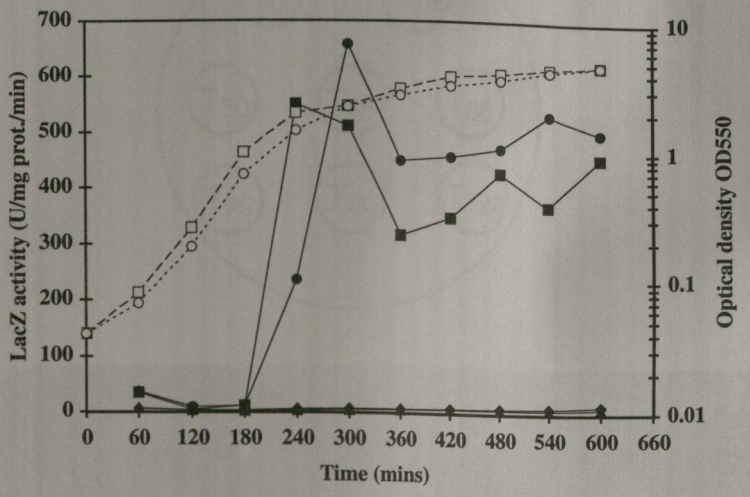


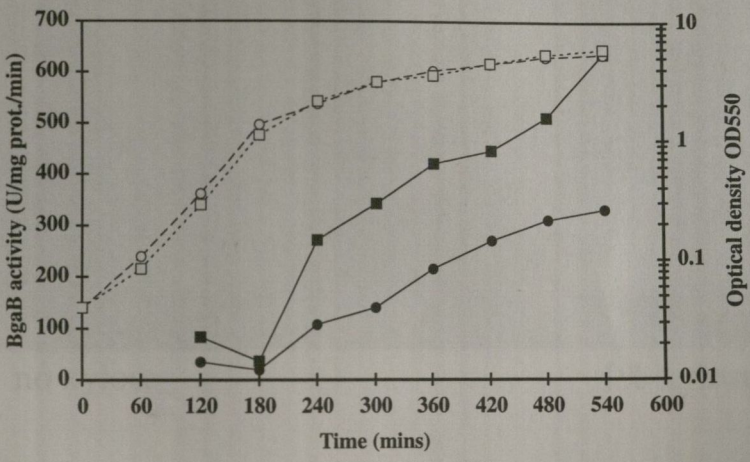
Figure 5.3. Expression of *yycK* following heat shock. Strain DN41 (*yycK-bgaB*) was grown to early log phase and half the culture transferred to a prewarmed flask at 48°C. The optical density at 37°C (□) and 48°C (O) are indicated by open symbols. BgaB activity measurements from the 37°C (■) and 48°C (●) cultures are shown by closed symbols.

Figure. 5.4. Profiles of growth and ß-galactosidase accumulation in the presence and absence of heterologous amylase production during the growth cycle in LB. Growth is represented by open symbols and ß-galactosidase accumulation by closed symbols. (A) Growth and LacZ accumulation for: strain RC010 (*xylR*::P*xyl-amyL* P_{spac}-*ykdA* P_{ykdA}-*lacZ*) (O, ●); RC011 (*xylR*::P*xyl-amyLQS* P_{spac}-*ykdA* P_{ykdA}-*lacZ*) (□, ■) both strains grown in the presence of 1% xylose and 1mM IPTG; strains RC010 (◆) and RC011 (▲) respectively, both grown in the absence of xylose with 1mM IPTG. (B) Growth and BgaB accumulation for strain DN201 (*xylR*::P*xyl-amyL yvtA::*pDN40 [*yvtA'-bgaB*]) in the absence (●) and presence (■) of 1% xylose inducer. (C) Growth and LacZ accumulation for strain DN3 (P_{spac}-*ykdA* P_{ykdA}-*lacZ*) in the absence (●) and presence (■) of 1% xylose inducer and 1mM IPTG respectively.

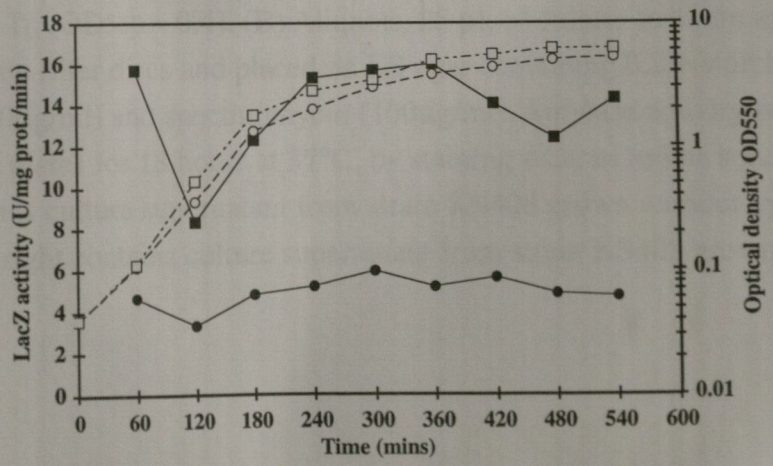
A


LacZ activity (U/mg prot./min)
Optical density OD550
Time (mins)
0
100
200
300
400
500
600
700
0.01
0.1
1
10
0
60
120
180
240
300
360
420
480
540
600
660

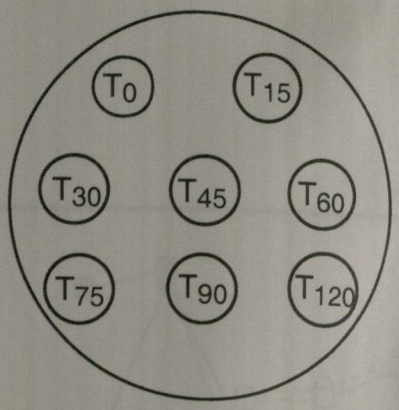
B


BgaB activity (U/mg prot./min)
Optical density OD550
Time (mins)
0
100
200
300
400
500
600
700
0.01
0.1
1
10
0
60
120
180
240
300
360
420
480
540
600

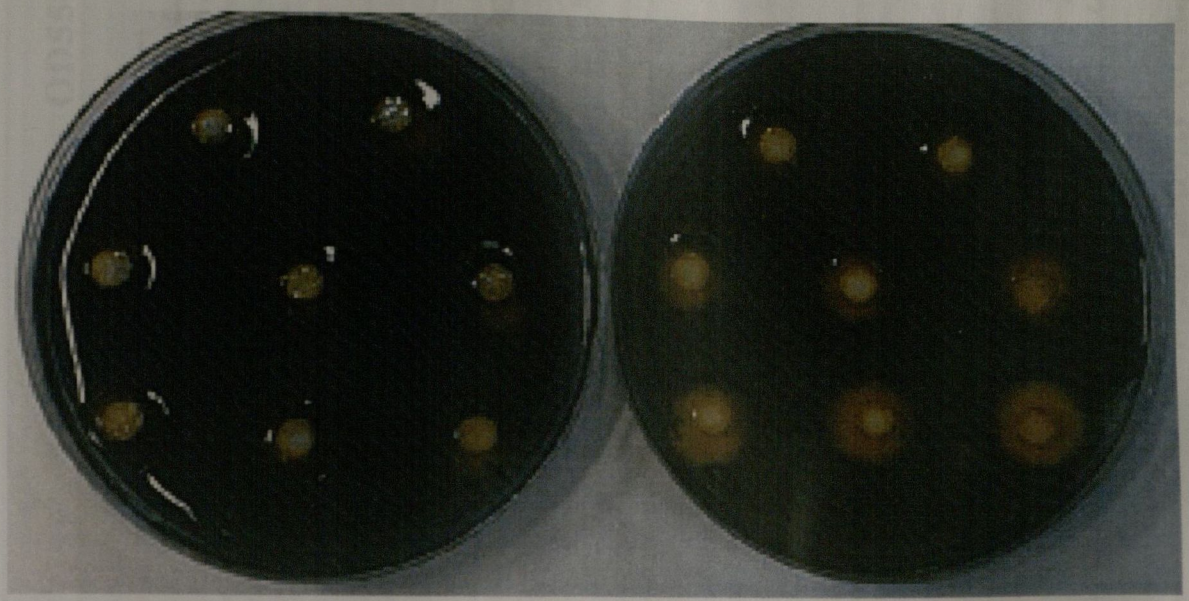
C


LacZ activity (U/mg prot./min)
Optical density OD550
Time (mins)
0
2
4
6
8
10
12
14
16
18
0.01
0.1
1
10
0
60
120
180
240
300
360
420
480
540
600

A



B



no xylose 1% xylose

Figure 5.5. Expression and secretion of AmyL in strain KS408 (*amyE xylR*::P*xyl-amyL*) in response to the presence or absence of xylose inducer. (A) Key of sampling time points in minutes after T_0 (OD550 ≈ 0.4). (B) Aliquots, 25 µl, of culture medium applied to sterile blank antibiotic filter discs and placed on LB agar containing 0.2% starch and both kanamycin(10µg/ml) and spectinomycin (100µg/ml). Amylase activity was recorded after incubation of plates for 18 hours at 37°C, by staining with an iodine solution. The plate on the left contains culture supernatant from strain KS408 grown without xylose, while the plates on the right contains culture supernatant from strain KS408 grown in the presence of xylose.

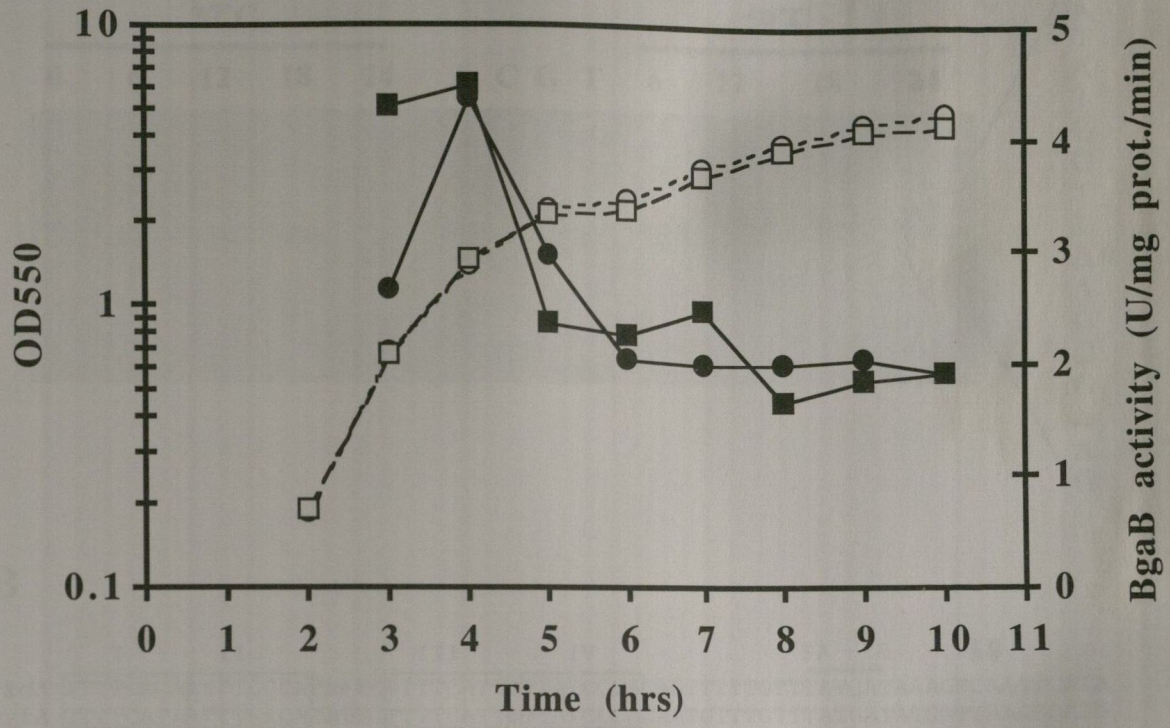


Figure 5.6. Expression of a *yycK'-bgaB* transcriptional fusion in response to heterologous amylase (AmyL) production. Strain DN203 (*amyE xylR* ::P*xyl-amyL yycK*::pDN41 [*yycK'-bgaB*]) was grown in LB broth in the presence and absence of 1% xylose inducer. Open symbols show growth (OD_{550}) in the presence (□) or absence (O) of xylose. Closed symbols show BgaB specific activity in the presence (■) or absence (●) of xylose.

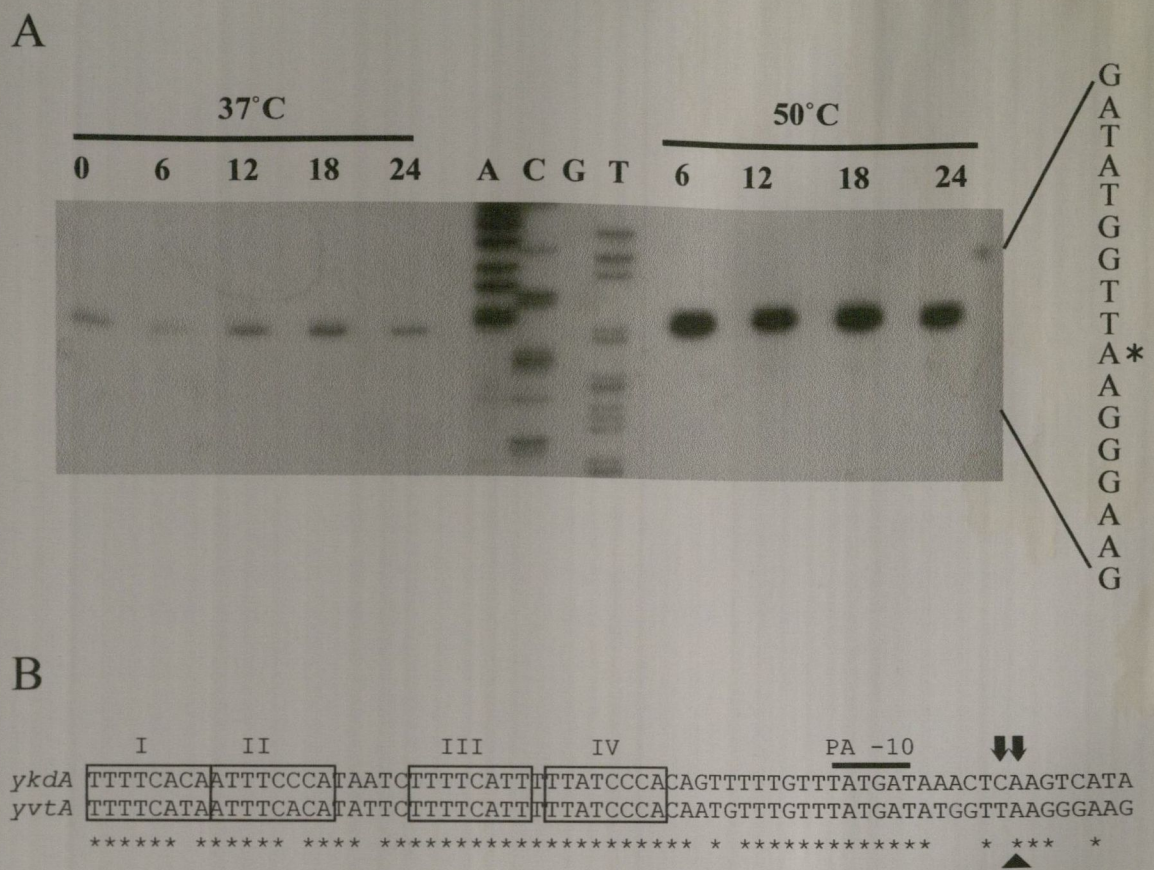
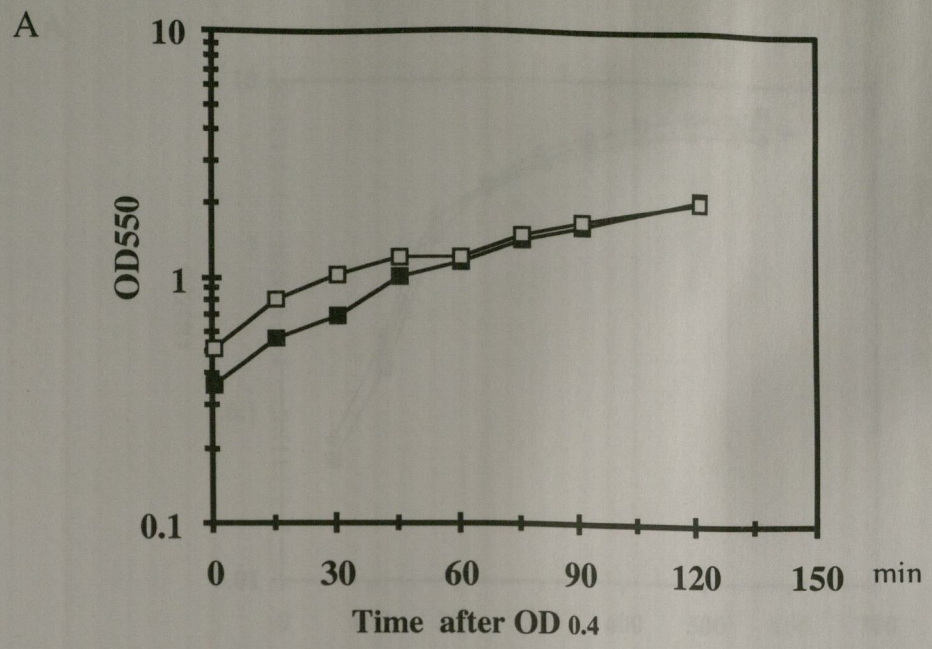
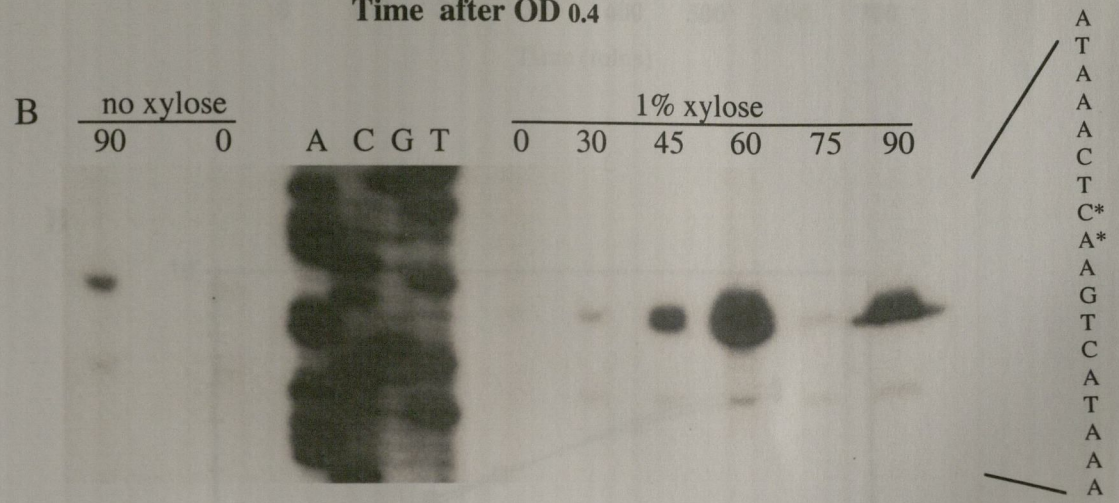
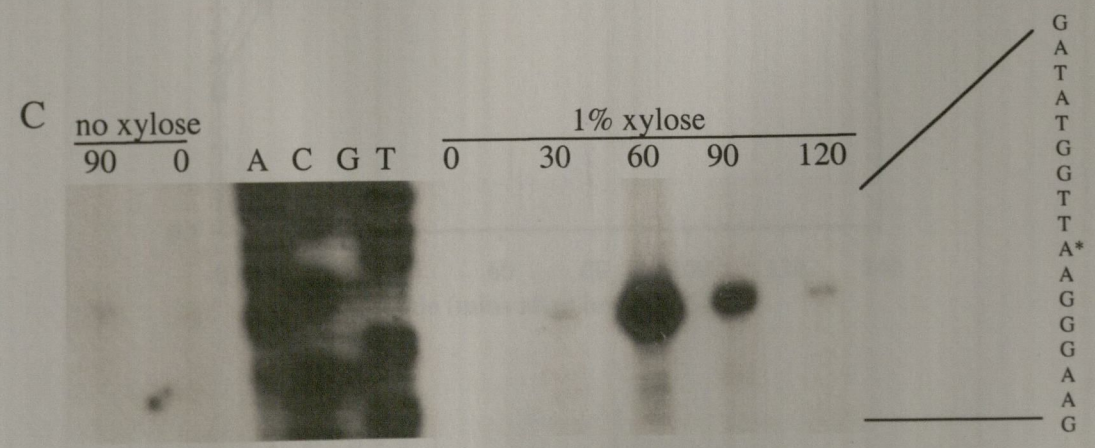


