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.

# An investigation of LytE in fulfilling the D,L-endopeptidase requirement for viability of *Bacillus subtilis*

by Laura Uelze

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

Supervised by Professor Kevin M. Devine

Smurfit Institute of Genetics

Trinity College Dublin

September 2016

TRINITY COLLEGE DUBLIN
18 JAN 2017
ACADEMIC REGISTRY

PhD
GENETICS
+ MICROBIOLOGY

TRINITY LIBRARY
0 8 MAR 2017
DUBLIN

Thesis 11166

## Statutory Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

# Acknowledgements

I would like to express my gratitude to my supervisor, Prof. Kevin Devine, whose lifelong expertise and insight into the world of *Bacillus subtilis* enabled this research project. Without your ideas and inspiration I would have long gotten lost in the scientific mist that lays between hypothesis and result.

I would also like to thank the other members of the lab, past and present: Dr. Eric Botella and Dr. Letal Salzberg, as well as Dr. Susanne Devine for the assistance and help they provided at all levels of my research project.

A very special thanks to Dr. David Noone who taught me the handicraft of bacterial genetics. How much my daily questions, regarding methods and instruments, slowed down your own research I do not know, but you never let me down. Thanks for being such a great tutor!

Thanks also to Michael Prunty, my good-natured guide to the Irish, the English language and the microscope. I wish you all the best for your own PhD and everything that follows. I very much hope we stay in touch.

I also want to thank Jennifer Lennon who completed her undergraduate project in the lab. I hope I have been a passable tutor to you and I am very grateful for the contribution which you have made to this project.

Appreciation also goes to the technical and administrative staff: Ms. Sue Holahan, Mr. David Sullivan, Ms. Brenda Campbell and Dr. Paul McDermott for all the instances in which their assistance helped me along the way.

A big thank you to Johannes Weihermüller, for equipping me with fantastic software and for priceless IT support in times of crisis and need.

Thanks also to Dr. Karsten Hokamp and Darren Fitzpatrick who not only provided me with statistical advice, but also sparked my interest in Python and R programming.

Similarly thanks also to Dr. Aoife McLysaght for her help in unravelling the evolutionary relationships between the LysM motifs.

I am grateful too, to Prof. Tony Kavanagh and Dr. Matthew Campbell who took time out of their busy schedules to oversee and assess my progress in the PhD programme.

A very special thanks to Prof. Thomas Wiegert, who introduced me to the world of microbial genetics and who opened a hidden door for me that led to Dublin.

Finally, thanks to everybody who supported me on my way, especially Dr. Detlef Thieme and the amazing team of the Institute of Doping Analysis and Sports Biochemistry Dresden, whom I hold in the highest esteem.

In conclusion, I recognize that this research would not have been possible without the financial assistance of the Science Foundation Ireland, the Trinity Postgraduate Scholarship and the ERASMUS programme and I am grateful for the funding provided.

# Contents

**List of Tables** ..... I

**List of Figures** ..... IV

**List of Abbreviations** ..... X

**Publications** ..... XIII

**Abstract** ..... XIV

**1 Chapter 1: Introduction** ..... 1

**1.1 *Bacillus subtilis*** ..... 1

1.1.1 *B. subtilis* as a model system for Gram-positive bacteria ..... 1
1.1.2 Genetics and genomics of *B. subtilis* ..... 7

**1.2 The cell envelope** ..... 7

1.2.1 Gram-positive cell wall structure and composition ..... 8
1.2.2 Cell wall metabolism ..... 9
1.2.3 Cell wall function ..... 13

**1.3 Two-component signal transduction systems in bacteria** ..... 13

1.3.1 Typical two-component system ..... 13
1.3.2 Two-component systems in *B. subtilis* ..... 17
1.3.3 The WalRK two-component system ..... 18

**1.4 Autolysins** ..... 22

1.4.1 Autolysins encoded by *B. subtilis* ..... 22
1.4.2 D,L-endopeptidases of *B. subtilis* ..... 25
1.4.2.1 CwlS ..... 25
1.4.2.2 CwlO ..... 27
1.4.2.3 LytE ..... 28
1.4.2.4 LytF ..... 28
1.4.2.5 PgdS ..... 29
1.4.2.6 YkfC ..... 29
1.4.2.7 YqgT ..... 30
1.4.3 Essentiality and synthetic lethality of autolysins ..... 30

**1.5 The LysM motif** ..... 31

**1.6 Scope of this study ..... 35**

**2 Chapter 2: Material and Methods ..... 37**

**2.1 Genetic modifications ..... 37**

2.1.1 Sequences and sequencing ..... 37
2.1.2 DNA manipulations ..... 37
2.1.3 Plasmid constructions ..... 40
2.1.3.1 Plasmid for HA-epitope tagging of LytE (with kanamycin resistance gene) ..... 40
2.1.3.2 Plasmids for HA-epitope tagging of LytF, CwlS and LytF ..... 40
2.1.3.3 Parental vector for the integration of a variant *lytE* gene at the *cwlO* locus ..... 41
2.1.3.4 Plasmids for the generation of variant LytE with two LysM motifs ..... 41
2.1.3.5 Plasmids for the generation of variant LytE with a single LysM motif ..... 43
2.1.3.6 Plasmids for the generation of variant LytE with three, four and five LysM motifs ..... 44
2.1.3.7 Plasmid for the generation of a variant LytE without LysM motif ..... 46
2.1.3.8 Plasmid for HA-epitope tagging of LytE (with spectinomycin resistance gene) ..... 46
2.1.3.9 Plasmids for the overexpression of variant LytE proteins in *E. coli* ..... 46
2.1.3.10 Plasmids for the generation of heterologous variant LytE with a single LysM motifs of a heterologous enzyme ..... 47
2.1.4 Strain constructions ..... 48
2.1.4.1 *lytE* null mutant strains ..... 48
2.1.4.2 Strains expressing a LytE protein from the *cwlO* promoter ..... 49
2.1.4.3 Strains expressing a HA-tagged LytE protein ..... 50
2.1.4.4 Markerless null mutant strains in *cwlS*, *lytF* and *yqgT* ..... 50
2.1.4.5 Double null mutant strains between *cwlS*, *lytF*, *yqgT* and *lytE, cwlO* ..... 51
2.1.4.6 Strains expressing a CwlS, LytF or YqgT protein from the *cwlO* promoter ..... 51
2.1.4.7 Strains expressing a HA-tagged LytE, CwlS, LytF or YqgT protein ..... 52
2.1.4.8 Strains containing a variant LytE expressed from the *cwlO* promoter ..... 53
2.1.4.9 Strains containing a variant LytE expressed from the *lytE* promoter (with intact *cwlO* gene) ..... 53
2.1.4.10 Strains containing a variant LytE expressed from the *lytE* promoter (with disrupted *cwlO* gene) ..... 54
2.1.4.11 Strains containing a HA-tagged variant LytE ..... 54
2.1.4.12 Strains containing a heterologous variant LytE expressed from the *cwlO* promoter ..... 54
2.1.4.13 Strains containing a HA-tagged heterologous variant LytE expressed from the *cwlO* promoter ..... 55

**2.2 Experimental procedures ..... 55**

2.2.1 Media and growth conditions ..... 55
2.2.2 Growth assays ..... 55
2.2.3 RNA extraction ..... 56
2.2.4 Northern analysis ..... 56
2.2.5 Quantitative Reverse Transcription PCR (qRT-PCR) ..... 56
2.2.6 Western analysis ..... 57
2.2.7 Microscopy ..... 58
2.2.8 Immunofluorescence microscopy ..... 58
2.2.9 Suppressor mutation test ..... 59
2.2.10 Antibiotic assay with Etest® stripes ..... 59

2.2.11 Autolysis assay ..... 59
2.2.12 Protein overexpression and purification ..... 60
2.2.13 Coomassie Blue staining ..... 60
2.2.14 Zymography ..... 60
2.2.15 D,L-endopeptidase enzymatic activity assay ..... 61
2.2.16 Insoluble peptidoglycan pull down assay ..... 61
2.2.17 Protein structure prediction ..... 62
2.2.18 Bioinformatics and computational biology ..... 62

**3 Chapter 3: Exploring the LytE / CwlO D,L-endopeptidase requirement for cell viability ..... 63**

**3.1 Objective ..... 63**

**3.2 Experimental Approach ..... 64**

**3.3 Results ..... 66**

3.3.1 Wild-type LytE expression in strains that are deleted for CwlO ..... 66
3.3.2 Expression of LytE from the *cwlO* promoter ..... 70
3.3.2.1 *lytE* transcript levels ..... 70
3.3.2.2 The level of cell wall stress in strains LU21 and LU25 that express *lytE* using the *cwlO* promoter ..... 72
3.3.2.3 Destabilizing effect of the *cwlO* leader sequence on the *lytE* transcript ..... 73
3.3.2.4 Determination of the cellular level of LytE protein when expressed using the *cwlO* promoter ..... 75
3.3.2.5 Localization of the LytE protein in cells that express it from the *cwlO* promoter ..... 77
3.3.2.6 Growth and cell morphology ..... 79

**3.4 Conclusion ..... 81**

**4 Chapter 4: The Potential of CwlS, LytF and YqgT to fulfil the D,L-endopeptidase requirement ..... 83**

**4.1 Objective ..... 83**

**4.2 Experimental Approach ..... 83**

**4.3 Results ..... 85**

4.3.1 The LytF, CwlS and YqgT D,L-endopeptidases individually are not synthetically lethal with either CwlO or LytE ..... 85
4.3.2 Phenotypes of strains carrying null mutations in *cwlS, lytF* and *yqgT* singly and in combination with either *cwlO* or *lytE* ..... 85
4.3.3 An investigation of strains expressing CwlS from the *cwlO* promoter ..... 87
4.3.3.1 Ability of CwlS to fulfil the D,L-endopeptidase requirement for viability in *B. subtilis* ..... 87
4.3.3.2 Establishment of CwlS protein levels in cells expressing it from the *cwlO* promoter ..... 88
4.3.3.3 Growth and morphology of strains LU106 and LU123 that express CwlS using the *cwlO* promoter ..... 90
4.3.3.4 Localization of the CwlS protein in cells of strain LU159 ..... 92
4.3.3.5 Conclusion ..... 92
4.3.4 An investigation of strains expressing LytF from the *cwlO* promoter ..... 92

4.3.4.1 Ability of LytF to fulfil the D,L-endopeptidase requirement for viability in *B. subtilis* ..... 92
4.3.4.2 Establishment of *lytF* transcript levels in cells expressing it from the *cwlO* promoter ..... 93
4.3.4.3 Establishment of LytF protein levels in cells expressing it from the *cwlO* promoter ..... 95
4.3.4.4 Growth and morphology of strains LU111, LU116, LU125 and LU134 that express LytF using the *cwlO* promoter ..... 98
4.3.4.5 Localization of the LytF protein in cells of strain LU151 ..... 101
4.3.4.6 Conclusion ..... 101
4.3.5 An investigation of strains expressing YqgT from the *cwlO* promoter ..... 101
4.3.5.1 Ability of YqgT to fulfil the D,L-endopeptidase requirement for viability in *B. subtilis* ..... 101
4.3.5.2 Establishment of YqgT protein levels in cells expressing it from the *cwlO* promoter ..... 102
4.3.5.3 Growth and morphology of strain LU112 that expresses YqgT using the *cwlO* promoter ..... 104
4.3.5.4 Localization of the YqgT protein in cells of strain LU163 ..... 105
4.3.5.5 Conclusion ..... 105
4.3.6 Endogenous autolysin activity ..... 105

**4.4 Conclusion ..... 108**

**5 Chapter 5: Exploring the functionality of the LysM motifs of LytE *in vivo* and *in vitro* ..... 111**

**5.1 Objective ..... 111**

**5.2 LytE structure and nomenclature ..... 112**

**5.3 Experimental Approach ..... 113**

5.3.1 Expression of variant LytE from the *cwlO* promoter ..... 113
5.3.2 Expression of heterologous variant LytE from the *cwlO* promoter ..... 120
5.3.3 Expression of variant LytE from the *lytE* promoter ..... 123
5.3.4 Purification of variant LytE proteins ..... 125

**5.4 Results ..... 128**

5.4.1 Ability of variant LytE to fulfil the D,L-endopeptidase requirement when expressed from the *cwlO* promoter ..... 128
5.4.2 Ability of heterologous variant LytE to fulfil the D,L-endopeptidase requirement when expressed from the *cwlO* promoter ..... 129
5.4.3 Ability of variant LytE to fulfil the D,L-endopeptidase requirement when expressed from the *lytE* promoter ..... 130
5.4.4 Verification of the successful disruption of the *cwlO* gene in strains that express a variant LytE from the *cwlO* promoter ..... 131
5.4.5 Verification of the absence of suppressor mutations in strains that express a variant LytE from the *cwlO* promoter ..... 133
5.4.6 Cell-associated protein levels of variant LytE proteins *in vivo* ..... 136
5.4.6.1 Effect of different promoter activities ($P_{cwlO}$ vs $P_{lytE}$) on variant LytE cell-associated protein levels ..... 137
5.4.6.2 Effect of the type of variant LytE on cell-associated protein levels ..... 139
5.4.6.3 Comparison between cell-associated and secreted variant LytE protein levels ..... 141
5.4.6.4 Comparison between cell-associated and secreted heterologous variant LytE protein levels ..... 144

5.4.7 Assessment of the extent to which variant LytE proteins fulfil the D,L-endopeptidase requirement by measuring induction of the SigI regulon ..... 147
5.4.8 Growth behaviour of strains containing a variant LytE under normal and stress conditions ..... 150
5.4.9 Colony Morphology of strains containing a variant LytE ..... 152
5.4.10 Cell Morphology of strains containing a variant LytE ..... 155
5.4.11 Sensitivity to salt stress of strains containing a variant LytE ..... 157
5.4.11.1 Growth behaviour of salt-stressed cells containing a variant LytE ..... 157
5.4.11.2 Cell morphology of salt-stressed cells containing a variant LytE ..... 163
5.4.12 Localization of variant LytE proteins under normal conditions and in salt stressed cells ..... 165
5.4.13 Antibiotic sensitivity of strains containing a variant LytE ..... 167
5.4.14 Characteristics of variant LytE proteins *in vitro* ..... 169
5.4.14.1 Predicted protein structures of variant LytE proteins ..... 169
5.4.14.2 Enzymatic activity of variant LytE proteins ..... 169
5.4.14.3 Binding affinity of variant LytE proteins to purified cell walls *in vitro* ..... 174
5.4.14.4 The effect of anionic polymers on binding affinity of variant LytE proteins to purified cell walls *in vitro* ..... 176

**5.5 Conclusion ..... 179**

**6 Chapter 6: Global analysis of LysM motifs and the evolutionary development of LytE ..... 185**

**6.1 Observations ..... 185**

**6.2 The LysM motifs of *B. subtilis 168* ..... 186**

**6.3 Global analysis of ≈ 30,000 LysM motifs ..... 189**

6.3.1 Objective ..... 189
6.3.2 Approach ..... 190
6.3.3 Results and Discussion ..... 191
6.3.3.1 Frequency distribution of LysM motifs per strain and the number of LysM motifs within proteins ..... 191
6.3.3.2 Function of LysM motif proteins ..... 193
6.3.3.3 Comparison of the frequency of LysM motif proteins among different bacterial families ..... 195
6.3.3.4 Duplicated LysM motifs ..... 201

**6.4 An evolutionary model for the development of CwlS, LytE and LytF ..... 204**

6.4.1 Objective ..... 204
6.4.2 Approach ..... 205
6.4.3 Results ..... 207
6.4.3.1 Identification of orthologs to CwlS, LytE and LytF by their catalytic domain ..... 207
6.4.3.2 The LysM motifs complement of CwlS, LytE and LytF orthologs ..... 209
6.4.3.3 Cluster analysis of individual LysM motifs of all CwlS, LytE and LytF orthologs ..... 212
6.4.4 Discussion ..... 215

**7 Summary ..... 223**

**8 Appendix ..... 231**

**8.1 Figures ..... 231**

**8.2 Sequences ..... 238**

8.2.1 Nucleotide and amino acid sequence of LytE ..... 238
8.2.2 Amino acid sequences of overexpressed and purified variant LytE proteins with 6xHis-tag ..... 239

**8.3 Strains used in this study ..... 241**

**8.4 Plasmids used in this study ..... 247**

**8.5 Primers and oligonucleotides used in this study ..... 250**

**9 References ..... 253**

# List of Tables

**Table 2.1: List of primer pairs for the amplification of PCR fragments encoding one LysM motif (see Figure 5.1 d) ..... 42**

**Table 2.2: List of PCR fragments that encode a variant LytE with two LysM motifs and the resulting plasmids after cloning of these fragments into the parental vector (see Figure 5.2 c) ..... 43**

**Table 2.3: List of PCR fragments that encode a variant LytE with one LysM motif and the resulting plasmids after cloning of these fragments into the parental vector (see Figure 5.2 b) ..... 44**

**Table 2.4: List of PCR fragments that encode a single LysM motif and the resulting plasmids after cloning of these fragments into the parental vector with two LysM motifs (see Figure 5.2 d) ..... 45**

**Table 2.5: List of primer pairs for the amplification of PCR fragments encoding a variant LytE ..... 47**

**Table 2.6: List of PCR fragments that encode a variant LytE and the resulting plasmids after cloning of these fragments into the parental vector ..... 47**

**Table 2.7: List of primer pairs for the amplification of PCR fragments encoding a heterologous variant LytE with a single heterologous LysM motif (see Figure 5.5 b) ..... 48**

**Table 2.8: List of PCR fragments that encode a heterologous variant LytE and the resulting plasmids after cloning of these fragments into the parental vector (see Figure 5.5 c) ..... 48**

**Table 4.1: List of null and double mutant strains between *cwlS, lytF, yqgT* and *lytE, cwlO* ..... 85**

**Table 4.2: List of strains expressing a CwlS protein from the *cwlO* promoter ..... 88**

**Table 4.3: List of strains expressing a LytF protein from the *cwlO* promoter ..... 93**

**Table 4.4: List of strains expressing a YqgT protein from the *cwlO* promoter ... 102**

**Table 4.5: List of strains investigated by autolysis assay (see Figure 4.10) ..... 106**

**Table 5.1: List of PCR fragments encoding variant LytE proteins and the plasmids used as template for their amplification ..... 125**

**Table 5.2: List of plasmids designed for overexpression of variant LytE proteins in *E. coli* ..... 126**

**Table 5.3: [Table continues on next page]: List of strains containing a variant LytE expressed from the *cwlO* promoter ..... 128**

**Table 5.4: List of strains containing a heterologous variant LytE expressed from the *cwlO* promoter ..... 130**

**Table 5.5: List of strains containing a variant LytE protein expressed from the *lytE* promoter ..... 131**

**Table 5.6: List of strains in which the successful disruption of the *cwlO* gene was analysed by western analysis (see Figure 5.9) ..... 132**

**Table 5.7: List of strains investigated by suppressor mutation assay (see Figure 5.10) ..... 135**

**Table 5.8: List of strains containing a HA-tagged variant LytE ..... 136**

**Table 5.9: List of strains investigated by western analysis (see Figure 5.11) ..... 138**

**Table 5.10: List of strains investigated by western analysis (see Figure 5.12) ... 140**

**Table 5.11: List of strains investigated by western analysis (see Figure 5.13) ... 141**

**Table 5.12: List of strains investigated by western analysis (see Figure 5.14) ... 145**

**Table 5.13: List of strains investigated by qRT-PCR analysis (see Figure 5.15) ..... 147**

**Table 5.14: List of strains investigated in growth assays (see Figure 5.16) ..... 150**

**Table 5.15: List of strains investigated in colony morphology assay (see Figure 5.17) ..... 153**

**Table 5.16: List of strains investigated by microscopy (see Figure 5.18) ..... 155**

**Table 5.17: List of strains investigated by growth inhibition assay with NaCl (see Figure 5.19) ..... 158**

**Table 5.18: List of strains investigated by microscopy (see Figure 5.20) ..... 163**

**Table 5.19: List of strains investigated by immunofluorescence microscopy (see Figure 5.21) ..... 166**

**Table 5.20: List of strains investigated by antibiotic sensitivity assay (see Figure 5.22) ..... 167**

**Table 5.21: List of strains from which purified cell walls was obtained (see Figure 5.27) ..... 178**

**Table 6.1: List of the 18 LysM motif proteins of *B. subtilis 168* ..... 187**

**Table 6.2: The 20 strains of the *Bacillus subtilis* species group ..... 206**

**Table 6.3: Categorisation of strains into 4 groups (A, B, C, D) on the basis of D,L-endopeptidase composition and LysM motif distribution .... 211**

**Table 8.1: Summary statistic of the frequency distribution of LysM motifs among different bacterial families ..... 231**

**Table 8.2: List of 45 proteins containing duplicated LysM motifs ..... 234**

**Table 8.3: List of the identified orthologs to LytE, CwlS and LytF ..... 236**

**Table 8.4: List of *B. subtilis* strains used in this study ..... 241**

**Table 8.5: List of *E. coli* strains used in this study ..... 246**

**Table 8.6: List of plasmids used in this study ..... 247**

**Table 8.7: List of primers used in this study ..... 250**

# List of Figures

**Figure 1.1: Vegetative and sporulation cell cycle of *B. subtilis*. ..... 4**

**Figure 1.2: Gram-positive cell envelope and cell wall metabolism. ..... 11**

**Figure 1.3: The WalRK two-component system. ..... 19**

**Figure 1.4: Autolytic enzyme cleavage sites and structures of the seven D,L-endopeptidases in *B. subtilis*. ..... 23**

**Figure 1.5: Alignment and phylogenetic tree of the catalytic domains of LytE, CwlS, LytF and CwlO. ..... 26**

**Figure 1.6: Structures of the three LysM motifs of the LytE D,L-endopeptidase and alignment of all thirty LysM motifs of *B. subtilis*. ..... 32**

**Figure 3.1: Schematic representation of the construction of strain LU17, LU21 and LU25, all of which express LytE from the *cwlO* promoter. ..... 65**

**Figure 3.2: Northern analysis of *lytE* transcript levels in WT168 and in LS103 (Δ*cwlO*) ..... 67**

**Figure 3.3: Transcript levels of *lytE*, *cwlO*, *mreBH*, *sigI* and *yocH* determined by qRT-PCR in strains carrying null mutations in *cwlO* and *lytE*. ..... 68**

**Figure 3.4: Transcript levels of *lytE*, *cwlO*, *mreBH*, *sigI* and *yocH* determined by qRT-PCR for strains that express LytE from the *cwlO* promoter. ..... 71**

**Figure 3.5: Assessment of the stability of the *lytE* transcript expressed from the *cwlO* promoter by northern analysis. ..... 74**

**Figure 3.6: Cell-associated and secreted LytE protein levels determined by western blot analysis. ..... 76**

**Figure 3.7: Localization of the LytE protein. ..... 78**

**Figure 3.8: Phenotypic characterization of strains that express LytE from the *cwlO* promoter. ..... 80**

**Figure 4.1: Schematic illustration of strains that express LytE, LytF, CwlS or YqgT from the *cwlO* promoter. ..... 84**

**Figure 4.2: Typical cell morphology of single (Δ*lytF*, Δ*cwlS*, Δ*yqgT*) and double mutants (in combination with Δ*lytE* and Δ*cwlO*). ..... 86**

**Figure 4.3: Cell-associated protein levels of CwlS in WT168 and when expressed from the *cwlO* promoter with and without *cwlO* leader sequence in comparison to LytE expressed from the *lytE* and the *cwlO* promoter. ..... 89**

**Figure 4.4: Growth curves for strains that express one of the following D,L-endopeptidase LytE, LytF, CwlS or YqgT from the *cwlO* promoter. ..... 91**

**Figure 4.5: Transcript levels of *lytE, lytF, cwlO* and *sigI* in strains that express LytF from the *cwlO* promoter. ..... 94**

**Figure 4.6: Cell-associated protein levels of LytE and LytF in WT168 and when expressed from the *cwlO* promoter with and without *cwlO* leader sequence. ..... 96**

**Figure 4.7: Typical cell morphology of strains that express either LytF, CwlS or YqgT from the *cwlO* promoter. ..... 99**

**Figure 4.8: Localization of the LytE, LytF, CwlS and YqgT proteins in strains in which the protein is expressed from the *cwlO* promoter. ..... 100**

**Figure 4.9: Cell-associated protein levels of YqgT in WT168 and when expressed from the *cwlO* promoter in comparison to other D,L-endopeptidases expressed from the *cwlO* promoter. ..... 103**

**Figure 4.10: Endogenous autolysin activity in strains that express either LytE, LytF, CwlS or YqgT from the *cwlO* promoter. ..... 107**

**Figure 5.1: Schematic structure of the WT168 *lytE* and the generation of PCR fragments encoding variant *lytE* without, with one or with two LysM motifs. ..... 114**

**Figure 5.2: Schematic representation of the construction of vectors designed to enable integration of variant *lytE* open reading frames into the chromosome at the *cwlO* locus. ..... 117**

**Figure 5.3: Determination of the total number of LysM motifs in a vector by agarose gel electrophoresis. ..... 118**

**Figure 5.4: Schematic representation of the integration of the variant *lytE* open reading frame into the chromosome at the *cwlO* locus. ..... 119**

**Figure 5.5 [Figure continues on next page]: Schematic representation of the construction of heterologous variant LytE. ..... 121**

**Figure 5.6: Schematic representation of the integration of a variant *lytE* open reading frame at the *lytE* locus. ..... 124**

**Figure 5.7: Protein expression and purification of the variant LytE[L31cat] protein monitored by SDS-PAGE and Coomassie Blue staining. ... 127**

**Figure 5.8: Purity of overexpressed variant LytE protein as determined by SDS-PAGE and Coomassie Blue staining. ..... 127**

**Figure 5.9: Verification of the absence of the *cwlO* gene, in strains that express a variant LytE protein from the *cwlO* promoter, by determining CwlO protein levels by western analysis. ..... 132**

**Figure 5.10: Suppressor mutation test for the detection of potential secondary mutations, enabling the expression of variant LytE constructs at the *cwlO* locus. ..... 135**

**Figure 5.11: The levels of cell-associated LytE variant proteins, expressed either from the *cwlO* promoter or from the *lytE* promoter. ..... 138**

**Figure 5.12: The levels of cell-associated LytE variant proteins expressed from the *lytE* promoter. ..... 140**

**Figure 5.13: The levels of LytE variant proteins expressed from the *cwlO* promoter, for cell-associated proteins (a) and secreted proteins (b). ..... 142**

**Figure 5.14: Levels of heterologous variant LytE proteins, expressed from the *cwlO* promoter for cell-associated proteins (a) and secreted proteins (b).....146**

**Figure 5.15: *sigI* and *mreBH* transcript levels as determined by qRT-PCR, for strains that express a variant LytE protein from the *cwlO* promoter. .....149**

**Figure 5.16: Growth curves for strains that express a variant LytE protein. .....151**

**Figure 5.17: Colony morphology of strains that express a variant LytE protein from the *cwlO* and the *lytE* promoter. .....154**

**Figure 5.18: Typical cell morphology of strains that express a variant LytE a) from the *cwlO* promoter (in combination with Δ*lytE*) and b) from the *lytE* promoter (in combination with Δ*cwlO*).....156**

**Figure 5.19 [Figure continues on next pages]: Growth curves of strains that express a variant LytE protein from the *cwlO* and the *lytE* promoter. .....160**

**Figure 5.20: Cell morphology of strains that express a variant LytE protein when exposed to salt stress.....164**

**Figure 5.21: LytE localization in strains that express a variant LytE protein when exposed to salt stress.....166**

**Figure 5.22: Susceptibility to fosfomycin of strains that express a variant LytE. .....168**

**Figure 5.23: Protein structure predictions for variant LytE.....170**

**Figure 5.24: Ability of purified variant LytE proteins to degrade purified cell walls established by zymography. .....172**

**Figure 5.25: Ability of purified variant LytE proteins to degrade live cells and purified cell walls. .....173**

**Figure 5.26: Binding affinity of purified variant LytE proteins to purified *B. subtilis* cell walls at different salt concentrations. .....175**

**Figure 5.27: Binding affinity of variant LytE proteins to different types of purified cell walls at different salt concentrations. ..... 178**

**Figure 6.1: Phylogenetic tree of the N-terminal localized LysM motifs of *B. subtilis 168*. ..... 188**

**Figure 6.2: Frequency distributions of the number of LysM motif proteins and individual LysM motifs per strain group, as well as the number of individual LysM motifs within proteins ..... 192**

**Figure 6.3: Frequency distribution of the function of LysM motif proteins. ..... 194**

**Figure 6.4: Summary statistic of the frequency distribution of LysM motifs among different bacterial families ..... 196**

**Figure 6.5: Summary statistic of the frequency distribution of LysM-rich proteins (proteins with two or more LysM motifs) among different bacterial families. ..... 198**

**Figure 6.6: Correlation between proteome size and the number of LysM motifs per strain. ..... 200**

**Figure 6.7: Four examples of proteins that contain LysM motifs with identical amino acid sequence. ..... 203**

**Figure 6.8: Full taxonomic lineage of *B. subtilis 168* ..... 205**

**Figure 6.9: Identification of orthologs to LytE, CwlS and LytF in the *B. subtilis* strain group by amino acid sequence analysis of their catalytic domains ..... 208**

**Figure 6.10: Metric multidimensional scaling of the catalytic domains of the LytE, CwlS and LytF orthologs ..... 210**

**Figure 6.11: Circular phylogenetic tree of the LysM motifs of CwlS, LytE and LytF orthologs. ..... 213**

**Figure 6.12: Metric multidimensional scaling of the LysM motifs of LytE, CwlS and LytF orthologs. ..... 214**

**Figure 6.13: A putative model of the evolutionary development of the CwlS, LytE and LytF D,L-endopeptidases and their LysM motifs. ..... 217**

**Figure 6.14: Cluster analysis of the second and third LysM motifs of CwlS, LytE and LytF orthologs. ..... 218**

**Figure 6.15: Proposed model for the evolutionary development of groups with different sets of paralogs and basic phylogenetic tree of all twenty strains of the *B. subtilis* strain group. ..... 220**

**Figure 8.1: Nucleotide and amino acid sequences of the wild-type LytE..... 238**

# List of Abbreviations

| | |
|---|---|
| ~P | Phosphorylated |
| aa | Amino acids |
| *amp*$^r$ | Ampicillin resistance gene |
| Apase | Alkaline phosphatase |
| ATP | Adenosine triphosphate |
| BLAST | Basic Local Alignment Search Tool |
| bp | Basepairs |
| C | Carboxyl group |
| CA domain | Catalytic and ATP-binding domain |
| cAMP | Cyclic adenosine 3',5'-monophosphate |
| cDNA | Complementary DNA |
| cGMP | Cyclic guanosine 3',5'-monophosphate |
| *cm*$^r$ | Chloramphenicol resistance gene |
| DAPI | 4´,6-diamidino-2-penylindole |
| DHp domain | Dimerisation and Histidine phosphotransfer domain |
| DNA | Deoxyribonucleic Acid |
| *erm*$^r$ | Erythromycin resistance gene |
| FITC | Fluorescein isothiocyanate |
| G+C | Guanine+Cytosine |
| HAMP | Histidine kinase, Adenylyl cyclase, Methyl accepting chemotaxis protein, Phosphatase |
| HK | Histidine Kinase |
| HPt | Histidine Phosphotransferase |
| IPTG | Isopropyl ß-D-1-1-thioglactopyranoside |
| *kan*$^r$ | Kanamycin resistance gene |

| | |
|---|---|
| Kb | Kilobasepairs |
| LB | Luria Bertani broth |
| LFH | Long-Flanking-Homology |
| LTA | Lipid Teichoic Acid |
| LysM | LysM motif |
| cat | Catalytic domain |
| MMDS | Metric Multidimensional scaling |
| mRNA | Messenger RNA |
| MW | Molecular Weight |
| NAG | N-acetylglucosamine |
| NAM | N-acetylmuramic acid |
| nc | Negative control |
| OD | Optical density |
| ORF | Open Reading Frame |
| PAS | Per-Arnt-Sim subdomain |
| PBP | Penicillin binding protein |
| PCR | Polymerase Chain Reaction |
| PG | Peptidoglycan |
| PGRP | Peptidoglycan Recognition Protein |
| qRT-PCR | Quantitative Reverse Transcription PCR |
| RBS | Ribosome Binding Site |
| RNA | Ribonucleic Acid |
| rpm | Revolutions per minute |
| RR | Response Regulator |
| Sd | Standard deviation |
| SDS | Sodium Dodecyl Sulfate |
| *spc*$^r$ | Spectinomycin resistance gene |
| TA | Teichoic Acid |

| | |
|---|---|
| TCS | Two-component Signal Transduction System |
| UPGMA | Unweighted Pair Group Method with Arithmetic Mean |
| WTA | Wall Teichoic Acid |
| WT168 | Wild-type strain *B. subtilis 168* |
| xyl | Xylose |

# Publications

Manuscript submitted

**Devine, K.M., Noone, D., Botella, E., Uelze, L., Smyrna, D., Devine, S.K., Salzberg, L., Szurmant, H.** (2016) Responsiveness to a range of diverse signals generates variability in WalRK activity during the growth cycle of *Bacillus subtilis*.

Manuscript in preparation

**Uelze, L., Devine, K.M., Noone, D.** (2016) An investigation of LytE in fulfilling the D,L-endopeptidase requirement for viability of *Bacillus subtilis*.

# Abstract

The bacterial cell wall determines cell shape, resists turgor pressure, provides a microclimate between the cell and the external milieu and acts as a platform for the exposure of cellular proteins. The cell wall is a dynamic structure with constant synthesis and turnover to accommodate the remodelling that is required during cell growth and division. Autolysins play a key role in cell wall remodelling, with *B. subtilis* encoding 35 such enzymes. Redundancy of enzymatic activity is a notable feature of this enzyme complement. For example, there are 7 D,L-endopeptidases (CwlO, CwlS, LytE, LytF, PgdS, YkfC, YqgT) encoded in the *B. subtilis* genome. Although none of the autolysins in *B. subtilis* is essential for viability, the CwlO and LytE D,L-endopeptidases are synthetically lethal. This is an especially interesting finding, since CwlO and LytE differ completely in the domains outside of the catalytic domain. The N-terminal domain of CwlO has a coiled-coil structure that interacts with the membrane embedded FtsEX complex, whereas LytE has 3 LysM motifs that bind to peptidoglycan.

The synthetic lethality of CwlO and LytE presents the opportunity to investigate experimentally the functional features of a D,L-endopeptidase necessary to fulfil the requirement for viability. To achieve this, an experimental system was established in which genes encoding different D,L-endopeptidases and variants of LytE were introduced at the *cwlO* locus, replacing the *cwlO* open reading frame. The capability of these enzymes to fulfil the D,L-endopeptidase requirement for viability was then tested by attempting to delete the endogenous *lytE* gene. Successful deletion of the endogenous *lytE* gene signifies that the investigated D,L-endopeptidase can fulfil the requirement for viability. A particular focus of this study was to examine the three LysM motifs of LytE to establish how many are needed for LytE function and the extent to which they are functionally equivalent.

Our results show that expression of LytE and LytF, containing 3 and 5 LysM motifs respectively, from the *cwlO* promoter can fulfil the D,L-endopeptidase requirement for viability. However, cells expressing LytF have a profound phenotype with aberrant morphology and very slow growth. Interestingly, expression of the CwlS, containing 4 LysM motifs and YqgT, that has an N-terminal domain of unknown function, cannot fulfil the D,L-endopeptidase requirement for viability. Our results show that a LytE protein with any one of its three LysM motifs can fulfil the D,L-endopeptidase requirement for viability, but that the three LysM motifs are not functionally equivalent. Cells expressing a LytE variant with LysM1 generally show a more robust and wild-type

like phenotype than those expressing motifs LysM2 and LysM3. However, all cells expressing a LytE variant with a single LysM motif exhibit a greater sensitivity to salt stress than wild-type.

Moreover, the association of LytE variant proteins with the cell wall correlated positively with the number of LysM motifs and negatively with the concentration of NaCl present in the medium both *in vivo* and *in vitro*. Furthermore, bioinformatics analysis allowed the formulation of an evolutionary model describing the development of the LytE, CwlS and LytF D,L-endopeptidases and the independent expansion of LysM motif repeats in these enzymes. In summary, these results show that the LytE enzyme has been optimized to fulfil the D,L-endopeptidase requirement for cell viability of *B. subtilis* with regard to the number and properties of LysM motifs that mediate peptidoglycan-binding.

# 1 Chapter 1: Introduction

## 1.1 *Bacillus subtilis*

*Bacillus subtilis* is a widely distributed, Gram-positive, mesophilic and facultative anaerobic bacterium of the *Firmicutes* phylum (Graumann, 2007). *B. subtilis* cells are rod-shaped and about 4-10 µm long and 0.25-1 µm in diameter and can form short chains (Yu *et al.*, 2014). The cell has a thick multi-layered cell wall that counteracts the internal turgor pressure that confers shape on the bacterium and that protects it from the conditions that prevail in the external environment. Furthermore, *B. subtilis* cells have flagella that confer motility and allow them to engage in swarming behaviour (Kearns and Losick, 2003, Guttenplan *et al.*, 2013).

Although it is possible to isolate *B. subtilis* from almost all environments, it is most prevalent in the upper layers of the soil, especially in the rhizosphere, surrounding plant roots and decaying plant material (Earl *et al.*, 2008). Interestingly it is also found as a commensal in the human gut (Hong *et al.*, 2009). The upper soil is a diverse environment, with a great variability in temperature, salt, pH and the availability of water and nutrients. This is reflected in the large number and variety of systems encoded in the *B. subtilis* genome that are involved in detecting, transducing and adapting to stress conditions. Of particular note in this regard are the 34 two-component signal transduction systems and 17 alternative sigma factors that confer the capability to sense and adapt to a wide range and variety of conditions (Fabret *et al.*, 1999; Hecker *et al.*, 2007; Mascher *et al.*, 2006; Staron *et al.*, 2009).

### 1.1.1 *B. subtilis* as a model system for Gram-positive bacteria

*B. subtilis* was one of the first studied Gram-positive bacteria. It has been intensively researched for more than fifty years and is now the model organism of choice for Gram-positive bacteria (Harwood, 1992; Sonenshein *et al.*, 1993). Today a great number of laboratories worldwide are engaged in all areas of *B. subtilis* research. Besides its long and successful history in research facilities, it is also a much valued 'workhorse' for the biotechnological industry. Several factors, some of which are discussed below, have contributed to this popularity, most of which stem from characteristics of *B. subtilis* that originally evolved to ensure its survival in the challenging environment that is the top layer of the soil.

**(i) The capacity to secrete proteins in the gram per litre range qualifies *B. subtilis* as a 'workhorse' for the biotechnological industry.**

The ability of *B. subtilis* to express and secrete high quantities of proteins and enzymes is of great industrial interest and commercial value. This is reflected in general market share estimates. The global market for microbial products was estimated at $144 billion in 2014 (BCC Research, 2015), with industrial enzymes reaching a share of $5 billion in 2013, which is expected to rise to $7 billion by 2018 (BCC Research, 2014; Gurung *et al.*, 2013). *B. subtilis* is widely used to produce industrial enzymes such as proteases, amylases and lipases, for the detergent, food and beverage sector as well as for the expression of recombinant proteins (Harwood and Cranenburgh, 2008; Schallmey *et al.*, 2004; van Dijl and Hecker, 2013).

*Bacillus subtilis* and other *Bacillus* species are among the best suited Gram-positive bacteria for enzyme production. This is, because *B. subtilis* has evolved to express and secrete enzymes when exposed to nutrient limitation in order to degrade nutrients outside of the cell (Harwood and Cranenburgh, 2008). *B. subtilis* can secrete an astonishing quantity of enzymes up to 20 to 25 gram protein per litre (van Dijl and Hecker, 2013). Furthermore, it can access a wide range of substrates like proteins, starch and pectin and therefore does not require expensive culture media.

Another beneficial factor for industrial applications is that *B. subtilis* is classified as a GRAS (Generally Regarded As Safe) bacterium (Hong *et al.*, 2008). Although it is closely related to human pathogens such as *Staphylococcus aureus* and *Bacillus anthracis* it is non-pathogenic and very rarely causes food poisoning. Enzymes and proteins produced in *B. subtilis* are therefore completely free of toxic by-products. Besides, studies have shown that *B. subtilis* can even be applied as a probiotic (Hong *et al.*, 2008).

Further applications for *B. subtilis* are the development of seeds coated with *B. subtilis* spores to foster their germination and subsequent growth (Tu *et al.*, 2016), since it has been shown that *B. subtilis* forms beneficial symbiotic associations with plants and promotes plant growth (Oldroyd, 2013).

**(ii) The ability of *B. subtilis* to become competent and to uptake and integrate exogenous DNA was the prerequisite for genetic research.**

An ability to manipulate the genome is a prerequisite for undertaking genetic analysis of an organism. This requires an ability to transform the organism with DNA. In 1960, Anagnostopoulus and Spizizen reported a method to transform *B. subtilis*. The discovery that a subpopulation of *B. subtilis* cells becomes naturally competent under

particular environmental and nutritional conditions has become a major research topic (Chen and Dubnau, 2004; Claverys *et al.*, 2006; Jakobs and Meinhardt, 2015). Competence is defined as the ability of bacteria to take up exogenous DNA and to (under certain conditions) integrate this DNA into the genome in order to acquire new traits, such as resistance to antibiotics (Chen and Dubnau, 2004). The ability of *B. subtilis* to become competent and transform exogenous DNA allows researchers to easily manipulate its genome in order to investigate its biology or to build bacterial strains that are superior in aspects such as enzyme production, which gives it a major advantage over bacteria that resist genetic engineering.

**(iii) Sporulation in *B. subtilis* is a simple and amenable developmental system.**

Another well studied adaptation strategy that is of great research interest is sporulation. Sporulation occurs if the survival of *B. subtilis* is threatened. Then the bacterium forms hardy endospores that can tolerate extreme conditions for prolonged periods with very little metabolic activity, a state also referred to as 'dormancy' (Higgins and Dworkin, 2012; Setlow, 2006; Tan and Ramamurthi, 2014). For example, it was shown that *B. subtilis* spores can survive in outer space and re-germinate afterwards (Wassmann *et al.*, 2012). Sporulation is a response to hostile conditions and nutrient starvation and begins with an asymmetric cell division, which results in two daughter cells with different cell fate. The spore daughter cell ('forespore') is then engulfed by the host cell ('mother cell') and develops an extremely rigid cell wall ('spore coat'). The maturation process also involves dehydration of the forespore as well as other metabolic adaptations. Finally the spore is released from the mother cell which for this purpose induces lysis of its own cell wall. The steps of the sporulation process are schematically pictured in Figure 1.1 (a).

Sporulation has become a model system for the understanding of developmental cell biology. Sporulation has many similarities to other developmental processes. First, like in most cell differentiations, *B. subtilis* cells that commit to sporulation have irreversibly determined their cell fate. Second, sporulating cells are required to regulate their transcription, manage intracellular signalling and undergo membrane remodelling accordingly (Tan and Ramamurthi, 2014).

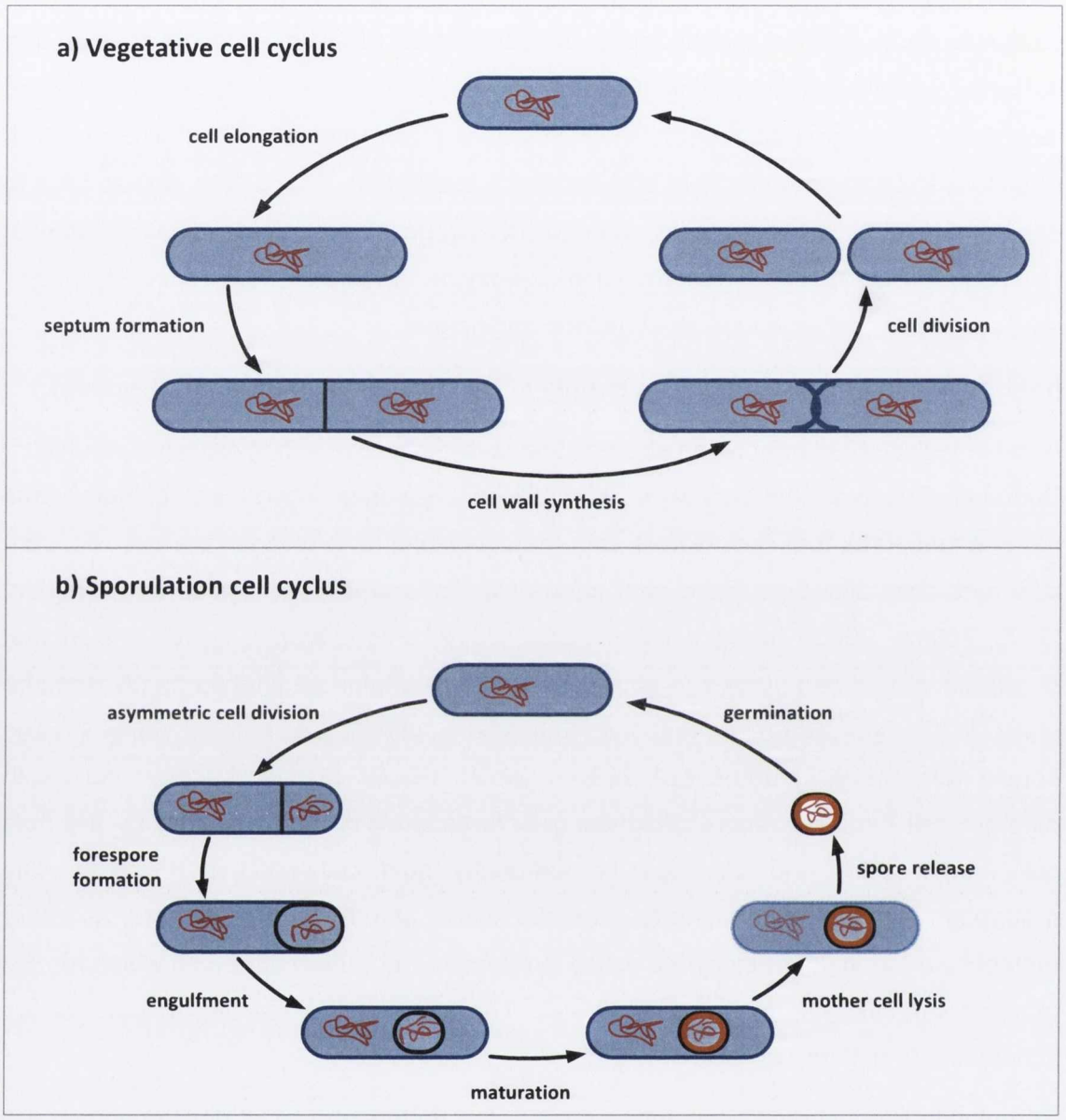



**Figure 1.1: Vegetative and sporulation cell cycle of *B. subtilis*. a)** In the vegetative cell cycle a period of cell elongation is followed by a symmetric cell division, which is initiated by the formation of the cell septum. New cell wall material is synthesized to form the poles of the two daughter cells before autolytic enzymes break down the remaining cell wall separating the two cells. **b)** The sporulation cell cycle begins with an asymmetric cell division, after which the smaller daughter cell is engulfed by the mother cell. The forespore then undergoes maturation, before the mother cell lyses and the spore is released. Figure adapted from Higgins and Dworkin (2012).

Studying developmental processes in *B. subtilis* has the advantage of a relatively simple system that is easily genetically manipulated. Furthermore, sporulation genes are often non-essential for vegetative growth which facilitates their investigation.

**(iv) *B. subtilis* forms biofilms which allow genetic analysis of this important medical phenomenon in a genetically tractable organism.**

*B. subtilis* has the capacity to form biofilms (Vlamakis *et al.*, 2013) that feature a dense association of different cell types. It is now recognized that biofilms are the predominant lifestyle of bacteria in natural environments (Mielich-Suss and Lopez, 2015). Biofilms protect the entire bacterial community by producing a complex extracellular matrix, made of polysaccharides, proteins and DNA (Kearns *et al.*, 2005) that lessens the impact of fluctuations in environmental conditions and offers protection from host defences (Cairns *et al.*, 2014; Roilides *et al.*, 2015). Biofilms have a complex structure and require elaborate cell-cell signalling and cell specialisation. They can form on living and non-living surfaces as well as on liquid surfaces. Biofilm formation begins with cells that lower their expression of flagella genes and increase the production of the extracellular matrix molecules. These sessile cells then form long chains, which expand the biofilm and reinforce its structure. The biofilm undergoes a maturation process, which is associated with spatially constrained cell death. This allows for the formation of channels and airspaces that facilitate the circulation of oxygen and nutrients. Finally the cells that are located at the top of the biofilm undergo sporulation and spores are released and distributed to start biofilm formation on other sides (Mielich-Suss and Lopez, 2015).

The understanding of biofilm formation is extremely significant in the medical field for the treatment of infections associated with biofilms (such as dental caries) and for the control of biofilm formation on medical instruments and implanted biomedical devices (e.g. catheters). Furthermore, biofilm formation is a major challenge in industrial settings, especially in food production. The formation of biofilms on industrial plants and equipment can decrease productivity, hinder sanitisation efficiency and can pose a health hazard (Bridier *et al.*, 2015). Although in most industrial and medical settings biofilm formation is detrimental, biofilms are crucial in many natural environments. Some applications even intentionally induce biofilm formation and exploit its associated characteristics, for example for the treatment of sewage or in biofilm reactors (Ercan and Demirci, 2015).

**(v) *B. subtilis* has become a model organism for the study of bistability in microbial populations.**

The processes of competence development, biofilm formation, cannibalism and sporulation all display bistability – i.e. only some cells in the population enter into these developmental pathways. Bistability refers to the fact that within any genetically identical bacterial community, individual cells do not usually develop in a uniform way, but express different genes at different times. For example, only approximately 20 % of *B. subtilis* cells become competent when exposed to the appropriate conditions. Within a biofilm the phenomena of bistability causes the differentiation of cells in different cell types, e.g. sessile and motile cells (Dubnau, 2010). During sporulation only a minority of cells become spores. Moreover, before *B. subtilis* undergoes sporulation it attempts to cannibalise a fraction of the bacterial community. In this process a subgroup of cells starts to secrete toxins that cause lysis in the surrounding cells. The lysed cells then serve as a nutrient source for the remaining bacteria that can so successfully delay sporulation (Lopez *et al.*, 2009).

All these examples display the phenomenon of bistability – i.e. there is heterogeneity in the bacterial population. Bistability is achieved by random switching mechanisms often with a positive feedback loop that divide a population into phenotypically-distinct subpopulations (Dubnau and Losick, 2006). Bistability is beneficial for the survival of a bacterium by allowing a clonal population to follow different stress response strategies at the same time, maximising its chances of survival. Given that *B. subtilis* displays at least four forms of bistability it is a good model organism for the studying of this phenomenon. The investigation of bistability processes in a relatively simple organism such as *B. subtilis* is also helpful for the understanding of more complex developmental processes in eukaryotes, especially the determination of cell fate and cellular differentiation.

All these characteristics, but especially the ability to generate endospores and become highly competent, make *B. subtilis* a very interesting and useful bacterium for research. Further advantages include the easy handling and its capacity to produce and secrete high quantities of enzymes and proteins. The latter, together with the ease with which it can be genetically engineered and its GRAS status, have established *B. subtilis* as a workhorse of the biotechnology industry.

The domesticated strains of *B. subtilis* favoured by research and industry, have in many cases lost abilities common to their ancestors still living in the wild, which made them unsuitable for laboratory handling (Afkham *et al.*, 2007). For example most domesticated strains do not exhibit swarming behaviour and are not capable of forming

complex biofilms. In contrast their ability for competence is increased and industrially used strains, broadly selected for high secretion capacities, often lack proteases. One of the strains most widely used in laboratories and also in this study, is *B. subtilis 168*, which is a tryptophan-requiring auxotroph.

### 1.1.2 Genetics and genomics of *B. subtilis*

The continuous scientific and industrial interest in *B. subtilis*, has led to it being one of the best understood organisms in the world. This insight is especially comprehensive on a genetic level. The *B. subtilis* genome is organized in one circular chromosome. Owing to the early scientific interest in *B. subtilis,* it was sequenced among the first full-genome sequencing projects. The size of its genome and the transformability of *B. subtilis* led to the completion of a genetic map and to the establishment of the complete genome sequence. In 1997 a network of European institutes and several research facilities in Korea and Japan performed the task of identifying its 4.2 Mbp large genome (Kunst *et al.*, 1997; Moszer, 1998). A seminal discovery from this study was the realisation that the function of more than 50 % of the genes encoded in the genome was unknown. This prompted a functional analysis by knocking out individual genes that found that 273 genes were essential (Albertini *et al.*, 2003). As a result of the consistent interest in *B. subtilis*, together with outstanding improvements of sequencing techniques the genome was sequenced a second time in 2009 (Barbe *et al.*, 2009). A subsequent study also found that of a total of 4100 genes, only 261 genes coding for 259 proteins and two functional RNAs are truly essential (Commichau et al., 2013). Initial attempts were made to minimise the genome (Ara et al., 2007) and projects to define the minimal possible *B. subtilis* genome (solely containing essential genes) are currently under way (Tanaka et al., 2013).

Multi-omics studies on systems analysis in cells growing on glucose and malate (Buescher *et al.*, 2012) and transcriptomic analysis of cells growing under 94 different environmental conditions (Nicolas *et al.*, 2012) give unprecedented insights into the physiology of *B. subtilis*. These studies have also generated a multitude of databases such as Subtiwiki (Michna *et al.*, 2013), SporeWeb (Eijlander et al., 2013) and DBTBS (Sierro *et al.*, 2008) that provide unparalleled information on studying physiological processes in *B. subtilis*.

## 1.2 The cell envelope

Most bacteria are surrounded by structures located external to the cytoplasmic membrane, a region known as the cell envelope. A notable exception is the genus

*Mycoplasma* which lacks structures outside the cytoplasmic membrane. Many such bacteria are pathogenic (e.g. *Mycoplasma genitalium, Mycoplasma pneumonia*) and are difficult to treat, since in lacking a cell wall, they are resistant to all antibiotics that target synthesis of this cell structure (Couldwell and Lewis, 2015).

The structure of the cell envelope is the basis for the differentiation of bacteria in Gram-positive and Gram-negative bacteria, a distinction that originated in a certain staining technique (Gram-staining). Gram-positive bacteria (including *B. subtilis*) are endowed with an inner plasma membrane and surrounded by a very thick cell wall made up of several peptidoglycan layers that form a thick structure of 15-80 nanometres. In contrast, Gram-negative bacteria, such as *E. coli*, have only a thin layer of peptidoglycan, but possess a second semi-permeable cell membrane (outer membrane), which is located outside of the cell wall and further stabilizes the cell. The cell wall of Gram-positive as well as Gram-negative bacteria is made up of peptidoglycan, which is intertwined with long anionic polymers, called teichoic acids. An extensive review describing the cell envelope is given by Silhavy *et al.* (2010).

### 1.2.1 Gram-positive cell wall structure and composition

Most Gram-positive bacterial cells are surrounded by a complex, rigid structure, called cell wall (Carballido-Lopez and Formstone, 2007; den Blaauwen *et al.*, 2008). The cell wall is composed of peptidoglycan and anionic polymers (Hayhurst *et al.*, 2008). A schematic cross section of the cell wall of *B. subtilis* is shown in Figure 1.2 (a) (see page 11).

The structure of the cell wall is predominantly determined by layers of peptidoglycan. Peptidoglycan is a biopolymer made of chains of alternating N-acetylmuramic acid (NAM) and N-acetylglucosamine (NAG) which are cross-linked by short peptide chains (Vollmer, Blanot *et al.*, 2008). These peptidoglycan chains run perpendicular to the length axis of the cell (Hayhurst *et al.*, 2008). The two sugars (NAM and NAG) are connected by β-(1,4)-glycosidic bonds. A short peptide chain is attached to each NAM molecule. The number of amino acid residues in this peptide chain varies between species. *B. subtilis* has peptide chains with four to five residues: L-alanine, D-glutamine, meso-diaminopimelic acid, two D-alanine, with the second D-alanine residue often being cleaved off by DD-endopeptidases. Amino acid residues in different peptide chains are linked to each other by transpeptidases. This cross-linking ensures that the peptidoglycan layers are attached to each other, which results in a strong and rigid structure. The exact site of the cross-linking is also species dependent. In

*B. subtilis* the cross-linking of the peptide chains occurs between the meso-diaminopimelic acid residue and the second D-alanine of two peptide chains (pictured in Figure 1.4 (a) (see page 23).

The peptidoglycan layers are not only cross-linked by peptide chains, but also densely interfused with other glycopolymers, called teichoic acids. The two main types of these teichoic acids are wall teichoic acid (WTA) that is covalently attached to peptidoglycan and lipoteichoic acid (LTA) that is attached to a lipid that anchors it in the membrane (Brown *et al.*, 2013). The most abundant teichoic acid is the WTA and studies have estimated that every ninth NAM molecule has an attached WTA (Brown *et al.*, 2013). The WTA polymer is made up of a disaccharide that is linked to the NAM molecule and a main chain of several phosphodiester-linked polyol repeats. These chains pervade several layers of peptidoglycan and can reach well beyond the peptidoglycan cell wall into the extracellular space. Teichoic acid polymers are made up of repeats of glycerol phosphate, glucosyl phosphate or ribitol phosphate. WTA contributes to the rigidity of the cell wall by further cross-linking peptidoglycan layers and also play important roles in the determination of cell shape, cell division, pathogenicity and antibiotic resistance (Brown *et al.*, 2013). Although LTA is likely to fulfil similar functions, the importance of LTA is much less understood (Percy and Grundling, 2014).

Unlike Gram-negative bacteria, Gram-positive bacteria have no outer membrane to display surface proteins. For this reason, all surface proteins that are secreted have to be attached covalently to the inner membrane, peptidoglycan or teichoic acids. Anchoring to the cell membrane can be achieved by membrane-spanning domains or lipid anchors. Proteins that are secreted carry an N-terminal secretion signal peptide and often a C-terminal sorting signal that is recognized by sortase enzymes, which then covalently link these proteins to the cell wall (Schneewind and Missiakas, 2012). Another possibility is the attachment of proteins to the cell wall by peptidoglycan-binding motifs, such as the LysM motif.

A crucial point to note is, that although the cell wall is a robust formation it is not a static structure, but undergoes continual processes of synthesis, turnover and recycling. These dynamic processes are required to remodel the cell wall during cell growth, division and sporulation.

### 1.2.2 Cell wall metabolism

The cell wall metabolism is a complex interplay of cell wall synthesis, cell wall turnover and cell wall recycling. These processes are essential for growth, cell division and sporulation and have to be well coordinated to prevent cell lysis or an inhibition of cell

elongation and division. A schematic overview of the vegetative and sporulation cell cycle is given in Figure 1.1 (see page 4).

The biosynthesis of the cell wall, necessary for growth and cell elongation, is described extensively in the review by Scheffers and Pinho (2005). A simple schematic representation of the peptidoglycan synthesis is shown in Figure 1.2 (b). In short, building blocks, which consist of a disaccharide formed by N-acetylglucosamine and N-acetylmuramic acid with an attached pentapeptide are assembled on the cytosolic side of the membrane. These units are then loaded on a polyisoprenoid carrier (undecaprenol phosphate), which generates Lipid II (Breukink and Kruijff, 2006; Kruijff *et al.*, 2008; van Heijenoort, 2007). Lipid II is flipped across the membrane by Lipid II flippases (Scheffers and Tol, 2015), a process that transports the peptidoglycan precursor to the outside of the cell wall. The disaccharide is then incorporated into the cell wall by Penicillin Binding Proteins (PBPs) (Macheboeuf *et al.*, 2006; Sauvage *et al.*, 2008). The carbohydrate chains of peptidoglycan are then polymerized by transglycosylases and cross-linking of the peptide chains is achieved by transpeptidases. Incorporation can require that the peptidoglycan matrix is first cleaved by autolytic enzymes. The lipid anchor is released and flipped back across the membrane, where it is recycled as a carrying unit. WTA and LTA are synthesized by a similar mechanism (Brown *et al.*, 2013; Percy and Grundling, 2014). Cell wall biosynthesis is an important antibiotic target and currently around 50 % of antibiotics target this pathway (Scheffers and Tol, 2015; Schneider and Sahl, 2010). It is thought that the complex interplay of cell wall synthesis enzymes is combined in a dynamic protein complex, called the elongasome (den Blaauwen *et al.*, 2008). The elongasome coordinates the integration of new disaccharide-pentapeptides into the cell wall and directs the elongation of the cell along the cylindrical axis. It has been established that the elongasome is closely associated with MreB, the cytoskeletal cell shape protein, but the direct interaction has not been uncovered yet (Laddomada *et al.*, 2016).

While the process of cell wall synthesis in growing cells is monitored by the elongasome, cell division is governed by the divisome. The divisome is a ring of polypeptides that are involved in cell division. The divisome is formed after chromosome replication and defines the two equally-sized daughter cells, which is why it is localized mid-cell.

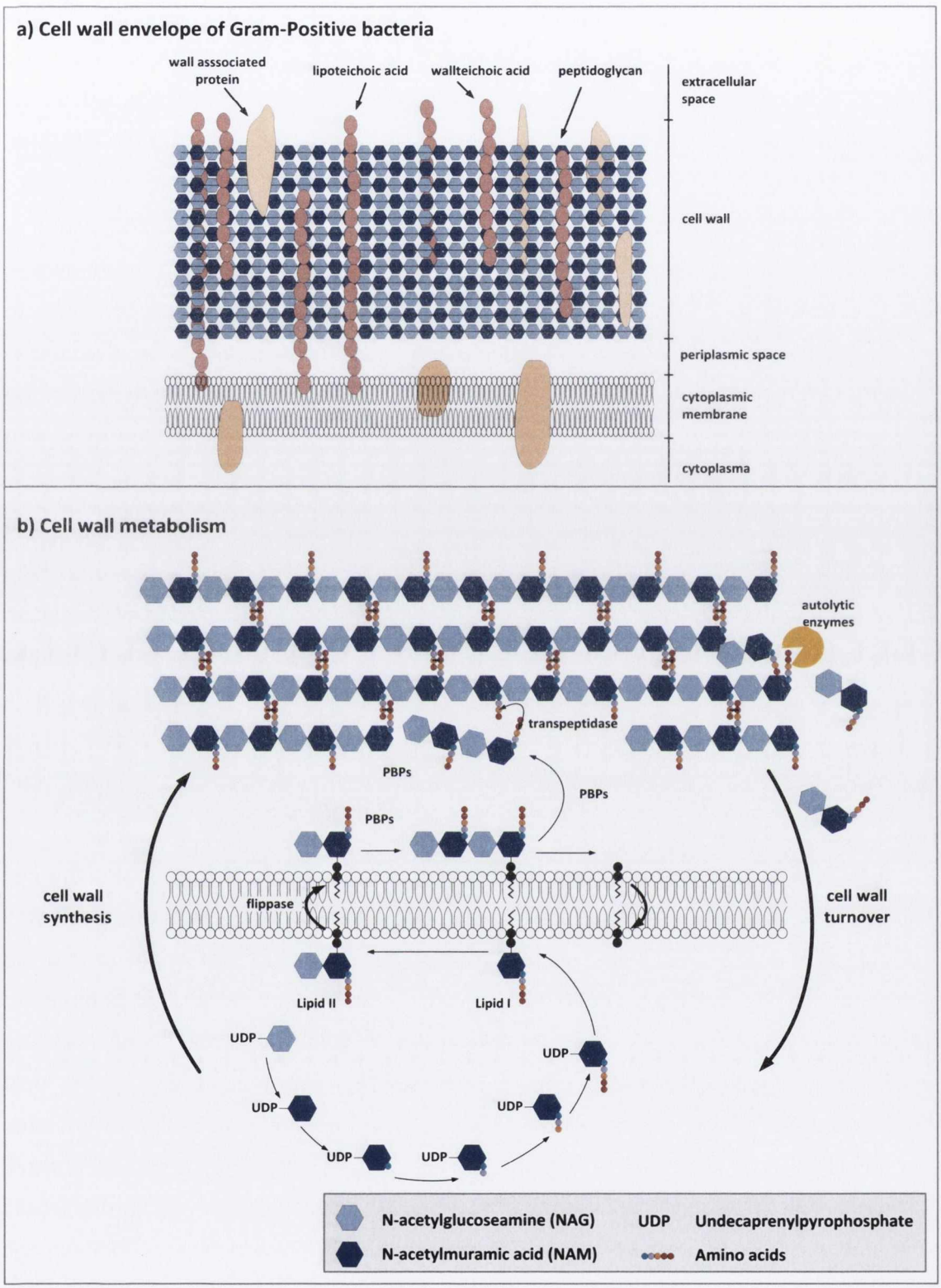



**Figure 1.2: Gram-positive cell envelope and cell wall metabolism. a)** Structure of the cell envelope of a Gram-positive bacterium. **b)** Peptidoglycan synthesis, turnover and recycling are part of the general cell wall metabolism. Disaccharides formed by N-acetylglucosamine and N-acetylmuramic acid with an attached pentapeptide are assembled in the cytoplasma and loaded on a polyisoprenoid undecaprenyl phosphate carrier (Lipid II), which is flipped across the membrane. Incorporation of the disaccharide into the cell wall is achieved by penicillin binding proteins (PBPs), after cleavage of the peptidoglycan matrix by autolytic enzymes. The carbohydrate chains of peptidoglycan are then integrated by PBPs and cross-linking of the peptide chains is achieved by transpeptidases. Figure adapted from Pinho *et al.* (2013).

The divisome is a dynamic protein complex that coordinates the formation of the septum, the synthesis of new peptidoglycan cell wall material, and eventually the degradation of the same to allow for the separation of the daughter cells, as well as the generation of new cell poles (Egan and Vollmer, 2013). The assembly of the divisome is initiated by the protein FtsZ, a tubulin homolog, which polymerizes to form a ring-like structure at the midpoint of the rod-shaped cell (Adams and Errington, 2009; Pinho *et al.*, 2013). The Z ring then functions as scaffolding onto which all required cell division proteins as well as PBPs bind, to form the divisome. The Z ring then constricts to enable cell division. At first, new cell wall material is synthesized along the divisome to form the septum. Later the peptidoglycan is degraded to allow for the separation of the cells and remodelled to create the new cell poles of the daughter cells (den Blaauwen *et al.*, 2008; Uehara and Bernhardt, 2011).

Both processes, cell wall synthesis and cell division, require an orderly degradation of the cell wall at certain sites. During cell wall synthesis the integration of new building blocks into the existing structure is not possible without prior cleavage of peptidoglycan bonds. During cell division the degradation of the cell wall material allows the two daughter cells to separate and form new cell poles. Therefore, cell wall degradation is as crucial for the cell wall metabolism as cell wall synthesis is. The degradation of the cell wall is achieved by autolytic enzymes that cleave bonds within the peptidoglycan polymer.

Although Gram-positive bacteria do not have an outer membrane that retains the cell wall turnover products, it is thought that the cell aims to recycle as much cell wall material as possible (Reith and Mayer, 2011). This allows the bacterium to use the peptidoglycan degradation products to synthesize new cell wall material or to tap them as an energy source (Park and Uehara, 2008). Furthermore, pathogenic bacteria recycle their peptidoglycan degradation products to prevent discovery by the host immune system, which can recognize peptidoglycan fragments as a sign of bacterial invasion (Boudreau *et al.*, 2012). Peptidoglycan degradation products can also serve as messenger substances, for example to promote the re-activation of dormant cells (Keep *et al.*, 2006; Shah *et al.*, 2008) or for the induction of antibiotic resistance in cell populations (Boudreau *et al.*, 2012; Johnson *et al.*, 2013). Cell wall recycling begins with the cleavage of bonds within the peptidoglycan by autolytic enzymes. This releases smaller peptidoglycan units from the cell wall mesh, which then have to be retained and transported to the periplasmic space. From there, the muropeptides are transported across the membrane by permeases and re-used for peptidoglycan biosynthesis (Johnson *et al.*, 2013).

Cell wall recycling is less well understood in Gram-positive than in Gram-negative bacteria, but many genes encoding cell wall recycling enzymes in Gram-negative bacteria have homologs in Gram-positive bacteria, indicating a similar function (Johnson *et al.*, 2013). These genes are murQ (an etherase that cleaves lactate from N-acetylmuramic acid), murR (transcriptional regulator of genes required for the utilization of N-acetylmuramic acid), murP (NAM-specific phosphotransferase), amiE (N-acetylmuramyl-L-alanine amidase) and nagZ (N-acetylglucosaminidase).

### 1.2.3 Cell wall function

The bacterial cell wall provides protection from pressure, turgor and toxins. The cell wall separates the cell from its surroundings and enables it to create its own microenvironment. It acts as a semi-permeable barrier to the exchange and transport of chemical compounds. Furthermore, it provides a platform for the display of flagella and pili as well as of proteins involved in communication, host recognition and signalling (Dramsi *et al.*, 2008; Guttenplan *et al.*, 2013; Jordan *et al.*, 2008; Navarre and Schneewind, 1999).

The understanding of the bacterial cell wall is important, because its structure and composition determines many characteristics of bacteria: pathogenicity (Frirdich and Gaynor, 2013), robustness, the attachment to surfaces and of course, their shape (Young, 2006). Furthermore, the synthesis of the cell wall is a primary antibiotic target.

## 1.3 Two-component signal transduction systems in bacteria

### 1.3.1 Typical two-component system

Two-component signal transduction systems (TCSs) enable organisms to receive and process extracellular and intracellular signals (Ulrich and Zhulin, 2010). This type of signal transduction system is found in almost all bacteria (Hoch, 2000; Stock *et al.*, 2000), approximately half of all archaea (Wuichet *et al.*, 2010) and some eukaryotes, especially plants (Alex and Simon, 1994). A basic TCS consists of a sensor kinase which detects a specific signal and a response regulator, usually a transcription factor that initiates a cellular response appropriate to the activating signal. Sensor kinases are often located in the cell membrane, but some are located within the cytoplasm e.g. KinA in *B. subtilis*.

The total number of TCSs in any given organism is in relation to the size of its genome, as well as its lifestyle (Galperin, 2005). Bacteria living in complex environments, where

conditions change often, have typically more TCSs than others. For example the pathogens *Mycoplasma genitalium and Mycoplasma hyopneumoniae* possess none of these signal transduction systems (Kiil *et al.*, 2005; Minzuno, 1998), whereas the ubiquitous occurring *B. subtilis* has 36 histidine kinases and 34 response regulators (Fabret *et al.*, 1999). TCSs not only vary in their total number in different species, but also in their structure and complexity, ranging from simple linear pathways to complex signalling circuits (Hoch, 2000; Jung *et al.*, 2012). Generally, TCSs are made up of two essential molecules; a histidine kinase (HK) (Carballido-Lopez *et al.*, 2006) that senses the signal and a response regulator (RR) that triggers the cellular response. The signal transduction between the HK and the RR is achieved by a phosphotransfer reaction. An example of a typical TCS (the WalRK TCS of *B. subtilis*) is presented in Figure 1.3 (b) (see page 19).

The HK is generally situated in the bacterial cell membrane, with a sensing domain reaching into the extracellular, transmembrane or cytoplasmic space and a kinase core inside the cell for the interaction with the RR. A cytoplasmic linker connects both domains of the HK. The HK is anchored in the membrane over two or more transmembrane segments, but besides the common membrane-bound HKs, soluble kinases are also known (MacFarlane and Merrick, 1985; Stock *et al.*, 1988). A HK can only be activated by the signal for which it is specific. According to the signal type, HKs can differ highly in size and in the number of sensing domains (Attwood, 2013). The sensing domain is the most variable part, which is responsible for recognizing physical signals such as pressure, temperature or light as well as for the detection of chemical compounds (ligands) e.g. nutrients, antibiotics and quorum-sensing signals. The wide spectrum of different sensing domains is described in a review by Cheung and Hendrickson (2010). HK coding sequences are easily identified on the basis of certain highly-conserved homologous regions, named H, N, G1, F and G2 boxes (Parkinson and Kofoid, 1992; Swanson *et al.*, 1994), which can be arranged in different ways. Furthermore, all HK have a C-terminal CA domain which binds the ATP molecule and catalyses the transfer of the phosphoryl group from the ATP to the conserved histidine amino acid located in the conserved H-box (Marina *et al.*, 2001; Song *et al.*, 2004).

A typical response regulator has two domains (Figure 1.3 b, see page 19) - a receiver domain containing a conserved aspartate which accepts the phosphoryl group from the His~P of the sensor kinase and an output domain which is usually, but not always a DNA-binding domain. Similar to HKs most RRs have two domains, from which one is responsible for receiving the signal from the associated HK and one initiates the cellular response. The receiver domain is highly conserved in all response regulators and shows typically a $(\beta\alpha)_5$ topology in combination with a divalent metal ion, which is

described in great detail in a review by Robert Bourret (2010). The active site of the receiver domain is formed by five conserved amino acids (Lee *et al.*, 2001; Stock *et al.*, 2000). Output domains are highly variable and although most bind to DNA and control gene expression, other functions are known as well. Some response regulators bind to RNA, interact with proteins or are enzymes themselves (Galperin, 2010). In contrast to HKs, RRs are less well understood, particularly due to their highly-variable output domains. The diversity of output domains also hampers their identification from the DNA sequence and their further investigation.

HK and RR communicate via a phosphotransfer cascade. If the HK senses a signal, a conserved histidine amino acid in the kinase core becomes phosphorylated via a ATP molecule. Thereafter the phosphoryl group is transferred to the RR and binds to a conserved aspartate within the received domain. In a last step the RR is dephosphorylated often in an auto-dephosphorylation reaction. The de-phosphorylation can also be catalysed by the HK itself (bifunctional HK), which quickens the reaction (Kanamaru *et al.*, 1989; Keener and Kustu, 1988; Lois *et al.*, 1993). The time frame, in which the phosphorylated RR is stable, varies greatly between different systems. In general the majority of TCSs have a short phosphorylation / dephosphorylation time to process any input as fast as possible (Pazy *et al.*, 2009; Thomas *et al.*, 2008). The activation of an RR is controlled on the N-terminal regulatory domain, which is highly conserved among all RRs. Depending on the system RRs can trigger a "yes or no" response (Msadek, 1999) or regulate the extent of the reaction, according to the strength of the signal (Goulian, 2010; Russo and Silhavy, 1993).

The reaction steps in a basic two-component signal transduction system can be described in the following way:

1. Autophosphorylation: $HK - His + ATP \rightleftharpoons HK - His{\sim}P + ADP$
2. Phosphotransfer: $HK - His{\sim}P + RR - Asp \rightleftharpoons HK - His + RR - Asp{\sim}P$
3. Dephosphorylation: $RR - Asp{\sim}P + H_2O \rightleftharpoons RR - Asp + P_i$

Especially in bacteria, most TCSs consist of one HK, which controls one RR, but more complex systems exist as well. The very same RRs can become phosphorylated by different HKs and some HKs have control over two or more RRs (Goulian, 2010). In *B. subtilis* any one of 5 HKs (namely KinA, KinB, KinC, KinD, KinE) phosphorylates the same RR (SpoOF) to initiate sporulation (Jiang *et al.*, 2000; Stephenson and Hoch, 2002). A phosphorelay system is a common variation of a TCS, in which the phosphotransfer is extended over a histidine-containing phosphotransfer domain (HPt) (Hoch and Varughese, 2001). HKs of a phosphorelay system have a C-terminal receiver domain, which is similar to a RR. This receiver domain becomes

phosphorylated by the HK, which then enables a histidine phosphotransferase to transfer the phosphoryl group from the receiver domain to the RR. HKs that have an attached receiver domain are called heterologous or multi-domain kinases (Bijlsma and Groisman, 2003).

The cognate HK and RR are highly specific to each other. This specificity is mediated on a molecular level by small, compatible specificity residues in the alpha helix of the HK and RR (Casino *et al.*, 2009; Skerker *et al.*, 2008). The conserved parts of both molecules facilitate close approximation between the active site of the receiver domain and the kinase core of the HK to enable the interaction (Tzeng and Hoch, 1997). Characteristics of the specificity residues are their high variability and the abundance of hydrophobic, polar amino acids (Hoch and Varughese, 2001; Zapf *et al.*, 2000). The connection between the amino acids is generated by hydrophobic and Van Der Waals forces (Podgornaia and Laub, 2013). Small kinase inhibitors can block the specificity residues on the HK and inhibit phosphorylation of the cognate RR. In *B. subtilis* the two proteins Sda and KipI are well known as kinase inhibitors, which hinder the initiation of sporulation, through blocking of the phosphorylation of SpoOF (Bick *et al.*, 2009; Rowland *et al.*, 2004). If, despite their high specificity, a HK phosphorylates an RR belonging to a homolog TCS, the resulting process is referred to as "cross-talk" (Wanner, 1992). Cells prevent this cross-talk to ensure that different systems do not inhibit each other (Laub and Goulian, 2007). Cross-talk occurs very rarely *in vivo*, but can be provoked by molecular disturbances such as the absence of the cognate RR (Groban *et al.*, 2009; Siryaporn and Goulian, 2008) or the overproduction of the HK (Ninfa *et al.*, 1988). Naturally TCSs rely on the excess of RRs, which outcompetes the RRs of other TCSs and ensures the specificity of the pathway (Cai and Inouye, 2002; Miyashiro and Goulian, 2008). If the normal HK-to-RR ratio is disturbed, RRs of other TCSs can use the HK of the disrupted system to become phosphorylated. Bifunctional HKs show another strategy to avoid cross-talk by quickly dephosphorylating the RR, when it becomes wrongly phosphorylated (Huynh and Stewart, 2011; Siryaporn and Goulian, 2008). RRs can also become phosphorylated by small acetyl-phosphate molecules which act as phosphor donors (Lukat *et al.*, 1992; McCleary *et al.*, 1993; Wolfe, 2005). In addition to the described cognate RR-HK pairs, which are often encoded within the same operon, genes encoding "orphan" RRs and HKs are sometimes found alone (i.e. with no candidate gene encoding the other partner in close approximation).

On grounds of the absence of TCSs in mammals, TCSs could turn out to be good drug targets in bacteria (Cai *et al.*, 2005; Chauhan and Calderone, 2008; Gotoh *et al.*, 2010; Stephenson, 2000; Stephenson and Hoch, 2004). However, until now most attempts in

this direction have not proved promising, especially, because most TCSs are not essential (Schreiber *et al.*, 2009). To this day the group of identified essential TCSs is small and comprises mostly TCSs relevant for cell division and replication processes. Examples for essential TCSs are CenKR (Skerker *et al.*, 2005), belonging to *Caulobacter crescentus*, MtrBA of *Mycobacterium tuberculosis* (Zahrt and Deretic, 2000) and the WalRK system of *B. subtilis* and other Gram-positive bacteria.