Figure 5.7. Primer extension analysis of *yvtA*. (A) RNA samples were prepared from an exponentially growing culture (OD550 = 0.3) of strain DN111(*yvtA*::kan) at 37°C and at varying times after temperature upshift to 50°C. Time is indicated in minutes, with 0 representing the sample taken before temperature upshift, and 6, 12, 18 and 24 being the time in minutes after upshift. The signal is that obtained from 12.5µg of total RNA. A sequencing ladder (A C G T) obtained with the same primer is shown and the complementary sequence is indicated on the right hand side. The base at which transcription initiation begins is indicated by an asterisk. (B) An alignment of the conserved regions of the *ykdA* and *yvtA* promoters. Nucleotide identity is indicated by an asterisk; the conserved octamer repeats (numbered I to IV) are boxed; the SigA-type –10 region is indicated by PA-10 and the initiation points of transcription are indicated by arrows for *ykdA* and an arrowhead for *yvtA*.

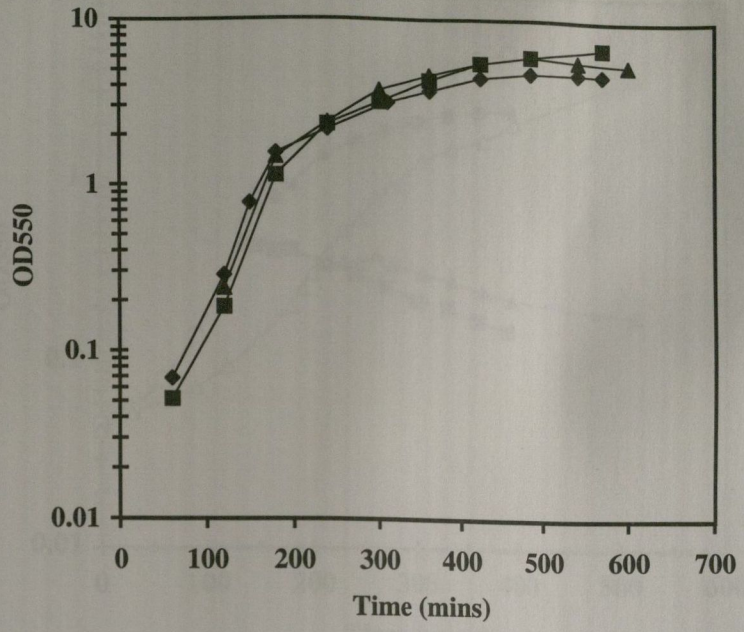
Figure 5.8. Primer extension analysis of *ykdA* and *yvtA* induction in response to production of the heterologous protein AmyL. (A) Cells of strain KS408 (*amyE xylR*::P*xyl-amyL*) were grown in the presence (■) and absence (□) of 1% xylose, the inducer of AmyL production. Samples were harvested from mid-exponentially growing cultures beginning at OD_{550} ≈0.4 (0) and at subsequent time points, indicated above each lane, in the presence or absence of xylose. (B) Primer extension of *ykdA* transcript using 12.5μg total RNA isolated from cells at the times indicated: A C G T indicate the lanes of the sequencing ladder, generated with the same primer as used for primer extension analysis, the complementary sequence is show to the right of the autoradiogram with an asterisk on the bases at which transcription initiates. The autoradiograph was exposed for 18h. (C) Primer extension of *yvtA* transcript using 12.5μg total RNA isolated from cells grown in the absence or presence of xylose at the times indicated. The sequencing ladder generated with the same primer as used for primer extension analysis and initiation points of transcription are indicated as outlined for (B). The autoradiogram was exposed for 18h.


A
10
OD550
1
0.1
0
30
60
90
120
150
min
Time after OD 0.4


B
no xylose
90
0
A C G T
1% xylose
0
30
45
60
75
90
A T A A A C T C* A* A G T C A T A A A


C
no xylose
90
0
A C G T
1% xylose
0
30
60
90
120
G A T A T G G T T A* A G G G A A G

A



B

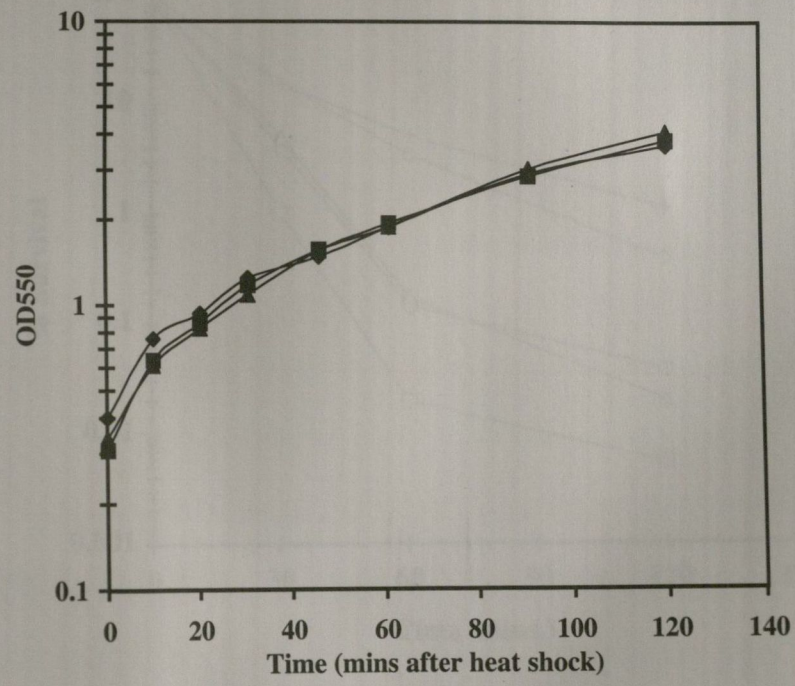


Figure 5.9. Growth of strains DN2 (WT), DN27 (*ykdA* Δ439) and DN112 (*yvtA* ::kan) at 37°C and 48°C.
(A) Strains DN2 (■), DN27 (▲) and DN112 (◆) were grown LB broth at 37°C and growth monitored at OD550.
(B) Growth of strains DN2, DN27 and DN112 after heat shock to 48°C in LB broth, same symbols apply to strains as in (A)

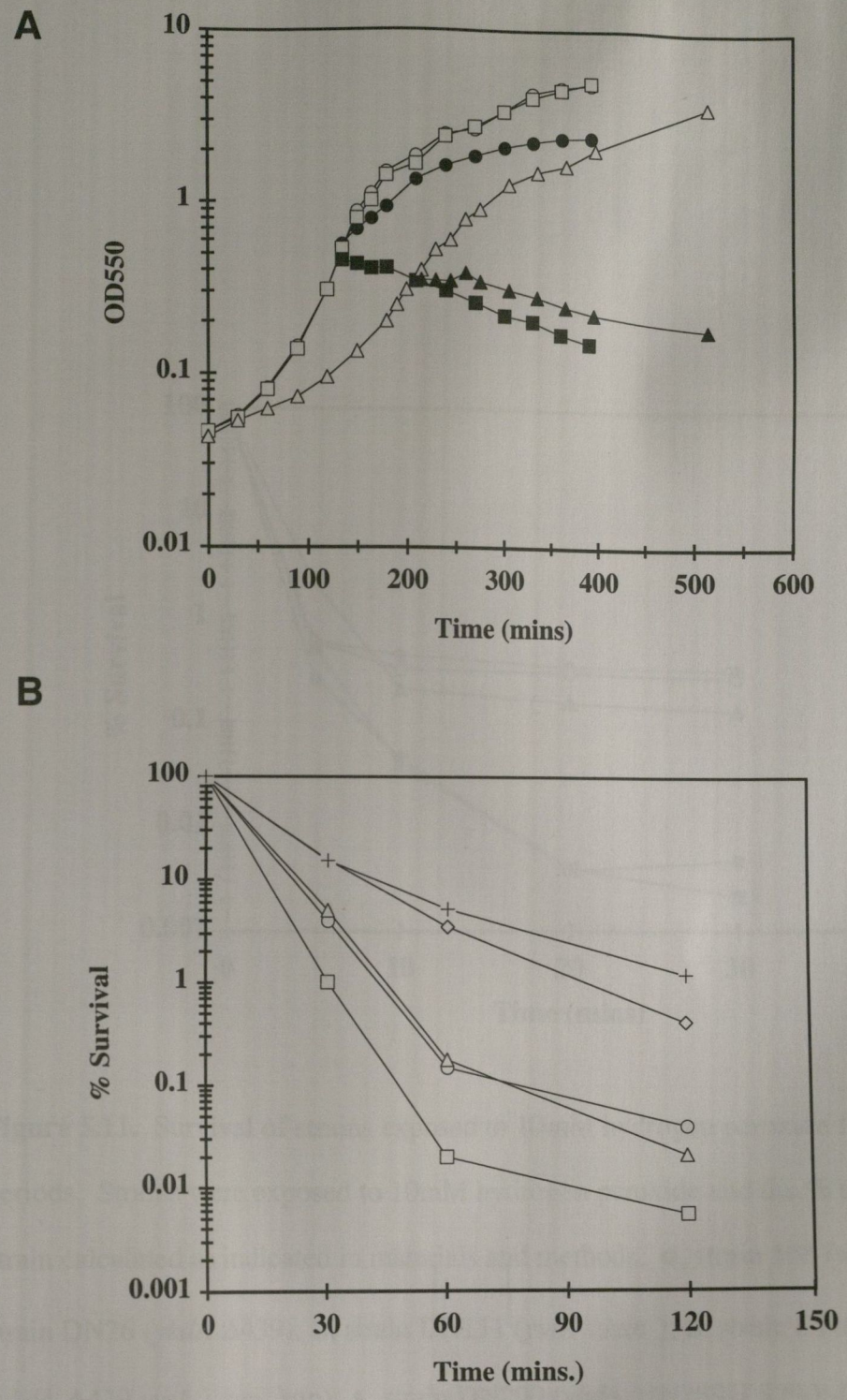


Figure 5.10. Growth and survival profiles of strains grown at elevated temperatures. (A) Growth profiles of strains grown at 37°C (open symbols) and 52°C (closed symbols). Growth was monitored by measuring optical density at 550 nm; o ●, strain 168 (wild-type); □ ■, strain DN115*sup* (*ykdA* Δ439 *yvtA* ::*kan sup*); Δ ▲, DN200 (*ykdA* 1[N289H S290M] *yvtA* ::*kan amyE* ::P*ykdA -bgaB*). (B) Survival of strains after exposure to heat stress at 54°C for varying time periods: circles, strain 168 (wild-type); diamonds, strain DN26 (*ykdA* Δ439); crosses, strain DN111 (*yvtA* ::*kan*); squares, strain DN115*sup* (*ykdA* Δ439 *yvtA* ::*kan sup*); triangles, strain DN200 (*ykdA* 1[N289H S290M] *yvtA* ::*kan amyE* ::P*ykdA -bgaB*).

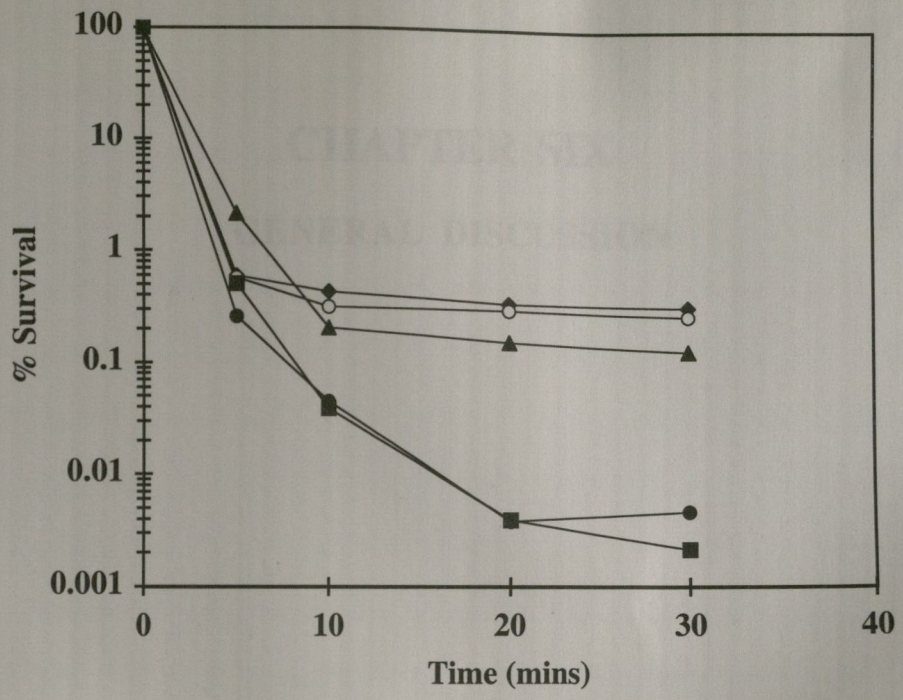


Figure 5.11. Survival of strains exposed to 10mM hydrogen peroxide for varying time periods. Strains were exposed to 10mM hydrogen peroxide and the % survival for each strain calculated as indicated in materials and methods. ●, strain 168 (wild-type); ◆, strain DN26 (*ykdA* Δ439); O, strain DN111 (*yvtA* ::*kan*); ■, strain DN115sup (*ykdA* Δ439 *yvtA* ::*kan* sup); ▲, strain DN200 (*ykdA* 1[N289H S290M] *yvtA* ::*kan* *amyE* ::P*ykdA* -*bgaB*).

CHAPTER SIX

GENERAL DISCUSSION

6.1. Summary of Results.

Sequencing of an approximately 15-kbp region of the *B. subtilis* chromosome identified an ORF, *ykdA*, encoding a putative protein of 449 amino acid which showed homology to DegP(HtrA) from *E. coli*. Analysis of the complete genome sequence of *B. subtilis* indicated that two other proteins, YvtA and YycK(YyxA), were also members of the HtrA family of serine proteases. All three proteins contain a central region with extensive homology to the conserved proteolytic domain of serine type proteases, a feature of all HtrA proteases (Barrett *et al.*, 1998). The three *B. subtilis* proteins also possess a protein-protein interacting or substrate recognition C-terminal PDZ domain. Unique to both *ykdA* and *yvtA* is a highly conserved upstream control sequence of 56-bp. In keeping with the occurrence of this control sequence it has been shown that both genes are induced, at the transcriptional level, by stimuli resulting from heat shock and secretion stress through overexpression of a secretable heterologous protein, AmyL. Furthermore, in strains mutated for either *ykdA* or *yvtA* the corresponding basal level of expression of the remaining gene is increased while the heat stress expression is hyperinduced. This increased expression is also evident from promoter fusions of the deleted gene and overall displays a system of auto- and cross-regulation. This mode of expression has not yet been shown for HtrA-like proteases from other organisms. The *yycK* gene, encoding the third HtrA-like protease and lacking any upstream motif similar to that associated with *ykdA* or *yvtA*, is not inducible by either heat or secretion stress stimuli. *yycK* expression is temporally regulated both by a σ^G dependent promoter located immediately upstream of the gene and from the promoter of the heptacistronic operon in which it is contained (Fabret and Hoch, 1998; Fukuchi *et al.*, 2000).

Strains mutated in either *ykdA* or *yvtA* display greatly increased resistance to hydrogen peroxide and a lesser but reproducible increase in resistance to lethal heat stress temperatures (Figures 4.10, 5.9 and 5.10). This phenotype is also evident in a strain mutated for *yvtA* and expressing a mutated YkdA protein putatively lacking protease activity (Figure 5.10). These results imply that both YkdA and YvtA have roles additional to their proteolytic functions possibly involving chaperone-like activity perhaps similar to DegP from *E. coli* (see below).

6.2. Structure and localisation of *B. subtilis* HtrA-like proteases.

The three HtrA-like proteases of *B. subtilis* all contain the same domain organisation, an N-terminal localisation domain followed by the proteolytic core domain and lastly the C-terminal PDZ substrate recognition and binding domain (Figure 3.7). The N termini preceding the transmembrane helix is the least conserved of the three domains among the proteins, an observation also applicable to HtrA proteases in general (Pallen and Wren, 1997). The N terminus of each of the *B. subtilis* proteins is predicted to be localised to the cytoplasm. This region of the protein is followed by a transmembrane motif in all three proteins (Fig. 3.7). It is noteworthy that no apparent signal peptide cleavage sequence is present, implying a membrane anchored localisation with the proteolytic domain on the *trans* side of the cytoplasmic membrane. A recent analysis of HtrA-like proteins in Gram-positive organisms showed that these HtrA-like proteases contain a single transmembrane motif and also lack any recognisable signal peptide (Jones *et al.*, 2001). Recently a comprehensive review of the predicted *B. subtilis* extracellular proteome was undertaken with emphasis on the identification of those proteins containing recognisable secretion signals (Tjalsma *et al.*, 2000). These authors compiled a listing of known and potential extracellular localised proteins. In addition, proteins with transmembrane motifs were also listed, among them the HtrA-like proteases of *B. subtilis*. Following on from this analysis the same authors in a collaborative study with the laboratory of Michael Hecker subsequently produced experimental data on the composition of the extracellular proteome of *B. subtilis*. This study confirmed much of the predicted analysis and also identified extracellular proteins devoid of any typical secretion signal (Antelmann *et al.*, 2001). In ongoing and recent work the presence of YkdA in the extracellular medium of *B. subtilis* cultures has been shown both in 2D-gel analysis (H. Antlemann, *et al.*, 2003) and by western blot analysis of extracellular fractions (D. Noone and C. Hough, unpublished data). The mechanism involved in release of YkdA to the medium is not known but might occur by a process know as "shaving" involving proteolysis (see Figure 8C Antelmann *et al.*, 2001). The medium associated YkdA protein was excised from the 2D-gel and subjected to MALDI-TOF analysis. This showed that the protein had been processed at residue 96, in the first serine rich region. Whether this indicates that YkdA functions in the medium, cell wall or remote from the cytoplasmic

membrane is still unknown. The data showing that *ykdA* is induced by secretion stress (see section 5.2.3 and Figures 5.4 & 5.7) and corroborated by a recently published report from Hyyrylainen *et al.*, (2001) lends emphasis to the fact that much of the stress from overproduction of secretable heterogeneous proteins is generated at the cell membrane/wall interface. In studies involving overexpression of AmyQ considerable amounts of the unprocessed preprotein were found to be cell associated (Hyyrylainen *et al.*, 2001; Vitikainen *et al.*, 2001). This may indicate that the primary role of YkdA is undertaken at this location and that release into the medium may be the result of this catalytic action. Although no data is available yet it is assumed from the similar positioning of the transmembrane domains of YvtA and YycK that these proteins are also membrane anchored.

Recently the crystal structure of the *E. coli* DegP protein was determined (Krojer *et al.*, 2002). DegP forms a hexamer composed of two tight trimeric subunits . The PDZ domains form the mobile side walls of the DegP proteolytic chamber. In one of two different crystal structures formed during this study, "molecule a" was shown to have a very similar structure to "molecule b" even without the apparent presence of the second PDZ domain. Given the conserved nature of secondary structure of the PDZ domains from YkdA, YvtA and YycK (Figure 3.9) to DegP, it could be speculated that the *B. subtilis* proteins may also adopt a similar tertiary structure. The importance of the PDZ domains to the correct functioning of the HtrA-like proteases was shown in the study of Sassoon *et al.* (1999). These authors showed that the positioning of a stop codon at the beginning of the proximal PDZ domain of DegP effectively rendered the protein functionless. A similar strain in which YkdA was truncated at the beginning of the PDZ domain showed expression characteristics similar to that of a YkdA mutant, DN27 (Table 4.3) (Bill Byrne, pers. comm.). Western blot analysis also failed to detect this truncated YkdA derivative indicating its extreme instability (D. Noone unpublished data).

6.3. Expression Analysis of *B. subtilis* HtrA Proteases.

The temporal expression of the HtrA-like proteases of *B. subtilis* was investigated with promoter fusions to the *bgaB* reporter gene. For *ykdA* and *yvtA* the pattern of expression was low and constant during the exponential phase of growth followed by a gradual accumulation during the transition and stationary phases of the growth cycle (Figure 5.1A, B). This pattern of expression suggested a role in stationary phase processes which indeed proved the case (see sections 6.3.1. and 6.3.2 below). In contrast, the expression of *yycK* (*yyxA*) was found to have a bimodal pattern of ß-galactosidase accumulation. Expression was evident during the exponential phase of growth of strains grown in rich medium (LB) which then decreased (Figure 5.2B). In cultures grown in SM broth a second postexponential induced expression was observed (Figure 5.2A). This pattern of expression agrees with that previously observed and shown to be mediated by a σ^G-dependent promoter (Fabret and Hoch, 1998).

6.3.1. Stress induced expression of *ykdA* and *yvtA*.

Since the *degP*(*htrA*) gene had been shown to be heat shock inducible in *E. coli* (Lipinska *et al.*, 1988) we sought to examine if this same general stress would affect the expression of the genes encoding HtrA-like proteases of *B. subtilis*. When strains containing promoter fusions of *ykdA* and *yvtA* to *bgaB* were subjected to a temperature shift from 37°C to 48°C the expression of both the *ykdA* and *yvtA* promoter fusions were seen to increase (Figure 4.2 and Table 5.2). The pattern of expression of *ykdA* showed an initial induction followed by a second period of accumulation. This biphasic pattern may be the result of an initial stress induction followed by the beginning of the temporal induction. *yvtA* heat shock induction was also seen to follow the same general pattern with the initial induction being more pronounced but also leading to a further and continuing accumulation of BgaB activity. In a recent study of the global transcriptional response to heat shock using array technology both *ykdA* and *yvtA* were identified as heat shock inducible genes (Helmann *et al.*, 2001). To identify the promoter(s) allowing for heat shock induction primer extension mapping analysis of both *ykdA* and *yvtA* was undertaken. This confirmed that both genes were heat shock induced at the transcriptional level and from very similar control sequences (Figures

4.1, 4.9 and 5.7, and see below). One notable aspect of the pattern of transcriptional induction for both *ykdA* and *yvtA* is the maintenance of an increased level of transcript from cells maintained at the heat shock temperature (50°C). This in contrast to "classical" heat shock induced responses where an initial burst of transcription is generally followed by an adaptation phase displaying a lower level of transcript, relative to the initial burst, but higher than the level observed in non stressed cells (see section 1.3.). This difference may be due to the fact that the primer extension analysis of *ykdA* and *yvtA* was undertaken in their respective mutant backgrounds where a higher level of signal may persist (see below). In support of this theory the study of Helmann *et al.*, (2001) undertaken in a wild type background found a more "classical type" induction profile with a rapid shut off phase following the initial induction.

In contrast to *ykdA* and *yvtA*, the expression of *yycK* (*yyxA*) was not induced by heat shock (Figure 5.3). In agreement with this result *yycK* (*yyxA*) was not shown to be induced by heat shock in a global heat shock analysis study (Helmann *et al.*, 2001).

6.3.2. Secretion Stress Induction of YkdA and YvtA.

In *E.coli* the *degP* gene is induced by a variety of stresses which generate increased amounts of misfolded proteins in the cell envelope (Raivio and Silhavy, 2000). This response is mediated through two regulatory systems both of which require the σ^E sigma factor or an unknown suppressor(s) mutation (Danese *et al.*, 1995; De Las Penas *et al.*, 1997a). Although *B. subtilis* does not contain a membrane bound periplasm it does have the ability to secrete large amounts of proteins both from wild type and industrially engineered strains (Ferrari *et al.*, 1993). To determine if the extracytoplasmically located HtrA-like proteases of *B. subtilis* performed any similar role we analysed the expression of *ykdA*, *yvtA* and *yycK* in strains overproducing the secretable heterologous protein AmyL (Stephenson *et al.*, 1998; Stephenson and Harwood, 1998). We found that both *ykdA* and *yvtA* genes were induced to a dramatic extent in strains overproducing AmyL (Figures 5.4, 5.7). The level of ß-galactosidase from the respective fusions was increased over 100-fold in these strains in

comparison the levels observed in a wild type background. Furthermore, in the secretion stressed strains AmyL is expressed and secreted in the exponential phase of the growth cycle but the induction of *ykdA* and *yvtA* is not seen until the transition phase. Another notable feature is that this induction is followed by a decrease in the level of accumulation of ß-galactosidase. To confirm this finding a transcriptional analysis using primer extension was undertaken with RNA isolated from the xylose inducible AmyL overproducing strain KS408. These experiments confirmed that expression of both genes occurred at the transcriptional level. Finally the primer extension analysis also demonstrated that the transcriptional induction in response to secretion stress utilised the same promoters employed for heat shock induction.(see Figures 4.9, 5.6 and 5.7).

Since the lack of a heat shock induction of *yycK* (*yyxA*) indicated a different mode of regulation for this gene, no induction of *yycK* was observed from strains overproducing AmyL (Figure 5.5). This result together with the absence of the upstream control sequence common to *ykdA* and *yvtA* implies *yycK* is regulated through a different mechanism(s).

6.4. Regulation of the HtrA-like proteases of *B. subtilis*.

Analysis of the upstream control sequences of the genes encoding the HtrA-like proteases of *B. subtilis* revealed a strikingly similar 56-bp sequence which is 91% identical between the *ykdA* and *yvtA* promoter sequences (Figure 3.10). This sequence is not present upstream of the *yycK* gene. Such a level of identity between the upstream regions of *ykdA* and *yvtA* suggested that both genes would be regulated in a similar manner. This indeed proved to be the case as documented above. We also noted the presence of a repeated octameric sequence common to both control regions (Figure 4.1). To analyse and identify the important positive and negative elements of this *cis*-acting sequence a sequential deletion analysis of the *ykdA* control sequence was carried out. The *ykdA* sequence was chosen as this gene was initially identified as having an autoregulatory component to its expression (see section 4.5 and below) and therefore higher levels of BgaB activity from the fusions used in this study made the analysis more amenable. The sequential 5' and 3' deletions of the *ykdA* control sequence were made, fused to *bgaB* and moved into either wild type or *ykdA* mutant strain

backgrounds (DN25 and DN26). This analysis allowed the identification of the important *cis* acting sequences for expression of *ykdA*. The main feature is that a 60-bp sequence delineated by the 5'Δ47 and 3'Δ61 end points allows for heat shock induction in both the wild type and *ykdA* null backgrounds (Figure 4.3 and Table 4.3). Four of the six identified octameric repeats are found within this sequence. Although a similar deletion analysis has not been undertaken on the *yvtA* control sequence, a point mutation in the fourth T nucleotide (T → G) of repeat I of the *yvtA* sequence (Figure 4.1) causes a 10-fold decrease in the expression of a *yvtA-bgaB* fusion in a *ykdA* or *yvtA* null backgrounds (Darmon *et al.*, 2002). A genome scan using the 56-bp sequence containing these motifs did not reveal any other homologous sequences present elsewhere on the *B. subtilis* chromosome.

The control sequences upstream of *ykdA* and *yvtA* do not contain any motifs similar to the operator sequences bound by the regulators of the class I and class III heats shock regulons HrcA and CtsR, respectively (Zuber and Schumann, 1994; Derre *et al.*, 1999b). Neither do they possess any σ^B type promoter sequences which controls the class II heat shock response. However, since both *ykdA* and *yvtA* were found to be heat shock inducible, signalling might proceed through a protein belonging to one of the above heat shock regulons. Analysis of the expression from a *ykdA-bgaB* fusion in strains mutated for each of the characterised regulators did not have any discernible effect on heat shock induction (Figure 4.3). Given that the Class I, II and III regulons for the most part control genes encoding cytosolic heat shock and general stress proteins, this result agrees with the probable extracytoplasmic localisation of YkdA and YvtA. Furthermore it would imply a transmembrane signal transduction system is part of the regulatory system controlling *ykdA* and *yvtA* expression.

6.4.1. Autoregulatory nature of *ykdA* and *yvtA* expression.

One of the findings in this work has been to show that there is an autoregulatory component involved in the control of expression of *ykdA* and *yvtA* (Tables 4.3 and 5.1). Autoregulation is commonly observed in genes encoding transcriptional regulators (Huang *et al.*, 1997; Huang *et al.*, 1998; Kobayashi *et al.*, 2001). However, since the products of the *ykdA* and

yvtA genes encode proteases the mechanism involving autoregulation most probably involves their action in controlling the level of inducing signal. The induction by heat and secretion stresses of *yvtA* and *ykdA* will cause an increase in the levels of denatured or misfolded proteins. In response the cell must deal with this situation by increasing the amount of proteases or chaperones to refold or degrade the damaged proteins, as occurs in all organisms through classical heat shock responses. In strains mutated for either *ykdA* or *yvtA* the level of misfolded proteins presumably increases. The regulatory system controlling expression of *ykdA* and *yvtA* must be sensitive to this increase and responds by increasing expression of its regulon promoters (Figure 5.1). In contrast to the strains mutated for *ykdA* or *yvtA* a strain engineered to increase the level of YkdA displays an attentuated response to heat shock (Figure 4.6), further evidence that the autoregulatory mechanism is linked to the functional role of the proteases. The increase in the level of signal in mutant backgrounds would appear to be influenced temporally as expression of *yvtA* or *ykdA* is dramatically increased in the postexponential phase of the growth cycle (Figure 5.1) This observation is further evidence for the role of YkdA and YvtA in monitoring the increased levels of proteins which are secreted in the stationary phase of the growth cycle.