### 1.3.2 Two-component systems in *B. subtilis*

There are 35 TCS in *B. subtilis*, which enable appropriate cellular response and adaptation to environmental conditions and nutrients (Fabret *et al.*, 1999; Ortet *et al.*, 2015). All but one are non-essential and the function of all but 11 TCS has been described. *B. subtilis* can access a wide range of different substrates and several TCSs are involved in sensing and responding to specific nutrients such as C4 dicarboxylic acids (DctRS), malate (MalRS), glutamine (GlnKL) and citrate (CitST) (Asai *et al.*, 2000; Satomura *et al.*, 2005; Tanaka *et al.*, 2003; Yamamoto *et al.*, 2000). A response to starvation, often linked to the bistability phenomenon, is regulated by three TCSs. DegSU regulates the expression of degradative enzymes and is also involved in competence, biofilm formation and swarming. ComPA modifies gene expression in response to cell density and regulates genetic competence and quorum sensing. Spo0A is a phosphorelay regulator that initiates sporulation and coordinates DNA replication (Molle, Fujita *et al.*, 2003). Furthermore, PhoPR regulates phosphate metabolism, upon phosphate limitation (Salzberg *et al.*, 2015) and ResDE aerobic and anaerobic respiration, upon oxygen limitation (Hartig and Jahn, 2012). A large number of TCSs is responsible for reacting to other stress signals. Response to cold shock is regulated by DesKR (Saita et al., 2015). Protein secretion stress is processed by CssRS (Darmon *et al.*, 2002). Response to antibiotic stress is activated by PsdRS, LiaRS, YxdJK and BceRS (Kesel *et al.*, 2013; Ohki *et al.*, 2003; Staron *et al.*, 2011; Zhang *et al.*, 2016). Some TCS systems, such as WalRK (Bisicchia *et al.*, 2007) are involved in cell wall metabolism, the maintenance of the cell surface by YvrGH (Serizawa *et al.*, 2014) and the regulation of autolysis rate by LytST.

Interestingly of the 35 TCSs in *B. subtilis* only one (WalRK) is essential for viability. Mutants in the other TCSs generally do not show a phenotype, unless exposed to the condition for which they are specific.

### 1.3.3 The WalRK two-component system

WalRK is the only essential TCS is *B. subtilis* and both the histidine kinase and the response regulator are essential for viability (Fabret and Hoch, 1998). The WalRK system is widely distributed among low G+C Gram-positive bacteria. The main group of bacteria possessing WalRK are families among the *Bacilli* class, particular the *Bacillaceae, Listeriaceae, Enterococcaceae, Steptococcaceae* and *Lactobacillaceae* family. The bacterium *Clostridium thermocellum* is the only member of the *Clostridia* class, which encodes a *walRK* operon. In most *Firmicutes*, including *Staphylococcus aureus* WalRK is also essential and therefore highly conserved (Albertini *et al.*, 2003; Fukuchi *et al.*, 2000; Martin *et al.*, 1999). However, in some species only the response regulator is essential (e.g. *Streptococci*). The essentiality of WalRK stems from its role as a major player in cell wall metabolism (Devine, 2012; Dubrac et al., 2008) and distinguishes it from most other TCSs that are confined to a specific environmental condition or a certain nutrient.

There are six genes, encoded within the WalRK operon of *B. subtilis*: *walR*, *walK*, *yycH*, *yycI*, *yycJ* and *yycK*. All genes of the operon are mainly expressed while cells undergo exponential growth and repressed from the moment they enter stationary phase (Fabret and Hoch, 1998; Howell *et al.*, 2003).

*WalR* and *walK* encode the RR and the HK respectively and are the only essential genes of the WalRK operon in *B. subtilis*. The next gene in the operon is *yycH* which has an inhibitory effect on the WalRK system. Its product is secreted and localized close to the cell membrane (Fukushima *et al.*, 2011). Knock-out mutants of *yycH* show a slower growth rate in the stationary phase, due to a cell wall defect (Szurmant *et al.*, 2005). The *yycI* gene encodes a similar protein, which interacts with YycH to control the expression of WalRK. Both proteins have an N-terminal transmembrane helix, that allows them to align with the HK and build a periplasmic complex (Szurmant, Mohan *et al.*, 2007). Deletion of *yycI* is likewise not lethal to *yycH*, but causes a growth defect. Santelli *et al.* (2007) analysed the structure of this complex and found that YycI forms a dimer, which is related to the C-terminal domains of YycH. The next gene in the operon encodes the YycJ protein, which directly or indirectly affects cell wall metabolism and is required for accurate coordination of cell division with DNA replication (Biller *et al.*, 2011). The last gene in the operon *yycK*, renamed *htrC,* encodes a control membrane protease similar to HtrA (Fabret and Hoch, 1998). A schematic representation of the WalRK operon in *B. subtilis* and in other species is shown in Figure 1.3 (a).

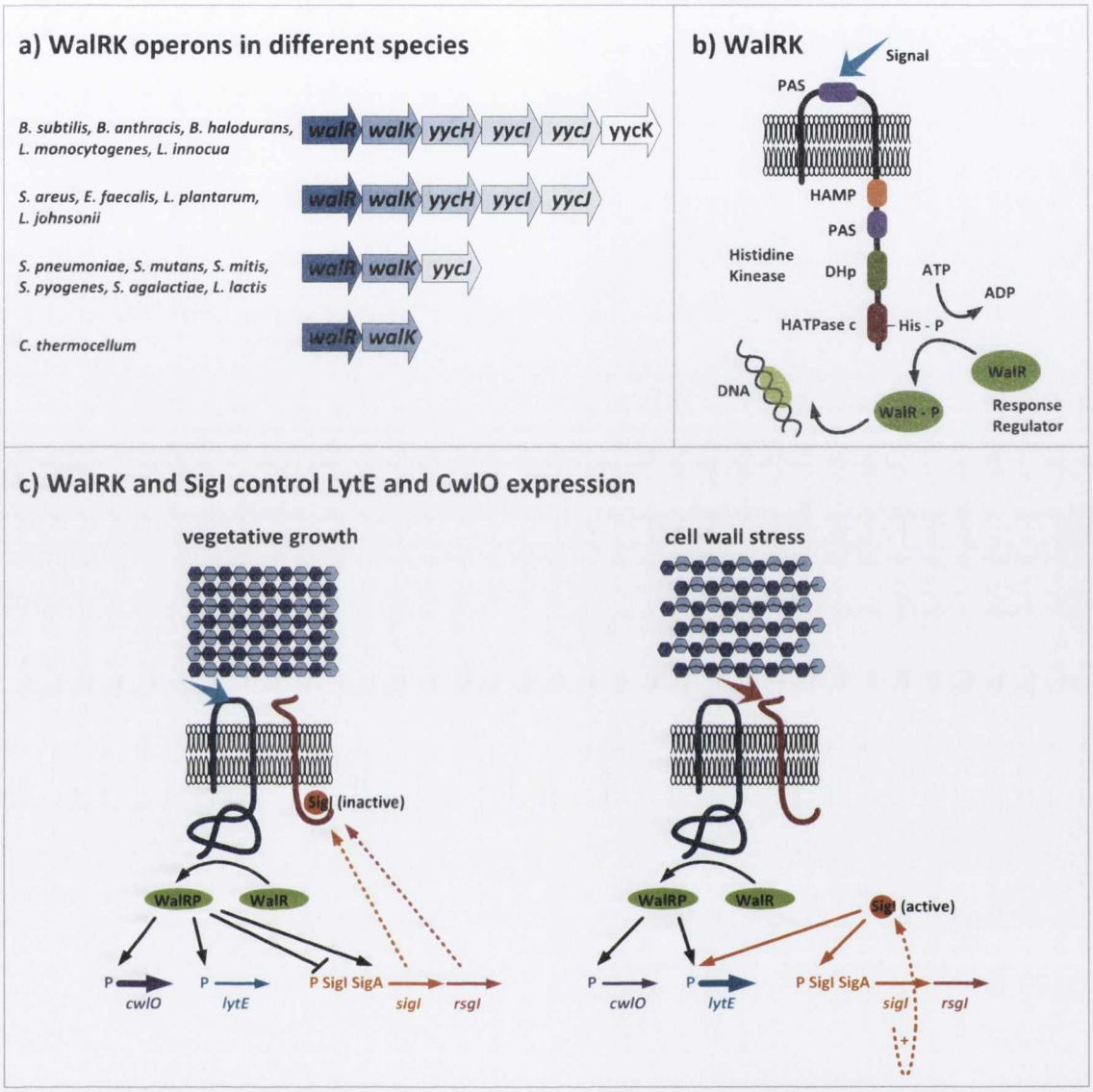



**Figure 1.3: The WalRK two-component system. a)** The WalRK operons of different species contain different set of genes: 6 genes (in *Bacillus subtilis, Bacillus anthracis, Bacillus halodurans, Listeria monocytogenes, Listeria innocua*), 5 genes (in *Staphylococcus aureus, Enterococcus faecalis, Lactobacillus plantarum, Lactobacillus johnsonii*), 3 genes (in *Streptococcus pneumoniae, Streptococcus mutans, Streptococcus mitis, Streptococcus pyogenes, Streptococcus agalactiae, Lactococcus lactis*) and 2 genes (in *Clostridium thermocellum*). **b)** Structure of the WalK Histidine Kinase and the WalR response regulator. Figure adapted from Dubrac *et al.* (2008). **c)** Interaction of WalRK and active SigI ensures expression of either CwlO or LytE in growing cells (left) and cells undergoing cell wall stress (right). Figure adapted from Salzberg *et al.* (2013).

The number of genes, encoded within the WalRK operon can be reduced in other species. Similar to the organization of the WalRK operon in *B. subtilis*, other *Bacillaceae* and *Listeriacea* also possess the full WalRK operon with six genes. However, the pathogen *S. aureus* and other bacteria, belonging to the *Staphylococcaceae* family have a reduced operon with five genes, which lack the *yycK* gene.

All *Streptococci* have a further reduced operon with only three genes (i.e. *walR, walK, yycJ). Clostridium thermocellum* is the sole member of the *Clostridia* class, which encodes a *walRK* operon and is further distinguished by the fact that its WalRK operon encodes only two genes: *walR* and *walK*.

The domain architecture of the WalK histidine kinase and the cognate WalR regulator is one of a typical TCS and well described in a review by Dubrac *et al.* (2008). WalRK is very similar to the PhoRP system, a TCS for the regulation of phosphate metabolism (Bisicchia et al., 2010; Howell et al., 2006). A schematic representation of the WalRK TCS of *B. subtilis* is presented in Figure 1.3 (b). The WalK histidine kinase is anchored in the cell membrane by two transmembrane sections, which connect the extra cytoplasmic loop that presumably contains the sensing domain. The intracellular part of the WalR histidine kinase holds the HAMP domain (Histidine kinases, Adenylyl cyclases, Methyl binding proteins, Phosphatases domain), the PAS domain (a PER-ARNT-SIM sensor domain), the HisKA domain (dimerisation and phosphoacceptor domain of histidine kinases) and the HATPase domain (Histidine kinase-like ATPases). The WalR response regulator consists of a CheY-homologous receiver domain and a DNA-binding output domain. The signal by which the WalK histidine kinase is activated is not known yet, but upon receipt of the unknown signal a conserved histidine amino acid in the kinase core of WalK becomes phosphorylated via using an ATP molecule. Thereafter the phosphoryl group is transferred to the WalR response regulator and binds to a conserved aspartate within the received domain, allowing the output domain to bind to DNA and regulate expression of the respective genes.

The WalRK TCS in *B. subtilis* controls cell wall metabolism by activating the expression of the CwlO and LytE D,L-endopeptidases and the YocH lytic transglycosylase. A schematic representation of the circumstances under which WalRK triggers the expression of LytE and CwlO is shown in Figure 1.3 (c). It also represses expression of the peptidoglycan deacetylase PdaC (encoded by *yjeA*) and the D,L-endopeptidase inhibitor IseA (Dubrac *et al.*, 2008; Howell *et al.*, 2003; Salzberg and Helmann, 2007). WalRK also activates the expression of YdjM, a protein which may be involved in cell wall metabolism. Another member of the WalR regulon is the actin homolog MreBH

which governs the localization of the LytE D,L-endopeptidase (Huang *et al.*, 2013). WalRK may also be involved in the expression of the cell division proteins FtsA and FtsZ (Fukuchi *et al.*, 2000), although this was not verified by the results of a ChiP on chip analysis of the WalR chromosomal binding loci performed by Salzberg *et al.* (2013). Instead a comprehensive WalR bindome was identified, which includes many genes that encode cell wall function proteins (e.g. *wapA, dacA, ggAB*). In cooperation with the SigI RsgI system, the WalRK TCS further fulfils the endopeptidase requirement, which states that the cell requires either the expression of LytE or CwlO for survival (Bisicchia *et al.*, 2007). While there is no shortage of one or the other during exponential growth, when the WalRK system is active and phosphorylated WalR binds to the SigA promoter of *lytE* and *cwlO* to initiate expression (Bisicchia *et al.*, 2007; Howell *et al.*, 2006), a different regulatory mechanism is required to uphold autolysin expression, when the WalRK system is turned off during heat stress and stationary phase. That is, because CwlO expression is completely dependent on WalRK and can only be controlled by its SigA promoter (Bisicchia *et al.*, 2007). In contrast LytE expression is not only regulated by a SigA promoter, but can in addition be controlled by a SigI promoter-like binding site upstream of the regulatory domain (Tseng *et al.*, 2011). The alternative sigma factor SigI is very important during heat survival and under stress conditions such as phosphate limitation and stationary phase (Tseng *et al.*, 2011; Zuber *et al.*, 2001). Its expression is controlled by active SigI through positive auto-regulation by a SigI promoter (Tseng and Shaw, 2008). An anti-sigma factor RsgI represses the activity of SigI and releases active SigI upon cell stress (Asai *et al.*, 2007). The SigI RsgI system and the WalRK system are indirectly coupled. WalR regulates SigI transcription by binding to sites overlapping the SigI promoter and upstream of the SigA promoter, presumably repressing expression of the SigI promoter and activating expression from the SigA promoter (Salzberg *et al.*, 2013), therefore interrupting the positive auto-regulation. Under stress conditions the repression of the SigI promoter is abolished and active SigI is released from RsgI, both supporting SigI expression through positive auto-regulation and initiating the transcription of MreBH and LytE, as well as BcrC (an undecaprenyl pyrophosphate phosphatase) and other SigI regulon genes. Overall the endopeptidase requirement can be translated into the statement that either an active WalRK TCS or active SigI factor is required for survival (Salzberg *et al.*, 2013). Both systems work together to ensure adequate levels of CwlO or LytE under different growth conditions.

In summary, the WalRK TCS controls the transcription of genes for autolysins and their inhibitors and manages cell wall remodelling during exponential growth, when frequent cell divisions, coupled with cell elongation are necessary. The peptidoglycan

metabolism is a complex interaction of different processes: the synthesis of new peptidoglycan material, the integration of new peptidoglycan building blocks in the existing cell wall, the degradation of the cell wall during sporulation and cell division and the recycling of old peptidoglycan. At the same time cell wall metabolism has to be adapted to environmental conditions such as nutrition and temperature which influence growth rate. The current model proposes that WalRK coordinates cylindrical wall synthesis with cell wall division, upon the receipt of an unknown signal, which is probably caused by an active cell wall synthesis (Bisicchia *et al.*, 2007; Cao *et al.*, 2002; Dubrac and Msadek, 2008)

## 1.4 Autolysins

### 1.4.1 Autolysins encoded by *B. subtilis*

A battery of autolytic enzymes function to remodel the cell wall by cleaving the peptidoglycan matrix in order to allow insertion of new peptidoglycan units during the elongation process or to degrade the cell wall during sporulation or cell division process (Ghuysen and Hakenbeck, 1994; Lee and Huang, 2013). Autolysins are also necessary to "trim" the cell wall of pathogenic bacteria, thereby removing outer fragments of peptidoglycan that otherwise would make recognition of the pathogen by the host immune system more likely (Atilano *et al.*, 2014). All bacteria that have a cell wall made of peptidoglycan possess autolysins with glycolytic and peptidolytic activities. The spectrum of enzymatic activities found among this battery of enzymes is sufficient to cleave the chemical bonds found in peptidoglycan. A schematic representation of the cleavage sites of autolytic enzymes is shown in Figure 1.4 (a). Autolysins are classified according to their cleavage specificity into four major groups: muramidases and glucosaminindases cleave the glycosidic linkages between the NAG and NAM residues in the sugar backbone, amidases (N-acetylmuramoyl-l-alanine amidases) cleave the linkage between the sugar and peptide side chain, while endopeptidases cleave the bonds in the peptide side chains (Vollmer, Joris et al., 2008). Similarly to muramidases, lytic transglycosylases cleave the same bonds, but produce anhydromuropeptides as degradation products. Peptidoglycan hydrolases that cannot on their own cause cell lysis are not classified as autolysins (Foster, 1994). These four major groups of autolysins in *B. subtilis* can further be split into subgroups on the basis of similarity of their amino acid sequence (Smith *et al.*, 2000).

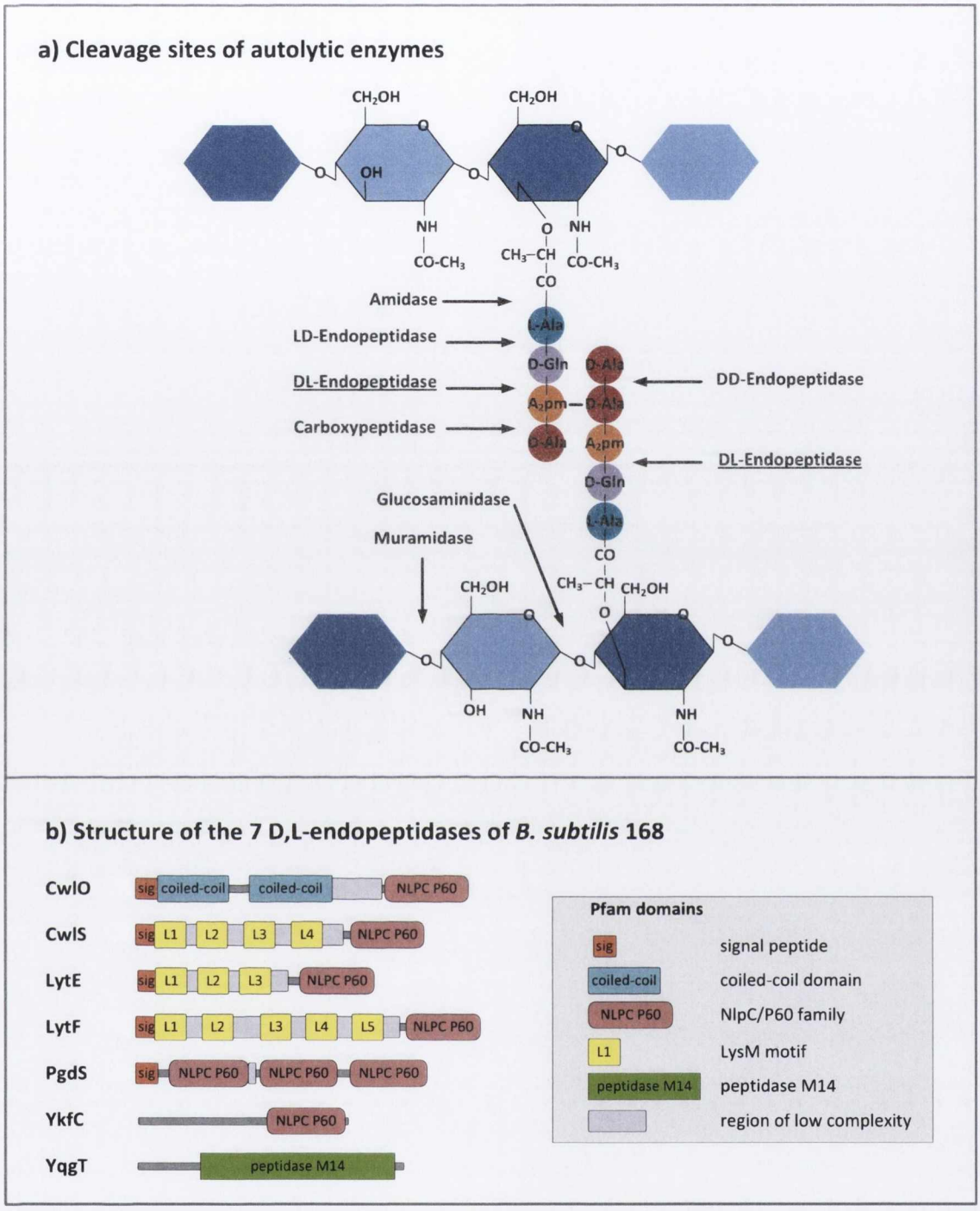



**Figure 1.4: Autolytic enzyme cleavage sites and structures of the seven D,L-endopeptidases in *B. subtilis*.** **a)** Enzymatic activities of different autolysins and the localization of the cleavage site. Figure adapted from Fukushima et al. (2006). **b)** The protein structures of the seven D,L-endopeptidases in *B. subtilis*. Schematic structures were adapted from the Pfam database and are approximately to scale.

Amidases can be subdivided into the LytC amidase family (LytC, YqiI, CwlC, CwlD and YrvJ) and the LytD glucosaminindases family (LytD). Several more amidases are similar to the amidase XlyA of the *B. subtilis* prophage PBSX (XlyA, XlyB, CwlH, CwlA, BlyA, YdhD and YocH). There are two types of endopeptidases: DL-endopeptidases and LD-endopeptidases. DL-endopeptidases cleave the peptide bond between d-Glu and $A_2$pm and are further subdivided into two groups.

The DL-endopeptidase I family is only made up of YqgT, while the DL-endopeptidase II family holds the remaining six D,L-endopeptidases (LytF, CwlS, LytE, CwlO, PgdS and YkfC). CwlK is an L,D-endopeptidase that cleaves the peptide bond between L-Ala and D-Glu. Five more groups were identified by their similarity to autolysins in other species. The Lysostaphin family (CwlP, LytH, SpoIIQ) shows similarity to lysostaphin from *Staphylococcus simulans*. The germination-specific lytic enzymes (SleB, YkvT and CwlJ) have similarity with SleB of *Bacillus cereus*. The LrgB family holds three homologs enzymes (YsbB, YwbG and YxaC) to LrgB an autolysin of *Staphylococcus aureus*. Furthermore, two homologous enzymes were identified to Slt70 of *E. coli* (CwlQ and YomI) and the LytG and YpBE enzymes of *B. subtilis* are homologous to muramidases found in the *enterococcus* species.

Autolytic enzymes are required for cell growth (Vollmer, 2012), because they play an important role in the process of cell wall synthesis by cleaving peptidoglycan bonds to allow the insertion of new peptidoglycan sub-units (Lee and Huang, 2013). Due to their destructive nature, autolytic enzymes have to be controlled tightly. Excessive cleavage of the peptidoglycan cell wall might otherwise lead to the formation of pores that would result in the rupturing of the cell envelope caused by the high internal tugor pressure. For this reasons it is thought that cell wall degradation and cell wall synthesis have to be coordinated in a way that new peptidoglycan material is available before old cell wall bonds are degraded. Indeed, cells in which the synthesis of peptidoglycan is impaired, undergo cell lysis, but not if they also have a defect in their autolytic enzyme activity (Goodwell *et al.*, 1976). However, although tight regulation of autolytic enzymatic activity is undoubtedly required, a study on *E. coli* suggests that the cell wall can tolerate more peptidoglycan damage than previously thought which would allow for some fluctuation in the autolytic activity in this bacterium (Huang *et al.*, 2008). The investigation of autolytic enzymes is one requirement for a thorough understanding of the complexity of the cell wall metabolism, but autolytic enzymes might also be useful on their own. Their specificity might allow them to be deployed as novel weapons against antibiotic resistant pathogenic bacteria (Fenton *et al.*, 2010; Gerstmans *et al.*, 2016; Szweda *et al.*, 2012; Wittekind and Schuch, 2016).

Interestingly most bacteria encode several redundant autolysins that are controlled by different regulatory systems. This makes an investigation of the function and regulation of an individual enzyme difficult, because often the lack of enzymatic activity of an individual enzyme can be offset by another one.

### 1.4.2 D,L-endopeptidases of *B. subtilis*

The *B. subtilis* genome encodes approximately 35 autolysins (Smith *et al.*, 2000). A crucial feature that is central to this area of study, is that there is very significant redundancy in the activities of these 35 autolytic enzymes. For example, there are five amidases of the LytC amidase family (LytC, YqiI, CwlC, CwlD, YrvJ) (Smith *et al.*, 2000). There are also seven D,L-endopeptidases that cleave the peptide bond between D-glutamate and diaminopimelate in the side chain (CwlO, CwlS, LytE, LytF, PgdS, YkfC, YqgT) (Ishikawa *et al.*, 1998). These enzymes are critical for expansion of the cell wall: in cleaving the peptide side chains, they allow the glycan chains to slide against one another and therefore allow expansion of the peptidoglycan. It is clear that the expression and activity of these enzymes must be tightly regulated to allow sufficient peptidoglycan hydrolysis for cell wall expansion, but not so much that would trigger cell lysis. One or other of the CwlO and LytE D,L-endopeptidases have been found to be required for cell wall expansion and growth (Bisicchia *et al.*, 2007). The following paragraphs describe the seven D,L-endopeptidases of *B. subtilis* in more detail. Their structures are illustrated schematically in Figure 1.4 (b) (see page 23).

#### 1.4.2.1 CwlS

CwlS was identified as an important autolysin for cell separation by Fukushima et al. (2006). The *cwlS* gene is localized in a single gene operon. It is part of several regulons (Abh regulon, AbrB regulon, CcpA regulon, SigD regulon, SigH regulon). CwlS expression is initiated by SigD and SigH, during vegetative growth (Hashimoto et al., 2012). CwlS is unlikely to be inhibited by IseA *in vivo*, which negatively regulates other cell separation enzymes at the post-translational level (e.g. LytE and LytF) (Yamamoto, Hashimoto et al., 2008), but is repressed by glucose (Blencke et al., 2003). The transcription of *cwlS* is repressed by the regulator proteins Abh and AbrB, which function synergistically (Chumsakul et al., 2011). The C-terminal domain of the CwlS protein consists of the enzymatic D,L-endopeptidase domain (Hashimoto et al., 2012), which hydrolyses the gamma-D-glutamyl bonds of meso-diaminopimelic acid

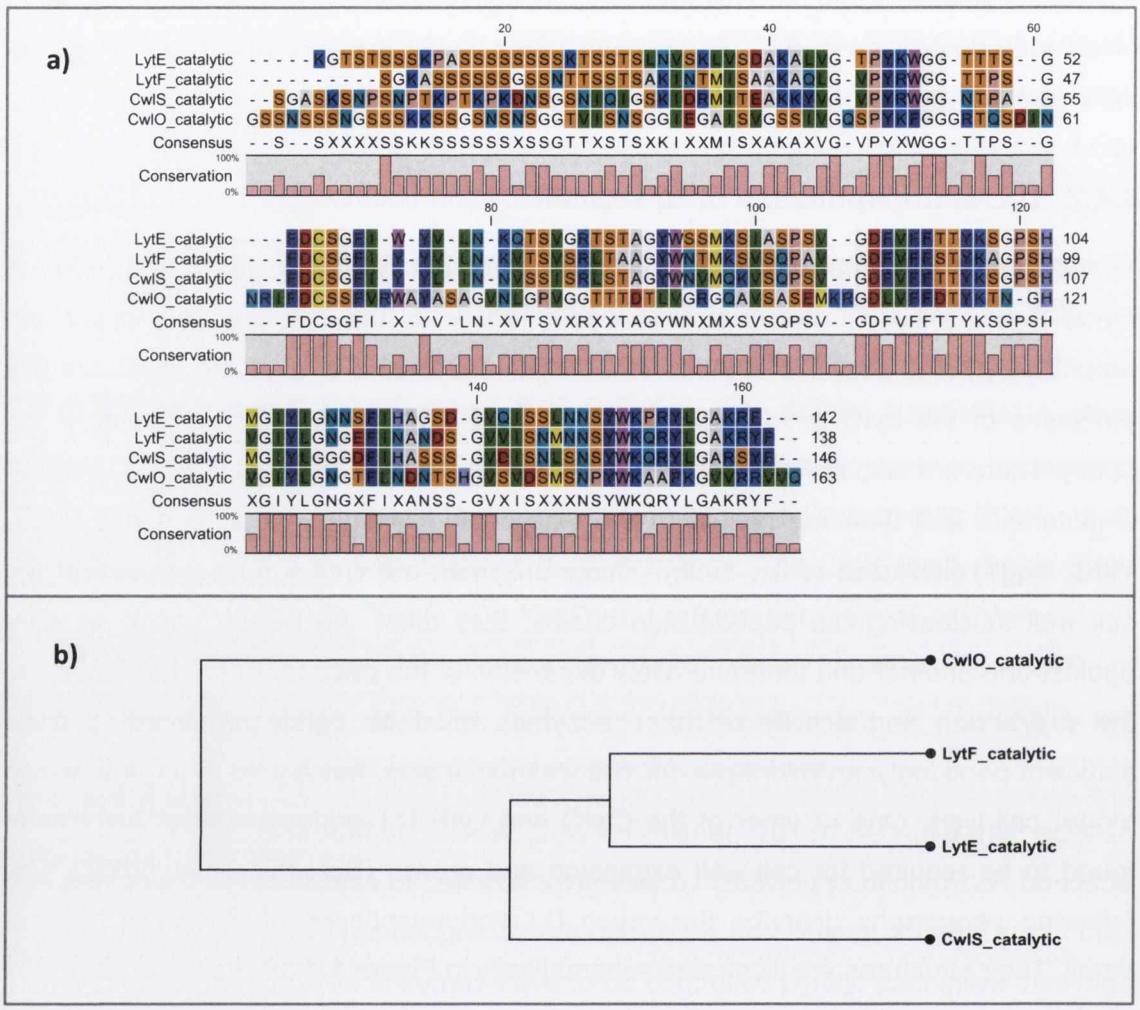



**Figure 1.5: Alignment and phylogenetic tree of the catalytic domains of LytE, CwlS, LytF and CwlO.** **a)** The amino acid sequence alignment was constructed with Clustal Omega. **b)** The phylogenetic tree was constructed with the UPGMA algorithm from the amino acid sequence alignment shown in (a).

The C-terminal catalytic domain has a strong sequence similarity to the D,L-endopeptidase domains of LytE, LytF and CwlO (see Figure 1.5).

The N-terminus of CwlS holds four LysM motifs (see Figure 1.4, page 23) (Buist *et al.*, 2008; Wong *et al.*, 2014). CwlS has a signal peptide, is secreted and localized to the cell membrane, cell septa and the poles (Hashimoto et al., 2012). Single *cwlS* mutants do not exhibit a phenotype, suggesting that absence of CwlS is compensated by the activity of other D,L-endopeptidase enzymes (Fukushima *et al.*, 2006).

#### 1.4.2.2 CwlO

CwlO was first described as a peptidoglycan hydrolase by Yamaguchi *et al.* (2004) and is involved in cell elongation. CwlO is part of a single gene operon and member of the WalR regulon. The regulation of CwlO expression and activity is well described by Noone et al. (2014). The TCS WalRK is the sole activator for CwlO expression, which is therefore only expressed during exponential growth, when the WalRK system is active. When cells enter stationary phase or experience cell wall stress, genes belonging to the WalR regulon are no longer expressed, which leads to a drop in CwlO expression (Salzberg *et al.*, 2013). Regulation of CwlO expression is also achieved at the transcriptional level. A 187 bp long leader sequence upstream of the *cwlO* open reading frame, causes a high instability of the *cwlO* transcript, due to a cleavage by RNase Y (Noone *et al.*, 2014). This allows CwlO expression to be highly responsive to the activating signal: in the absence of a signal, the existing mRNA is rapidly degraded. CwlO activity requires the protein to be associated with the membrane located FtsE-FtsX complex and ATP hydrolysis (Dominguez-Cuevas *et al.*, 2013; Meisner *et al.*, 2013; Yamaguchi *et al.*, 2004). It is also associated with the actin-like protein Mbl (Dominguez-Cuevas *et al.*, 2013). These regulatory features ensure that the CwlO enzymatic activity is restricted to the inner side of the cell wall. Another distinctive feature is the existence of an anti-sense transcript, with unknown function, which is induced by a SigB promoter during the stationary phase (Noone *et al.*, 2014). The C-terminal domain of the CwlO protein consists of the enzymatic D,L-endopeptidase domain, with strong sequence similarity to the D,L-endopeptidase domains of LytE, LytF and CwlS. The N-terminal domain has a coiled-coil structure, which is presumably required for the interaction with FtsEX (see Figure 1.4 b, page 23). The CwlO protein has a signal peptide and is secreted. CwlO is localized at the lateral sidewall of the cell (in the cylinder) (Hashimoto *et al.*, 2012; Voigt *et al.*, 2009). CwlO mutants display slightly shorter and thicker cells, a phenotype that is rescued by addition of $Mg^{2+}$ (Dominguez-Cuevas *et al.*, 2013; Meisner *et al.*, 2013). Combined mutants in CwlO and LytE are synthetically lethal (Bisicchia *et al.*, 2007; Hashimoto *et al.*, 2012).

### 1.4.2.3 LytE

LytE is a major autolysin, with DL-endopeptidase activity and was identified by Margot *et al.* (1998). LytE is important for cell elongation and separation and contributes to survival under heat stress. It is localized in a single gene operon and member of several regulons (SigH regulon, SigI regulon, Spo0A regulon, WalR regulon). LytE expression is activated by the WalRK system during exponential growth, when the WalRK system is active. However, LytE expression can also be triggered by active SigI. The expression of LytE is normally regulated by a SigA and a SigH promoter (Hashimoto *et al.*, 2012), but under stress conditions (i.e. heat stress), when the RsgI system is activated, active SigI is released, which in turn triggers LytE expression from a SigI promoter-like sequence upstream of the SigA promoter (Tseng *et al.*, 2011). LytE is on this account the main autolysin produced in stationary phase and under stress conditions (Tseng *et al.*, 2011; Tseng and Shaw, 2008). The C-terminal domain of the LytE protein consist of the enzymatic D,L-endopeptidase domain, with strong sequence similarity to the D,L-endopeptidase domains of CwlS, LytF and CwlO (see Figure 1.5, page 26). The N-terminal domain of LytE consists of three LysM motifs which are important for the interaction with the peptidoglycan substrate (see Figure 1.4, page 23). The LytE protein has a signal peptide, is secreted and localized at cell separation sites and the cell poles (Scheffers and Pinho, 2005). It shows an interaction with the membrane-associated actin-like protein MreBH that directs the localization of LytE to the cylindrical part of the cell (Carballido-Lopez *et al.*, 2006). Resistance to ß-lactam in a SigM mutant can be restored by inactivation of LytE. It was suggested that this effect is due to an decreased lysis behaviour in the affected cells (Luo and Helmann, 2012). Single *lytE* mutants are displaying a growth defect when grown under heat stress (Tseng *et al.*, 2011). Combined mutants in LytE and CwlO are synthetic lethal (Bisicchia *et al.*, 2007; Hashimoto *et al.*, 2012). For the same reason combined mutants in LytE and FtsE and FtsX are also synthetically lethal (since CwlO requires the activation by the FtsE-FtsX complex) (Dominguez-Cuevas *et al.*, 2013; Meisner *et al.*, 2013). Combined mutants in LytE and LytF have an increased cell separation defect (Meisner *et al.*, 2013).

### 1.4.2.4 LytF

Another D,L-endopeptidase of *B. subtilis* is LytF which was also identified by Margot *et al.* (1999). LytF is a major, non-essential autolysin which is important for cell separation. It is part of a single gene operon and member of the SigD regulon and SlrR regulon. Expression of LytF is exclusively triggered by the sigma factor SigD, during the late exponential phase (Ohnishi *et al.*, 1999). LytF activity can be inhibited by

interaction with IseA, which binds to its catalytic domain (Yamamoto, Miyake *et al.*, 2008). LytF is also negatively controlled by SlrR, which acts in complex with SinR (Chai et al., 2010). The LytF protein consist of a C-terminal enzymatic domain, which exhibits the D,L-endopeptidase activity (Hashimoto *et al.*, 2012), as well as a N-terminal domain with five N-acetylglucosamine-polymer binding LysM motifs (see Figure 1.4, page 23). Its C-terminal catalytic domains has strong sequence similarity to the D,L-endopeptidase domains of LytE, CwlS and CwlO (see Figure 1.5, page 26). LytF has a signal peptide, is secreted into the media and localizes to the cell septa and poles on the vegetative cell surface, if the lipoteichoic acid synthase LtaS is present (Kiriyama *et al.*, 2014). It is suggested that the presence of WTA is important for the correct localization of LytF (Kiriyama *et al.*, 2014; Yamamoto, Miyake *et al.*, 2008). LytF mutants have cell separation defects, resulting in long chains of cells, an effect which is amplified in combination with a LytE mutant (Meisner *et al.*, 2013).

#### 1.4.2.5 PgdS

The PgdS autolysin was first described as an extracellular protein by Hirose *et al.* (2000), although its function remained unknown until Suzuki and Tahara (2003) defined it as a novel enzyme involved in the degradation of polyglutamic acid. It is part of a single gene operon and member of the SigD regulon, as shown by Serizawa *et al.* (2004) in an extensive analysis of SigD-regulated genes in *B. subtilis*. The PgdS protein belongs to the nlpC/p60 family (see Figure 1.4, page 23) has a signal sequence and is secreted (Voigt *et al.*, 2009). The disruption of pgdS does not affect cell morphology or growth, but results in a strain with higher yield of poly ($\gamma$-glutamic acid), since it is no longer degraded by PgdS (Scoffone *et al.*, 2013). This effect could be of interest for biotechnological applications (Candela and Fouet, 2006).

#### 1.4.2.6 YkfC

YkfC is one of the least investigated D,L-endopeptidases of *B. subtilis*. YkfC is involved in cell wall metabolism, by cleaving L-Ala-γ-D-Glu dipeptides from cell wall peptides (Xu *et al.*, 2010). It is member of the CodY regulon and repressed by CodY, a GTP-binding repressor of several genes during exponential growth (Molle, Nakaura *et al.*, 2003), indicating that YkfC is expressed during cell wall stress or sporulation. The YkfC protein belongs to the nlpC/p60 family and has no signal peptide (see Figure 1.4, page 23). Its localization has not been demonstrated and no phenotype of *ykfC* mutants has been described so far.

#### 1.4.2.7 YqgT

YqgT is a poorly described autolysin of *B. subtilis*. The *yqgT* gene was identified by Mizuno *et al.* (1996), who described it as being similar to the endopeptidase I in *B. sphaericus*. Park and Uehara (2008) classified YqgT as ortholog of the MpaA $\gamma$-D-Glu-Dap amidase of *E. coli*. A study about the bacterium *Epulopiscium* (Miller et al., 2012) listed YqgT as one of the core sporulation genes identified in *B. subtilis*. Its regulation remains unknown and it has not been assigned to a certain regulon. The YqgT protein belongs to the peptidase M14 family (see Figure 1.4, page 23) and has both endopeptidase and carboxypeptidase activities (Vollmer, Joris *et al.*, 2008). YqgT has no signal peptide and Tjalsma *et al.* (2000), who conducted an extensive genome-based survey of the secretome, stated that YqgT is localized in the cytoplasma. No phenotype of a YqgT mutant has been described so far.

### 1.4.3 Essentiality and synthetic lethality of autolysins

The *B. subtilis* genome belongs to the best understood bacterial genomes and contains approximately 4,100 genes. A first project in 2003, identified 271 of these genes as essential (Albertini *et al.*, 2003), whereas a more recent study lists no more than 261 essential genes (Commichau *et al.*, 2013). Interestingly none of the seven D,L-endopeptidases in *B. subtilis* is essential for viability and nearly all of them can be deleted, without generating an atypical morphological or growth phenotype. Thus, due to redundancy in these enzymatic activities a disruption of one or more autolysins genes can be normally compensated by the activity of others. However, as discovered by Bisicchia *et al.* (2007), *lytE* and *cwlO* that both encode D,L-endopeptidases are synthetically lethal, i.e. the cell requires one or other of these enzymes for viability. This indicates that they perform an overlapping essential function for the cell, probably in elongation of the cell wall during growth, as reported by Hashimoto *et al.* (2012). It is important to note that although LytE and CwlO possess similar C-terminal domains (the region that encodes the D,L-endopeptidase activity), they have completely different N-terminal domains. While the N-terminal domain of CwlO is necessary for the interaction with the FtsEX ATP transporter (Dominguez-Cuevas *et al.*, 2013; Meisner *et al.*, 2013; Yamaguchi *et al.*, 2004), the N-terminal domain of LytE is composed of three LysM motifs, which bind to peptidoglycan (Ishikawa *et al.*, 1998). Considering these differences it is even more astonishing that they are able to substitute for each other.

The synthetic lethality of endopeptidase-type autolysins, first shown by Bisicchia *et al.* (2007) has since been demonstrated in *E. coli* and in *Vibrio cholera*. Dörr *et al.* (2013) showed that of three putative periplasmic proteins of *Vibrio cholera* (ShyA, ShyB, and

ShyC) that bind peptidoglycan and belong to the M23 peptidase family, ShyA and ShyC are synthetically lethal. It was further shown that ShyA and ShyC are important for cell elongation and that ShyA is a D,D-endopeptidase. Similarly Singh *et al.* (2012) discovered that *E. coli* requires at least one of three redundant murein endopeptidases (Spr, YdhO and YebA) for growth.

In general the phenomenon of synthetic lethality is thought to be related to genetic robustness and buffering (Nijman, 2011). Functional redundancy assures survival by providing a "back-up" system. Loss or mutation of genes or the inhibition of pathways by environmental substances such as toxins can so be circumvented.

## 1.5 The LysM motif

The LysM motif is found in the N-terminal region of three (CwlS, LytE and LytF) of the seven D,L-endopeptidases encoded in the *B. subtilis* genome (see Figure 1.4, page 23). The LysM motif is a protein domain that binds to the N-acetylglucosamino moiety of peptidoglycan or chitin (for reviews see Buist *et al.*, 2008 and Visweswaran *et al.*, 2011). The LysM motif was first identified as a repeat motif in bacterial lysins, which is why it was named LysM motif (Ponting *et al.*, 1999). The motif consists of 44 to 65 amino acids and proteins can have one or more of these motifs. Proteins with up to 12 LysM motifs have been identified (Zhang *et al.*, 2009). LysM motifs can be located in the N-terminal as well as the C-terminal domain of a protein and sometimes also in the central part, where they often connect two catalytic sites. The localization of the LysM motif within the protein is presumably important for the correct positioning of the enzymatic domain to the substrate. The secondary structure of the LysM motif is one of beta-alpha-alpha-beta confirmation, in which the two alpha helixes are on the same side of an anti-parallel beta sheet. The predicted protein structure of the three LysM motifs of LytE is shown in Figure 1.6 (a, b, c). So far only two secondary structure of LysM motifs have been investigated experimentally (Bateman and Bycroft, 2000; Chen *et al.*, 2008; Shao *et al.*, 2009). The sequence of LysM motifs is conserved (Buist *et al.*, 2008) and an alignment of the amino acid sequence of all available LysM motifs of *B. subtilis* is shown in Figure 1.6 (d). As can be seen from this sequence alignment, the LysM motif seems to be well conserved over the first fifteen and somewhat less over the last six amino acids residues. The middle part is poorly conserved with the exception of an asparagine residue at position 32.

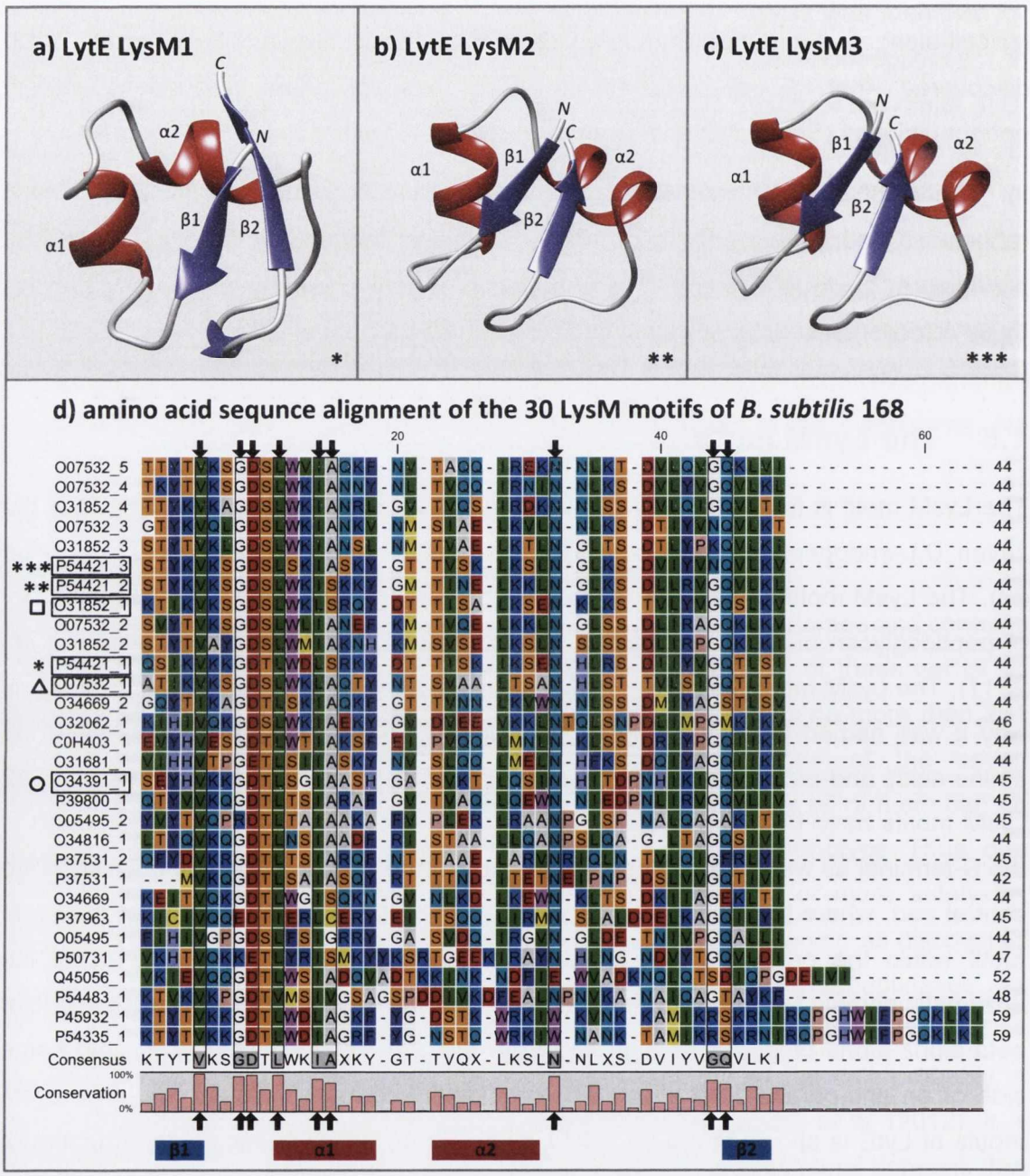



**Figure 1.6: Structures of the three LysM motifs of the LytE D,L-endopeptidase and alignment of all thirty LysM motifs of *B. subtilis*.** Structure of the **a)** first ($NH_2$-terminal), **b)** second and **c)** third LysM motif of LytE. LysM motif structures were predicted from the amino acid sequence with the I-TASSER server for protein structure prediction. The confidence scores estimating the quality of the model (range 2 to -5, with a higher score predicted a higher quality) are 1.09, 1.24 and 1.20. Secondary structures are identified and coloured (α helix = red, β strand = blue). **d)** Amino acid alignment of all thirty LysM motifs identified in *B. subtilis 168*. LysM motifs are identified with their UniProt accession number and _number, referring to the position of the LysM motif within the protein (i.e. P54421_3 is the third LysM motif of the LytE enzyme). Conserved amino acid residues are indicated by arrows and the regions of the secondary structures (determined from the predicted secondary structures of the three LysM motifs of LytE and of YkuD) are indicated by blue and red boxes. The three LytE LysM motifs are indicated by one, two and three asterisks respectively. The N-terminal LysM motif of CwlS is highlighted by a rectangle, the N-terminal LysM motif of LytF by a triangle and the single LysM motif of XlyB by a circle.

The LysM motif is ancient and widely spread in prokaryotes, eukaryotes and bacteriophages, but has not been detected in archaea. In eukaryotes the motif is often found in plant sensor kinases, which use the structure to sense bacteria and fungi and to recognize them either as symbionts or parasites. These plant kinase proteins are of great interest and many studies have been conducted to understand how exactly they contribute to the special interactions between plants and microorganisms (Antolín-Llovera *et al.*, 2012; Greeff *et al.*, 2012; Gust *et al.*, 2012; Oldroyd, 2013; Spaink, 2004). In prokaryotes the majority of proteins containing LysM motifs, are secreted and bind to the cell envelope.

Besides truly secreted proteins there is a range of proteins that bind to the cell wall in a non-covalent manner as outer-membrane proteins or lipoproteins that also contain LysM motifs. Bacterial LysM motifs are especially common in autolysins whose substrate is peptidoglycan, but are also found in peptidases, esterases, reductases and nucleotidases. Some pathogenic bacteria possess LysM containing proteins that can act as virulence factors for example the N-acetylmuramyl-L-alanine amidase Sle1 of *S. aureus* that has three LysM motifs (Kajimura *et al.*, 2005).

Although the LysM motif amino acid sequence is conserved, the number of motifs is not. If proteins have more than one LysM motif, the motifs are spaced by linker sequences which have a high number of Ser, Thr and Asp or Pro residues, but no sequence similarity and can vary in length (Buist *et al.*, 1995; Ohnuma *et al.*, 2008). It is unknown if the number of copies of the LysM motifs is optimized for a specific substrate or if it is in direct relation with the binding affinity *in vivo*. Several studies have shown that a total loss of LysM motifs greatly decreases the peptidoglycan-binding capacity as well as the enzyme activity *in vitro* (Eckert *et al.*, 2006; Steen, Buist *et al.*, 2005; Wong *et al.*, 2014), while only one investigation proved that a total number of three LysM motifs in the AcmA autolysin of *L. lactis* is indeed optimal for the binding to cell wall *in vitro* (Steen, Buist *et al.*, 2005). An *in vitro* study by Wong *et al.* (2014) came to the conclusion that the four LysM motifs of the D,L-endopeptidase CwlS in *B. subtilis* have different binding capacities and contribute additionally to peptidoglycan-binding. A similar conclusion was drawn by Mesnage *et al.* (2014), who investigated the additive effect of six LysM motifs to the binding capacity of AtlA (an autolysin of the bacterial pathogen *Enterococcus faecalis*) *in vitro* and who suggested that the individual LysM motifs do not form organized quaternary structures. Furthermore, it was observed that LysM motifs bind more easily to peptidoglycan when it is organized in long glycan chains as it is in the cell wall, compared to short oligosaccharides. It was also suggested that a gradual loss of LysM motifs caused by degradation, could be one way to control enzymatic activity (Eckert *et al.*, 2006; Steen, Palumbo *et al.*, 2005). Overall

the binding of LysM motifs as well as their structure and the importance of the number of LysM motifs is not fully understood yet.

The evolutionary origin of the LysM motif is unclear, but two hypotheses have been offered, to explain why so far no LysM motif has been detected in archaea. First it is possible that LysM motifs were existent in the common ancestor of bacteria, archaea and eukaryotes and that they were subsequently lost in ancestral archaea. Another possibility is that LysM motifs originally arose in bacteria and were at an early evolutionary time point transferred from bacteria to eukaryotes through horizontal gene transfer (or vice-versa) (Ponting *et al.*, 1999). Interestingly although LysM motif proteins make one of the smaller groups of PFAM domain families, they come in a great variety of domain architectures (Zhang *et al.*, 2009). Zhang *et al.* (2009) state that there is only one simple core LysM motif architecture that is conserved throughout all kingdoms, while each kingdom has specific architectures that are not shared with others. This suggests that LysM motif proteins have undergone a great diversification and distinct evolution while at the same time the LysM motif itself remained essentially unchanged as a peptidoglycan or chitin binding structure.

Interestingly, it has been shown that the binding of LysM motif proteins in Gram-positive bacteria only occurs at specific sites, although the whole cell envelope is made up of peptidoglycan. Likewise the CwlS (4 LysM motifs) and the LytF D,L-endopeptidases of *B. subtilis* (5 LysM motifs) are mainly localized at the cell poles and septa (Hashimoto *et al.*, 2012; Kiriyama *et al.*, 2014). However, site-specific binding is not necessarily conferred by LysM motifs, but can be due to protein-protein interactions or to properties of the cell wall. Indeed it has been found that binding can be hindered by other cell wall components such as lipoteichoic acid or by modification of the peptidoglycan for example through acetylation (Andre *et al.*, 2008, Frankel and Schneewind, 2012, 2012; Steen *et al.*, 2003; Veiga *et al.*, 2007). It is known that LytE activity (another D,L-endopeptidases of *B. subtilis* with 3 LysM motifs) requires interaction with the membrane-associated MreBH protein, which determines the extracellular localization of LytE (Carballido-Lopez *et al.*, 2006). Another interesting question is what determines whether secreted LysM motif proteins attach to the inner or the outer side of the cell wall. It is also unclear if LysM-containing enzymes are active at the precise location at which they are secreted or if the enzymatic activity occurs at a different site and what influence LysM motifs have on enzyme activity.

Current research activity not only focuses on an understanding of the molecular mechanisms that confer LysM motif binding, but also on to how this specific property could be deployed in biotechnological applications. Binding of LysM motifs to

peptidoglycan or chitin is fast, robust and strong and could therefore be utilized to bind specifically designed LysM containing proteins to the surface of Gram-positive bacteria or to particles coated with peptidoglycan or chitin. A review of possible applications is given by Visweswaran *et al.* (2014). The following options are currently under development: LysM motifs could be used to bind antigens to the surface of non-genetically modified Gram-positive bacteria for immunization purposes as exploited by Bosma *et al.* (2006) and Visweswaran *et al.* (2014). Similarly LysM motifs could prove useful for cell surface displays (Okano *et al.*, 2008). Another possibility is to exploit the LysM motif binding capacity for purification purposes or for cell immobilisation (Simsek *et al.*, 2013). LysM motif fusion proteins could also be used to immuno-stain or detect fungi or bacteria (Frankel and Schneewind, 2012; Shao *et al.*, 2012; Xu *et al.*, 2011).

## 1.6 Scope of this study

This study set out to establish the role of D,L-endopeptidases in the elongation of the cell wall of *B. subtilis*. This is complicated by the redundancy in D,L-endopeptidases in *B. subtilis* with seven D,L-endopeptidases in total. I propose that the fact that these enzymes are located outside the cell makes it difficult to control their activities through activation / inhibition. Therefore, the cell may circumvent this problem by producing multiple autolysins, each with a slightly altered peptidoglycan-binding or enzymatic activity. I propose that producing just the right amount of autolytic activity, i.e. enough to remodel the cell wall, but not so much to cause cell lysis, requires a fine-tuning of the autolysin function, which is achieved by a wide variety of expression patterns, protein interactions and peptidoglycan-binding motifs, which results in different cellular location, peptidoglycan-binding affinity or enzymatic activity of the autolytic enzymes.

The synthetic lethality of the CwlO and LytE D,L-endopeptidases presents an experimental system that allows examination of the precise requirements for the cellular functioning and activity of these two enzymes. Moreover, the synthetic lethality of the CwlO / LytE autolysins allows investigation of whether any of the other five D,L-endopeptidase encoded in the genome can complement their activity. Thus, it can be investigated which D,L-endopeptidase can support growth and viability when expressed under the control of the *cwlO* promoter. The variance in LysM motifs among these D,L-endopeptidases can also be investigated: i.e. can LytE (three LysM motifs) support growth with three LysM motifs from another endopeptidase enzyme or when its own three LysM motifs are permuted. Several properties of these heterologous enzymes can be assessed - enzymatic activity, ability to bind peptidoglycan, cellular

location and distribution of enzymes, in addition to measuring their ability to complement CwlO function.

I hypothesize that different cell wall binding domains (e.g. as in CwlO and LytE) cause a different distribution of autolytic enzymes within the cell wall and that different numbers of LysM peptidoglycan-binding domains facilitate a differential affinity for peptidoglycan. I also expect to gain a better understanding of the regulation of these enzymes and how they are expressed at different times and under different growth conditions and to describe a model of the highly-sensitive regulatory mechanisms which are necessary to control autolysin activity of D,L-endopeptidases.

This study is of importance, because of the fact that autolysins function outside the cell to support cell growth, what makes them an attractive target for vaccines and antibacterial agents. Both their enzymatic activities and their peptidoglycan-binding capacities can be targeted for antibacterial development. A better understanding of the structures and regulation of autolysins will help to establish the possibility of developing a successful antimicrobial agent based on these principles. The investigation of LysM motifs and how they contribute to peptidoglycan-binding could prove useful for the development of cell displays. Binding of LysM motifs to peptidoglycan is fast, robust and strong and therefore LysM motifs are an attractive tool to bind proteins, peptides and enzymes to the surface of Gram-positive bacteria or peptidoglycan coated particles. The binding characteristics of LysM motifs also create an opportunity to use non-GMO bacteria as vectors in vaccine development.

# 2 Chapter 2: Material and Methods

## 2.1 Genetic modifications

### 2.1.1 Sequences and sequencing

If not stated differently all sequences used in this work were taken from the database GenoList, which includes the sequenced *B. subtilis* genome (Barbe *et al.*, 2009) as well as 742 other genomes of different bacteria and archaea species [as at 02/10/13] (Lechat *et al.*, 2007). The database includes and develops the sequencing information, which was previously published in SubtiList (Moszer *et al.*, 2002). DNA was sequenced by Source BioScience.

### 2.1.2 DNA manipulations

A list containing all *B. subtilis* and *E. coli* strains used in this study is presented in Table 8.4 (appendix, page 241) and Table 8.5 (appendix, page 246). Standard molecular cloning methods were used (Sambrook et al., 1989). Plasmids (see Table 8.6, appendix, page 247) were cloned and propagated in *E. coli* strains TG-1 or TP611, which were made competent by the calcium chloride method (Sambrook et al., 1989). All transformations in *B. subtilis* were carried out following the protocol of Anagnostopoulus and Spizizen (1960). Construction details for each strain are specified in the following sections. Different methods (outlined below) were chosen to manipulate genes. If not stated differently, all DNA fragments were PCR amplified from chromosomal DNA of *B. subtilis 168*, which was isolated with the GenElute™ Bacterial Genomic DNA Kit (Sigma-Aldrich). Oligonucleotides were obtained from IDT Belgium and are listed in Table 8.7 (appendix, page 250). Transformants were selected on LB plates with respective antibiotics. Positive transformants which showed the expected antibiotic resistance were discriminated from spontaneous mutants through PCR screening and sequencing. Strains were stored long term in 25 % glycerol at -72°C. Screening PCRs for positive transformants were designed with appropriate primer pairs chosen to confirm the insertion of the linear DNA or the plasmid at the right location on the chromosome. Screening PCRs were undertaken with MyTaq™ Mix (Bioline). Linear DNA fragments used in molecular cloning were amplified with Phusion™ Hot Start II High-Fidelity DNA Polymerase (Fisher Scientific). PCR products were purified with illustra GFX PCR DNA and the Gel Band Purification Kit (GE Healthcare). Strand

overlap extension PCRs (Horton *et al.*, 1989) were undertaken with the Expand Long Template PCR System (Roche).

(i) **Double cross-over recombination with PCR amplified long flanking homology fragments (Wach, 1996)** was the method of choice for most gene disruptions and gene manipulations (e.g. the generation of a *lytE* null mutants and the expression of LytE by the *cwlO* promoter). In short, homologous regions upstream and downstream of the gene or region of interest were PCR amplified and joined to a middle fragment. The middle fragment was designed to replace the gene or region of interest and either encompassed an antibiotic resistance cassette (for simple gene disruption) or in addition to the antibiotic resistance cassette an open reading frame or a region of interest (for gene manipulations). Upstream and downstream fragments were amplified with specific primers that in addition to their respective annealing sequence possessed overhanging sequences at their 5'-end (of about 25 bp length). These overhanging sequences were designed to generate overlapping regions with the subsequent fragment, enabling the joining of two fragments by strand overlap extension PCR (Horton *et al.*, 1989). The final fragment was then transformed in *B. subtilis* by double cross-over recombination (Thomson, 1988).

(ii) **Markerless deletion of genes** (for Δ*cwlS,* Δ*lytF* and Δ*yqgT*) was achieved with the Xer recombination method of Bloor and Cranenburgh (2006). In a first step, an antibiotic resistance cassette was amplified that is flanked by *dif*-sites and regions homologous to the gene of interest. The antibiotic cassette was then integrated on the chromosome by double cross-over and replaced the gene or region of interest. Secondly the antibiotic cassette was excised with the help of the *dif*-sites. These specific sites are the recognition sequence for the native Xer site-specific recombinase RipX/CodV which resolves the two directly followed repeats of the *dif*-sites by excision of the intermediate region. In this work a chloramphenicol antibiotic cassette was amplified from the plasmid pDG1661 with primers, the 5′ end of which incorporated a region of homology flanking as well as the 28 bp long *B. subtilis dif*-site (*dif*-site sequence: 5′ - ACT TCC TAG AAT ATA TAT TAT GTA AAC T). The chloramphenicol antibiotic cassette was then joined, by strand overlap extension PCR (Horton *et al.*, 1989), to upstream and downstream fragments of the gene of interest. The final fragment was then transformed into *B. subtilis* by double cross-over recombination (Thomson, 1988) thereby replacing the respective gene. Positive transformants were selected on LB agar plates containing chloramphenicol. Strains were then

grown 24 hours in LB media without chloramphenicol, which allowed the excision of the chloramphenicol resistance cassette. Successful excision was confirmed by replicate platting (+/- chloramphenicol) and a screening PCR.

(iii) **Plasmid integration by single cross-over recombination (Dowds *et al.*, 1988)** was carried out to achieve the HA-epitope tagging of various proteins (CwlS-HA, LytE-HA, LytF-HA, YqgT-HA). The C-terminal regions of these genes were PCR amplified with primers that incorporated the HA-tag sequence (HA-tag sequence: 5′ - TAT CCG TAT GAT GTT CCG GAT TAT GCA TAA TAA) before the stop codon and that flanked the fragments with restrictions sites. The parental vector (containing an antibiotic resistance gene) and the PCR fragments were then digested with the compatible enzymes which were purchased from New England Biolabs and Thermo Scientific (FastDigest® restriction enzymes), following the instructions of the manufacturer. T4 DNA ligase (Roche) was used to ligate the fragments and plasmids at 16°C overnight. Plasmids were then transformed and propagated in competent *E. coli* (Sambrook *et al.*, 1989). Positive transformants were selected on ampicillin LB agar plates and the successful ligation of the plasmid was confirmed by screening PCR. The GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich) was used to isolate plasmids, which were then transformed into *B. subtilis* and integrated over single cross-over recombination (Thomson, 1988).

(iv) **Plasmid integration by double cross-over recombination (Dowds *et al.*, 1988)** was carried out to integrate a *lytE* open reading frame with permuted LysM motifs at the *cwlO* locus. One advantage of this method was that LysM motifs could be easily permutated. A parental plasmid was generated by cloning a fragment with homology upstream and downstream to the *cwlO* gene to both sites of an antibiotic cassette into a vector as described previously. The open reading frame of the *lytE* gene with permuted LysM motifs was then cloned in frame into this vector (between the *cwlO* upstream homology and the antibiotic cassette). To ensure a double cross-over recombination the plasmid was linearized by PCR amplification before transformation into *B. subtilis*.

### 2.1.3 Plasmid constructions

#### 2.1.3.1 Plasmid for HA-epitope tagging of LytE (with kanamycin resistance gene)

Plasmid **pLU2** enables a C-terminal translational fusion between the *lytE* open reading frame and a HA-tag. The HA peptide sequence is YPYDVPDYA (nucleotide sequence: 5´ - TAT CCG TAT GAT GTT CCG GAT TAT GCA TAA TAA) (AddGene). The plasmid was constructed by cloning a fragment, containing a 267 bp homologous region to the C-terminus of the *lytE* gene with the HA-tag sequence inserted before the stop codon, into the parental vector pDG780. The fragment was amplified with the following primer pair LytE_Tag_F_BamHI / LytE_Tag_R_HA_EcoRI and cloned into pDG780 by use of the *Bam*HI and *Eco*RI restriction sites.

#### 2.1.3.2 Plasmids for HA-epitope tagging of LytF, CwlS and LytF

The plasmid **pLU6** allows the generation of a C-terminal translational fusion between the LytF protein and a HA-tag. To construct plasmid pLU6, a fragment encompassing a 427 bp homologous region to the C-terminus of the *lytF* gene with the HA-tag sequence inserted before the stop codon, was cloned into the parental vector pDG780. The fragment was amplified with the primer pair LytF_Tag_F_BamHI / LytF_Tag_R_HA_EcoRI and cloned into pDG780 by use of the *Bam*HI and *Eco*RI restriction sites.

Plasmid **pLU7** allows the generation of a C-terminal translational fusion between the CwlS protein and a HA-tag. To construct plasmid pLU7, a fragment encompassing a 405 bp homologous region to the C-terminus of the *cwlS* gene with the HA-tag sequence inserted before the stop codon, was cloned into the parental vector pDG780. The fragment was amplified with the primer pair CwlS_Tag_F_BamHI / CwlS_Tag_R_HA_EcoRI and cloned into pDG780 by use of the *Bam*HI and *Eco*RI restriction sites.

Plasmid **pLU8** allows the generation of a C-terminal translational fusion between the YqgT protein and a HA-tag. To construct plasmid pLU8, a fragment encompassing a 402 bp homologous region to the C-terminus of the *yqgT* gene with the HA-tag sequence inserted before the stop codon, was cloned into the parental vector pDG780. The fragment was amplified with the primer pair YqgT_Tag_F_BamHI / YqgT_Tag_R_HA_EcoRI and cloned into pDG780 by use of the *Bam*HI and *Eco*RI restriction sites.

### 2.1.3.3 Parental vector for the integration of a variant *lytE* gene at the *cwlO* locus

A parental plasmid was required for convenient cloning of variant *lytE* open reading frames into a cassette that would allow integration of variant *lytE* genes at the *cwlO* locus. Furthermore, integration of the plasmid by double-cross-over recombination should result in the exchange of the *cwlO* open reading frame for the open reading frame of the variant *lytE*, which would then be expressed from the *cwlO* promoter.

To obtain such a parental plasmid, two PCR fragments were amplified with up- and downstream homology to the *cwlO* locus and cloned into a vector consecutively. First the upstream homology was amplified with the primer pair CwlO_UF_LysM and CwlO_UR_LysM which generated a 1071 bp long fragment encompassing the *cwlO* promoter, the *cwlO* leader sequence and part of the *lytE* signal sequence. The *lytE* signal sequence was altered to contain a silent *Bam*HI restriction site (i.e. not altering the amino acid sequence), which was achieved by adding this sequence to the 5´-end overhang of the reverse primer. The fragment and the parental vector pdG780 were then digested with *Not*I and *Bam*HI and the fragment was cloned into the vector to create the plasmid **pLU9**. Second a 953 bp long fragment encompassing the downstream homology to the *cwlO* locus was amplified with the primer pair CwlO_DF_LysM and CwlO_DR _LysM. The fragment was then digested with *Xho*I and *Kpn*I and cloned into the vector pLU9 to generate plasmid **pLU10.** The correct sequence of the upstream homology containing the *cwlO* promoter and *cwlO* leader sequence was verified by sequencing.

A schematic representation of the cloning process conducted to obtain the vectors pLU9 and pLU10 is shown in Figure 5.2 (a) (see page 117).

### 2.1.3.4 Plasmids for the generation of variant LytE with two LysM motifs

Nine plasmids were generated, harbouring a variant *lytE* open reading frame, encoding a variant LytE protein with two LysM motifs, in a cassette that would allow integration of these constructs at the *cwlO* locus. The LysM motifs of these variant *lytE* genes were altered in a way that they contained two instead of three LysM motifs and that a rearrangement of the two LysM motifs was possible. The plasmids were obtained by cloning dimer fragments (fragments of the *lytE* open reading frame with two LysM motifs) in a parental vector with up- and downstream homology to the *cwlO* locus. The construction of the nine dimer fragments is illustrated in Figure 5.1 (d) (see page 114). A schematic representation of the cloning process conducted to obtain the vectors is

shown in Figure 5.2 (c) (see page 117). The nomenclature used (e.g. linker, spacer) refers to certain domains of LytE which are defined in section 5.2 (page 112).

*LytE* dimer variant fragments were generated by amplifying the three LysM motifs of *lytE* each with two primer pairs each (creating six PCR fragments in total). One primer pair generated part of the *lytE* signal sequence, followed by the LysM motif, followed by the linker sequence (i.e. '*signal – LysM – linker*', 219 bp). The second primer pair generated the linker sequence followed by the LysM motif, followed by parts of the spacer sequence (i.e. '*linker – LysM – spacer*', 250 bp). For cloning purposes a silent *Bam*HI restriction site was introduced in the signal sequence and a silent *Spe*I restriction site was integrated in the linker sequence. The signal, linker and spacer sequences were generated through the 5´ overhang sequences of the primers.

Table 2.1 lists the respective primer pairs to amplify the six described fragments. All reactions were performed with WT168 chromosomal DNA as template.

**Table 2.1: List of primer pairs for the amplification of PCR fragments encoding one LysM motif (see Figure 5.1 d)**

| Forward primer | Reverse primer | Template | Resulting PCR fragment |
|---|---|---|---|
| L1_fwd_sig_BamHI | L1_rev_link | chrom. DNA of WT168 | *signal – L1 – linker* |
| L1_fwd_link | L1_rev_spac | chrom. DNA of WT168 | *linker – L1 – spacer* |
| L2_fwd_sig_BamHI | L2_rev_link | chrom. DNA of WT168 | *signal – L2 – linker* |
| L2_fwd_link | L2_rev_spac | chrom. DNA of WT168 | *linker – L2 – spacer* |
| L3_fwd_sig_BamHI | L3_rev_link | chrom. DNA of WT168 | *signal – L3 – linker* |
| L3_fwd_link | L3_rev_spac | chrom. DNA of WT168 | *linker – L3 – spacer* |

Any one of the three '*signal – LysM – linker*' fragments was then combined with any one of the three '*linker – LysM – spacer*' fragments by strand overlap extension PCR using the linker sequences of both fragments as an overlapping region. The strand overlap extension PCR was performed with the forward primer of the '*signal – LysM – linker*' fragment and with the reverse primer of the '*linker – LysM – spacer*' fragment. This amplification method generated nine dimer fragments of 421 bp length, with each fragment containing two LysM motifs in total (i.e. '*signal – LysM – linker – LysM – spacer*').

A second fragment of 431 bp length was then amplified with the primer pair Catalytic_LytE_Fwd_Spac / Catalytic_LytE_Rev_XmaI. This fragment encompassed parts of the spacer sequence (generated by the 5´ overhang of the forward primer), the *lytE* catalytic sequence and the *lytE* terminator sequence ('*spacer – catalytic*'). All nine

dimer fragments were then combined with this fragment by strand overlap extension PCR using the spacer sequence as overlapping region. The PCR was performed with the forward primer used to amplify the respective '*signal – LysM – linker*' fragment and the reverse primer Catalytic_LytE_Rev_XmaI for all reactions likewise. In this way nine fragments of 826 bp length were generated that contained the *lytE* signal sequence, two LysM motifs separated by a linker sequence (with a silent SpeI restriction site), followed by the spacer sequence, followed by the *lytE* catalytic domain and the *lytE* terminator ('*signal – LysM – linker – LysM – spacer – catalytic*'). The fragments were then digested with *Bam*HI and *Xma*I and cloned into the parental vector pLU10.

Table 2.2 lists which PCR fragments were cloned into the parental vector pLU10, to generate any of the following plasmids: pLU14 to pLU22.

**Table 2.2: List of PCR fragments that encode a variant LytE with two LysM motifs and the resulting plasmids after cloning of these fragments into the parental vector (see Figure 5.2 c)**

| PCR fragment | Parental plasmid | Resulting plasmid |
|---|---|---|
| *signal – L1 – linker – L3 – spacer – catalytic* | pLU10 | pLU14 (*lytE[L13cat]*) |
| *signal – L1 – linker – L2 – spacer – catalytic* | pLU10 | pLU15 (*lytE[L12cat]*) |
| *signal – L1 – linker – L1 – spacer – catalytic* | pLU10 | pLU16 (*lytE[L11cat]*) |
| *signal – L2 – linker – L1 – spacer – catalytic* | pLU10 | pLU17 (*lytE[L21cat]*) |
| *signal – L2 – linker – L2 – spacer – catalytic* | pLU10 | pLU18 (*lytE[L22cat]*) |
| *signal – L2 – linker – L3 – spacer – catalytic* | pLU10 | pLU19 (*lytE[L23cat]*) |
| *signal – L3 – linker – L1 – spacer – catalytic* | pLU10 | pLU20 (*lytE[L31cat]*) |
| *signal – L3 – linker – L2 – spacer – catalytic* | pLU10 | pLU21 (*lytE[L32cat]*) |
| *signal – L3 – linker – L3 – spacer – catalytic* | pLU10 | pLU22 (*lytE[L33cat]*) |

### 2.1.3.5 Plasmids for the generation of variant LytE with a single LysM motif

Three plasmids were generated, harbouring a variant *lytE* open reading frame, encoding a variant LytE protein with a single LysM motif, in a cassette that would allow the integration of these construct at the *cwlO* locus. These plasmids were obtained by cloning monomer fragments (fragments of the *lytE* open reading frame with only one LysM motif) in a parental vector with up- and downstream homology to the *cwlO* locus. The construction of the three monomer fragments is illustrated in Figure 5.1 (c) (see page 114). A schematic representation of the cloning process conducted to obtain these vectors is shown in Figure 5.2 (b) (see page 117). The nomenclature used (e.g. linker, spacer) refers to certain domains of LytE which are defined in section 5.2 (page

112). The three LysM motifs sequences were amplified with the following primer pairs L1_Fwd_Sig_BamHI / L1_Rev_Spac, L2_Fwd_Sig_BamHI / L2_Rev_Spac and L3_Fwd_Sig_BamHI / L3_Rev_Spac. The resulting fragments contained parts of the *lytE* signal sequence (as generated by the 5´ overhang of the forward primer), one LysM motif sequence and parts of the spacer sequence (generated by the 5´ overhang of the reverse primer) and can be described as 'linker – LysM – spacer'. The three fragments were then joined to another fragment encompassing the *lytE* catalytic sequence. This fragment of 431 bp length was amplified with the primer pair catalytic_*lytE*_fwd_spac / catalytic_*lytE*_rev_XmaI. The fragment encompasses part of the spacer sequence (generated by the 5´ overhang of the forward primer), the *lytE* catalytic sequence and the *lytE* terminator sequence (i.e. '*spacer – catalytic*'). The joining of the fragments was achieved by strand overlap extension PCR with the spacer sequence as overlapping region. The final fragments were then cloned into pLU10 after digestion of this vector and the fragments with *Xma*I and *Bam*HI restriction enzymes to generate the plasmids listed in Table 2.3.

**Table 2.3: List of PCR fragments that encode a variant LytE with one LysM motif and the resulting plasmids after cloning of these fragments into the parental vector (see Figure 5.2 b)**

| PCR fragment | Parental plasmid | Resulting plasmid |
|---|---|---|
| *signal – LysM 1 – spacer – catalytic* | pLU10 | pLU23 (*lytE[L1cat]*) |
| *signal – LysM 2 – spacer – catalytic* | pLU10 | pLU24 (*lytE[L2cat]*) |
| *signal – LysM 3 – spacer – catalytic* | pLU10 | pLU25 (*lytE[L3cat]*) |

### 2.1.3.6 Plasmids for the generation of variant LytE with three, four and five LysM motifs

Plasmids were generated, harbouring a variant *lytE* open reading frame, encoding a variant LytE protein with three, four or five LysM motifs, in a cassette that would allow the integration of these construct at the *cwlO* locus. These plasmids were obtained by cloning PCR fragments of single LysM motifs into a vector already holding a *lytE* open reading frame with two LysM motifs. The construction of these vectors is illustrated in Figure 5.2 (d) (see page 117). The nomenclature used (e.g. linker, spacer) refers to certain domains of LytE which are defined in section 5.2 (page 112). The three individual LysM motif sequences of *lytE* were amplified with the following primer pairs: L1_fwd_ link / L1_rev_ link, L2_fwd_link / L2_rev_link, L3_fwd_ link / L3_rev_link. The 5´ overhang sequences of these primers were designed to create linker sequences

(with a silent *Spe*I restriction site) up- and downstream of the LysM motif. The individual LysM motifs could then be cloned into the parental vectors after digestion of the vectors and the fragments with *Spe*I.

Since the same (*Spe*I) restriction site upstream and downstream of the fragment was used to clone the fragment into the vector, multiple integrations events of one fragment were possible. It was also possible for the fragment to integrate in the wrong direction. The number of integrated motifs was determined by a screening PCR with the primer pair LytE_DF / Spac_Check_Rev, followed by agarose gel electrophoresis. It was then possible to deduct the number of motifs from the size of the PCR product as shown in Figure 5.3 (see page 118). Plasmids were further sequenced to confirm the correct nucleotide sequence. The obtained plasmids are listed in Table 2.4.

**Table 2.4: List of PCR fragments that encode a single LysM motif and the resulting plasmids after cloning of these fragments into the parental vector with two LysM motifs (see Figure 5.2 d)**

| PCR fragment | Parental plasmid | Resulting plasmid |
|---|---|---|
| *linker – LysM 1 – linker* | pLU20 (*lytE[L31cat]*) | pLU26 (*lytE[L311cat]*) |
| *linker – LysM 1 – linker* | pLU19 (*lytE[L23cat]*) | pLU29 (*lytE[L213cat]*) |
| *linker – LysM 1 – linker* | pLU18 (*lytE[L22cat]*) | pLU31 (*lytE[L212cat]*) |
| *linker – LysM 2 – linker* | pLU18 (*lytE[L22cat]*) | pLU30 (*lytE[L222cat]*) |
| *linker – LysM3 – linker* | pLU20 (*lytE[L31cat]*) | pLU27 (*lytE[L331cat]*) |
| *linker – LysM3 – linker* | pLU19 (*lytE[L23cat]*) | pLU28 (*lytE[L233cat]*) |
| *linker – LysM3 – linker* | pLU20 (*lytE[L31cat]*) | pLU32 (*lytE[L3331cat]*) |
| *linker – LysM3 – linker* | pLU19 (*lytE[L23cat]*) | pLU34 (*lytE[L2333cat]*) |
| *linker – LysM3 – linker* | pLU20 (*lytE[L31cat]*) | pLU33 (*lytE[L33331cat]*) |
| *linker – LysM 1 – linker* | pLU14 (*lytE[L13cat]*) | pLU38 (*lytE[L113cat]*) |
| *linker – LysM 1 – linker* | pLU14 (*lytE[L13cat]*) | pLU39 (*lytE[L1113cat]*) |
| *linker – LysM 2 – linker* | pLU14 (*lytE[L13cat]*) | pLU40 (*lytE[L123cat]*$^{R}$) |
| *linker – LysM 1 – linker* | pLU16 (*lytE[L11cat]*) | pLU42 (*lytE[L11111cat]*) |
| *linker – LysM 2 – linker* | pLU16 (*lytE[L11cat]*) | pLU43 (*lytE[L1221cat]*) |
| *linker – LysM 2 – linker* | pLU16 (*lytE[L11cat]*) | pLU44 (*lytE[L12221cat]*) |
| *linker – LysM 3 – linker* | pLU14 (*lytE[L13cat]*) | pLU41 (*lytE[L1333cat]*) |

### 2.1.3.7 Plasmid for the generation of a variant LytE without LysM motif

One plasmid was generated, harbouring a variant *lytE* open reading frame, encoding a variant LytE protein without a LysM motif, in a cassette that would allow transforming this construct at the *cwlO* locus. The nomenclature used (e.g. linker, spacer) refers to certain domains of LytE which are defined in section 5.2 (page 112).