There are examples in *B. subtilis* where the levels of proteins, having extracytoplasmic functions, can be altered without causing a feedback response to increase expression. The expression of the genes encoding cell wall protease WprA and of the *dlt* operon controlling the degree of D-alanylation of teichoic acid is unaffected by lack of these proteins (Stephenson and Harwood, 1998; Thwaite *et al.*, 2002). Similarly expression from the promoter of the *sipS* gene encoding an signal peptidase with a major role in secretion is unaffected by deletion of the *sipS* structural gene (Bolhuis *et al.*, 1996). These observations could be due to redundancy of function by other systems. The fact that both *yvtA* and *ykdA* display auto- and reciprocal cross regulation implies that these genes encode proteases whose functions cannot be compensated by other proteins in *B. subtilis*. This conclusion is supported by the severe phenotype observed in strains mutated for *ykdA* and *yvtA* (see sections 5.2.2., 5.2.5. and Figure 5.9).

In contrast to the autoregulatory nature of *ykdA* and *yvtA* expression, there appears to be no effect from the loss of DegP function on a P*degP-lacZ* fusion or increased activation of the σ^E or Cpx regulons (Mecsas *et al.*, 1993; Raina *et al.*, 1995) . This may be an indication that the *Bacillus* proteases undertake a more significant housekeeping role than the *E. coli* DegP protease under physiological conditions.

6.4.2. The CssR/S two component system regulates expression of *ykdA*(*htrA*) and *yvtA*.

Shortly after completion of this work a paper was published by laboratories involved in the European *Bacillus* Secretion Group (EBSG) dealing with the response of *B. subtilis* to severe secretion stress (Hyyrylainen *et al.*, 2001). These authors identified the *yvqAB* operon as encoding homologues of the Cpx system which regulates *degP*(*htrA*) in *E. coli* (see section 1.5.8.). The *yvqAB* operon was renamed *cssRS* to reflect its role in controlling the cells response to secretion stress. The importance of the Css system to the survival of the cell when challenged with the overproduced model secretory protein AmyQ was shown unambiguously in the study of Hyyrylainen *et al.* (2001). In addition it also demonstrated that levels of a mutant version of a native extracytoplasmically located foldase, PrsA, is influenced by the presence or absence of the Css system. Assuming that proteolysis was a major factor in mediating these responses, the expression of *htrA*(*ykdA*) was investigated in cells mutated for *cssS*. These analyses showed that *htrA*(*ykdA*) expression is controlled by the Css system.

The study described above corroborated our findings that *ykdA*(*htrA*) was induced by secretion stress. It is noteworthy that the basal levels of expression from the *htrA-lacZ* fusion used by Hyyrylainen and co-workers were significantly increased compared to our levels (compare Table 1,Hyyrylainen *et al.*, 2001 to Figure 5.4 [this thesis]) . This provides further evidence of the autoregulatory nature of *ykdA* expression. However, the induced level of expression is further increased in the study of Hyyrylainen *et al.* (2001) which may be due to the nature or amount of the stress from overproduced AmyQ protein. We used a single copy xylose inducible *amyL* construct while the AmyQ protein was produced constitutively from a

multicopy construct. Use of the multicopy system with either our *bgaB* or *lacZ* fusion constructs has also demonstrated increased level of reporter activity similar to those reported by Hyyrylainen *et al.*(2001) (C. Hough pers comm.). These data may indicate that the expression of *ykdA*(*htrA*) is dose responsive to the levels of overproduced and presumably misfolded or unprocessed amylases.

In an ongoing collaboration between our laboratory and those of Bron, van Dijl and Kuipers, it has now been shown that *yvtA* is also regulated by the Css system. In addition the *cssRS* operon has been shown to be subject to positive autoregulation by the CssRS system (Darmon *et al.*, 2002) and independently in a major transcriptome analysis (Kobayashi *et al.*, 2001). This latter study also suggests that the Css regulon has additional members. It is interesting to note that the transcriptional induction of the *cssRS* operon is much less spectacular than that of *ykdA* or *yvtA* following secretion stress. Additionally heat shock induction at the transcriptional level of *cssRS* is only observed in a *ykdA* mutant background (see below) (Darmon *et al.*, 2002). These results may reflect the major response of the two-component system to stress is mediated at the post-translational level culminating in the phosphorylation of the response regulator CssR → CssR~P.

6.4.3. Another regulator(s) for HtrA-like proteases in *B. subtilis* ?

Some important differences arise from a comparison of the results contained in this thesis and those of Hyyrylainen *et al.* (2001) In constructing a strain (DN115) doubly mutated for *ykdA* and *yvtA* we noted the accumulation of suppressors by transformants to allow normal wild-type growth. This same phenotype was not reported when the *cssS* or *cssR* genes were inactivated by insertion of pMUTIN (Hyyrylainen *et al.*, 2001). We also found that strains lacking the protease activities of both *ykdA* and *yvtA* in strains DN115 or DN200 cannot grow at increased yet sublethal temperatures. This thermosensitive phenotype was not reported from strains containing mutations in *cssS* (BFA2462) or *cssR* (BFA2461) and analysed during the *Bacillus* Functional Analysis programme.

The differences noted above could be accounted for by a number of possibilities; (1) residual expression of *ykdA*, *yvtA* or both genes in the absence of the CssRS system, whereby there may be sufficient expression from the σ^A-type promoter motif to allow growth at high temperatures, and (2) the existence of a second system positively regulating expression of *ykdA* and *yvtA*. Our initial attempts to identify a regulator for *ykdA* focused on the analysis of *ykdA* expression in strains bearing mutations in genes encoding ECF-type sigma factors (see Table 4.2). This was promoted by the presence of generic-type ECF promoter motifs in the control regions of *ykdA* and *yvtA* (see Figure 4.1). Apart from a small decrease in the level of heat shock induction in the strain (DN52) bearing a mutation in *sigX* (σ^X) this analysis indicated no role for ECF σ's in the expression of *ykdA*. In agreement with this result the promoters shown to be regulated by the *B. subtilis* ECF sigma factors σ^X and σ^W have yielded consensus promoter motifs for both these regulators neither of which are found upstream of *ykdA* of *yvtA* (Huang and Helmann, 1998; Huang *et al.*, 1999). The regulation by an ECF-type sigma factor may only become apparent under certain specialised conditions not yet tested. Given the proven antagonistic relationship between σ^X and σ^W controlled promoters there could be conceivable competition for binding sites between CssR and any of the ECF sigma factors which might regulate the HtrA proteases. In *E. coli* a number of genes have now been shown to possess such dual regulation by the Cpx system and $E\sigma^E$ (Dartigalongue *et al.*, 2001b; De Wulf *et al.*, 2002). Therefore regulation by an ECF sigma factor might only become apparent in a *cssRS* mutant strain. The current wave of global transcriptome analyses emerging from the *Bacillus* scientific community may identify secondary regulators if they exist.

6.5. Roles of *B. subtilis* HtrA-like proteases.

The work presented in this thesis confirms the initial homology based predictions of a role for YkdA and YvtA as stress responsive proteins of the HtrA-like family of proteases. In many other bacteria similar stress responsive roles has been shown for HtrA homologues by expression analysis or implied from phenotypic studies. The data in this thesis shows that *ykdA* and *yvtA* are expressed in response to the general stress of heat shock. Both genes possess a similar control sequence, not present in *yycK* a third gene encoding a HtrA-like

protease, which is not heat shock inducible. In a wild type background both *yvtA* and *ykdA* are induced between three to ten fold by heat shock (Table 5.2).

The importance of having a proteolytically functional copy of either YkdA or YvtA is indicated by the phenotype of transformants resulting from the creation of a *ykdA yvtA* double mutant (DN115). This strain displays a small round mucoid colony morphology which rapidly accumulates a suppressor(s) with wild type colony morphology and growth characteristics. Strain DN115 grows similarly to the wild type or single mutant strains in rich medium at 37°C (see Figures 5.8, 5.9). However, when strain DN115 is shifted to a temperature of 52°C it rapidly ceases to grow and starts to lyse. Therefore in keeping with its role as a heat shock protease the presence of a HtrA-type protease is required by *B. subtilis* to survive prolonged heat shock exposure. Furthermore since these experiments were performed in strains with an intact *yycK* gene it implies that YycK cannot fulfil the roles undertaken by YkdA or YvtA.

The second stress responsive role for the YkdA and YvtA proteases occurs when a heterologous secretory protein is overproduced. For this series of experiments the model protein AmyL from *B. licheniformis* was used. It has been demonstrated that the AmyL protein when overproduced from an inducible promoter is susceptible to considerable cell associated degradation (Stephenson and Harwood, 1998; Jensen *et al.*, 2000). The induction of *ykdA* and *yvtA* in response to secretion stress is of the order of a hundred fold (Figure 5.4) and is in marked contrast to the induction observed following heat stress. It should also be noted that the secretion stress induction observed for *ykdA* was in a strain (RC010) in which the level of YkdA was putatively elevated from the presence of IPTG in the medium. This preloading of the cell with increased YkdA was not apparently sufficient to deal with the secretion stress and a considerable induction of the *ykdA* promoter was seen. This observation may indicate that the level of YkdA in strain RC010 was not comparable to that in the wild type strain as no direct measure of the amount of the YkdA protein has been undertaken. However in strain DN33 the need for induction of *ykdA* following heat shock

was effectively abrogated by increasing expression of *ykdA* from an unregulated Pspac promoter (Figure 4.6).

The pattern of transcriptional induction following heat stress and secretion stress also differs. Heat shock induction appears to generate a sustained increase in transcription while a definite peak of transcription followed by an adaptation phase is observed in the case of secretion stress (compare Figures 4. 2 and 5. 4). This difference in the profiles may provide some clues as to the primary roles and sensing mechanisms employed for expression of YkdA and YvtA proteases. An important aspect of the induction by secretion stress is the link to the transition phase of the growth cycle. Many proteins involved in the chaperoning (PrsA), translocation (SecA) and processing (Sips) of secreted proteins are induced in a temporal dependent manner, at the transition phase, similar to that of *ykdA* and *yvtA* (Tjalsma *et al.*, 1997; Herbort *et al.*, 1999; Hyyrylainen *et al.*, 2001). At this stage of the growth cycle *B. subtilis* begins to secrete large amounts of enzymes to enable the breakdown and subsequent utilisation of complex nutrient sources in the medium. Therefore it may be that under conditions of secretion stress when the cell is preloaded with a large amount of a single heterologous protein (AmyL) much of the secretory machinery is rapidly saturated or titrated by this protein. This presumably leads to incorrect folding, aggregation, and a general blocking of the entire general secretion pathway both by native proteins and AmyL. This may in turn prevent the export of important extracellular enzymes such as the autolysins necessary for proper cell division and cell wall integrity. Therefore it is important that a system exists to remove misfolded or aggregated proteins from the secretion gateways. The greater induction profile seen under conditions of secretion stress may reflect a more intimate contact between the components of the secretion machinery and the sensing system inducing *ykdA* and *yvtA* expression. In contrast, during heat stress the degree of protein misfolding and the species of proteins affected could be quite different. Many of the proteins damaged may be remote from the sensing system while the endogenous proteins of the secretion machinery may be less prone to thermal denaturation thereby functioning almost normally under these conditions. However, as seen in *ykdA* or *yvtA* mutant strains, the level of induction of either *ykdA* or *yvtA* is considerably increased by heat stress possibly due to

higher levels of misfolded proteins at the secretion gateways. The increased level of expression even under non heat shock conditions in the single mutant strains suggests that while the cells grow similarly to wild type the cells are still experiencing increased stress conditions. This observation implies a less than complete functional redundancy between YkdA and YvtA or that co-operation is required between the two proteases for optimum functioning.

DegP in *E. coli* has been shown to possess chaperone activity which predominates at lower temperatures (Spiess *et al.*, 1999). There are a number of results in this thesis which also suggest that YkdA and YvtA may also perform chaperone roles. In either of the *ykdA* or *yvtA* single mutant strains there is an increase in the hydrogen peroxide resistance of these strains (Figures 4.10 and 5.10). No apparent role for the Per regulon in mediating this increased resistance is implied from the analysis of an *ahpCF-lacZ* fusion in the *ykdA* mutant strain (Figure 4.11). A plausible explanation for this resistance may be enhanced proteolysis of oxidatively damaged proteins by increased amounts of the remaining protease. However in strain DN200, containing a null mutation in *yvtA* and a mutation in the serine active site of YkdA1[S290M], there is an equally high level of hydrogen peroxide resistance (Figure 5.10). Secondly, in strain AH24 also containing the YkdA1 protein but with a wild type *yvtA* allele, both the basal uninduced level and the heat shock induced level of expression from a *ykdA-bgaB* fusion are very similar to a wild type strain rather than that of a *ykdA* mutant strain (compare Figure 4.8B to Table 4.3). This implies that the YkdA1[S290M] protein is to some degree functioning to influence the level of signal. We now know that the YkdA1 protein is produced as demonstrated by western blot and 2D-gel proteome analysis of strain DN200 (D. Noone, unpublished data; H. Antlemann, E Darmon *et al* unpublished data). Whether the YkdA1 protein possesses true chaperone activity has yet to be determined. Alternatively the mutant protein may simply bind any damaged proteins thereby lowering the overall level sensed through CssS directly or indirectly (see below). Sequestration of unfolded substrates by proteolytically inactive proteases leading to observable phenotypes has been demonstrated previously for Lon and DegP in *E coli* (Van Melderen and Gottesman, 1999; Misra *et al.*, 2000).

It would also be interesting to know if either or both YvtA and YkdA had non-stress responsive roles. It is noteworthy that HtrA in *L. lactis* is suggested to be the sole extracellular protease responsible for processing and maturation of secreted proteins. One way to investigate these possibilities would be through a screen using YkdA or YvtA as bait proteins in a yeast two hybrid system. We are currently involved in a collaboration with the laboratory of Dr. Philippe Noirot to undertake such a screen in an effort to identify potential substrates for YkdA and YvtA. This approach may yield some indications of housekeeping functions performed by YvtA or YkdA.

The results of the data presented in this thesis and the recent finding of the major role for the CssRS two-component system in the regulation of *ykdA* and *yvtA* allow for the formation of a model of the roles performed by and mechanisms governing YkdA and YvtA expression (Figure 6.1). It is assumed that the major role of the two proteases is to prevent a build up of misfolded or aggregated proteins at the cell membrane/wall interface. This may be accomplished by either chaperoning misfolded proteins or degrading aggregated complexes. Since a much greater degree of induction is observed under secretion stress conditions it may be that the primary role of YkdA and YvtA is to ensure that the translocation and secretory gateways are functional. The lower level of induction seen under heat shock presumably reflects a lesser insult to the secretion machinery or that additional mechanisms are available to deal with this more general stress. It is not known yet if the extracytoplasmic loop of CssS directly senses the signal or if any intermediary protein whose titration away from a (putative) interaction with CssS may determine activation of the CssRS system.

6.6 Concluding Remarks

It is clear that *ykdA* and *yvtA* respond to two types of stress; (i) heat shock and (ii) secretion stress. Both types of stress are know to generate misfolded and aggregated proteins. The heat shock induction of both genes is rather restrained but reproducible. This may reflect the putative localisation of YkdA and YvtA on the outside of the cytoplasmic membrane. In support of this probable localisation, treatment of *B. subtilis* with the antibiotic puromycin,

known to induce the cytosolic heat shock response (Kruger *et al.*, 1996; Gerth *et al.*, 1998; Mogk *et al.*, 1998; Derre *et al.*, 1999a), did not cause increased expression of *ykdA* or *yvtA*. A much greater induction of expression of both *ykdA* and *yvtA* is seen to occur following secretion stress, a condition where increased amounts of secretory proteins accumulate at the cell membrane/wall interface (Jensen *et al.*, 2000; Hyyrylainen *et al.*, 2001; Vitikainen *et al.*, 2001). In contrast the third HtrA-type protease, YycK, does not respond to either stress. This difference in expression pattern is correlated with the absence of the sequence common to the control regions of both *ykdA* and *yvtA*. At least one system know to control expression of *ykdA* and *yvtA* has now been identified, CssRS, which itself is responsive at the transcriptional level to secretion stress and to heat stress in a *ykdA* or *yvtA* mutant background. Therefore this work has identified the first members of a regulon whose primary role appears to be to provide a "heat shock type response" at the extracytoplasmic side of the cell envelope of *B. subtilis*.

Current and future work will focus on a number of areas to further expand the understanding of the roles and regulation of the HtrA-type proteases in *B. subtilis*. (1) The precise localisation of YkdA, YvtA and YycK using western immunoblotting and proteomic analysis; (2) identification of the natural substrates of YkdA and YvtA through yeast two-hybrid screening which may also identify other possible roles of the HtrA-type proteases; (3) identification of other possible regulatory components in the induction mechanism controlled by CssRS; (4) the identification of other stimuli and mechanisms inducing *ykdA* or *yvtA* operating through or independent of CssRS.

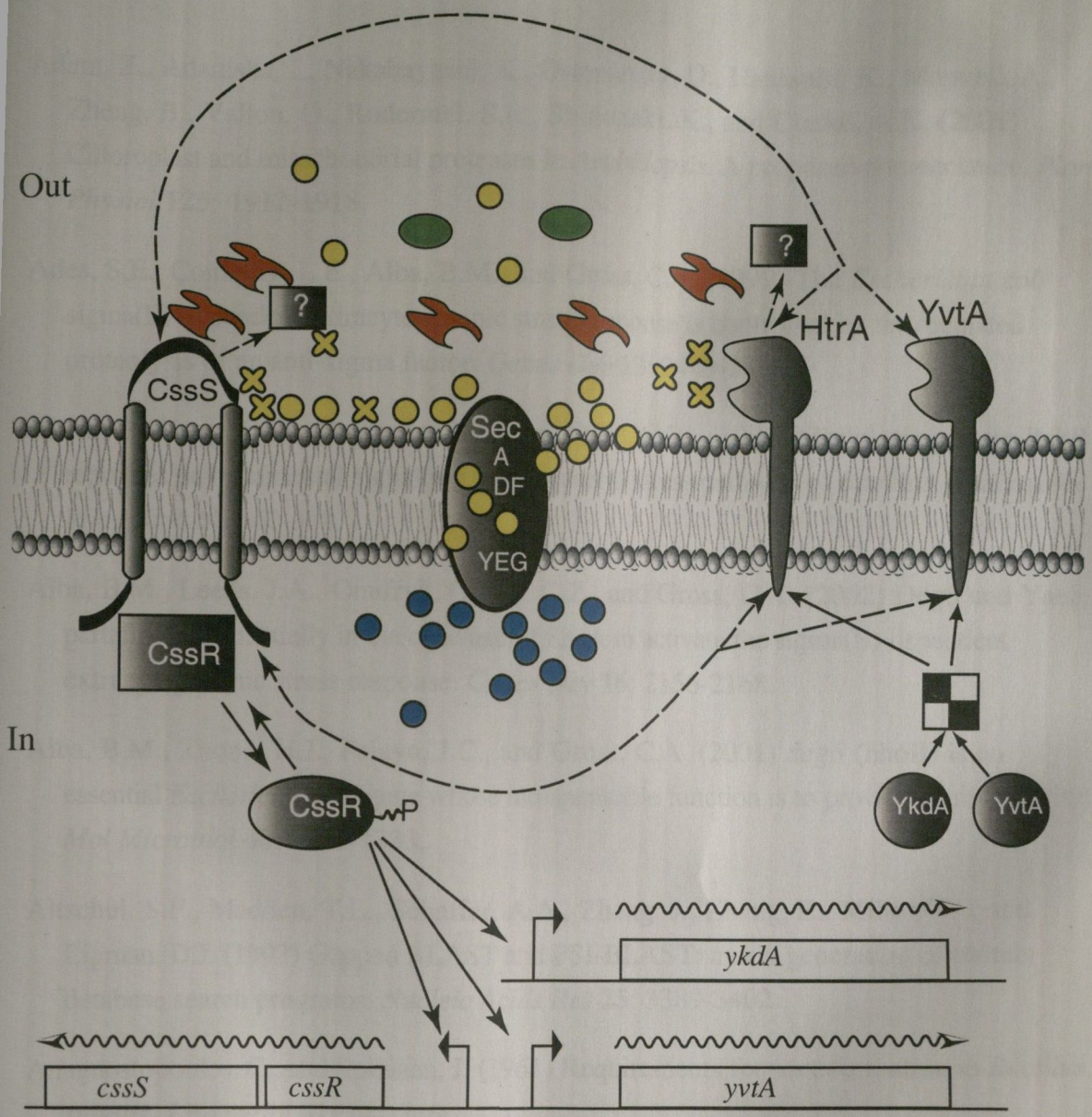


Figure 6.1. Model for regulation of expression of YkdA and YvtA by the CssRS two-component system. The YkdA and YvtA proteins are anchored in the cytoplasmic membrane via their single transmembrane domains. The Sec translocon, composed of its multiple integral membrane subunits, SecA,E, D, F, Y and G, is indicated by a single oval.

The CssS histidine sensor kinase containing two transmembrane domains, an extracytoplasmically localised loop and cytoplasmic N and C termini is also shown. Other known components of the secretion machinery, PrsA, SipSTVW, Lsp, BdbB/C and SppA have been omitted for clarity. The green and red symbols indicate thermolabile proteins which necessitate the proteolytic action of YkdA and/or YvtA after heat stress. The blue and yellow symbols represent overexpression of heterologous amylases which can saturate the secretion machinery causing misfolding both of the amylase and other endogenous secretory proteins indicated by yellow star shapes at the cell membrane /wall interface. The extracytoplasmic loop of CssS is presumed to sense these misfolded proteins either directly or through an intermediary molecule (rectangle) possibly through titration which results in activation of CssR by phosphorylation. The CssR~P regulator binds to and activates transcription (wavy lines) of *ykdA*, *yvtA* and *cssRS*. YkdA and YvtA are produced and transported to the membrane by an unknown mechanism (checkered box). Dashed lines indicate other possible signaling circuits whereby CssRS could sense stress communicated through, (i) the N-terminal domains of YkdA and/or YvtA (ii) the extracytoplasmic domains of these same proteases. It is assumed that *in vivo* the proteins shown above are in closer physical proximity which may facilitate this latter sensing mechanism. This diagram is based upon a more detailed schematic contained within the review of Tjalsma *et al.*, (2000).

REFERENCES

Adam, Z., Adamska, I., Nakabayashi, K., Ostersetzer, O., Haussuhl, K., Manuell, A., Zheng, B., Vallon, O., Rodermel, S.R., Shinozaki, K., and Clarke, A.K. (2001) Chloroplast and mitochondrial proteases in *Arabidopsis*. A proposed nomenclature. *Plant Physiol* **125**: 1912-1918.

Ades, S.E., Connolly, L.E., Alba, B.M., and Gross, C.A. (1999) The *Escherichia coli* sigma(E)-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti-sigma factor. *Genes Dev* **13**: 2449-2461.

Akbar, S., Kang, C.M., Gaidenko, T.A., and Price, C.W. (1997) Modulator protein RsbR regulates environmental signalling in the general stress pathway of *Bacillus subtilis*. *Mol Microbiol* **24**: 567-578.

Alba, B.M., Leeds, J.A., Onufryk, C., Lu, C.Z., and Gross, C.A. (2002) DegS and YaeL participate sequentially in the cleavage of RseA to activate the sigma(E)-dependent extracytoplasmic stress response. *Genes Dev* **16**: 2156-2168.

Alba, B.M., Zhong, H.J., Pelayo, J.C., and Gross, C.A. (2001) *degS* (hhoB) is an essential *Escherichia coli* gene whose indispensable function is to provide sigma activity. *Mol Microbiol* **40**: 1323-1333.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.

Anagnostopoulos, C., and Spizizen, J. (1961) Requirements for transformation in *Bacillus subtilis*. *J Bacteriol* **81**: 741-746.

Antelmann, H., Darmon,E., Noone, D., Veening, J.M., Westers, H., Bron, S., Kuipers, O.P., Devine, K.M., Hecker, M., and van Dijl, J.M. (2003) The extracellular proteome of *Bacillus subtilis* under secretion stress conditions. *Mol Microbiol* in press.

Antelmann, H., Tjalsma, H., Voigt, B., Ohlmeier, S., Bron, S., van Dijl, J.M., and Hecker, M. (2001) A proteomic view on genome-based signal peptide predictions. *Genome Res* **11**: 1484-1502.

Archibald, A.R., Hancock, I.C., and Harwood, C.R. (1993) Cell Wall Structure, Synthesis, and Turnover. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*. Sonenshein, A.L., Hoch, J.A. and Losick, R. (eds). Washington: American Society for Microbiology.

Arnold, F., Bedouet, L., Batina, P., and Robreau, G. (1999) Cloning and sequencing of the central region of the flagellin gene from the Gram-positive bacterium *Clostridium tyrobutyricum* ATCC 25755. *Microbiol Immunol* **43**: 1-8.

Baneyx, F., and Georgiou, G. (1990) *In vivo* degradation of secreted fusion proteins by the *Escherichia coli* outer membrane protease OmpT. *J Bacteriol* **172**: 491-494.

Barrett, A.J., Rawlings, N.D., and Woessner, J.F., (eds) (1998) *Handbook of Proteolytic Enzymes.* London: Academic Press.

Bass, S., Gu, Q., and Christen, A. (1996) Multicopy suppressors of *prc* mutant *Escherichia coli* include two HtrA (DegP) protease homologs (HhoAB), DksA, and a truncated R1pA. *J Bacteriol* **178**: 1154-1161.

Beebe, K.D., Shin, J., Peng, J., Chaudhury, C., Khera, J., and Pei, D. (2000) Substrate recognition through a PDZ domain in tail-specific protease. *Biochemistry* **39**: 3149-3155.

Behrens, S., Maier, R., de Cock, H., Schmid, F.X., and Gross, C.A. (2001) The SurA periplasmic PPIase lacking its parvulin domains functions *in vivo* and has chaperone activity. *Embo J* **20**: 285-294.

Belitsky, B.R., Brill, J., Bremer, E., and Sonenshein, A.L. (2001) Multiple genes for the last step of proline biosynthesis in *Bacillus subtilis*. *J Bacteriol* **183**: 4389-4392.

Bernhardt, J., Volker, U., Volker, A., Antelmann, H., Schmid, R., Mach, H., and Hecker, M. (1997) Specific and general stress proteins in *Bacillus subtilis*--a two- deimensional protein electrophoresis study. *Microbiology* **143**: 999-1017.

Biswas, I., Gruss, A., Ehrlich, S.D., and Maguin, E. (1993) High-efficiency gene inactivation and replacement system for gram- positive bacteria. *J Bacteriol* **175**: 3628-3635.

Blaszczak, A., Georgopoulos, C., and Liberek, K. (1999) On the mechanism of FtsH-dependent degradation of the sigma 32 transcriptional regulator of *Escherichia coli* and the role of the Dnak chaperone machine. *Mol Microbiol* **31**: 157-166.

Bol, D.K., and Yasbin, R.E. (1991) The isolation, cloning and identification of a vegetative catalase gene from *Bacillus subtilis*. *Gene* **109**: 31-37.

Bolhuis, A., Sorokin, A., Azevedo, V., Ehrlich, S.D., Braun, P.G., de Jong, A., Venema, G., Bron, S., and van Dijl, J.M. (1996) *Bacillus subtilis* can modulate its capacity and specificity for protein secretion through temporally controlled expression of the *sipS* gene for signal peptidase I. *Mol Microbiol* **22**: 605-618.

Bolhuis, A., Tjalsma, H., Smith, H.E., de Jong, A., Meima, R., Venema, G., Bron, S., and van Dijl, J.M. (1999) Evaluation of bottlenecks in the late stages of protein secretion in *Bacillus subtilis*. *Appl Environ Microbiol* **65**: 2934-2941.

Boucher, J.C., Martinez-Salazar, J., Schurr, M.J., Mudd, M.H., Yu, H., and Deretic, V. (1996) Two distinct loci affecting conversion to mucoidy in *Pseudomonas aeruginosa* in cystic fibrosis encode homologs of the serine protease HtrA. *J Bacteriol* **178**: 511-523.

Boylan, S.A., Redfield, A.R., Brody, M.S., and Price, C.W. (1993) Stress-induced activation of the sigma B transcription factor of *Bacillus subtilis*. *J Bacteriol* **175**: 7931-7937.

Boylan, S.A., Rutherford, A., Thomas, S.M., and Price, C.W. (1992) Activation of *Bacillus subtilis* transcription factor sigma B by a regulatory pathway responsive to stationary-phase signals. *J Bacteriol* **174**: 3695-3706.

Boylan, S.A., Thomas, M.D., and Price, C.W. (1991) Genetic method to identify regulons controlled by nonessential elements: isolation of a gene dependent on alternate transcription factor sigma B of *Bacillus subtilis*. *J Bacteriol* **173**: 7856-7866.

Brenner, S. (1988) The molecular evolution of genes and proteins: a tale of two serines. *Nature* **334**: 528-530.

Buchberger, A., and Bukau, B. (1997) *Escherichia coli* DnaK. In *Guidebook to Molecular Chaperones and Protein-Folding Catalysts*. Gething, M.-J. (ed). Oxford: Oxford University Press, pp. 22-25.

Bukau, B., Deuerling, E., Pfund, C., and Craig, E.A. (2000) Getting newly synthesized proteins into shape. *Cell* **101**: 119-122.

Bukau, B., and Horwich, A.L. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* **92**: 351-366.

Cavard, D., Lazdunski, C., and Howard, S.P. (1989) The acylated precursor form of the colicin A lysis protein is a natural substrate of the DegP protease. *J Bacteriol* **171**: 6316-6322.

Cheggour, A., Fanuel, L., Duez, C., Joris, B., Bouillenne, F., Devreese, B., Van Driessche, G., Van Beeumen, J., Frere, J.M., and Goffin, C. (2000) The *dppA* gene of *Bacillus subtilis* encodes a new D-aminopeptidase. *Mol Microbiol* **38**: 504-513.

Collinet, B., Yuzawa, H., Chen, T., Herrera, C., and Missiakas, D. (2000) RseB binding to the periplasmic domain of RseA modulates the RseA:sigmaE interaction in the cytoplasm and the availability of sigmaE.RNA polymerase. *J Biol Chem* **275**: 33898-33904.

Connolly, L., De Las Penas, A., Alba, B.M., and Gross, C.A. (1997) The response to extracytoplasmic stress in *Escherichia coli* is controlled by partially overlapping pathways. *Genes Dev* **11**: 2012-2021.

Cortes, G., de Astorza, B., Benedi, V.J., and Alberti, S. (2002) Role of the *htrA* gene in Klebsiella pneumoniae virulence. *Infect Immun* **70**: 4772-4776.

Cosma, C.L., Crotwell, M.D., Burrows, S.Y., and Silhavy, T.J. (1998) Folding-based suppression of extracytoplasmic toxicity conferred by processing-defective LamB. *J Bacteriol* **180**: 3120-3130.

Cosma, C.L., Danese, P.N., Carlson, J.H., Silhavy, T.J., and Snyder, W.B. (1995) Mutational activation of the Cpx signal transduction pathway of *Escherichia coli* suppresses the toxicity conferred by certain envelope- associated stresses. *Mol Microbiol* **18**: 491-505.

Danese, P.N., Oliver, G.R., Barr, K., Bowman, G.D., Rick, P.D., and Silhavy, T.J. (1998) Accumulation of the enterobacterial common antigen lipid II biosynthetic intermediate stimulates *degP* transcription in *Escherichia coli*. *J Bacteriol* **180**: 5875-5884.

Danese, P.N., and Silhavy, T.J. (1997) The sigma(E) and the Cpx signal transduction systems control the synthesis of periplasmic protein-folding enzymes in *Escherichia coli*. *Genes Dev* **11**: 1183-1193.

Danese, P.N., and Silhavy, T.J. (1998) CpxP, a stress-combative member of the Cpx regulon. *J Bacteriol* **180**: 831-839.

Danese, P.N., Snyder, W.B., Cosma, C.L., Davis, L.J., and Silhavy, T.J. (1995) The Cpx two-component signal transduction pathway of *Escherichia coli* regulates transcription of the gene specifying the stress-inducible periplasmic protease, DegP. *Genes Dev* **9**: 387-398.

Darmon, E., Noone, D., Masson, A., Bron, S., Kuipers, O.P., Devine, K.M., and van Dijl, J.M. (2002) A novel class of heat and secretion stress-responsive genes is controlled by the autoregulated CssRS two-component system of *Bacillus subtilis*. *J Bacteriol* **184**: 5661-5671.

Dartigalongue, C., Loferer, H., and Raina, S. (2001a) EcfE, a new essential inner membrane protease: its role in the regulation of heat shock response in *Escherichia coli*. *Embo J* **20**: 5908-5918.

Dartigalongue, C., Missiakas, D., and Raina, S. (2001b) Characterization of the *Escherichia coli* sigma E regulon. *J Biol Chem* **276**: 20866-20875.

Dartigalongue, C., and Raina, S. (1998) A new heat-shock gene, *ppiD*, encodes a peptidyl-prolyl isomerase required for folding of outer membrane proteins in *Escherichia coli*. *Embo J* **17**: 3968-3980.

De Las Penas, A., Connolly, L., and Gross, C.A. (1997a) SigmaE is an essential sigma factor in *Escherichia coli*. *J Bacteriol* **179**: 6862-6864.

De Las Penas, A., Connolly, L., and Gross, C.A. (1997b) The sigmaE-mediated response to extracytoplasmic stress in *Escherichia coli* is transduced by RseA and RseB, two negative regulators of sigmaE. *Mol Microbiol* **24**: 373-385.