A fragment spanning the signal sequence, the spacer and the sequence of the catalytic domain was amplified with the primers LytE_Spacer_Fwd_Sig / Catalytic_LytE_Rev_XmaI (see Figure 5.1 b, page 114). The fragment was then cloned into pLU10 after digestion of this vector and of the fragment with *Xma*I and *Bam*HI restriction enzymes to generate the plasmid **pLU37.**

### 2.1.3.8 Plasmid for HA-epitope tagging of LytE (with spectinomycin resistance gene)

Plasmid **pLU46** enables a C-terminal translational fusion between the C-terminus of *lytE* and a HA-tag that can be selected for in *B. subtilis* by spectinomycin. The plasmid was constructed as a derivative from the parental vector pLU2, by exchanging the kanamycin resistance cassette of this vector against a spectinomycin cassette. To do so, a fragment of 1154 bp length containing the spectinomycin antibiotic resistance cassette was amplified from LU1 with the primer pair Spec_Fwd_EcoRI and Spec_Rev_SalI. The parental vector pLU2 and the PCR fragment were then digested with *Sal*I and *Eco*RI enzymes to release the kanamycin cassette and to allow the insertion of the spectinomycin resistance cassette into the digested vector.

### 2.1.3.9 Plasmids for the overexpression of variant LytE proteins in *E. coli*

To create plasmids that would allow the overexpression of variant LytE proteins in *E. coli,* the respective sequence of the proteins (excluding the start codon and the *lytE* signal sequence) were amplified as described in Table 2.5. PCR fragments were then cloned into the *Nde*I/*Xho*I sites of the vector pET21b+ thereby re-generating the start codon and generating a C-terminal fusion between the protein and a 6xHis-tag to obtain the plasmids listed in Table 2.6.

**Table 2.5: List of primer pairs for the amplification of PCR fragments encoding a variant LytE**

| Forward primer | Reverse primer | Template | Resulting PCR fragment |
|---|---|---|---|
| Exp_L1_Fwd | Exp_Cat_Rev | pLU40 | *L1 – linker – L2 – linker – L3 – spacer – catalytic – 6xHis* |
| Exp_L3_Fwd | Exp_Cat_Rev | pLU20 | *L1 – linker – L3 – spacer – catalytic – 6xHis* |
| Exp_L1_Fwd | Exp_Cat_Rev, | pLU14 | *L3 – linker – L1 – spacer – catalytic – 6xHis* |
| Exp_L1_Fwd | Exp_Cat_Rev | pLU23 | *L1 – spacer – catalytic – 6xHis* |
| Exp_L3_Fwd | Exp_Cat_Rev | pLU25 | *L3 – spacer – catalytic – 6xHis* |
| Exp_Spac_Fwd | Exp _Cat_ Rev | pLU37 | *spacer – catalytic – 6xHis* |
| Exp_XlyB_Fwd | Exp_Cat_Rev | pJL04 | $L1^{lytF}$ *– linker – L3 – spacer – catalytic – 6xHis* |
| Exp_LytF_Fwd | Exp_Cat_Rev | pJL02 | $L1^{xlyB}$ *– linker – L3 – spacer – catalytic – 6xHis* |

**Table 2.6: List of PCR fragments that encode a variant LytE and the resulting plasmids after cloning of these fragments into the parental vector**

| PCR fragment | Parental plasmid | Resulting plasmid |
|---|---|---|
| *L1 – linker – L2 – linker – L3 – spacer – catalytic – 6xHis* | pET21b+ | pLU47 *lytE[L123cat]*$^{R}$ |
| *L1 – linker – L3 – spacer – catalytic – 6xHis* | pET21b+ | pLU48 *lytE[L31cat]* |
| *L3 – linker – L1 – spacer – catalytic – 6xHis* | pET21b+ | pLU49 *lytE[L13cat]* |
| *L1 – spacer – catalytic – 6xHis* | pET21b+ | pLU50 *lytE[L1cat]* |
| *L3 – spacer – catalytic – 6xHis* | pET21b+ | pLU51 *lytE[L3cat]* |
| *spacer – catalytic – 6xHis* | pET21b+ | pLU52 *lytE[cat]* |
| $L1^{lytF}$ *– linker – L3 – spacer – catalytic – 6xHis* | pET21b+ | pLU53 *lytE[*$L1^{LytF}ca^{lytE}$*]* |
| $L1^{xlyB}$ *– linker – L3 – spacer – catalytic – 6xHis* | pET21b+ | pLU53 *lytE[*$L1^{XlyB}cat^{lytE}$*]* |

### 2.1.3.10 Plasmids for the generation of heterologous variant LytE with a single LysM motifs of a heterologous enzyme

Four plasmids were generated, harbouring the *lytE* open reading frame with a single heterologous LysM motif in a cassette that would allow transforming these constructs at the *cwlO* locus. These plasmids were obtained by cloning monomer fragments (fragments of the *lytE* open reading frame with a single heterologous LysM motif) in a parental vector with up- and downstream homology to the *cwlO* locus. The construction of the three monomer fragments is illustrated in Figure 5.5 (b) (see page 121). The nomenclature used (e.g. linker, spacer) refers to certain domains of LytE which are defined in section 5.2 (page 112).

Four LysM motifs were selected for this process. The first ($NH_2$-terminal) LysM motif of CwlS (D,L-endopeptidase of *B. subtilis 168*), the first LysM motif of LytF (D,L-endopeptidase of *B. subtilis 168*), the single LysM motif of XlyB (PBSX prophage protein of *B. subtilis 168*) and the first LysM motifs of AAA (amidase of *Staphylococcus aureus* strain Newman). The structure of the open reading frames of these genes is given in Figure 5.5 (a) (see page 121).

The monomer fragments were constructed in the same way as described in section 2.1.3.5, with the difference that four heterologous LysM motifs were amplified (as described in Table 2.7). After the joining of these PCR fragments to the *lytE* catalytic sequence by strand overlap PCR, the fragments were then cloned into pLU10 digested with *Xma*I and *Bam*HI restriction sites to obtain the plasmids listed in Table 2.8. These plasmids were generated by Jennifer Lennon under my supervision.

**Table 2.7: List of primer pairs for the amplification of PCR fragments encoding a heterologous variant LytE with a single heterologous LysM motif (see Figure 5.5 b)**

| Forward primer | Reverse primer | Template | Resulting PCR fragment |
|---|---|---|---|
| CwlS_LysM1_Fwd | CwlS_LysM1_Rev | chrom. DNA of WT168 | *signal* – $L1^{cwlS}$ – *spacer* |
| LytF_LysM1_Fwd | LytF_LysM1_Rev | chrom. DNA of WT168 | *signal* – $L1^{lytF}$ – *spacer* |
| XlyB_LysM1_Fwd | XlyB_LysM1_Rev | chrom. DNA of WT168 | *signal* – $L1^{xlyB}$ – *spacer* |
| Aaa_LysM1_Fwd | Aaa_LysM1_Rev | chrom. DNA of *S. aureus Newman* | *signal* – $L1^{aaa}$ – *spacer* |

**Table 2.8: List of PCR fragments that encode a heterologous variant LytE and the resulting plasmids after cloning of these fragments into the parental vector (see Figure 5.5 c)**

| PCR fragment | Parental plasmid | Resulting plasmid |
|---|---|---|
| *signal* – $L1^{cwlS}$ – *spacer – catalytic* | pLU10 | pJL1 (*lytE*[$L1^{cwlS}$*cat*$^{lytE}$]) |
| *signal* – $L1^{lytF}$ – *spacer – catalytic* | pLU10 | pJL2 (*lytE*[$L1^{lytF}$*cat*$^{lytE}$]) |
| *signal* – $L1^{xlyB}$ – *spacer – catalytic* | pLU10 | pJL4 (*lytE*[$L1^{xlyB}$*cat*$^{lytE}$]) |
| *signal* – $L1^{aaa}$ – *spacer – catalytic* | pLU10 | pJL5 (*lytE*[$L1^{aaa}$*cat*$^{lytE}$]) |

## 2.1.4 Strain constructions

### 2.1.4.1 *lytE* null mutant strains

A *lytE* null mutant strain (**LU1**) was constructed by replacing the *lytE* gene with a spectinomycin resistance cassette by double cross-over with a LFH fragment. The

spectinomycin resistance cassette was amplified from the plasmid pDG1727 with the primer pair Spec_Fwd / Spec_Rev. Fragments upstream and downstream of *lytE* were amplified with the primer pairs LytE_Up_Fwd / LytE_Up_Rev_Spec and LytE_Do_Fwd_Spec / LytE_Do_Rev. The fragments were then joined together by strand overlap extension PCR and transformed into WT168 to create strain LU1 by double cross-over recombination.

A second *lytE* null mutant strain (**LU331**) was generated by amplifying up- and downstream homology to the *lytE* locus with the primer pairs: LytE_Up_Fwd_2 / Signal_LytE_Rev_Spec and LytE_Do_Fwd_Spec / LytE_Do_Rev. Another fragment encompassing the spectinomycin resistance gene was amplified with the primer pair Spec_Fwd and Spec_Rev from the plasmid pDG1727. The three fragments were then joined together by strand overlap long flanking homology PCR and the PCR fragment of 2749 bp length was transformed into WT168 to generate the strain LU331. The construction of strain LU331 is shown schematically in Figure 5.6 (see page 124).

#### 2.1.4.2 Strains expressing a LytE protein from the *cwlO* promoter

The generation of the strains LU17, LU21 and LU25 is illustrated in Figure 3.1 (see page 65). A schematic structure of the *lytE* and *cwlO* locus of strain LU17, LU21, LU23 and LU25 is given in Figure 4.1 (see page 84).

Strain **LU17** has a transcriptional fusion between the *cwlO* promoter and the *lytE* gene. The *cwlO* gene in this strain is replaced by the *lytE* open reading frame at the *cwlO* locus. Furthermore, this strain does contain the *cwlO* leader region, which was shown to cause the *cwlO* transcript to be highly unstable (Noone *et al.*, 2014). To generate strain LU17, four PCR fragments were amplified individually and joined together and transformed into the strain EB35. Upstream and downstream fragments to generate homology to the *cwlO* locus were amplified with the primer pairs CwlO_UF_Sma / CwlO_UR and QE93_Fwd_Ery / QE_94_Rev. The *lytE* open reading frame was amplified with the primer pair LytE_DF and LytE_DR_Ery. An erythromycin resistance cassette was amplified from the plasmid pDG6464 with the primer pair Mls_Fwd / Mls_Rev. The individual fragments were joined together by strand overlap extension PCR (Horton *et al.*, 1989) and transformed into the strain EB35 to create strain LU17.

Strain LU17 carries two copies of the *lytE* gene (one at the *cwlO* locus and one at the *lytE* locus). The second copy (at the *lytE* locus) was deleted in strain **LU21** by transforming a PCR fragment amplified from the chromosomal DNA of LU1 with the primer pair LytE_Up_Fwd / LytE_Do_Rev into LU17 to generate the strain LU21.

To generate strain **LU23**, which is a *cwlOleader*-less version of LU17, a similar approach was chosen. All fragments were amplified as described previously, with the difference that the upstream fragment to the *cwlO* locus, was amplified from chromosomal DNA of the strain DN3000, which has a 141 bp markerless deletion of the leader region (Noone *et al.*, 2014). The fragments were then fused together and transformed into EB35 to generate strain LU23. The deletion of the endogenous *lytE* gene in LU23 was achieved as described earlier by transforming a PCR fragment amplified from the chromosomal DNA of LU1 with the primer pair LytE_Up_Fwd / LytE_Do_Rev into LU17 to generate the strain **LU25**.

#### 2.1.4.3 Strains expressing a HA-tagged LytE protein

Strains **LU45, LU47, LU61** and **LU63** are derivatives of strains WT168, LS103, LU21 and LU25 respectively, in which the *lytE* gene is fused with a C-terminal HA-tag. These strains were constructed by transforming plasmid pLU2 (see section 2.1.3.1) into WT168, LS103, LU21 and LU25. The successful integration of the plasmid into a *cwlO* mutant strain (LS103) served as proof for the functionality of the HA-tag, given that a failure would lead to a combined *cwlO* / *lytE* mutant, which is synthetically lethal (Bisicchia *et al.*, 2007).

#### 2.1.4.4 Markerless null mutant strains in *cwlS*, *lytF* and *yqgT*

Markerless null-mutants in *cwlS*, *lytF* and *yqgT* (**LU72, LU74, LU76**) were generated by the Xer recombination method (Bloor and Cranenburgh, 2006), which allows the clean excision of a region of interest on the basis of *dif*-sites. First upstream and downstream regions of the *cwlS*, *lytF* and *yqgT* gene were amplified with the following primer sets: CwlS_Up_Fwd / CwlS_Up_Rev, CwlS_Do_Fwd / CwlS_Do_Rev; LytF_Up_Fwd / LytF_Up_Rev, LytF_Do_Fwd / LytF_Do_Rev; YqgT_Up_Fwd / Yqg_Up_Rev, YqgT_Do_Fwd / YqgT_Do_Rev. Second the primer pair Cat_Diff_Fwd / Cat_Diff_Rev (which contain 5´ tails to generate the *dif*-sites) was used to amplify the *cat* gene for chloramphenicol resistance from the plasmid pDG1661. Upstream and downstream fragments were then fused with the chloramphenicol antibiotic cassette and transformed into WT168. Double cross-over of the fragment into the bacterial chromosome caused the exchange of the gene of interest against the chloramphenicol cassette. To gain a markerless deletion, positive transformants were grown for 24 hours in LB media without a selective antibiotic, which induced the excision of the chloramphenicol cassette based on the *dif*-sites. Screening for chloramphenicol sensitive colonies yielded in strains LU72 (Δ*cwlS*), LU74 (Δ*lytF*), LU76 (Δ*yqgT*).

### 2.1.4.5 Double null mutant strains between *cwlS*, *lytF*, *yqgT* and *lytE, cwlO*

The single mutants in the genes *cwlS* (LU72), *lytF* (LU74) and *yqgT* (LU76) were combined with disruption of the *lytE* or the *cwlO* gene to create the following double mutants **LU90** (Δ*cwlS* Δ*cwlO*::*kan*$^r$), **LU92** (Δ*cwlS* Δ*lytE*::*kan*$^r$), **LU96** (Δ*yqgT* Δ*cwlO*::*kan*$^r$), **LU98** (Δ*yqgT* Δ*lytE*::*kan*$^r$), **LU100** (Δ*lytF* Δ*cwlO*::*kan*$^r$) and **LU102** (Δ*lytF* Δ*lytE*::*kan*$^r$). A kanamycin resistance cassette (flanked by upstream and downstream homologues regions to the *cwlO* and *lytE* gene) was transformed into the strains LU72, LU74 and LU76 to disrupt either the *cwlO* or the *lytE* open reading frame. The Δ*cwlO*::*kan*$^r$ cassette was amplified from the chromosomal DNA of strain EB35 with the primer pair CwlO_UF / QE94_Rev. The Δ*lytE*::*kan*$^r$ cassette was amplified from the chromosomal DNA of strain DN3061 with the primer pair LytE_Up_Fwd / LytE_Do_Rev. The PCR fragments were then transformed by double cross-over into the respective parental strains (LU72, LU74, LU76) to create the double null mutant strains.

### 2.1.4.6 Strains expressing a CwlS, LytF or YqgT protein from the *cwlO* promoter

A schematic structure of the *lytE* and *cwlO* loci of the strains LU106, LU111, LU112, LU123, LU125, LU116 and LU134 is presented in Figure 4.1 (see page 84).

The open reading frames of the autolysins *cwlS*, *lytF* and *yqgT* were cloned into the *cwlO* expression cassette to create a transcriptional fusion between the *cwlO* promoter (containing the *cwlO* leader region) with the respective gene. The procedure followed the approach chosen to create strain LU17 (see section 2.1.4.2). In short, these genetic modifications were achieved by transforming a PCR fragment, made up of four individual fragments (joined together by strand overlap extension PCR) into a parental strain over double-cross-over, thereby replacing the *cwlO* gene with the open reading frame of either *cwlS*, *lytF* or *yqgT*.

The four individual fragments consisted of up- and downstream fragments (encompassing the homology to the *cwlO* locus), an antibiotic resistance cassette (erythromycin) and the gene of interest (*cwlS*, *lytF* or *yqgT*). Downstream fragments and the erythromycin resistance cassette were amplified as described previously using the primer pair QE93_Fwd_Ery / QE94_Rev to amplify the downstream fragment from WT168 chromosomal DNA and the primer pair Mls_Fwd / Mls_Rev to amplify the erythromycin resistance cassette from the plasmid pDG6464. To generate strain **LU106** the upstream fragment was amplified with the primers CwlO_UF and CwlO_UR_CwlS and a middle fragment (containing the *cwlS* open

reading frame) was amplified with the following primers CwlS_Fwd / CwlS_Rev_Ery. The fragments were then joined with the erythromycin resistance cassette fragment and the downstream fragment to generate a PCR fragment which was transformed into LU72 to create strain LU106. To generate strain **LU111** the upstream fragment was amplified with the primers CwlO_UF and CwlO_UR_LytF and a middle fragment (containing the *lytF* open reading frame) was amplified with the following primers LytF_Fwd / LytF_Rev_Ery. The fragments were then joined with the erythromycin resistance cassette fragment and the downstream fragment to generate a PCR fragment which was transformed into LU74 to generate strain LU111. To generate strain LU112 the upstream fragment was amplified with the primers CwlO_UF and CwlO_UR_YqgT and a middle fragment (containing the *yqgT* open reading frame) was amplified with the following primers YqgT_Fwd / YqgT_Rev_Ery. The fragments were joined with the erythromycin resistance cassette fragment and the downstream fragment to generate a PCR fragment which was transformed into LU76 to generate strain **LU112**.

An attempt was made to create *cwlOleader* less versions of the strain LU106, LU111 and LU112 following the approach chosen for the strain LU23 (see section 2.1.4.2), in which the upstream fragment was amplified from the chromosomal DNA of DN3000. The *cwlO* leaderless fragments were transformed into LU72 to create the strain **LU123** and into LU74 to create the strain **LU125**. It was not possible to obtain a leaderless version of the strain LU112 by transforming the specific DNA fragment into strain LU76.

Strains **LU116** and **LU134** were obtained by transforming a PCR fragment, containing a spectinomycin resistance cassette flanked by homologues region to the *lytE* locus, into the parental strains LU111 and LU125. The fragment was amplified with the primer pair LytE_Up_Fwd / LytE_Do_Rev from the chromosomal DNA of LU1.

### 2.1.4.7 Strains expressing a HA-tagged LytE, CwlS, LytF or YqgT protein

To achieve a C-terminal fusion between either the *lytE, cwlS, lytF* or the *yqgT* gene and a HA-tag, strain **LU143** was acquired by transforming the plasmid pLU2 (see section 2.1.3.1) into strain LU17. Strains **LU145, LU159** and **LU161** were obtained by transforming the plasmid pLU7 (see section 2.1.3.2) into WT168, LU106 and LU123. Strains **LU147, LU151, LU153, LU155** and **LU157** were generated by transforming the plasmid pLU6 (see section 2.1.3.2) into WT168, LU111 and LU116, LU125 and LU134. The strains **LU149** and **LU163** were generated by transforming the plasmid pLU8 (see section 2.1.3.2) into WT168 and into LU112.

### 2.1.4.8 Strains containing a variant LytE expressed from the *cwlO* promoter

The plasmids pLU16, pLU15, pLU14, pLU17, pLU18, pLU19, pLU20, pLU21, pLU22, pLU23, pLU24, pLU25, pLU26, pLU32, pLU27, pLU33, pLU28, pLU34, pLU30, pLU31, pLU29, pLU37, pLU38, pLU40, pLU41, pLU42, pLU43 and pLU44 (see sections 2.1.3.4 to 2.1.3.7) were transformed by double cross-over (with homology up- and downstream of the *cwlO* locus) into WT168 to generate the strains: **LU239, LU207, LU212, LU216, LU220, LU223, LU227, LU232, LU235, LU261, LU265, LU269, LU272, LU274, LU276, LU278, LU280, LU282, LU286, LU288, LU312, LU314, LU315, LU318, LU322, LU324, LU326 and LU328.**

The integration of the variant *lytE* open reading frame on the chromosome of WT168 at the *cwlO* locus is shown schematically in Figure 5.4 (a) (see page 119).

The plasmids pLU16, pLU15, pLU14, pLU17, pLU18, pLU19, pLU20, pLU21, pLU22, pLU23, pLU24, pLU25, pLU26, pLU32, pLU27, pLU33, pLU28, pLU34, pLU30, pLU31, pLU29, pLU38, pLU40, pLU41, pLU42, pLU43 and pLU44 (see sections 2.1.3.4 to 2.1.3.7) were transformed by double cross-over (with homology up- and downstream of the *cwlO* locus) into the *lytE* null mutant LS089 to generate the strains: **LU205, LU209, LU213, LU217, LU222, LU226, LU229, LU233, LU238, LU262, LU266, LU270, LU273, LU275, LU277, LU279, LU281, LU283, LU287, LU289, LU313, LU316, LU320, LU323, LU325, LU327 and LU329.**

The integration of the variant *lytE* open reading frame on the chromosome of LS089 at the *cwlO* locus is shown schematically in Figure 5.4 (b) (see page 119).

### 2.1.4.9 Strains containing a variant LytE expressed from the *lytE* promoter (with intact *cwlO* gene)

The following strains express a variant LytE protein with permuted LysM motifs from the *lytE* promoter. These strains were constructed by transforming a PCR fragment encoding the variant LytE, by double cross-over into a *lytE* null mutant strain (LU331), thereby reconstituting the *lytE* gene. The final PCR fragment was generated by joining four individual fragments by strand overlap extension PCR. The upstream and the downstream fragments flanking the *lytE* locus were amplified with the primer pairs LytE_Up_Fwd_2 / Signal_LytE_Rev and LytE_Do_Fwd_Ery / LytE_Do_Rev from chromosomal DNA of WT168. The downstream fragment was joined to an erythromycin resistance cassette which was amplified from pDG646 with the primers Mls_Fwd / Mls_Rev. The upstream fragment was joined to fragments with the altered sequence of the *lytE* gene, which were amplified with the primers Signal_LytE_Fwd / Catalytic_LytE_Rev_Ery from the plasmids pLU40, pLU23, pLU24, pLU25, pLU20,

pLU27, pLU37, pLU16, pLU17, pLU14 and pLU29 (see section 2.1.3.4 to 2.1.3.7). After joining of the upstream fragment (upstream homology to the *lytE* locus plus altered *lytE* gene sequence) to the downstream fragment (erythromycin resistance cassette plus downstream homology to the *lytE* locus), the fragments were then transformed into the strain LU331 to create the strains **LU335, LU337, LU339, LU341, LU372, LU374, LU398, LU399, LU401, LU403 and LU405.**

The integration of the PCR fragment on the chromosome of strain LS331 over double cross-over is shown schematically in Figure 5.6 (see page 124).

#### 2.1.4.10 Strains containing a variant LytE expressed from the *lytE* promoter (with disrupted *cwlO* gene)

To disrupt the *cwlO* gene in a series of strains, a fragment encompassing the *cwlO* locus was amplified with the primer pair CwlO_UF / QE94_Rev from chromosomal DNA of the strain EB35, in which the *cwlO* gene was disrupted by a kanamycin resistance gene. The PCR fragment was then transformed by double cross-over into the strains LU335, LU337, LU339, LU341, LU372, LU374, LU399, LU401, LU403 and LU405 to obtain the strains **LU343, LU350, LU352, LU354, LU383, LU385, LU407, LU409, LU411 and LU413.**

The disruption of the *cwlO* gene over double cross-over is shown schematically in Figure 5.6 (see page 124).

#### 2.1.4.11 Strains containing a HA-tagged variant LytE

To enable a C-terminal fusion of the variant LytE proteins with a HA-tag, a plasmid pLU46 (see section 2.1.3.8) was designed. This plasmid holds homology to the *lytE* catalytic domain, followed by the HA-tag and integrates by single cross-over at the *lytE* locus.

The plasmid pLU46 was transformed into the strains LU205, LU213, LU217, LU229, LU262, LU266, LU320, LU343, LU385, LU407, LU411, LU313, LU350, LU270, LU352 and LU413 to generate the strains **LU414, LU415, LU416, LU417, LU418, LU419, LU421, LU422, LU424, LU425, LU427, LU432, LU435, LU440, LU442** and **LU443**.

#### 2.1.4.12 Strains containing a heterologous variant LytE expressed from the *cwlO* promoter

The plasmids pJL1, pJL2, pJL4, pJL5 (see section 2.1.3.10) were transformed by double cross-over (with homology up- and downstream of the *cwlO* locus) into WT168 to generate the strains: **JL11, JL13, JL15 and JL17.**

The plasmid pJL4 was transformed by double cross-over (with homology up- and downstream of the *cwlO* locus) into the *lytE* null mutant LS089 to generate the strain **JL19.** The integration of the fragment over double cross-over is shown schematically in Figure 5.5 (d) (see page 121).

These strains were generated by Jennifer Lennon under my supervision.

#### 2.1.4.13 Strains containing a HA-tagged heterologous variant LytE expressed from the *cwlO* promoter

To enable a C-terminal fusion of the heterologous variant LytE proteins with a HA-tag, plasmid pLU46 (see section 2.1.3.8) was transformed by single cross-over into the strains JL13, JL15 and JL19 to generate the strains **JL21, JL22 and JL24** respectively. These strains were generated by Jennifer Lennon under my supervision.

## 2.2 Experimental procedures

### 2.2.1 Media and growth conditions

*E. coli* and *B. subtilis* were routinely grown in Luria-Bertani (LB) media (Sambrook *et al.*, 1989). Experiments with *B. subtilis* were performed in LPDM (low phosphate-defined medium, 0.75 MM $K_3(PO)_4$) and HPDM (high phosphate-defined medium, 4.5 MM $K_3(PO)_4$) (Müller *et al.*, 1997). Growth studies with varying amounts of salt were conducted in salt-free buffered LB (Boylan *et al.*, 1993), to which NaCl was added up to the desired concentration. $OD_{600}$ measurements to monitor growth were taken with a UV mini-1240 UV-visible spectrophotometer (Shimadzu Instruments). Liquid cultures were grown in an orbital shaker at 37°C and 220 rpm (New Brunswick, Edison NJ). Antibiotics were added as required at the following concentrations: 100 mg / ml ampicillin, 10 mg / ml chloramphenicol, 2 µg or 150 mg / ml erythromycin, 10 mg / ml kanamycin, 125 µg / ml rifampicin and 100 mg / ml spectinomycin.

### 2.2.2 Growth assays

Growth studies were conducted with the BioTek Synergy™2 system (BioTek Instruments). A volume of 100 µl media in a CELLSTAR® 96 well plate (Greiner Bio-One) was inoculated from an overnight culture to an $OD_{600}$ of 0.03 and incubated at 37°C with slow shaking. The $OD_{600}$ was measured every 10 minutes over a time period of 8 hours. Three technical repeats were performed for each growth curve.

### 2.2.3 RNA extraction

RNA extraction was performed as described in Salzberg *et al.* (2013). In short, samples were harvested at the designated $OD_{600}$ and centrifuged for 3 minutes at 4 000 x g and 4°C. Cell pellets were immediately frozen in a dry-ice ethanol bath and could be stored at -70°C or processed on the same day. The bacterial pellet was resuspended in 200 µl distilled water and transferred to a tube containing 0.5 g glass beads and 1 ml TRI reagent (RNA Isolation Reagent, Sigma-Aldrich). The cells were broken in a bead-beater (FastPrep®-24 Instrument, MP Biomedicals) with 3 pulses of 40 seconds and a speed of 6 m/s and cooled on ice before addition of 200 µl chloroform. Phases were separated through centrifugation for 3 minutes at 14 500 x g and the top phase was removed to a new tube and the same volume of 200 proof ethanol was added. The mixture was then bound to a column of the GenElute Mammalian Total RNA Miniprep Kit (Sigma) and processed following the manufacturer's instructions, with the exception that DNase was added on the column (10U RQ1 DNase, Promega) and incubated for 15 minutes after the first wash.

### 2.2.4 Northern analysis

RNA samples for northern blot analyses were prepared as described in section 2.2.3. For mRNA half-life experiments cultures were grown to the required $OD_{600}$, a sample removed prior to rifampicin addition and further samples harvested after rifampicin addition at the required time intervals (final rifampicin concentration: 125 µg/ml). Northern analysis was performed as described previously by Noone *et al.* (2000). In short, RNA was separated on an agarose gel (1.2 % [wt / vol]) and transferred to a positively charged nylon membrane (Pall Gelman, Ann Arbor, Mich) through capillary blotting. The membrane was then hybridized with DIG labelled RNA probes. Northern blot probes were synthesized with the following primer pairs: *lytE* probe F / *lytE* probe T7 R (*lytE* probe), yycE probe fwd / yycE probe rev T7 (*cwlO* probe). The T7 reverse primer carries the T7 RNA polymerase sequence which allows synthesis of dig-UTP labelled RNA probes by the T7 RNA polymerase. The probes were DIG labelled with the DIG Northern Starter Kit (Roche). Hybridized probe was detected with CDP-Star chemiluminescencent substrate (Roche).

### 2.2.5 Quantitative Reverse Transcription PCR (qRT-PCR)

RNA samples for qRT-PCR were prepared as described in section 2.2.3. One microgram total RNA was transcribed to cDNA with the Transcriptor High Fidelity cDNA synthesis kit (Roche). The cDNA was synthesized following the instructions of the

manufacturer with random hexamer primers. qRT-PCR was carried out in 96 well plates (LightCycler 480 multiwell 96, Roche) with the LightCycler 480 SYBR Green I Master Kit (Roche) on a LightCycler 480 instrument (Roche). The qRT-PCR programme was followed by a melt curve, which assured the specificity of each amplification. The absence of DNA was confirmed by analysing the initial RNA samples in one qRT-PCR reaction. All reactions were conducted in duplicate. Cp´s (crossing points) were identified with the Second Derivative Maximum Method of the LightCycler 480 software (version 1.5.0). A qRT-PCR reaction with 16 S primers (16S Bsu F / 16S Bsu R) served as normalization. All qRT-PCR primers were designed with the aid of the PrimerExpress 3.0 software (Applied Biosystems). The relative expression ratios were calculated with the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). The following primer were used to amplify the respective transcripts: *16s* (16S_Bsu_F / 16S_Bsu_R), *cwlO* (YvcE_5´_qPCR_orf*2 / YvcE_3´_qPCR_orf*2), *lytE* (*lytE*_qPCR_F2 / *lytE*_qPCR_R2), *lytF* (*lytF*_qPCR_fwd / *lytF*_qPCR_rev), *mreBH* (*mreBH*_qPCR_F / *mreBH*_qPCR_R), *sigI* (SigI_Chip_fwd / SigI_Chip_rev), *sigX* (SigX_qPCR_F / SigX_qPCR_R) and *yocH* (yocH_northern_F / yocH_northern_R).

### 2.2.6 Western analysis

Western blot analysis was carried out as previously described (Howell *et al.*, 2006). Protein samples for western blot analysis were obtained by growing *B. subtilis* strains to specified stages of the growth cycle. Samples were harvested by centrifugation at 4000 x g to separate the cells from the medium. Cells were then lysed in a lysis solution (10mM Tris pH = 8, 0.5mM EDTA, 100 µg/ml lysozyme, 10 µg/ml Dnase, 2 % [v/v] protease inhibitor cocktail set III, Calbiochemie). The proteins were resuspended in King-Laemmli sample buffer (King and Laemmli, 1971). Protein concentrations were determined by bicinchoninic acid assay (BCA assay) (Calbiochem). To isolate secreted proteins in the media, culture supernatant was mixed with ice-cold TCA (trichloroacetic acid, 10 % final concentration [vol / vol]). Proteins were then harvested by repeated centrifugation. The resulting pellet was washed in ice-cold acetone and resuspended in King-Laemmli sample buffer. All samples were diluted to the desired protein concentration in King-Laemmli sample buffer with 0.025 % (final concentration) ß-mercaptoethanol. Samples were thoroughly vortexed and boiled for 10 minutes before loading 5 µg to 10 µg protein on a 12 % SDS-PAGE Tris-Glycine polyacrylamide gel. Proteins isolated from the cell supernatant were loaded as OD equivalents (units of optical density at $OD_{600}$ multiplied by volume). After separating the proteins by polyacrylamide gel electrophoresis the proteins were transferred to a PVDF membrane (Roche) by electro blotting. To decrease the background the membrane

was blocked with skimmed milk powder (10 % [w/v]) in TBS buffer (0.05 M Tris, 0.15 M NaCl, pH = 7.5). The primary antibody was applied in a dilution of 1:5000 for 2 hours, before the membrane was washed in TBS buffer with 0.1 % TWEEN 20 (Sigma-Aldrich) to remove unspecific antibody binding. The secondary antibody (dilution 1:40.000) was incubated for 1 hour. Specific protein-antibody interactions were detected with the WesternBright™ Sirius Western analysis detection kit (Advansta). The primary CwlO antibody was affinity purified from polyclonal antiserum produced in New Zealand White rabbits according to the standard procedure used by Dundee Laboratories (Dundee, Scotland) (Noone *et al.*, 2014). The HA primary antibody (raised in rabbit) was purchased from Sigma-Aldrich. Horse-radish peroxidases secondary anti-mouse and anti-rabbit antibodies were purchased from Sigma-Aldrich.

### 2.2.7 Microscopy

Cells were heat-fixed on glass slides and examined with an Olympus BX61 microscope. Antifade solution containing 4´,6´diamidino-2-phenylidole (DAPI, SlowFade Gold antifade, Invitrogen) was added to visualize the nucleoids. The CellSens 1.9 Digital Imaging Software® was used to capture and process pictures.

### 2.2.8 Immunofluorescence microscopy

Immunofluorescence microscopy was performed essentially as described in Noone *et al.* (2012). In brief, strains were grown in LB or salt free buffered LB to an $OD_{600}$ of 0.2. Cells were harvested and washed three times in PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 1.8 mM $KH_2PO_4$). Cells were then fixed in 2.6 % paraformaldehyde first for 20 minutes at room temperature, followed by 30 minutes on ice and washed three more times in PBS buffer. Cells were then resuspended to an $OD_{600}$ of 0.4 and an aliquots of the cell suspension were applied on a multiwell microscopic slide pre-treated with poly-L-lysine as described in Szurmant, Fukushima *et al.* (2007). Cells were given 30 minutes at room temperature to attach to the slide, before excess suspension was removed and each well was washed three times with GTE buffer (50 mM, 20 mM Tris-HCl pH = 7.5, 10 mM EDTA). Each well was then incubated with lysozyme (2 mg / ml in GTE buffer) for 2 minutes. Wells were then washes 10 times with PBS buffer and slides immersed in ice-cold methanol for 5 minutes and allowed to dry completely. Wells were rehydrated for 5 minutes with PBS buffer and blocked for 20 minutes with 2 % BSA solved in PBS buffer. Wells were individually incubated with the primary antibody (rabbit anti-HA 1:500, rabbit serum anti-CwlO 1:1000) in PBS-BSA for 1 hour at room temperature. Each well was then washed 10 times with PBS-BSA before incubation with the secondary antibody (an

anti-rabbit IgG conjugated to fluorescein purchased from Sigma) at a 1:200 dilution in PBS-BSA in the dark. Afterwards each well was washed again 10 times with PBS buffer before antifade solution containing 4´,6´diamidino-2-phenylidole (DAPI) (SlowFade Gold antifade, Invitrogen) was applied. Samples were then observed with an Olympus BX61 microscope.

### 2.2.9 Suppressor mutation test

The presence of extragenic mutations was tested by examining transformation frequencies. Chromosomal DNA of the strain, to be tested for extragenic mutations (donor strain), was transformed in increasing concentrations (from 0.05 µg to 1.5 µg) in a recipient strain. Successful transformants were selected for the antibiotic present in the donor strain. If the genetic modification of the donor strain requires extragenic mutations for stability, then transformation frequencies will be reduced by orders of magnitude compared to a control strain.

### 2.2.10 Antibiotic assay with Etest® stripes

Etest® stripes were used to compare antibiotic sensitivity of bacterial strains to vancomycin, ampicillin and fosfomycin. Etest® stripes were purchased from bioMérieux at the following concentrations: Vancomycin (0.016-256 µg/ml), Ampicillin (0.016-256 µg/ml), Fosfomycin (0.064-1024 µg/ml). Cells were grown in LB media to an $OD_{600}$ of 0.2. Thirty microliter of cell culture were then added to 3 ml of 0.7 % (w/v) agar which was applied on top of 25 ml 1 % (w/v) agar plates pre-warmed to 45°C. Etest® strips were applied to the inoculated agar plate following the instructions of the manufacturer. The plates were incubated 16 hours at 37°C. The minimum inhibitory concentration (MIC) was determined where the ellipse intersects the reading scale.

### 2.2.11 Autolysis assay

Autolysis rates were measured as previously described in Bisicchia *et al.* (2010). In short *B. subtilis* strains were grown in HPDM medium in shake flasks at 37°C until exponential phase ($OD_{600}$ = 0.5). Cells were then harvested and washed once in 10 mM Tris-HCl (pH 8.0) and resuspended in 10 mM Tris-HCl (pH 8.0). The $OD_{600}$ was adjusted to $OD_{600}$ = 0.5 and cell lysis was monitored by measuring the decrease in $OD_{600}$ every 10 minutes, using a plate reader with slow orbital shaking at 37°C (BioTek Synergy™2 system). Three technical replicates were conducted for each strain.

### 2.2.12 Protein overexpression and purification

The variant LytE proteins were overexpressed in *E. coli* strain BL21(DE3) which was transformed with the respective expression plasmid (pLU47 to pLU53). Protein expression and purification was performed as described by Howell *et al.* (2006). In short BL21(DE3), harbouring the respective plasmid, was grown to an $OD_{600}$ of 0.6, at which point protein expression was induced with 1 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside). After 3 hours incubation at 28°C cells were harvested, incubated for 1 hour with lysozyme (1 mg/ml) and sonicated. The lysate was mixed with His-beads to capture the Hisx6 tagged protein and the beads washed repeatedly with wash buffer (20 mM Imidazole, 300 mM NaCl, 50 mM $Na_2HPO_4$, 50 mM $NaH_2PO_4$) (four times with 0.1 % [v/v] TWEEN20 and four times without) to remove unspecific binding. After elution with 250 mM imidazole, dialysis was used to transfer the protein to a storage buffer containing 50 % glycerol (20 mM Tris pH 8, 300 mM NaCl, 50 % glycerol).

### 2.2.13 Coomassie Blue staining

Coomassie Blue staining was used to visualize proteins. Proteins were first separated on a 12 % SDS-PAGE Tris-Glycine polyacrylamide gel by electrophoresis. The gel was then stained with Coomassie Brilliant Blue R-250 for 20 minutes at room temperature with shaking (0.1 % Coomassie Brilliant Blue R-250, 40 % Methanol, 10 % Glacial acetic acid). Destaining was achieved by incubating the gel at room temperature in destaining solution (Methanol 10 %, Glacial acetic acid, 10 % [vol/vol]) which was replenished four times during the first 2 hours and then incubated for a further 16 hours.

### 2.2.14 Zymography

Zymography was performed as described by Leclerc and Asselin (1989). Proteins were separated on a 12 % SDS PAGE gel, containing 30 µl of previously purified cell walls of *B. subtilis 168*. Renaturation was conducted overnight at 37°C with shaking in renaturation buffer (containing 25 MM Tris-HCl [pH 7.5], 0.1 % [vol/vol] Triton X-100, 10 mM MgCl2). The gel was stained for 3 hours in 0.1 % Methylene Blue, 0.01 % KOH at 37°C with shaking. Destaining was performed at room temperature in distilled water until clearing zones became visible.

### 2.2.15 D,L-endopeptidase enzymatic activity assay

Cell wall lytic activity of purified autolytic enzymes was measured as the decrease in $OD_{600}$ for purified cell walls or live cells to which purified proteins were added. Cell wall of *B. subtilis 168* was prepared as described in Salzberg and Helmann (2007) (the number of washes was increased to 15). Cell wall lytic activity was measured by preparing a suspension of purified cell walls in HPDM (to a final absorbance of $OD_{600}$ = 0.5) to which 60, 85 or 100 pmole of the purified enzyme were added per 100 µl reaction volume. Alternatively WT168 cells were grown in HPDM to an $OD_{600}$ of 0.5 and harvested and washed three times in ice-cold 10mM Tris-HCl (pH 8.0) before they were resuspended to an $OD_{600}$ of 0.5 in ice-cold 10mM Tris-HCl (pH 8.0). To each 100 µl of this suspension 60, 85 or 100 pmole of purified enzyme were added. The reaction mixture was then incubated at 37°C in a 96 well plate and the decrease in $OD_{600}$ was monitored every 2 minutes over a period of 3 hours, using a plate reader with slow orbital shaking at 37°C (BioTek Synergy™2 system). Three technical replicates were conducted for each strain.

### 2.2.16 Insoluble peptidoglycan pull down assay

To determine the binding capacity of variant LytE proteins to peptidoglycan an insoluble peptidoglycan pull down assay was conducted. The assay was performed essentially as described in Wong *et al.* (2014). Forty-four pmole of variant LytE protein were mixed with purified cell walls of *B. subtilis* (to a final absorbance of $OD_{600}$ = 10) in the assay buffer (50 mM Tris/HCl, pH 8, 0.05 M NaCl, 5 mM ß-mercaptoethanol, 1 % TWEEN20) in a 25 µl reaction volume. The mixture was incubated for 10 minutes at 10° C and centrifuged for 10 minutes at 10500 g. The supernatant was collected and the pellet resuspended in 15 µl assay buffer. The supernatant and the resuspended pellet as well as a control sample, containing 44 pmole of the protein and no cell wall to ensure that proteins do not precipitate on their own, were analysed on a 12.5 % SDS / PAGE gel. The proteins were visualized by Coomassie Blue staining and the image was quantified with the ImageJ software (Schneider *et al.*, 2012).To test the influence of salt on the binding capacity of variant LytE proteins, varying amounts of NaCl were added to the assay buffer (final concentrations: 0.05 M, 0.1 M, 0.2 M, 0.3 M, 0.4 M, 1 M). Two replicates of the pull down assay at 1 M NaCl were conducted to ensure the reproducibility of the assay. A WT168 LytE protein sample was performed in duplicate at all other salt concentrations.

### 2.2.17 Protein structure prediction

All protein structures presented in this work were predicted with I-TASSER (Iterative Threading ASSEmbly Refinement), an online server for protein structure predictions that scores highly in community-wide trials (Yang *et al.*, 2014). I-TASSER follows a hierarchical approach to protein structure prediction. It first identifies structural templates from protein databases and then constructs full-length atomic models by iterative template fragment assembly simulations. Protein models were then visualized with the UCSF ChimeraX software (Pettersen *et al.*, 2004).

### 2.2.18 Bioinformatics and computational biology

Bioinformatics analyses were conducted with various programs, all of which are currently available online and freely accessible. Simple sequence alignments and phylogenetic trees were constructed with the CLC Sequence Viewer programme, which was also used to determine primer sequences and restriction sites. Silent restriction sites were identified with the SerialCloner programme. The CellProfiler cell image analysis software (Carpenter *et al.*, 2006; Lamprecht *et al.*, 2007) was used to determine the number of excessive colonies on plates. The ImageJ software (Schneider *et al.*, 2012) was applied to quantify bands on gels and to measure cell parameters such as cell length and width. The statistical computing programming language R (R Core Team, 2015) was used to conduct statistical analysis of large data volumes such as the global LysM motif analysis and to design diagrams and graphs. R is extended by user-generated packages that fulfil specialized functions. The most important R packages used in this work were *ggplot2* – for visualisations (Wickham, 2009), *dplyr* and *tidyr* – for supplying the 'grammar' of data manipulation (Wickham, 2016; Wickham and Francois, 2015), *msa* – for multiple sequence alignments (Bodenhofer *et al.*, 2015) and *bios2mds* – for multi-dimensional scaling analysis (Pele *et al.*, 2012).

# 3 Chapter 3: Exploring the LytE / CwlO D,L-endopeptidase requirement for cell viability

## 3.1 Objective

The discovery of the LytE / CwlO synthetic lethality by Bisicchia *et al.* (2007), i.e. the fact that *B. subtilis* requires one or the other of these enzymes for viability, indicates that they perform an overlapping essential function for the cell. This function is probably related to the elongation of the cell wall during growth as reported by Bisicchia *et al.* (2007) and Hashimoto *et al.* (2012). In view of the requirement of one or other enzyme for viability, it is interesting that the structures of LytE and CwlO differ significantly. Although they possess similar C-terminal domains (the region that encodes the D,L-endopeptidase activity), they have completely different N-terminal domains. The N-terminal domain of CwlO has a coiled-coil structure that mediates the interaction with the FtsEX ATP transporter-type protein complex that is required for enzyme activity (Dominguez-Cuevas *et al.*, 2013; Meisner *et al.*, 2013; Yamaguchi *et al.*, 2004). In contrast, the N-terminal domain of LytE is composed of three LysM motifs, which bind to peptidoglycan (Ishikawa *et al.*, 1998). Furthermore, these two D,L-endopeptidases are expressed at different times of the growth cycle and under different conditions. CwlO and LytE are both expressed in growing cells under the control of the WalRK two-component system. However, LytE is also expressed during cell wall stress and in stationary phase under the control of the SigI RsgI sigma factor / antisigma factor (Tseng *et al.*, 2011). Furthermore, CwlO expression is also regulated on a posttranscriptional level. The *cwlO* transcript has a 187 bp leader sequence that makes it very unstable (half-life of less than 1 minute) (Noone *et al.*, 2014). When this leader sequence is deleted, the *cwlO* transcript is greatly stabilized (half-life of ~6 minutes). Considering the differences in both structure and expression profile between these two D,L-endopeptidases, it is very curious indeed that they are able to complement for each other.

It is known that the CwlO and LytE D,L-endopeptidases can complement for each other under normal growth conditions. However, the synthetic lethality of these two enzymes affords the opportunity to investigate the cellular requirement for D,L-endopeptidase activity in greater detail, because the activity and function of one D,L-endopeptidase can be examined in the absence of the other (i.e. in strains expressing one enzyme

with a null mutation in the other). This approach overcomes the issue of enzyme redundancy (there are 35 autolysins in *B. subtilis*) that has hampered investigations of autolysins activity and function in bacteria. Thus, the following questions will be addressed in this section:

(i) How effectively does the LytE D,L-endopeptidase complement the loss of CwlO D,L-endopeptidase enzymatic activity in *Bacillus subtilis*?

(ii) Can LytE complement the function of CwlO, when it is expressed in exponentially growing cells under the control of the *cwlO* promoter instead of the *lytE* promoter?

(iii) Does the *cwlO* leader have a similar destabilizing effect on the *lytE* transcript as it does on the *cwlO* transcript?

## 3.2 Experimental Approach

To determine whether the LytE D,L-endopeptidase, when expressed from the *cwlO* promoter, can complement the loss of the CwlO D,L-endopeptidase in exponential and stationary phase, strain LU17 was constructed. In this strain the *cwlO* open reading frame is replaced by the 1005 bp *lytE* open reading frame at the *cwlO* locus, resulting in the following genotype: $P_{cwlOleader}$*lytE* Δ*cwlO* (see Figure 3.1). The endogenous *lytE* gene was then deleted from the chromosome of this strain to generate a second strain (LU21) which encodes a single copy of *lytE* located at the *cwlO* locus ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*). LytE expression in this strain is solely dependent on the *cwlO* promoter. A second feature of this study was to determine if the *cwlO* leader sequence destabilizes the newly constructed hybrid *cwlOleaderlytE*$_{ORF}$ mRNA. Therefore, two strains were constructed (Figure 3.1) with *lytE* expressed using the *cwlO* promoter at the normal *cwlO* locus – (i) in the first strain the *lytE* transcript contains the leader sequence (LU21, $P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) and (ii) in the second strain (LU25, $P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*) the *lytE* transcript does not contain the leader sequence. The *cwlO* locus of both LU21 and LU25 strains was sequenced to confirm the successful genetic modifications and a test for a *lytE* suppressor mutation was conducted, to rule out the possibility of a re-located *lytE* gene on the chromosome. A detailed description of how these strains were constructed is given in section 2.1.4.2. A schematic illustration of the genetic modifications undertaken and the resulting *cwlO* and *lytE* loci of both strains are shown in Figure 3.1.

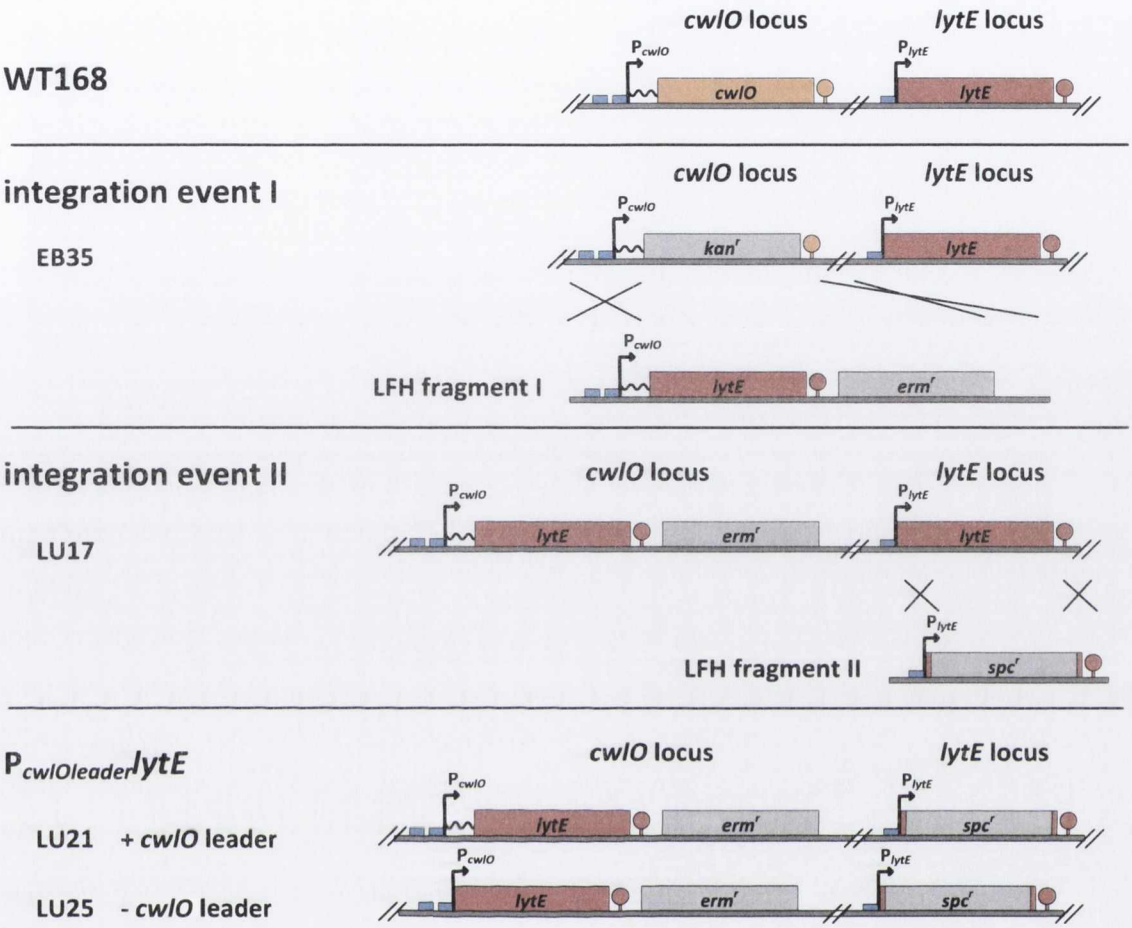



**Figure 3.1: Schematic representation of the construction of strain LU17, LU21 and LU25, all of which express LytE from the *cwlO* promoter.** CwlO expression in WT168 is regulated by the WalRK recognition sites and the *cwlO* leader sequence, while LytE expression is regulated by the *lytE* promoter. To construct strain LU21 (**$P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE***) two integration events were required. In **integration event I**, a strain EB35 (in which *cwlO* is replaced by a kanamycin resistance cassette) is transformed with a long flanking homology fragment (LFH fragment I), which leads to the exchange of the kanamycin cassette with the *lytE* open reading frame and an erythromycin cassette resulting in strain LU17 ($P_{cwlOleader}$*lytE* Δ*cwlO*). A second **integration event II**, disrupts the *lytE* gene in strain LU17 through transformation with the LFH fragment II, resulting in strain LU21. Strain LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*) is constructed in the same manner as LU21, but lacks the 187 bp *cwlO* leader sequence. The presence or absence of the *cwlO* leader sequence is indicated with +/-. Open reading frames for *lytE* and *cwlO* are not contiguous.

It is important to note the following: (i) that the *lytE* gene is present in single copy in strains LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) and LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*); (ii) that *lytE* is expressed using the heterologous *cwlO* promoter in both strains and (iii) that the *lytE* transcript contains a (destabilizing) leader sequence in strain LU21, but is without the *cwlO* leader sequence in strain LU25 and (iv) that the endogenous *cwlO* is deleted in strains LU21 and LU25.

## 3.3 Results

### 3.3.1 Wild-type LytE expression in strains that are deleted for CwlO

Bisicchia *et al.* (2007) discovered that *B. subtilis* requires either the D,L-endopeptidase activity of CwlO or LytE for viability. It was also observed that LytE and CwlO are both expressed in exponential growing cells, under the control of the WalRK TCS system. Furthermore, it has been reported that LytE expression can also be induced by SigI under stress conditions (Salzberg *et al.*, 2013). To address the question of how effectively the LytE D,L-endopeptidase can complement for loss of the CwlO D,L-endopeptidase enzymatic activity under normal growth conditions, strain LS103 (Δ*cwlO*) was constructed in which the *cwlO* gene was deleted. Expression of *lytE* was then compared between WT168 and LS103 (Δ*cwlO*) by northern blot analysis and by qRT-PCR. The results of the northern analysis (Figure 3.2) showed that deletion of *cwlO* results in expression of *lytE* being significantly increased. In wild-type strain WT168 (WT168, Figure 3.2 left panel), the *lytE* transcript was strongly expressed at the early phase of the growth cycle (T1, $OD_{600} \approx 0.2$). The transcript was then present at a lower level in exponentially growing cells, but increased when cells entered stationary phase (T5, $OD_{600} \approx 2.0$). The profile and level of *lytE* transcript was greatly increased in strain LS103 (Δ*cwlO*) that has a null mutation in the *cwlO* gene. This increase was especially strong during the exponential phase of the growth cycle i.e. $OD_{600} \approx 0.2$ to $OD_{600} \approx 1.5$ (T1 – T4).

The general trend of increased *lytE* transcript level in a Δ*cwlO* mutant was verified when quantitated by qRT-PCR (Figure 3.3). The level of *lytE* transcript in strain EB35 (Δ*cwlO*) was increased approximately 2 to 6-fold in the *cwlO* mutant (strain EB35, Figure 3.3) depending on the stage of growth. However, it would appear that a 6-fold increase in expression is somewhat below what one would have expected from that the northern analysis. Therefore, to ascertain if the increased *lytE* transcript levels also leads to an increase in the cellular level of LytE protein, a western blot analysis was performed on strains expressing LytE with a HA-epitope.

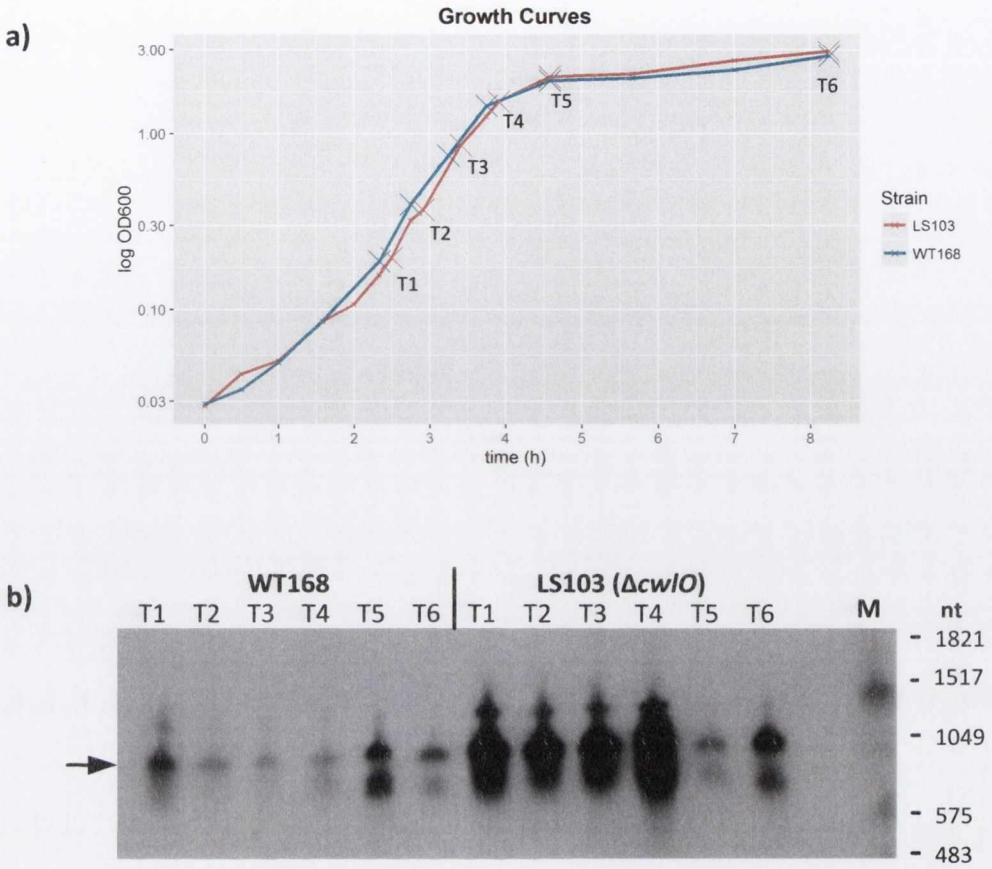



**Figure 3.2: Northern analysis of *lytE* transcript levels in WT168 and in LS103 (Δ*cwlO*).** Cells were grown in LPDM and harvested at the designated time points corresponding to the following $OD_{600}$: T1 ≈ 0.2, T2 ≈ 0.4, T3 ≈ 0.8, T4 ≈ 1.5, T5 ≈ 2.0, T6 ≈ 3.0. **a)** The growth curves of both strains are presented and the time points, at which samples were harvested are indicated. RNA was isolated from the samples and analysed by northern analysis. **b)** The northern blot shows transcript levels of *lytE* in WT168 (left) compared to strain LS103 (Δ*cwlO*, right) during the growth cycle. Five ng of total RNA was loaded onto each lane. An RNA size marker is shown (M) and the migration position and size (in nucleotides, nt) of each marker is shown. The film was exposed for 1 minute. The arrow indicates the position of the *lytE* transcript (*lytE* = 1002 nt).

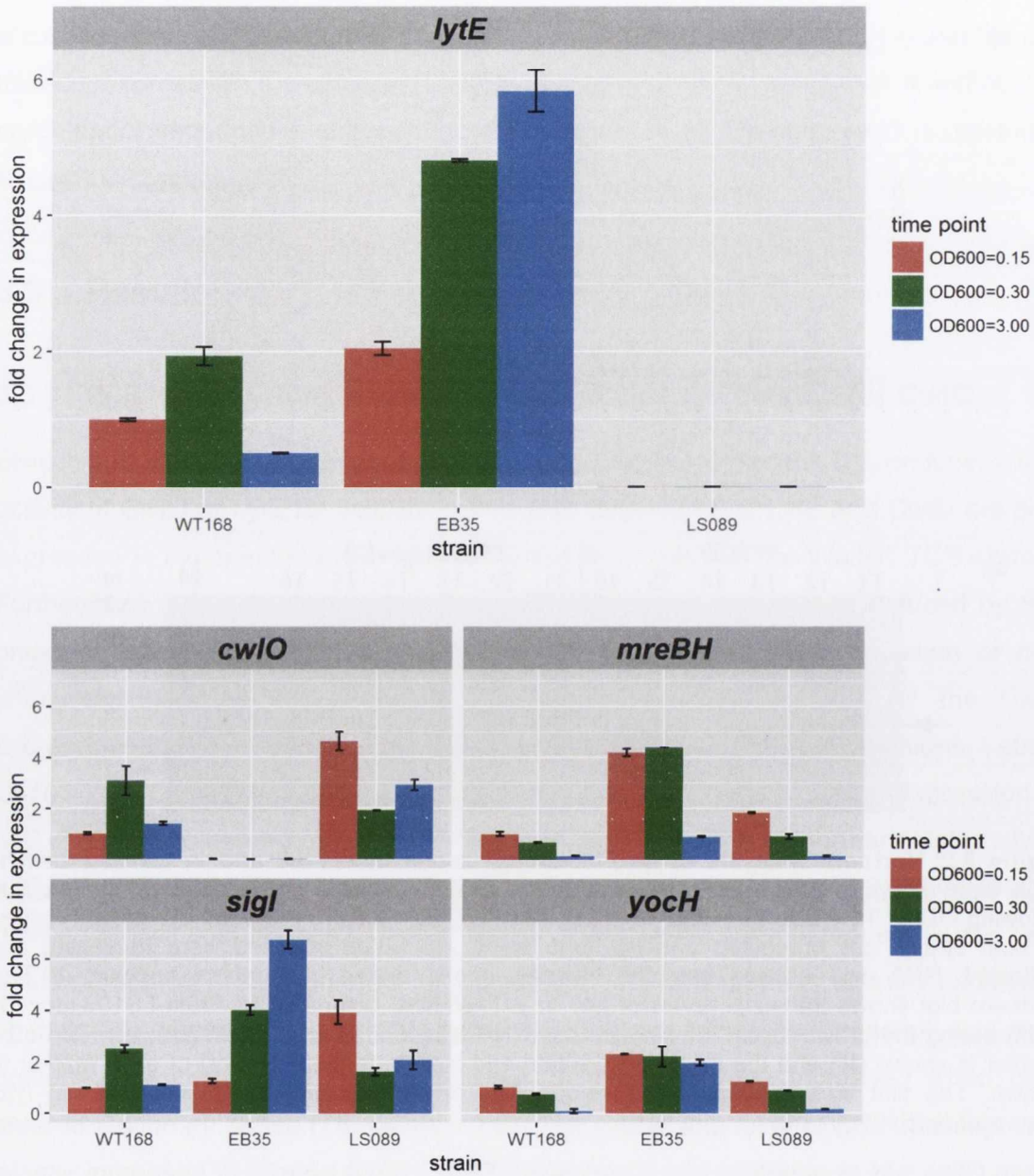



**Figure 3.3: Transcript levels of *lytE*, *cwlO*, *mreBH*, *sigI* and *yocH* determined by qRT-PCR in strains carrying null mutations in *cwlO* and *lytE*.** Cells of the strains WT168, EB35 (Δ*cwlO*) and LS089 (Δ*lytE*) were grown in HPDM and samples harvested at the indicated time points. RNA was prepared from the samples and transcribed into cDNA, which was analysed by qRT-PCR. Transcript levels were normalised relative to the level of transcript for the wild-type sample at time point $OD_{600}$=0.15, which was assigned a value of 1. Error bars are shown.

Strains LU45 (*lytE-HA*) and LU47 (*lytE-HA ΔcwlO*) are derivates of WT168 and LS103 in which the *lytE* open reading frame was tagged with a HA epitope. The strains were constructed as detailed in section 2.1.4.3 and the results of the western blot analysis are shown in Figure 3.6 (see page 76). The western blot analysis (Figure 3.6) shows the LytE protein levels for cell-associated protein and protein in the media for strains LU45 and LU47 for two time points $OD_{600} \approx 0.2$ and $OD_{600} \approx 1.0$ (T1 and T2). As is obvious from the western analysis it is clear that the level of LytE protein is greatly increased when the *cwlO* gene is deleted. The level of cell-associated LytE protein is high in cells of strain LU47 (*lytE-HA ΔcwlO*) at both T1 and T2, but is almost undetectable in strain LU45 (*lytE-HA*) at these time points (Figure 3.6, left hand side of panels a and b). Interestingly for strain LU47, the level of LytE protein did not vary between the two time points (T1 and T2) for cell-associated samples, but significantly less LytE protein was detected in the media for T2 than for T1. These results show that the LytE protein levels, both cell-associated and accumulated in the medium, are greatly elevated in strains that carry a null mutation in the *cwlO* gene (LU47) when compared with similar strains (LU45) in which the *cwlO* gene is intact.

The *lytE* gene is a member of the SigI regulon (Tseng *et al.*, 2011). Therefore, a possible mechanism by which *lytE* expression and LytE protein levels are increased is through activation of the SigIRsgI transcriptional regulators as proposed by Salzberg *et al.* (2013). To verify this hypothesis, the transcript levels of SigI and of members of the SigI regulon were measured by qRT-PCR in strains EB35 (*ΔcwlO*) and LS089 (*ΔlytE*) and compared to those observed in wild-type strain WT168. The results are presented in Figure 3.3. *SigI* transcript levels were elevated, when compared to WT168, in the *cwlO* mutant (EB35) at $OD_{600} = 0.30$ and $OD_{600} = 3.0$ by about 4 and 6.5-fold, although the levels at the first time point ($OD_{600} = 0.15$) did not appear to be increased. The level of *mreBH* transcript, a constituent gene of the SigI regulon that encodes an actin-like protein, were also elevated in strain EB35 (*ΔcwlO*) at time points $OD_{600} = 0.15$ and $OD_{600} = 0.3$ by ~4-fold when compared to WT168. The fact that *sigI* and *mreBH* transcript levels were both elevated in the *cwlO* mutant, indicates that the SigI regulon is upregulated. This confirms that deletion of the *cwlO* gene causes activation of the SigIRsgI transcriptional regulators and leads to increased expression of LytE.

Furthermore, qRT-PCR analysis also verified that strain EB35 (*ΔcwlO*) is indeed null mutant for the *cwlO* gene as no *cwlO* transcript was detected (Figure 3.3). Interestingly qRT-PCR analysis also detected a very small increase in *yocH* transcript levels by about 2-fold for EB35 (*ΔcwlO*) (Figure 3.3). YocH is an amidase, important for cell wall turnover and part of the WalR regulon (Salzberg *et al.*, 2013).

In summary, it is evident that although LytE can fulfil the D,L-endopeptidase requirement in a *cwlO* mutant, the absence of CwlO leads to transcriptional changes that alter *lytE* gene expression. Disruption of the *cwlO* gene creates a stress stimulus that is (probably) sensed by RsgI, resulting in elevated levels of activated SigI which in turn increase LytE expression. This suggests that the standard low level of LytE protein normally expressed by the WalRK system in wild-type cells is likely not sufficient to fulfil the D,L-endopeptidase requirement when CwlO is absent.

### 3.3.2 Expression of LytE from the *cwlO* promoter

#### 3.3.2.1 *lytE* transcript levels

It was shown in the previous section that LytE expression is increased through activation of SigI in strains deleted for *cwlO*. Since (i) *cwlO* expression is dependent on WalRK activity and (ii) WalRK activity varies significantly throughout the growth cycle (Noone *et* al., 2016, submitted) it is evident that the WalRK and SigIRsgI systems work in tandem to ensure adequate expression of D,L-endopeptidase activity (i.e. CwlO or LytE) under normal growth and stress conditions. To further address these issues, it was decided to establish if LytE could fulfil the D,L-endopeptidase requirement when expressed from the *cwlO* promoter (that is completely dependent on WalRK activity). Therefore, *lytE* expression was established in strains LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*), and LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*), in which expression is directed by the *cwlO* promoter. The results of a qRT-PCR based determination of *lytE* transcript levels in strains LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) and LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*) are presented in Figure 3.4. The results show that the level of *lytE* transcript in strain LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) was 20 to 30-fold elevated in exponentially growing cells ($OD_{600}$ = 0.15 & 0.3) and up to 8-fold in stationary phase cells ($OD_{600}$ = 3.0) when compared to the level expressed in wild-type strain WT168 at similar time-points (Figure 3.4). It is also evident that the *cwlO* leader sequence affects the level of *lytE* transcript (Figure 3.4). The level of *lytE* transcript in exponentially growing cells of LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*) was increased 35 to 40-fold when compared with that in cells of WT168. This compares with a 20 to 25-fold increase in *lytE* transcript in cells of strain LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*). The increase in the cellular level of *lytE* transcript in strain LU25 strongly suggests that the *cwlO* leader sequence destabilizes the hybrid mRNA ($_{cwlOleader}$*lytE*) in a manner similar to its action on the *cwlO* transcript.

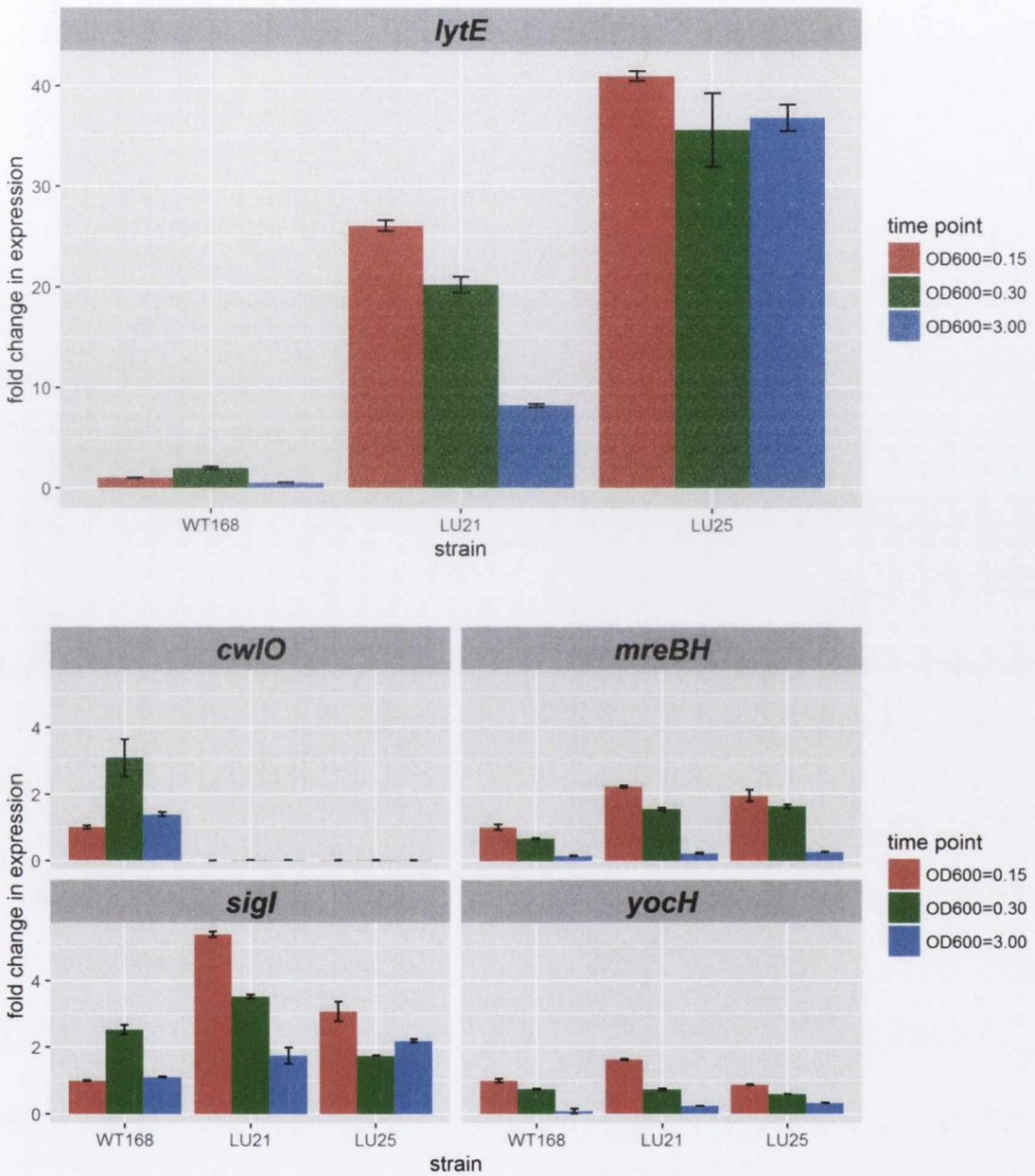



**Figure 3.4: Transcript levels of *lytE*, *cwlO*, *mreBH*, *sigI* and *yocH* determined by qRT-PCR for strains that express LytE from the *cwlO* promoter**. Cells of the strains WT168, LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) and LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*) were grown in HPDM and samples harvested at the indicated time points. RNA was prepared from the samples and transcribed into cDNA, which was analysed by qRT-PCR. Transcript levels were normalised relative to the level of transcript for the wild-type sample at time point $OD_{600}$=0.15, which was assigned a value of 1. Error bars are shown.

In summary, these data show that the natural *cwlO* promoter is more active than the *lytE* promoter both in wild-type cells growing exponentially and in cells in which $P_{lytE}$ activity is increased due to SigI activation caused by deletion of the *cwlO* gene. Furthermore, the *cwlO* leader sequence destabilizes the hybrid $_{cwlOleader}$*lytE* transcript in a manner similar to its action on the *cwlO* transcript.

#### 3.3.2.2 The level of cell wall stress in strains LU21 and LU25 that express *lytE* using the *cwlO* promoter

It has been shown (see section 3.3.1) that deletion of *cwlO* leads to activation of SigIRsgI and increased LytE expression that fulfils the D,L-endopeptidase requirement. The constituent genes of the SigI regulon encode proteins involved in cell wall functions (e.g. LytE D,L-endopeptidase, BcrC undecaprenyl pyrophosphate phosphatase and the actin-like protein MreBH (Tseng *et al.*, 2011; Tseng and Shaw, 2008).

Together these results suggest that the SigIRsgI regulators are sensitive to cell wall stress, particularly that associated with D,L-endopeptidase deficiency. Thus, we sought to use SigIRsgI activation in strains LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) and LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*) as a measure of the extent to which (i) the D,L-endopeptidase requirement is fulfilled by placing *lytE* expression under the control of the *cwlO* promoter and (ii) the level of cell wall stress generated by the very high level of LytE expression in both strains.

Transcript levels of *cwlO*, *mreBH*, *sigI* and *yocH* were measured by qRT-PCR in total RNA samples prepared from cells grown in HPDM media. Samples were taken at three points throughout the growth curve with two samples taken from exponentially growing cells ($OD_{600}$ = 0.15 and $OD_{600}$ = 0.3) and one sample taken from late stationary phase cells ($OD_{600}$ = 3.0). The level of *sigI* transcript (expression of this operon is positively autoregulated) and of *mreBH* (a constituent gene of the SigI regulon) was measured and the results are presented in Figure 3.4. As expected, no *cwlO* transcript could be detected in either strain LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) or LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*), verifying that the *cwlO* open reading frame has been replaced by the *lytE* gene and that a copy of *cwlO* is not present elsewhere on the chromosome. Measurements of SigI transcript levels showed an increase in strain LU21 of ~5-fold at $OD_{600}$ = 0.15, but only a very slight increase of ~2 to 3-fold at later time points ($OD_{600}$ = 0.3 and $OD_{600}$ = 3.0). *SigI* transcript levels were also slightly elevated in strain LU25 with the greatest increase of ~3-fold being detected at $OD_{600}$ = 0.15. The level of *mreBH* transcript was elevated ~2-fold, especially during the exponential phases ($OD_{600}$ = 0.15

and $OD_{600}$ = 0.3) of the growth cycle (Figure 3.4). However, unlike the case of *sigI*, there is little difference in *mreBH* transcript levels between strain LU21 and LU25. Transcript levels of *yocH* (a member of the WalR regulon) were comparable to WT168 levels (Figure 3.4). The fact that transcript levels of SigI regulon genes (*sigI* and *mreBH*) were slightly elevated in strains LU21 and LU25 suggests that the expression of LytE from the *cwlO* promoter in a Δ*cwlO* Δ*lytE* background, although sufficient for ensuring viability and growth, generates a modest stress stimulus in the cell. Interestingly, based on *sigI* transcript levels in the cell, the stress stimulus would appear to be smaller in strain LU25 than it is in strain LU21, suggesting that removal of the *cwlO* leader sequence that leads to elevated levels of LytE transcript and protein, is beneficial. A notable feature of these data is that the elevation in *sigI* transcript in exponentially growing cells of strains EB35 (Δ*cwlO*) and LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) are comparable, but is reduced when *lytE* expression is further increased in LU25 by removal of the *cwlO* leader sequence. These data suggest that although expression of LytE from the *cwlO* promoter can fulfil the D,L-endopeptidase requirement for cell viability, it does not complement CwlO function completely and that further elevation of transcript is required (by removal of the leader sequence) to achieve a state without a cell wall stress (i.e. no elevation of *sigI* expression) being generated.

#### 3.3.2.3 Destabilizing effect of the *cwlO* leader sequence on the *lytE* transcript

Noone *et al.* (2014) demonstrated that *cwlO* expression is regulated at the post-transcriptional level by a 187 bp long leader sequence positioned upstream of the open reading frame on the mRNA transcript. This leader sequence causes the *cwlO* transcript to be highly unstable due to cleavage by RNase Y (Noone *et al.*, 2014). To establish if the leader sequence also destabilizes the hybrid *lytE* transcript (i.e. $_{cwlOleader}$*lytE*) that is expressed using the $P_{cwlO}$ promoter, the half-life of the *lytE* transcript in strain LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) and LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*) was assessed by northern analysis. Strain LU21 and LU25 were grown in HPDM and transcription was stopped by the addition of rifampicin at an $OD_{600}$ = 0.5 (T0). The degradation of *lytE* mRNA was monitored by harvesting samples every minute following rifampicin addition. RNA was prepared from the samples and analysed by northern analysis. The results are shown in Figure 3.5.