De Wulf, P., Kwon, O., and Lin, E.C. (1999) The CpxRA signal transduction system of *Escherichia coli*: growth- related autoactivation and control of unanticipated target operons. *J Bacteriol* **181**: 6772-6778.

De Wulf, P., McGuire, A.M., Liu, X., and Lin, E.C. (2002) Genome-wide profiling of promoter recognition by the two-component response regulator CpxR-P in *Escherichia coli*. *J Biol Chem* **12**: 12.

Derre, I., Rapoport, G., Devine, K., Rose, M., and Msadek, T. (1999a) ClpE, a novel type of HSP100 ATPase, is part of the CtsR heat shock regulon of *Bacillus subtilis*. *Mol Microbiol* **32**: 581-593.

Derre, I., Rapoport, G., and Msadek, T. (1999b) CtsR, a novel regulator of stress and heat shock response, controls *clp* and molecular chaperone gene expression in gram-positive bacteria. *Mol Microbiol* **31**: 117-131.

Derre, I., Rapoport, G., and Msadek, T. (2000) The CtsR regulator of stress response is active as a dimer and specifically degraded in vivo at 37 degrees C. *Mol Microbiol* **38**: 335-347.

Devine, K.M. (1995) The *Bacillus subtilis* genome project: aims and progress. *Trends Biotechnol* **13**: 210-216.

Dong, J., Iuchi, S., Kwan, H.S., Lu, Z., and Lin, E.C. (1993) The deduced amino-acid sequence of the cloned *cpxR* gene suggests the protein is the cognate regulator for the membrane sensor, CpxA, in a two-component signal transduction system of *Escherichia coli*. *Gene* **136**: 227-230.

Dowds, B.C.A., O'Kane, C., Gormley, E., McConnell, D.J., and Devine, K.M. (1988) Integrating Plasmids In The Genetic Engineering Of *Bacilli*. In *Recombinant DNA and Bacterial Fermentation*. Thomson, J.A. (ed). Boca Raton, FA: CRC Press, pp. 137-156.

Doyle, D.A., Lee, A., Lewis, J., Kim, E., Sheng, M., and MacKinnon, R. (1996) Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell* **85**: 1067-1076.

Dubnau, D. (1999) DNA uptake in bacteria. *Annu Rev Microbiol* **53**: 217-244.

Dubnau, D., and Provvedi, R. (2000) Internalizing DNA. *Res Microbiol* **151**: 475-480.

Dworkin, J., and Losick, R. (2001) Linking nutritional status to gene activation and development. *Genes Dev* **15**: 1051-1054.

Ehrlich, S.D. (2001) Aims of the Functional Analysis Project. In *Functional Analysis of Bacterial Genes: A Practical Manual*. Schumann, W., Ehrlich, S.D. and Ogasawara, N. (eds). Chichester UK.: Wiley, pp. 1-3.

Elzer, P.H., Hagius, S.D., Robertson, G.T., Phillips, R.W., Walker, J.V., Fatemi, M.B., Enright, F.M., and Roop, R.M.n. (1996) Behaviour of a high-temperature-requirement A (HtrA) deletion mutant of Brucella abortus in goats. *Res Vet Sci* **60**: 48-50.

Elzer, P.H., Phillips, R.W., Kovach, M.E., Peterson, K.M., and Roop, R.M.n. (1994) Characterization and genetic complementation of a *Brucella abortus* high- temperature-requirement A (*htrA*) deletion mutant. *Infect Immun* **62**: 4135-4139.

Engelmann, S., Lindner, C., and Hecker, M. (1995) Cloning, nucleotide sequence, and regulation of *katE* encoding a sigma B- dependent catalase in *Bacillus subtilis*. *J Bacteriol* **177**: 5598-5605.

Erickson, J.W., and Gross, C.A. (1989) Identification of the sigma E subunit of *Escherichia coli* RNA polymerase: a second alternate sigma factor involved in high- temperature gene expression. *Genes Dev* **3**: 1462-1471.

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

Fabret, C., and Hoch, J.A. (1998) A two-component signal transduction system essential for growth of *Bacillus subtilis*: implications for anti-infective therapy. *J Bacteriol* **180**: 6375-6383.

Faccio, L., Fusco, C., Chen, A., Martinotti, S., Bonventre, J.V., and Zervos, A.S. (2000) Characterization of a novel human serine protease that has extensive homology to bacterial heat shock endoprotease HtrA and is regulated by kidney ischemia. *J Biol Chem* **275**: 2581-2588.

Fanning, A.S., and Anderson, J.M. (1996) Protein-protein interactions: PDZ domain networks. *Curr Biol* **6**: 1385-1388.

Ferrari, E., Jarnagin, A.S., and Schmidt, B.F. (1993) Commercial production of extracellular enzymes. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*. Sonenshein, A.L., Hoch, J.A. and Losick, R. (eds). Washington, D.C.: American Society for Microbiology, pp. 917-937.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., and et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496-512.

Foster, S.M., and Popham, D. (2002) Structure and Synthesis of Cell Wall, Spore Cortex, Teichoic Acids, S-Layers and Capsules. In *Bacillus subtilis and Its Closest Relatives*. Sonenshein, A.L., Hoch, J.A. and Losick, R. (eds). Washington: ASM Press, pp. 21-41.

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

Fukuchi, K., Kasahara, Y., Asai, K., Kobayashi, K., Moriya, S., and Ogasawara, N. (2000) The essential two-component regulatory system encoded by *yycF* and *yycG* modulates expression of the ftsAZ operon in *Bacillus subtilis*. *Microbiology* **146**: 1573-1583.

Gentzsch, M., and Tanner, W. (1996) The PMT gene family: protein O-glycosylation in *Saccharomyces cerevisiae* is vital. *Embo J* **15**: 5752-5759.

Gerth, U., Kruger, E., Derre, I., Msadek, T., and Hecker, M. (1998) Stress induction of the *Bacillus subtilis clpP* gene encoding a homologue of the proteolytic component of the Clp protease and the involvement of ClpP and ClpX in stress tolerance. *Mol Microbiol* **28**: 787-802.

Gill, R.T., DeLisa, M.P., Shiloach, M., Holoman, T.R., and Bentley, W.E. (2000) OmpT expression and activity increase in response to recombinant chloramphenicol acetyltransferase overexpression and heat shock in *E. coli*. *J Mol Microbiol Biotechnol* **2**: 283-289.

Glaser, P., Kunst, F., Arnaud, M., Coudart, M.P., Gonzales, W., Hullo, M.F., Ionescu, M., Lubochinsky, B., Marcelino, L., Moszer, I., and et al. (1993) *Bacillus subtilis* genome project: cloning and sequencing of the 97 kb region from 325 degrees to 333 degrees. *Mol Microbiol* **10**: 371-384.

Gonzalez-Pastor, J.E., San Millan, J.L., Castilla, M.A., and Moreno, F. (1995) Structure and organization of plasmid genes required to produce the translation inhibitor microcin C7. *J Bacteriol* **177**: 7131-7140.

Gottesman, S. (1996) Proteases and their targets in *Escherichia coli*. *Annu Rev Genet* **30**: 465-506.

Gottesman, S. (1999) Regulation by proteolysis: developmental switches. *Curr Opin Microbiol* **2**: 142-147.

Gottesman, S., Roche, E., Zhou, Y., and Sauer, R.T. (1998) The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **12**: 1338-1347.

Gottesman, S., Wickner, S., and Maurizi, M.R. (1997) Protein quality control: triage by chaperones and proteases. *Genes Dev* **11**: 815-823.

Gray, C.W., Ward, R.V., Karran, E., Turconi, S., Rowles, A., Viglienghi, D., Southan, C., Barton, A., Fantom, K.G., West, A., Savopoulos, J., Hassan, N.J., Clinkenbeard, H., Hanning, C., Amegadzie, B., Davis, J.B., Dingwall, C., Livi, G.P., and Creasy, C.L. (2000) Characterization of human HtrA2, a novel serine protease involved in the mammalian cellular stress response. *Eur J Biochem* **267**: 5699-5710.

Gross, C. (1996) Function and Regulation of the Heat Shock Proteins. In *Escherichia coli and Salmonella cellular and molecular biology*. Neidhardt, F. and others (eds). Washington: ASM Press, pp. 1382-1399.

Guerout-Fleury, A.M., Shazand, K., Frandsen, N., and Stragier, P. (1995) Antibiotic-resistance cassettes for *Bacillus subtilis*. *Gene* **167**: 335-336.

Guigueno, A., Dassa, J., Belin, P., and Boquet, P.L. (2001) Oversynthesis of a new *Escherichia coli* small RNA suppresses export toxicity of DsbA'-PhoA unfoldable periplasmic proteins. *J Bacteriol* **183**: 1147-1158.

Haldenwang, W.G. (1995) The sigma factors of *Bacillus subtilis*. *Microbiol Rev* **59**: 1-30.

Haldenwang, W.G., and Losick, R. (1980) Novel RNA polymerase sigma factor from *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **77**: 7000-7004.

Harwood, C.R., and Cutting, S.M., (eds) (1990) *Molecular Biological Methods for Bacillus*. Chichester, U.K.: Wiley.

Haussuhl, K., Andersson, B., and Adamska, I. (2001) A chloroplast DegP2 protease performs the primary cleavage of the photodamaged D1 protein in plant photosystem II. *Embo J* **20**: 713-722.

He, X.S., Shyu, Y.T., Nathoo, S., Wong, S.L., and Doi, R.H. (1991) Construction and use of a *Bacillus subtilis* mutant deficient in multiple protease genes for the expression of eukaryotic genes. *Ann N Y Acad Sci* **646**: 69-77.

Hecker, M., Schumann, W., and Volker, U. (1996) Heat-shock and general stress response in *Bacillus subtilis*. *Mol Microbiol* **19**: 417-428.

Hecker, M., and Volker, U. (1998) Non-specific, general and multiple stress resistance of growth- restricted *Bacillus subtilis* cells by the expression of the sigmaB regulon. *Mol Microbiol* **29**: 1129-1136.

Helmann, J.D., Wu, M.F., Kobel, P.A., Gamo, F.J., Wilson, M., Morshedi, M.M., Navre, M., and Paddon, C. (2001) Global Transcriptional Response of *Bacillus subtilis* to Heat Shock. *J Bacteriol* **183**: 7318-7328.

Henderson, T.A., Dombrosky, P.M., and Young, K.D. (1994) Artifactual processing of penicillin-binding proteins 7 and 1b by the OmpT protease of *Escherichia coli*. *J Bacteriol* **176**: 256-259.

Herbig, A.F., and Helmann, J.D. (2002) Metal Iron Uptake and Oxidative Stress. In *Bacillus subtilis and Its Closest Relatives*. Sonenshein, A.L., Hoch, J.A. and Losick, R. (eds). Washington: ASM Press, pp. 405-414.

Herbort, M., Klein, M., Manting, E.H., Driessen, A.J., and Freudl, R. (1999) Temporal expression of the *Bacillus subtilis secA* gene, encoding a central component of the preprotein translocase. *J Bacteriol* **181**: 493-500.

Herman, C., Thevenet, D., D'Ari, R., and Bouloc, P. (1995) Degradation of sigma 32, the heat shock regulator in *Escherichia coli*, is governed by HflB. *Proc Natl Acad Sci U S A* **92**: 3516-3520.

Hirata, H., Negoro, S., and Okada, H. (1984) Molecular basis of isozyme formation of beta-galactosidases in *Bacillus stearothermophilus*: isolation of two beta-galactosidase genes, *bgaA* and *bgaB*. *J Bacteriol* **160**: 9-14.

Hirata, H., Negoro, S., and Okada, H. (1985) High production of thermostable beta-galactosidase of *Bacillus stearothermophilus* in *Bacillus subtilis*. *Appl. Environ. Microbiol.* **49**: 1547-1549.

Hoa, N.T., Brannigan, J.A., and Cutting, S.M. (2001) The PDZ domain of the SpoIVB serine peptidase facilitates multiple functions. *J Bacteriol* **183**: 4364-4373.

Hoch, J.A., and Silhavy, T.J., (eds) (1995) *Two-Component Signal Transduction.* Washington, DC: ASM Press.

Holmes, D.S., and Quigley, M. (1981) A rapid boiling method for the preparation of bacterial plasmids. *Anal Biochem* **114**: 193-197.

Horsburgh, M.J., and Moir, A. (1999) Sigma M, an ECF RNA polymerase sigma factor of *Bacillus subtilis* 168, is essential for growth and survival in high concentrations of salt. *Mol Microbiol* **32**: 41-50.

Horwich, A.L., Weber-Ban, E.U., and Finley, D. (1999) Chaperone rings in protein folding and degradation. *Proc Natl Acad Sci U S A* **96**: 11033-11040.

Hoskins, J.R., Pak, M., Maurizi, M.R., and Wickner, S. (1998) The role of the ClpA chaperone in proteolysis by ClpAP. *Proc Natl Acad Sci U S A* **95**: 12135-12140.

Hu, S.I., Carozza, M., Klein, M., Nantermet, P., Luk, D., and Crowl, R.M. (1998) Human HtrA, an evolutionarily conserved serine protease identified as a differentially expressed gene product in osteoarthritic cartilage. *J Biol Chem* **273**: 34406-34412.

Huang, X., Decatur, A., Sorokin, A., and Helmann, J.D. (1997) The *Bacillus subtilis* sigma(X) protein is an extracytoplasmic function sigma factor contributing to survival at high temperature. *J Bacteriol* **179**: 2915-2921.

Huang, X., Fredrick, K.L., and Helmann, J.D. (1998) Promoter recognition by *Bacillus subtilis* sigmaW: autoregulation and partial overlap with the sigmaX regulon. *J Bacteriol* **180**: 3765-3770.

Huang, X., Gaballa, A., Cao, M., and Helmann, J.D. (1999) Identification of target promoters for the *Bacillus subtilis* extracytoplasmic function sigma factor, sigma W. *Mol Microbiol* **31**: 361-371.

Huang, X., and Helmann, J.D. (1998) Identification of target promoters for the *Bacillus subtilis* sigma X factor using a consensus-directed search. *J Mol Biol* **279**: 165-173.

Hyyrylainen, H.L., Bolhuis, A., Darmon, E., Muukkonen, L., Koski, P., Vitikainen, M., Sarvas, M., Pragai, Z., Bron, S., van Dijl, J.M., and Kontinen, V.P. (2001) A novel two-component regulatory system in *Bacillus subtilis* for the survival of severe secretion stress. *Mol Microbiol* **41**: 1159-1172.

Hyyrylainen, H.L., Vitikainen, M., Thwaite, J., Wu, H., Sarvas, M., Harwood, C.R., Kontinen, V.P., and Stephenson, K. (2000) D-Alanine substitution of teichoic acids as a modulator of protein folding and stability at the cytoplasmic membrane/cell wall interface of *Bacillus subtilis*. *J Biol Chem* **275**: 26696-26703.

Itzhaki, H., Naveh, L., Lindahl, M., Cook, M., and Adam, Z. (1998) Identification and characterization of DegP, a serine protease associated with the luminal side of the thylakoid membrane. *J Biol Chem* **273**: 7094-7098.

Jensen, C.L., Stephenson, K., Jorgensen, S.T., and Harwood, C. (2000) Cell-associated degradation affects the yield of secreted engineered and heterologous proteins in the *Bacillus subtilis* expression system. *Microbiology* **146**: 2583-2594.

Johnson, K., Charles, I., Dougan, G., Pickard, D., O'Gaora, P., Costa, G., Ali, T., Miller, I., and Hormaeche, C. (1991) The role of a stress-response protein in *Salmonella typhimurium* virulence. *Mol Microbiol* **5**: 401-407.

Jones, C.H., Bolken, T.C., Jones, K.F., Zeller, G.O., and Hruby, D.E. (2001) Conserved DegP protease in gram-positive bacteria is essential for thermal and oxidative tolerance and full virulence in *Streptococcus pyogenes*. *Infect Immun* **69**: 5538-5545.

Jones, C.H., Danese, P.N., Pinkner, J.S., Silhavy, T.J., and Hultgren, S.J. (1997) The chaperone-assisted membrane release and folding pathway is sensed by two signal transduction systems. *Embo J* **16**: 6394-6406.

Jones, C.H., Dexter, P., Evans, A.K., Liu, C., Hultgren, S.J., and Hruby, D.E. (2002) *Escherichia coli* DegP Protease Cleaves between Paired Hydrophobic Residues in a Natural Substrate: the PapA Pilin. *J Bacteriol* **184**: 5762-5771.

Kalman, S., Duncan, M.L., Thomas, S.M., and Price, C.W. (1990) Similar organization of the *sigB* and *spoIIA* operons encoding alternate sigma factors of *Bacillus subtilis* RNA polymerase. *J Bacteriol* **172**: 5575-5585.