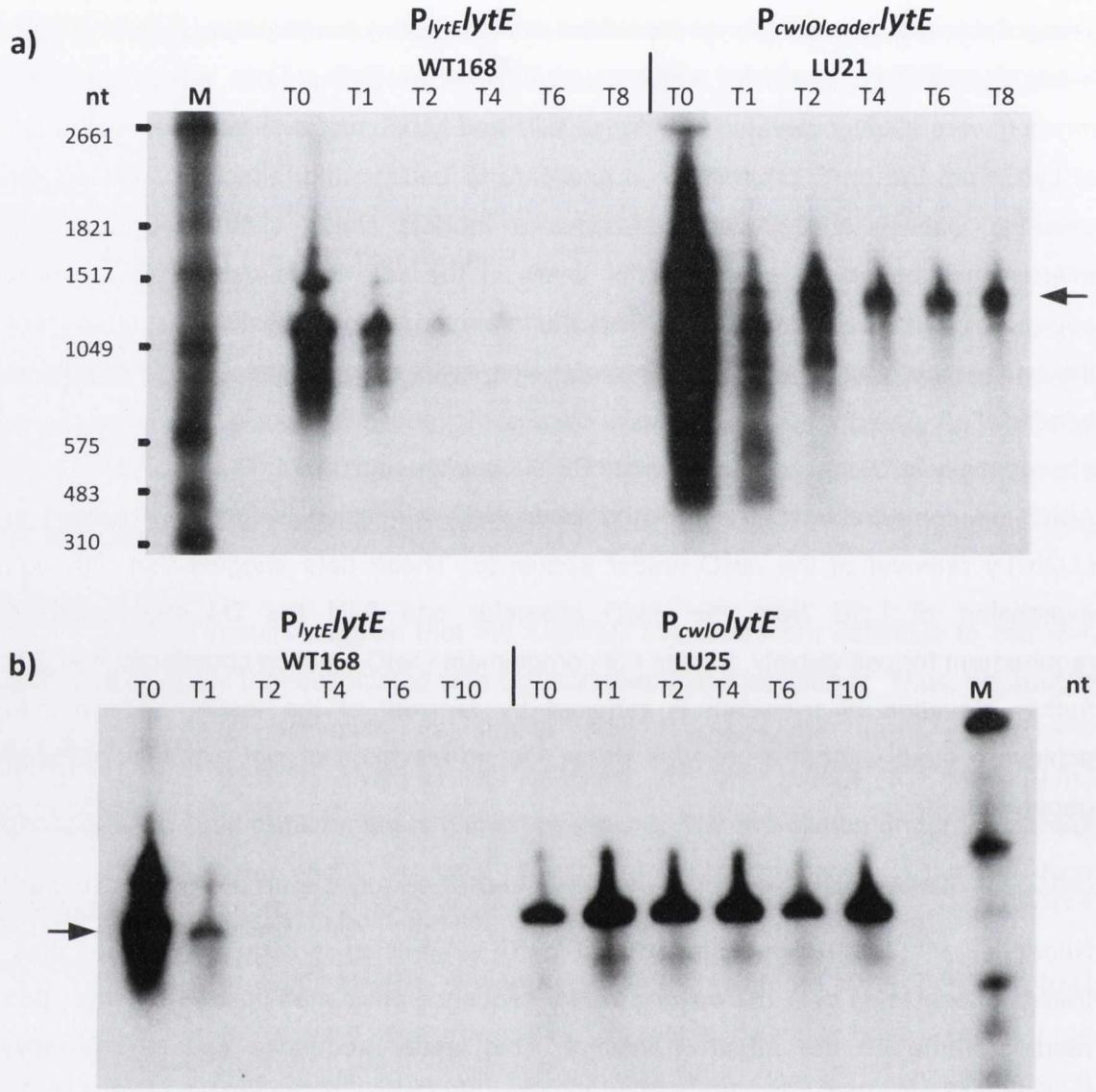



**Figure 3.5: Assessment of the stability of the *lytE* transcript expressed from the *cwlO* promoter by northern analysis.** The stability of the *lytE* transcript expressed with the *cwlO* leader sequence (LU21, $P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) **(a)** and without the *cwlO* leader sequence (LU25, $P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*) **(b)** was established by northern analysis and compared to the stability of the wild-type *lytE* transcript (WT168). Cells were grown in HPDM to an $OD_{600}$ = 0.5 (T0) at which point transcription was terminated by addition of rifampicin (final concentration 125 µg/ml). Time points indicate minutes after rifampicin addition. RNA was prepared from the samples and analysed by northern analysis. Ten ng of total RNA was loaded onto each lane for all samples except LU25 (2 ng RNA per lane). An RNA size marker is shown (M) and the migration position and size (in nucleotides, nt) of each marker is shown. The arrow indicates the *lytE* transcript (*lytE*, 1002 nt). The film was exposed for 1 minute.

Northern analysis shows (i) that the level of *lytE* transcript was elevated with respect to wild-type strain WT168 as expected (panel a); (ii) that the *lytE* transcript in strains WT168 and LU21 were extensively smeared (especially at T0 and T1) suggesting that it is processed and (iii) that the *lytE* transcript was very unstable in strain LU21 with the transcript being rapidly degraded and reduced to a very low level over a short time period (≈ 8 minutes). Removal of the *cwlO* leader sequence (strain LU25, $P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*) affects the *lytE* transcript profile in several ways [compare profiles of LU21 (contains *cwlO* leader sequence) with LU25 (no leader sequence)]: (i) there is less smearing of the hybrid $_{cwlO}$*lytE* transcript and (ii) the level of transcript does not decrease during the 10 minute period of the experiment (Figure 3.5). These data show that when present, the *cwlO* leader sequence destabilizes the transcript of the $_{cwlOleader}$*lytE* hybrid.

### 3.3.2.4 Determination of the cellular level of LytE protein when expressed using the *cwlO* promoter

Previous results have shown a high level of *lytE* transcript in cells of strain LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) when the gene is expressed under the control of the *cwlO* promoter at that locus (Figure 3.4, see page 71). Moreover, the level of transcript is further elevated when *lytE* is expressed without the *cwlO* leader sequence (strain LU25; $P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*) (section 3.3.2.3). Therefore, to establish whether the higher level of *lytE* transcript in these strains results in a higher level of LytE protein, a western blot analysis was performed as described in Chapter 2. In order to detect the LytE protein by western analysis, it was necessary to generate derivative strains in which LytE was tagged with the HA-epitope. The following strains were constructed as detailed in Chapter 2 (section 2.1.4.3): LU45 (*lytE-HA*); LU47 (*lytE-HA* Δ*cwlO*); LU61 ($P_{cwlOleader}$*lytE-HA* Δ*cwlO* Δ*lytE*) and LU63 ($P_{cwlO}$*lytE-HA* Δ*cwlO* Δ*lytE*). These represent LytE-HA tagged version of respectively WT168, LS103, LU21 and LU25.

The results are shown in Figure 3.6. It is clear that the level of cell-associated LytE protein in cells of strain LU45 is barely detectable, but that it is greatly increased in strain LU47 in which *cwlO* is deleted (Figure 3.6, panel a). Interestingly, the level of cell-associated LytE-HA protein in cells of strain LU61 ($P_{cwlOleader}$*lytE-HA* Δ*cwlO* Δ*lytE*) is significantly higher than that observed in cells of strain LU47 (*lytE-HA* Δ*cwlO*). Thus, the level of cell-associated LytE-HA protein produced by expression from the *cwlO* promoter in strain LU61 is higher than that produced upon induction of SigI by deletion of *cwlO* (strain LU47).

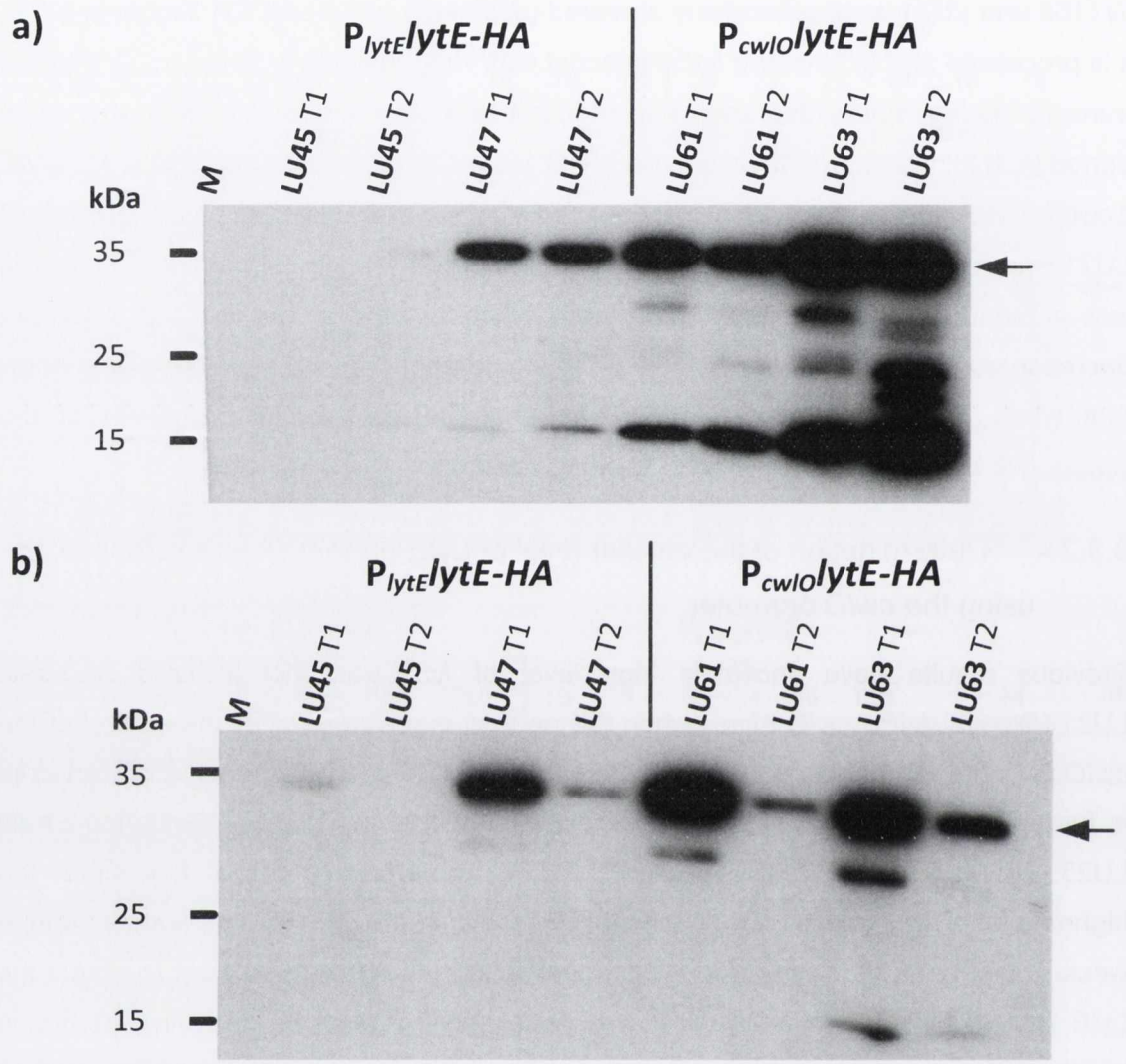



**Figure 3.6: Cell-associated and secreted LytE protein levels determined by western blot analysis.** The level of LytE protein associated with the cell **(a)** and accumulated in the medium **(b)** was established by western blot analysis in the following strains: LU45 (*lytE-HA*), LU47 (*lytE-HA* Δ*cwlO*), LU61 ($P_{cwlOleader}$*lytE-HA* Δ*cwlO* Δ*lytE*) and LU63 ($P_{cwlO}$*lytE-HA* Δ*cwlO* Δ*lytE*). Cells were grown in HPDM and harvested at two time points which correspondent to the following $OD_{600}$: T1 = 0.2, T2 = 1. Total cell protein was prepared from the samples and the HA-tagged LytE protein was detected by western analysis. Twenty µg of total cell protein extract was loaded onto each lane. The amount of protein loaded for medium samples was equivalent to 1 ml medium at an $OD_{600}$ = 0.3. Molecular weight markers are shown (M, kDa) and the film was exposed for 1 minute. The arrow indicates the position at which the full-length LytE migrates (LytE, MW = 37 kDa).

It is also clear that the level of cell-associated LytE-HA protein is very significantly higher when the *cwlO* leader sequence is removed from the hybrid gene - compare cell-associated LytE protein levels in strain LU63 ($P_{cwlO}$*lytE-HA* Δ*cwlO* Δ*lytE*, no leader sequence) with that in LU61 ($P_{cwlOleader}$*lytE-HA* Δ*cwlO* Δ*lytE*) that contains a leader sequence. There is also a high level of processed/degraded protein associated with cells of strain LU63, commensurate with the very high level of full-length LytE protein associated with the cell. A comparison of the level of LytE-HA protein that is cell-associated (Figure 3.6, panel a) and accumulated in the medium (Figure 3.6, panel b) shows there to be a direct correlation in exponentially growing cells at T1 ($OD_{600}$ = 0.2) for strain LU45, LU47 and LU61. However, the increased level of cell-associated LytE-HA in strain LU63 does not lead to an increase in level of protein that accumulates in the medium (Figure 3.6, panel a and b). It is also evident that LytE-HA protein in the medium is degraded as cells approach the stationary phase of the growth cycle (compare samples T1 and T2 in panel b).

In summary these studies show that LytE protein is expressed at a high level from the *cwlO* promoter and accumulates to a high level both in association with the cell and in the medium. Significantly, the level of protein expressed using the heterologous *cwlO* promoter is higher than that expressed by the endogenous *lytE* promoter when the latter is induced by deletion of *cwlO*.

#### 3.3.2.5 Localization of the LytE protein in cells that express it from the *cwlO* promoter

Previous studies found the LytE protein to be secreted and bound to cell septa, poles and the lateral sidewall of the cell (Hashimoto *et al.*, 2012). Furthermore, its localization is determined through interaction with the actin-like protein MreBH (Carballido-Lopez *et al.*, 2006). To establish the cellular level and location of LytE when expressed from the *cwlO* promoter, strain LU143 ($P_{cwlOleader}$*lytE-HA* Δ*cwlO*) was constructed in which a HA-epitope was attached to LytE for immuno detection purposes.

The results are presented in Figure 3.7. It is clear that there is a much higher level of LytE protein associated with the cell, when LytE is expressed from the *cwlO* promoter (LU143, Figure 3.7), then when expressed from the *lytE* promoter (LU45, Figure 3.7). The LytE protein seems to be distributed in a punctate, uneven manner along the length of the cell. Some of the LytE foci are of a very high intensity while others are significantly lower (Figure 3.7).

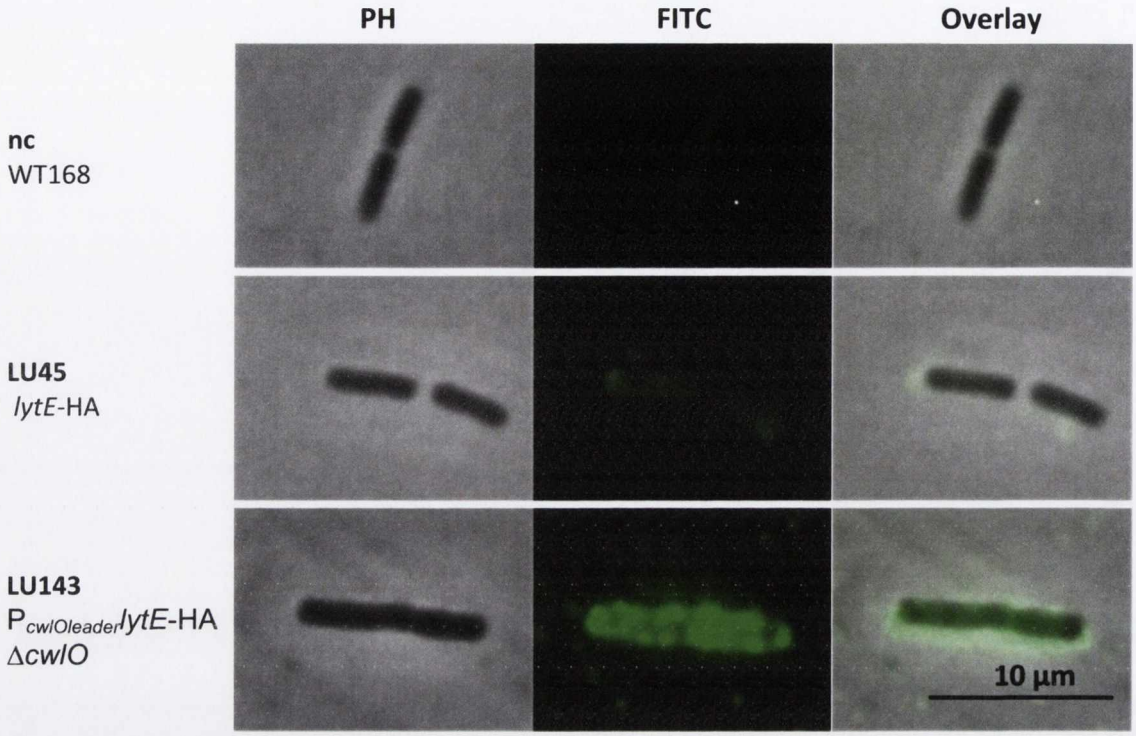



**Figure 3.7: Localization of the LytE protein.** Strains were grown in buffered LB to an $OD_{600}$ of 0.2 at which point cells were harvested, fixed in 2.6 % paraformaldehyde, immunostained and examined with an Olympus BX61 microscope. Immunostaining was achieved by endowing the LytE protein with a HA-tag and treating cells first with a primary anti-HA antibody followed by binding of the primary antibody with a fluorescein isothiocyanate (FITC)-labelled secondary antibody. Proteins were visualized by phase contrast microscopy (PH) and by immunofluorescence microscopy (FITC). Magnification is 1000x. A negative control (nc) is shown (WT168, no HA-tag). The genotypes of the examined strains as well as a scale bar are shown.

Morphologically, LytE is present along the cell cylinder and at the septum, in accordance with the observations reported by Hashimoto *et al.* (2012). From this it can be concluded that the LytE protein is located along the cell cylinder and at the septum at a high level and in a punctate manner, when expressed from the *cwlO* promoter.

### 3.3.2.6 Growth and cell morphology

There is a high level of LytE protein both associated with the cell and accumulated in the medium in the case of strains LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) and LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*), in both of which the *lytE* gene is expressed from the *cwlO* promoter (Figure 3.6, see page 76). Therefore, we sought to compare the growth, morphology and sensitivity to cell wall-acting antibiotics of these strains with the parental strain LU17 ($P_{cwlOleader}$*lytE* Δ*cwlO*), and to strains with null mutations in *cwlO* (LS103) and *lytE* (LS089), and to the wild-type strain WT168.

The observed cell morphology of these strains is presented in Figure 3.8 (a). Wild-type cells (WT168, Figure 3.8 a) were straight and rod-shaped. A *lytE* mutant (LS089, Figure 3.8 a) featured slightly twisted cells. Cells of a *cwlO* mutant (LS103, Figure 3.8 a) were wild-type like, as were cells of strain LU17. The majority of analysed cells for strain LU21 exhibited no abnormalities, but a few cells featured small bumps and mini-cells. Cells of strain LU25 were the most aberrant with a substantial fraction of cells showing bulges and thickenings.

The results of the growth studies are presented in Figure 3.8 (b). It was observed that all strains with the exception of LU25 showed a wild-type like growth behaviour. However, strain LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*) had a minor growth inhibition in all tested media (LB, LPDM, HPDM). The growth inhibition of strain LU25 was apparent during lag phase and led to a delayed onset of the exponential growth phase (LU25, Figure 3.8 b).

In addition none of the investigated strains had an increased antibiotic sensitivity for fosfomycin, bacitracin, vancomycin, d-cycloserine, penicillin or lysozyme (data not shown).

From this it can be noted that not only the absence of LytE, but also an excess of LytE protein affects cell morphology. Both Δ*lytE* mutants as well as the strain LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*) exhibited difficulties to maintain cell shape and longitudal growth.

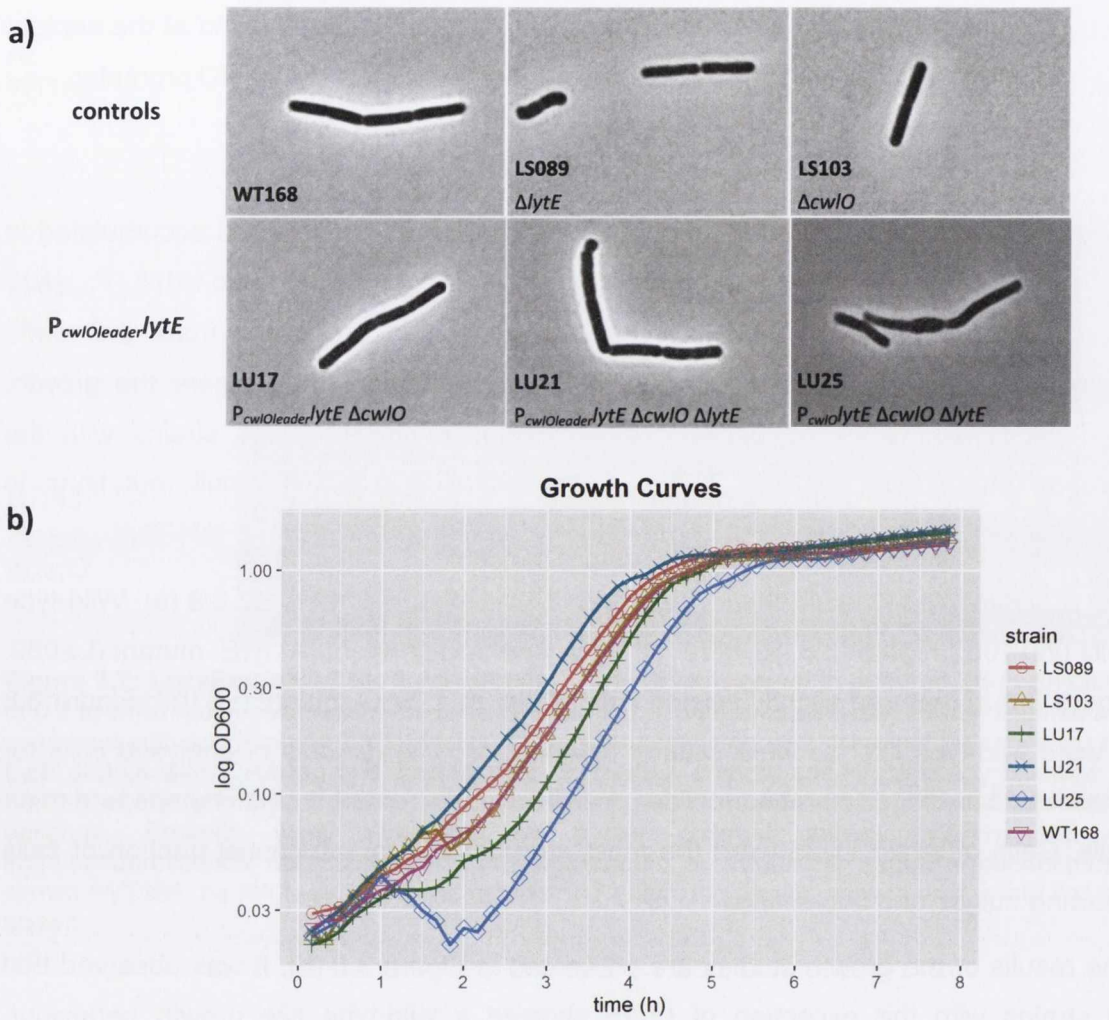



**Figure 3.8: Phenotypic characterization of strains that express LytE from the *cwlO* promoter. a) Typical cell morphology.** Cells were grown in LB to an $OD_{600}$ of 0.1, heat-fixed and examined with an Olympus BX61 microscope. Magnification is 1000x. The genotypes of the examined strains as well as a scale bar are shown. **b) Growth curves.** Growth curves were analysed for the following strains: WT168, LS089 (*ΔlytE*), LS103 (*ΔcwlO*), LU17 ($P_{cwlOleader}$*lytE ΔcwlO*), LU21 ($P_{cwlOleader}$*lytE ΔcwlO ΔlytE*) and LU25 ($P_{cwlO}$*lytE ΔcwlO ΔlytE*). Overnight cultures of those strains were used to inoculate 100 µl LPDM media in a CELLSTAR® 96 well plate to an $OD_{600}$ of 0.03 and incubated at 37°C with slow shaking, the $OD_{600}$ was measured every 10 minutes over a time period of 8 hours. Three technical repeats for each strain were conducted. One representative example is shown.

## 3.4 Conclusion

It has previously been established that the CwlO and LytE D,L-endopeptidase form a synthetically lethal pair (Bisicchia *et al.*, 2007; Hashimoto *et al.*, 2012). Therefore, the aims of this study were (i) to investigate how effectively LytE can fulfil the D,L-endopeptidase requirement for viability of *B. subtilis* in the absence of *cwlO* and (ii) to establish the extent to which LytE can complement CwlO function when it is expressed at the *cwlO* locus under the control of the *cwlO* promoter. Furthermore, the effect of the *cwlO* leader sequence on *lytE* transcript stability and proteins levels was determined.

It was established that the disruption of the *cwlO* open reading frame up-regulates the SigI regulon, which results in an increased expression of LytE in *cwlO* mutants. Furthermore, that a strain, expressing LytE under the control of the *cwlO* promoter and without the endogenous *lytE* gene, could be successfully generated (LU21, $P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*). Removal of the *cwlO* leader sequence in this strain led to an increase in *lytE* transcript and LytE protein levels (LU25, $P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*). This is because the *cwlO* leader sequence has a destabilizing effect on the *lytE* transcript, in a manner similar to that observed with the *cwlO* transcript. Growth and morphology of cells expressing LytE, under the control of the *cwlO* promoter with *cwlO* leader sequence was indistinguishable from WT168, however those in which the leader sequence was removed, exhibited a bulged cell morphology in lag phase and early exponential phase. Furthermore, a strain without the *cwlO* leader sequence also had a growth inhibition which was most apparent during the lag phase. Interestingly the time point at which these phenotypic defects were observed, correlates with the high level of WalRK observed during these growth phases (Noone *et al.*, 2016 submitted).

In both strains LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) and LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*) high levels of LytE protein were found both cell-associated and secreted into the medium. However, the LytE protein appeared to be degraded in the medium as cultures approached stationary phase. This may be caused by the proteases that are expressed at this stage of the growth cycle. The level of LytE-HA protein in cells with *lytE* expressed using the heterologous *cwlO* promoter was higher, than that produced from the SigI promoter when it was induced due to loss of CwlO function.

The extent to which LytE complements CwlO activity when expressed from the *cwlO* promoter was assessed by measuring activation of the SigI regulon (naturally induced when CwlO is not expressed). The qRT-PCR analysis demonstrated that SigI is activated when *lytE* is expressed from the *cwlO* promoter in the presence of the *cwlO* leader sequence, but not when expressed from the *cwlO* promoter without the leader

sequence. This is consistent with the fact that LytE transcript and protein levels are higher when LytE is expressed from the *cwlO* promoter without *cwlO* leader sequence.

In summary, this study confirms that the LytE D,L-endopeptidase can complement CwlO when expressed using *cwlO* expression signals - both transcriptional (i.e. promoter) and post-transcriptional (i.e. *cwlO* leader sequence). It is clear however that the two enzymes are not equivalent since the SigI regulon, which is induced by cell envelope stress, is induced when *lytE* is expressed using the normal *cwlO* expression signals (i.e. promoter and leader sequence).

# 4 Chapter 4: The Potential of CwlS, LytF and YqgT to fulfil the D,L-endopeptidase requirement

## 4.1 Objective

Bisicchia *et al.* (2007) and Hashimoto *et al.* (2012) showed that the D,L-endopeptidase activity of either CwlO or LytE is required for the viability of *B. subtilis*. The CwlO enzyme is expressed under normal growth conditions, controlled by the WalRK TCS, while LytE is expressed under stress conditions, controlled by the SigIRsgI regulatory system (Bisicchia *et al.*, 2007; Salzberg *et al.*, 2013). It was shown in the previous section that *B. subtilis* cells are also viable when expression of the LytE D,L-endopeptidase is controlled by the *cwlO* promoter. This result is especially interesting since the N-terminal domains of the CwlO and LytE D,L-endopeptidases that direct how the enzymes interact with the cell envelope, are different as detailed in the introduction to this thesis. In view of these findings, it was of interest to determine if other D,L-endopeptidases encoded by *B. subtilis* such as LytF (5 LysM motifs in the N-terminal domain), CwlS (4 LysM motifs in the N-terminal domain) and YqgT (N-terminal domain of unknown function) can also support the viability of *B. subtilis* when expressed using the *cwlO* promoter. Therefore, the aim of this project was to establish if the CwlS, LytF or YqgT D,L-endopeptidases encoded in the *B. subtilis* genome can alone fulfil the D,L-endopeptidase requirement for viability when expressed using the *cwlO* promoter.

## 4.2 Experimental Approach

The approach adopted to fulfil the stated objective was to construct a series of strains in which the *cwlS, lytF* and *yqgT* genes were individually expressed at the *cwlO* locus under the control of the natural *cwlO* promoter. The level of transcript of each gene was varied by expressing each gene both with the *cwlO* leader sequence (giving an unstable transcript) and without the leader sequence (giving a stable transcript) (see section 2.1.4.6 and Figure 4.1).

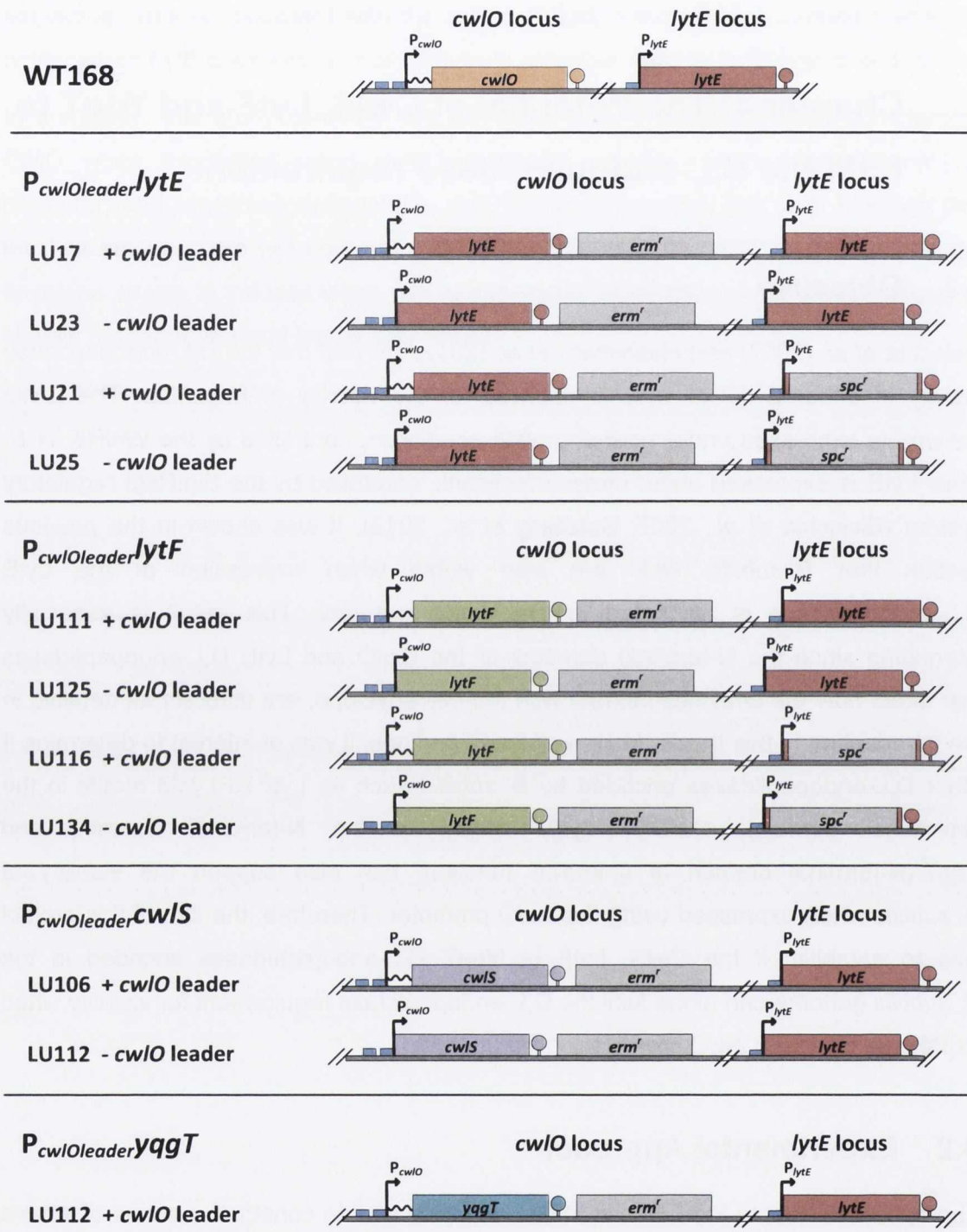



**Figure 4.1: Schematic illustration of strains that express LytE, LytF, CwlS or YqgT from the *cwlO* promoter.** The *cwlO* gene is replaced by an D,L-endopeptidase open reading frame in all featured strains (except WT168) and the *lytE* gene is disrupted by an spectinomycin resistance cassette in strains LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*), LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*), LU116 ($P_{cwlOleader}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE*) and LU134 ($P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE*). The presence or absence of the *cwlO* leader sequence is indicated with +/-. Open reading frames for *lytE* and *cwlO* are not contiguous.

The ability of each enzyme to fulfil the D,L-endopeptidase requirement for viability was then established by attempting to delete the *lytE* gene in each of the strains. Therefore, if the expressed enzyme can fulfil the D,L-endopeptidase requirement, a viable strain should be obtained with *lytE* deleted. Conversely, an inability to delete the *lytE* gene would strongly indicate that the enzyme expressed at the *cwlO* locus is unable to fulfil the D,L-endopeptidase requirement for viability of *B. subtilis*.

## 4.3 Results

### 4.3.1 The LytF, CwlS and YqgT D,L-endopeptidases individually are not synthetically lethal with either CwlO or LytE

In order to rule out the possibility of synthetic lethality between the D,L-endopeptidases encoded by *cwlS, lytF* and *yqgT* and those encoded by either *lytE* or *cwlO*, a series of strains were constructed as detailed in Table 4.1.

**Table 4.1: List of null and double mutant strains between *cwlS, lytF, yqgT* and *lytE, cwlO***

| Single mutants | Double mutants (with Δ*cwlO*) | Double mutants (with Δ*lytE*) |
|---|---|---|
| LU72 (Δ*cwlS*) | LU90 (Δ*cwlS* Δ*cwlO*) | LU92 (Δ*cwlS* Δ*lytE*) |
| LU74 (Δ*lytF*) | LU100 (Δ*lytF* Δ*cwlO*) | LU102 (Δ*lytF* Δ*lytE*) |
| LU76 (Δ*yqgT*) | LU96 (Δ*yqgT* Δ*cwlO*) | LU98 (Δ*yqgT* Δ*lytE*) |

The mutants were constructed by first creating three strains individually deleted in the *cwlS* (strain LU72), *lytF* (strain LU74) and *yqgT* (strain LU76) genes (these are markerless null mutants). The *cwlO* and *lytE* genes were then individually deleted in each of these three strains (see Table 4.1). The results showed that it was possible to construct all nine strains with *cwlS, lytF* and *yqgT* individual deleted and with each of these genes deleted in combination with either *cwlO* or with *lytE*. Thus, we can conclude that the LytF, CwlS and YqgT D,L-endopeptidases individually are not synthetically lethal with either CwlO or LytE.

### 4.3.2 Phenotypes of strains carrying null mutations in *cwlS, lytF* and *yqgT* singly and in combination with either *cwlO* or *lytE*

The phenotypes of each of the strains listed in Table 4.1 were examined. As previously stated, all these strains were obtained signifying that each is viable.

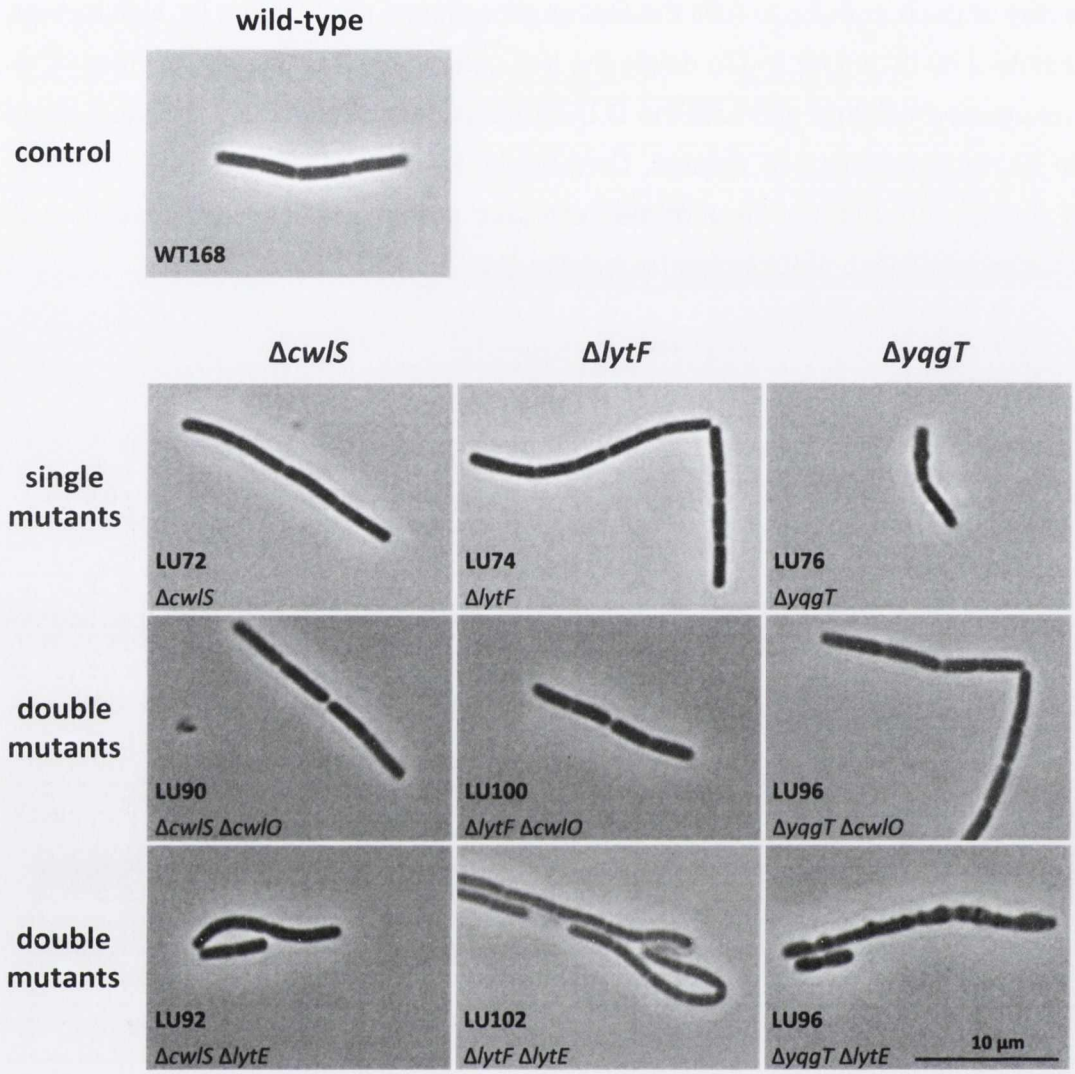



**Figure 4.2: Typical cell morphology of single (Δ*lytF*, Δ*cwlS*, Δ*yqgT*) and double mutants (in combination with Δ*lytE* and Δ*cwlO*).** Cells were grown in LB to an $OD_{600}$ of 0.1, heat-fixed and examined with an Olympus BX61 microscope. Magnification is 1000x. The genotypes of the examined strains as well as a scale bar are shown.

Morphological differences were observed when cells were grown in LB that were especially noticeable during the lag and early exponential phases of the growth cycle (see Figure 4.2). The morphology of cells of strain LU72 (Δ*cwlS*) was similar to that of wild-type strain WT168, as previously reported (Fukushima *et al.*, 2006). The morphology of cells of strain LU74 (Δ*lytF*) showed that cells were longer and exhibited more chaining in agreement with the results of Ohnishi *et al.* (1999). The morphology of cells of strain LU76 (Δ*yqgT*) showed that there were bulges in the middle and at the ends of the cell and some minicells were also observed. The morphology of cells of strain LU90 (Δ*cwlS* Δ*cwlO*) was similar to wild-type. However, cells of strain LU92 (Δ*cwlS* Δ*lytE*) were more bent and curled than cells of WT168. Cells of strain LU100 (Δ*lytF* Δ*cwlO*) were similar to wild-type cells. However, double mutants in *lytF* and *lytE* (LU102 Δ*lytF* Δ*lytE*) exhibited highly filamentous cells, an effect that had been observed before by Ohnishi *et al.* (1999). Cells of strain LU96 (Δ*yqgT* Δ*cwlO*) exhibited wild-type cell morphology, but a *yqgT* mutant combined with a *lytE* mutant (LU98 Δ*yqgT* Δ*lytE*) showed a strongly abnormal phenotype with a great number of bulges along the cell cylinder. In summary it can be concluded that phenotypes observed in the single mutants of *cwlS*, *lytF* or *yqgT* remained unaffected when combined with a *cwlO* mutant, but were amplified when combined with a *lytE* mutant.

Furthermore, the growth profiles for all mutants were examined and growth of all strains was found similar to wild-type strain WT168 in both LB and in LPDM (see Figure 4.4 a / b, page 91).

### 4.3.3 An investigation of strains expressing CwlS from the *cwlO* promoter

#### 4.3.3.1 Ability of CwlS to fulfil the D,L-endopeptidase requirement for viability in *B. subtilis*

The ability of the CwlS D,L-endopeptidase, when expressed from the *cwlO* promoter, to fulfil the D,L-endopeptidase requirement for viability was established by attempting to construct the series of strains listed in Table 4.2. The *cwlS* gene, with or without the *cwlO* leader sequence, was inserted into the chromosome of strain LU72 (Δ*cwlS*) in a manner that replaced the entire *cwlO* gene, generating strains LU106 and LU123. An attempt was then made to delete the endogenous *lytE* gene in strains LU106 and LU123, an event that made cell viability dependent on the CwlS D,L-endopeptidase expressed at the *cwlO* locus. A detailed description of how these strains were constructed is given in section 2.1.4.6.

**Table 4.2: List of strains expressing a CwlS protein from the *cwlO* promoter**

| Strain | Genotype | *cwlS* transcript | * |
|---|---|---|---|
| LU106 | $P_{cwlOleader}$*cwlS* Δ*cwlS* Δ*cwlO* | unstable *cwlS* transcript | V |
| | $P_{cwlOleader}$*cwlS* Δ*cwlS* Δ*cwlO* Δ*lytE* | unstable *cwlS* transcript | - |
| LU123 | $P_{cwlO}$*cwlS* Δ*cwlS* Δ*cwlO* | stable *cwlS* transcript | V |
| | $P_{cwlO}$*cwlS* Δ*cwlS* Δ*cwlO* Δ*lytE* | stable *cwlS* transcript | - |

* viable (V) strain constructed or strain not obtained (-)

Only two of the strains listed in Table 4.2, LU106 and LU123, could be constructed. The *lytE* gene could not be disrupted in either of these strains despite repeated attempts to do so. These results indicate that CwlS is unable to fulfil the D,L-endopeptidase requirement for viability under these conditions.

#### 4.3.3.2 Establishment of CwlS protein levels in cells expressing it from the *cwlO* promoter

The inability of CwlS to fulfil the D,L-endopeptidase requirement for viability could be due to (i) some inherent property of the enzyme e.g. turnover, number or cellular locations or (ii) the enzyme being poorly expressed or rapidly degraded under the conditions tested. Therefore, to examine CwlS expression, the enzyme was tagged with a HA-epitope (for detection by western blot analysis) in wild-type strain WT168 (generating strain LU145 *cwlS-HA*), in strain LU106 (generating strain LU159 $P_{cwlOleader}$*cwlS-HA* Δ*cwlS* Δ*cwlO*) and in strain LU123 (generating strain LU161 $P_{cwlO}$*cwlS-HA* Δ*cwlS* Δ*cwlO*).

The results of the western blot analysis are shown in Figure 4.3. It is clear that cell-associated CwlS protein cannot be detected in extracts of strain LU145 (*cwlS-HA*) indicating that its expression is very low under these growth conditions. When expression of CwlS is driven by the *cwlO* promoter in strain LU159 ($P_{cwlOleader}$*cwlS-HA* Δ*cwlS* Δ*cwlO*), the cell-associated CwlS protein level was increased and full length CwlS protein (44 kDa) was detected (full arrow, Figure 4.3.). Furthermore, the level of CwlS protein associated with cells of strain LU161 ($P_{cwlO}$*cwlS-HA* Δ*cwlS* Δ*cwlO* - expressing the *cwlS* gene without the leader sequence) was increased when compared to that observed in strain LU169. This indicates that the *cwlO* leader sequence stabilizes the *cwlS* transcript and leads to higher levels of CwlS expression as was also observed for the *lytE* gene.

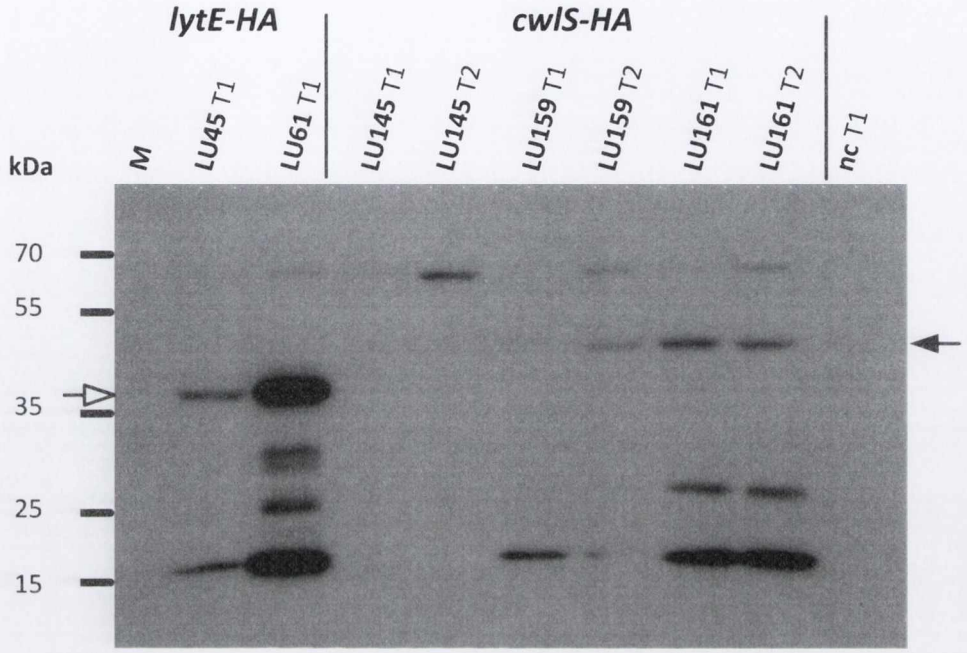



**Figure 4.3: Cell-associated protein levels of CwlS in WT168 and when expressed from the *cwlO* promoter with and without *cwlO* leader sequence in comparison to LytE expressed from the *lytE* and the *cwlO* promoter.** Cell-associated protein samples of the following strains are shown: LU45 (*lytE-HA*), LU61 ($P_{cwlOleader}$*lytE-HA* Δ*cwlO* Δ*lytE*), LU145 (*cwlS-HA*), LU159 ($P_{cwlOleader}$*cwlS-HA* Δ*cwlS* Δ*cwlO*) and LU161 ($P_{cwlO}$*cwlS-HA* Δ*cwlS* Δ*cwlO*). A negative control (nc, WT168, without HA-tag) is also displayed. Cells were grown in HPDM and harvested at two time points which correspondent to the following $OD_{600}$: T1 = 0.2, T2 = 1. Total cell protein was prepared from the samples and the HA-tagged proteins were detected by western analysis. Twenty µg total cell protein extract was deployed per lane. A molecular weight marker is shown (M, kDa). The film was exposed for 10 minutes. The full arrow indicates the CwlS protein (CwlS, MW = 44 kDa) and the open arrow indicates the LytE protein (LytE, MW = 37 kDa).

It is notable however that there is a very significant increase in the levels of two smaller bands (MW ≥ 25 kDa and ≥ 15 kDa) suggesting that the CwlS enzyme is proteolytically processed when expressed at a high level (see LU161, Figure 4.3).

The level of cell-associated CwlS protein in strains LU145, LU159 and LU161 was then compared with that of LytE protein expressed under normal growth conditions (LU45 *lytE-HA*) and expressed from the *cwlO* promoter in the absence of both *cwlO* and of the endogenous *lytE* gene (LU61, $P_{cwlOleader}$*lytE-HA ΔcwlO ΔlytE*). The levels of these proteins can be roughly compared since both are visualized using the same HA-epitope and antibody. It is clear that the level of full-length CwlS is strain LU161 is approximately equal to that of LytE when the latter is expressed, from its natural promoter under normal growth conditions (see LU45, Figure 4.3), but is very much lower than the level of cell-associated LytE protein that is observed when it is expressed from the *cwlO* promoter without the leader sequence (see LU61, Figure 4.3). In conclusion this data shows that the level of cell-associated CwlS protein is undetectable under normal growth conditions. However, although CwlS is expressed at a high level in strain LU161 under the control of the *cwlO* promoter, a very significant portion of the protein appears to be processed (bands ≥25 kDa and ≥15 kDa) leaving only a relatively low level of full-length enzyme. This low level of cell-associated CwlS is very much lower than that observed in strain LU61, when *lytE* is expressed from the *cwlO* promoter. Therefore, it is possible that the inability of CwlS to fulfil the D,L-endopeptidase requirement for viability is due to significant processing of the enzyme upon expression.

#### 4.3.3.3 Growth and morphology of strains LU106 and LU123 that express CwlS using the *cwlO* promoter

The growth and morphology of strains LU106 ($P_{cwlOleader}$*cwlS ΔcwlS, ΔcwlO*) and LU123 ($P_{cwlO}$*cwlS ΔcwlS ΔcwlO*) were investigated. A representative growth curve is presented in Figure 4.4 c).

Pictures of cells with the typical morphology of each strain are shown in Figure 4.7 (see 99). Strain LU106 ($P_{cwlOleader}$*cwlS ΔcwlS ΔcwlO*) exhibited slightly shorter cells than wild type and no growth defect. However, cells of strain LU123 ($P_{cwlO}$*cwlS ΔcwlS ΔcwlO*) featured an altered cell morphology: cells were somewhat bent and some had small excrescences (like horns) on their ends. Furthermore, there was a significantly increased lag phase associated with strain LU123.

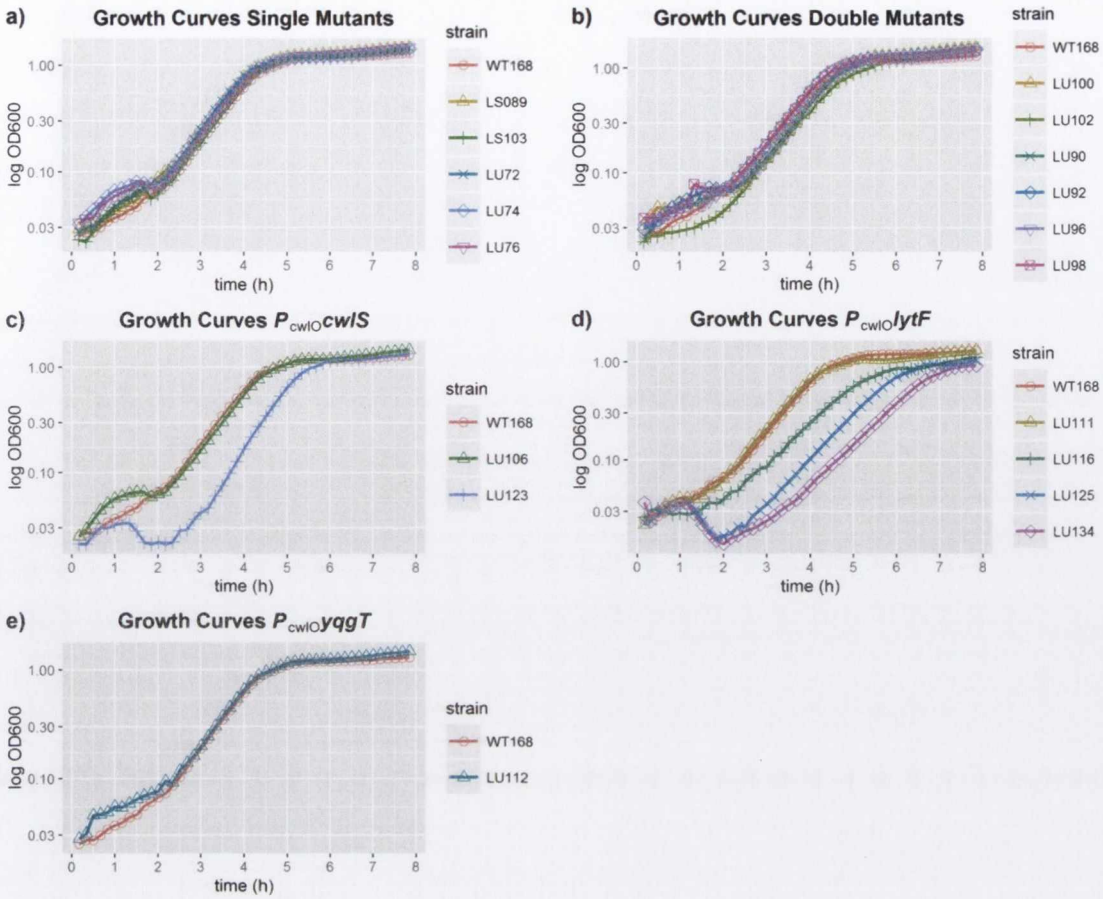



**Figure 4.4: Growth curves for strains that express one of the following D,L-endopeptidase LytE, LytF, CwlS or YqgT from the *cwlO* promoter.** Growth curves were analysed for the following strains: **a)** LS089 (Δ*lytE*), LS103 (Δ*cwlO*), LU72 (Δ*cwlS*), LU74 (Δ*lytF*), LU76 (Δ*yqgT*); **b)** LU90 (Δ*cwlS* Δ*cwlO*), LU92 (Δ*cwlS* Δ*lytE*), LU96 (Δ*yqgT* Δ*cwlO*), LU98 (Δ*yqgT* Δ*lytE*), LU100 (Δ*lytF* Δ*cwlO*), LU102 (Δ*lytF* Δ*lytE*); **c)** LU106 ($P_{cwlOleader}$*cwlS* Δ*cwlS* Δ*cwlO*), LU123 ($P_{cwlO}$*cwlS* Δ*cwlS* Δ*cwlO*); **d)** LU111 ($P_{cwlOleader}$*lytF* Δ*lytF* Δ*cwlO*), LU116 ($P_{cwlOleader}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE*), LU125 ($P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO*), LU134 ($P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE*); **e)** LU112 ($P_{cwlOleader}$*yqgT* Δ*yqgT* Δ*cwlO*). The wild-type strain (WT168) is shown as a control. Overnight cultures of those strains were used to inoculate 100 µl LPDM media in a CELLSTAR® 96 well plate to an $OD_{600}$ of 0.03 and incubated at 37°C with slow shaking, the $OD_{600}$ was measured every 10 minutes over a time period of 8 hours. Three technical repeats for each strain were conducted. One representative example is shown.

In summary, it is concluded that when positioned at the *cwlO* locus under the control of the *cwlO* promoter, CwlS protein is expressed and remains cell-associated. However, the level of cell-associated CwlS protein is low. Although the level can be increased by stabilizing the *cwlS* mRNA through removal of the *cwlO* leader sequence, this results in significant degradation of the CwlS protein. Expressing CwlS instead of CwlO in strain LU106 does not affect the growth profile, but leads to smaller cells. However, increasing CwlS expression (LU123) leads to a growth defect and to severe morphological abnormality in some cells. These phenotypes indicate that the CwlS protein produced in strains LU106 and LU123 is functional.

#### 4.3.3.4 Localization of the CwlS protein in cells of strain LU159

Previous studies found that the CwlS protein is secreted and bound to cell septa and cell poles (Hashimoto *et al.*, 2012). To establish its cellular location when expressed at elevated levels using the *cwlO* promoter the CwlS protein was visualized in exponentially growing cells by immunostaining in strain LU159 ($P_{cwlOleader}$*cwlS-HA* Δ*cwlS* Δ*cwlO*). The results are presented in Figure 4.8 (page 100). The data showed that CwlS was localized predominantly at the cell septa and cell poles, which is in agreement with the observations reported by Hashimoto *et al.* (2012). However, a low level of punctate staining is also observed along the cell cylinder in strain LU159. This cellular distribution is indeed interesting: CwlS protein containing 4 LysM motifs appears to be restricted to the septa and poles of cells, whereas LytE protein with 3 LysM motifs is distributed throughout the cell cylinder and at the septa and poles.

#### 4.3.3.5 Conclusion

In summary, these data show that the CwlS D,L-endopeptidase protein is expressed using the *cwlO* promoter and is probably functional. However, CwlS cannot alone fulfil the D,L-endopeptidase requirement in *B. subtilis*. Furthermore, high CwlS protein levels affect both growth and cellular morphology.

### 4.3.4 An investigation of strains expressing LytF from the *cwlO* promoter

#### 4.3.4.1 Ability of LytF to fulfil the D,L-endopeptidase requirement for viability in *B. subtilis*

The ability of the LytF D,L-endopeptidase, when expressed from the *cwlO* promoter, to fulfil the D,L-endopeptidase requirement for viability was established by attempting to construct the series of strains listed in Table 4.3. The *lytF* gene, either with or without the *cwlO* leader sequence, was inserted into the chromosome of strain LU74 (Δ*lytF*) in

a manner that replaced the entire *cwlO* gene, generating strains LU111 and LU125. An attempt was then made to delete the endogenous *lytE* gene in strains LU111 and LU125, an event that made cell viability dependent on the LytF D,L-endopeptidase expressed at the *cwlO* locus. A detailed description of how these strains were constructed is given in section 2.1.4.6.

**Table 4.3: List of strains expressing a LytF protein from the *cwlO* promoter**

| Strain | Genotype | *lytF* transcript | * |
|---|---|---|---|
| LU111 | $P_{cwlOleader}$*lytF* Δ*lytF* Δ*cwlO* | unstable *lytF* transcript | V |
| LU116 | $P_{cwlOleader}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE* | unstable *lytF* transcript | V |
| LU125 | $P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO* | stable *lytF* transcript | V |
| LU134 | $P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE* | stable *lytF* transcript | V |

* viable (V) strain constructed or strain not obtained (-)

All of the listed strains (Table 4.3) were successfully constructed. Thus, the *lytE* gene could be disrupted in strain LU111 (generating strain LU116) and in strain LU125 (generating strain LU134).

Both the *cwlO* and *lytE* genes are deleted in strains LU116 and LU134 and the *lytF* gene is expressed from the *cwlO* promoter in the presence (LU116) and absence (LU134) of the *cwlO* leader sequence. A test was performed (as detailed in Chapter 2 section 2.2.9) to assess if the generation of strains LU116 and LU134 was accompanied by the occurrence of extragenic suppressors. The results of these transformation experiments indicate that generation of extragenic suppressors did not occur during the process of *lytE* gene deletion (data not shown). It is concluded therefore that LytF can fulfil the D,L-endopeptidase requirement for viability of *B. subtilis* when it is expressed using the *cwlO* promoter.

#### 4.3.4.2 Establishment of *lytF* transcript levels in cells expressing it from the *cwlO* promoter

The capability of LytF to fulfil the D,L-endopeptidase requirement for viability, when expressed using the *cwlO* promoter prompted us to establish *lytF* transcript levels in a selection of the strains listed in Table 4.3 and to compare them with those observed in cells of wild-type strain WT168.

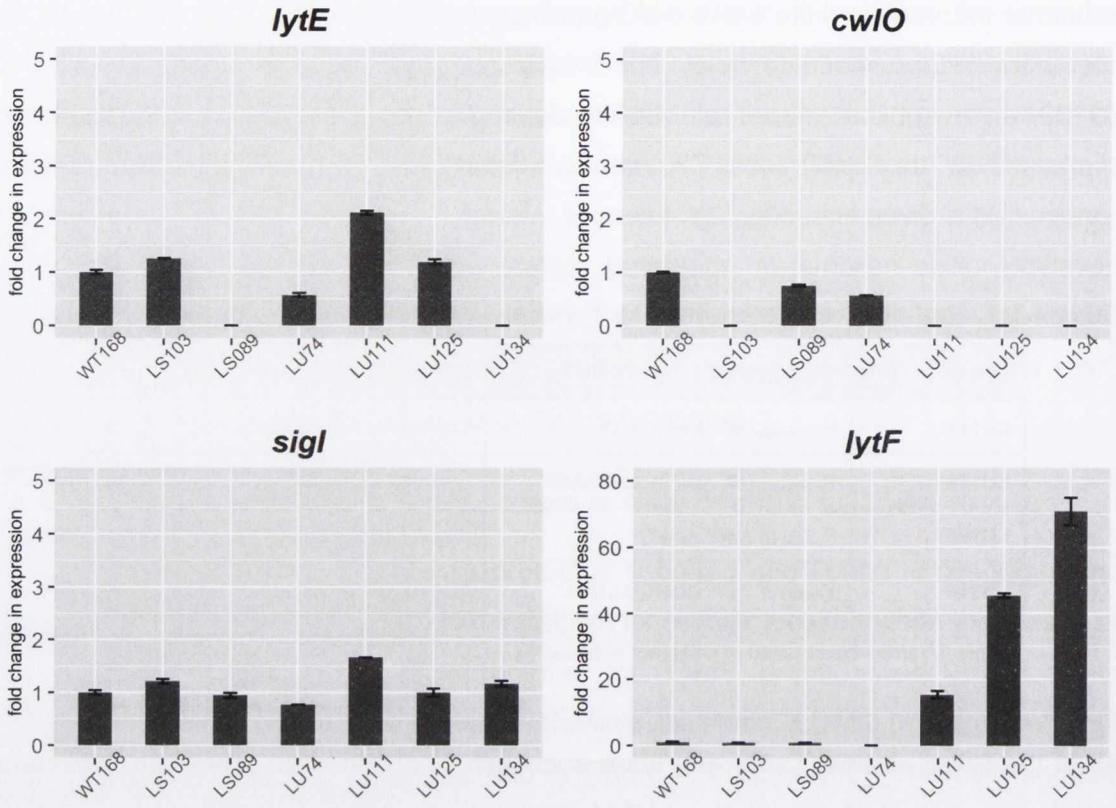



**Figure 4.5: Transcript levels of *lytE, lytF, cwlO* and *sigI* in strains that express LytF from the *cwlO* promoter.** Transcript levels were determined for the following strains: WT168, LS103 (Δ*cwlO*), LS089 (Δ*lytE*), LU74 (Δ*lytF*), LU111 ($P_{cwlOleader}$*lytF* Δ*lytF* Δ*cwlO*), LU125 ($P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO*) and LU134 ($P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE*). Cells were grown in HPDM and samples harvested at $OD_{600}$=0.15. RNA was prepared from the samples and transcribed into cDNA, which was analysed by qRT-PCR. Transcript levels were normalised relative to the level of transcript for the WT168 sample, which was assigned a value of 1. Error bars are shown.

Thus, the levels of *lytE*, *cwlO*, *lytF* and *SigI* transcripts were measured by qRT-PCR in strains LU111 ($P_{cwlOleader}$*lytF* Δ*lytF* Δ*cwlO*), LU125 ($P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO*) and LU134 ($P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE*) and were compared with those found in wild-type strain WT168 and in strains LU74 (Δ*lytF*), LS089 (*ΔlytE*) and LS103 (Δ*cwlO*). The data is presented in Figure 4.5.

When expressed using the *cwlO* promoter (strains LU111, LU125 and LU134), the level of *lytF* mRNA transcript was highly elevated when compared to that observed in strains in which it was expressed from its natural promoter (WT168, LS103 and LS089) (Figure 4.5). It is clear that the *lytF* transcript is elevated >15-fold in strain LU111, >40-fold in strain LU125 and ~70-fold in strain LU134. The increase in *lytF* mRNA level in LU125 ($P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO*) over that observed in LU111 ($P_{cwlOleader}$*lytF* Δ*lytF* Δ*cwlO*) indicates that the *cwlO* leader sequence destabilizes the *lytF* transcript.

To assess the level of cell wall stress generated, the level of *sigI* and *lytE* transcripts were also measured. It is clear that the levels of *sigI* and *lytE* transcripts were slightly elevated in strain LU111 (~2-fold), but were present at normal levels in strain LU125 and LU134. This suggests that significant cell wall stress is not generated in these strains. The absence of the *cwlO* and *lytE* genes from strain LU134 is also verified by this qRT-PCR determination of transcript levels. The mRNA transcript of neither gene was detected in strain LU134 by this analysis.

In summary, these data show that *lytF* is expressed at a very significantly higher level under the control of the *cwlO* promoter than it is under the control of its natural promoter. It is clear that as expected, transcript from neither *cwlO* nor *lytE* is present in strain LU134 indicating that expression of *lytF* at a high level can fulfil the D,L-endopeptidase requirement for viability in *B. subtilis*.

#### 4.3.4.3 Establishment of LytF protein levels in cells expressing it from the *cwlO* promoter

The level of *lytF* transcript is very high in cells of strains LU125 and LU134 in which the gene is expressed using the *cwlO* promoter (see previous section). We then sought to establish the extent to which LytF protein accumulated in the cells of these strains. To achieve this, it was necessary to tag the LytF protein in strains LU125 and LU134 with a HA-epitope in order to detect the protein by western blot analysis.

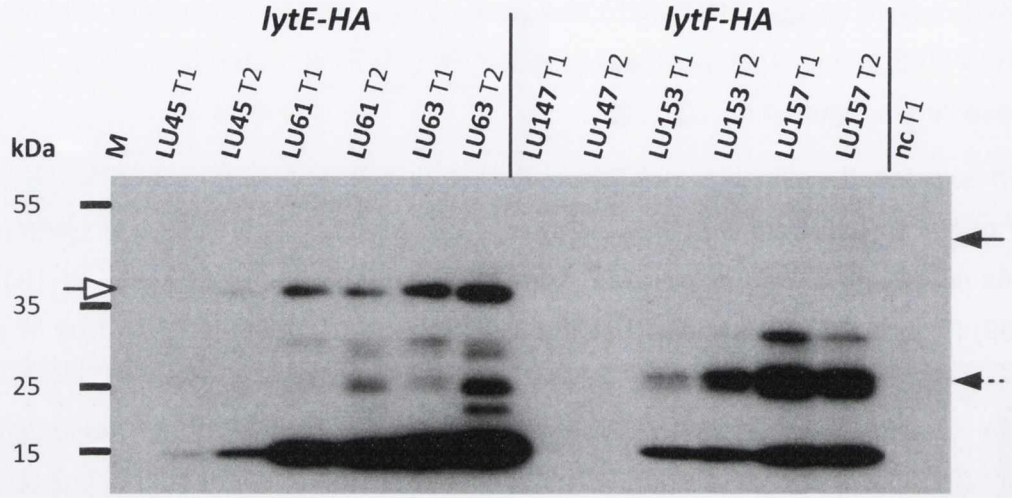



**Figure 4.6: Cell-associated protein levels of LytE and LytF in WT168 and when expressed from the *cwlO* promoter with and without *cwlO* leader sequence.** Cell-associated protein samples of the following strains are shown: LU45 (*lytE-HA*), LU61 ($P_{cwlOleader}$*lytE-HA* Δ*cwlO* Δ*lytE*), LU63 ($P_{cwlO}$*lytE-HA* Δ*cwlO* Δ*lytE*), LU147 (*lytF-HA*), LU153 (Δ*lytF* $P_{cwlOleader}$*lytF-HA* Δ*cwlO* Δ*lytE*) and LU157 (Δ*lytF* $P_{cwlO}$*lytF-HA* Δ*cwlO* Δ*lytE*). A negative control (WT168 without HA-tag) is also displayed. Cells were grown in HPDM and harvested at two time points which correspondent to the following $OD_{600}$: $t_1 = 0.2$, $t_2 = 1$. Total cell protein was prepared from the samples and the HA-tagged proteins were detected by western analysis. Ten µg total cell protein extract was deployed per lane. Molecular weight markers are shown (M, kDa). The film was exposed for one minute. The full arrow indicates the LytF protein (LytF, MW = 51 kDa), the open arrow indicates LytE protein (LytE, MW = 37 kDa), and the broken arrow indicates the degraded LytF protein.

This was achieved as detailed in Chapter 2 (section 2.1.4.7) and resulted in the generation of strains LU153 (*$P_{cwlOleader}$lytF-HA ΔcwlO ΔlytF ΔlytE*) and LU157 (*$P_{cwlO}$lytF-HA ΔcwlO ΔlytF ΔlytE*) in which the LytF protein is expressed with and without the *cwlO* leader sequence respectively. Furthermore, strain LU147 (*lytF-HA*) was constructed so that the levels of LytF proteins that accumulated in strains expressing it from the *cwlO* promoter could be compared to the level of LytF protein normally present in cells. The results of the western blot analysis are shown in Figure 4.6.

Western analysis shows that the level of LytF protein expressed under these conditions (strain LU147, *lytF-HA*) is very low and beneath detection (Figure 4.6). However, full-length LytF protein was detected in another experiment in which the film was exposed for longer (see Figure 4.9, page 103). When expression of LytF is driven by the *cwlO* promoter in strains LU153 (*$P_{cwlOleader}$lytF-HA ΔcwlO ΔlytF ΔlytE*) and LU157 (*$P_{cwlO}$lytF-HA ΔcwlO ΔlytF ΔlytE*), the levels of cell-associated LytF protein were increased. However, the levels of full length LytF protein (MW = 51 kDa) remained very low and this protein form is perhaps barely detectable in sample LU157 T2. Instead there is a very large increase in the level of proteins that represent processed forms of LytF. The two major processed forms of LytF migrate with a size of ~25 and 15 kDa. A further band is present at a lower level in strain LU157 migrating with a size of ~30 kDa (Figure 4.6). Again it is clear that removing the *cwlO* leader sequence results in an increased level of these bands supporting the view that these proteins are indeed processed forms of LytF.

A comparison was made between the cellular levels of LytE and LytF proteins present in wild-type cells and in cells in which these proteins are individually expressed from the *cwlO* promoter (Figure 4.6). It is possible to make a rough comparison since both LytE and LytF are tagged with the same HA-epitope in all strains. It is clear that full-length LytE protein is detectable at reasonably high levels in strains LU61 and LU63, whereas no full-length protein is detectable in strain LU153 and LU157. However, it is very clear indeed that overexpression of these two proteins leads to a very high level of processing of the full-length protein (Figure 4.6).

A further interesting feature of these results is that since cells of strain LU153 and LU157 are viable when expressing the LytF-HA protein (in the absence of both *cwlO* and *lytE*), it can be concluded that (i) the tagged protein must be functional and (ii) that some level of functional full-length protein must be present in cell, albeit at a level beneath our detection levels.

In summary, these results show that overexpression of LytF-HA protein results in significant processing of the protein. However, a very low level of full-length protein must be present in cells to maintain viability in the absence of both CwlO and LytE.

#### 4.3.4.4 Growth and morphology of strains LU111, LU116, LU125 and LU134 that express LytF using the *cwlO* promoter

Cell morphology and growth of strains LU111, LU116, LU125 and LU134 were investigated and the results are presented in Figure 4.7 and Figure 4.4 d (see page 91).

The data shows that strains LU111, LU116, LU125 and LU134 that express LytF using the *cwlO* promoter showed anomalous growth and cell morphology. The least affected strain was strain LU111 ($P_{cwlOleader}$*lytF* Δ*lytF* Δ*cwlO*), which had wild-type like growth and displayed only minor cell morphology defects mostly on the cell poles which were slightly more bulged or flatter than usual. Removal of the *cwlO* leader sequence in the same background (LU125 $P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO*) resulted in major cell morphology defects with cells exhibiting excrescences as well as a growth defects with delayed lag phase.

All observed effects were equally or more pronounced in strains in which the endogenous *lytE* gene was disrupted. Strain LU116 ($P_{cwlOleader}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE*) and strain LU134 ($P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE*) displayed major growth defects (delayed lag phase) and cell morphology defects. These included bent cells, thickenings on the cell poles, minicells and excrescences. All cell morphology defects were most pronounced during exponential growth and early lag phase. In addition the strains lacking *lytE* e.g. LU116 ($P_{cwlOleader}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE*) and LU134 ($P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE*) developed suppressors when grown on plates, which turned up within a week when incubated at room temperature (data not shown).

In conclusion, expression of LytF using the *cwlO* promoter leads to very serious growth and cell morphology defects, which are exacerbated by increased LytF expression (removal of the leader sequence) or by deletion of the endogenous *lytE* gene.

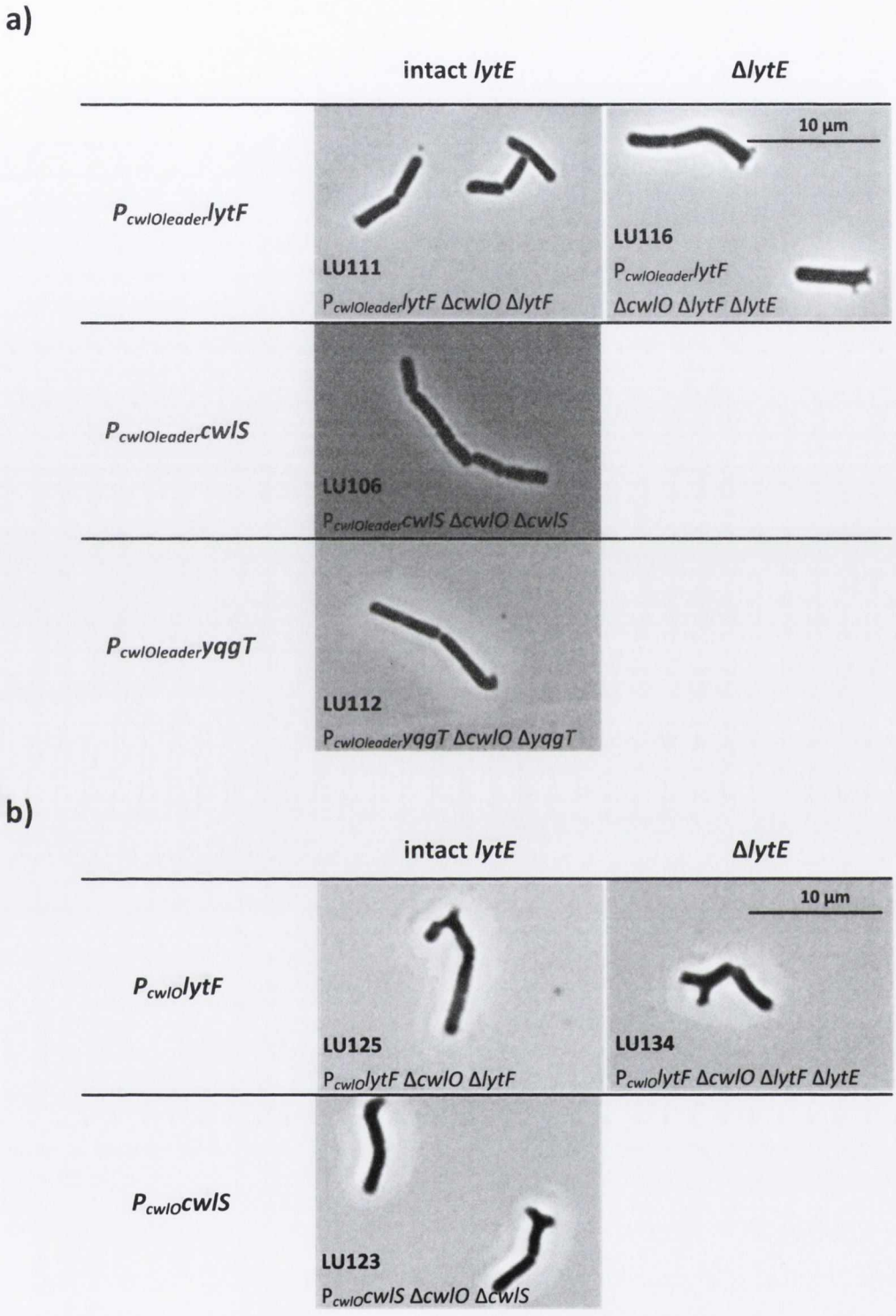



**Figure 4.7: Typical cell morphology of strains that express either LytF, CwlS or YqgT from the *cwlO* promoter.** D,L-endopeptidases are either expressed from the *cwlO* promoter with *cwlO* leader sequence **(a)** or from the *cwlO* promoter without *cwlO* leader sequence **(b)**. Cells were grown in LB to an $OD_{600}$ of 0.1, heat-fixed and examined with an Olympus BX61 microscope. Magnification is 1000x. The genotypes of the examined strains as well as a scale bar are shown.

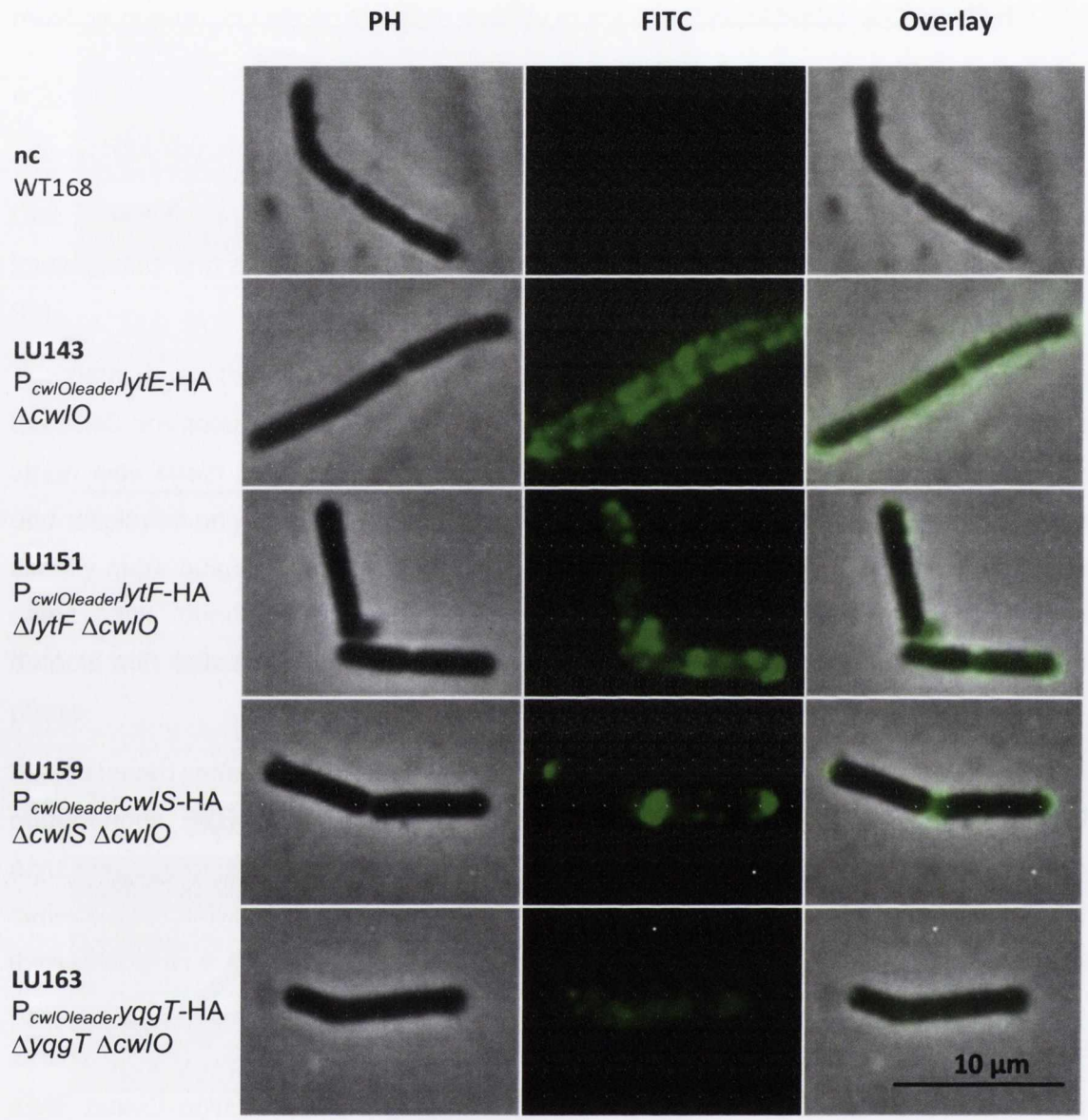



**Figure 4.8: Localization of the LytE, LytF, CwlS and YqgT proteins in strains in which the protein is expressed from the *cwlO* promoter.** Strains were grown in buffered LB to an $OD_{600}$ of 0.2 at which point cells were harvested, fixed in 2.6 % paraformaldehyde, immunostained and examined with an Olympus BX61 microscope. Immunostaining was achieved by endowing the respective protein with a HA-tag and treating cells first with a primary anti-HA antibody followed by binding of the primary antibody with a fluorescein isothiocyanate (FITC)-labelled secondary antibody. Proteins were visualized by phase contrast microscopy (PH) and by immunofluorescence microscopy (FITC). Magnification is 1000x. A negative control (nc) is shown (WT168, no HA-tag). The genotypes of the examined strains as well as a scale bar are shown.