Kanehara, K., Ito, K., and Akiyama, Y. (2002) YaeL (EcfE) activates the sigma(E) pathway of stress response through a site-2 cleavage of anti-sigma(E), RseA. *Genes Dev* **16**: 2147-2155.

Kanemori, M., Yanagi, H., and Yura, T. (1999) Marked instability of the sigma(32) heat shock transcription factor at high temperature. Implications for heat shock regulation. *J Biol Chem* **274**: 22002-22007.

Keiler, K.C., and Sauer, R.T. (1996) Sequence determinants of C-terminal substrate recognition by the Tsp protease. *J Biol Chem* **271**: 2589-2593.

Kenklies, J., Ziehn, R., Fritsche, K., Pich, A., and Andreesen, J.R. (1999) Proline biosynthesis from L-ornithine in *Clostridium sticklandii*: purification of delta1-pyrroline-5-carboxylate reductase, and sequence and expression of the encoding gene, *proC*. *Microbiology* **145**: 819-826.

Kennedy, M.B. (1995) Origin of PDZ (DHR, GLGF) domains. *Trends Biochem Sci* **20**: 350.

Kim, K.I., Park, S.C., Kang, S.H., Cheong, G.W., and Chung, C.H. (1999) Selective degradation of unfolded proteins by the self- compartmentalizing HtrA protease, a periplasmic heat shock protein in *Escherichia coli*. *J Mol Biol* **294**: 1363-1374.

Klein, C., and Entian, K.D. (1994) Genes involved in self-protection against the lantibiotic subtilin produced by *Bacillus subtilis* ATCC 6633. *Appl Environ Microbiol* **60**: 2793-2801.

Kobayashi, K., Ogura, M., Yamaguchi, H., Yoshida, K., Ogasawara, N., Tanaka, T., and Fujita, Y. (2001) Comprehensive DNA microarray analysis of *Bacillus subtilis* two-component regulatory systems. *J Bacteriol* **183**: 7365-7370.

Kolmar, H., Waller, P.R., and Sauer, R.T. (1996) The DegP and DegQ periplasmic endoproteases of *Escherichia coli*: specificity for cleavage sites and substrate conformation. *J Bacteriol* **178**: 5925-5929.

Kontinen, V.P., and Sarvas, M. (1993) The PrsA lipoprotein is essential for protein secretion in *Bacillus subtilis* and sets a limit for high-level secretion. *Mol Microbiol* **8**: 727-737.

Koonin, E., Aravind, L., and Galperin, M. (2000) A Comparative-Genomic View of the Microbial Stress Response. In *Bacterial Stress Responses*. Hengee-Aronis, S.G.a.R. (ed). Washington: ASM Press, pp. 417-444.

Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M., and Clausen, T. (2002) Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature* **416**: 455-459.

Kruger, E., and Hecker, M. (1998) The first gene of the *Bacillus subtilis clpC* operon, *ctsR*, encodes a negative regulator of its own operon and other class III heat shock genes. *J Bacteriol* **180**: 6681-6688.

Kruger, E., Msadek, T., and Hecker, M. (1996) Alternate promoters direct stress-induced transcription of the *Bacillus subtilis clpC* operon. *Mol Microbiol* **20**: 713-723.

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

Kruger, E., Zuhlke, D., Witt, E., Ludwig, H., and Hecker, M. (2001) Clp-mediated proteolysis in Gram-positive bacteria is autoregulated by the stability of a repressor. *Embo J* **20**: 852-863.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B.,

Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Danchin, A., and et al. (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**: 249-256.

Laskowska, E., Kuczynska-Wisnik, D., Skorko-Glonek, J., and Taylor, A. (1996) Degradation by proteases Lon, Clp and HtrA, of *Escherichia coli* proteins aggregated in vivo by heat shock; HtrA protease action in vivo and in vitro. *Mol Microbiol* **22**: 555-571.

Lazar, S.W., and Kolter, R. (1996) SurA assists the folding of *Escherichia coli* outer membrane proteins. *J Bacteriol* **178**: 1770-1773.

Levine, M.M., Galen, J., Barry, E., Noriega, F., Chatfield, S., Sztein, M., Dougan, G., and Tacket, C. (1996) Attenuated Salmonella as live oral vaccines against typhoid fever and as live vectors. *J Biotechnol* **44**: 193-196.

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

Li, S.R., Dorrell, N., Everest, P.H., Dougan, G., and Wren, B.W. (1996) Construction and characterization of a Yersinia enterocolitica O:8 high- temperature requirement (htrA) isogenic mutant. *Infect Immun* **64**: 2088-2094.

Li, W., Srinivasula, S.M., Chai, J., Li, P., Wu, J.W., Zhang, Z., Alnemri, E.S., and Shi, Y. (2002) Structural insights into the pro-apoptotic function of mitochondrial serine protease HtrA2/Omi. *Nat Struct Biol* **9**: 436-441.

Lin, W.J., Huang, S.W., and Chou, C.P. (2001) DegP-coexpression minimizes inclusion-body formation upon overproduction of recombinant penicillin acylase in *Escherichia coli*. *Biotechnol Bioeng* **73**: 484-492.

Lindner, C., Stulke, J., and Hecker, M. (1994) Regulation of xylanolytic enzymes in *Bacillus subtilis*. *Microbiology* **140**: 753-757.

Lipinska, B., Fayet, O., Baird, L., and Georgopoulos, C. (1989) Identification, characterization, and mapping of the *Escherichia coli htrA* gene, whose product is essential for bacterial growth only at elevated temperatures. *J Bacteriol* **171**: 1574-1584.

Lipinska, B., Sharma, S., and Georgopoulos, C. (1988) Sequence analysis and regulation of the *htrA* gene of *Escherichia coli*: a sigma 32-independent mechanism of heat-inducible transcription. *Nucleic Acids Res* **16**: 10053-10067.

Lipinska, B., Zylicz, M., and Georgopoulos, C. (1990) The HtrA (DegP) protein, essential for *Escherichia coli* survival at high temperatures, is an endopeptidase. *J Bacteriol* **172**: 1791-1797.

Lonetto, M.A., Brown, K.L., Rudd, K.E., and Buttner, M.J. (1994) Analysis of the Streptomyces coelicolor *sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *Proc Natl Acad Sci U S A* **91**: 7573-7577.

Loosmore, S.M., Yang, Y.P., Oomen, R., Shortreed, J.M., Coleman, D.C., and Klein, M.H. (1998) The *Haemophilus influenzae* HtrA protein is a protective antigen. *Infect Immun* **66**: 899-906.

Martin, D.W., Schurr, M.J., Mudd, M.H., Govan, J.R., Holloway, B.W., and Deretic, V. (1993) Mechanism of conversion to mucoidy in *Pseudomonas aeruginosa* infecting cystic fibrosis patients. *Proc Natl Acad Sci U S A* **90**: 8377-8381.

Martins, L.M., Iaccarino, I., Tenev, T., Gschmeissner, S., Totty, N.F., Lemoine, N.R., Savopoulos, J., Gray, C.W., Creasy, C.L., Dingwall, C., and Downward, J. (2002) The serine protease Omi/HtrA2 regulates apoptosis by binding XIAP through a reaper-like motif. *J Biol Chem* **277**: 439-444.

Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**: 911-940.

Mathiopoulos, C., Mueller, J.P., Slack, F.J., Murphy, C.G., Patankar, S., Bukusoglu, G., and Sonenshein, A.L. (1991) A *Bacillus subtilis* dipeptide transport system expressed early during sporulation. *Mol Microbiol* **5**: 1903-1913.

Matsuo, E., Sampei, G., Mizobuchi, K., and Ito, K. (1999) The plasmid F OmpP protease, a homologue of OmpT, as a potential obstacle to *E. coli*-based protein production. *FEBS Lett* **461**: 6-8.

McGuffin, L.J., Bryson, K., and Jones, D.T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics* **16**: 404-405.

McLaughlin, J.R., Murray, C.L., and Rabinowitz, J.C. (1981) Unique features in the ribosome binding site sequence of the gram-positive *Staphylococcus aureus* beta-lactamase gene. *J Biol Chem* **256**: 11283-11291.

Mecsas, J., Rouviere, P.E., Erickson, J.W., Donohue, T.J., and Gross, C.A. (1993) The activity of sigma E, an *Escherichia coli* heat-inducible sigma- factor, is modulated by expression of outer membrane proteins. *Genes Dev* **7**: 2618-2628.

Meerman, H.J., and Georgiou, G. (1994) Construction and characterization of a set of *E. coli* strains deficient in all known loci affecting the proteolytic stability of secreted recombinant proteins. *Biotechnology (N Y)* **12**: 1107-1110.

Merchante, R., Pooley, H.M., and Karamata, D. (1995) A periplasm in *Bacillus subtilis*. *J Bacteriol* **177**: 6176-6183.

Miller, J.H. (1972) *Experiments in molecular genetics.* Cold Spring Harbor NY: Cold Spring Harbor Laboratory Press.

Misra, R., CastilloKeller, M., and Deng, M. (2000) Overexpression of protease-deficient DegP(S210A) rescues the lethal phenotype of *Escherichia coli* OmpF assembly mutants in a *degP* background. *J Bacteriol* **182**: 4882-4888.

Missiakas, D., Betton, J.M., and Raina, S. (1996) New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/OmpH. *Mol Microbiol* **21**: 871-884.

Missiakas, D., and Raina, S. (1998) The extracytoplasmic function sigma factors: role and regulation. *Mol Microbiol* **28**: 1059-1066.

Miyoshi, A., Poquet, I., Azevedo, V., Commissaire, J., Bermudez-Humaran, L., Domakova, E., Le Loir, Y., Oliveira, S.C., Gruss, A., and Langella, P. (2002) Controlled production of stable heterologous proteins in *Lactococcus lactis*. *Appl Environ Microbiol* **68**: 3141-3146.

Mogk, A., Homuth, G., Scholz, C., Kim, L., Schmid, F.X., and Schumann, W. (1997) The GroE chaperonin machine is a major modulator of the CIRCE heat shock regulon of *Bacillus subtilis*. *Embo J* **16**: 4579-4590.

Mogk, A., Volker, A., Engelmann, S., Hecker, M., Schumann, W., and Volker, U. (1998) Nonnative proteins induce expression of the *Bacillus subtilis* CIRCE regulon. *J Bacteriol* **180**: 2895-2900.

Morita, M.T., Tanaka, Y., Kodama, T.S., Kyogoku, Y., Yanagi, H., and Yura, T. (1999) Translational induction of heat shock transcription factor sigma32: evidence for a built-in RNA thermosensor. *Genes Dev* **13**: 655-665.

Mourez, M., Skouloubris, S., Betton, J.M., and Dassa, E. (1997) Heat shock induction by a misassembled cytoplasmic membrane protein complex in *Escherichia coli*. *Mol Microbiol* **26**: 821-831.

Msadek, T. (1999) When the going gets tough: survival strategies and environmental signaling networks in *Bacillus subtilis*. *Trends Microbiol* **7**: 201-207.

Msadek, T., Dartois, V., Kunst, F., Herbaud, M.L., Denizot, F., and Rapoport, G. (1998) ClpP of *Bacillus subtilis* is required for competence development, motility, degradative enzyme synthesis, growth at high temperature and sporulation. *Mol Microbiol* **27**: 899-914.

Msadek, T., Kunst, F., and Rapoport, G. (1994) MecB of *Bacillus subtilis*, a member of the ClpC ATPase family, is a pleiotropic regulator controlling competence gene expression and growth at high temperature. *Proc Natl Acad Sci U S A* **91**: 5788-5792.

Nagai, H., Yuzawa, H., Kanemori, M., and Yura, T. (1994) A distinct segment of the sigma 32 polypeptide is involved in DnaK-mediated negative control of the heat shock response in *Escherichia coli*. *Proc Natl Acad Sci U S A* **91**: 10280-10284.

Noone, D., Howell, A., Collery, R., and Devine, K.M. (2001) YkdA and YvtA, HtrA-like serine proteases in *Bacillus subtilis*, engage in negative autoregulation and reciprocal cross-regulation of *ykdA* and *yvtA* gene expression. *J Bacteriol* **183**: 654-663.

Noone, D., Howell, A., and Devine, K.M. (2000) Expression of *ykdA*, encoding a *Bacillus subtilis* homologue of HtrA, is heat shock inducible and negatively autoregulated. *J Bacteriol* **182**: 1592-1599.

Ohnishi, R., Ishikawa, S., and Sekiguchi, J. (1999) Peptidoglycan hydrolase LytF plays a role in cell separation with CwlF during vegetative growth of *Bacillus subtilis*. *J Bacteriol* **181**: 3178-3184.

Oke, V., Shchepetov, M., and Cutting, S. (1997) SpoIVB has two distinct functions during spore formation in *Bacillus subtilis*. *Mol Microbiol* **23**: 223-230.

Page, R.D. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**: 357-358.

Pallen, M.J., and Ponting, C.P. (1997) PDZ domains in bacterial proteins. *Mol Microbiol* **26**: 411-413.

Pallen, M.J., and Wren, B.W. (1997) The HtrA family of serine proteases. *Mol Microbiol* **26**: 209-221.

Pedersen, L.L., Radulic, M., Doric, M., and Abu Kwaik, Y. (2001) HtrA homologue of *Legionella pneumophila*: an indispensable element for intracellular infection of mammalian but not protozoan cells. *Infect Immun* **69**: 2569-2579.

Perego, M. (1993) Integrational Vectors for Genetic Manipulation in *Bacillus subtilis*. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*. Sonenshein, A.L., Hoch, J.A. and Losick, R. (eds). Washington: American Society for Microbiology.

Petersohn, A., Brigulla, M., Haas, S., Hoheisel, J.D., Volker, U., and Hecker, M. (2001) Global analysis of the general stress response of *Bacillus subtilis*. *J Bacteriol* **183**: 5617-5631.

Petersohn, A., Engelmann, S., Setlow, P., and Hecker, M. (1999) The katX gene of *Bacillus subtilis* is under dual control of sigmaB and sigmaF. *Mol Gen Genet* **262**: 173-179.

Phillips, R.W., Elzer, P.H., and Roop, R.M., II (1995) A *Brucella melitensis* high temperature requirement A (*htrA*) deletion mutant demonstrates a stress response defective phenotype in vitro and transient attenuation in the BALB/c mouse model. *Microb Pathog* **19**: 227-284.

Phillips, R.W., and Roop, R.M., 2nd (2001) Brucella abortus HtrA functions as an authentic stress response protease but is not required for wild-type virulence in BALB/c mice. *Infect Immun* **69**: 5911-5913.

Pogliano, J., Lynch, A.S., Belin, D., Lin, E.C., and Beckwith, J. (1997) Regulation of *Escherichia coli* cell envelope proteins involved in protein folding and degradation by the Cpx two-component system. *Genes Dev* **11**: 1169-1182.

Ponting, C.P. (1997) Evidence for PDZ domains in bacteria, yeast, and plants. *Protein Sci* **6**: 464-468.

Ponting, C.P., Phillips, C., Davies, K.E., and Blake, D.J. (1997) PDZ domains: targeting signalling molecules to sub-membranous sites. *Bioessays* **19**: 469-479.

Poquet, I., Saint, V., Seznec, E., Simoes, N., Bolotin, A., and Gruss, A. (2000) HtrA is the unique surface housekeeping protease in *Lactococcus lactis* and is required for natural protein processing. *Mol Microbiol* **35**: 1042-1051.

Porankiewicz, J., Wang, J., and Clarke, A.K. (1999) New insights into the ATP-dependent Clp protease: *Escherichia coli* and beyond. *Mol Microbiol* **32**: 449-458.

Pragai, Z., and Harwood, C.R. (2000) YsxC, a putative GTP-binding protein essential for growth of *Bacillus subtilis* 168. *J Bacteriol* **182**: 6819-6823.

Price, C.W. (2002) General Stress Response. In *Bacillus subtilis and Its Closest Relatives*. Sonenshein, A.L., Hoch, J.A. and Losick, R. (eds). Washington: ASM Press, pp. 369-384.

Price, C.W., Fawcett, P., Ceremonie, H., Su, N., Murphy, C.K., and Youngman, P. (2001) Genome-wide analysis of the general stress response in *Bacillus subtilis*. *Mol Microbiol* **41**: 757-774.

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

Raivio, T.L., Laird, M.W., Joly, J.C., and Silhavy, T.J. (2000) Tethering of CpxP to the inner membrane prevents spheroplast induction of the cpx envelope stress response. *Mol Microbiol* **37**: 1186-1197.

Raivio, T.L., Popkin, D.L., and Silhavy, T.J. (1999) The Cpx envelope stress response is controlled by amplification and feedback inhibition. *J Bacteriol* **181**: 5263-5272.

Raivio, T.L., and Silhavy, T.J. (1997) Transduction of envelope stress in *Escherichia coli* by the Cpx two- component system. *J Bacteriol* **179**: 7724-7733.