#### 4.3.4.5 Localization of the LytF protein in cells of strain LU151

Previous studies found the LytF protein to be secreted and bound to cell septa and cell poles (Hashimoto *et al.*, 2012; Kiriyama *et al.*, 2014). To establish whether a change in the regulation of LytF expression (by driving high level expression from the *cwlO* promoter) alters these localization patterns, LytF protein was visualized by immunostaining in strain LU151 ($P_{cwlOleader}$*lytF-HA* Δ*lytF* Δ*cwlO*). The result is presented in Figure 4.8.

It is clear that LytF is localized at the cell septa and cell poles, but that it is also present along the lateral cell cylinder. Localization along the cell cylinder differs from what was observed by Hashimoto *et al.* (2012). However, it is possible that the very high LytF protein levels in strain LU151 (see Figure 4.6, page 96) might cause binding to regions of the cell not normally found in cells of WT168.

#### 4.3.4.6 Conclusion

In summary, the LytF protein alone can fulfil the D,L-endopeptidase requirement for viability of *B. subtilis* when expressed using the *cwlO* promoter. However, the level of full-length protein in such cells is very low, while the level of processed protein is greatly increased. The protein is localized at both the septum and along the cell cylinder, with the highest intensity at the former site. Furthermore, cells expressing only the LytF endopeptidase display aberrant growth and cell morphology. Thus, while cell viability can be achieved by replacing the CwlO and LytE D,L-endopeptidases with expression of LytF, the cells are clearly very 'sick' in all of the aspects tested.

### 4.3.5 An investigation of strains expressing YqgT from the *cwlO* promoter

#### 4.3.5.1 Ability of YqgT to fulfil the D,L-endopeptidase requirement for viability in *B. subtilis*

The ability of the YqgT D,L-endopeptidase, when expressed from the *cwlO* promoter, to fulfil the D,L-endopeptidase requirement for viability was established by attempting to construct the series of strains listed in Table 4.4. The *yqgT* gene, either with or without the *cwlO* leader sequence, was inserted into the chromosome of strain LU76 in a manner that replaced the entire *cwlO* gene, generating strain LU112. An attempt was then made to delete the endogenous *lytE* gene in strains LU112 an event that made cell viability dependent on the YqgT D,L-endopeptidase expressed at the *cwlO* locus. A detailed description of how this strain was constructed is given in section 2.1.4.6.

**Table 4.4: List of strains expressing a YqgT protein from the *cwlO* promoter**

| Strain | Genotype | *yqgT* transcript | * |
|---|---|---|---|
| LU112 | $P_{cwlOleader}$*yqgT* Δ*yqgT* Δ*cwlO* | unstable *yqgT* transcript | V |
| | $P_{cwlOleader}$*yqgT* Δ*yqgT* Δ*cwlO* Δ*lytE* | unstable *yqgT* transcript | - |
| | $P_{cwlO}$*yqgT* Δ*yqgT* Δ*cwlO* | stable *yqgT* transcript | - |
| | $P_{cwlO}$*yqgT* Δ*yqgT* Δ*cwlO* Δ*lytE* | stable *yqgT* transcript | - |

* viable (V) strain constructed or strain not obtained (-)

Only strain LU112 ($P_{cwlOleader}$*yqgT* Δ*yqgT* Δ*cwlO*) could be constructed. The *lytE* gene could not be disrupted in strain LU112, showing that the YqgT D,L-endopeptidase cannot complement CwlO function in the absence of the *lytE* gene. This indicates that YqgT cannot fulfil the D,L-endopeptidase requirement for cell viability. Furthermore, it was also not possible to construct a strain in which the YqgT D,L-endopeptidase is expressed using the *cwlO* promoter, but in the absence of the *cwlO* leader sequence, suggesting that such elevated levels of this protein are not compatible with cell viability.

In summary, YqgT cannot fulfil the D,L-endopeptidase requirement for viability in *B. subtilis* with a suggestion that overexpression of this enzyme is toxic to the cell.

#### 4.3.5.2 Establishment of YqgT protein levels in cells expressing it from the *cwlO* promoter

The inability of YqgT to fulfil the D,L-endopeptidase requirement may be due to an inherent property of the enzyme or to a failure to be expressed or maintained in cells using the *cwlO* promoter. To distinguish between these possibilities, strain LU163 ($P_{cwlOleader}$*yqgT-HA* Δ*cwlO* Δ*yqgT*) was constructed in which an epitope-tagged YqgT protein is expressed using the *cwlO* promoter and leader sequence as described in section 2.1.4.7.

To compare the level of YqgT protein expressed from the natural promoter with that expressed from the *cwlO* promoter, strain LU149 (*yqgT-HA*) was also constructed in which the *yqgT* gene is tagged with the HA-epitope. The presence and level of the YqgT-HA-tagged protein was then established in strains LU149 and LU163 by western analysis and the results are shown in Figure 4.9.

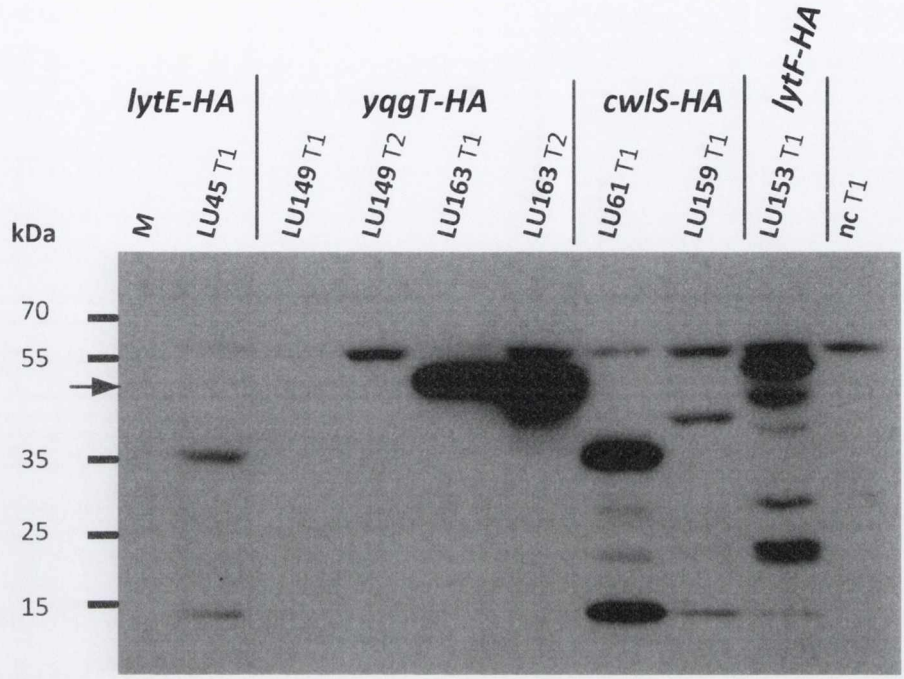



**Figure 4.9: Cell-associated protein levels of YqgT in WT168 and when expressed from the *cwlO* promoter in comparison to other D,L-endopeptidases expressed from the *cwlO* promoter.** Cell-associated protein samples of the following strains are shown: LU45 (*lytE-HA*), LU149 (*yqgT*-HA), LU163 ($P_{cwlOleader}$*yqgT-HA* Δ*yqgT* Δ*cwlO*), LU61 ($P_{cwlO}$*cwlS-HA* Δ*cwlS* Δ*cwlO*), LU159 ($P_{cwlOleader}$*cwlS-HA* Δ*cwlS* Δ*cwlO*) and LU153 ($P_{cwlOleader}$*lytF-HA* Δ*lytF* Δ*cwlO* Δ*lytE*). A negative control (WT168 without HA-tag) is also displayed. Cells were grown in HPDM and harvested at two time points which correspondent to the following $OD_{600}$: T1 = 0.2, T2 = 1. Total cell protein was prepared from the samples and the HA-tagged proteins were detected by western analysis. Twenty µg total cell protein extract was deployed per lane. A molecular weight marker is shown (M, kDa). The film was exposed for 10 minutes. The arrow indicates the predicted full-length YqgT protein (YqgT, MW = 43 kDa).

Expression of YqgT protein in strain LU149 (*yqgT-HA*) was low and could not be detected by western analysis. When expression of YqgT was driven by the *cwlO* promoter in LU163 ($P_{cwlOleader}$*yqgT-HA* Δ*cwlO* Δ*yqgT*), YqgT cell-associated protein levels were significantly increased and full length YqgT protein (43 kDa, full arrow) was detected. Interestingly, no degradation products of YqgT were detected. To compare cell-associated YqgT protein levels with those of LytE, LytF and CwlS proteins, four strains were analysed by western analysis, in each of which the protein is tagged with the same HA-epitope. Strain LU45 (*lytE-HA*) allowed the comparison of natural LytE and YqgT protein levels, by comparing strain LU45 in which the LytE protein is expressed from its natural *lytE* promoter to strain LU149 in which the YqgT protein is expressed from its natural *yqgT* promoter. It can be seen that in a WT168 background LytE protein levels were greater than YqgT protein levels. In the following strains protein expression was induced by the *cwlO* promoter: LU61 ($P_{cwlOleader}$*lytE-HA* Δ*cwlO* Δ*lytE*), LU159 ($P_{cwlOleader}$*cwlS-HA* Δ*cwlS* Δ*cwlO*), LU153 ($P_{cwlOleader}$*lytF-HA* Δ*lytF* Δ*cwlO* Δ*lytE*) and LU163 ($P_{cwlOleader}$*yqgT-HA* Δ*yqgT* Δ*cwlO*). In these strains either LytE (LU61), CwlS (LU159), LytF (LU153) or YqgT (LU163) were expressed from the *cwlO* locus with the *cwlO* leader sequence. Interestingly under these circumstances cell-associated YqgT protein levels were the highest, followed by LytE and LytF protein levels which were similar to each other and CwlS protein level were the lowest. The western blot also featured an unspecific band occurring at 65 kDa, which can be observed in the control sample (nc T1, a lysate from wild-type strain WT168 that does not express any HA-tagged protein).

It can be concluded from these data that the YqgT protein is expressed weakly in wild-type. Furthermore, it is highly expressed at both periods of the growth cycle, and remains cell-associated at a very high level, when expressed using the *cwlO* promoter. Thus, the inability to fulfil the D,L-endopeptidase requirement is not caused by failure of the protein to be expressed, to accumulate or to remain associated with the cell.

#### 4.3.5.3 Growth and morphology of strain LU112 that expresses YqgT using the *cwlO* promoter

The phenotype of strain LU112 ($P_{cwlOleader}$*yqgT* Δ*yqgT* Δ*cwlO*) was investigated by microscopy (see Figure 4.7, page 99) and growth studies (see Figure 4.4 e, page 91) to test the effect of expressing YqgT at a high level.

A representative growth curve is presented in Figure 4.4 (e) (page 91). The data shows that LU112 ($P_{cwlOleader}$*yqgT* Δ*yqgT* Δ*cwlO*) had a similar growth behaviour to WT168.

The morphology of cells of strain LU112 expressing YqgT using the *cwlO* promoter is shown in Figure 4.7 (see page 99). A small proportion of cells displayed unusual cell poles and slightly bent cells. This altered cell morphology was only apparent in early exponential phase. This observation is interesting with regard to the question whether the expressed YqgT protein is functional. Since the strain LU112 displays a different cell morphology than the double mutant LU96 (*ΔyqgT ΔcwlO*, no cell morphology defect) (see Figure 4.2, page 86) it can be concluded that the YqgT protein expressed from the *cwlO* promoter in strain LU112 is indeed functional and that an increased level of YqgT protein in this strain affects cell morphology.

#### 4.3.5.4 Localization of the YqgT protein in cells of strain LU163

The localization of the YqgT protein was visualized by immunostaining in strain LU163 ($P_{cwlOleader}$*yqgT-HA ΔyqgT ΔcwlO*) that expressed YqgT using the *cwlO* promoter. The results presented in Figure 4.8 (page 100) show that YqgT is present along the cell cylinder in a smeared pattern instead of the punctate pattern observed for autolysins such as LytE and LytF (Figure 4.8). It is also evident that YqgT does not localize at the cell septa.

#### 4.3.5.5 Conclusion

In summary, these data show that while the YqgT D,L-endopeptidase protein is expressed, accumulates and remains cell-associated in *B. subtilis* cells, when it is expressed using the *cwlO* promoter. However, YqgT cannot alone fulfil the D,L-endopeptidase requirement for viability in *B. subtilis*.

### 4.3.6 Endogenous autolysin activity

A measurement of autolysis rates of exponentially growing cells was conducted to test whether the overexpression of autolysins negatively influences viability by causing excessive degradation of the cell wall. To test this hypothesis, strains were grown in HPDM until reaching the exponential phase of growth ($OD_{600}$ = 0.5) and then transferred into 10 mM Tris-HCl (pH 8). The rate of cell autolysis was then determined by monitoring the $OD_{600}$ at 10 minute intervals (see Chapter 2 section 2.2.11). This analysis was performed on the strains listed in Table 4.5.

**Table 4.5: List of strains investigated by autolysis assay (see Figure 4.10)**

| Strain | Genotype | Strain | Genotype |
|---|---|---|---|
| LU61 | $P_{cwlOleader}$*lytE-HA ΔcwlO ΔlytE* | WT168 | |
| LU63 | $P_{cwlO}$*lytE-HA ΔcwlO ΔlytE* | LU76 | *ΔyqgT* |
| LU163 | $P_{cwlOleader}$*yqgT-HA ΔyqgT ΔcwlO* | LU72 | *ΔcwlS* |
| LU159 | $P_{cwlOleader}$*cwlS-HA ΔcwlS ΔcwlO* | LU74 | *ΔlytF* |
| LU161 | $P_{cwlO}$*cwlS-HA ΔcwlS ΔcwlO* | LS103 | *ΔcwlO* |
| LU153 | $P_{cwlOleader}$*lytF-HA ΔlytF ΔcwlO ΔlytE* | LS089 | *ΔlytE* |
| LU157 | $P_{cwlO}$*lytF-HA ΔlytF ΔcwlO ΔlytE* | | |

The results presented in Figure 4.10 show that autolysis was not affected in the strains carrying single mutants in *cwlS*, *lytF* and *yqgT* (LU72, LU74, LU76) when compared to wild-type WT168 (Figure 4.10 a). A *ΔcwlO* strain (LS103) exhibited a slightly accelerated endogenous autolysis, especially in the first two hours of the experiment (Figure 4.10 a). Surprisingly, the experiment testified that a *ΔlytE* strain (LS089) had an extremely reduced autolysis and still reached about 50 % of the original $OD_{600}$ after 6 hours (Figure 4.10 a).

This finding was confirmed by a second independent repeat of the experiment with the strain LU1 (*ΔlytE*) and EB35 (*ΔcwlO*), which have the same genotype as LS089 (*ΔlytE*) and LU103 (*ΔcwlO*), but were constructed independently (data not shown). This result is in accordance with Luo and Helmann (2012) who hypothesized that inactivation of *lytE* presumably increases β-lactam resistance by delaying cell lysis.

The two strains expressing LytE from the *cwlO* promoter, LU61 ($P_{cwlOleader}$*lytE-HA ΔcwlO ΔlytE*) and LU63 ($P_{cwlO}$*lytE-HA ΔcwlO ΔlytE*), showed a more rapid endogenous autolysis in the first two hours of the experiment than WT168, similarly to that observed with a *ΔcwlO* strain (LS103) (Figure 4.10 b).

Strains expressing CwlS from the *cwlO* promoter (LU159, LU161) behaved in a similar way and lysed more quickly than wild-type (Figure 4.10 c). The same was true for a strain expressing YqgT from the *cwlO* promoter (LU163, Figure 4.10 e).

Those strains expressing LytF from the *cwlO* promoter (LU153, LU157) exhibited the most rapid endogenous autolysis (Figure 4.10 d). Removal of the *cwlO* leader sequence caused accelerated autolysis further in strain LU157 ($P_{cwlO}$*lytF-HA ΔlytF ΔcwlO ΔlytE*) compared to strain LU153 ($P_{cwlOleader}$*lytF-HA ΔlytF ΔcwlO ΔlytE*).

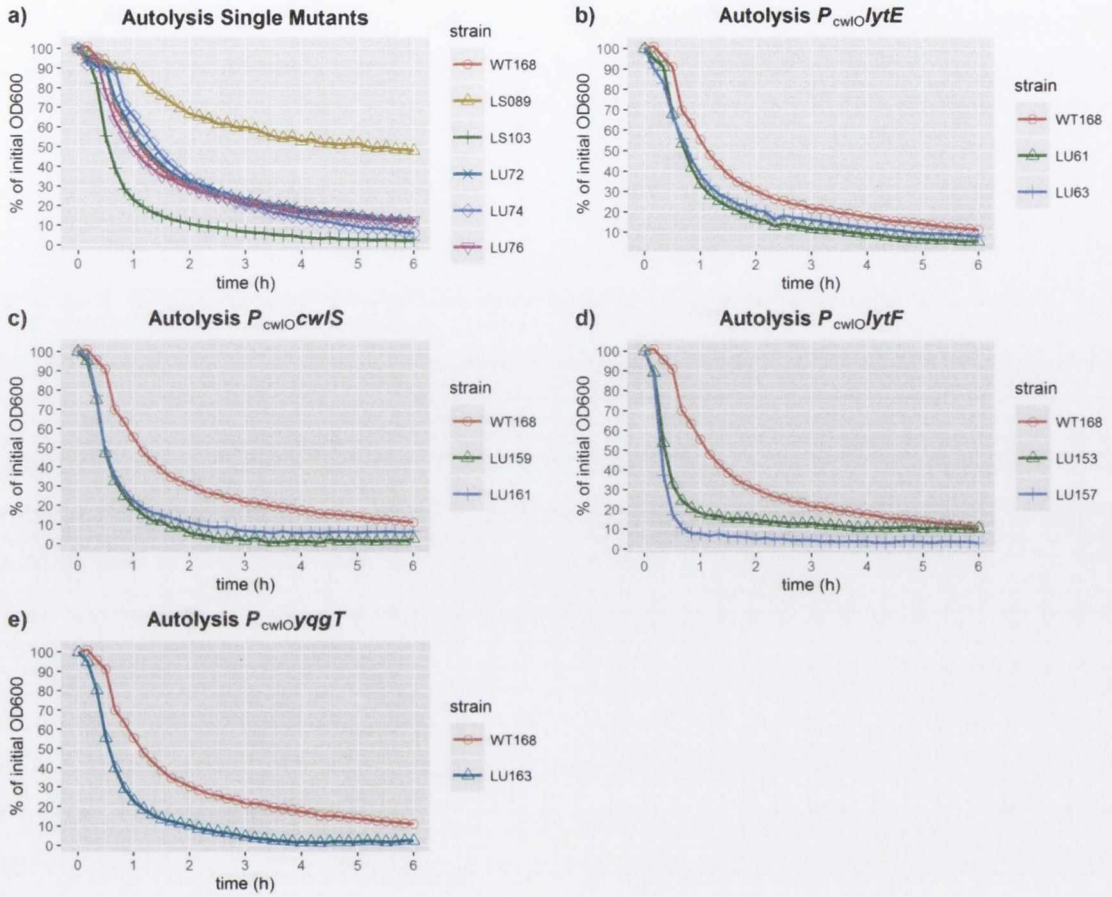



**Figure 4.10: Endogenous autolysin activity in strains that express either LytE, LytF, CwlS or YqgT from the *cwlO* promoter.** The endogenous autolysin activity was analysed for the following strains: **a)** LS089 (Δ*lytE*), LS103 (Δ*cwlO*), LU72 (Δ*cwlS*), LU74 (Δ*lytF*), LU76 (Δ*yqgT*); **b)** LU61 ($P_{cwlOleader}$*lytE-HA* Δ*cwlO* Δ*lytE*), LU63 ($P_{cwlO}$*lytE-HA* Δ*cwlO* Δ*lytE*); **c)** LU159 ($P_{cwlOleader}$*cwlS-HA* Δ*cwlS* Δ*cwlO*), LU161 ($P_{cwlO}$*cwlS-HA* Δ*cwlS* Δ*cwlO*); **d)** LU153 ($P_{cwlOleader}$*lytF-HA* Δ*lytF* Δ*cwlO* Δ*lytE*), LU157 ($P_{cwlO}$*lytF-HA* Δ*lytF* Δ*cwlO* Δ*lytE*); **e)** LU163 ($P_{cwlOleader}$*yqgT-HA* Δ*yqgT* Δ*cwlO*). The wild-type strain (WT168) is shown as a control. Strains were grown in HPDM medium and a sample harvested at mid-exponential phase ($OD_{600}$ = 0.5). Cells were then washed and resuspended in 10 mM Tris-HCl buffer (pH 8). The decrease in optical density was measured every 10 minutes over a time period of 6 hours. The experiment was performed in three technical repeats and similar profiles were obtained. One representative example is shown.

Taken together this results show that disruption of the endogenous *cwlO* gene leads to increased autolysis and that *lytE* mutants feature an extremely decelerated autolysis. Overexpression of a D,L-endopeptidase from the *cwlO* promoter (such as LytE, CwlS, LytF or YqgT) does not seem to affect endogenous autolysis, independent of the effect that the disruption of the endogenous *cwlO* gene has.

## 4.4 Conclusion

That there are seven D,L-endopeptidases encoded in the *B. subtilis* genome, indicates the importance of this type of enzyme activity for the cell. This relevance of D,L-endopeptidase activity is further emphasized by the discovery of Bisicchia *et al.* (2007) that *lytE* and *cwlO* are synthetically lethal, i.e. the cell requires the enzyme activity encoded by one or other of these genes for viability. Furthermore, it was shown that LytE and CwlO are able to complement for each other despite their having different N-terminal domains, and that this occurs when they are expressed either from their natural promoters (Bisicchia *et al.*, 2007) or when LytE is expressed from the *cwlO* promoter (previous chapter of this thesis). Therefore, this part of the project set out to investigate if other cellular D,L-endopeptidases such as LytF (5 N-terminal LysM motifs), CwlS (4 N-terminal LysM motifs) and YqgT (N-terminal domain of unknown function) can also fulfil the D,L-endopeptidase requirement for viability of *B. subtilis*. Of course the synthetic lethality of *cwlO* and *lytE* implies that LytF, CwlS and YqgT cannot complement the D,L-endopeptidase requirement under normal growth conditions. It could be that some inherent property of the LytF, CwlS and YqgT enzymes precludes fulfilment of the D,L-endopeptidase activity requirement for viability. For example cell wall binding capability or affinity or the localization of the enzyme within the cell may such as to prevent complementation of CwlO and LytE activity. It could also be the case that LytF, CwlS and YqgT cannot complement, because the profile or level of their expression is inappropriate. Therefore, to test this latter possibility, strains were constructed in which expression of each of the *lytF, cwlS* and *yqgT* genes were placed under the control of the natural *cwlO* promoter. The level of expression at this locus was varied by either inclusion of the *cwlO* leader sequence (that destabilizes the mRNA and reduces protein levels) or removal of the *cwlO* leader sequence (that stabilizes the mRNA and leads to increased cellular levels of the protein). The strategy adopted to test for the ability of LytF, CwlS and YqgT to fulfil the D,L-endopeptidases for viability was to attempt to delete the endogenous *lytE* gene from the strain in which the enzyme is expressed at the *cwlO* locus (a process that deletes the endogenous *cwlO* gene).

The results showed that of the three enzymes tested (LytF, CwlS and YqgT), only LytF is capable of fulfilling the D,L-endopeptidase requirement for viability in *B. subtilis,* when both *lytE* and *cwlO* are deleted. However, although viable, these cells were not robust, showing significant defects in morphology (Figure 4.7, page 99), growth profile (Figure 4.4, page 91) and level of cell-associated protein (Figure 4.6, page 96). For example, despite a very high level of cellular *lytF* mRNA transcript (Figure 4.5, page 94), only a very low level of full-length LytF protein was cell-associated, but a very high level of processed LytF cell-associated protein was observed (Figure 4.6, page 96). LytF-HA protein was observed to be distributed along the length of the cylinder and at the septum (Figure 4.8, page 100).

It is interesting however that some level of complementation of *cwlO / lytE* function was achieved by the LytF protein that has the same D,L-endopeptidase enzymatic domain and 5 LysM motifs with which to bind to the cell wall peptidoglycan. Perhaps it is somewhat surprising considering the similarities between the LytE protein (3 LysM motifs) and LytF (5 LysM motifs) that the former can fulfil the D,L-endopeptidase with comfort, while the latter achieves some level of viability, but the cells are aberrant in several respects and are definitely not robust.

A somewhat surprising result of this study is that while LytF is able to complement for CwlO (albeit with a low level of effectiveness and yielding non-robust cells with aberrations in cell morphology, growth and autolysis), CwlS is unable to fulfil the D,L-endopeptidase requirement. This is especially curious since there is a high level of sequence and structural similarity between LytF, CwlS and LytE. The C-terminal D,L-endopeptidase domains of LytF, CwlS and LytE exhibit strong sequence similarity and all three enzymes possess N-terminally-located LysM motifs (but varying numbers of LysM motifs, with 5, 4 and 3 respectively). This poses the question as to what causes the inability of CwlS to fulfil the D,L-endopeptidase requirement. We have shown that there is a high level of CwlS associated with the cell, when expressed from the *cwlO* promoter (Figure 4.3, page 89), although significant processing of the enzyme was observed. However, it may be significant that the CwlS protein produced is primarily located at the cell septa with only a low presence along the cell cylinder (Figure 4.8, page 100). This is a significant finding since it has been established that the D,L-endopeptidase requirement is associated with facilitating cell cylinder elongation (Bisicchia *et al.*, 2007). Thus, there may not be a sufficient level of CwlS located at the cylinder to facilitate cell elongation. This is supported by the observations that cells expressing *cwlS* using the *cwlO* promoter appear to be shorter than wild-type cells and to have unusual structures at the poles of the cell (Figure 4.7, page 99). An interesting feature of the CwlS story is how LytE and LytF proteins with 3 and 5 LysM

motifs respectively, are both located along the cell cylinder and at the septum and poles, whereas CwlS with 4 LysM motifs appears to have a different subcellular localization, being predominantly located at the septum and poles. This suggests that features in addition to the LysM motif number / type are required for correct cellular localization of autolysins.

The YqgT enzyme also could not fulfil the D,L-endopeptidase requirement for viability of *B. subtilis*. Very little is known about the YqgT D,L-endopeptidase. Of particular note is that is does not have LysM motifs (as do LytE, LytF and CwlS) with which to bind to the cell wall, but instead a domain of unknown function. However, we show that it does bind at a low level to cell walls in a smeared pattern, rather than the punctate pattern observed with the other enzymes. However, it does not appear to bind to cell septa and poles (Figure 4.8. page 100). Deletion of YqgT appears to yield mild phenotypic effects that are exacerbated when combined with a null *lytE* mutation (Figure 4.2, page 86). It is interesting (i) that YqgT was expressed and remains cell-associated at a very high level and (ii) that overexpression does not appear to lead to processing of the enzyme (Figure 4.9, page 103). This high level expression also leads to unusual structures at the poles of cells (Figure 4.7, page 99). However, it is unclear from the analysis presented why this enzyme does not complement the D,L-endopeptidase requirement.

Interesting issues that emerge from this and other studies is the nature of autolysins binding to the cell envelope: how they bind; to what cellular component they bind and with what affinity; how they are processed and how they are released from the cell. The LytE protein, containing 3 LysM motifs, can fulfil the D,L-endopeptidase requirement very effectively when expressed from its natural promoter, through the SigI promoter in response to cell envelope stress (Salzberg *et al.*, 2013) or when expressed from the *cwlO* promoter (this study). However, the LytF protein, containing 5 LysM motifs, can only barely fulfil this requirement giving cells with severe defects while the CwlS enzyme, with 4 LysM motifs cannot fulfil the D,L-endopeptidase requirement for viability. This interesting observations suggest that an investigation of LysM motif function in enzymes that fulfil the D,L-endopeptidase requirement would give significant insight into the cellular function of autolysins in cell wall metabolism.

# 5 Chapter 5: Exploring the functionality of the LysM motifs of LytE *in vivo* and *in vitro*

## 5.1 Objective

The fact that either CwlO or LytE or to a lesser extent LytF, can fulfil the D,L-endopeptidase requirement (this study) is interesting, given that their N-terminal domains are different and that a different cellular function has been ascribed to them. The N-terminal domain of CwlO binds to membrane-bound FtsEX, an event that is required for enzymatic activity, whereas LytE and LytF have three and five LysM motifs respectively that mediate binding to peptidoglycan.

LysM motifs are widely distributed, conserved protein motifs that mediate binding to peptidoglycan, which is why they are commonly found in autolytic enzymes. There are 18 LysM motif proteins in *B. subtilis,* with a total of 30 individual LysM motifs. There is variance among proteins both in the number of LysM motifs and in the amino acid sequences of the individual LysM motifs. Twelve LysM motifs are contained within the three D,L-endopeptidases CwlS, LytE and LytF that have four, three and five motifs respectively. Three more proteins, the amidase YocH, an aminidase YaaH and YdhD (a protein of unknown function) have two LysM motifs each, while the remaining 12 LysM motifs are present in a variety of proteins that have a single LysM motif each.

The LysM motif is generally conserved and the LysM motifs of CwlS, LytE and LytF have an especially strong sequence similarity, although the number of LysM motifs is different. It was shown earlier that LytE (3 LysM motifs) and LytF (5 LysM motifs), but not CwlS (4 LysM motifs) when expressed from the *cwlO* promoter can support growth of *B. subtilis*. There is little known about how individual LysM motifs bind to peptidoglycan and the significance of the fact that the number of LysM motif varies in proteins with similar enzyme activity.

In this section functions of individual LysM motifs will be examined within the context of LytE D,L-endopeptidase function to address the following questions:

(i) Is LytE still functional and capable of CwlO complementation when the number of LysM motifs is varied, and when its three LysM motifs are permuted?

(ii) Can LysM motifs from other enzymes complement those of LytE?

(iii) How do variations in LysM motifs affect LytE enzyme activity, its binding affinity for peptidoglycan and the localization of these modified enzymes within the cell wall?

## 5.2 LytE structure and nomenclature

To address the questions outlined above, a system was designed to alter both the number and the order of LysM motifs of the LytE D,L-endopeptidase. For example, so that LytE proteins could be generated with one, two, three, four or more LysM motifs and that the order of motifs could be varied. The following paragraph describes the detailed structure of the wild-type LytE protein which has three LysM motifs and the nomenclature we have devised to describe variant LytE proteins. A schematic diagram of this structure is given in Figure 5.1 (a) (page 114). The nucleotide and amino acid sequence of LytE, colour encoded by the respective structural elements, is specified in the appendix 8.2.1 (page 238).

The wild-type LytE protein (LytE[L123cat]$^{WT168}$) contains three N-terminal LysM motifs that are of the same length (44 aa) and similar in their amino acid sequence (compare Figure 1.6, page 32). The three LysM motifs have been numbered in this study, to differentiate between the $NH_2$-terminal motif (L1) and the following LysM motifs (L2 and L3). The first LysM motif (L1) follows directly after the signal peptide (signal peptide length: 26 aa). The three individual LysM motifs are separated by short serine rich sequences, which henceforth will be referred to as the 'linker sequences'. The two linker sequences vary slightly in their sequence and length. The first linker (the sequence between L1 and L2) has the amino acid sequence GKSTSSKSSSSSSSSS and the second linker (between L2 and L3) has the amino acid sequence GSTSSSSSSSSKVSSSSTS. To simplify the cloning process, it was decided to link all LysM motifs with the first linker sequence (linker sequence length: 16 aa, sequence above). The C-terminal catalytic domain (catalytic domain length: 103 aa) follows the third LysM motif (L3), and similarly to the linker regions between the individual LysM motifs, L3 is separated from the C-terminal catalytic domain by a third serine rich sequence which is longer than the other two linker sequences. In this study this sequence has been dubbed the 'spacer' (this spacer is 38 aa in length).

Following this nomenclature the wild-type LytE can be described as

signal – L1 – linker – L2 – linker – L3 – spacer – catalytic domain

However, since all variant LytE constructed include the signal sequence, the chosen linker, the spacer and catalytic domain, in the remaining part of this study their

descriptions have been omitted. According to this nomenclature system, the wild-type LytE will be described to as:

$$LytE[L123cat]^{WT168}$$

In the same way, LytE[L13cat] is a variant LytE with two LysM motifs (L1 and L3) and is missing the second LysM motif while LytE[L111cat] is a variant LytE with three L1-type LysM motifs. Thus, it is to be understood that both constructs (LytE[L13cat] and LytE[L111cat]) have the signal, linker and spacer sequences. Furthermore, a variant LytE was constructed that does not contain LysM motifs. In this variant the signal is directly followed by the spacer and the catalytic domain. This variant LytE is described as LytE[cat]. Lastly, it is necessary to differentiate between the wild-type LytE protein and the reconstituted LytE protein. They are basically the same LytE protein containing three LysM motifs (LytE[L123cat]), but they differ in their linker sequences, as well as in a few silent point mutations introduced to aid the cloning process (which do not affect the amino acid sequence). The wild-type LytE will be referred to as $LytE[L123cat]^{WT168}$ (the two linker sequences vary slightly) and the reconstituted LytE will be referred to as $LytE[L123cat]^{R}$ (in which the two linker sequences are identical). Also the description LytE[L1cat] describes a variant LytE protein with one (L1) LysM motif, while the same description written in italics *lytE [L1cat]* refers to the gene that encodes the variant *lytE*.

## 5.3 Experimental Approach

### 5.3.1 Expression of variant LytE from the *cwlO* promoter

A system was devised to allow expression of variant LytE with various numbers and permutations of the LysM motifs, using the *cwlO* promoter. A graphic representation of the cloning process is shown in Figure 5.1. In a first step, PCR fragments encoding variant *lytE* with one LysM motifs were constructed (see Figure 5.1 c). To generate PCR fragments encoding variant *lytE* with one LysM motif, three separate PCR reactions were necessary. First, each of the three individual LysM motifs of *lytE* were amplified by PCR, with a primer pair, flanked by part of the signal and spacer sequence. Secondly the *lytE* catalytic domain was also amplified with a primer pair, in which the forward primer was flanked by the spacer sequence.

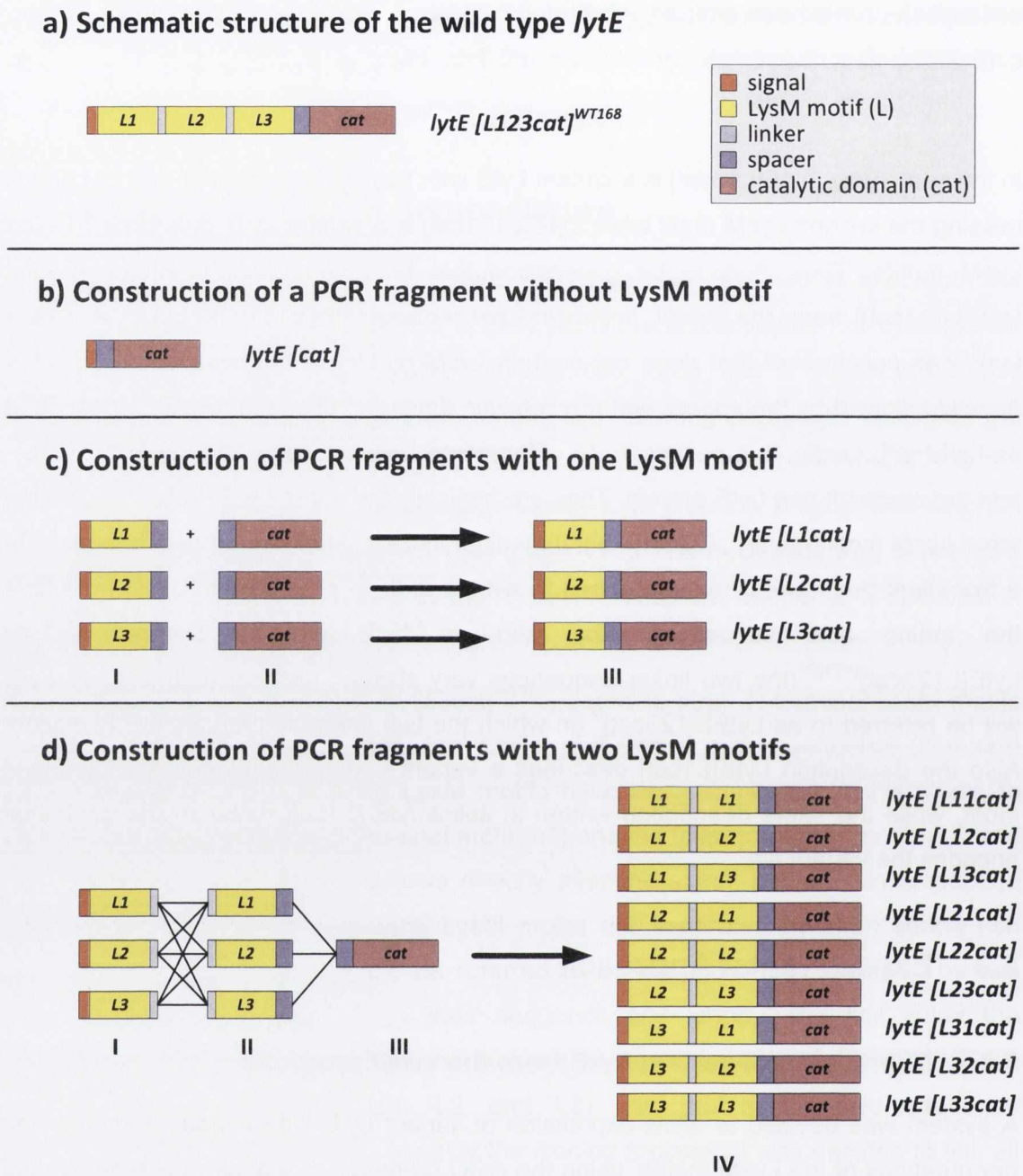



**Figure 5.1: Schematic structure of the WT168 *lytE* and the generation of PCR fragments encoding variant *lytE* without, with one or with two LysM motifs. a)** Schematic structure of the WT168 *lytE*, which has a signal sequence, three LysM motifs (numbered L1, L2 and L3), separated by linker sequences and a catalytic domain (cat) which is separated from L3 by a spacer sequence. **b)** A PCR fragment encoding a variant *lytE* without a LysM motif was generated by amplifying the catalytic domain with primers flanked by the signal sequence. **c)** PCR fragments encoding variant *lytE* with one LysM motif (III) are constructed by amplifying individual LysM motifs (I), with primers flanked by the signal and the spacer sequence, and joining them to the catalytic domain (II) by long-flanking homology PCR utilizing the spacer sequence for strand overlap. **d)** PCR fragments encoding variant *lytE* with two LysM motifs (IV), are constructed by amplifying individual LysM motifs with two set of primers (I and II) (one primer set is flanked by the signal and the linker sequence and the other by the linker and spacer sequence) and joining them alternating to the catalytic domain (III) by long-flanking homology PCR utilizing the linker and spacer sequence for strand overlap to generate all nine possible permutations.

Finally, the three PCR fragments were joined by strand overlap extension PCR to the fragment that encodes the catalytic domain, utilizing the spacer sequence for strand overlap and thereby re-generating the spacer in the complete single DNA fragment. In this way the following three DNA fragments were generated that encode the named *lytE* variants:

*lytE[L1cat]* *lytE[L2cat]* *lytE[L3cat]*

To create a PCR fragment encoding a variant *lytE*, without a LysM motif a single-step PCR reaction was sufficient (see Figure 5.1 b). The catalytic domain was amplified in one reaction with a primer pair that flanked the resulting fragment with part of the signal and restriction sites. In this way the following DNA fragment was generated:

*lytE[cat]*

A similar method was applied to construct PCR fragments encoding variant *lytE* with two LysM motifs (see Figure 5.1 d). In this case, DNA encoding each individual LysM motif of *lytE* was amplified by generating DNA fragments with two different primer pairs. The first primer pair flanked the resulting fragment with part of the sequence of the signal peptide and the linker sequence, while the second primer pair flanked the resulting fragment with the linker sequence and part of the spacer sequence. In this way, by alternatively joining these two fragments and joining the so generated fragment of two LysM motifs to the catalytic domain as described earlier, the nine possible variant *lytE* with two LysM motifs were generated:

| | | |
|---|---|---|
| *lytE[L11cat]* | *lytE[L21cat]* | *lytE[L31cat]* |
| *lytE[L12cat]* | *lytE[L22cat]* | *lytE[L32cat]* |
| *lytE[L13cat]* | *lytE[L23cat]* | *lytE[L33cat]* |

For subsequent cloning purposes, all fragments (encoding *lytE* variants without, one and two LysM motifs) were flanked by restrictions sites and a silent *SpeI* restriction site was introduced into each linker sequence. All DNA fragments generated could then be cloned into a vector, which was designed to enable integration of the variant *lytE* open reading frames at the *cwlO* locus by a double-cross over event. All resulting vectors were sequenced to verify the correct joining of the DNA fragments as well as the number and order of the LysM motifs. Thus, all variant *lytE* genes could be placed in single copy at the *cwlO* locus and each could be expressed from the normal *cwlO* promoter. The cloning of the fragment into the vector is represented in Figure 5.2 (b / c) (page 117). When inserted into the plasmid, each variant *lytE* gene was placed downstream of the $P_{cwlO}$ and was seamlessly joined to a signal sequence already present on the plasmid.

The next objective was to generate variant *lytE* with three or more LysM motifs. This was achieved using the series of vectors that were constructed to generate variant *lytE* with two LysM motifs. To achieve this, individual LysM motifs were amplified with primer pairs which generated LysM motifs flanked to both sides with the linker sequence. As mentioned earlier, a single silent *SpeI* restriction site was introduced into all linker sequences. By utilizing this restriction site, it was therefore possible to clone a single LysM motif fragments into a vector with two LysM motifs connected by the linker sequence (see Figure 5.2 d). Since this cloning step allowed insertion of multiple DNA fragments it was necessary to verify the number of integrated LysM motifs by PCR and agarose gel electrophoresis, taking advantage of the fact that all LysM motifs are of the same length (44 aa). An example of a typical agarose gel is shown in Figure 5.3 (page 118). Vectors with the desired number of LysM motifs were then sequenced to establish the number, type and particularly orientation of the inserted DNA fragments and to verify that the DNA sequences were correct. In this way the following fifteen DNA fragments were generated that encode the named *lytE* variants:

| | | | |
|---|---|---|---|
| *lytE[L113cat]* | *lytE[L212cat]* | *lytE[L222cat]* | *lytE[L311cat]* |
| *lytE[L123cat]* | *lytE[L213cat]* | *lytE[L233cat]* | *lytE[L331cat]* |
| *lytE[L1221cat]* | *lytE[L1333cat]* | *lytE[L2333cat]* | *lytE[L3331cat]* |
| *lytE[L11111cat]* | *lytE[L12221cat]* | *lytE[L33331cat]* | |

After generation of PCR fragments using the desired plasmids as template, the PCR fragments were inserted by a double-cross-over recombination event into the *cwlO* locus of wild-type strain WT168 and into a strain (LS089) in which the resident *lytE* gene was replaced by the chloramphenicol resistance gene (*cm*$^r$) (Figure 5.4, page 119). The resulting series of strains express variant LytE proteins from the *cwlO* promoter (with the *cwlO* leader sequence) and have either the endogenous natural *lytE* gene (WT168 as parental strain) or the endogenous *lytE* gene is deleted (all integrations into strain LS089 Δ*lytE*). In the series of strains that had the wild-type strain WT168 as parental, (i.e. that contained the endogenous *lytE* gene) expression of the variant *lytE* at the *cwlO* locus could be investigated. The ability of each variant LytE (when expressed from the *cwlO* promoter) to fulfil the D,L-endopeptidase requirement could be determined in the series of strains that had LS089 Δ*lytE* as the parental strain.

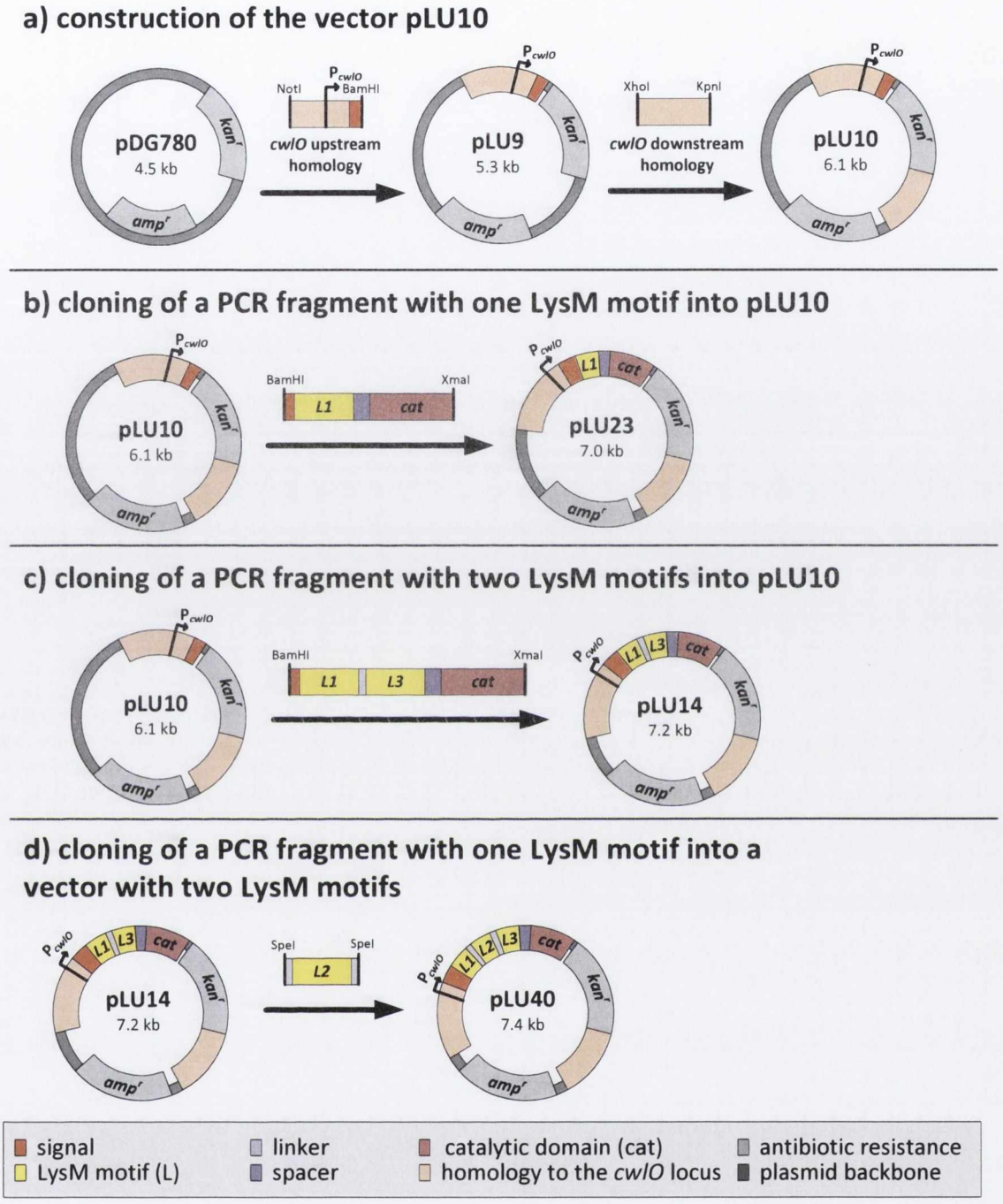



**Figure 5.2: Schematic representation of the construction of vectors designed to enable integration of variant *lytE* open reading frames into the chromosome at the *cwlO* locus.** **a)** A parental vector pLU10 was constructed by cloning ≈ 800 bp of up- and downstream homology to the *cwlO* locus into the vector pDG780. The upstream homology encompasses the *cwlO* promoter sequence, the *cwlO* leader sequence and part of the *lytE* signal sequence. **b)** A LFH fragment encompassing part of the *lytE* signal sequence, one LysM motif (L) and the *lytE* catalytic domain (cat), is cloned in frame, into the vector pLU10, thereby regenerating the *lytE* signal sequence. **c)** A LFH fragment encompassing part of the *lytE* signal sequence, two LysM motif and the *lytE* catalytic domain, is cloned in frame into the vector pLU10, thereby regenerating the *lytE* signal sequence. **d)** Plasmids generated in (c) are used as parental vectors, in which a single LysM motif, flanked by linker sequences, is cloned in frame into a silent restriction site within the linker sequence, thereby regenerating the linker sequences. Due to the fact that multiple integration events can take place the resulting vector can have three or more LysM motifs.

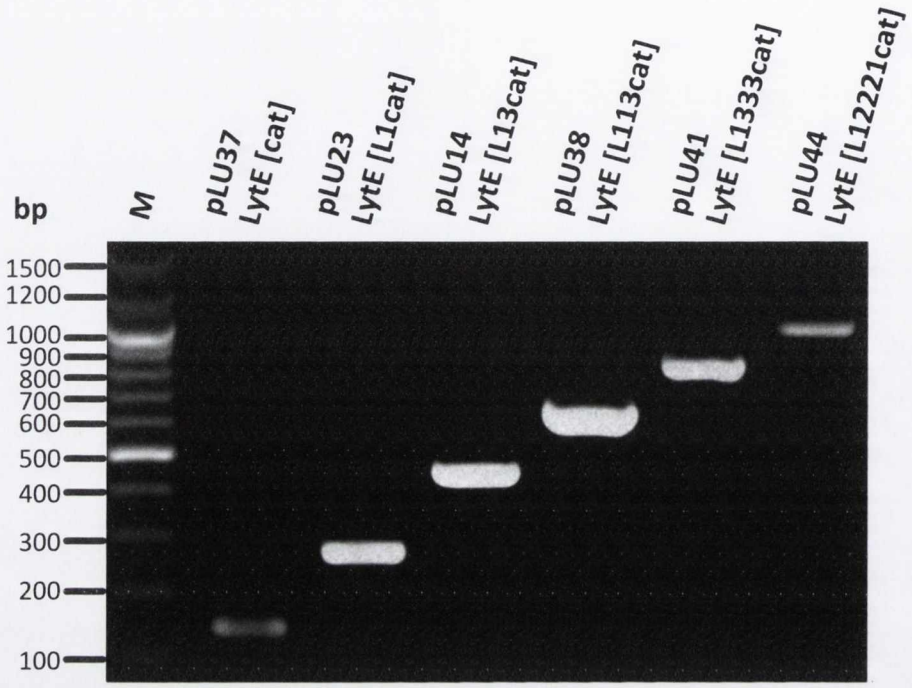



**Figure 5.3: Determination of the total number of LysM motifs in a vector by agarose gel electrophoresis.** The total number of LysM motifs within a vector could be determined by agarose gel electrophoresis, taking advantage of the fact that all LysM motifs are of the same length (132 bp). Plasmids were screened by PCR with the primer pair LytE_DF / Spac_Check_Rev. The following plasmids were screened and their PCR product is displayed on the gel: pLU37 (without LysM motif), pLU23 (1 LysM motif), pLU14 (2 LysM motifs), pLU38 (3 LysM motifs), pLU41 (4 LysM motifs) and pLU44 (5 LysM motifs). A size maker (M, bp) is shown).

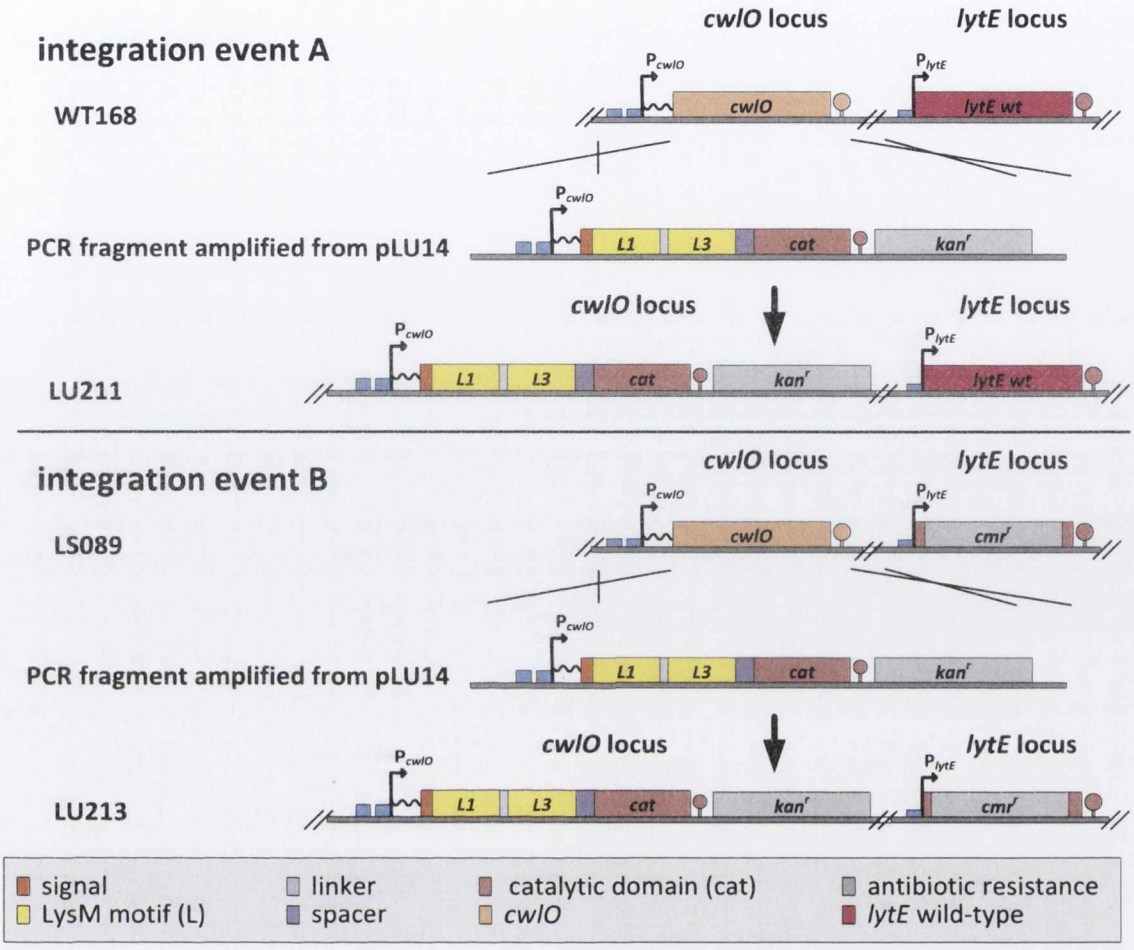



**Figure 5.4: Schematic representation of the integration of the variant *lytE* open reading frame into the chromosome at the *cwlO* locus. Integration event A:** A PCR product, amplified from the vector pLU14 (encompassing up- and downstream homology to the *cwlO* locus, the variant *lytE* open reading frame with the combination *lytE[L13cat]* and a kanamycin resistance gene), is transformed by double cross-over into WT168. This results in strain LU211, in which the *cwlO* gene is replaced by the open reading frame of the variant *lytE*, which is now under the control of the *cwlO* promoter. Strain LU211 does not express a CwlO protein, but two LytE proteins: the WT168 LytE protein expressed from the *lytE* promoter and the variant LytE[L13cat] protein expressed from the *cwlO* promoter. **Integration event B:** The same PCR fragment as in (A) is transformed by double cross-over into the strain LS089 (Δ*lytE*) to generate strain LU213. The *cwlO* gene in strain LU213 is replaced by the open reading frame *lytE[L13cat]* which is under the control of the *cwlO* promoter. Strain LU211 does not express a CwlO protein and only one LytE proteins: the variant LytE[L13cat] expressed from the *cwlO* promoter. Open reading frames for *lytE* and *cwlO* are not contiguous.

A viable strain can only be obtained in the LS089 strain (Δ*lytE* background) if the integrated variant LytE expressed using the *cwlO* promoter can fulfil the D,L-endopeptidase requirement. Thus, if no viable strain is obtained upon transformation, then the variant LytE being integrated at the *cwlO* locus is incapable of fulfilling that requirement. Thus, the nature and requirement of D,L-endopeptidase activity for viability of *B. subtilis* can be addressed using these series of strains. A detailed description of how these strains were constructed is given in Chapter 2 (section 2.1.4.8).

### 5.3.2 Expression of heterologous variant LytE from the *cwlO* promoter

To further investigate LysM motif function and the D,L-endopeptidase requirement for viability in *B. subtilis*, it was decided to construct heterologous variant LytE in which the LysM motifs is from heterologous *B. subtilis* proteins and from a *S. aureus* autolysin. Given that LysM motif sequence are well conserved this is a similar approach to mutating one of the LytE LysM motifs in order to obtain variant *lytE* with slightly altered LysM motif sequence (compare LysM motif amino acid sequences in Figure 1.6, page 32).

The following four LysM motifs were selected in order to construct variant LytE that contained just one heterologous LysM motif (also see Figure 5.5 a):

- the first ($NH_2$-terminal) LysM motif of CwlS ($L1^{CwlS}$),
  – CwlS is a D,L-endopeptidase of *B. subtilis 168* with a total of 4 LysM motifs
- the first ($NH_2$-terminal) LysM motif of LytF ($L1^{LytF}$),
  – LytF is a D,L-endopeptidase of *B. subtilis 168* with a total of 5 LysM motifs
- the single LysM motif of XlyB ($L1^{XlyB}$),
  – XlyB is an autolysin with one LysM motif that is encoded in the PBSX prophage protein of *B. subtilis 168*
- the first ($NH_2$-terminal) LysM motifs of Aaa ($L1^{Aaa}$),
  – Aaa is an amidase encoded in *Staphylococcus aureus* strain Newman with a total of 3 LysM motifs

Each LysM motifs was amplified as described earlier using chromosomal DNA of the respective strain as template and using forward and reverse primers that contained sequence encoding parts of the *lytE* signal sequence and spacer sequence respectively at their 5'-ends.

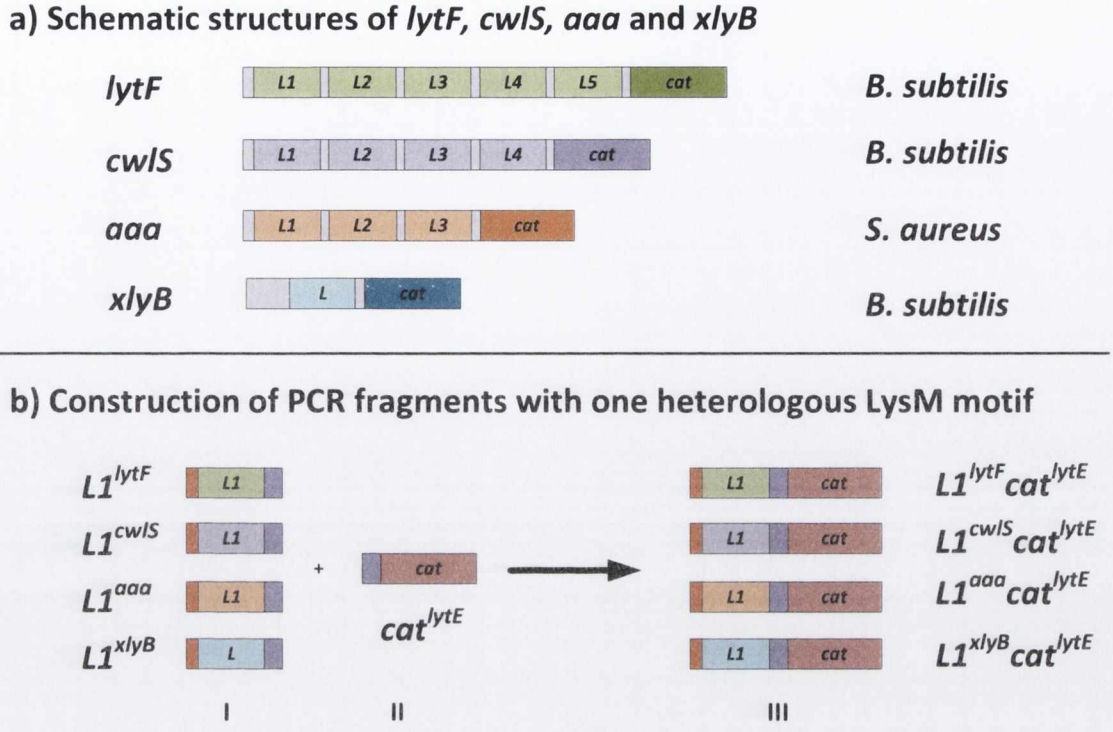



**Figure 5.5 [Figure continues on next page]: Schematic representation of the construction of heterologous variant LytE. a)** Schematic structures are shown for *lytF* (*B. subtilis 168*), *cwlS* (*B. subtilis 168*), *aaa* (*S. aureus Newman*) and *xlyB* (*B. subtilis 168*), with five, four, three and one LysM motifs respectively. **b)** Two PCR fragments (fragment I encompasses the *lytE* signal sequence, a heterologous LysM motif and part of the *lytE* spacer sequence and fragment II encompasses part of the *lytE* spacer sequence and the *lytE* catalytic domain) are combined by strand-overlap extension PCR to create a fragment (III). The resulting PCR fragment III encompasses one heterologous LysM motif and the catalytic domain of *lytE*. Four different fragments encoding heterologous variant *lytE* were generated this way: *lytE*[$L1^{lytF}cat^{lytE}$], *lytE*[$L1^{cwlS}cat^{lytE}$], *lytE*[$L1^{aaa}cat^{lytE}$] and *lytE*[$L1^{xlyB}cat^{lytE}$].

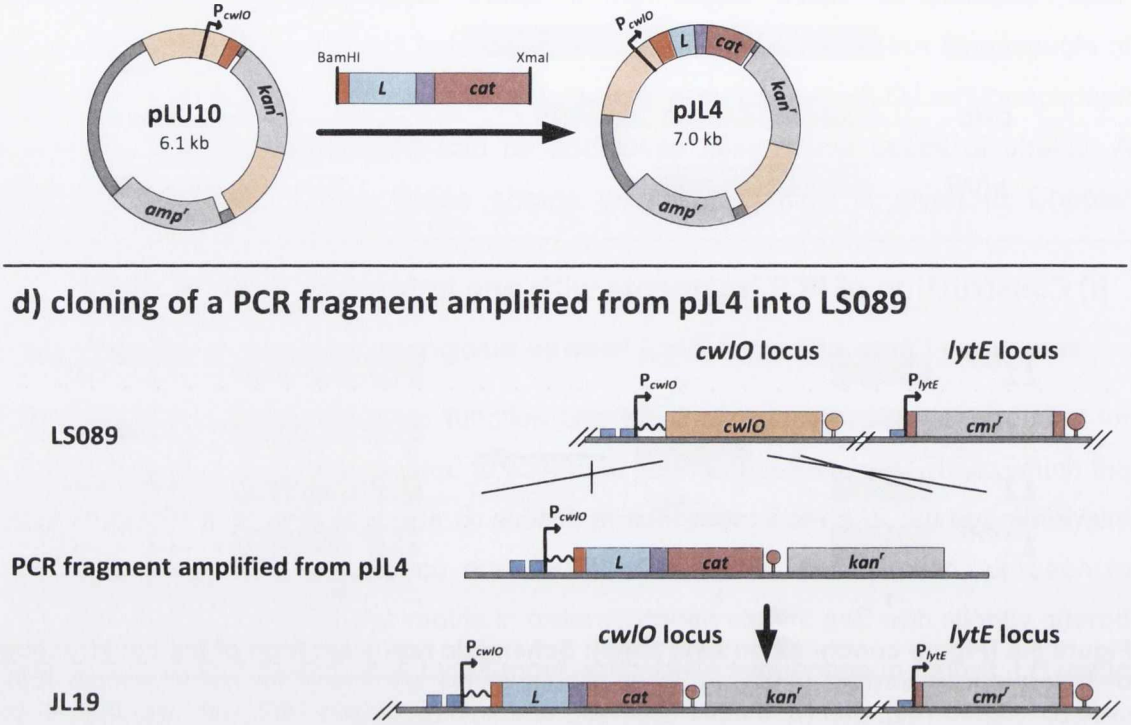



**Figure 5.5 (continued). Schematic representation of the construction of heterologous variant LytE. c)** A PCR fragment generated in (b) is cloned in frame into the vector pLU10, thereby regenerating the *lytE* signal sequence (resulting in pJL4). d) A PCR product amplified from the vector pJL4 encompassing up- and downstream homology to the *cwlO* locus, the variant *lytE[L1*$^{XlyB}$*cat*$^{lytE}$*]* open reading frame and a kanamycin resistance gene, is transformed by double cross-over into the strain LS089 (Δ*lytE*). In the resulting strain JL19, the *cwlO* gene is replaced by the open reading frame *lytE[L1*$^{XlyB}$*cat*$^{lytE}$*]*, which is under the control of the *cwlO* promoter. Strain JL19 does not express a CwlO or a LytE[L123cat]$^{WT168}$ protein, but one heterologous variant LytE[L1$^{XlyB}$cat$^{LytE}$] protein expressed from the *cwlO* promoter. Open reading frames for *lytE* and *cwlO* are not contiguous.

These PCR fragments were then joined by strand overlap extension PCR to a DNA fragment of the catalytic domain (that included part of the spacer sequence), utilizing the spacer sequence for strand overlap and thereby re-generating the spacer sequence. In this way the following four DNA fragments were generated (see Figure 5.5 b):

*lytE*[*L1*$^{cwlS}$*cat*$^{lytE}$] *lytE*[*L1*$^{lytF}$*cat*$^{lytE}$] *lytE*[*L1*$^{xlyB}$*cat*$^{lytE}$] *lytE*[*L1*$^{aaa}$*cat*$^{lytE}$]

The constructs were then cloned into plasmid pLU10 (see Figure 5.5 c), a vector designed to enable integration by a double cross-over event at the *cwlO* locus. To ensure a double cross-over event, integration plasmids were linearized by PCR and then integrated by a double-cross-over event at the *cwlO* locus of the chromosomes of strains WT168 (wild-type) and LS089 (Δ*lytE*), thereby replacing the endogenous *cwlO* open reading frame (see Figure 5.5 d). This allowed the generation of a set of strains in which the heterologous variant LytE proteins are expressed from the *cwlO* promoter (with the *cwlO* leader sequence) and that either possess the endogenous *lytE* gene (WT168 as parental strain), or do not possess the endogenous *lytE* gene (*lytE* mutant LS089 as parental strain). The ability of each heterologous variant LytE to fulfil the D,L-endopeptidase requirement was investigated by attempting to obtain these two strains. A viable strain would be obtained in the Δ*lytE* background only if the heterologous variant LytE protein expressed using the *cwlO* promoter ($P_{cwlO}$) can fulfil the D,L-endopeptidase requirement. A detailed description of how these strains were constructed is given in Chapter 2 section 2.1.4.12. These strains were generated by Jennifer Lennon under my supervision.

### 5.3.3 Expression of variant LytE from the *lytE* promoter

To address the effect of expression level on the ability of the variant LytE proteins to fulfil the D,L-endopeptidase requirement for viability in *B. subtilis,* a selection of variant LytE were chosen to be expressed from the *lytE* ($P_{lytE}$) promoter. The *lytE* promoter is weaker than the *cwlO* promoter (this study). Therefore, it was decided to test if variant LytE that were able to fulfil the D,L endopeptidase requirement for *B. subtilis* viability at an high expression level (when expressed using $P_{cwlO}$) were also able to do so when their expression level was lowered (when expressed using $P_{lytE}$). To express the variant LytE from the *lytE* promoter the variant *lytE* open reading frame sequence was amplified from the appropriate plasmids listed in Table 5.1.

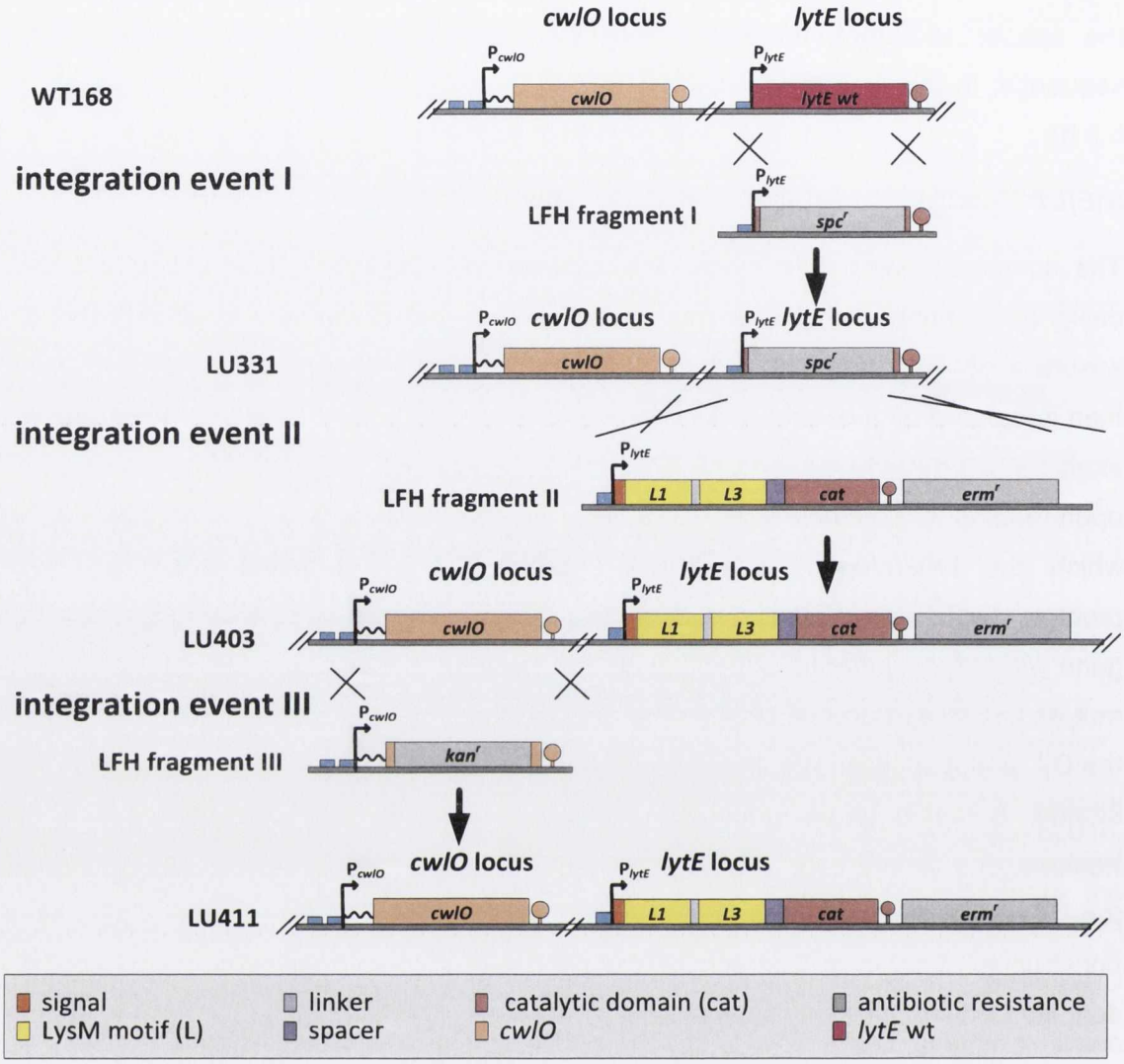



**Figure 5.6: Schematic representation of the integration of a variant *lytE* open reading frame at the *lytE* locus. integration event I)** The first transformation step disrupts the endogenous *lytE* gene in WT168, through transformation with a long flanking homology fragment (LFH fragment I), carrying a spectinomycin resistance gene, which integrates at the *lytE* locus by double cross-over, resulting in strain L331. **integration event II)** In a second transformation step, a LFH fragment II encompassing up- and downstream homology to the *lytE* locus, the variant *lytE* open reading frame *L13cat* and a erythromycin resistance cassette, is transformed into the strain LU331 and integrates by double cross-over at the *lytE* locus, thereby replacing the spectinomycin resistance gene and creating the strain LU403. In **integration event III)**, the endogenous *cwlO* gene in LU403 is disrupted, through transformation with a LFH fragment III, carrying a kanamycin resistance gene, which integrates at the *cwlO* locus by double cross-over, resulting in strain LU411. Open reading frames for *lytE* and *cwlO* are not contiguous.

**Table 5.1: List of PCR fragments encoding variant LytE proteins and the plasmids used as template for their amplification**

| Without / One LysM motifs | | Two LysM motifs | | Three LysM motifs | |
|---|---|---|---|---|---|
| **PCR fragment** | **Template** | **PCR fragment** | **Template** | **PCR fragment** | **Template** |
| *lytE[cat]* | pLU37 | *lytE[L11cat]* | pLU16 | *lytE[L213cat]* | pLU29 |
| *lytE[L1cat]* | pLU23 | *lytE[L21cat]* | pLU17 | *lytE[L331cat]* | pLU27 |
| *lytE[L2cat]* | pLU24 | *lytE[L31cat]* | pLU20 | *lytE[L123cat]*$^{R}$ | pLU40 |
| *lytE[L3cat]* | pLU25 | *lytE[L13cat]* | pLU14 | | |

The PCR fragments were then joined by strand overlap extension PCR to sequences that were homologous to upstream and downstream regions of the *lytE* locus. These DNA fragments were then integrated by a double cross-over homologous recombination at the *lytE* locus of strain LU331 (Δ*lytE*), thereby regenerating a variant *lytE* gene in single copy at the *lytE* locus to be expressed using the endogenous *lytE* promoter. A graphic presentation of these steps is shown in Figure 5.6. The ability of each variant LytE protein to fulfil the D,L-endopeptidase requirement for viability in *B. subtilis* was then tested by attempting to delete the *cwlO* gene in each of these strains. A viable strain is obtained only if the endogenous *cwlO* gene can be deleted. In this event, the variant LytE expressed from the *lytE* promoter can fulfil the D,L-endopeptidase requirement for viability. A detailed description of how these strains were constructed is given in Chapter 2 (section 2.1.4.9 and 2.1.4.10).

### 5.3.4 Purification of variant LytE proteins

In order to study the properties of the variant LytE proteins *in vitro*, they were expressed in an *E. coli* host and purified using a 6xHis-tag.

Proteins were expressed using the pET system (Novagen). In this system the open reading frame encoding the recombinant protein sequences is cloned into a vector with inducible expression, which is then transformed into an *E. coli* strain BL21(DE3) that is optimized for expression. In brief, the *E. coli* host strain encodes a *T7* RNA polymerase gene which is under the control of a *lac* promoter. Induction of the *lac* promoter with IPTG results in high level expression of the desired protein. The expressed protein was then purified using the 6xHis-tag (see Chapter 2, section 2.2.12).

Plasmids for protein expression were designed by amplifying the variant *lytE* open reading frame, excluding the *lytE* signal peptide, the start codon and the stop codon. The 5′-tail sequences of the primers were designed to generated *Nde*I / *Xho*I restriction sites up- and downstream of the fragment, in order to allow for cloning the fragment

into the vector pET21b+. The vector is conveniently equipped with a 6xHis-tag as well as a stop codon. Ligation of the fragment with the vector re-generated the start codon. A detailed description of how these plasmids were constructed is given in Chapter 2 (section 2.1.3.9).

In this way the plasmids listed in Table 5.2, carrying the open reading frames of different variant *lytE*, were constructed. The amino acid sequences of the variant LytE proteins including the 6xHis-tag are specified in the appendix 8.2.2 (page 239).

**Table 5.2: List of plasmids designed for overexpression of variant LytE proteins in *E. coli***

| LysM motifs | Plasmid | Variant LytE |
|---|---|---|
| No LysM motif | pLU52 | LytE[cat] |
| One LysM motif | pLU50 | LytE[L1cat] |
| | pLU51 | LytE[L3cat] |
| | pLU54 | LytE[L1$^{XlyB}$cat$^{LytE}$] |
| | pLU53 | LytE[L1$^{LytF}$ca$^{LytE}$] |
| Two LysM motifs | pLU49 | LytE[L13cat] |
| | pLU48 | LytE[L31cat] |
| Three LysM motifs | pLU47 | LytE[L123cat]$^{R}$ |
| | pETLytE | LytE[L123cat]$^{WT168}$ |

All variant LytE proteins were successfully expressed and purified. Successful protein expression and purity was monitored by polyacrylamide gel electrophoresis with subsequent Coomassie Blue staining. An example of the resulting gel, prepared to monitor the protein expression and purification of the variant LytE[L31cat], is shown in Figure 5.7. The purity of all expressed variant LytE proteins was determined in a similar manner and the resulting Coomassie Blue stained gel is shown in Figure 5.8. As can be seen from the gel, individual variant LytE proteins formed distinct bands, whose position corresponds to their molecular weight. Generally a high purity was achieved, but three variant LytE proteins (LytE[L123cat]$^{R}$, LytE[L1cat] and LytE[L1$^{LytF}$cat$^{LytE}$]) showed additional unspecific bands.

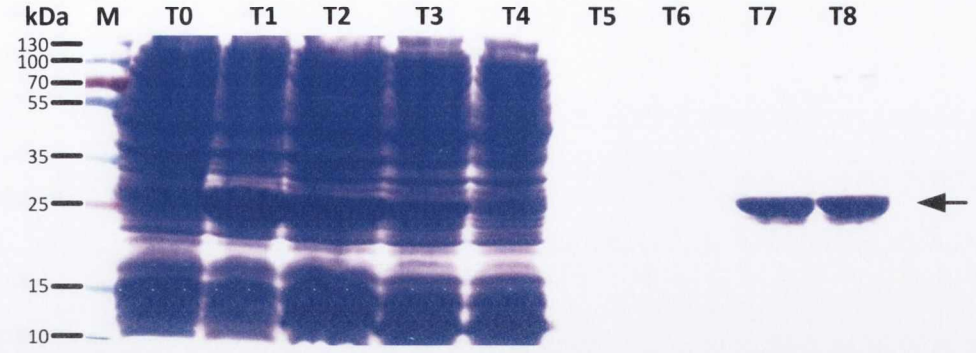



**Figure 5.7: Protein expression and purification of the variant LytE[L31cat] protein monitored by SDS-PAGE and Coomassie Blue staining.** Samples were harvested from the *E. coli* culture harbouring the plasmid pLU48 before induction of the protein expression with IPTG (T0), 3 hours after induction of protein expression (T1), after sonication of the cells (T2), from the supernatant after the centrifugation of the sonicated cells (T3), from the supernatant after binding of the variant LytE protein to His-beads (T4), after 4 washes (T5), after 8 washes (T6), after elution from the His-beads (T7) and after dialysis (T8). The arrow indicates the variant LytE[L31cat] protein which has a molecular weight of 27.4 kDa. A molecular weight marker is shown (M, kDa).




**Figure 5.8: Purity of overexpressed variant LytE protein as determined by SDS-PAGE and Coomassie Blue staining.** The following variant LytE proteins are shown: LytE[L123cat]$^{WT168}$ (34.09 kDa), LytE[L123cat]$^{R}$ (33.85 kDa), LytE[L31cat] (27.38 kDa), LytE[L13cat] (27.38 kDa), LytE[L1cat] (21.15 kDa), LytE[L3cat] (20.87 kDa), LytE[cat] (16.1 kDa), LytE[L1$^{LytF}$cat$^{LytE}$] (20.62 kDa) and LytE[L1$^{XlyB}$cat$^{LytE}$] (20.77 kDa). One µg total protein was deployed per lane. A molecular weight marker is shown (M, kDa). R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

## 5.4 Results

### 5.4.1 Ability of variant LytE to fulfil the D,L-endopeptidase requirement when expressed from the *cwlO* promoter

Disruption of the endogenous *lytE* genes served as a test for the ability of the variant LytE to fulfil the D,L-endopeptidase requirement, when expressed from the *cwlO* promoter. The following series of strains (Table 5.3) were obtained / not obtained.

**Table 5.3: [Table continues on next page]: List of strains containing a variant LytE expressed from the *cwlO* promoter**

| | Intact endogenous *lytE* gene | | | Disrupted endogenous *lytE* gene | | |
|---|---|---|---|---|---|---|
| **LysM** | **Strain** | **Genotype** | ***** | **Strain** | **Genotype** | ***** |
| No LysM | LU314 | $P_{cwlOleader}$*lytE [cat] ΔcwlO* | V | - | $P_{cwlOleader}$*lytE [cat] ΔcwlO ΔlytE* | - |
| One LysM | LU261 | $P_{cwlOleader}$*lytE [L1cat] ΔcwlO* | V | LU262 | $P_{cwlOleader}$*lytE [L1cat] ΔcwlO ΔlytE* | V |
| | LU265 | $P_{cwlOleader}$*lytE [L2cat] ΔcwlO* | V | LU266 | $P_{cwlOleader}$*lytE [L2cat] ΔcwlO ΔlytE* | V |
| | LU269 | $P_{cwlOleader}$*lytE [L3cat] ΔcwlO* | V | LU270 | $P_{cwlOleader}$*lytE [L3cat] ΔcwlO ΔlytE* | V |
| Two LysM | LU239 | $P_{cwlOleader}$*lytE [L11cat] ΔcwlO* | V | LU205 | $P_{cwlOleader}$*lytE [L11cat] ΔcwlO ΔlytE* | V |
| | LU207 | $P_{cwlOleader}$*lytE [L12cat] ΔcwlO* | V | LU209 | $P_{cwlOleader}$*lytE [L12cat] ΔcwlO ΔlytE* | V |
| | LU212 | $P_{cwlOleader}$*lytE [L13cat] ΔcwlO* | V | LU213 | $P_{cwlOleader}$*lytE [L13cat] ΔcwlO ΔlytE* | V |
| | LU216 | $P_{cwlOleader}$*lytE [L21cat] ΔcwlO* | V | LU217 | $P_{cwlOleader}$*lytE [L21cat] ΔcwlO ΔlytE* | V |
| | LU220 | $P_{cwlOleader}$*lytE [L22cat] ΔcwlO* | V | LU222 | $P_{cwlOleader}$*lytE [L22cat] ΔcwlO ΔlytE* | V |
| | LU223 | $P_{cwlOleader}$*lytE [L23cat] ΔcwlO* | V | LU226 | $P_{cwlOleader}$*lytE [L23cat] ΔcwlO ΔlytE* | V |
| | LU227 | $P_{cwlOleader}$*lytE [L31cat] ΔcwlO* | V | LU229 | $P_{cwlOleader}$*lytE [L31cat] ΔcwlO ΔlytE* | V |
| | LU232 | $P_{cwlOleader}$*lytE [L32cat] ΔcwlO* | V | LU233 | $P_{cwlOleader}$*lytE [L32cat] ΔcwlO ΔlytE* | V |
| | LU235 | $P_{cwlOleader}$*lytE [L33cat] ΔcwlO* | V | LU238 | $P_{cwlOleader}$*lytE [L33cat] ΔcwlO ΔlytE* | V |
| Three LysM | LU315 | $P_{cwlOleader}$*lytE [L113cat] ΔcwlO* | V | LU316 | $P_{cwlOleader}$*lytE [L113cat] ΔcwlO ΔlytE* | V |
| | LU318 | $P_{cwlOleader}$*lytE [L123cat]$^{R}$ ΔcwlO* | V | LU320 | $P_{cwlOleader}$*lytE [L123cat]$^{R}$ ΔcwlO ΔlytE* | V |
| | LU288 | $P_{cwlOleader}$*lytE [L212cat] ΔcwlO* | V | LU289 | $P_{cwlOleader}$*lytE [L212cat] ΔcwlO ΔlytE* | V |
| | LU312 | $P_{cwlOleader}$*lytE [L213cat] ΔcwlO* | V | LU313 | $P_{cwlOleader}$*lytE [L213cat] ΔcwlO ΔlytE* | V |
| | LU286 | $P_{cwlOleader}$*lytE [L222cat] ΔcwlO* | V | LU287 | $P_{cwlOleader}$*lytE [L222cat] ΔcwlO ΔlytE* | V |
| | LU280 | $P_{cwlOleader}$*lytE [L233cat] ΔcwlO* | V | LU281 | $P_{cwlOleader}$*lytE [L233cat] ΔcwlO ΔlytE* | V |
| | LU272 | $P_{cwlOleader}$*lytE [L311cat] ΔcwlO* | V | LU273 | $P_{cwlOleader}$*lytE [L311cat] ΔcwlO ΔlytE* | V |
| | LU276 | $P_{cwlOleader}$*lytE [L331cat] ΔcwlO* | V | LU277 | $P_{cwlOleader}$*lytE [L331cat] ΔcwlO ΔlytE* | V |

**Table 5.3: [continued]: List of strains containing a variant LytE expressed from the *cwlO* promoter**

| | Intact endogenous *lytE* gene | | | Disrupted endogenous *lytE* gene | | |
|---|---|---|---|---|---|---|
| **LysM** | **Strain** | **Genotype** | * | **Strain** | **Genotype** | * |
| Four LysM | **LU326** | $P_{cwlOleader}$*lytE [L1221cat]* Δ*cwlO* | V | **LU327** | $P_{cwlOleader}$*lytE [L1221cat]* Δ*cwlO* Δ*lytE* | V |
| | **LU322** | $P_{cwlOleader}$*lytE [L1333cat]* Δ*cwlO* | V | **LU323** | $P_{cwlOleader}$*lytE [L1333cat]* Δ*cwlO* Δ*lytE* | V |
| | **LU282** | $P_{cwlOleader}$*lytE [L2333cat]* Δ*cwlO* | V | **LU283** | $P_{cwlOleader}$*lytE [L2333cat]* Δ*cwlO* Δ*lytE* | V |
| | **LU274** | $P_{cwlOleader}$*lytE [L3331cat]* Δ*cwlO* | V | **LU275** | $P_{cwlOleader}$*lytE [L3331cat]* Δ*cwlO* Δ*lytE* | V |
| Five LysM | **LU324** | $P_{cwlOleader}$*lytE [L11111cat]* Δ*cwlO* | V | **LU325** | $P_{cwlOleader}$*lytE [L11111cat]* Δ*cwlO* Δ*lytE* | V |
| | **LU328** | $P_{cwlOleader}$*lytE [L12221cat]* Δ*cwlO* | V | **LU329** | $P_{cwlOleader}$*lytE [L12221cat]* Δ*cwlO* Δ*lytE* | V |
| | **LU278** | $P_{cwlOleader}$*lytE [L33331cat]* Δ*cwlO* | V | **LU279** | $P_{cwlOleader}$*lytE [L33331cat]* Δ*cwlO* Δ*lytE* | V |

* viable (V) strain constructed or strain not obtained (-)

All of the strains listed in Table 5.3 could be constructed with one (expected) exception. It was not possible to obtain a viable strain expressing a variant LytE with no LysM motif in parental strain LS089 (Δ*lytE*) that does not have the endogenous *lytE* gene (i.e. disrupting the endogenous *lytE* gene in strain LU314). Therefore, all tested variant LytE that possess at least one LysM motif can fulfil the D,L-endopeptidase requirement, when expressed from the $P_{cwlO}$ promoter in a Δ*lytE* background. The open reading frames of variant LytE at the *cwlO* chromosomal locus were sequenced in a selection of the listed strains (LU314, LU262, LU266, LU270, LU205, LU209, LU213, LU217, LU222, LU226, LU229, LU233, LU238, LU320, LU313, LU287 and LU277) to verify the correct integration of the fragments and that no mutations had arisen during the integration process.

### 5.4.2 Ability of heterologous variant LytE to fulfil the D,L-endopeptidase requirement when expressed from the *cwlO* promoter

Disruption of the endogenous *lytE* genes served as a test for the ability of the heterologous variant LytE to fulfil the D,L-endopeptidase requirement when expressed from the *cwlO* promoter. The following series of strains (Table 5.4) were obtained / not obtained.

**Table 5.4: List of strains containing a heterologous variant LytE expressed from the *cwlO* promoter**

| Intact endogenous *lytE* gene | | | Disrupted endogenous *lytE* gene | | |
|---|---|---|---|---|---|
| **Strain** | **Genotype** | * | **Strain** | **Genotype** | * |
| **JL11** | $P_{cwlOleader}$*lytE* [*L1*$^{cwlS}$*cat*$^{lytE}$] Δ*cwlO* | V | | $P_{cwlOleader}$*lytE* [*L1*$^{cwlS}$*cat*$^{lytE}$] Δ*cwlO* Δ*lytE* | - |
| **JL13** | $P_{cwlOleader}$*lytE* [*L1*$^{lytF}$*cat*$^{lytE}$] Δ*cwlO* | V | | $P_{cwlOleader}$*lytE* [*L1*$^{lytF}$*cat*$^{lytE}$] Δ*cwlO* Δ*lytE* | - |
| **JL15** | $P_{cwlOleader}$*lytE* [*L1*$^{xlyB}$*cat*$^{lytE}$] Δ*cwlO* | V | **JL19** | $P_{cwlOleader}$*lytE* [*L1*$^{xlyB}$*cat*$^{lytE}$] Δ*cwlO* Δ*lytE* | V |
| **JL17** | $P_{cwlOleader}$*lytE* [*L1*$^{aaa}$*cat*$^{lytE}$] Δ*cwlO* | V | | $P_{cwlOleader}$*lytE* [*L1*$^{aaa}$*cat*$^{lytE}$] Δ*cwlO* Δ*lytE* | - |

* viable (V) strain constructed or strain not obtained (-)

Three of the listed strains could be not be obtained. It was not possible to obtain a viable strain in which a heterologous variant LytE protein with the N-terminal LysM motif of CwlS, LytF or Aaa is expressed using the *cwlO* promoter in strains that lack the endogenous *lytE* gene. However, it was possible to obtain a strain that expresses (using the $P_{cwlO}$) a heterologous variant LytE protein with the LysM motif of XlyB in a Δ*lytE* background. The *cwlO* locus of strains JL11, JL13, JL17 and JL19 was sequenced to verify the correct integration of the DNA fragments and that no mutations had occurred during the integration process. These strains were generated by Jennifer Lennon under my supervision.

### 5.4.3 Ability of variant LytE to fulfil the D,L-endopeptidase requirement when expressed from the *lytE* promoter

Disruption of the endogenous *cwlO* gene served as a test for the ability of the variant LytE to fulfil the D,L-endopeptidase requirement when expressed from the *lytE* promoter. The following series of strains (Table 5.5) were obtained / not obtained. All, but one, of the listed strains could be constructed. It was not possible to obtain a viable strain that expresses a variant LytE protein with no LysM motif from the *lytE* promoter and that does not possess the endogenous *cwlO* gene. All other tested variant LytE, possessing at least one LysM motif, can fulfil the D,L-endopeptidase requirement, when expressed from the *lytE* promoter in a Δ*cwlO* background.

The *lytE* locus of each strain was sequenced to verify that the correct integration event had occurred and that there were no mutations in the inserted genes.

**Table 5.5: List of strains containing a variant LytE protein expressed from the *lytE* promoter**

| | Intact *cwlO* gene | | | Disrupted *cwlO* gene | | |
|---|---|---|---|---|---|---|
| | **Strain** | **Genotype** | * | **Strain** | **Genotype** | * |
| No LysM | **LU398** | $P_{lytE}$*lytE [cat] ΔlytE* | V | | $P_{lytE}$*lytE [cat] ΔlytE ΔcwlO* | - |
| One LysM | **LU337** | $P_{lytE}$*lytE [L1cat] ΔlytE* | V | **LU350** | $P_{lytE}$*lytE [L1cat] ΔlytE ΔcwlO* | V |
| | **LU339** | $P_{lytE}$*lytE [L2cat] ΔlytE* | V | **LU352** | $P_{lytE}$*lytE [L2cat] ΔlytE ΔcwlO* | V |
| | **LU341** | $P_{lytE}$*lytE [L3cat] ΔlytE* | V | **LU354** | $P_{lytE}$*lytE [L3cat] ΔlytE ΔcwlO* | V |
| Two LysM | **LU399** | $P_{lytE}$*lytE [L11cat] ΔlytE* | V | **LU407** | $P_{lytE}$*lytE [L11cat] ΔlytE ΔcwlO* | V |
| | **LU403** | $P_{lytE}$*lytE [L13cat] ΔlytE* | V | **LU411** | $P_{lytE}$*lytE [L13cat] ΔlytE ΔcwlO* | V |
| | **LU401** | $P_{lytE}$*lytE [L21cat] ΔlytE* | V | **LU409** | $P_{lytE}$*lytE [L21cat] ΔlytE ΔcwlO* | V |
| | **LU372** | $P_{lytE}$*lytE [L31cat] ΔlytE* | V | **LU383** | $P_{lytE}$*lytE [L31cat] ΔlytE ΔcwlO* | V |
| Three LysM | **LU235** | $P_{lytE}$*lytE [L123cat]*$^{R}$ *ΔlytE* | V | **LU343** | $P_{lytE}$*lytE [L123cat]*$^{R}$ *ΔlytE ΔcwlO* | V |
| | **LU405** | $P_{lytE}$*lytE [L213cat] ΔlytE* | V | **LU413** | $P_{lytE}$*lytE [L213cat] ΔlytE ΔcwlO* | V |
| | **LU374** | $P_{lytE}$*lytE [L331cat] ΔlytE* | V | **LU385** | $P_{lytE}$*lytE [L331cat] ΔlytE ΔcwlO* | V |

* viable (V) strain constructed or strain not obtained (-)

### 5.4.4 Verification of the successful disruption of the *cwlO* gene in strains that express a variant LytE from the *cwlO* promoter

A series of strains were successfully constructed that express variant LytE proteins with variation in the number, type and combination of LysM motifs. It many cases it was possible to delete the endogenous *cwlO* gene, showing that these variant LytE D,L-endopeptidases are capable of fulfilling the D,L-endopeptidase requirement for viability in *B. subtilis.*

However, it is important to establish that this requirement is fulfilled by the variant protein alone, and that the CwlO protein is indeed not expressed in these strains, and that for example, merodiploid strains with a duplication of the *cwlO* gene have not been generated.

Therefore, to verify the correct deletion of the *cwlO* gene in strains in which the variant *lytE* open reading frame is cloned onto the *cwlO* locus, it was decided to perform a western blot analysis on cell lysates and detect CwlO with a specific polyclonal antibody as detailed in experimental procedures.

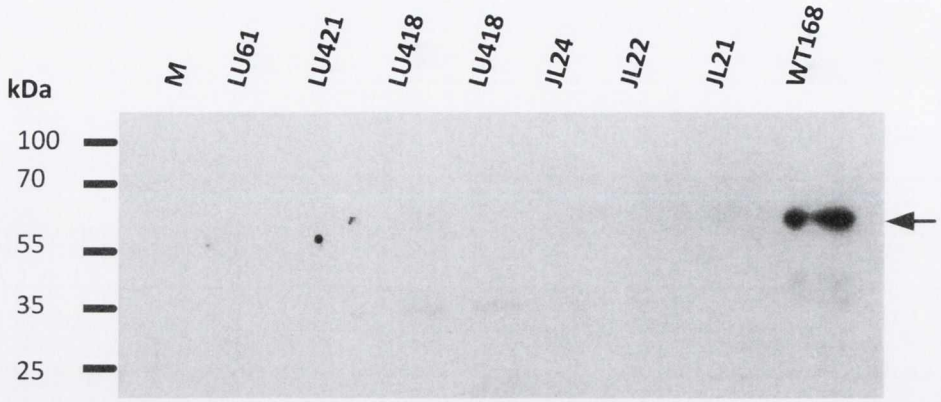



**Figure 5.9: Verification of the absence of the *cwlO* gene, in strains that express a variant LytE protein from the *cwlO* promoter, by determining CwlO protein levels by western analysis.** Cell-associated protein samples of the following strains are shown: LU61 (Δ*cwlO* $P_{cwlOleader}$*lytE-HA* Δ*lytE*), LU421 ($P_{cwlOleader}$*lytE-HA [L123cat]*$^{R}$ Δ*cwlO* Δ*lytE*), LU418 ($P_{cwlOleader}$*lytE-HA [L1cat]* Δ*cwlO* Δ*lytE*), JL21 ($P_{cwlOleader}$*lytE-HA [L1*$^{lytF}$*cat*$^{lytE}$*]* Δ*cwlO*), JL21 ($P_{cwlOleader}$*lytE-HA [L1*$^{xlyB}$*cat*$^{lytE}$*]* Δ*cwlO*), JL24 ($P_{cwlOleader}$*lytE-HA [L1*$^{xlyB}$*cat*$^{lytE}$*]* Δ*cwlO* Δ*lytE*) and WT168. Cells were grown in salt-free buffered LB and harvested at an $OD_{600}$ of 0.3. Total cell protein was prepared from the samples and the CwlO protein was detected by western analysis. Five µg total cell protein extract was deployed per lane. Molecular weight markers are shown (M, kDa). The film was exposed for five minutes. The arrow indicates the full length CwlO protein (CwlO, MW = 50 kDa). This western blot analysis was conducted by Jennifer Lennon under my supervision. R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

**Table 5.6: List of strains in which the successful disruption of the *cwlO* gene was analysed by western analysis (see Figure 5.9)**

| Strain | Genotype |
|---|---|
| **WT168** | |
| **LU61** | $P_{cwlOleader}$*lytE-HA* Δ*cwlO* Δ*lytE* |
| **LU421** | $P_{cwlOleader}$*lytE-HA [L123cat]*$^{R}$ Δ*cwlO* Δ*lytE* |
| **LU418** | $P_{cwlOleader}$*lytE-HA [L1cat]* Δ*cwlO* Δ*lytE* |
| **JL21** | $P_{cwlOleader}$*lytE-HA [L1*$^{lytF}$*cat*$^{lytE}$*]* Δ*cwlO* |
| **JL22** | $P_{cwlOleader}$*lytE-HA [L1*$^{xlyB}$*cat*$^{lytE}$*]* Δ*cwlO* |
| **JL24** | $P_{cwlOleader}$*lytE-HA [L1*$^{xlyB}$*cat*$^{lytE}$*]* Δ*cwlO* Δ*lytE* |

The strains, listed in Table 5.6, were chosen for this experiment, because they represent a selection of strains that express different variant LytE proteins from the *cwlO* promoter: a variant LytE protein with three LysM motifs (LU421), one LysM motif (LU418) and a heterologous LysM motifs (JL24) which are expressed in strains in which the endogenous *lytE* gene is disrupted (all previous strains) or intact (JL21, JL22). Strain LU61 was chosen as a negative control, since it was confirmed previously (this study) that this strain does not produce a CwlO protein. WT168 was chosen as a positive control and contains the endogenous *cwlO* gene.

The results are presented in Figure 5.9. A single band can be seen in WT168 that migrated at a MW of 50 kDa which is the correct size and migration pattern for the CwlO protein. It is also clear that this band was not observed in strains LU61, LU421, LU418, JL21, JL22 or JL24. It can be concluded that these strains do not express a CwlO protein. Therefore, the variant LytE proteins expressed in these strain have properties that are sufficient to fulfil the D,L-endopeptidase requirement in *B. subtilis*. This experiment was conducted by Jennifer Lennon under my supervision.

### 5.4.5 Verification of the absence of suppressor mutations in strains that express a variant LytE from the *cwlO* promoter

In the previous section it was shown that the CwlO protein is not produced in strains LU61, LU421, LU418, JL21, JL22 or JL24 expressing variant LytE proteins using the *cwlO* promoter ($P_{cwlO}$) and that lack the endogenous *lytE* gene. However, it is possible that suppressor mutations may have arisen during the construction of these strains that compensate for loss of the CwlO D,L-endopeptidase. It is known that *B. subtilis* is capable of responding to stress by accumulating beneficial mutations (Gunka *et al.*, 2013). Such suppressor mutations may for example upregulate the expression of other D,L-endopeptidases (to counteract a lack of D,L-endopeptidase activity) or down regulate the expression of the variant LytE (which might be expressed at too high level, causing cell wall damage).

Therefore, it was decided to test for the presence of such suppressor mutations by examining transformation frequencies, by transforming increasing concentrations (from 0.05 µg to 1.5 µg) of chromosomal DNA of strains that express a variant LytE from the *cwlO* promoter into the recipient strain LS089 (Δ*lytE*::*cm*$^r$), and selecting for the presence of the antibiotic resistance gene in the donor strains (kanamycin) at the *cwlO* locus. If viability requires only expression of the variant LytE protein at the *cwlO* locus, then transformation frequency should be the same as with chromosomal DNA of strain DN3061 (Δ*lytE*::*kan*$^r$). If however viability requires expression of the variant LytE

proteins at the *cwlO* locus AND the presence of a mutation at another chromosomal site (i.e. two DNA fragments must be transformed), then the transformation frequency will be reduced by orders of magnitude. The strains, listed in Table 5.7, were chosen for this experiment, because it was observed that strains that express a variant LytE protein with a single LysM motifs from the *cwlO* promoter in a Δ*lytE* background (e.g. JL19, LU262, LU270) exhibited lysing colonies after prolonged incubation on plates (see Figure 5.17, page 154). This indicates that these strains are experiencing cell wall stress and are therefore more likely to develop suppressor mutations, than strains that are not exposed to cell wall stress (i.e. LU313, LU320). Strain DN3061 was chosen as a negative control for the reasons that in this strain the *lytE* open reading frame is replaced by a kanamycin resistance cassette. Transforming chromosomal DNA of DN3061 into the recipient strain LS089 (Δ*lytE*::*cm*$^r$) therefore leads to a simple exchange of antibiotic resistance cassettes without altering the genotype of the strain (i.e. not likely to generate suppressor mutations).

The observed transformation frequencies expressed as number of colonies per µg DNA are presented in Figure 5.10. The results show that transformation frequencies with chromosomal DNA from strains JL19, LU262, LU270, LU313 and LU320 are in the same order of magnitude as transformation frequencies with DN3061. For all tested strains, transformation frequencies were high, and roughly correlated with the amount of deployed chromosomal DNA. In general, if more chromosomal DNA was used in the transformation, more transformants were obtained (a maximum of over 3000 colonies per plate for 1.5 µg of chromosomal DNA of strain LU262).

The transformation frequency of all strains was similar to that of the control (DN3061). Although on average, transformation frequencies were lowest for chromosomal DNA of strain JL19, they were still high (over 700 colonies per plate for the lowest amount of chromosomal DNA deployed 0.05 µg).

These results indicate that expression of the variant LytE proteins alone is sufficient to fulfil the D,L-endopeptidase requirement in strains JL19, LU262, LU270, LU313 and LU320. Thus, it is unlikely that mutations at other sites have occurred. Given this high transformation frequencies and their closeness to the transformation frequencies of the control, it is unlikely that any of the investigated strains carries a suppressor mutation.

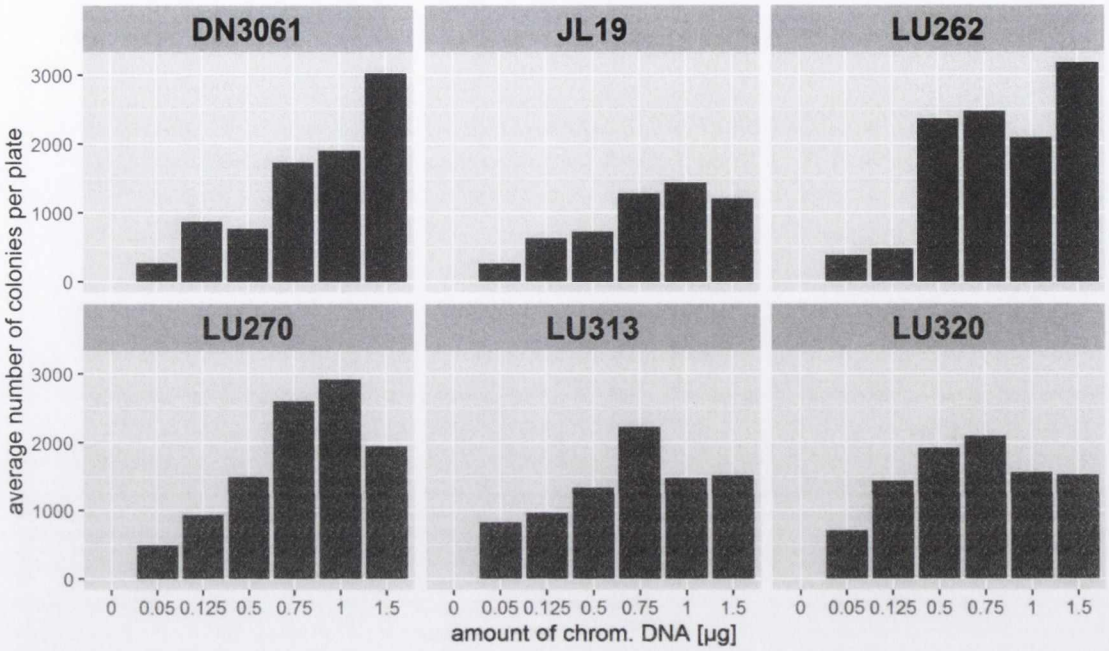



**Figure 5.10: Suppressor mutation test for the detection of potential secondary mutations, enabling the expression of variant LytE constructs at the *cwlO* locus.** The suppressor mutation test was conducted by transforming varying amounts of chromosomal DNA (0 µg, 0.05 µg, 0.125 µg, 0.5 µg, 0.75 µg, 1 µg, 1.5 µg), obtained of the following strains DN3061 (Δ*lytE*::*kan*$^r$), JL19 (Δ*cwlO*::P$_{cwlOleader}$*lytE [L1*$^{xlyB}$*cat*$^{lytE}$*] kan*$^r$ Δ*lytE*::*cm*$^r$), LU262 (Δ*cwlO*::P$_{cwlOleader}$*lytE [L1cat] kan*$^r$ Δ*lytE*::*cm*$^r$), LU270 (Δ*cwlO*::P$_{cwlOleader}$*lytE [L3cat] kan*$^r$ Δ*lytE*::*cm*$^r$), LU313 (Δ*cwlO*::P$_{cwlOleader}$*lytE [L213cat] kan*$^r$ Δ*lytE*::*cm*$^r$) and LU320 (Δ*cwlO*::P$_{cwlOleader}$*lytE [L123cat]*$^R$ *kan*$^r$ Δ*lytE*::*cm*$^r$) into strain LS089 (Δ*lytE*::*cm*$^r$). Positive transformants were selected on kanamycin plates. The bar graph shows the average number of colonies per plate, in relation to the amount of chromosomal DNA deployed in the transformation. R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

**Table 5.7: List of strains investigated by suppressor mutation assay (see Figure 5.10)**

| Strain | Genotype |
|---|---|
| **DN3061** | Δ*lytE*::*kan*$^r$ |
| **JL19** | P$_{cwlOleader}$*lytE [L1*$^{xlyB}$*cat*$^{lytE}$*]* Δ*cwlO kan*$^r$ Δ*lytE*::*cm*$^r$ |
| **LU262** | P$_{cwlOleader}$*lytE [L1cat]* Δ*cwlO kan*$^r$ Δ*lytE*::*cm*$^r$ |
| **LU270** | P$_{cwlOleader}$*lytE [L3cat]* Δ*cwlO kan*$^r$ Δ*lytE*::*cm*$^r$ |
| **LU313** | P$_{cwlOleader}$*lytE [L213cat]* Δ*cwlO kan*$^r$ Δ*lytE*::*cm*$^r$ |
| **LU320** | P$_{cwlOleader}$*lytE [L123cat]*$^R$ Δ*cwlO kan*$^r$ Δ*lytE*::*cm*$^r$ |

### 5.4.6 Cell-associated protein levels of variant LytE proteins *in vivo*

Our results show that all investigated variant LytE constructs (possessing at least one LysM motif) were able to fulfil the D,L-endopeptidase requirement indicating that they are expressed and functional. However, it is possible that the type of variant LytE protein may affect the cellular protein level of the variant LytE protein. To address this issue, the level of variant LytE protein was determined by western blot analysis in a selection of strains in which the variant LytE protein was HA-tagged (Table 5.8).

**Table 5.8: List of strains containing a HA-tagged variant LytE**

| | HA-tagged variant *lytE* at the *cwlO* locus | | | HA-tagged variant *lytE* at the *lytE* locus | | |
|---|---|---|---|---|---|---|
| **LysM** | **Strain** | **Genotype** | ***** | **Strain** | **Genotype** | ***** |
| One LysM | LU418 | $P_{cwlOleader}$*lytE-HA [L1cat] ΔcwlO ΔlytE* | V | LU435 | $P_{lytE}$*lytE-HA [L1cat] ΔcwlO ΔlytE* | V |
| | LU419 | $P_{cwlOleader}$*lytE-HA [L2cat] ΔcwlO ΔlytE* | V | LU442 | $P_{lytE}$*lytE-HA [L2cat] ΔcwlO ΔlytE* | V |
| | LU440 | $P_{cwlOleader}$*lytE-HA [L3cat] ΔcwlO ΔlytE* | V | | | |
| | JL21 | $P_{cwlOleader}$*lytE-HA [L1*$^{lytF}$*cat*$^{lytE}$*] ΔcwlO* | V | | | |
| | JL22 | $P_{cwlOleader}$*lytE-HA [L1*$^{xlyB}$*cat*$^{lytE}$*] ΔcwlO* | V | | | |
| | JL24 | $P_{cwlOleader}$*lytE-HA [L1*$^{xlyB}$*cat*$^{lytE}$*] ΔcwlO ΔlytE* | V | | | |
| Two LysM | LU414 | $P_{cwlOleader}$*lytE-HA [L11cat] ΔcwlO ΔlytE* | V | LU425 | $P_{lytE}$*lytE-HA [L11cat] ΔcwlO ΔlytE* | V |
| | LU415 | $P_{cwlOleader}$*lytE-HA [L13cat] ΔcwlO ΔlytE* | V | LU427 | $P_{lytE}$*lytE-HA [L13cat] ΔcwlO ΔlytE* | V |
| | LU416 | $P_{cwlOleader}$*lytE-HA [L21cat] ΔcwlO ΔlytE* | V | | | |
| | LU417 | $P_{cwlOleader}$*lytE-HA [L31cat] ΔcwlO ΔlytE* | V | | | |
| Thee LysM | LU421 | $P_{cwlOleader}$*lytE-HA [L123cat]*$^{R}$ *ΔcwlO ΔlytE* | V | LU422 | $P_{lytE}$*lytE-HA [L123cat]*$^{R}$ *ΔcwlO ΔlytE* | V |
| | LU432 | $P_{cwlOleader}$*lytE-HA [L213cat] ΔcwlO ΔlytE* | V | LU443 | $P_{lytE}$*lytE-HA [L213cat] ΔcwlO ΔlytE* | V |
| | | | | LU424 | $P_{lytE}$*lytE-HA [L331cat] ΔcwlO ΔlytE* | V |

* viable (V) strain constructed or strain not obtained (-)

Lysates were prepared for the strains listed in Table 5.8 that each express a LytE-HA protein i.e. tagged with a HA-epitope as described in experimental procedures (section 2.1.4.11 and 2.1.4.13.). It is important to note that the variant LytE in all these strains could be tagged with HA-epitope. Since expression of a LytE-HA variant in each of these strains is essential for fulfilling the D,L-endopeptidase requirement, it can be concluded that the HA-tag does not impair the functionality of the variant LytE. Protein levels of the HA-tagged variant LytE proteins in these strains were established by western blot analysis and the results are shown in Figure 5.11 (page 138), Figure 5.12 (page 140) and Figure 5.13 (page 142), each of which is discussed in detail in the following sections.

#### 5.4.6.1 Effect of different promoter activities ($P_{cwlO}$ vs $P_{lytE}$) on variant LytE cell-associated protein levels

It has been established that the *cwlO* promoter ($P_{cwlO}$) is stronger than the *lytE* promoter ($P_{lytE}$), even when the latter is expressed under the control of SigI under cell wall stress conditions (this study). Therefore, it is expected that the level of cell-associated variant LytE protein will be higher when expressed using the $P_{cwlO}$ that with the $P_{lytE}$ promoter. To test this hypothesis, the level of cell-associated variant LytE protein was determined in the strains listed in Table 5.9. Four different variant LytE were selected for this study (LytE[L2cat], LytE[L13cat], LytE[L123cat]$^R$, LytE[L213cat]) and protein levels were compared between strains that express them either from the $P_{cwlO}$ or the $P_{lytE}$ promoters. Lysates were prepared from exponentially growing cells of the strains listed in Table 5.9 and western blot analysis was carried out as described in experimental procedures.

The result is shown in Figure 5.11. All LytE-HA variant proteins were detected and bands migrated according to the predicted MW. No non-specific bands were detected (see nc), but a degradation product of ~15kDa was detected for some proteins (e.g. LytE[L213cat] of LU432 and LytE[L13cat] of LU415). Interestingly, although the bands for both variant LytE proteins LytE[L123cat]$^R$ (LU421 and LU422) and LytE[L213cat] (LU432 and LU443) migrated to the same position (approx. 37 kDa), an additional band was observed in strains that expressed LytE[L213] that migrated to a position slightly above the correct sized band of LytE[L213cat]. The presence of this band may indicate that there is a defect in processing LytE[L213cat] during secretion – the larger band perhaps retaining the signal peptide. The level of variant LytE protein varied significantly in the strains tested.

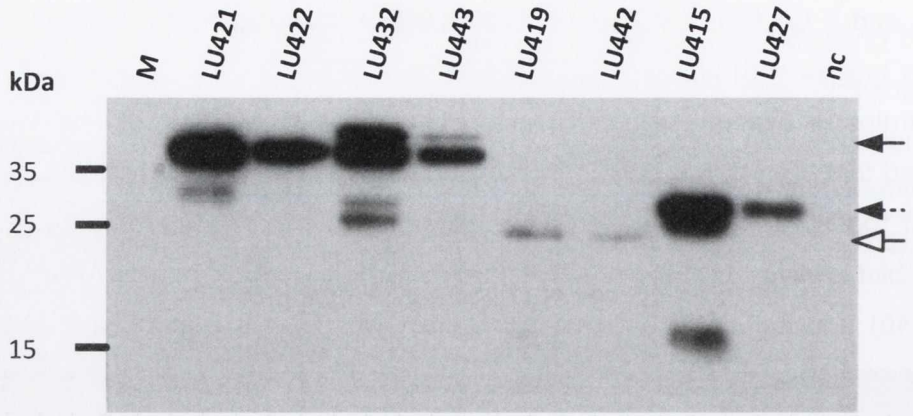



**Figure 5.11: The levels of cell-associated LytE variant proteins, expressed either from the *cwlO* promoter or from the *lytE* promoter.** Cell-associated protein samples of the following strains are shown: LU421 ($P_{cwlOleader}$*lytE-HA [L123cat]*$^R$ Δ*cwlO* Δ*lytE*), LU422 ($P_{lytE}$*lytE-HA [L123cat]*$^R$ Δ*cwlO* Δ*lytE*), LU432 ($P_{cwlOleader}$*lytE-HA [L213cat]* Δ*cwlO* Δ*lytE*), LU443 ($P_{lytE}$*lytE-HA [L213cat]* Δ*cwlO* Δ*lytE*), LU419 ($P_{cwlOleader}$*lytE-HA [L2cat]* Δ*cwlO* Δ*lytE*), LU442 ($P_{lytE}$*lytE-HA [L2cat]* Δ*cwlO* Δ*lytE*), LU415 ($P_{cwlOleader}$*lytE-HA [L13cat]* Δ*cwlO* Δ*lytE*) and LU427 ($P_{lytE}$*lytE-HA [L13cat]* Δ*cwlO* Δ*lytE*). A negative control nc (WT168 without HA-tag) is also displayed. Cells were grown in salt-free buffered LB and harvested at an $OD_{600}$ of 0.25. Total cell protein was prepared from the samples and the HA-tagged proteins were detected by western analysis. Seven µg total cell protein extract was deployed per lane. Molecular weight markers are shown (M, kDa). The film was exposed for seven minutes. The full arrow indicates the full length LytE protein (LytE, MW = 37 kDa), the broken arrow indicates a full length dimer protein (variant LytE with two LysM motifs, MW = 30 kDa) and the open arrow indicates a full length monomer protein (variant LytE with one LysM motif, MW = 23 kDa). R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

**Table 5.9: List of strains investigated by western analysis (see Figure 5.11)**

| variant *lytE* at *cwlO* locus | | variant *lytE* at *lytE* locus | |
|---|---|---|---|
| **Strain** | **Genotype** | **Strain** | **Genotype** |
| **LU421** | $P_{cwlOleader}$*lytE-HA [L123cat]*$^R$ Δ*cwlO* Δ*lytE* | **LU422** | $P_{lytE}$*lytE-HA [L123cat]*$^R$ Δ*cwlO* Δ*lytE* |
| **LU432** | $P_{cwlOleader}$*lytE-HA [L213cat]* Δ*cwlO* Δ*lytE* | **LU443** | $P_{lytE}$*lytE-HA [L213cat]* Δ*cwlO* Δ*lytE* |
| **LU419** | $P_{cwlOleader}$*lytE-HA [L2cat]* Δ*cwlO* Δ*lytE* | **LU442** | $P_{lytE}$*lytE-HA [L2cat]* Δ*cwlO* Δ*lytE* |
| **LU415** | $P_{cwlOleader}$*lytE-HA [L13cat]* Δ*cwlO* Δ*lytE* | **LU427** | $P_{lytE}$*lytE-HA [L13cat]* Δ*cwlO* Δ*lytE* |

For all four investigated variant LytE proteins there was more cell-associated protein in strains in which the expression of the variant LytE was initiated by $P_{cwlO}$ than $P_{lytE}$ (compare LU421 to LU422, LU432 to LU443, LU419 to LU442, LU415 to LU427).

Furthermore, the level of cell-associated variant LytE proteins was higher for proteins with two or more LysM motifs (LytE[L13cat], LytE[L123cat]$^{R}$ and LytE[L213cat]) than for proteins with a single LysM motif (LytE[L2cat]) (compare LU421, LU422, LU432, LU443, LU415, LU427 to LU419 and LU442). From this it can be concluded that the level of cell-associated variant LytE protein is higher when expression is driven by the *cwlO* promoter than when expression is driven by the *lytE* promoter. Secondly, it is clear that a higher level of variant LytE remains cell-associated for those proteins with two or more LysM motifs than those with only one LysM motif.

### 5.4.6.2 Effect of the type of variant LytE on cell-associated protein levels

Our results suggest that the number and order of LysM motifs in each variant LytE influences the level of protein that is cell-associated. To further explore this observation, the level of cell-associated LytE protein was compared between strains (listed in Table 5.10) expressing variant LytE proteins with three (LytE[L123cat], LytE[L213cat]), two (LytE[L11cat], LytE[L13cat]) and one LysM motifs (LytE[L1cat], LytE[L2cat]) and with all proteins being expressed from the *lytE* promoter. Two strains were used as controls: LU45 (*lytE-HA*) and LU47 (*lytE-HA* Δ*cwlO*) which both express the HA-tagged wild-type LytE[L123cat]$^{WT168}$ protein, but a higher level of cell-associated protein would be expected in LU47 where the *lytE* promoter is upregulated by activation of SigI due to the absence of *cwlO* (Salzberg *et al.*, 2013).

The result of the western blot analysis is shown in Figure 5.12. With one exception, all tested proteins could be detected and their bands migrated at the positions expected from their predicted MW. The exception was the variant LytE[L1cat] (LU435), where the full length protein (predicted MW ~21 kDa) could not be detected. However, detection of this specific degradation product (predicted MW ~15 kDa), which is also specifically detected in the lysate prepared from strain LU425 (LytE[L11cat]) indicates that the HA-tagged variant LytE protein L1cat is synthesized, but is rapidly degraded. No unspecific bands were detected (nc). The level of variant LytE protein varied significantly in the strains tested. As observed earlier levels of cell-associated proteins were higher in a Δ*cwlO* background (compare LU45 to LU47). Cell-associated protein levels were comparable between LytE[L123cat]$^{WT168}$ and LytE[L123cat]$^{R}$ (compare LU47 and LU422), but slightly lower for LytE[L213cat] (LU443).

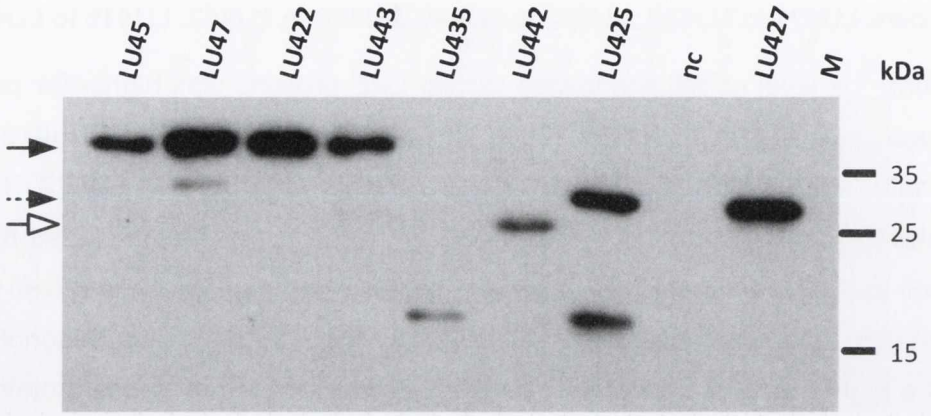



**Figure 5.12: The levels of cell-associated LytE variant proteins expressed from the *lytE* promoter.** Cell-associated protein samples of the following strains are shown: LU45 (*lytE-HA*), LU47 (*ΔcwlO lytE-HA*), LU422 ($P_{lytE}$*lytE-HA [L123cat]*$^R$ *ΔcwlO ΔlytE*), LU443 ($P_{lytE}$*lytE-HA [L213cat] ΔcwlO ΔlytE*), LU435 ($P_{lytE}$*lytE-HA [L1cat] ΔcwlO ΔlytE*), LU442 ($P_{lytE}$*lytE-HA [L2cat] ΔcwlO ΔlytE*), LU425 ($P_{lytE}$*lytE-HA [L11cat] ΔcwlO ΔlytE*) and LU427 ($P_{lytE}$*lytE-HA [L13cat] ΔcwlO ΔlytE*). A negative control nc (WT168 without HA-tag) is also displayed. Cells were grown in salt-free buffered LB and harvested at an $OD_{600}$ of 0.25. Total cell protein was prepared from the samples and the HA-tagged proteins were detected by western analysis. Ten µg total cell protein extract was deployed per lane. Molecular weight markers are shown (M, kDa). The film was exposed for two minutes. The full arrow indicates the full length LytE protein (LytE, MW = 37 kDa), the broken arrow indicates a full length dimer protein (variant LytE with two LysM motifs, MW = 30 kDa) and the open arrow indicates a full length monomer protein (variant LytE with one LysM motif, MW = 23 kDa). R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

**Table 5.10: List of strains investigated by western analysis (see Figure 5.12)**

| LysM motif | Strain | Genotype |
|---|---|---|
| One LysM | LU435 | $P_{lytE}$*lytE-HA [L1cat] ΔcwlO ΔlytE* |
| | LU442 | $P_{lytE}$*lytE-HA [L2cat] ΔcwlO ΔlytE* |
| Two LysM | LU425 | $P_{lytE}$*lytE-HA [L11cat] ΔcwlO ΔlytE* |
| | LU427 | $P_{lytE}$*lytE-HA [L13cat] ΔcwlO ΔlytE* |
| Three LysM | LU422 | $P_{lytE}$*lytE-HA [L123cat]*$^R$ *ΔcwlO ΔlytE* |
| | LU443 | $P_{lytE}$*lytE-HA [L213cat] ΔcwlO ΔlytE* |
| Control | LU45 | *lytE-HA* |
| | LU47 | *lytE-HA ΔcwlO* |
| | WT168 | No HA-tag [control] |

Similarly cell-associated protein levels were detected for the two variant LytE with two LysM motifs LytE[L13cat] (LU427) and LytE[L11cat] (LU425). However, variant LytE proteins with a single LysM motif exhibited less cell-associated protein (compare LytE[L1cat] of LU435 and LytE[L2cat] of LU442). This result confirms that cell-associated protein levels are indeed affected by the type of variant LytE protein. Variant LytE proteins possessing two or three LysM motifs have considerably higher cell-associated protein levels than variant LytE proteins with a single LysM motif.

#### 5.4.6.3 Comparison between cell-associated and secreted variant LytE protein levels

Our previous results show that the level of cell-associated variant LytE protein is affected by the number of LysM motifs in the protein: cells expressing variant LytE with more than one LysM motif have higher levels of cell-associated protein. Similar observations have been made in *in vitro* studies (Mesnage *et al.*, 2014; Wong *et al.*, 2014), which concluded that individual LysM motifs contribute additively to peptidoglycan-binding. Thus, this may account for our observation that variant LytE with only one LysM motif are less capable of binding to the cell wall than those with two or more LysM motifs. If it is assumed that the lower level of cell-associated LytE is caused by lowered binding to the cell wall, and not lowered expression, then we predict the level of variant LytE protein found in the medium will have the reverse profile of that for cell-associated protein. Thus, to address this issue, we compare the levels of cell-associated variant LytE proteins with those in the culture supernatant.

**Table 5.11: List of strains investigated by western analysis (see Figure 5.13)**

| LysM motif | Strain | Genotype |
|---|---|---|
| One LysM | LU418 | $P_{cwlOleader}$*lytE-HA [L1cat] ΔcwlO ΔlytE* |
| | LU419 | $P_{cwlOleader}$*lytE-HA [L2cat] ΔcwlO ΔlytE* |
| | LU440 | $P_{cwlOleader}$*lytE-HA [L3cat] ΔcwlO ΔlytE* |
| Two LysM | LU415 | $P_{cwlOleader}$*lytE-HA [L13cat] ΔcwlO ΔlytE* |
| | LU417 | $P_{cwlOleader}$*lytE-HA [L31cat] ΔcwlO ΔlytE* |
| | LU414 | $P_{cwlOleader}$*lytE-HA [L11cat] ΔcwlO ΔlytE* |
| Three LysM | LU421 | $P_{cwlOleader}$*lytE-HA [L123cat]*$^{R}$ *ΔcwlO ΔlytE* |
| | LU432 | $P_{cwlOleader}$*lytE-HA [L213cat] ΔcwlO ΔlytE* |
| Control | LU61 | $P_{cwlOleader}$*lytE-HA ΔcwlO ΔlytE* |

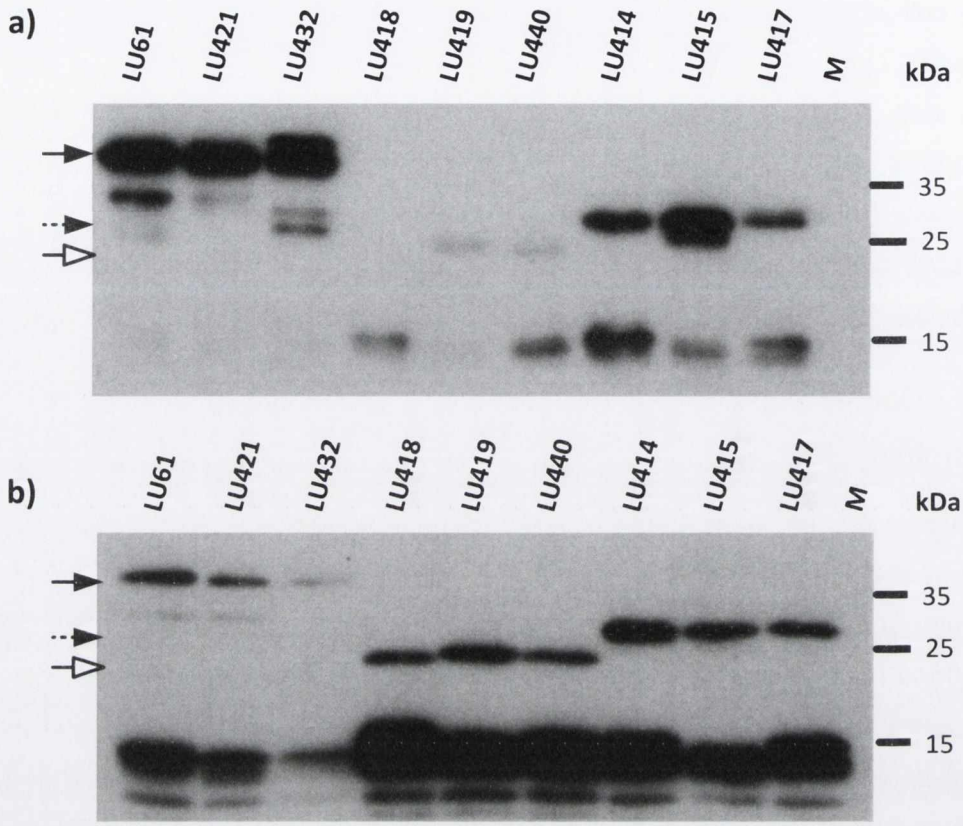



**Figure 5.13: The levels of LytE variant proteins expressed from the *cwlO* promoter, for cell-associated proteins (a) and secreted proteins (b).** Protein samples of the following strains are shown: LU61 (*ΔcwlO* $P_{cwlOleader}$*lytE-HA ΔlytE*), LU421 ($P_{cwlOleader}$*lytE-HA [L123cat]*$^{R}$ *ΔcwlO ΔlytE*), LU432 ($P_{cwlOleader}$*lytE-HA [L213cat] ΔcwlO ΔlytE*), LU418 ($P_{cwlOleader}$*lytE-HA [L1cat] ΔcwlO ΔlytE*), LU419 ($P_{cwlOleader}$*lytE-HA [L2cat] ΔcwlO ΔlytE*), LU440 ($P_{cwlOleader}$*lytE-HA [L3cat] ΔcwlO ΔlytE*), LU414 ($P_{cwlOleader}$*lytE-HA [L11cat] ΔcwlO ΔlytE*), LU415 ($P_{cwlOleader}$*lytE-HA [L13cat] ΔcwlO ΔlytE*) and LU417 ($P_{cwlOleader}$*lytE-HA [L31cat] ΔcwlO ΔlytE*). A negative control nc (WT168 without HA-tag) is also displayed. Cells were grown in salt-free buffered LB and harvested at an $OD_{600}$ of 0.25. Total cell protein was prepared from the samples and the HA-tagged proteins were detected by western analysis. Five µg total cell protein extract was deployed per lane **(a)** and 0.15 OD equivalent media proteins were deployed per lane **(b)**. Molecular weight markers are shown (M, kDa). Both films were exposed for seven minutes. The full arrow indicates the full length LytE protein (LytE, MW = 37 kDa), the broken arrow indicates a full length dimer protein (variant LytE with two LysM motifs, MW = 30 kDa) and the open arrow indicates a full length monomer protein (variant LytE with one LysM motif, MW = 23 kDa). R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

Nine variant LytE were selected for this experiment with varying number and combinations of LysM motifs (LytE[L123cat]$^{R}$, LytE[L213cat], LytE[L1cat], LytE[L2cat], LytE[L3cat], LytE[L11cat], LytE[L13cat], LytE[L31cat]) and protein samples were harvested from the strains listed in Table 5.11 expressing these proteins from the *cwlO* promoter. Furthermore, strain LU61 was included as a control. Thus, for each strain the level of variant LytE protein that is cell-associated and that present in the culture supernatant was established by western blot analysis as described in experimental procedures.

The result of this comparison is shown in Figure 5.13. With one exception, all proteins were expressed and a band, migrating at the expected position of each full length protein, could be detected. The single exception was the variant LytE protein expressed in strain LU418 (LytE[L1cat]), where a band corresponding to the full length protein (predicted to be ~ 21 kDa) could not be detected. Instead a specific band (presumed degradation product) with an estimated size of around 15 kDa was observed which suggests that this strain did indeed express the HA-tagged variant LytE protein. A similar band was observed for most samples and the presumed degradation product was generally stronger in the samples harvested from the media than in the cell-associated protein samples.

As observed earlier, although the bands for variant LytE proteins with three LysM motifs LytE[L123cat] (LU421, LU61) and LytE[L213cat] (LU432) migrated at the same position (around 37 kDa), there was an additional band that migrated above the full length LytE[L213cat] variant LytE protein in strain LU432 that may indicate a defect in processing the signal peptide in this strain. Importantly, a strong inverse correlation in the level of variant LytE was observed between protein that is cell-associated and protein found in the media. Strains that had a high level of cell-associated variant LytE protein e.g. LU61 (LytE WT168), LU421 (LytE[L123cat]$^{R}$) and LU432 (LytE[LU213cat]), had low levels in the media while proteins with low levels for the cell-associated variant LytE proteins e.g. strains LU418 (LytE[L1cat]), LU419 (LytE[L2cat]) and LU440 (LytE[L3cat]) had high levels of these proteins in the culture media. This correlation between number of LysM motifs and cell association / presence in the medium is supported by the distribution of variant LytE with two LysM motifs which show profiles intermediate between proteins with three and one LysM motifs respectively. Strains expressing variant LytE with two LysM motifs (e.g. LU415 (LytE[L13cat]), LU417 (LytE[L31cat]) and L440 (LytE[L11cat]) show a lower level of cell-associated protein than those with three LysM motifs, but a higher level than those with one LysM motif. Conversely, the level of LysM protein in the culture

medium is lower than those proteins with one LysM motif, but higher than those with three LysM motifs.

This data indicates that the number of LysM motifs in variant LytE proteins is a determinant of the extent to which the protein is cell-associated or shed into the culture media. Variant LytE proteins with one LysM motif were predominantly found in the culture supernatant while variant LytE proteins with three LysM motifs were predominantly found in association with the cell wall. Moreover the distribution of variant LytE proteins with two LysM motifs takes an intermediate between that found for those with three and one LysM motifs. It is concluded therefore that the observed differences in distribution of proteins observed *in vivo* is due to difference in the number of LysM motifs in the different variant LytE with increasing number of LysM motifs leading to increased binding, in agreement with the *in vitro* experiments reported by Mesnage *et al.* (2014) and Wong *et al.* (2014).

#### 5.4.6.4 Comparison between cell-associated and secreted heterologous variant LytE protein levels

Our previous results suggest that proteins with different numbers of LysM motifs are probably expressed at similar levels, but that the number of LysM motifs determines the extent to which the protein is distributed between the cell wall and the culture supernatant. We show that variant LytE proteins with one LysM motif bind poorly to peptidoglycan with the expressed protein being found predominantly in the culture supernatant. The converse holds for proteins with higher numbers of LysM motifs (see Figure 5.13, page 142). Thus, we sought to establish the distribution of variant LysM proteins that have a heterologous LysM motif.

The distribution of two heterologous variant LytE proteins in strains JL21 (LytE[$L1^{LytF}cat^{LytE}$]) and JL22 and JL24 (both LytE[$L1^{XlyB}cat^{LytE}$]) was compared with the wild-type LytE protein (LU61, LU421) and LytE[L1cat] (LU418). It is important to note that strains JL22 and JL24 express the same heterologous variant LytE protein, but that strain LU24 does not have the endogenous *lytE* gene. An added impetus for this experiment was the finding that LytE[$L1^{XlyB}cat^{LytE}$] was able to fulfil the D,L-endopeptidase requirement, but LytE[$L1^{LytF}cat^{LytE}$] was not. It is therefore interesting to establish if this lack of complementation was due to lack of expression or deficiency in binding.

**Table 5.12: List of strains investigated by western analysis (see Figure 5.14)**

| LysM motif | Strain | Genotype |
|---|---|---|
| One LysM | LU418 | $P_{cwlOleader}$*lytE-HA [L1cat] ΔcwlO ΔlytE* |
| | JL21 | $P_{cwlOleader}$*lytE-HA* [$L1^{lytF}cat^{lytE}$] *ΔcwlO* |
| | JL22 | $P_{cwlOleader}$*lytE-HA* [$L1^{xlyB}cat^{lytE}$] *ΔcwlO* |
| | JL24 | $P_{cwlOleader}$*lytE-HA* [$L1^{xlyB}cat^{lytE}$] *ΔcwlO ΔlytE* |
| Three LysM | LU421 | $P_{cwlOleader}$*lytE-HA [L123cat]*$^{R}$ *ΔcwlO ΔlytE* |
| Control | LU61 | $P_{cwlOleader}$*lytE-HA ΔcwlO ΔlytE* |

Protein samples were harvested from the strains listed in Table 5.12 and processed according to Chapter 2. The result of the western blot is shown in Figure 5.14. Our results show that a band corresponding to the full length protein in strains JL22 and JL24 (LytE[$L1^{XlyB}cat^{LytE}$]) was not detected and that the band corresponding to the full length variant LytE protein in strain JL21 (LytE[$L1^{LytF}cat^{LytE}$]) was weakly detected. However, that they were expressed was indicated by the presence of a specific band of ~15 kDA in size that is likely to be a degradation product of these proteins. No non-specific bands were observed (nc). The western blot therefore confirms that all heterologous variant LytE proteins, including LytE[$L1^{LytF}cat^{LytE}$] (JL21) were expressed.

As observed earlier, the LytE[L123cat] protein was mainly detected in the cell-associated sample with little protein present in the media sample. Also there was a very low amount of variant LytE protein with one LysM motif found to be cell-associated, but there was a high level of a putative LytE degradation observed in samples prepared from the culture supernatant. A similar scenario was found for the distribution of the heterologous variant LytE proteins. Proteins LytE[$L1^{LytF}cat^{LytE}$] and LytE[$L1^{XlyB}cat^{LytE}$] were not detected in samples prepared from cells indicating that only a very low level of proteins was cell-associated. However, there was a strong band of the putative degradation product (~15 kDa) present in samples prepared from the culture supernatant. Furthermore, the distribution of the variant LytE[$L1^{XlyB}cat^{LytE}$] protein was not different in samples prepared from strains JL22 and JL24 - both express the heterologous variant LytE[$L1^{XlyB}cat^{LytE}$] protein from the $P_{cwlO}$, *but* vary in having (JL22) or not having (JL24) a functional *lytE*. It is important to note that the level of cell-associated LytE[$L1^{XlyB}cat^{LytE}$] was below detection, despite this protein being capable of fulfilling the D,L-endopeptidase requirement for cell viability.

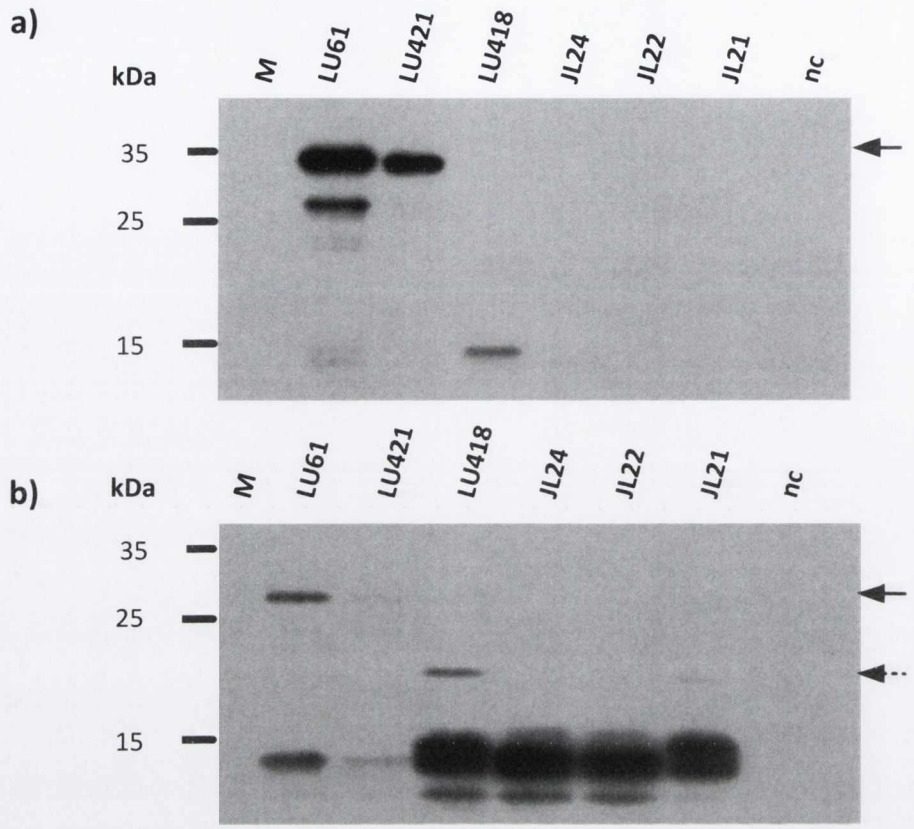



**Figure 5.14: Levels of heterologous variant LytE proteins, expressed from the *cwlO* promoter for cell-associated proteins (a) and secreted proteins (b).** Protein samples of the following strains are shown: LU61 (Δ*cwlO* $P_{cwlOleader}$*lytE-HA* Δ*lytE*), LU421 ($P_{cwlOleader}$*lytE-HA [L123cat]*$^{R}$ Δ*cwlO* Δ*lytE*), LU418 ($P_{cwlOleader}$*lytE-HA [L1cat]* Δ*cwlO* Δ*lytE*), JL21 ($P_{cwlOleader}$*lytE-HA [L1*$^{lytF}$*cat*$^{lytE}$*]* Δ*cwlO*), JL22 ($P_{cwlOleader}$*lytE-HA [L1*$^{xlyB}$*cat*$^{lytE}$*]* Δ*cwlO*) and JL24 ($P_{cwlOleader}$*lytE-HA [L1*$^{xlyB}$*cat*$^{lytE}$*]* Δ*cwlO* Δ*lytE*). A negative control nc (WT168 without HA-tag) is also displayed. Cells were grown in salt-free buffered LB and harvested at an $OD_{600}$ of 0.3. Total cell protein was prepared from the samples and the HA-tagged proteins were detected by western analysis. Five µg total cell protein extract was deployed per lane **(a)** and 0.15 OD equivalent media proteins were deployed per lane **(b)**. Molecular weight markers are shown (M, kDa). Both films were exposed for ten minutes. The full arrow indicates the full length LytE protein (LytE, MW = 37 kDa) and the broken arrow indicates a full length dimer protein (variant LytE with two LysM motifs, MW = 30 kDa). Western blot analysis was conducted by Jennifer Lennon under my supervision. R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

In summary it can be concluded that heterologous variant LytE[L1$^{LytF}$cat$^{LytE}$] and LytE[L1$^{XlyB}$cat$^{LytE}$] have a cellular distribution similar to that of variant LytE with a single homologous LysM motif. However, the LytE[L1$^{XlyB}$cat$^{LytE}$] protein is capable of fulfilling the D,L-endopeptidase requirement despite a very low level of cell-associated protein. This experiment was conducted by Jennifer Lennon under my supervision.

### 5.4.7 Assessment of the extent to which variant LytE proteins fulfil the D,L-endopeptidase requirement by measuring induction of the SigI regulon

It was established earlier that the Sig regulon is upregulated in a *cwlO* mutant, which results in higher expression levels of LytE protein (*lytE* is a member of the SigI regulon) (see Figure 3.3, page 68). Previous results show that the SigI regulon is upregulated in strain LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) and that this can be suppressed by increasing *lytE* expression through removal of the *cwlO* leader sequence (strain LU25; $P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*), - both expresses *lytE* from the $P_{cwlO}$ promoter (see Figure 3.4, page 71). Thus, the extent to which the SigI regulon is induced is an indirect measure of the extent to which the D,L-endopeptidase requirement is fulfilled and hence also of cell wall stress. It was therefore hypothesized that the extent to which each variant LytE protein fulfils the D,L-endopeptidase requirement could be assessed by determining the extent to which the SigI regulon was induced. This was determined by establishing the mRNA transcript levels of SigI regulon genes in strains expressing a variant LytE protein (listed in Table 5.13) by qRT-PCR.

**Table 5.13: List of strains investigated by qRT-PCR analysis (see Figure 5.15)**

| LysM motif | Strain | Genotype |
|---|---|---|
| Control | WT168 | |
| | EB35 | Δ*cwlO* |
| | LS089 | Δ*lytE* |
| One LysM | LU262 | $P_{cwlOleader}$*lytE [L1cat]* Δ*cwlO* Δ*lytE* |
| | LU266 | $P_{cwlOleader}$*lytE [L2cat]* Δ*cwlO* Δ*lytE* |
| | LU270 | $P_{cwlOleader}$*lytE [L3cat]* Δ*cwlO* Δ*lytE* |
| | JL19 | $P_{cwlOleader}$*lytE* [L1$^{xlyB}$cat$^{lytE}$] Δ*cwlO* Δ*lytE* |
| Three LysM | LU320 | $P_{cwlOleader}$*lytE [L123cat]*$^{R}$ Δ*cwlO* Δ*lytE* |

The results of the qRT-PCR analysis are shown in Figure 5.15. SigI / MreBH transcripts were measured relative to those in wild-type strain WT168. As observed earlier the two

members of the SigI regulon (*sigI* and *mreBH*) were both upregulated (by about 3 and 9-fold) in the *cwlO* mutant (EB35), but not in the *lytE* mutant (LS089). Therefore, deletion of *lytE* alone when CwlO is expressed from its own promoter does not lead to SigI induction. Interestingly a strain that expresses a variant LytE protein with the wild-type combination (LU320, LytE[L123cat]$^{R}$) from the *cwlO* promoter and does not possess an endogenous *lytE* gene also showed no upregulated *sigI* and only slightly upregulated *mreBH* transcript levels (3.5-fold). Similarly a strain that expresses a variant LytE protein with the combination L1cat in the same background also exhibited wild-type transcript levels for *sigI* and *mreBH.* However, both transcript levels were elevated if a strain contained a variant LytE with a single LysM motif with the combination LytE[L2cat], LytE[L3cat] or LytE[L1$^{XlyB}$cat$^{LytE}$]. *SigI* and *mreBH* transcript levels were upregulated in strain LU266 (LytE[L2cat]) by about 3.5-fold and 11-fold respectively. In strain LU270 (LytE[L3cat]) both transcripts were elevated by about 3-fold (*sigI*) and 7-fold (*mreBH*). In contrast *sigI* transcript levels were not significantly elevated in strain JL19 (LytE[L1$^{XlyB}$cat$^{LytE}$]). However, *mreBH* transcript levels in the same strain were upregulated by 4-fold.

These results indicate that: (i) The three LysM motifs of LytE are not functionally equivalent. (ii) L1 function is capable of conferring on the catalytic domains the ability to fulfil the D,L-endopeptidase requirement to a greater extent than L2 and L3. This suggests that L1 may bind more strongly to peptidoglycan than L2 or L3. (iii) SigI is not induced in a strain expressing LytE[L1$^{XlyB}$cat$^{LytE}$] indicating that this combination is also capable of complementation. This experiment was conducted by Jennifer Lennon under my supervision.

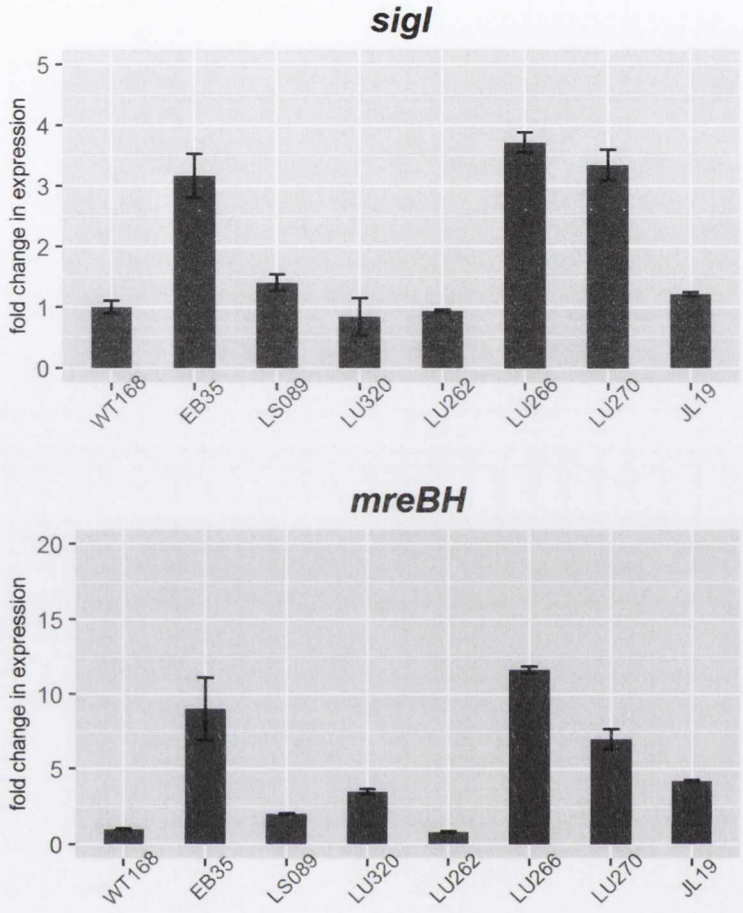



**Figure 5.15: *sigI* and *mreBH* transcript levels as determined by qRT-PCR, for strains that express a variant LytE protein from the *cwlO* promoter.** Transcript levels were established for the following strains: WT168, EB35 (Δ*cwlO*), *LS089* (Δ*lytE*), LU320 ($P_{cwlOleader}$*lytE [L123cat]*$^{R}$ Δ*cwlO* Δ*lytE*), LU262 ($P_{cwlOleader}$*lytE [L1cat]* Δ*cwlO* Δ*lytE*), LU266 ($P_{cwlOleader}$*lytE [L2cat]* Δ*cwlO* Δ*lytE*) and LU270 ($P_{cwlOleader}$*lytE [L3cat]* Δ*cwlO* Δ*lytE*), JL19 ($P_{cwlOleader}$*lytE [L1*$^{xlyB}$*cat*$^{lytE}$*]* Δ*cwlO* Δ*lytE*). Cells were grown in HPDM and samples harvested at $OD_{600}$=0.15. RNA was prepared from the samples and transcribed into cDNA, which was analysed by qRT-PCR with specific primer sets. Transcript levels were normalised relative to the level of transcript for the WT168 sample, which was assigned a value of 1. Error bars are shown. qRT-PCR analysis was conducted by Jennifer Lennon under my supervision. R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

### 5.4.8 Growth behaviour of strains containing a variant LytE under normal and stress conditions

It was shown that cell growth is altered when D,L-endopeptidase expression is varied. For example altered growth is observed in strain LU116, in which the D,L-endopeptidase requirement is fulfilled by expressing LytF from the $P_{cwlO}$ and the endogenous *cwlO* and *lytE* genes are deleted (see Figure 4.4, page 91). Therefore, it was decided to test whether the growth of strains that express variant LytE proteins (either from the *cwlO* promoter or from the *lytE* promoter) was affected. Growth profiles in LB, LPDM and HPDM media were established for representative strains (listed in Table 5.14) that express variant LytE as well as the following controls: WT168, EB35 (Δ*cwlO*), LS089 (Δ*lytE*), LU17 ($P_{cwlOleader}$*lytE* Δ*cwlO*) and LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*).

**Table 5.14: List of strains investigated in growth assays (see Figure 5.16)**

| | variant *lytE* at *cwlO* locus | | variant *lytE* at *lytE* locus | |
|---|---|---|---|---|
| **LysM** | **Strain** | **Genotype** | **Strain** | **Genotype** |
| One LysM | LU262 | $P_{cwlOleader}$*lytE [L1cat]* Δ*cwlO* Δ*lytE* | LU350 | $P_{lytE}$*lytE [L1cat]* Δ*lytE* Δ*cwlO* |
| | LU266 | $P_{cwlOleader}$*lytE [L2cat]* Δ*cwlO* Δ*lytE* | LU352 | $P_{lytE}$*lytE [L2cat]* Δ*lytE* Δ*cwlO* |
| | LU270 | $P_{cwlOleader}$*lytE [L3cat]* Δ*cwlO* Δ*lytE* | LU354 | $P_{lytE}$*lytE [L3cat]* Δ*lytE* Δ*cwlO* |
| Two LysM | LU205 | $P_{cwlOleader}$*lytE [L11cat]* Δ*cwlO* Δ*lytE* | LU407 | $P_{lytE}$*lytE [L11cat]* Δ*lytE* Δ*cwlO* |
| | LU217 | $P_{cwlOleader}$*lytE [L21cat]* Δ*cwlO* Δ*lytE* | LU409 | $P_{lytE}$*lytE [L21cat]* Δ*lytE* Δ*cwlO* |
| | LU229 | $P_{cwlOleader}$*lytE [L31cat]* Δ*cwlO* Δ*lytE* | LU383 | $P_{lytE}$*lytE [L31cat]* Δ*lytE* Δ*cwlO* |
| | LU213 | $P_{cwlOleader}$*lytE [L13cat]* Δ*cwlO* Δ*lytE* | LU411 | $P_{lytE}$*lytE [L13cat]* Δ*lytE* Δ*cwlO* |
| Three LysM | LU277 | $P_{cwlOleader}$*lytE [L331cat]* Δ*cwlO* Δ*lytE* | LU385 | $P_{lytE}$*lytE [L331cat]* Δ*lytE* Δ*cwlO* |
| | LU313 | $P_{cwlOleader}$*lytE [L213cat]* Δ*cwlO* Δ*lytE* | LU413 | $P_{lytE}$*lytE [L213cat]* Δ*lytE* Δ*cwlO* |
| | LU320 | $P_{cwlOleader}$*lytE [L123cat]*$^{R}$ Δ*cwlO* Δ*lytE* | LU343 | $P_{lytE}$*lytE [L123cat]*$^{R}$ Δ*lytE* Δ*cwlO* |

A representative growth curve for growth of each strain in LB is shown in Figure 5.16. All strains with one exception exhibited similar growth behaviour to WT168. However, growth of strain LU354 ($P_{lytE}$, LytE[L3cat]) was altered. Strain LU354 grew more slowly than WT168, reached stationary phase prematurely and plateaued quicker. The same growth inhibition for this strain was also observed in LPDM and HPDM media (data not shown). Strains were further tested for sensitivity to changes in pH (pH = 6 / 7 / 8.5), EtOH (4 % [vol / vol]), temperature (48°C) and salt (0.3 M NaCl).