Raivio, T.L., and Silhavy, T.J. (1999) The sigmaE and Cpx regulatory pathways: overlapping but distinct envelope stress responses. *Curr Opin Microbiol* **2**: 159-165.

Raivio, T.L., and Silhavy, T.J. (2000) Sensing and Responding to Envelope Stress. In *Bacterial Stress Responses*. Storz, G. and Hengge-Aronis, R. (eds). Washington: ASM Press, pp. 19-32.

Raivio, T.L., and Silhavy, T.J. (2001) Periplasmic stress and ECF sigma factors. *Annu Rev Microbiol* **55**: 591-624.

Ratnayake-Lecamwasam, M., Serror, P., Wong, K.W., and Sonenshein, A.L. (2001) *Bacillus subtilis* CodY represses early-stationary-phase genes by sensing GTP levels. *Genes Dev* **15**: 1093-1103.

Resto-Ruiz, S.I., Sweger, D., Widen, R.H., Valkov, N., and Anderson, B.E. (2000) Transcriptional activation of the htrA (High-temperature requirement A) gene from *Bartonella henselae*. *Infect Immun* **68**: 5970-5978.

Rietsch, A., and Beckwith, J. (1998) The genetics of disulfide bond metabolism. *Annu Rev Genet* **32**: 163-184.

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

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual.* Cold Spring Habor, N.Y.: Cold Spring Habor Laboratory Press.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**: 5463-5467.

Sassoon, N., Arie, J.P., and Betton, J.M. (1999) PDZ domains determine the native oligomeric structure of the DegP (HtrA) protease. *Mol Microbiol* **33**: 583-589.

Scanlan, E., O'Reilly, M., Krogh Andersen, K., Noone, D., and Devine, K.M. (2001) Screening for Increased Sensitivity or Resistance and Level of LacZ Reporter Gene

Expression in Response to the Oxidising Agents Hydrogen Peroxide and Methyl Viologen. In *Functional Analysis of Bacterial Genes: A Practical Manual*. Schumann, W., Ehrlich, S.D. and Ogasawara, N. (eds). Chichester UK.: Wiley, pp. 227-230.

Schaeffer, P., Miller, J., and Aubert, J. (1965) Catabolite repression of bacterial repression. *Proc. Natl. Acad. Sci. USA* **54**: 701-711.

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

Schrogel, O., and Allmansberger, R. (1997) Optimisation of the BgaB reporter system: determination of transcriptional regulation of stress responsive genes in *Bacillus subtilis*. *FEMS Microbiol Lett* **153**: 237-243.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95**: 5857-5864.

Schulz, A., and Schumann, W. (1996) *hrcA*, the first gene of the *Bacillus subtilis dnaK* operon encodes a negative regulator of class I heat shock genes. *J Bacteriol* **178**: 1088-1093.

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

Schumann, W., Ehrlich, S.D., and Ogasawara, N., (eds) (2001) *Functional Analysis of Bacterial Genes: A Practical Manual.* Chichester, UK.: Wiley.

Schumann, W., Hecker, M., and Msadek, T. (2002) Regulation and Function of Heat-Inducible Genes in *Bacillus subtilis*. In *Bacillus subtilis and Its Closest Relatives*. Sonenshein, A.L., Hoch, J.A. and Losick, R. (eds). Washington: ASM Press, pp. 359-368.

Schurr, M.J., Yu, H., Martinez-Salazar, J.M., Boucher, J.C., and Deretic, V. (1996) Control of AlgU, a member of the sigma E-like family of stress sigma factors, by the negative regulators MucA and MucB and *Pseudomonas aeruginosa* conversion to mucoidy in cystic fibrosis. *J Bacteriol* **178**: 4997-5004.

Scott, J.M., and Haldenwang, W.G. (1999) Obg, an essential GTP binding protein of *Bacillus subtilis*, is necessary for stress activation of transcription factor sigma(B). *J Bacteriol* **181**: 4653-4660.

Scott, J.M., Ju, J., Mitchell, T., and Haldenwang, W.G. (2000) The *Bacillus subtilis* GTP binding protein obg and regulators of the sigma(B) stress response transcription factor cofractionate with ribosomes. *J Bacteriol* **182**: 2771-2777.

Scott, J.M., Smirnova, N., and Haldenwang, W.G. (1999) A *Bacillus*-specific factor is needed to trigger the stress-activated phosphatase/kinase cascade of sigmaB induction. *Biochem Biophys Res Commun* **257**: 106-110.

Sebert, M.E., Palmer, L.M., Rosenberg, M., and Weiser, J.N. (2002) Microarray-based identification of *htrA*, a *Streptococcus pneumoniae* gene that is regulated by the CiaRH two-component system and contributes to nasopharyngeal colonization. *Infect Immun* **70**: 4059-4067.

Shotland, Y., Koby, S., Teff, D., Mansur, N., Oren, D.A., Tatematsu, K., Tomoyasu, T., Kessel, M., Bukau, B., Ogura, T., and Oppenheim, A.B. (1997) Proteolysis of the phage lambda CII regulatory protein by FtsH (HflB) of *Escherichia coli*. *Mol Microbiol* **24**: 1303-1310.

Skorko-Glonek, J., Lipinska, B., Krzewski, K., Zolese, G., Bertoli, E., and Tanfani, F. (1997) HtrA heat shock protease interacts with phospholipid membranes and undergoes conformational changes. *J Biol Chem* **272**: 8974-8982.

Skorko-Glonek, J., Wawrzynow, A., Krzewski, K., Kurpierz, K., and Lipinska, B. (1995) Site-directed mutagenesis of the HtrA (DegP) serine protease, whose proteolytic activity is indispensable for *Escherichia coli* survival at elevated temperatures. *Gene* **163**: 47-52.

Skorko-Glonek, J., Zurawa, D., Kuczwara, E., Wozniak, M., Wypych, Z., and Lipinska, B. (1999) The *Escherichia coli* heat shock protease HtrA participates in defense against oxidative stress. *Mol Gen Genet* **262**: 342-350.

Slack, F.J., Mueller, J.P., Strauch, M.A., Mathiopoulos, C., and Sonenshein, A.L. (1991) Transcriptional regulation of a *Bacillus subtilis* dipeptide transport operon. *Mol Microbiol* **5**: 1915-1925.

Smeds, A., Varmanen, P., and Palva, A. (1998) Molecular characterization of a stress-inducible gene from *Lactobacillus helveticus*. *J Bacteriol* **180**: 6148-6153.

Smith, C.K., Baker, T.A., and Sauer, R.T. (1999) Lon and Clp family proteases and chaperones share homologous substrate- recognition domains. *Proc Natl Acad Sci U S A* **96**: 6678-6682.

Smith, T.J., Blackman, S.A., and Foster, S.J. (2000) Autolysins of *Bacillus subtilis*: multiple enzymes with multiple functions. *Microbiology* **146**: 249-262.

Snyder, W.B., Davis, L.J., Danese, P.N., Cosma, C.L., and Silhavy, T.J. (1995) Overproduction of NlpE, a new outer membrane lipoprotein, suppresses the toxicity of periplasmic LacZ by activation of the Cpx signal transduction pathway. *J Bacteriol* **177**: 4216-4223.

Sonenshein, A.L., Hoch, J.A., and Losick, R., (eds) (1993) *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics.* Washington: American Society for Microbiology.

Sonenshein, A.L., Hoch, J.A., and Losick, R., (eds) (2002) *Bacillus subtilis and Its Closest Relatives.* Washington: ASM Press.

Spiess, C., Beil, A., and Ehrmann, M. (1999) A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* **97**: 339-347.

Staden, R. (1986) The current status and portability of our sequence handling software. *Nucleic Acids Res* **14**: 217-231.

Staden, R., Beal, K.F., and Bonfield, J.K. (2000) The Staden package, 1998. *Methods Mol Biol* **132**: 115-130.

Stephenson, K., Carter, N.M., Harwood, C.R., Petit-Glatron, M.F., and Chambert, R. (1998) The influence of protein folding on late stages of the secretion of alpha-amylases from *Bacillus subtilis*. *FEBS Lett* **430**: 385-389.

Stephenson, K., and Harwood, C.R. (1998) Influence of a cell-wall-associated protease on production of alpha-amylase by *Bacillus subtilis*. *Appl Environ Microbiol* **64**: 2875-2881.

Storz, G. (1999) An RNA thermometer. *Genes Dev* **13**: 633-636.

Stragier, P., and Losick, R. (1996) Molecular genetics of sporulation in *Bacillus subtilis*. *Annu Rev Genet* **30**: 297-241.

Strauch, K.L., and Beckwith, J. (1988) An *Escherichia coli* mutation preventing degradation of abnormal periplasmic proteins. *Proc Natl Acad Sci U S A* **85**: 1576-1580.

Strauch, K.L., Johnson, K., and Beckwith, J. (1989) Characterization of *degP*, a gene required for proteolysis in the cell envelope and essential for growth of *Escherichia coli* at high temperature. *J Bacteriol* **171**: 2689-2696.

Straus, D., Walter, W., and Gross, C.A. (1990) DnaK, DnaJ, and GrpE heat shock proteins negatively regulate heat shock gene expression by controlling the synthesis and stability of sigma 32. *Genes Dev* **4**: 2202-2209.

Suzuki, Y., Imai, Y., Nakayama, H., Takahashi, K., Takio, K., and Takahashi, R. (2001) A serine protease, HtrA2, is released from the mitochondria and interacts with XIAP, inducing cell death. *Mol Cell* **8**: 613-621.

Tatsuta, T., Tomoyasu, T., Bukau, B., Kitagawa, M., Mori, H., Karata, K., and Ogura, T. (1998) Heat shock regulation in the *ftsH* null mutant of *Escherichia coli*: dissection of stability and activity control mechanisms of sigma32 in vivo. *Mol Microbiol* **30**: 583-593.

Thwaite, J.E., Baillie, L.W., Carter, N.M., Stephenson, K., Rees, M., Harwood, C.R., and Emmerson, P.T. (2002) Optimization of the cell wall microenvironment allows increased production of recombinant *Bacillus anthracis* protective antigen from B. subtilis. *Appl Environ Microbiol* **68**: 227-234.

Tjalsma, H., Bolhuis, A., Jongbloed, J.D., Bron, S., and van Dijl, J.M. (2000) Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol Mol Biol Rev* **64**: 515-547.

Tjalsma, H., Noback, M.A., Bron, S., Venema, G., Yamane, K., and van Dijl, J.M. (1997) *Bacillus subtilis* contains four closely related type I signal peptidases with overlapping substrate specificities. Constitutive and temporally controlled expression of different sip genes. *J Biol Chem* **272**: 25983-25992.

Tomoyasu, T., Mogk, A., Langen, H., Goloubinoff, P., and Bukau, B. (2001) Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the *Escherichia coli* cytosol. *Mol Microbiol* **40**: 397-413.

Tomoyasu, T., Ogura, T., Tatsuta, T., and Bukau, B. (1998) Levels of DnaK and DnaJ provide tight control of heat shock gene expression and protein repair in *Escherichia coli*. *Mol Microbiol* **30**: 567-581.

Vagner, V., Dervyn, E., and Ehrlich, S.D. (1998) A vector for systematic gene inactivation in *Bacillus subtilis*. *Microbiology* **144**: 3097-3104.

Van Melderen, L., and Gottesman, S. (1999) Substrate sequestration by a proteolytically inactive Lon mutant. *Proc Natl Acad Sci U S A* **96**: 6064-6071.

van Wely, K.H., Swaving, J., Freudl, R., and Driessen, A.J. (2001) Translocation of proteins across the cell envelope of Gram-positive bacteria. *FEMS Microbiol Rev* **25**: 437-454.

Vandeputte-Rutten, L., Kramer, R.A., Kroon, J., Dekker, N., Egmond, M.R., and Gros, P. (2001) Crystal structure of the outer membrane protease OmpT from *Escherichia coli* suggests a novel catalytic site. *Embo J* **20**: 5033-5039.

Verhagen, A.M., Silke, J., Ekert, P.G., Pakusch, M., Kaufmann, H., Connolly, L.M., Day, C.L., Tikoo, A., Burke, R., Wrobel, C., Moritz, R.L., Simpson, R.J., and Vaux, D.L. (2002) HtrA2 promotes cell death through its serine protease activity and its ability to antagonize inhibitor of apoptosis proteins. *J Biol Chem* **277**: 445-454.

Vieira, J., and Messing, J. (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**: 259-268.

Vitikainen, M., Pummi, T., Airaksinen, U., Wahlstrom, E., Wu, H., Sarvas, M., and Kontinen, V.P. (2001) Quantitation of the capacity of the secretion apparatus and requirement for PrsA in growth and secretion of alpha-amylase in *Bacillus subtilis*. *J Bacteriol* **183**: 1881-1890.

Volker, U., Andersen, K.K., Antelmann, H., Devine, K.M., and Hecker, M. (1998) One of two osmC homologs in *Bacillus subtilis* is part of the sigmaB- dependent general stress regulon. *J Bacteriol* **180**: 4212-4218.

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

Volker, U., Maul, B., and Hecker, M. (1999) Expression of the sigmaB-dependent general stress regulon confers multiple stress resistance in *Bacillus subtilis*. *J Bacteriol* **181**: 3942-3948.

Vollmer, M.D., and Schlomann, M. (1995) Conversion of 2-chloro-*cis*,*cis*-muconate and its metabolites 2-chloro- and 5-chloromuconolactone by chloromuconate cycloisomerases of pJP4 and pAC27. *J Bacteriol* **177**: 2938-2941.

Wakeley, P.R., Dorazi, R., Hoa, N.T., Bowyer, J.R., and Cutting, S.M. (2000) Proteolysis of SpolVB is a critical determinant in signalling of Pro-sigmaK processing in *Bacillus subtilis*. *Mol Microbiol* **36**: 1336-1348.

Waller, P.R., and Sauer, R.T. (1996) Characterization of *degQ* and *degS*, *Escherichia coli* genes encoding homologs of the DegP protease. *J Bacteriol* **178**: 1146-1153.

Wang, Q.P., and Kaguni, J.M. (1989) A novel sigma factor is involved in expression of the *rpoH* gene of *Escherichia coli*. *J Bacteriol* **171**: 4248-4253.

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

Wickner, S., and Maurizi, M.R. (1999) Here's the hook: similar substrate binding sites in the chaperone domains of Clp and Lon. *Proc Natl Acad Sci U S A* **96**: 8318-8320.

Williams, K., Oyston, P.C., Dorrell, N., Li, S., Titball, R.W., and Wren, B.W. (2000) Investigation into the role of the serine protease HtrA in *Yersinia pestis* pathogenesis. *FEMS Microbiol Lett* **186**: 281-286.

Womble, D.D. (2000) GCG: The Wisconsin Package of sequence analysis programs. *Methods Mol Biol* **132**: 3-22.

Wu, Q.L., Kong, D., Lam, K., and Husson, R.N. (1997) A mycobacterial extracytoplasmic function sigma factor involved in survival following stress. *J Bacteriol* **179**: 2922-2929.

Yamamoto, T., Hanawa, T., Ogata, S., and Kamiya, S. (1996) Identification and characterization of the *Yersinia enterocolitica gsrA* gene, which protectively responds to intracellular stress induced by macrophage phagocytosis and to extracellular environmental stress. *Infect Immun* **64**: 2980-2987.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103-119.

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

Ye, R., Kim, J.H., Kim, B.G., Szarka, S., Sihota, E., and Wong, S.L. (1999) High-level secretory production of intact, biologically active staphylokinase from *Bacillus subtilis*. *Biotechnol Bioeng* **62**: 87-96.

Yuan, G., and Wong, S.L. (1995a) Isolation and characterization of *Bacillus subtilis groE* regulatory mutants: evidence for *orf39* in the *dnaK* operon as a repressor gene in regulating the expression of both *groE* and *dnaK*. *J Bacteriol* **177**: 6462-6468.

Yuan, G., and Wong, S.L. (1995b) Regulation of *groE* expression in *Bacillus subtilis*: the involvement of the sigma A-like promoter and the roles of the inverted repeat sequence (CIRCE). *J Bacteriol* **177**: 5427-5433.

Yura, T., Kanemori, T., and Morita, M. (2000) The Heat Shock Response: Regulation and Function. In *Bacterial Stress Responses*. Hengee-Aronis, S.G.a.R. (ed). Washington: ASM Press, pp. 3-18.

Yura, T., Nagai, H., and Mori, H. (1993) Regulation of the heat-shock response in bacteria. *Annu Rev Microbiol* **47**: 321-350.

Zhang, S., Scott, J.M., and Haldenwang, W.G. (2001) Loss of ribosomal protein L11 blocks stress activation of the *Bacillus subtilis* transcription factor sigma(B). *J Bacteriol* **183**: 2316-2321.

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

Zumbrunn, J., and Trueb, B. (1996) Primary structure of a putative serine protease specific for IGF- binding proteins. *FEBS Lett* **398**: 187-192.