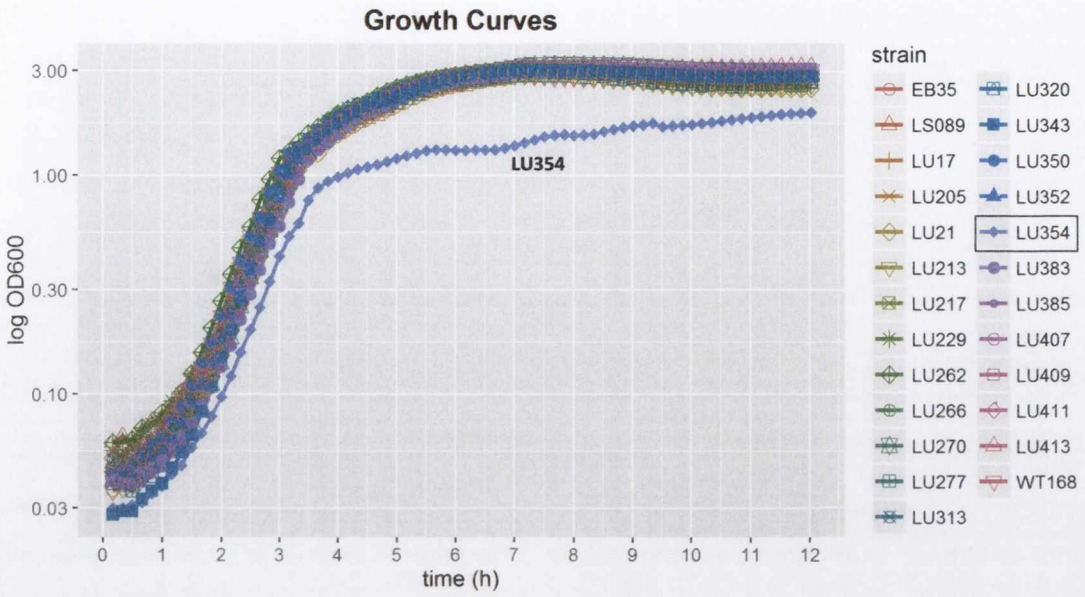



**Figure 5.16: Growth curves for strains that express a variant LytE protein.** Overnight cultures were used to inoculate 100 µl LB media in a CELLSTAR® 96 well plate to an $OD_{600}$ of 0.03 and incubated at 37°C with slow shaking, the $OD_{600}$ was measured every 10 minutes over a time period of 12 hours. Three technical repeats for each strain were conducted. One representative example is shown. The following strains were characterized: EB35(Δ*cwlO*), LS089 (Δ*lytE*), LU17 ($P_{cwlOleader}$*lytE* Δ*cwlO*), LU205 ($P_{cwlOleader}$*lytE*), *[L11cat]* Δ*cwlO* Δ*lytE*), LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*), LU213 ($P_{cwlOleader}$*lytE*), *[L13cat]* Δ*cwlO* Δ*lytE*), LU217 ($P_{cwlOleader}$*lytE*), *[L21cat]* Δ*cwlO* Δ*lytE*), LU229 ($P_{cwlOleader}$*lytE*), *[L31cat]* Δ*cwlO* Δ*lytE*),LU262 ($P_{cwlOleader}$*lytE*), *[L1cat]* Δ*cwlO* Δ*lytE*), LU266 ($P_{cwlOleader}$*lytE*), *[L2cat]* Δ*cwlO* Δ*lytE*), LU270 ($P_{cwlOleader}$*lytE*), *[L3cat]* Δ*cwlO* Δ*lytE*), LU277 ($P_{cwlOleader}$*lytE*), *[L331cat]* Δ*cwlO* Δ*lytE*), LU313 ($P_{cwlOleader}$*lytE*), *[L213cat]* Δ*cwlO* Δ*lytE*), LU320 ($P_{cwlOleader}$*lytE*), *[L123cat]*$^R$ Δ*cwlO* Δ*lytE*), LU343 ($P_{lytE}$*lytE [L123cat]*$^R$ Δ*lytE* Δ*cwlO*), LU350 ($P_{lytE}$*lytE [L1cat]* Δ*lytE* Δ*cwlO*), LU352 ($P_{lytE}$*lytE [L2cat]* Δ*lytE* Δ*cwlO*), **LU354 ($P_{lytE}$*lytE [L3cat]* Δ*lytE* Δ*cwlO*),** LU383 ($P_{lytE}$*lytE [L31cat]* Δ*lytE* Δ*cwlO*), LU385 ($P_{lytE}$*lytE [L331cat]* Δ*lytE* Δ*cwlO*), LU407 ($P_{lytE}$*lytE [L11cat]* Δ*lytE* Δ*cwlO*), LU409 ($P_{lytE}$*lytE [L21cat]* Δ*lytE* Δ*cwlO*), LU411 ($P_{lytE}$*lytE [L13cat]* Δ*lytE* Δ*cwlO*), LU413 ($P_{lytE}$*lytE [L213cat]* Δ*lytE* Δ*cwlO*) and WT168. Strain LU354 is highlighted. R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

None of the investigated strains showed an increased or decreased sensitivity to the first three conditions pH, EtOH and temperature (data not shown). However, some strains exhibited growth defects when exposed to salt stress (data not shown). This was investigated in more detail in a follow-up experiment described in section 5.4.11.

In summary it can be concluded that cell growth is not affected in strains that express a variant LytE. However, LysM motif L3 appears to complement to a lower extent leading to a less robust growth, but only when expressed from the $P_{lytE}$ (normal growth under $P_{cwlO}$). Furthermore, variant LytE are not affected in sensitivity to pH, EtOH and temperature. However, some variant LytE exhibited an increased sensitivity to salt, which was investigated in more detail in section 5.4.11.

### 5.4.9 Colony Morphology of strains containing a variant LytE

It was observed that colonies of the strain LU134 ($P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE*) that expresses LytF from the $P_{cwlO}$ promoter (in a genetic background with Δ*lytE* and Δ*cwlO* deleted) developed an aberrant morphology with small colonies appearing after prolonged incubation on a plate (this study). We posit that such small colonies arise from cells that have developed suppressor mutations and that their presence is an indication of a defect in complementation of the D,L-endopeptidase requirement. It was therefore decided to establish the colony morphology of strains that express variant LytE proteins using either the $P_{cwlO}$ or $P_{lytE}$ promoters. To address this question, colonies of strains were spotted onto LB agar plates (with no antibiotic) with a 96 replicator pin. The plates were then incubated for 12 hours at 37°C followed by incubation at room temperature. The change in growth and morphology of the colonies was then monitored after 1 day, 4 days and 7 days. The investigated strains are listed in Table 5.15 and strains WT168, LS089 (Δ*lytE*) and EB35 (Δ*cwlO*) were included as controls.

The observed colony morphologies are presented in Figure 5.17 (see page 154). All controls exhibited large and dry colonies of irregular form that differentiated over the course of the two weeks. The colony of strain EB35 (Δ*cwlO*) appeared to be slightly more round and less dry. Differences in the colony morphology of strains expressing variant LytE proteins with three LysM motifs, correlated with whether expression was induced by the *lytE* or by the *cwlO* promoter. For example the colony morphology of strains expressing variant LytE proteins LytE[L331cat] and LytE[L213cat] was similar to that of WT168 when the gene for the variant protein was expressed using the $P_{cwlO}$ promoter (although slightly translucent for LytE[L213cat]), but the colonies were smaller, translucent and exhibited lysis when the gene for the variant proteins was

expressed using the $P_{lytE}$ promoter. Furthermore, the colony morphology of strains LU320 and LU343 that express a variant LytE protein with the LytE[L123cat]$^{R}$ combination using the $P_{cwlO}$ or $P_{lytE}$ promoters displayed a morphology very similar to that of strain WT168. The colony morphology of strains that express variant LytE[L11cat] and LytE[L13cat] proteins with two LysM motifs using the $P_{cwlO}$ promoter *was* similar to that of WT168, but translucent colonies that developed suppressors were observed when the same two variant proteins were expressed using the $P_{lytE}$ promoter. However, small translucent colonies were observed for strains expressing proteins LytE[L21cat] and LytE[L31cat] using either the $P_{cwlO}$ or the $P_{lytE}$ promoters. The morphology of strains expressing variant LytE proteins with one LysM motif was one of smaller and rounder colonies than that of wild-type strain WT168. While a non-lysing type colony morphology was seen for strains that express the variant LytE[L1cat] using the $P_{cwlO}$ promoter, colonies lysed when this construct was expressed from the $P_{lytE}$ promoter. However, a lysing-type colony morphology was observed when LytE[L2cat] and LytE[L3cat] were expressed using either the $P_{cwlO}$ or $P_{lytE}$ promoters. Similarly a lysis-type colony morphology with small suppressor-type colonies was observed for strain JL19 that expresses the heterologous LytE[$L1^{XlyB}cat^{LytE}$] variant protein using the $P_{cwlO}$ promoter.

**Table 5.15: List of strains investigated in colony morphology assay (see Figure 5.17)**

| | variant *lytE* at *cwlO* locus | | variant *lytE* at *lytE* locus | |
|---|---|---|---|---|
| **LysM** | **Strain** | **Genotype** | **Strain** | **Genotype** |
| One LysM | LU262 | $P_{cwlOleader}$*lytE [L1cat] ΔcwlO ΔlytE* | LU350 | $P_{lytE}$*lytE [L1cat] ΔlytE ΔcwlO* |
| | LU266 | $P_{cwlOleader}$*lytE [L2cat] ΔcwlO ΔlytE* | LU352 | $P_{lytE}$*lytE [L2cat] ΔlytE ΔcwlO* |
| | LU270 | $P_{cwlOleader}$*lytE [L3cat] ΔcwlO ΔlytE* | LU354 | $P_{lytE}$*lytE [L3cat] ΔlytE ΔcwlO* |
| | JL15 | $P_{cwlOleader}$*lytE* [$L1^{xlyB}cat^{lytE}$] *ΔcwlO* | | |
| | JL19 | $P_{cwlOleader}$*lytE* [$L1^{xlyB}cat^{lytE}$] *ΔcwlO ΔlytE* | | |
| Two LysM | LU205 | $P_{cwlOleader}$*lytE [L11cat] ΔcwlO ΔlytE* | LU407 | $P_{lytE}$*lytE [L11cat] ΔlytE ΔcwlO* |
| | LU217 | $P_{cwlOleader}$*lytE [L21cat] ΔcwlO ΔlytE* | LU409 | $P_{lytE}$*lytE [L21cat] ΔlytE ΔcwlO* |
| | LU229 | $P_{cwlOleader}$*lytE [L31cat] ΔcwlO ΔlytE* | LU383 | $P_{lytE}$*lytE [L31cat] ΔlytE ΔcwlO* |
| | LU213 | $P_{cwlOleader}$*lytE [L13cat] ΔcwlO ΔlytE* | LU411 | $P_{lytE}$*lytE [L13cat] ΔlytE ΔcwlO* |
| Three LysM | LU277 | $P_{cwlOleader}$*lytE [L331cat] ΔcwlO ΔlytE* | LU385 | $P_{lytE}$*lytE [L331cat] ΔlytE ΔcwlO* |
| | LU313 | $P_{cwlOleader}$*lytE [L213cat] ΔcwlO ΔlytE* | LU413 | $P_{lytE}$*lytE [L213cat] ΔlytE ΔcwlO* |
| | LU320 | $P_{cwlOleader}$*lytE [L123cat]*$^{R}$ *ΔcwlO ΔlytE* | LU343 | $P_{lytE}$*lytE [L123cat]*$^{R}$ *ΔlytE ΔcwlO* |

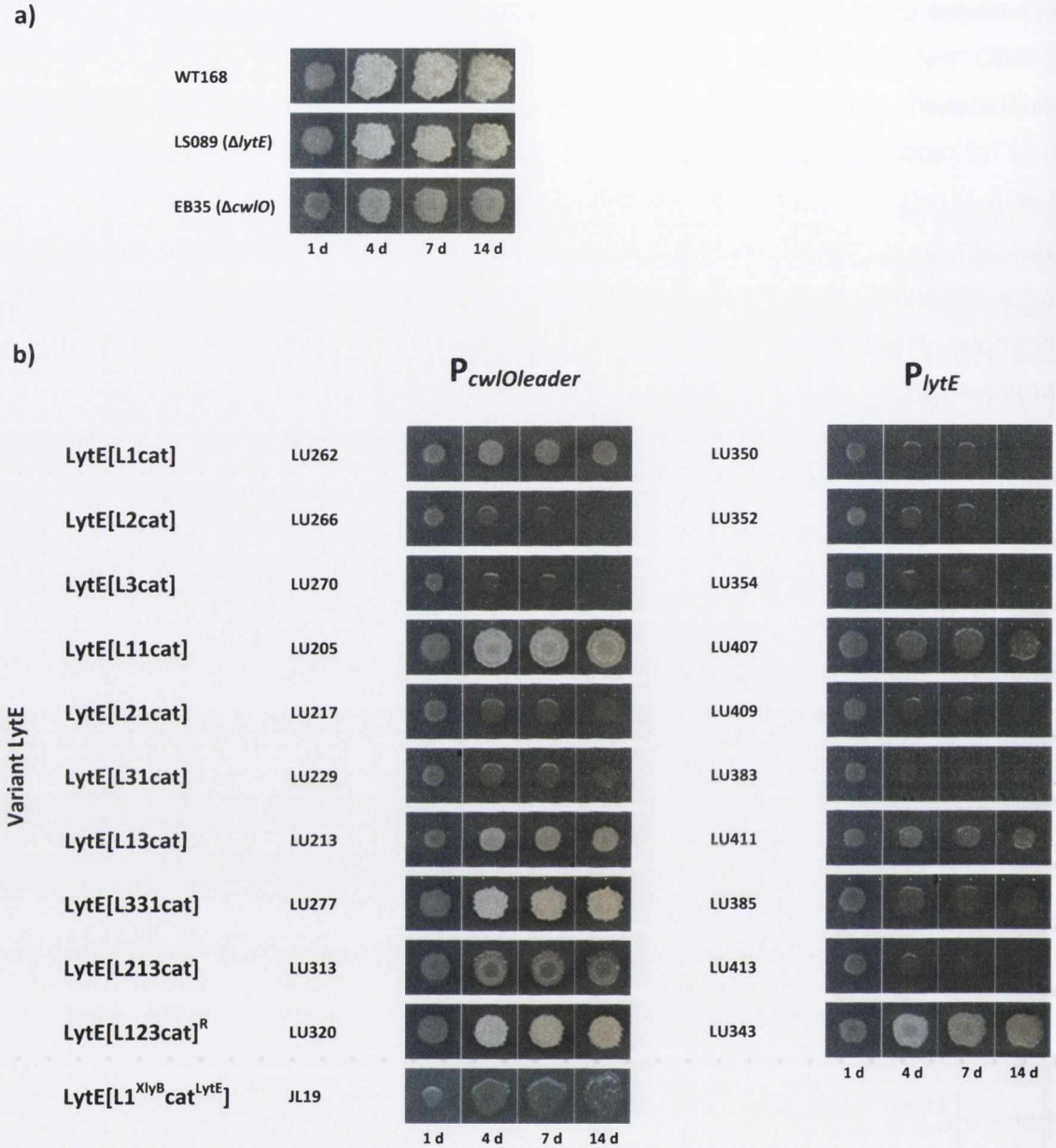



**Figure 5.17: Colony morphology of strains that express a variant LytE protein from the *cwlO* and the *lytE* promoter.** LB agar plates (no antibiotic) were inoculated with a toothpick from an overnight LB culture and incubated overnight at 37°C followed by incubation for 14 days at room temperature. Growth of each colony was documented after 1 day, 4 days, 7 days and 14 days. The following strains were characterized: **a)** controls: WT168, LS089 (Δ*lytE*), EB35 (Δ*cwlO*); **b)** strains with variant LytE: LU262 ($P_{cwlOleader}$*lytE [L1cat]* Δ*cwlO* Δ*lytE*), LU266 ($P_{cwlOleader}$*lytE [L2cat]* Δ*cwlO* Δ*lytE*), LU270 ($P_{cwlOleader}$*lytE [L3cat]* Δ*cwlO* Δ*lytE*), LU205 ($P_{cwlOleader}$*lytE [L11cat]* Δ*cwlO* Δ*lytE*), LU217 ($P_{cwlOleader}$*lytE [L21cat]* Δ*cwlO* Δ*lytE*), LU229 ($P_{cwlOleader}$*lytE [L31cat]* Δ*cwlO* Δ*lytE*), LU213 ($P_{cwlOleader}$*lytE [L13cat]* Δ*cwlO* Δ*lytE*), LU277 ($P_{cwlOleader}$*lytE [L331cat]* Δ*cwlO* Δ*lytE*), LU313 ($P_{cwlOleader}$*lytE [L213cat]* Δ*cwlO* Δ*lytE*), LU320 ($P_{cwlOleader}$*lytE [L123cat]*$^R$ Δ*cwlO* Δ*lytE*), LU350 ($P_{lytE}$*lytE [L1cat]* Δ*lytE* Δ*cwlO*), LU352 ($P_{lytE}$*lytE [L2cat]* Δ*lytE* Δ*cwlO*), LU354 ($P_{lytE}$*lytE [L3cat]* Δ*lytE* Δ*cwlO*), LU407 ($P_{lytE}$*lytE [L11cat]* Δ*lytE* Δ*cwlO*), LU409 ($P_{lytE}$*lytE [L21cat]* Δ*lytE* Δ*cwlO*), LU383 ($P_{lytE}$*lytE [L31cat]* Δ*lytE* Δ*cwlO*), LU411 ($P_{lytE}$*lytE [L13cat]* Δ*lytE* Δ*cwlO*), LU385 ($P_{lytE}$*lytE [L3311cat]* Δ*lytE* Δ*cwlO*), LU413 ($P_{lytE}$*lytE [L213cat]* Δ*lytE* Δ*cwlO*), LU343 ($P_{lytE}$*lytE [L123cat]*$^R$ Δ*lytE* Δ*cwlO*) and JL19 ($P_{cwlOleader}$*lytE [L1*$^{XlyB}$*cat*$^{lytE}$*]* Δ*cwlO* Δ*lytE*). R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

In summary strains expressing variant LytE proteins with three LysM motifs produced bigger, healthier colonies than those that expressed variant LytE proteins with one or two LysM motifs. In general colonies were stronger, bigger and exhibited less lysis regardless of the variant LytE protein when they were expressed using the stronger $P_{cwlO}$ rather than the weaker $P_{lytE}$ promoter with two exceptions: LytE[L2cat] and LytE[L3cat] produced equally small lysing colonies when expressed using either promoters.

### 5.4.10 Cell Morphology of strains containing a variant LytE

It was shown that manipulating D,L-endopeptidase regulation can affect cell morphology. For example, cells with *lytE* deleted are frequently bent (Bisicchia *et al.*, 2007) and cells of strain LU25 ($P_{cwlO}$*lytE* Δ*cwlO* Δ*lytE*) frequently exhibit bulged cells (see Figure 3.8, page 80), and a strain LU134 ($P_{cwlO}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE*) that expresses LytF from the *cwlO* promoter featured cells with excrescences (see Figure 4.7, page 99). Therefore, we decided to establish if cells of strains that express a variant LytE protein using either the $P_{cwlO}$ or $P_{lytE}$ have altered cell morphology. Microscopy was performed on the strains listed in Table 5.16.

**Table 5.16: List of strains investigated by microscopy (see Figure 5.18)**

| **variant *lytE* at *cwlO* locus** | | | |
|---|---|---|---|
| **With intact endogenous *lytE* gene** | | **With disrupted endogenous *lytE* gene** | |
| **Strain** | **Genotype** | **Strain** | **Genotype** |
| **LU265** | $P_{cwlOleader}$*lytE [L2cat]* Δ*cwlO* | **LU266** | $P_{cwlOleader}$*lytE [L2cat]* Δ*cwlO* Δ*lytE* |
| **LU269** | $P_{cwlOleader}$*lytE [L3cat]* Δ*cwlO* | **LU270** | $P_{cwlOleader}$*lytE [L3cat]* Δ*cwlO* Δ*lytE* |
| **LU220** | $P_{cwlOleader}$*lytE [L22cat]* Δ*cwlO* | **LU222** | $P_{cwlOleader}$*lytE [L22cat]* Δ*cwlO* Δ*lytE* |
| **LU318** | $P_{cwlOleader}$*lytE [L123cat]*$^{R}$ Δ*cwlO* | **LU320** | $P_{cwlOleader}$*lytE [L123cat]*$^{R}$ Δ*cwlO* Δ*lytE* |
| **JL15** | $P_{cwlOleader}$*lytE [L1*$^{xlyB}$*cat*$^{lytE}$*]* Δ*cwlO* | **JL19** | $P_{cwlOleader}$*lytE [L1*$^{xlyB}$*cat*$^{lytE}$*]* Δ*cwlO* Δ*lytE* |
| **variant *lytE* at *lytE* locus** | | | |
| **With intact *cwlO* gene** | | **With disrupted *cwlO* gene** | |
| **Strain** | **Genotype** | **Strain** | **Genotype** |
| **LU337** | $P_{lytE}$*lytE [L1cat]* Δ*lytE* | **LU350** | $P_{lytE}$*lytE [L1cat]* Δ*lytE* Δ*cwlO* |
| **LU339** | $P_{lytE}$*lytE [L2cat]* Δ*lytE* | **LU352** | $P_{lytE}$*lytE [L2cat]* Δ*lytE* Δ*cwlO* |
| **LU341** | $P_{lytE}$*lytE [L3cat]* Δ*lytE* | **LU354** | $P_{lytE}$*lytE [L3cat]* Δ*lytE* Δ*cwlO* |
| **LU335** | $P_{lytE}$*lytE [L123cat]*$^{R}$ Δ*lytE* | **LU343** | $P_{lytE}$*lytE [L123cat]*$^{R}$ Δ*lytE* Δ*cwlO* |

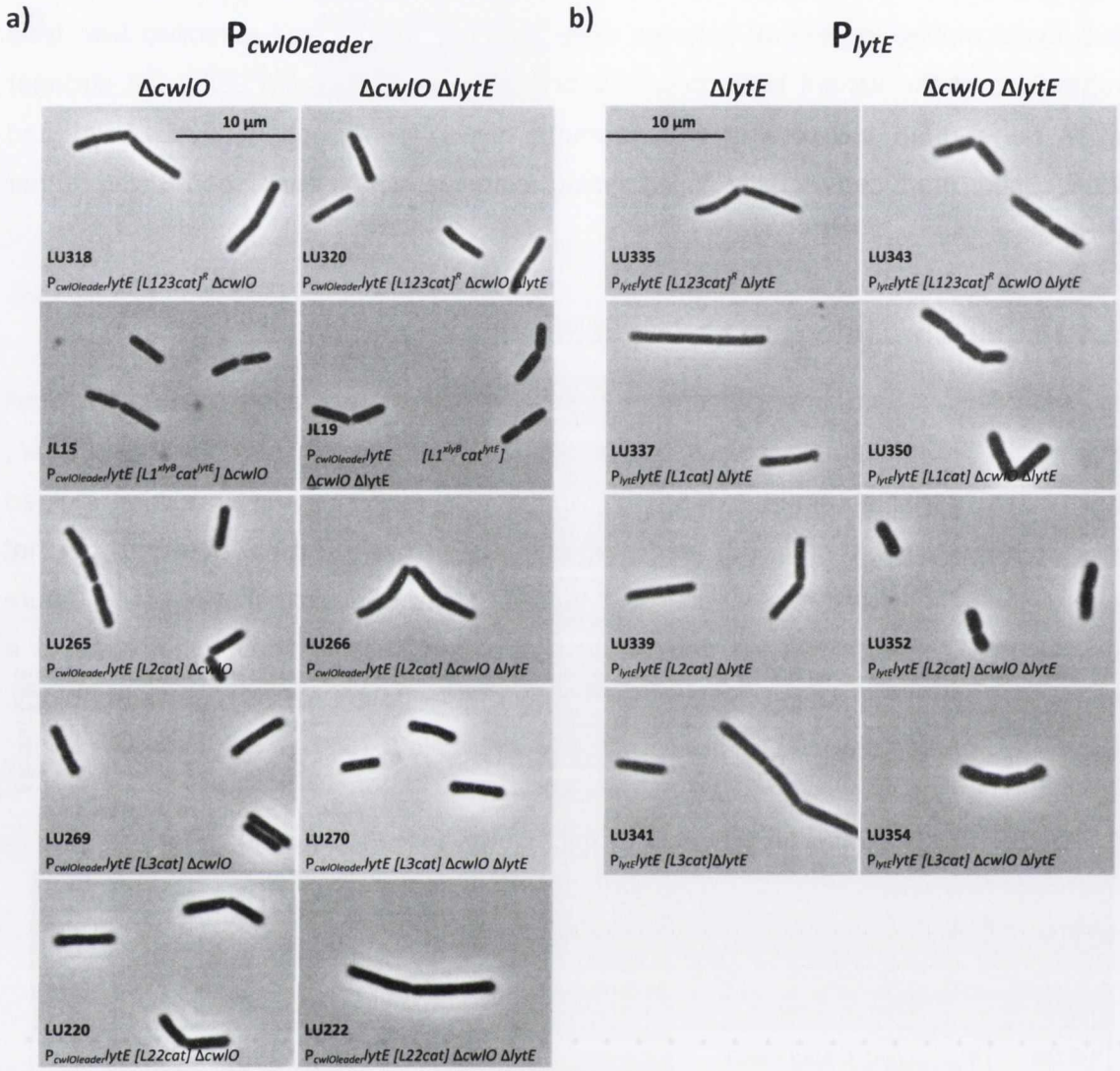



**Figure 5.18: Typical cell morphology of strains that express a variant LytE a) from the *cwlO* promoter (in combination with Δ*lytE*) and b) from the *lytE* promoter (in combination with Δ*cwlO*).** Cells were grown in LB to an $OD_{600}$ of 0.1, heat-fixed and examined with an Olympus BX61 microscope. Magnification is 1000x. The genotypes of the examined strains as well as a scale bar are shown. R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

Typical cell morphologies of the investigated strains are presented in Figure 5.18. The cellular morphology of only three of all the strains tested differed consistently from that of cells of wild-type strain WT168. The three strains LU350 ($P_{lytE}$*lytE[L1cat]* Δ*lytE* Δ*cwlO*), LU352 ($P_{lytE}$*lytE[L2cat]* Δ*lytE* Δ*cwlO*) and LU354 ($P_{lytE}$*lytE [L3cat]* Δ*lytE* Δ*cwlO*) that displayed a distinct cell morphology have three distinguishing characteristics: (i) these strains express a variant LytE protein with a single LysM motif, (ii) all three variant proteins were expressed using the (weaker) $P_{lytE}$ and (iii) the *cwlO* gene was deleted such that the D,L-endopeptidase requirement can be fulfilled only by this variant LytE protein. Cells of these strains were shorter and thicker than WT168 cells (right hand panels of Figure 5.18). It is important to note that cells of strains that express a variant LytE protein with these single LysM motifs using the *lytE* promoter, but in which the endogenous *cwlO* gene is intact i.e. LU337 ($P_{lytE}$*lytE [L1cat]* Δ*lytE*), LU339 ($P_{lytE}$*lytE [L2cat]* Δ*lytE*), LU341 ($P_{lytE}$*lytE [L3cat]* Δ*lytE*) did not show aberrant cell morphology. Furthermore, cells of strains LU266 ($P_{cwlOleader}$*lytE [L2cat]* Δ*cwlO* Δ*lytE*) and LU270 ($P_{cwlOleader}$*lytE [L3cat]* Δ*cwlO* Δ*lytE*) that express a variant LytE protein with a single LysM motif using the (stronger) $P_{cwlO}$ promoter, but in a background with *lytE* deleted, also have no apparent morphological defects. It is concluded from these data that cells that express a variant LytE protein with any one of the three LytE LysM motifs can fulfil the D,L-endopeptidase requirement. However, our data suggest that the level of protein expression is important: cells that express the variant LytE[L1cat], LytE[L2cat] or LytE[L3] using the stronger $P_{cwlO}$ promoter have normal cell morphology while cells that express them from the weaker $P_{lytE}$ promoter are shorter and fatter than are wild-type cells.

### 5.4.11 Sensitivity to salt stress of strains containing a variant LytE

#### 5.4.11.1 Growth behaviour of salt-stressed cells containing a variant LytE

It was observed that some strains that express variant LytE proteins, had increased sensitivity to salt stress (this study). To further explore this observation a selection of strains (listed in Table 5.17) and their appropriate controls were grown overnight in salt-free buffered LB. These overnight cultures were used to inoculate buffered LB containing NaCl at the following concentrations: 0.05 M, 0.1 M, 0.2 M, 0.3 M and 0.4 M NaCl). Cell growth was monitored and resulting growth profiles were compared to those observed during growth in LB with no added NaCl. The following strains were included as controls: WT168, LS089 (Δ*lytE*), EB35 (Δ*cwlO*) and LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*).

**Table 5.17: List of strains investigated by growth inhibition assay with NaCl (see Figure 5.19)**

| | variant *lytE* at *cwlO* locus | | variant *lytE* at *lytE* locus | |
|---|---|---|---|---|
| **LysM** | **Strain** | **Genotype** | **Strain** | **Genotype** |
| One LysM | LU262 | $P_{cwlOleader}$*lytE [L1cat] ΔcwlO ΔlytE* | LU350 | $P_{lytE}$*lytE [L1cat] ΔlytE ΔcwlO* |
| | LU266 | $P_{cwlOleader}$*lytE [L2cat] ΔcwlO ΔlytE* | LU352 | $P_{lytE}$*lytE [L2cat] ΔlytE ΔcwlO* |
| | LU270 | $P_{cwlOleader}$*lytE [L3cat] ΔcwlO ΔlytE* | LU354 | $P_{lytE}$*lytE [L3cat] ΔlytE ΔcwlO* |
| Two LysM | LU205 | $P_{cwlOleader}$*lytE [L11cat] ΔcwlO ΔlytE* | LU407 | $P_{lytE}$*lytE [L11cat] ΔlytE ΔcwlO* |
| | LU217 | $P_{cwlOleader}$*lytE [L21cat] ΔcwlO ΔlytE* | LU409 | $P_{lytE}$*lytE [L21cat] ΔlytE ΔcwlO* |
| | LU229 | $P_{cwlOleader}$*lytE [L31cat] ΔcwlO ΔlytE* | LU383 | $P_{lytE}$*lytE [L31cat] ΔlytE ΔcwlO* |
| | LU213 | $P_{cwlOleader}$*lytE [L13cat] ΔcwlO ΔlytE* | LU411 | $P_{lytE}$*lytE [L13cat] ΔlytE ΔcwlO* |
| Three LysM | LU277 | $P_{cwlOleader}$*lytE [L331cat] ΔcwlO ΔlytE* | LU385 | $P_{lytE}$*lytE [L331cat] ΔlytE ΔcwlO* |
| | LU313 | $P_{cwlOleader}$*lytE [L213cat] ΔcwlO ΔlytE* | LU413 | $P_{lytE}$*lytE [L213cat] ΔlytE ΔcwlO* |
| | LU320 | $P_{cwlOleader}$*lytE [L123cat]*$^{R}$ *ΔcwlO ΔlytE* | LU343 | $P_{lytE}$*lytE [L123cat]*$^{R}$ *ΔlytE ΔcwlO* |

The result of these growth inhibition assays is presented in Figure 5.19 (page 160). All control strains: WT168, a Δ*lytE* mutant (LS089) and a Δ*cwlO* mutant (EB35) grew normally at NaCl concentrations up to 0.4 M. This observation is in agreement with Sévin *et al.* (2016), who reported that *B. subtilis* can grow at a concentration of 0.4 M NaCl. This normal growth profile was also observed for strain LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*). All variant LytE strains grew normally under the absence of salt or at a salt concentration of 0.1 M as was observed earlier. However, growth inhibitions were observed at a salt concentration of 0.2 M and higher.

The extent to which salt stress caused a change in growth behaviour correlated with three factors: (i) the total number of LysM motifs in the variant LytE, (ii) the type of LysM motifs and (iii) the expression level (*cwlO* vs *lytE* promoter).

*Variant LytE with three LysM motifs:* A normal growth profile was observed for strains expressing variant LytE proteins with three LysM motifs. For example, strains expressing variant LytE[L123cat]$^{R}$ and LytE[L331cat] grew normally at salt concentrations up to 0.4 M. However, strains expressing the variant LytE[L213cat] grew normally when the protein was expressed using the *cwlO* promoter, but minor growth defects were observed in 0.4 M salt when expressed using the *lytE* promoter.

*Variant LytE with two LysM motifs:* In general growth of strains expressing variant LytE proteins with two LysM motifs was normal up to 0.4 M NaCl. However, the specific combination of LysM motif was an important feature in these growth profiles.

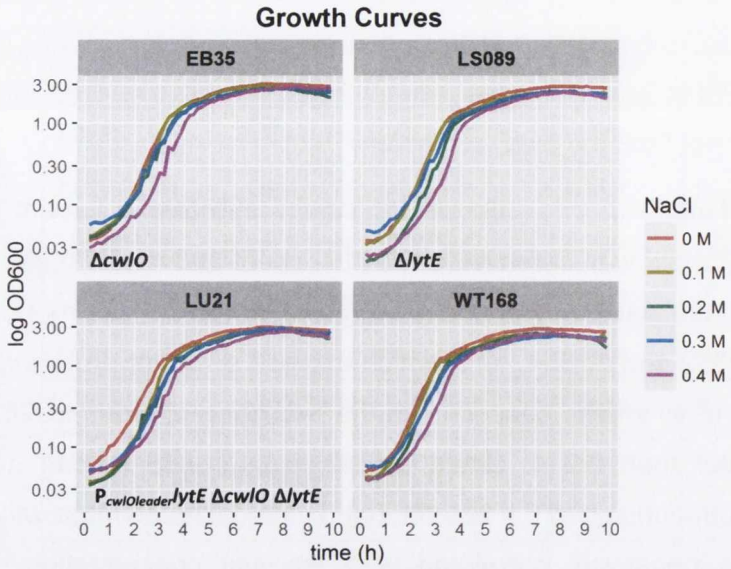
Growth was normal up to 0.4 M NaCl for strains expressing the variant LytE[L11cat] and LytE[L13cat]. A prolonged lag phase was observed for strains expressing LytE[L21cat] and LytE[L31cat] when grown at NaCl concentrations greater than 0.2 M. However, this defect was observed when the variant LytE proteins were expressed from the *lytE* promoter, but not from the *cwlO* promoter.

*Variant LytE with one LysM motif:* Growth of strains expressing variant LytE proteins with a single LysM motif was defective at NaCl concentrations of ≥ 0.2M. The onset of the exponential phase was delayed and growth was slower for strains LU262, LU266, LU270, LU350, LU352 and LU354, all of which express a variant protein with a single LysM motif. Growth of all strains ceased (with the exception of strain LU262) at NaCl concentrations greater than 0.2 M. Strain LU262 ($P_{cwlOleader}$*lytE* [*L1cat*] Δ*cwlO* Δ*lytE*) grew at NaCl concentrations up to 0.4 M, although some decrease in growth rate was observed which was negatively correlated with the salt concentration. It is important to note that strains that express a variant LytE proteins with a single LysM motif, but in which the endogenous *cwlO* or *lytE* gene was present, did not show any growth defect at 0.4 M NaCl (data not shown). Thus, the observed defects were entirely due to the properties of the variant LytE protein (i.e. either the level of expression ($P_{cwlO}$ or $P_{lytE}$) or the number or combination of LysM motifs).

In summary, it can be conclude from this data that the total number of LysM motifs in a variant LytE, has the greatest impact on the resistance to salt stress, of a strain containing a variant LytE protein. Strains with a variant LytE protein with only one LysM motif, showed the greatest sensitivity to salt stress, while increasing the number of LysM motifs had a positive impact on resistance to salt stress. Furthermore, the type of permutation mattered. Strains that possess a variant LytE that contained L1 at the first position had a greater resistance to salt stress than strains in which the variant LytE did not contain L1, or in which it was not localized adjacent to the signal peptide.

Since sensitivity to salt stress was not altered in strains that in addition to the variant LytE protein also had an intact endogenous *lytE* or *cwlO* gene, it can be deduced that it is not the presence of the variant LytE that caused the increased sensitivity to salt stress, but rather a lack of D,L-endopeptidase enzymatic activity. The lack of D,L-endopeptidase activity is most likely caused by the decreased binding capacity of a variant LytE protein with fewer LysM motif, as Mesnage *et al.* (2014) and Wong *et al.* (2014) have shown that LysM motifs contribute additively to binding *in vitro*.

**a) Growth of control strains**




**b) Growth of strains expressing variant LytE proteins with a single LysM motif**

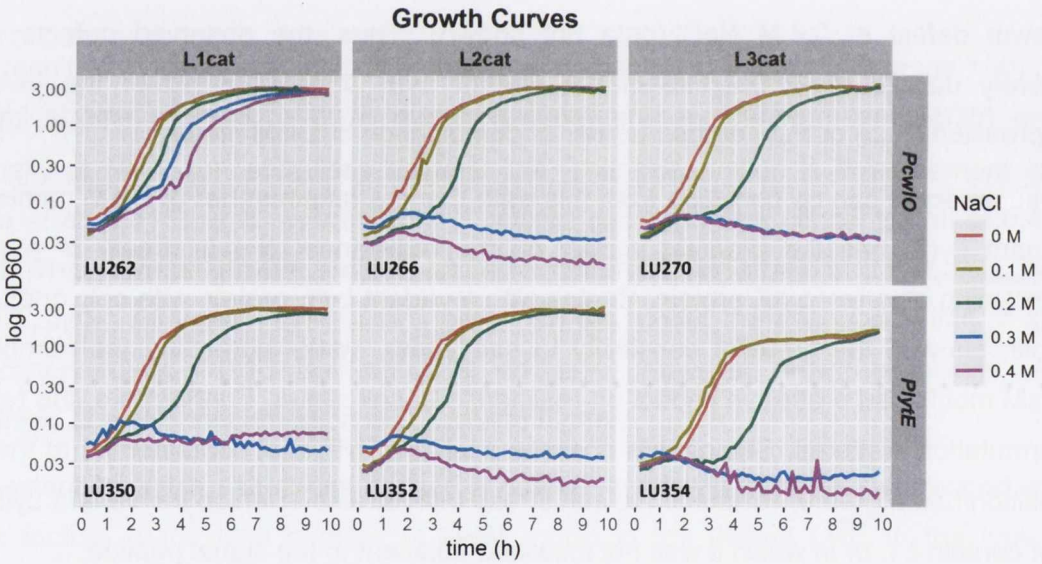
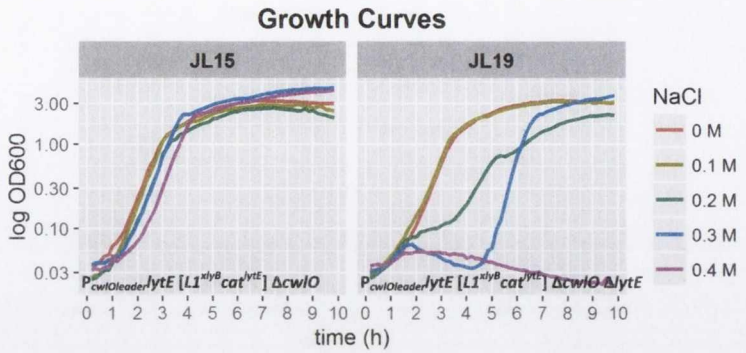



**Figure 5.19 [Figure continues on next pages]: Growth curves of strains that express a variant LytE protein from the *cwlO* and the *lytE* promoter.** Growth curves were obtained in the following way: Overnight cultures of the strains were grown in salt-free buffered LB and used to inoculate 100 µl media in a CELLSTAR® 96 well plate to an $OD_{600}$ of 0.03. NaCl was added as desired to a final concentration of 0.1 M, 0.2 M, 0.3 M and 0.4 M. The 96 well plate was incubated at 37°C with slow shaking and the $OD_{600}$ was measured every 10 minutes over a time period of 10 hours. Three technical repeats for each strain were conducted. One representative example is shown. Growth curves were obtained for the following strains: **a)** WT168, LS089 (Δ*lytE*), EB35 (Δ*cwlO*), LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*); **b)** LU262 ($P_{cwlOleader}$*lytE [L1cat]* Δ*cwlO* Δ*lytE*), LU266 ($P_{cwlOleader}$*lytE [L2cat]* Δ*cwlO* Δ*lytE*), LU270 ($P_{cwlOleader}$*lytE [L3cat]* Δ*cwlO* Δ*lytE*), LU350 ($P_{lytE}$*lytE [L1cat]* Δ*cwlO* Δ*lytE*), LU352 ($P_{lytE}$*lytE [L2cat]* Δ*cwlO* Δ*lytE*), LU354 ($P_{lytE}$*lytE [L3cat]* Δ*cwlO* Δ*lytE*);

**c) Growth of strains expressing variant LytE proteins with a single heterologous LysM motif**




**d) Growth of strains expressing variant LytE proteins with two LysM motifs**

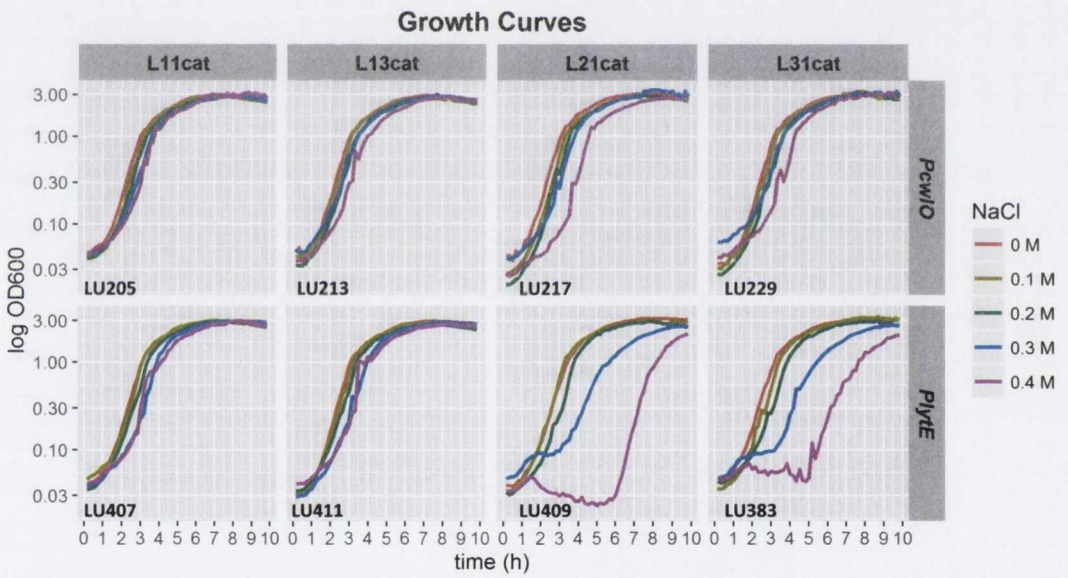



**Figure 5.19 (continued).** Growth curves were obtained for the following strains: **c)** JL15 ($P_{cwlOleader}$*lytE [L1*$^{xlyB}$*cat*$^{lytE}$*]* Δ*cwlO*), JL19 ($P_{cwlOleader}$*lytE [L1*$^{xlyB}$*cat*$^{lytE}$*]* Δ*cwlO* Δ*lytE*); Growth curves for JL15 and JL19 were established by Jennifer Lennon under my supervision. **d)** LU205 ($P_{cwlOleader}$*lytE [L11cat]* Δ*cwlO* Δ*lytE*), LU213 ($P_{cwlOleader}$*lytE [L13cat]* Δ*cwlO* Δ*lytE*), LU217 ($P_{cwlOleader}$*lytE [L21cat]* Δ*cwlO* Δ*lytE*), LU229 ($P_{cwlOleader}$*lytE [L31cat]* Δ*cwlO* Δ*lytE*), LU407 ($P_{lytE}$*lytE [L11cat]* Δ*cwlO* Δ*lytE*), LU411 ($P_{lytE}$*lytE [L13cat]* Δ*cwlO* Δ*lytE*), LU409 ($P_{lytE}$*lytE [L21cat]* Δ*cwlO* Δ*lytE*), LU383 ($P_{lytE}$*lytE [L31cat]* Δ*cwlO* Δ*lytE*);

**e) Growth of strains expressing variant LytE proteins with three LysM motifs**

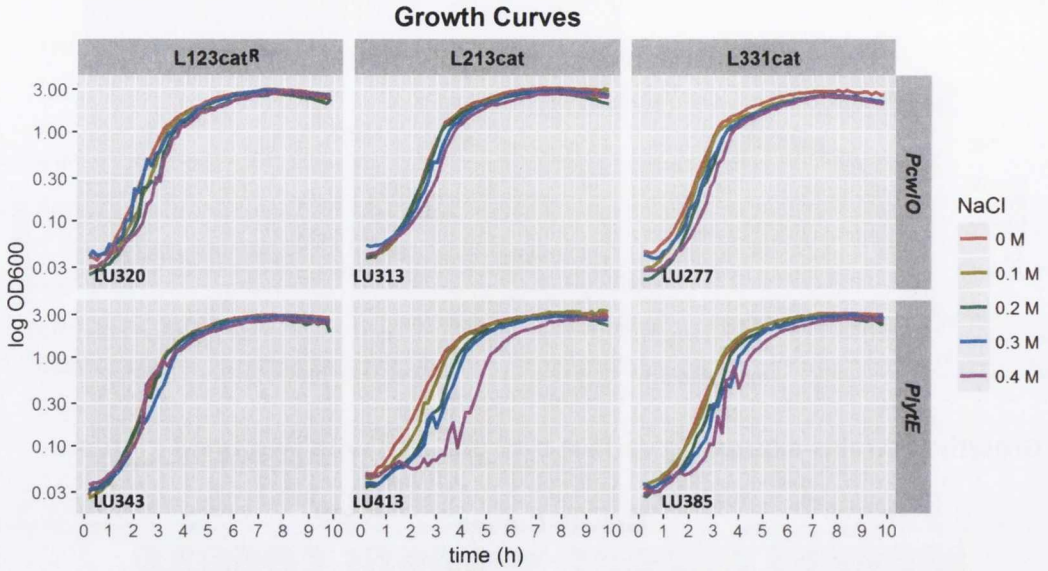



**Figure 5.19 (continued).** Growth curves were obtained for the following strains: **e)** LU320 ($P_{cwlOleader}$*lytE [L123cat]*$^R$ Δ*cwlO* Δ*lytE*), LU313 ($P_{cwlOleader}$*lytE [L213cat]* Δ*cwlO* Δ*lytE*), LU277 ($P_{cwlOleader}$*lytE [L331cat]* Δ*cwlO* Δ*lytE*), LU343 ($P_{lytE}$*lytE [L123cat]*$^R$ Δ*cwlO* Δ*lytE*), LU413 ($P_{lytE}$*lytE [L213cat]* Δ*cwlO* Δ*lytE*), LU385 ($P_{lytE}$*lytE [L331cat]* Δ*cwlO* Δ*lytE*). R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

### 5.4.11.2 Cell morphology of salt-stressed cells containing a variant LytE

In view of the observation that some strains expressing variant LytE proteins with different numbers and combinations of LysM motifs displayed sensitivity to elevated levels of NaCl (Figure 5.19, page 160), we sought to examine the morphology of selected strains under such salt stress conditions. The strains selected for investigation are shown in Table 5.18 and include control strains WT168, EB35 (Δ*cwlO*) and LS089 (Δ*lytE*). Strains were grown in buffered LB either with or without 0.2 M NaCl to an $OD_{600}$ of 0.1 and analysed by microscopy.

**Table 5.18: List of strains investigated by microscopy (see Figure 5.20)**

| With intact endogenous *lytE* gene | | With disrupted endogenous *lytE* gene | |
|---|---|---|---|
| **Strain** | **Genotype** | **Strain** | **Genotype** |
| **LU207** | $P_{cwlOleader}$*lytE [L12cat]* Δ*cwlO* | **LU209** | $P_{cwlOleader}$*lytE [L12cat]* Δ*cwlO* Δ*lytE* |
| **LU261** | $P_{cwlOleader}$*lytE [L1cat]* Δ*cwlO* | **LU262** | $P_{cwlOleader}$*lytE [L1cat]* Δ*cwlO* Δ*lytE* |
| **LU265** | $P_{cwlOleader}$*lytE [L2cat]* Δ*cwlO* | **LU266** | $P_{cwlOleader}$*lytE [L2cat]* Δ*cwlO* Δ*lytE* |
| **LU269** | $P_{cwlOleader}$*lytE [L3cat]* Δ*cwlO* | **LU270** | $P_{cwlOleader}$*lytE [L3cat]* Δ*cwlO* Δ*lytE* |
| **LU318** | $P_{cwlOleader}$*lytE [L123cat]*$^R$ Δ*cwlO* | **LU320** | $P_{cwlOleader}$*lytE [L123cat]*$^R$ Δ*cwlO* Δ*lytE* |

Examples of the characteristic cell morphology of each strain under both conditions are shown in Figure 5.20. It was found that with the exception of wild-type strain WT168, cells of all strains were slightly thicker when cells were exposed to salt stress. However, the extent of this effect varied according to the LysM motif composition of the variant LytE protein expressed.

The effect was marginal in LS089 (Δ*lytE*) and in the variant LytE strains LU318 and LU320 that both expressed the variant LytE[L123cat]$^R$ from the *cwlO* promoter (with or without the endogenous *lytE* gene). In comparison the variant LytE strains LU207 and LU209, (both LytE[L12cat]) featured distinctively thicker and shorter cells when exposed to salt stress. The greatest change in cell morphology was observed in strains expressing a variant LytE protein with one LysM motif. Among these strains, further distinctions could be found between the type of LysM motif (L1, L2 or L3) and between strains possessing or lacking the endogenous *lytE* gene. Strain LU261 and LU262 (both LytE[L1cat]) showed thicker cells when grown with 0.2 M NaCl, but the effect was more pronounced for strain LU262 (Δ*lytE*), which also showed much shorter cells.

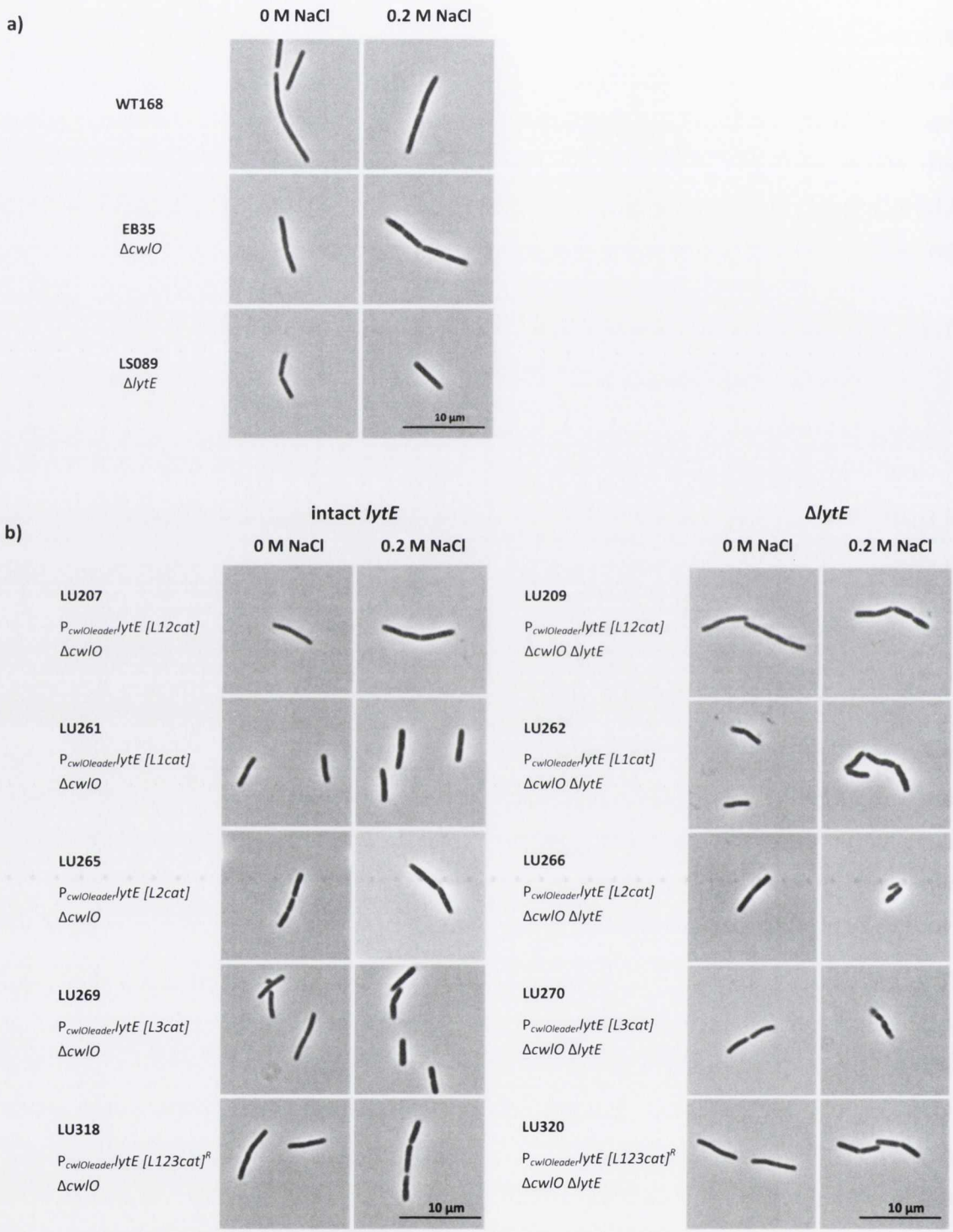



**Figure 5.20: Cell morphology of strains that express a variant LytE protein when exposed to salt stress.** Two separate media (buffered LB and buffered LB with 0.2 M NaCl) were inoculated from the same overnight culture and the culture grown to an $OD_{600}$ of 0.1 at which point cells were harvested, heat-fixed and examined with an Olympus BX61 microscope. Magnification is 1000x. The genotypes of the examined strains as well as a scale bar are shown. R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

Strains LU265 and LU266 (both LytE[L2cat]) and strains LU269 and LU270 (both LytE[L3cat]) exhibited thicker cells when grown with 0.2 M NaCl, an effect greatly amplified upon disruption of the endogenous *lytE* gene (LU266, LU270). Both LU266 and LU270 exhibited very small, almost round cells that showed strong signs of lysis.

In summary, this experiment established that salt stress affects cell morphology. Most strains, when grown at 0.2 M NaCl exhibited shorter and thicker cells. This effect was especially evident in strains in which the D,L-endopeptidase requirement was fulfilled by a variant LytE protein with a single LysM motif. Low expression levels (expression by the *lytE* promoter) further impaired cell morphology, indicating that the cause is a lack of D,L-endopeptidase activity. Interestingly the type of LysM motif of the variant LytE protein is also important. A strain expressing a variant LytE protein with the combination LytE[L1cat] from the *lytE* promoter showed a less impaired cell morphology than strains with the variant LytE combination LytE[L2cat] or LytE[L3cat].

### 5.4.12 Localization of variant LytE proteins under normal conditions and in salt stressed cells

It was found that the LytE autolysin is secreted and binds to cell septa, poles and the lateral sidewall of the cell, a location that is directed by the actin-like protein MreBH (Carballido-Lopez *et al.*, 2006; Hashimoto *et al.*, 2012). Thus, we sought to establish the cellular localization of variant LytE proteins of a selection of strains (listed in Table 5.19) during normal growth and in salt-stressed cells. Variant LytE proteins were localized by immunofluorescence microscopy in cells grown in buffered LB medium without salt and with 0.2 M NaCl as described in Chapter 2.

The results (presented in Figure 5.21) show that in strain LU421 (LytE[L123cat]$^{R}$) the LytE protein is present at a very high level and is distributed along the entire length of the cell. It is also evident that expressing LytE from the $P_{cwlO}$ promoter results in a very significantly higher level of protein being cell-associated than is observed in cells expressing LytE from the normal $P_{lytE}$ promoter (compare strains LU421 and LU45).

A similar localization pattern was observed with the other tested variant LytE. However, it is clear that the level of cell-associated protein correlates with the number of LysM motifs in the LytE protein: the amount of labelling is highest with 3 LysM motifs (strain LU421, LytE[L123cat]$^{R}$), lower with 2 LysM motifs (strain LU415, LytE[L13cat]) and very low with only 1 LysM motif (strain LU418 and LU440, LytE[L1cat] and LytE[L3cat]).

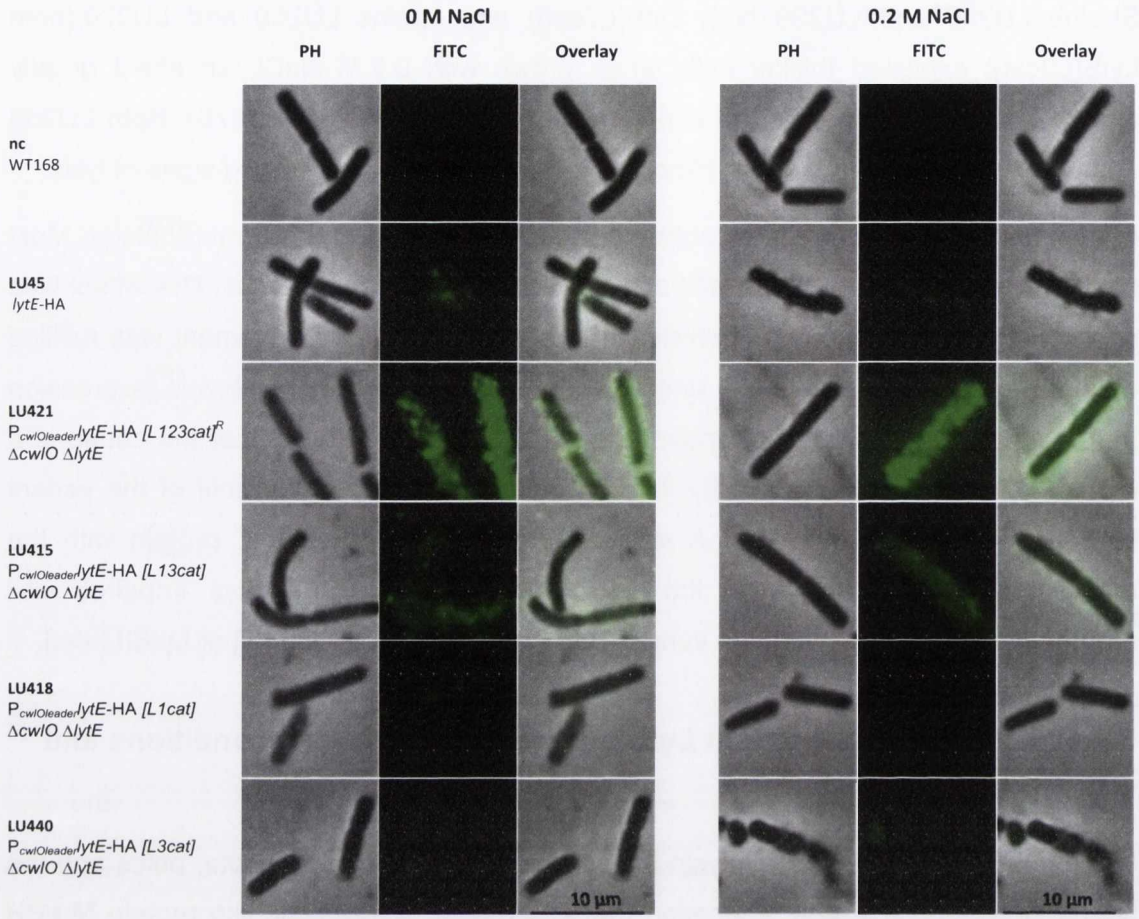



**Figure 5.21: LytE localization in strains that express a variant LytE protein when exposed to salt stress.** Two separate media (buffered LB and buffered LB with 0.2 M NaCl) were inoculated from the same overnight culture and the culture grown to an $OD_{600}$ of 0.2 at which point cells were harvested, fixed in 2.6 % paraformaldehyde, immunostained and examined with an Olympus BX61 microscope. Immunostaining was achieved by endowing the respective variant LytE protein with a HA-tag and treating cells first with a primary anti-HA antibody, followed by binding of the primary antibody with a fluorescein isothiocyanate (FITC)-labelled secondary antibody. Proteins were visualized by phase contrast microscopy (PH) and by immunofluorescence microscopy (FITC). Magnification is 1000x. A negative control (nc) is shown (WT168, no HA-tag). The genotypes of the examined strains as well as a scale bar are shown. R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

**Table 5.19: List of strains investigated by immunofluorescence microscopy (see Figure 5.21)**

| LysM motif | Strain | Genotype |
|---|---|---|
| One LysM | LU418 | $P_{cwlOleader}$*lytE-HA [L1cat] ΔcwlO ΔlytE* |
| | LU440 | $P_{cwlOleader}$*lytE-HA [L3cat] ΔcwlO ΔlytE* |
| Two LysM | LU415 | $P_{cwlOleader}$*lytE-HA [L13cat] ΔcwlO ΔlytE* |
| Three LysM | LU421 | $P_{cwlOleader}$*lytE-HA [L123catf]*$^{R}$ *ΔcwlO ΔlytE* |
| Control | WT168 | |
| | LU421 | *lytE-HA* |

It is also evident that that different LysM motifs have different binding capabilities – a higher level of binding is observed in cells of strain LU418 (expressing LytE[L1cat] than in strain LU440 (expressing LytE[L3cat]). However, exposure to salt stress did not lead to detectable differences in the binding of variant LytE proteins to the cell wall.

In conclusion, this study establishes that number and identity of LysM motifs greatly affects binding of variant LytE proteins to the cell wall. However, the pattern of their localization along the cell cylinder is not affected.

### 5.4.13 Antibiotic sensitivity of strains containing a variant LytE

The results presented show that strains expressing a variant LytE have different levels of cell-associated protein (Figure 5.13 – page 142, Figure 5.21 – page 166), grow differently (Figure 5.16, page 151), have different colony morphologies (Figure 5.17, page 154) and show some difference in growth inhibition when exposed to salt stress (Figure 5.19 – page 160, Figure 5.20 – page 164). Such differences probably correlate with differences in cell wall structure and metabolism. This suggests that cells may show differences in their susceptibility to cell wall acting antibiotics.

**Table 5.20: List of strains investigated by antibiotic sensitivity assay (see Figure 5.22)**

| With intact endogenous *lytE* gene | | With disrupted endogenous *lytE* gene | |
|---|---|---|---|
| **Strain** | **Genotype** | **Strain** | **Genotype** |
| **LU261** | $P_{cwlOleader}$*lytE [L1cat] ΔcwlO* | **LU262** | $P_{cwlOleader}$*lytE [L1cat] ΔcwlO ΔlytE* |
| **LU265** | $P_{cwlOleader}$*lytE [L2cat] ΔcwlO* | **LU266** | $P_{cwlOleader}$*lytE [L2cat] ΔcwlO ΔlytE* |
| **LU269** | $P_{cwlOleader}$*lytE [L3cat] ΔcwlO* | **LU270** | $P_{cwlOleader}$*lytE [L3cat] ΔcwlO ΔlytE* |
| **LU207** | $P_{cwlOleader}$*lytE [L12cat] ΔcwlO* | **LU209** | $P_{cwlOleader}$*lytE [L12cat] ΔcwlO ΔlytE* |
| **LU286** | $P_{cwlOleader}$*lytE [L222cat] ΔcwlO* | **LU287** | $P_{cwlOleader}$*lytE [L222cat] ΔcwlO ΔlytE* |

Antibiotic sensitivity was determined for the strains listed in Table 5.20 by the use of Etest® stripes with antibiotics of the following concentrations: vancomycin (0.016-256 µg/ml), ampicillin (0.016-256 µg/ml) and fosfomycin (0.064-1024 µg/ml) as described in Chapter 2 (section 2.2.10).

None of the investigated strains showed an altered sensitivity to vancomycin or ampicillin (data not shown), but increased sensitivity to fosfomycin was detected for strains that express a variant LytE protein with one LysM motif in a Δ*lytE* background (i.e. LU262, LU266 and LU270).

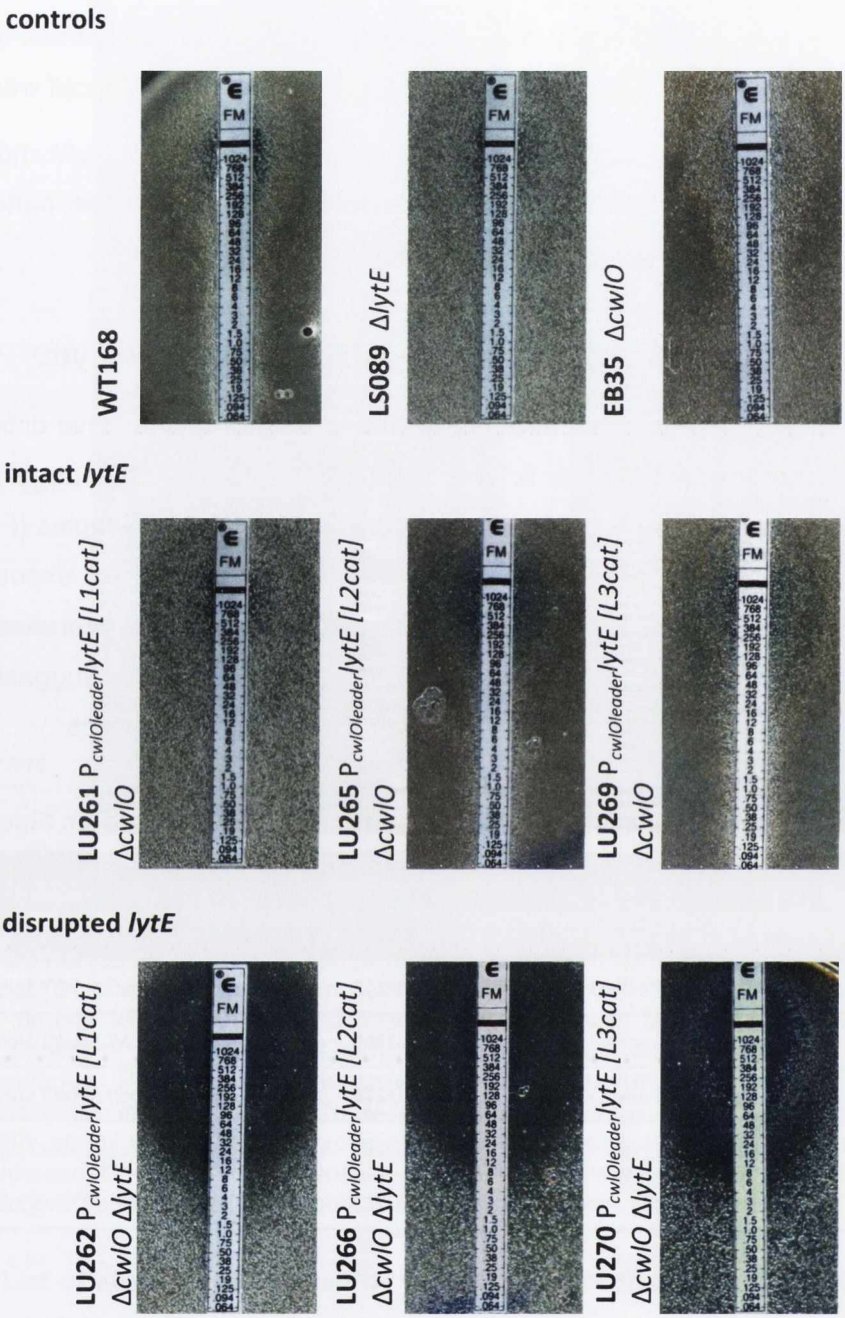



**Figure 5.22: Susceptibility to fosfomycin of strains that express a variant LytE.** Susceptibility to fosfomycin was established for the following strains: WT168, LS089 (Δ*lytE*), EB35 (Δ*cwlO*), LU261 ($P_{cwlOleader}$*lytE [L1cat]* Δ*cwlO*), LU265 ($P_{cwlOleader}$*lytE [L2cat]* Δ*cwlO*), LU269 ($P_{cwlOleader}$*lytE [L3cat]* Δ*cwlO*), LU262 ($P_{cwlOleader}$*lytE [L1cat]* Δ*cwlO* Δ*lytE*), LU266 ($P_{cwlOleader}$*lytE [L2cat]* Δ*cwlO* Δ*lytE*) and LU270 ($P_{cwlOleader}$*lytE [L3cat]* Δ*cwlO* Δ*lytE*). Susceptibility was determined with Etest® stripes (0.064-1024 µg/ml fosfomycin). Strains LU262, LU266 and LU270 have an increased susceptibility to fosfomycin, shown by the ellipse around the higher end of the scale which is absent in the controls.

Figure 5.22 shows the growth limitations induced by fosfomycin in these strains. Unfortunately, it was not possible to determine the exact MIC (Minimal Inhibitory Concentration), since no homogenous growth could be obtained from these strains. However, the approximate MIC for these strains was 8 µg/ml fosfomycin. Interestingly strains that express the same variant LytE, but that also possessed the endogenous *lytE* gene did not exhibit a change in sensitivity to fosfomycin (compare LU261 to LU262, LU265 to LU266 and LU269 to LU270).

It is concluded that resistance to fosfomycin is impaired in strains that express variant LytE proteins with a single LysM motif (i.e. LU262, LU266, LU270), but that resistance to ampicillin and vancomycin in these strains is unaffected.

### 5.4.14 Characteristics of variant LytE proteins *in vitro*

#### 5.4.14.1 Predicted protein structures of variant LytE proteins

Protein structures of the purified variant LytE proteins were predicted with the I-TASSER server (Yang *et al.*, 2014). The predicted protein structures are shown in Figure 5.23. The predicted structures demonstrate that the catalytic domain is separated from the LysM motifs by the spacer for all the variant LytE proteins in this study. Furthermore, in proteins with more than one LysM motif, it is predicted that these domains form independent structures.

#### 5.4.14.2 Enzymatic activity of variant LytE proteins

Several studies have shown that a total loss of LysM motifs greatly decreases not only peptidoglycan-binding, but also the enzyme activity of LysM motif containing autolysins *in vitro* (Eckert *et al.*, 2006; Steen, Buist *et al.*, 2005; Wong *et al.*, 2014). In contrast Steen *et al.* (2003) demonstrated that the removal of the three LysM motifs of the major autolysin AcmA of *Lactococcus lactis* did not alter enzymatic activity (Steen *et al.*, 2003). To address the effect of varying the number and combination of LysM motifs on LytE activity the enzymatic (autolytic) activity of selected variant LytE proteins (listed in Table 5.2, page 126) was established *in vitro* and *in vivo* as described in Chapter 2. The results are shown in Figure 5.24 (page 172) and Figure 5.25 (page 173).

The enzymatic activity of variant LytE proteins was established in a non-quantitative way by testing their ability to degrade *in vitro* purified cell walls fixed in a SDS PAGE gel. The result is presented in Figure 5.24. It is evident that all purified LytE protein produced a visible clearing at the position that correlates with the molecular weight of each variant LytE (Figure 5.24).

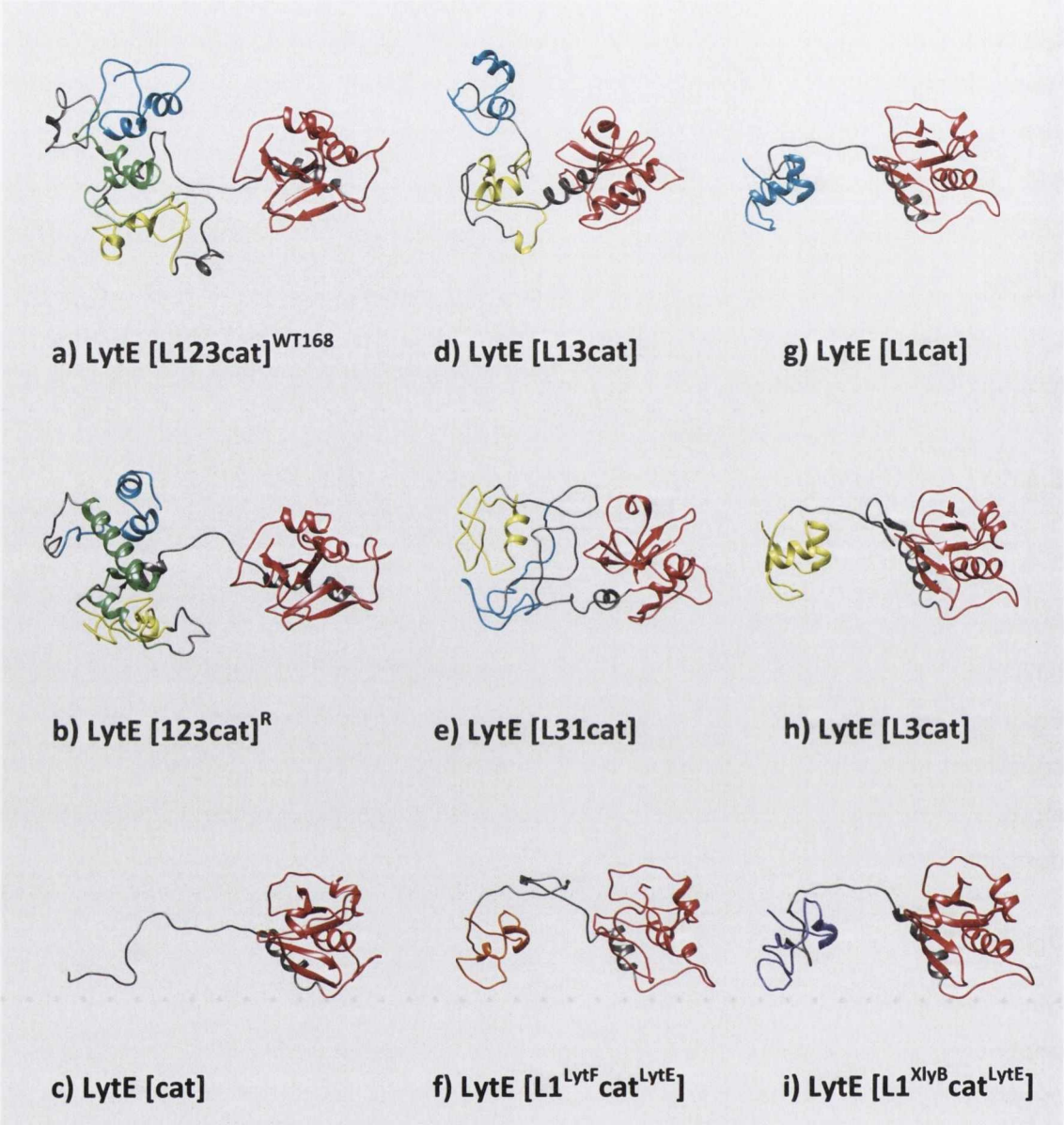



**Figure 5.23: Protein structure predictions for variant LytE.** Protein structures were predicted for the following variant LytE proteins: **(a)** LytE[L123cat]$^{WT168}$, **(b)** LytE[L123cat]$^{R}$, **(c)** LytE[cat], **(d)** LytE[L13cat], **(e)** LytE[L31cat], **(f)** LytE[L1$^{LytF}$cat$^{LytE}$], **(g)** LytE[L1cat], **(h)** LytE[L3cat], **(i)** LytE[L1$^{XlyB}$cat$^{LytE}$]. Protein structures were predicted from the amino acid sequence with the I-TASSER server for protein structure and function prediction. The confidence scores estimating the quality of the model (range -5 to 2, with a higher score predicted a higher quality) are -3.10 , -2.91, -0.11, -2.79, -2.59, 0.22, -0.46, -0.05, -0.46 (from a to i). Models were coloured with the Chimera software. Colours indicate the following domains: Red = catalytic domain, Black = linker / spacer, Blue = LytE LysM motif 1, Green = LytE LysM motif 2, Yellow = LytE LysM motif 3, Orange = LytF LysM motif 1, Purple = XlyB LysM motif 1. R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

Interestingly the two wild-type LytE proteins (LytE[L123cat]$^{WT168}$ and LytE[L123cat]$^{R}$) have a second clearing below the position predicted for the intact protein, indicating perhaps that a degradation product also was enzymatically active (Figure 5.24). Moreover, it is evident that the LytE catalytic domain alone (LytE[cat]) was also enzymatically active. Thus, it was demonstrated that all variant proteins were enzymatically active.

The enzymatic activity of variant LytE proteins was further established by testing their ability to degrade *in vitro* purified cell walls (Figure 5.25 a) and live cells of *B. subtilis* (Figure 5.25 b). For the degradation of *in vitro* purified cell walls (Figure 5.25 a) the greatest decrease in $OD_{600}$ was observed with LytE[L123cat]$^{WT168}$ whereas no degradation was observed when no protein was added (control). All variant LytE proteins degraded cell walls to a lesser extent that did the wild-type protein, including the protein with only the catalytic domain (LytE[cat]). However, all variant LytE proteins degraded the purified cell walls at similar rates. Thus, it is concluded that all proteins have approximately the same enzymatic activity when measured by this assay *in vitro*.

The assay with live *B. subtilis* cells (Figure 5.25 b) measures the ability of the added LytE protein to induce cell lysis. It can be seen that the cells gradually lysed even without addition of LytE protein during the experiment (control). However, addition of LytE[L123cat]$^{WT168}$ increased the rate and extent of lysis. Interestingly, the 'apparent initial rate' of cells lysis was greater for all the variant LytE proteins tested (including the catalytic domain) than it was for the LytE[L123cat]$^{WT168}$ protein.

A few variant LytE proteins degraded the live *B. subtilis* cells somewhat slower (i.e. LytE[L1$^{XlyB}$cat$^{LytE}$], LytE[L1$^{LytF}$cat$^{LytE}$], LytE[cat]) and the most rapid degradation was obtained with LytE[L123cat]$^{R}$.

For unknown reasons the LytE[L123cat]$^{WT168}$ protein showed the greatest enzymatic activity when deployed to degrade purified cell walls, but the lowest enzymatic activity when deployed to degrade live cells. This was confirmed in an independent repeat of the experiment with a LytE[L123cat]$^{WT168}$ protein purified in a different experiment. It is possible that the LytE[L123cat]$^{WT168}$ protein has some interaction with the cell-associated proteins of the live cells, which are not present for the variant LytE proteins. However, this does not explain why then, the LytE[L123cat]$^{WT168}$ protein shows a higher activity in the degradation of peptidoglycan than other variant LytE proteins.

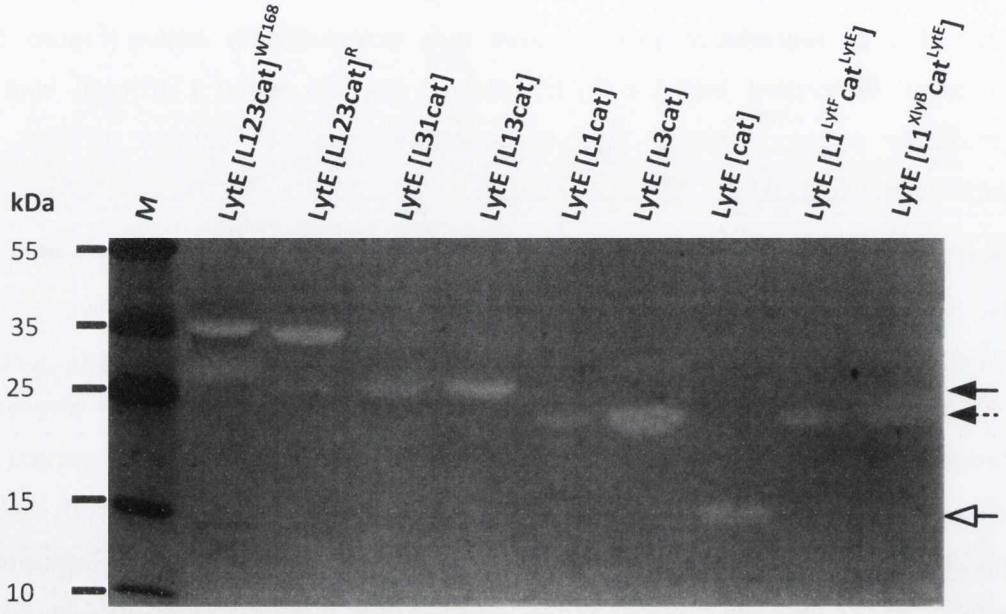



**Figure 5.24: Ability of purified variant LytE proteins to degrade purified cell walls established by zymography.** The following variant LytE proteins were analysed by zymography: LytE[L123cat]$^{WT168}$ (34.09 kDa), LytE[L123cat]$^{R}$ (33.85 kDa), LytE[L31cat] (27.38 kDa), LytE[L13cat] (27.38 kDa), LytE[L1cat] (21.15 kDa), LytE[L3cat] (20.87 kDa), LytE[cat] (16.1 kDa), LytE[L1$^{LytF}$cat$^{LytE}$] (20.62 kDa) and LytE[L1$^{XlyB}$cat$^{LytE}$] (20.77 kDa). 500 pmole of each variant LytE protein were separated on a 12 % SDS PAGE gel, containing purified cell walls of *Bacillus subtilis 168*. Proteins were then re-naturated and given time to degrade the cell walls, before the gel was stained with 0.1 % methylene blue and then destained until clearing zones became visible. Molecular weight markers are shown (M, kDa). R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

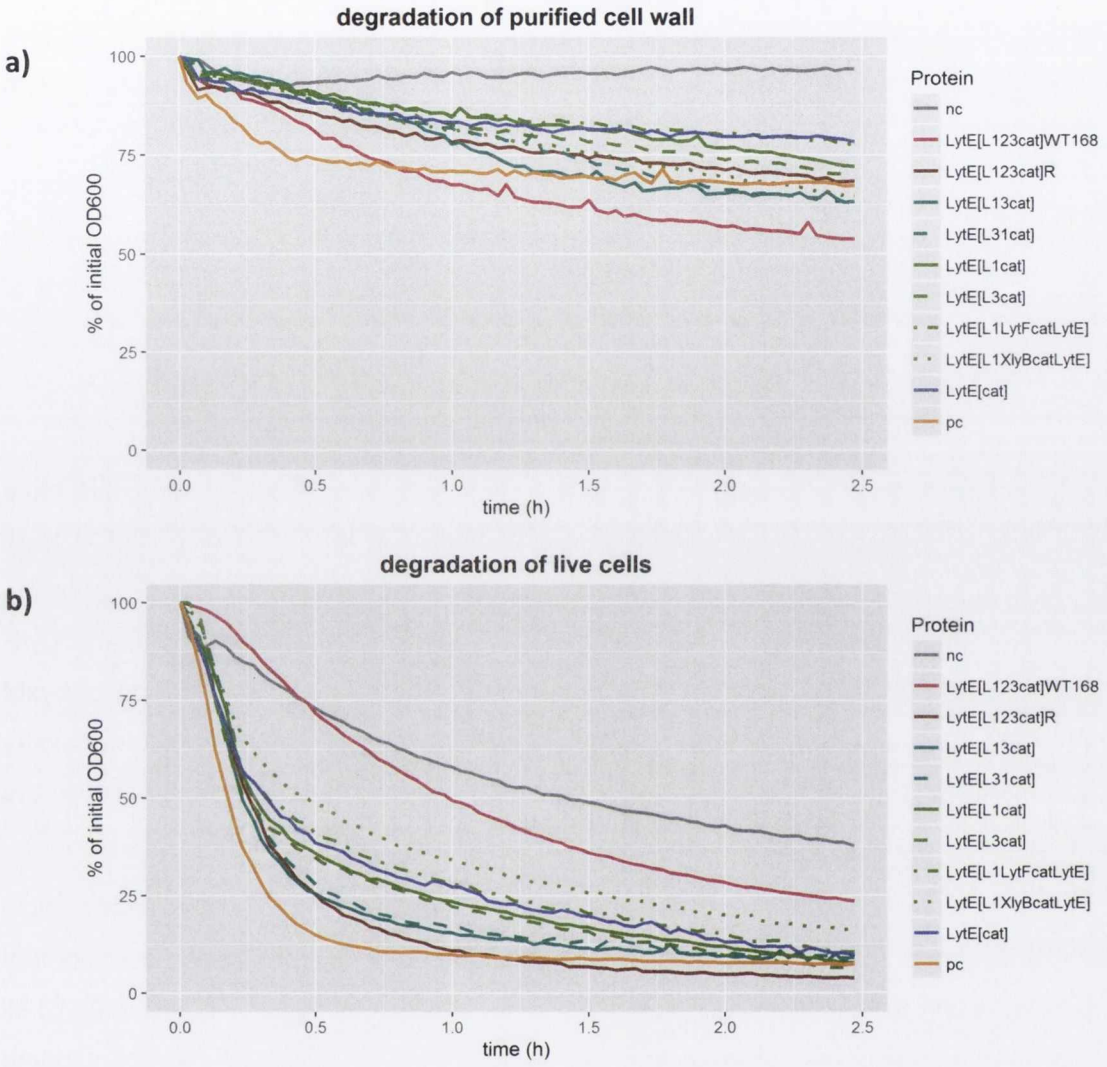



**Figure 5.25: Ability of purified variant LytE proteins to degrade live cells and purified cell walls. a)** The diagram shows the enzymatic activity of variant LytE proteins (each deployed at 85 pmole per 100 µl reaction volume) against purified cell walls (obtained from *B. subtilis 168*) suspended in HPDM. **b)** The diagram shows the enzymatic activity of variant LytE protein (each deployed at 60 pmole per 100 µl reaction volume) against live *B. subtilis* WT168 cells, resuspended in 10mM Tris-HCl (pH 8.0), which were prior grown in HPDM to an $OD_{600}$ of 0.5. The mixtures were incubated a 37°C in a 96 well plate and cell wall lytic activity was measured as the decrease in $OD_{600}$ every 2 minutes over a period of 3 hours, using a plate reader with slow orbital shaking at 37°C.Three technical replicates were conducted for each strain and the results showed a very high level of reproducibility. One representative example is displayed. A negative control (nc, no purified protein added) and a positive control (pc, lysozyme) are shown. R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

In summary, it can be concluded that all variant LytE proteins (including the variant LytE without LysM motifs) have enzymatic activities and that all purified variant LytE proteins are overall similarly active, although some variation was observed in the live cell assay. Interestingly, there appears to be an inverse correlation with complementation and binding – therefore it could be that those proteins that have higher binding affinity to peptidoglycan permanently associate with the cell wall and can access a limited range of substrate, whereas those that have lower binding affinity only temporarily associate with peptidoglycan and can therefore access a higher number of substrates.

#### 5.4.14.3 Binding affinity of variant LytE proteins to purified cell walls *in vitro*

Results presented in this work suggest that variant LytE with fewer LysM motifs have lesser binding capacity as observed by western analysis assay (Figure 5.13 – page 142, Figure 5.14 – page 146) and immunofluorescence microscopy (Figure 5.21, page 166). Therefore, it was hypothesized that the binding capacity of variant LytE proteins (listed in Table 5.2, page 126) is dependent on the number and type of LysM motifs. This hypothesis was tested using an insoluble peptidoglycan pull down assay as described in Chapter 2 (section 2.2.16) with varying amounts of NaCl added to the assay buffer (0.05 M, 0.1 M, 0.2 M, 0.3 M, 0.4 M and 1 M NaCl).

The results of these experiments are shown in Figure 5.26. The results clearly show that the peptidoglycan-binding capacity is subject to two factors: the type of variant LytE protein and the salt concentration. With a low salt concentration of 0.05 M NaCl all variant LytE proteins with at least one LysM motif showed an equally strong binding capacity of approximately 100 % (proportion of protein found in the pellet). However, no binding could be detected for the variant LytE without LysM motifs (LytE[cat]).

An increase in the salt concentration (≥ 0.1 M NaCl) resulted in a decreased binding capacity for all variant LytE proteins. However, the rate at which binding capacity declined varied between proteins and correlated with the number of LysM motifs present. The LytE[L123cat]$^{WT168}$ protein and the wild-type like protein LytE[L123cat]$^{R}$ showed the greatest binding capacity with nearly 100 % binding up to 0.3 M salt. Binding decreased slightly at higher salt concentrations, but still reached > 90 % at 1 M salt. The variant LytE proteins possessing two LysM motifs (i.e. LytE[L13cat] and LytE[L31cat]) exhibited a less robust binding capacity that declined to 80 % with 0.4 M NaCl and 30 % to 40 % at 1 M NaCl. The variant LytE with one LysM motif (i.e. LytE[L1cat] and LytE[L3cat]) exhibited a sharply declining binding capacity.

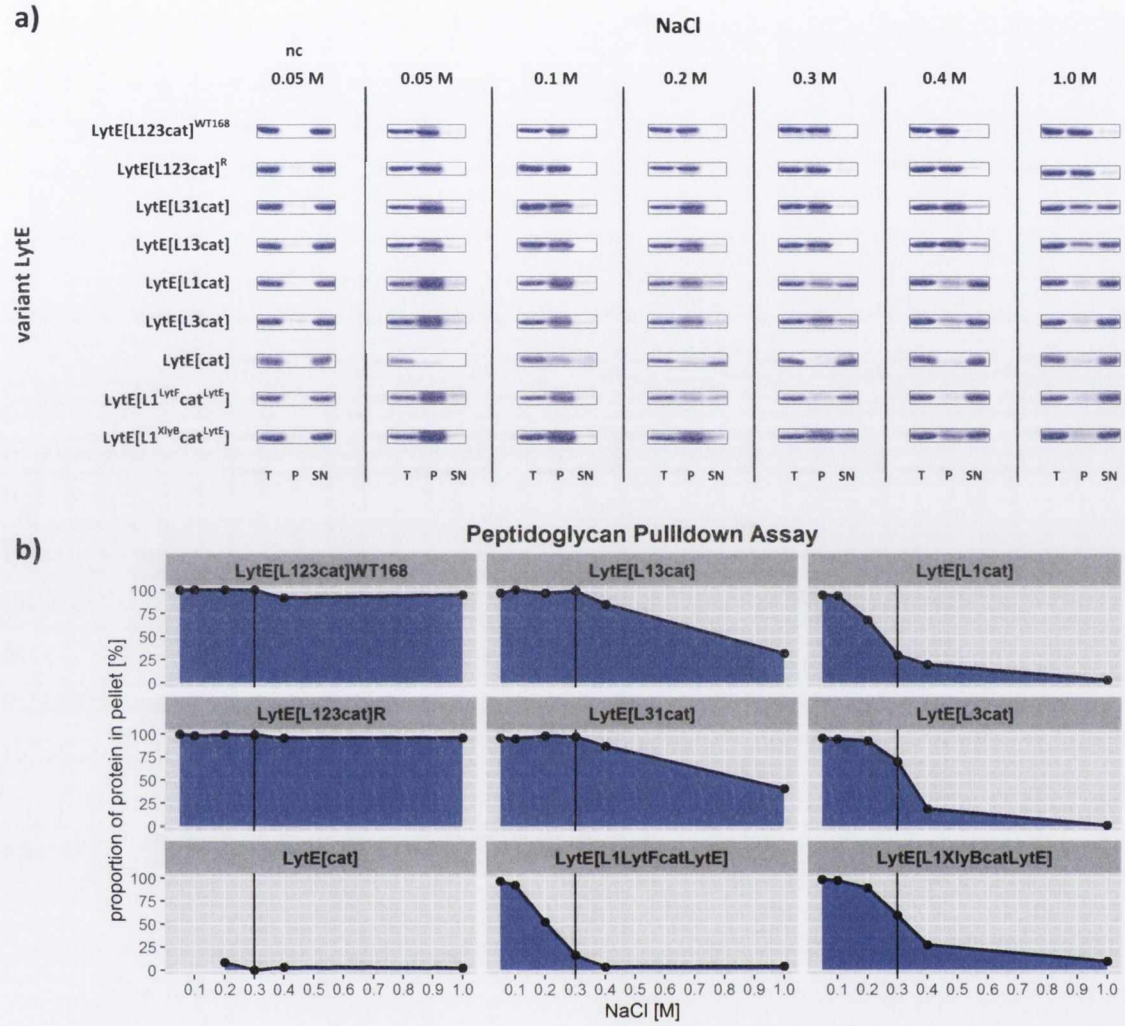



**Figure 5.26: Binding affinity of purified variant LytE proteins to purified *B. subtilis* cell walls at different salt concentrations.** The table **(a)** and graph **(b)** show the results of peptidoglycan pull down assays for variant LytE proteins. The following proteins were analysed: LytE[L123cat]$^{WT168}$, LytE[L123cat]$^{R}$, LytE[cat], LytE[L13cat], LytE[L31cat], LytE[L1$^{LytF}$cat$^{LytE}$], LytE[L1cat], LytE[L3cat], LytE[L1$^{XlyB}$cat$^{LytE}$]. Forty-four pmole of variant LytE protein were mixed with purified cell walls of *B. subtilis* (to a final absorbance of $OD_{600}$ = 10) in a 25 µl reaction volume. A negative control (nc) with without cell wall was performed to ensure that proteins do not precipitate on their own. Pulldown assays were conducted for five different NaCl concentrations (0.05 M, 0.1 M, 0.2 M, 0.3 M, 0.4 M and 1 M). The table **(a)** shows the Coomassie Blue stained proteins (T = Total protein, P = Protein in pellet, SN = protein in supernatant). The graph **(b)** shows the quantified protein binding capacity defined by proportion of protein found in the pellet. R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

Interestingly binding capacity decreased slightly faster for LytE[L1cat] than for LytE[L3cat]. At 0.2 M NaCl approximately 70 % of the LytE[L1cat] protein was found in the pellet, but about 90 % for LytE[L3cat]. With 0.3 M NaCl LytE[L1cat] achieved 30 % binding and LytE[L3cat] 70 % binding. Both reached around 20 % binding with 0.4 M NaCl and almost no bound protein could be detected with 1 M NaCl (< 5 %). The heterologous variant LytE showed similar binding patterns with LytE[L1$^{XlyB}$cat$^{LytE}$] exhibiting a slightly higher binding capacity than LytE[L3cat] and LytE[L1$^{LytF}$cat$^{LytE}$] having a slightly lower binding capacity than LytE[L1cat].

Overall this experiment attests to the ability of LysM motifs to confer binding to peptidoglycan. All variant LytE proteins, with at least one LysM motif bound equally well to peptidoglycan at a salt concentration of 0.5 M NaCl. Interestingly differences in binding capacity between different variant LytE proteins could be detected at a salt concentration greater than 0.1 M NaCl. While all proteins showed similar binding capacity at lower salt concentrations, an increase in the salt concentration impaired binding. The extent to which binding was hindered depended on the number of LysM motifs in a protein. Proteins with more LysM motifs exhibited a greater robustness of binding affinity than proteins with a single LysM motif. This attests to the ability of individual LysM motifs to contribute in an additive manner to peptidoglycan-binding. However, individual LysM motifs were not equal in their binding capacities, which was evident when comparing variant LytE with a single LysM motif.

#### 5.4.14.4 The effect of anionic polymers on binding affinity of variant LytE proteins to purified cell walls *in vitro*

It had been demonstrated that all variant LytE protein bound well to purified cell walls of *B. subtilis 168* at low salt concentration. However, binding was hindered by an increase in salt and the extent to which the binding capacity of different proteins declined was subjected to the number of LysM motifs (Figure 5.26). It was hypothesized that binding might also depend on the type of cell walls and whether they contain (i) no anionic polymer WTA; (ii) normal WTA and (iii) WTA replaced by teichuronic acid.

For this reason, cell walls of different strains (listed in Table 5.21) with defects in teichoic metabolism were purified, and the binding of the variant LytE proteins (listed in Table 5.2, page 126) to these purified cell walls was tested and compared to binding capacities achieved with peptidoglycan of *B. subtilis 168* (as described in Chapter 2 section 2.2.16).

Under normal growth conditions wild-type strain WT168 produces glycerol phosphate wall teichoic acid. However, since wall teichoic acids are rich in phosphates their

biosynthesis is restricted upon phosphate starvation. Under these circumstances PhoPR (an important Two-component signal transduction system in *B. subtilis* that mediates response to phosphate starvation) initiates synthesis of teichuronic acid as a replacement anionic polymer that does not contain phosphate (Botella *et al.*, 2014).

CK1 is a control strain with a reconstituted PhoR and has therefore the same cell wall metabolism as WT168. However, the PhoR in the strain CK6 has a point mutation rendering it constitutively active. As a result of this, strain CK6 is locked in a phosphate limitation response even when phosphate is abundant. Its overexpression of PhoP represses the tagAB operon which in turn blocks the biosynthesis of wall teichoic acid. Simultaneously, activation of the teichuronic synthesis operon by PhoR leads to the overexpression of teichuronic acid (an anionic polymer sugar), which results in the generation of an abnormal cell wall structure. Strain EB156 has a xylose inducible promoter that controls expression of TagO, an enzyme essential for the biosynthesis of teichoic acid. Therefore, teichoic acid is produced in this strain only when it is grown with xylose, but not when it is grown without xylose.

The anomalies in the cell wall metabolism of strain CK6 and EB156 (grown without xylose) can be seen as changes in cell morphology as shown in Figure 5.27. The peptidoglycan-binding affinity of three investigated variant LytE proteins (LytE[L1cat], LytE[L13cat] and LytE[L123cat]$^{R}$) as well as LytE[L123cat]$^{WT168}$ to the different types of purified cell walls was determined in an insoluble peptidoglycan pull-down assay as described earlier. The result is shown in Figure 5.27 and demonstrates that the binding capacity is not dependent on anionic polymers. All tested variant LytE proteins showed the same binding capacities to all four purified cell wall types. As observed earlier, binding was subject to the number of LysM motif within a protein and indirectly correlated with the salt concentration. However, no differences could be detected for the binding capacity of a specific protein to different anionic polymers.

**Table 5.21: List of strains from which purified cell walls was obtained (see Figure 5.27)**

| Strain | Genotype | Cell wall metabolism |
|---|---|---|
| **WT168** | | biosynthesis of glycerol phosphate wall teichoic acid |
| **CK1** | *ΔphoR phoP::phoR* | biosynthesis of glycerol phosphate wall teichoic acid |
| **CK6** | *ΔphoR phoP::phoR [Met249Lys]* | no biosynthesis of glycerol phosphate wall teichoic acid, active biosynthesis of teichuronic acid |
| **EB156 [1 % xylose]** | *ΔxylR ΔxylA ΔxylB $P_{xyl}$tagO* | biosynthesis of glycerol phosphate wall teichoic acid |
| **EB156 [no xylose]** | *ΔxylR ΔxylA ΔxylB $P_{xyl}$tagO* | no biosynthesis of glycerol phosphate wall teichoic acid |

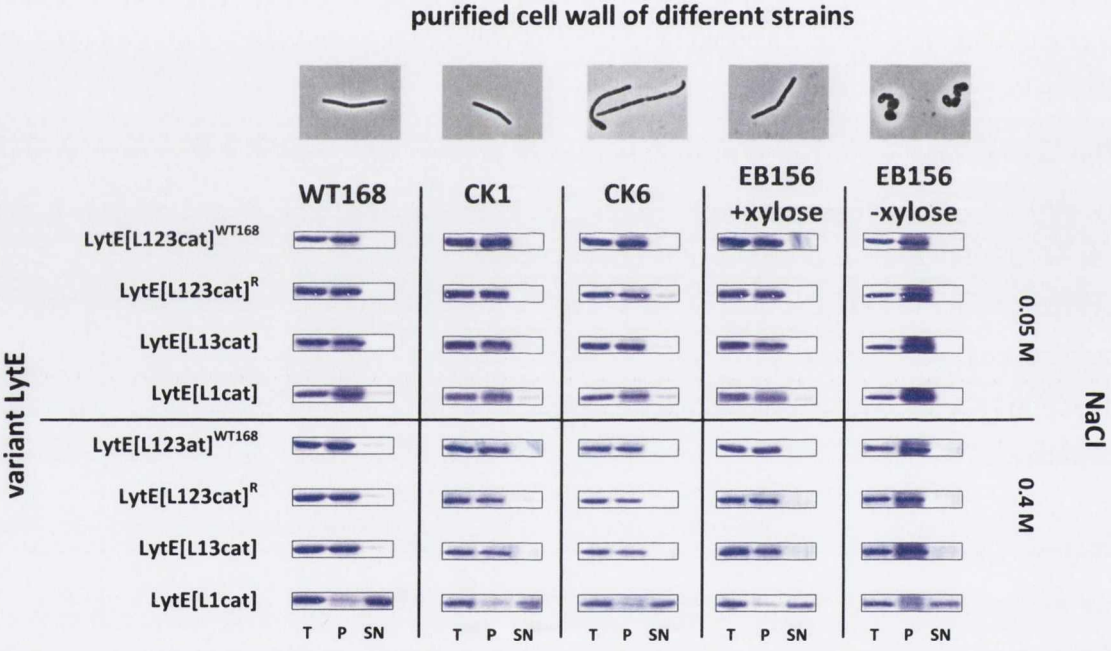



**Figure 5.27: Binding affinity of variant LytE proteins to different types of purified cell walls at different salt concentrations.** The table shows the Coomassie Blue stained proteins which are the result of the peptidoglycan pull down assay for variant LytE proteins with different types of purified cell walls (obtained from the strains pictured above). The following proteins were analysed: LytE[L123cat]$^{WT168}$, LytE[L123cat]$^{R}$, LytE[L13cat], LytE[L31cat] and LytE[L1cat]. Forty-four pmole of variant LytE protein were mixed with purified cell walls (to a final absorbance of $OD_{600}$ = 10) in a 25 µl reaction volume. The peptidoglycan pull down assay was conducted for two different NaCl concentrations (0.05 M and 0.4 M). Five different types of purified cell walls were tested: WT168 (obtained from the WT168 strain), CK1 (obtained from the strain CK1), CK6 (obtained from the strain CK6), EB156+ (obtained from the strain EB156 grown with 1% xylose) and EB156- (obtained from the strain EB156 grown without xylose). The strains have the following genotypes: WT168, CK1 (*ΔphoR phoP::phoR*), CK6 (*ΔphoR phoP::phoR [Met249Lys]*) and EB156 (*ΔxylR ΔxylA ΔXlyB $P_{xyl}$tagO*). T = Total protein, P = Protein in pellet, SN = protein in supernatant. R superscript denotes the reconstituted wild-type protein i.e. constructed according to the methodology of the recombinant proteins.

## 5.5 Conclusion

The viability of *B. subtilis* cells requires D,L-endopeptidase activity for cleavage of peptide cross-links in the peptidoglycan of the cell wall. This requirement is fulfilled either by CwlO (in exponentially growing cells) or by LytE (in cells exposed to cell wall stress). An interesting feature of these two synthetically lethal enzymes is that the protein domains and mechanism through which they are cell-associated are completely different. CwlO has a coiled-coil N-terminal domain and interacts with the FtsEX protein complex that is embedded in the membrane, an interaction that is required for its activity. In contrast, LytE has three LysM motifs that bind to peptidoglycan (see Figure 1.4, page 23). We sought therefore to further investigate how these LysM motifs function to allow LytE fulfil the D,L-endopeptidase requirement *in vivo*. This was especially interesting since we have shown (Chapter 4) that the D,L-endopeptidase requirement for viability can be fulfilled by LytE (3 LysM motifs) and to a lesser extent by LytF (5 LysM motifs) when they are expressed using the $P_{cwlO}$ promoter, but not by CwlS (4 LysM motifs). The focus of this investigation was to investigate the LysM motif requirement that enables LytE to fulfil the D,L-endopeptidase *in vivo*. These studies are complemented by those of other groups who have investigated LysM motif binding to peptidoglycan *in vitro* (Eckert *et al.*, 2006; Steen *et al.*, 2003; Steen, Buist *et al.*, 2005; Wong *et al.*, 2014). Therefore, the synthetic lethality of CwlO and LytE presents a unique opportunity to examine the LytE / CwlO D,L-endopeptidase requirement in *B. subtilis* and to investigate how the three LysM motifs of LytE contribute to the functionality of the enzyme *in vivo*.

To address this questions a system was designed that allowed the generation of genes encoding variant LytE proteins with different numbers and combinations of LysM motifs. The resulting variant LytE proteins varied greatly from the WT168 combination (LytE[L123cat]$^{WT168}$) and included proteins with a single LysM motif (3 different variants), with two LysM motifs (9 different variants), with three LysM motifs (8 different variants), with four LysM motifs (4 different variants) and with five LysM motifs (3 different variants). In addition four heterologous variant LytE were designed that contained a single LysM motif of a heterologous enzyme (L1$^{CwlS}$, L1$^{LytF}$, L1$^{XlyB}$, L1$^{Aaa}$). Two controls were included in the study: a reconstituted wild-type LytE as positive control (LytE[L123cat]$^{R}$), and a variant LytE without LysM motifs (LytE[cat]) as a negative control.

All variant *lytE* (32 variants in total) were cloned into the *cwlO* locus, so that the variant LytE protein was expressed from the *cwlO* promoter, thereby replacing the *cwlO* gene (strains listed in Table 5.3 – page 128 and Table 5.4 – page 130). A selection of

variant *lytE* (11 variants in total) were cloned onto the *lytE* locus, so that the variant LytE protein was expressed from the *lytE* promoter, thereby replacing the endogenous *lytE* gene (strains listed in Table 5.5, page 131).

Furthermore, a selection of variant LytE proteins (listed in Table 5.2, page 126), was overexpressed in *E. coli*, purified and analysed *in vitro*. All purified variant LytE proteins including a variant LytE protein without LysM motifs (LytE[cat]) were shown to be enzymatically active (Figure 5.24, page 172), although rates of activity varied slightly (Figure 5.25, page 173). This result is in contrast to several studies that have shown that a total loss of LysM motifs greatly decreases the enzymatic activity of LysM motif containing autolysins *in vitro* (Eckert *et al.*, 2006; Steen, Buist *et al.*, 2005; Wong *et al.*, 2014), but in accordance with Steen *et al.* (2003), who demonstrated that the removal of the three LysM motifs of the major autolysin AcmA of *Lactococcus lactis* did not alter enzymatic activity.

Expression of the variant LytE from the *cwlO* promoter resulted in higher cell-associated protein levels than expression from the *lytE* promoter (Figure 5.11, page 138), indicating that the *cwlO* promoter is a stronger promoter than the *lytE* promoter as observed earlier in this study (Figure 3.6, page 76). It was further shown that the number of LysM motifs in variant LytE proteins is a determinant of the extent to which the protein is cell-associated or shed into the culture media (Figure 5.13 – page 142, Figure 5.14 – page 146). Variant LytE proteins with one LysM motif were predominantly found in the culture supernatant while variant LytE proteins with three LysM motifs were predominantly found in association with the cell wall. Moreover the distribution of variant LytE proteins with two LysM motifs was an intermediate between that found for those with three and one LysM motifs. It is concluded therefore that the observed differences in distribution of these proteins, observed *in vivo*, is due to differences in the number of LysM motifs in the different variant LytE; with increasing number of LysM motifs leading to increased binding, in agreement with the *in vitro* experiments reported by Mesnage *et al.* (2014) and Wong *et al.* (2014). Western analysis further showed that there may be a defect in processing the signal sequence in variant LytE that do not have the LysM1 Motif in its normal position (adjacent to the signal sequence) (Figure 5.11, page 138).

In agreement with the observed distribution of variant LytE proteins *in vivo*, *in vitro* binding studies of the purified proteins showed that variant LytE with two or three LysM motifs exhibit a more robust peptidoglycan-binding affinity than variant LytE with a single LysM motif (Figure 5.26 – page 175, Figure 5.27 – page 178). Although all variant LytE proteins bound equally well to purified cell walls at a low salt

concentration, binding affinity was impaired at higher salt concentrations (> 0.1 M NaCl). Variant LytE with three LysM motifs were able to maintain high binding affinity at increased salt concentrations, however variant LytE with a single LysM motif were not. Again variant LytE with two LysM motifs took an intermediate position. This indicates that individual LysM motif contribute to binding to peptidoglycan additively. This observation is contrary to an *in vitro* study by Hu *et al.* (2010), who found that 0.5 M NaCl is the best concentration for LysM mediated cell wall binding. As expected, a variant LytE without LysM motifs exhibited no binding affinity to purified cell walls under all tested salt concentrations, thus LysM motifs are essential for cell wall binding. This observation is supported by several previous studies that have shown that a total loss of LysM motifs greatly decreases the peptidoglycan-binding capacity *in vitro* (Eckert *et al.*, 2006; Steen, Buist *et al.*, 2005; Wong *et al.*, 2014). It was also shown that the anionic polymer composition of the purified cell walls, deployed in the binding affinity assay, does not influence binding of the variant LytE proteins (Figure 5.27, page 178).

Thus, the binding affinity of variant LytE is clearly influenced by the number of LysM motifs present (*in vitro* and *in vivo*). However, this seems to have no influence on the localization of the protein. Localization studies demonstrated that a variant LytE with three LysM motifs and a variant LytE with a single LysM motif show the same localization pattern (Figure 5.21, page 166) and are equally distributed along the cell septa, poles and the lateral sidewall of the cell, as observed by Carballido-Lopez *et al.* (2006) and Hashimoto *et al.* (2012). Once more, localization studies also found variant LytE with three LysM motifs to be cell-associated to a greater extent than variant LytE with a single LysM motif, again attesting to the ability of LysM motifs to contribute to peptidoglycan affinity in an additive manner.

The disruption of either the *lytE* gene (variant LytE expressed from $P_{cwlO}$) or the *cwlO* gene (variant LytE expressed from $P_{lytE}$) served as a test for the ability of the individual LysM motif permutations, to confer functionality to the LytE enzyme. The viability of the resulting strains was investigated and the following established: (i) All strains that express a variant LytE with at least one LysM motif (L1, L2 or L3) are viable. (ii) A variant LytE without a LysM motif cannot fulfil the D,L-endopeptidase requirement. A strain that expresses such a variant LytE from the $P_{cwlO}$ or $P_{lytE}$ promoter in a Δ*lytE* or Δ*cwlO* background is not viable. Therefore, LysM motifs are essential for the functionality of the LytE enzyme. (iii) A variant LytE with a single LysM motif of a heterologous enzyme cannot fulfil the D,L-endopeptidase requirement in the case of the $NH_2$-terminal LysM motif of CwlS (*B. subtilis*), LytE (*B. subtilis*) or Aaa (*S. aureus*).

However, a variant LytE with the single LysM motif of XlyB (an autolysin encoded in the PBSX genome of *B. subtilis*), can fulfil the D,L-endopeptidase requirement when expressed from the *cwlO* promoter.

Therefore, of a total of 32 tested variant LytE, only 4 variant LytE (LytE[cat]), LytE[L1$^{CwlS}$cat$^{LytE}$], LytE[L1$^{LytF}$cat$^{LytE}$], LytE[L1$^{Aaa}$cat$^{LytE}$]) were not able to fulfil the D,L-endopeptidase requirement. However, almost all strains in which the D,L-endopeptidase requirement was fulfilled by any of the other 28 variant LytE, although viable, exhibited an impaired robustness. We therefore sought to establish which properties of the variant LytE protein were related to the characteristics observed in these strains.

An attempt was made to investigate the extent to which a variant LytE can fulfil the D,L-endopeptidase requirement by measuring the transcript levels of SigI regulon genes. It was established earlier that the extent to which the SigI regulon is induced, is an indirect measure of the extent to which the D,L-endopeptidase requirement is fulfilled and hence also of cell wall stress (Figure 3.3 – page 68, Figure 3.4 – page 71). It was shown that SigI is not induced in the strain expressing the reconstituted wild-protein (LytE[L123cat]$^{R}$), but to a varying degree in strains expressing variant LytE with a single LysM motif (Figure 5.15, page 149).

None of the strains containing variant LytE exhibited major growth or cell morphology defects (Figure 5.16 – page 151, Figure 5.18 – 156). However, a strain expressing the variant LytE[L3cat] from the *lytE* promoter consistently reached stationary phase prematurely and plateaued more sharply than wild-type (Figure 5.16). Similarly, strains expressing the variants LytE[L1cat], LytE[L2cat] and LytE[L3cat] exhibited shorter and thicker cells, when the variant LytE was expressed at a low level ($P_{lytE}$), but not when expressed a high level ($P_{cwlO}$) (Figure 5.18). It was further observed that colony morphology, which is an indicator for the ability to grow multicellularly, was altered for strains expressing certain variant LytE, especially for those containing a single LysM motif and those expressed from the *lytE* promoter (Figure 5.17, page 154). These strains had small, translucent colonies that showed a strong sign of lysis and developed suppressors. Interestingly, even so, strains expressing the variant LytE[L123cat]$^{R}$ and LytE[L213cat] did not differ from each other in growth behaviour and cell morphology. However they exhibited different colonies, with wild-type like colonies for LytE[L123cat]$^{R}$ and more translucent colonies for LytE[L213cat] (Figure 5.17). Although sensitivity to antibiotics was not increased for strains expressing a variant LytE, in the case of vancomycin and ampicillin, strains expressing a variant

LytE with a single LysM motif from the *cwlO* promoter exhibited an increased sensitivity to fosfomycin (Figure 5.22, page 168).

Finally, growth and cell morphology defects were initiated or amplified when strains were exposed to salt stress (≥ 0.2 M NaCl) (Figure 5.19, page 160). Growth of cells expressing variant LytE with a single LysM motif was inhibited at a salt concentration above 0.1 M NaCl in a dosage-depended manner. All strains exhibited slightly shorter and thicker cell when grown with 0.2 M NaCl, but extremely short and round cells were observed for the variant LytE[L2cat] and LytE[L3cat], when expressed from the *lytE* promoter (Figure 5.20, page 164).

Therefore, although all investigated strains containing a variant LytE were viable, most strains differed from wild-type in at least one characteristic (the three strains that varied the least from wild-type expressed the variants LytE[L11cat], LytE[L123cat]$^{R}$, LytE[L331cat] (all $P_{cwlO}$). The extent to which strains showed a robust and wild-type like growth behaviour and cell morphology depended on the number and the type of LysM motifs present in the variant LytE as well as the expression level of the variant LytE ($P_{cwlO}$ vs $P_{lytE}$).

We conclude that given that all variant LytE have similar enzymatic activity *in vivo* and are expressed at similar levels (depending on the promoter) the observed characteristics of strains containing a variant LytE must be due to the different binding capacities of the variant LytE proteins as discussed above. For example it was shown that a purified LytE[L3cat] protein has a lower binding affinity than LytE[L31cat] which in turn has a lower binding affinity than LytE[L123cat] *in vitro* (Figure 5.26, page 175) and *in vivo* (Figure 5.13, page 142). In accordance with this, strain LU354 (LytE[L3cat], $P_{lytE}$) has a less robust growth, than LU383 (LytE[L31cat], $P_{lytE}$), which is outperformed by LU343 (LytE[L123cat]$^{R}$, $P_{lytE}$) (Figure 5.19, page 160), with similar implementations for colony growth (Figure 5.17, page 154) and cell morphology (Figure 5.18, page 156).

In summary, this study shows that although a variant LytE with a single LysM motif can fulfil the requirement for viability, this appears to be a minimum requirement, given that almost all of the strains containing a variant LytE differed from wild-type. Therefore, in order to fulfil the optimal requirement (robust growth under normal and stress conditions) a LytE protein with more than one LysM motif is required. This is especially evident when analysing growth inhibitions induced by salt stress. Given that *B. subtilis* is a soil-dwelling organism, its capability to withstand salt stress is essential for its survival. Therefore, cells that possess a LytE enzyme with less than three LysM motifs would not be able to survive natural fluctuations in the salt concentrations in the soil. Similarly the ability to form colonies is an indicator of the ability to grow in a

multicellular environment – the natural habitat for *B. subtilis*. Therefore, cells that have a sub-optimal ability to grow multicellularly would again be impaired in their ability to survive in their natural environment.

Importantly this study also shows that a variant LytE can fulfil the D,L-endopeptidase requirement, despite very little protein being cell-associated, while substantial amounts of protein are shed into the medium. This suggests that LytE does not need to bind tightly to peptidoglycan to fulfil the requirement for D,L-endopeptidase activity.

In addition some evidence was found that individual LysM motifs are functionally distinct and differ in their ability to confer functionality to the LytE enzyme with L1 being superior to L2 and L3. A likely explanation for this phenomenon is that the N-terminal LysM motif has some evolutionary developed interaction with the adjacent signal sequence domain, which perhaps causes the enzyme to be expressed, exported or bound to the cell wall more efficiently. A second possible explanation is that some LysM motifs possess innate or inherent characteristics that differentiate them from other LysM motifs, for example a greater peptidoglycan-binding capacity.

# 6 Chapter 6: Global analysis of LysM motifs and the evolutionary development of LytE

## 6.1 Observations

Previous work (Chapter 5) has demonstrated that a LytE enzyme with one LysM motif can complement the LytE[L123cat]$^{WT168}$ enzyme with three LysM motifs. However, the reduction in the number of LysM motifs causes an increased sensitivity to salt stress that correlates with reduced peptidoglycan-binding affinity *in vivo* as well as *in vitro.* Interestingly although there is significant sequence similarity in sequence and structure between the three LysM motifs of LytE (see Figure 1.6, page 32), they were not identical in function, as shown by the different properties of LytE proteins with each of the LysM motifs in single copy. Variant LytE proteins with the N-terminal LysM motif of LytE (LytE[L1cat]) generally showed higher peptidoglycan-binding capacity. In addition, strains in which the D,L-endopeptidase requirement was fulfilled by LytE[L1cat] showed greater stress resistance than strains expressing LytE[L2cat] or LytE[L3cat]. Furthermore, of the four heterologous LysM motifs tested: (i) the N-terminal LysM motif of the CwlS and LytF D,L-endopeptidases of *B. subtilis 168*; (ii) the single LysM motif of XlyB, a PBSX prophage of *B. subtilis 168* and (iii) the N-terminal LysM motif of Aaa – an amidase of *Staphylococcus aureus* Newman, only the LysM motif of XlyB was capable of complementing the LysM motif of LytE (i.e. when combined with the LytE catalytic domain). This demonstrates that although LysM motifs are highly conserved structurally, they are functionally distinct.

These observations raise questions about the evolutionary development of individual LysM motifs and their functional relevance in bacterial species.

(i) How does the number of LysM motifs vary between proteins?

(ii) Is the number of LysM motifs within a protein optimized for peptidoglycan-binding?

(iii) Does the number of LysM motifs control autolytic enzymatic activity?

(iv) How do proteins acquire different numbers of LysM motifs?

(v) Is the first (proximal to the N-terminus) or last (proximal to the C-terminus) LysM motif of a protein functionally different from the others, because of their being adjacent to the signal peptide and catalytic domains respectively?

(vi) Is the number and type of LysM motifs within a protein more important for bacteria found in the environment, where physical conditions will fluctuate significantly more, than for pathogenic bacteria? This possibility is suggested by the fact that a variant LytE protein with only one LysM motif complements the LytE[L123cat]$^{WT168}$ enzyme. However, a strain expressing such variant LytE is more sensitive to salt stress.

(vii) Do LysM motif proteins occur more frequently in Gram-positive than in Gram-negative bacteria, given that Gram-positive bacteria have a thicker peptidoglycan cell wall?

## 6.2 The LysM motifs of *B. subtilis 168*

To address the question of whether the first (proximal to the N-terminus) or last (proximal to the C-terminus) LysM motif of a protein is functionally different from the others, all LysM motifs of *B. subtilis 168* were identified and compared to each other by cluster analysis. The LysM motifs of *B. subtilis 168* were identified using the online bioinformatics database UniProt (UniProt, 2015).

Eighteen LysM motif proteins were identified in *B. subtilis* strain 168. A majority of these proteins contain only one LysM motif (12 proteins), but three proteins contain two LysM motifs (YaaH, YdhD and YocH). Furthermore, D,L-endopeptidases LytE, CwlS and LytF have three, four and five LysM motifs respectively. There are a total of 30 LysM motifs encoded in the *B. subtilis 168* chromosome. All LysM motifs encoded in the *B. subtilis 168* chromosome are listed in Table 6.1. In this study LysM motifs are numbered (e.g. LysM_1, LysM_2) according to their position relative to the N-terminus of the protein. Most LysM motifs of *B. subtilis* are localized in the N-terminal region (e.g. CwlS, LytE and LytF), but some (e.g. XkdP, YqbP, YqfZ) are located in the C-terminal region and others (e.g. XlyA, XlyB, YneA) are centrally located. An alignment of all 30 LysM motifs is shown in Figure 1.6 (page 32), however the 10 C-terminally located LysM motifs were excluded from the analysis below.

The 20 N-terminally located LysM motifs were analysed by multiple sequence alignment and a phylogenetic tree was constructed, which is presented in Figure 6.1. These data show that there is significant similarity among the N-terminal located LysM motifs of the three D,L-endopeptidases (LytE, CwlS and LytF).

**Table 6.1: List of the 18 LysM motif proteins of *B. subtilis 168***

| Protein | Accession number | Protein function | LysM motif position* / number** | LysM motif range |
|---|---|---|---|---|
| **CwlS** | O31852 | D,L-endopeptidase, peptidoglycan hydrolase | N / 4 | 27-70, 88-131, 157-200, 225-268 |
| **LytE** | P54421 | cell wall hydrolase (major autolysin), D,L-endopeptidase-type autolysin | N / 3 | 26-69, 86-129, 149-192 |
| **LytF** | O07532 | gamma-D-glutamate-meso-diaminopimelate muropeptidase (major autolysin) | N / 5 | 27-70, 92-135, 174-217, 240-283, 307-350 |
| **SafA** | O32062 | morphogenetic protein | N / 1 | 2-47 |
| **spoVID** | P37963 | spore coat assembly | C / 1 | 523-567 |
| **XkdP** | P54335 | unknown | C / 1 | 159-217 |
| **XlyA** | P39800 | N-acetylmuramoyl-L-alanine amidases (PBSX prophage) | Centr / 1 | 159-203 |
| **XlyB** | O34391 | N-acetylmuramoyl-L-alanine amidase (PBSX prophage) | Centr / 1 | 177-221 |
| **YaaH** | P37531 | N-acetylglucosaminidase | N / 2 | 1-42, 47-91 |
| **YdhD** | O05495 | unknown | N / 2 | 2-45, 48-92 |
| **YkuD/Ltd** | O34816 | L,D-Transpeptidases | N / 1 | 2-45 |
| **YkvP** | O31681 | unknown | C / 1 | 343-386 |
| **YkzQ** | C0H403 | unknown | C / 1 | 29-72 |
| **YneA** | Q45056 | cell division control protein | Centr / 1 | 38-89 |
| **YocH** | O34669 | amidase | N / 2 | 26-69, 78-121 |
| **YpBE** | P50731 | unknown | C / 1 | 189-235 |
| **YqbP** | P45932 | similar to phage-related protein | C / 1 | 159-217 |
| **YqfZ** | P54483 | unknown | C / 1 | 48-95 |

* N = N-terminal, C = C-terminal, Centr = central
** total number of LysM motifs within a protein

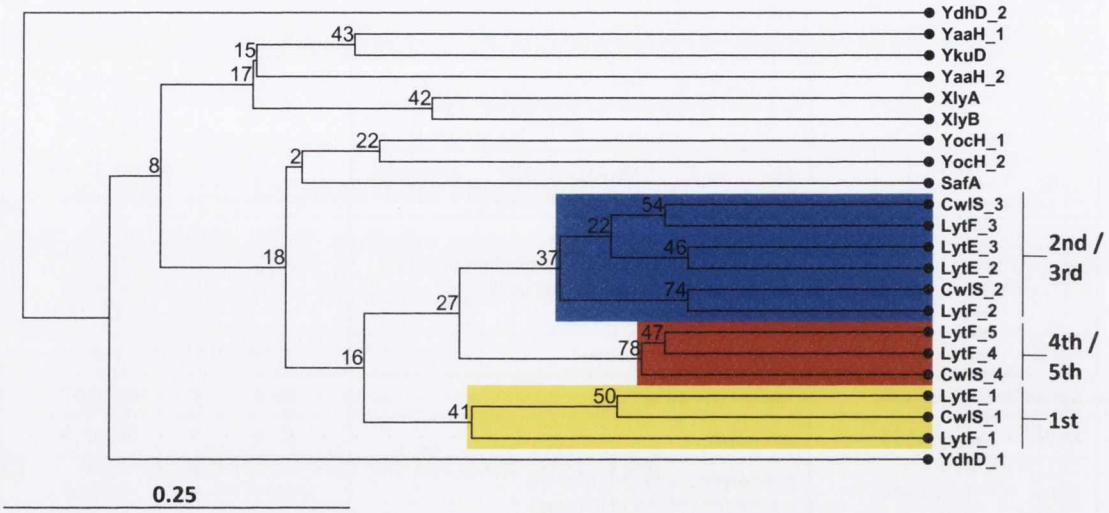



**Figure 6.1: Phylogenetic tree of the N-terminal localized LysM motifs of *B. subtilis 168*.** The phylogenetic tree was constructed from a multiple sequence alignment (generated with Clustal Omega) of the amino acid sequence of the N-terminal localized LysM motifs of *B. subtilis 168*. The phage proteins XlyA and XlyB are included as outgroup. LysM motifs are identified by protein name and their position within the protein (e.g. LytE_1 refers to the $NH_2$-terminal LysM motif of the LytE enzyme). Colours highlight clusters which are formed by the twelve LysM motifs of CwlS, LytE and LytF that cluster according to their position within the protein, not the respective protein (Yellow = First Position, Blue = Second and Third Position, Red = Fourth and Fifth Position). The phylogenetic tree was constructed with the UPGMA algorithm. Distance measures are given in Jukes-Cantor units. Bootstrap values are shown (1000 replicates).

Interestingly, the LysM motifs of these three D,L-endopeptidases do not cluster according to the protein in which they are present, but instead according to their position within these proteins. The three LysM motifs at position one (Figure 6.1 yellow, - CwlS_1, LytE_1, LytF_1) are clustered, while the LysM motifs at position two, three, four and five also form a cluster that splits into two subgroups: one contains CwlS_4, LytF_4 and LytF_5 (Figure 6.1 red), and the second contains the remaining motifs CwlS_2, CwlS_3, LytE_2, LytE_3, LytF_2 and LytF_3 (Figure 6.1 blue). A word of caution is that the bootstrap values for these clusters are low, indicating a low reproducibility of the phylogenetic tree.

However, this cluster analysis indicates that the first (proximal to the N-terminus) LysM motif of each of the three D,L-endopeptidases (CwlS, LytE, LytF) are structurally distinct from those at subsequent positions perhaps, because of their location adjacent to the signal peptide.

## 6.3 Global analysis of ≈ 30,000 LysM motifs

### 6.3.1 Objective

The LysM motif is the most common protein motif for peptidoglycan hydrolases and is widely spread in the bacterial and eukaryotic kingdoms. Buist *et al.* (2008) reports that over 4000 proteins encoded by prokaryotes and eukaryotes contain one or more LysM motifs. Studies have focused either on identifying types of proteins that contain LysM motifs for example Nod factor-binding LysM receptor-like kinases of plants (Antolín-Llovera *et al.*, 2012; Gust *et al.*, 2012) or studying the functions of LysM-contain proteins such as recognizing symbiotic and parasitic bacteria (Oldroyd, 2013) or exploring the importance of individual LysM motifs in specific proteins (Steen, Buist *et al.*, 2005).

The frequency with which LysM motif proteins occur within different bacterial subgroups and the frequency with which proteins with multiple LysM motifs occur, have not been reported. It is proposed that determining the occurrence of LysM motifs in different bacterial phyla and families might reveal interesting correlations such as an increased importance for LysM containing proteins in some bacterial groups, but not in others. For example it is proposed that Gram-positive bacteria should have more LysM motif proteins than do Gram-negative bacteria, because of the structure of their cell envelope, and perhaps that soil-dwelling bacteria may have more LysM motifs than do pathogenic bacteria. Therefore, this study set out to identify all bacterial LysM motif

proteins listed on UniProt (a protein database) and address the question of how frequently they occur in different bacterial families.

### 6.3.2 Approach

To get an overview of the frequency and distribution of LysM motif containing proteins in different bacterial species, all LysM motif proteins listed on UniProt (UniProt, 2015) were identified. A UniProt search for all entries containing the term 'LysM' yielded 55,014 hits [January 2016]. The UniProt database contains reviewed entries (entries that were manually annotated and reviewed) and un-reviewed entries which were computer-annotated. For the purpose of this study both types of entry were considered.

All 55,014 search results were analysed with R, an open source programming language and software environment for statistical computing (R Core Team, 2015). The amino acid sequence of the presumed LysM motif was obtained from the UniProt database using the Retrieve/ID mapping tool provided by UniProt. The combined data was then further screened to separate entries describing *bona fide* LysM motifs of proteins from entries otherwise associated with the term 'LysM' The criteria used to identify LysM proteins motifs were: (i) must be from a bacterial species; (ii) sequence must be specific to an individual bacterial strain (e.g. *Bacillus subtilis 168*, *Lactobacillus fermentum CECT5716*), but not unspecific (e.g. *Bacillus subtilis*); (iii) only putative motifs within the range of 40 – 70 amino acids were considered, (iv) a test for duplicate entries was conducted and only unique results were kept. Finally 30 entries were sampled randomly and verified manually to confirm the accuracy of the data.

The resulting dataset contained information relating to 38,576 LysM motifs. These LysM motifs were found in a total of 26,078 proteins encoded by 6,971 different bacterial strains. Strains could be classified into 68 phyla and bacterial 264 families, but only 11 phyla and 61 bacterial families contained 20 or more individual strains. To determine how frequent LysM motifs are in different taxonomic orders, the total number of LysM motifs and the total number of LysM motif containing proteins was summed up for each bacterial phyla and bacterial class. The count was then set in relation to the total number of strains contained within this order to arrive at a LysM motif / per strain ratio.

One problem to overcome was the cumulative occurrence of very closely related strains, which biases a given LysM motif / per strain ratio towards the overrepresented strain. To avoid this effect, closely related strains were pooled in strain groups and a mean LysM motif / per strain ratio was assigned to each strain group.

This approach has some limitations that mainly arise from the fact that only proteins that contain LysM motifs were considered. Therefore, the results give no information about the general frequency with which bacteria have LysM motifs (since bacteria that do not encode LysM motif proteins were not included in the analysis). Furthermore, although it is possible that not all LysM motif proteins of a species are registered in UniProt, since the majority of sequences represented in the database are derived from whole genome sequencing projects (UniProt Collaborators, 2015) it seems reasonable to assume that the information for each species is complete.

### 6.3.3 Results and Discussion

#### 6.3.3.1 Frequency distribution of LysM motifs per strain and the number of LysM motifs within proteins

The general frequency of LysM motifs and LysM motif containing proteins was calculated for the bacterial strains present in the compiled dataset. Only those strains that belonged to one of the following eleven bacterial phyla (which contained at least 20 individual strains each) were considered: *Deinococcus-Thermus*, *Spirochaetes*, *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, *Thermotogae*, *Chlamydiae*, *Actinobacteria*, *Cyanobacteria*, *Planctomycetes* and *Fusobacteria*.

A total number of 3,707 strain groups (pooled from 6,187 individual strains) were analysed for this statistic. The mean number of LysM motif proteins or individual LysM motifs was calculated for all strains of a strain group (e.g. although *B. subtilis 168* has 18 LysM motif proteins with 30 LysM motifs, a mean of 15 LysM motif proteins with 26 LysM motifs per strain was calculated for the strain group *B. subtilis*).

The result of this analysis is presented in Figure 6.2 (a, b). This analysis showed that 97.4 % of strain groups have on average 1 – 10 LysM motif proteins; 2.6 % of strains groups have 11 – 20 LysM motif proteins while only 0.06 % of strains have more than 20 to 23 (the highest number) LysM motif proteins. The number of individual LysM motifs encoded per strain group shows a similar profile: 88.7 % of strain groups have on average a total of 1 - 10 LysM, while approximately 10 % of strain groups have a total of 11 - 20 LysM motifs; 1.2 % of strain groups had 21 - 30 LysM motifs and 0.2 % of strain groups had more than 30 LysM motifs. No strain group had more than 42 LysM motifs.

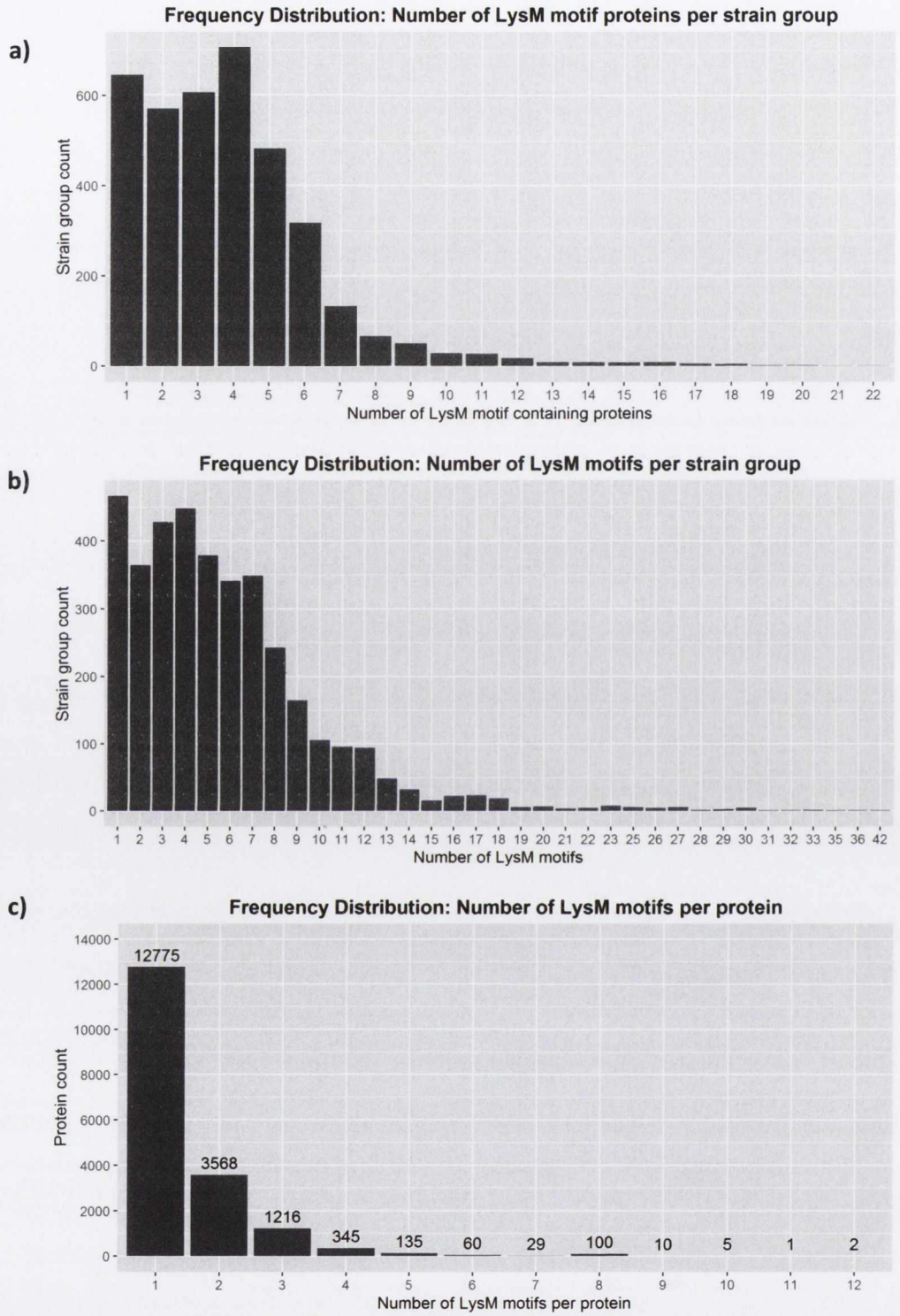



**Figure 6.2: Frequency distributions of the number of LysM motif proteins and individual LysM motifs per strain group, as well as the number of individual LysM motifs within proteins.** The two top bar graphs visualize **a)** the total number of LysM motif proteins per strain group and **b)** the total number of individual LysM motifs per strain group. The same dataset was used for both statistics. The dataset contains information to 6.187 individual strains, which were pooled into 3.707 strain groups (e.g. *B. subtilis 168* and *Bacillus subtilis XF-1* were united in the strain group *Bacillus subtilis*). **c)** A similar, but slightly more comprehensive dataset was used to calculate the frequency distribution that illustrates the number of LysM motifs within proteins. A total of 19.078 bacterial LysM motif proteins were compared for this statistic. A correction for the heaped occurrence of orthologous proteins of closely related species with the same number of LysM motifs was conducted.

In a second similar analysis the frequency of LysM motifs per protein was determined. A slightly different dataset was used for this statistic, because it was attempted to remove homologous proteins from the dataset. Homologous proteins were defined as those of identical length, name, number of LysM motifs and identical functions that are present in strains that belong to the same strain group. In total 18,246 proteins were analysed and the frequency of LysM motifs per protein is presented also in Figure 6.2 (c). Results show that ~67 % of proteins have one LysM motif; ~19 % of proteins have two LysM motifs; ~6 % of proteins have 3 LysM motifs and ~2 % of proteins have 4 LysM motifs. Proteins with five to eight LysM motifs are rare (~2) and proteins with nine to twelve LysM motifs are very rare (< 0.1 %). However, one protein had eleven LysM motifs (*Carnobacterium* U5S7Z8) and two proteins had twelve LysM motifs (e.g. *Carnobacterium* F4BMM5), the maximum found.

These results are supported by Bjorklund *et al.* (2006), who demonstrated that most bacterial proteins with protein motif repeats do not possess more than two repeats and that very few bacterial proteins contain more than 10 repeats of a protein motif. On the other hand, eukaryotic proteins were found to frequently possess more than 10 repeats and two more studies have shown previously that repeats of protein motifs principally occur more frequently in multi-cellular than in unicellular organism (Apic *et al.*, 2001; Ekman *et al.*, 2005). The fact that LysM motif repeats occur in all regular and irregular numbers indicates that there is no selection for certain units of LysM motifs that would suggest a functional or structural advantage or constraint.

In summary this analysis shows that the majority of strains (90 %) do not possess more than 6 LysM motifs proteins. It was further identified that 67 % of proteins have only a single LysM motif and that only 10 % of proteins have more than 2 LysM motifs.

#### 6.3.3.2 Function of LysM motif proteins

The dataset was further analysed to identify the function of LysM motif proteins, in so far as this was possible, considering the variance in the functional analysis of proteins in the database. As described previously homologues proteins were removed from the analysis (leaving 18,246 proteins for statistical analysis). Over half of these proteins (56.8 %) had no function ascribed to them. However, the frequency of ascribed functions to the remaining proteins is presented in Figure 6.3. The data show that a majority of the proteins have peptidase activity with the M23 family featuring prominently. A particularly notable feature is the high percentage of proteins whose enzyme activity is associated with cell wall function (e.g. transglycosylases, muramidases, amidases).

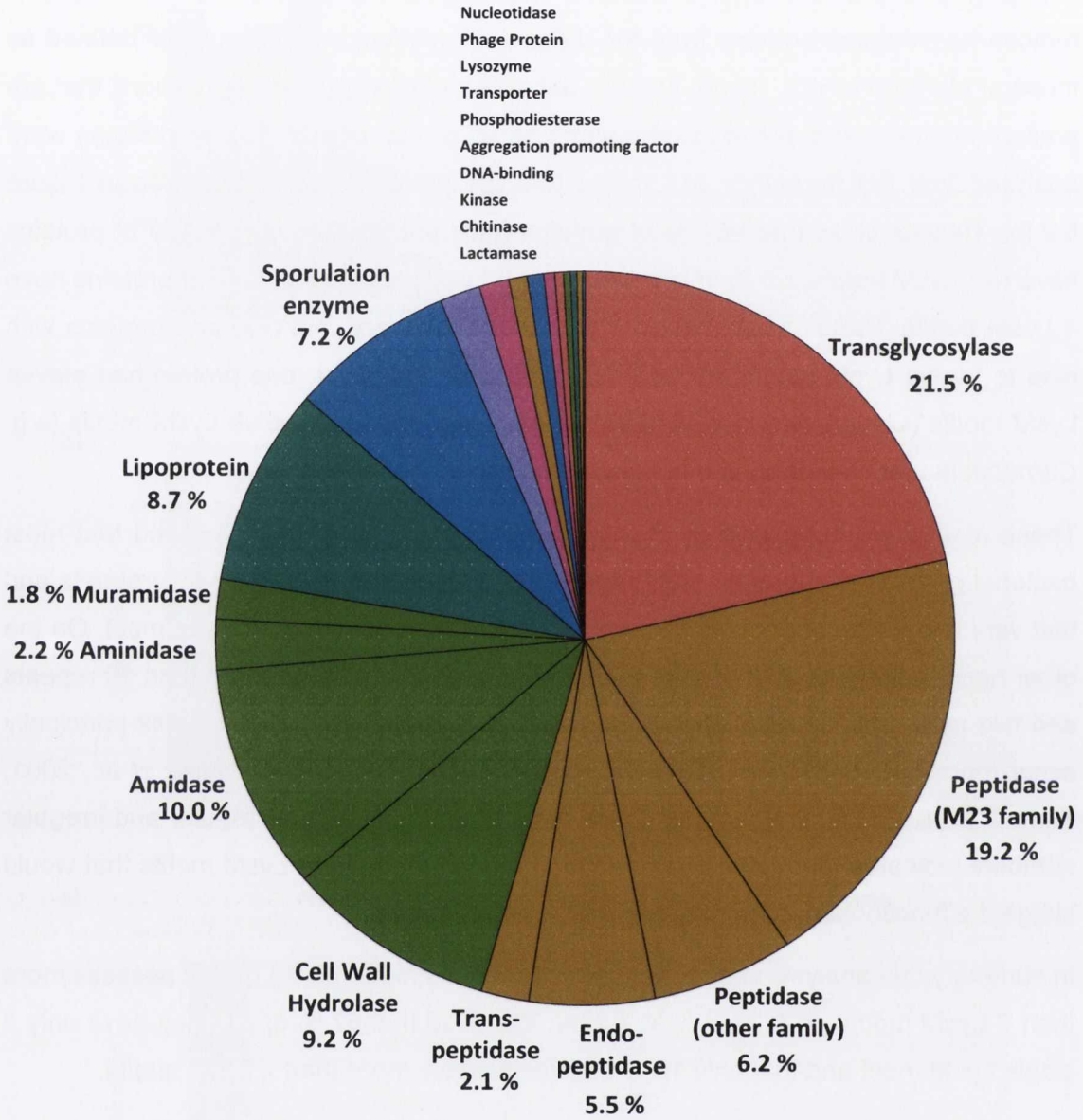



**Figure 6.3: Frequency distribution of the function of LysM motif proteins.** The schematic representation shows the predicted functions for LysM motif proteins. Of a total of 19.078 proteins, protein functions were predicted for 8.247 proteins. The pie chart represents the function of those 8.247 proteins. A correction for the heaped occurrence of orthologous proteins of closely related species with the same number of LysM motifs was conducted.

Also interesting is the occurrence of LysM motifs in proteins associated with bacteriophages. However, it is clear that the majority of activities of LysM containing proteins are relevant to its function, in binding to peptidoglycan. Of interest are also a few of the enzymatic functions associated with a minority of LysM motif proteins such as chitinases (possible binding to chitin cell walls of fungi), aggregation promoting factor (perhaps involved in biofilm formation) and DNA-binding. This result is confirmed by Buist *et al.* (2008) who found the LysM motif to be present in proteins with functions similar to the ones described here.

#### 6.3.3.3 Comparison of the frequency of LysM motif proteins among different bacterial families

The compiled dataset allowed the computation of summary statistics such as the frequency of LysM motifs within different bacterial subgroups. Although the same 3,707 strain groups (pooled from 6,187 individual strains) investigated in section 6.3.3.1 were originally considered in the analysis, the results presented only show 2,793 strain groups (pooled form 5.176 individual strains). This is, because bacterial families with less than 20 individual strains were not included in the analysis

Figure 6.4 and Table 8.1 (appendix, page 231) present the number of LysM motifs per strain group for 61 different bacterial families. The result shows that the *Bacillaceae* family had by far the highest number of LysM motifs per strain (median 15.5 LysM motifs per strain). The lowest median was one LysM motif per strain (for *Actinomycetaceae, Bartonellaceae, Bifidobacteriaceae, Caulobacteraceae, Gordoniaceae, Mycobacteriaceae, Veillonellaceae*). Most families possessed 2 ($Q_{0.25}$) to 7 ($Q_{0.75}$) LysM motifs per strain. The total number of strains analysed per bacterial family varied. The most strains were analysed for *Enterobacteriaceae* (655 strains, 176 strain groups) and the fewest for *Peptostreptococcaceae* (26 strains, 9 strain groups). The greatest standard deviation (sd) was computed for *Bacillaceae* (sd = 8.3), followed by *Desulfovibrionaceae* (sd = 6.9), *Ruminococcaceae* (sd = 5.2), *Peptococcaceae* (sd = 4.9) and *Lactobacillaceae* (sd = 4.4) while the standard deviation was mostly between 1 and 3 for all other bacterial families. A high stand deviation indicates a high diversity within a bacterial family. This is shown for instance in the *Bacillaceae* family (contains 135 strain groups, sd = 8.3) which encompasses strain groups with a high number of LysM motifs such as *Bacillus subtilis* (12 individual strains, average 25.75 LysM motifs per strain) and strain groups with a relatively low number of LysM motif such as *Bacillus cereus* (116 individual strains, average 8.06 LysM motifs per strain).

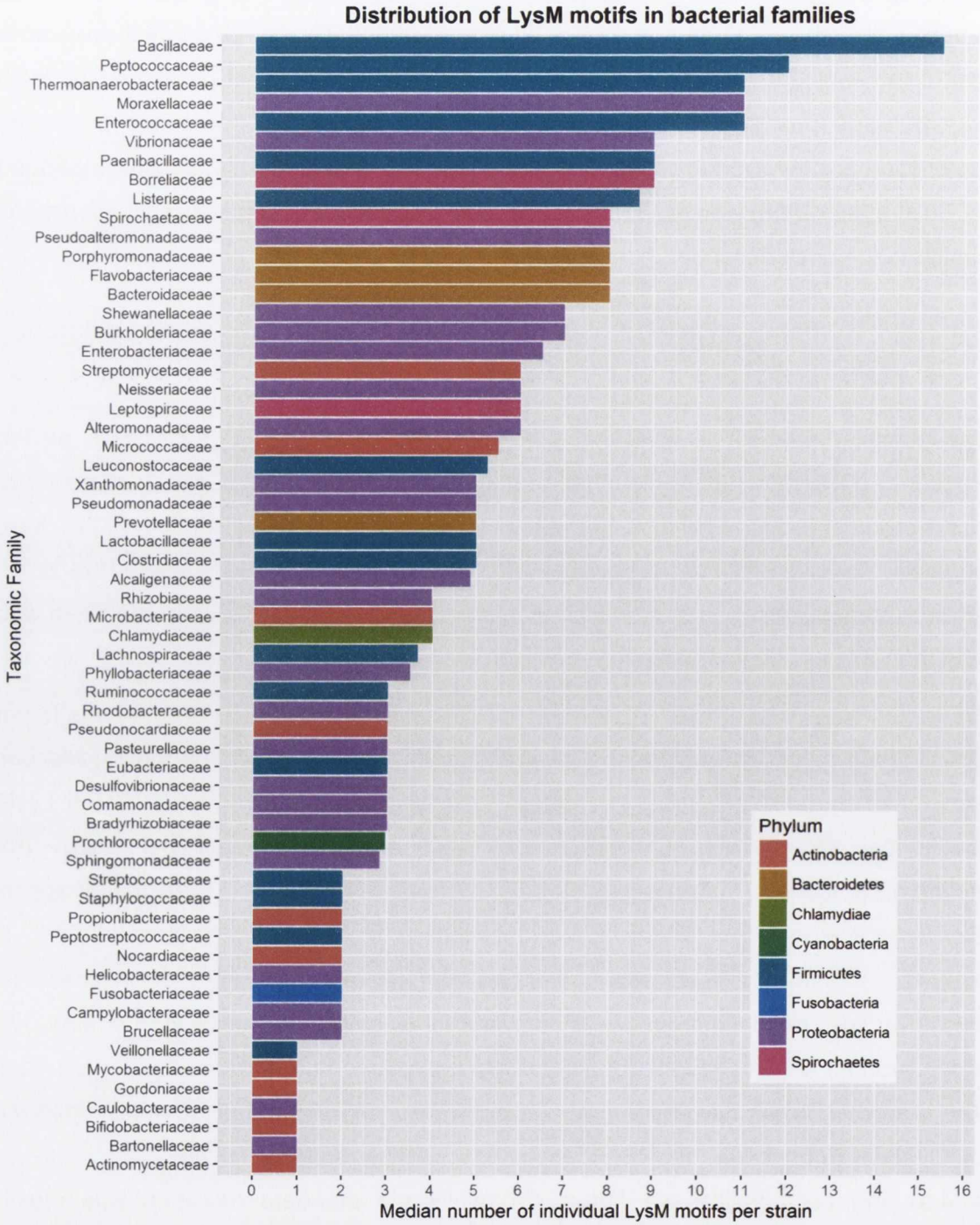



**Figure 6.4: Summary statistic of the frequency distribution of LysM motifs among different bacterial families.** The bar diagram shows the median number of LysM motifs per strain for different bacterial families. The median number of LysM motifs per strain was corrected for the cumulative occurrence of closely related strains that were pooled in 'strain groups' (e.g. *B. cereus AH187* and *B. cereus AH820* both belong to the strain group *B. cereus*). The colour indicates the bacterial phylum to which each family belongs.

It was observed earlier that LysM motif proteins rarely possess more than two LysM motifs (see Figure 6.2 c, page 192). For this reason it was further analysed which of the here described bacterial families possessed proteins with two or more LysM motifs. The result of this analysis is presented in Figure 6.5. LysM-rich proteins (defined as proteins that possess more than two LysM motifs) were mainly present in members of the bacterial families *Vibrionaceae* (average of 1.92 LysM motif rich proteins per strain), *Bacillaceae* (average of 1.86 motif rich proteins per strain), *Flavobacteriaceae* (average of 1.63 per strain), *Listeriaceae* (average of 1.50 per strain) and *Streptococcaceae* (average of 1.50 motif rich proteins per strain).

In a second step, an attempt was made to correlate the observed distribution of LysM motifs with characteristics of the investigated strains such as Gram-stain, habitat or genome size. It was hypothesized that bacteria with thick peptidoglycan cell walls (Gram-positive bacteria) would on average have more LysM motifs than bacteria with thin peptidoglycan cell walls (Gram-negative bacteria), given that LysM motifs bind to peptidoglycan. The grouping of bacterial strains into phyla allowed an approximate division into Gram-positive and Gram-negative bacteria, although this distinction is not always clear-cut. The investigated bacterial strains could be grouped in 8 different phyla: *Actinobacteria* (Gram-positive), *Bacteroidetes* (Gram-negative), *Chlamydiae* (presumed Gram-negative (Elwell *et al.*, 2016)), *Cyanobacteria* (combination of features (Hoiczyk and Hansel, 2000)), *Firmicutes* (mainly Gram-positive), *Fusobacteria* (Gram-negative), *Proteobacteria* (Gram-negative), *Spirochaetes* (distinctive diderm (Cullen *et al.*, 2004)). However, no correlation could be found between the number of LysM motifs and the Gram-stain. Gram-positive phyla were found to have many LysM motifs (e.g. *Firmicutes*, median 5 LysM motifs / strain group) as well as fewer LysM motifs (e.g. *Actinobacteria* median 3 LysM motifs / strain group). Similarly Gram-negative phyla had many LysM motifs (e.g. *Bacteroidetes*, median 7 LysM motifs / strain group) and fewer LysM motifs (e.g. *Fusobacteria*, median 2 LysM motifs / strain group). Interestingly members of the distinctive diderm (double-membrane) *Spirochaetes* phylum held on average the most LysM (8 LysM motifs / strain). Also even bacteria with presumed rudimentary cell walls such as *Chlamydiae* (Elwell *et al.*, 2016) possessed LysM motifs (median 4 LysM motifs / strain group).

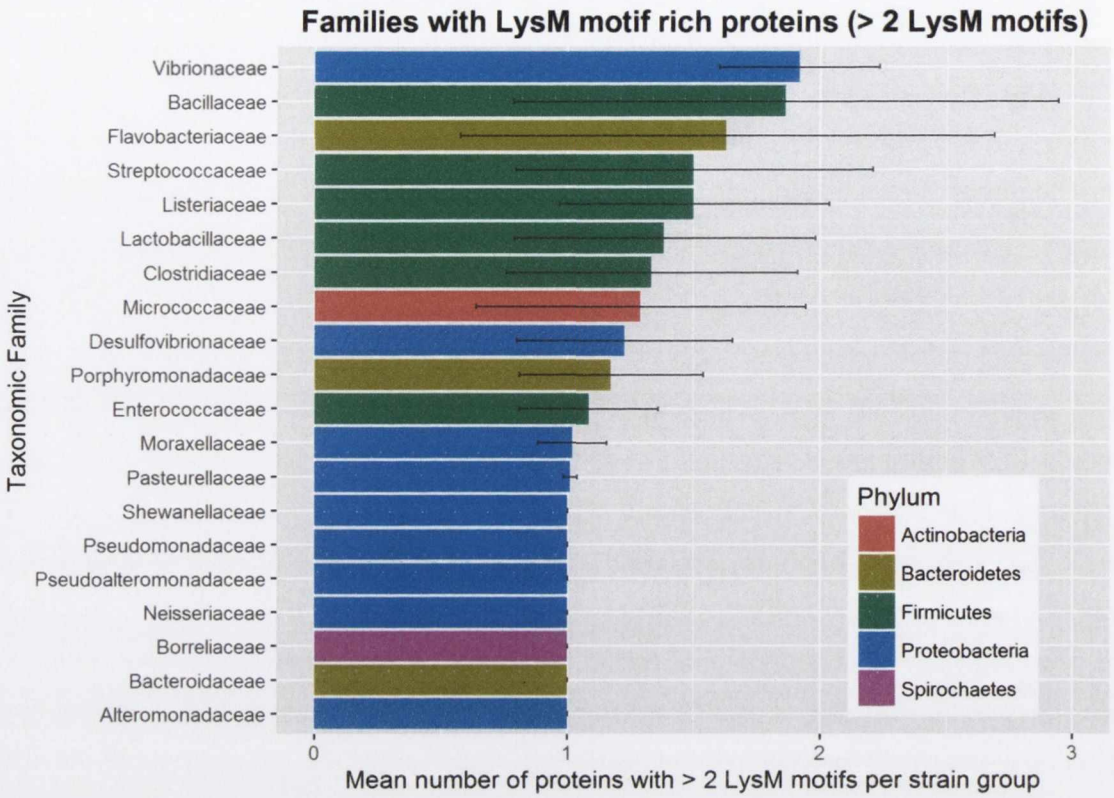



**Figure 6.5: Summary statistic of the frequency distribution of LysM-rich proteins (proteins with two or more LysM motifs) among different bacterial families.** The bar diagram shows the average number of LysM-rich motifs per strain for twenty different bacterial families (but only those strains were included in the analysis that possess at least one LysM-rich protein and families with less than 20 individual strains were not considered). The average number of LysM-rich proteins per strain was corrected for the cumulative occurrence of closely related strains that were pooled in 'strain groups' (e.g. *B. cereus* AH187 and *B. cereus* AH820 both belong to the strain *group B. cereus*). The colour indicates the bacterial phylum to which each family belongs. Standard deviation is shown as error bar.

It was further hypothesized that environmental bacteria, especially those living in the upper layer of the soil and in contact with the rhizosphere would have more LysM motif than pathogenic bacteria, for the reason that LysM motif can also bind to other polymers such as chitin (component of the cell wall of fungi) (Akcapinar *et al.*, 2015). However, it proved difficult to determine a connection between the number of LysM motifs and the habitat, since most bacterial phyla and families contain free-living environmental bacteria as well as pathogenic bacteria, however a small correlation may have been found. The first quartile of bacterial families (median LysM / per strain group > 7) has an excess of bacteria which are found environmentally while the approximate last quartile (median LysM / per strain group < 2.5) features mainly bacteria which are living as pathogens. Examples for bacterial families of environmental bacteria are: *Bacillaceae* (soil, water, 15.5 LysM / per strain), *Moraxellaceae* (soil, water, 11 LysM / per strain), *Vibrionaceae* (water, 9 LysM / per strain), *Spirochaetaceae* (water, 8 LysM / per strain). Examples for bacterial families mainly containing pathogenic bacteria are: *Staphylococcaceae* (staphylococcal infection, 2 LysM / per strain), *Brucellaceae* (brucellosis, 2 LysM / per strain), *Mycobacteriaceae* (tuberculosis and leprosy, 1 LysM / per strain), *Actinomycetaceae* (nocardiosis, 1 LysM / per strain), *Helicobacteraceae* (gastritis, 2 LysM / per strain) and *Bartonellaceae* (cat scratch disease, 1 LysM / per strain). However, this correlation is weak and pathogenic bacteria can also be found among families within the higher range as well as environmental bacteria are found among families of the lower range. Furthermore, many bacterial families (e.g. *Bacillaceae*) contain bacteria found both environmentally (*Bacillus subtilis*) as well as pathogens (*Bacillus anthracis*), which makes distinctions difficult. Likewise Jernigan and Bordenstein (2015) found no correlation between the occurrence of tetratricopeptide motifs repeats (associated with virulence) or armadillo repeats and the lifestyle or pathogenicity of bacteria.

Furthermore, the possibility was explored that the number of LysM motifs is in relation to the genome size of a bacterium. Bacteria that live environmentally generally encounter a multitude of physical conditions and nutrient availability that require a larger genome that encodes signal transduction systems and metabolic enzymes, while pathogenic bacteria live in more restricted environments (smaller genome). For example, bacteria of the family *Chlamydiae* grow by infecting eukaryotic host cells and have substantially reduced genomes (Elwell *et al.*, 2016). Reshef *et al.* (2010) suggest that the expansion of protein repeats has the potential to significantly alter proteome size and that for this reason the number of repeats will be tightly regulated in bacteria.

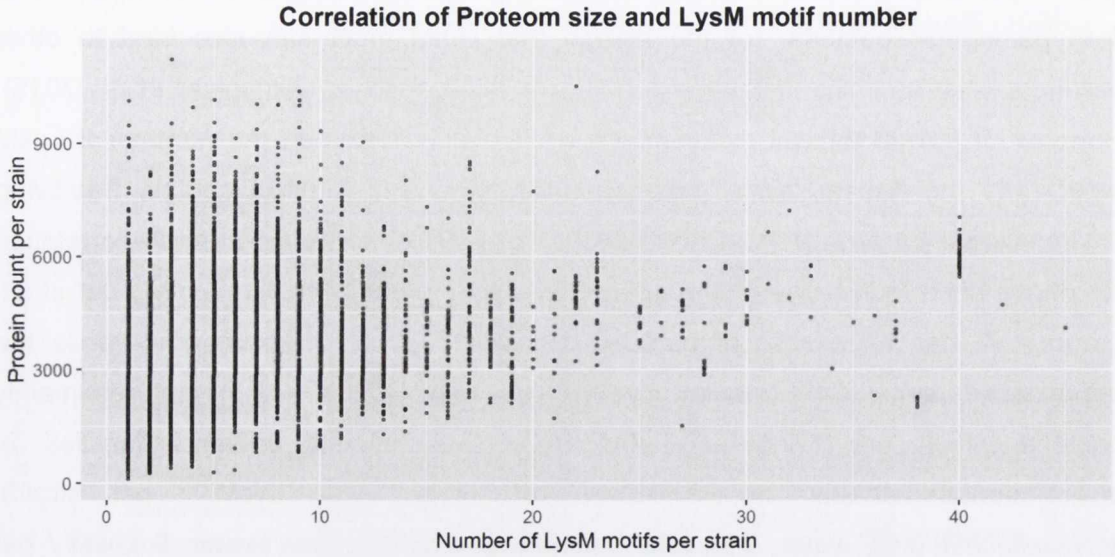



**Figure 6.6: Correlation between proteome size and the number of LysM motifs per strain.** The scatter plot visualizes the correlation between the proteome size (in total protein count) and the number of LysM motif in a strain. A total of 6,730 strains were compared for this statistic.

Therefore, bacteria with a large genome may well encode more LysM motifs than bacteria with a small genome. The proteome size could be determined for 6,730 of 6,971 strains in the dataset. However, no correlation between the proteome size and the number of LysM motifs per species was found (Figure 6.6). This is a contrary result to Bjorklund *et al.* (2006) who showed that more complex organisms with a larger proteome size contain more protein motif repeats. However, in the study conducted by Bjorklund *et al.* (2006) a greater spectrum of genome sizes was compared (prokaryotes vs eukaryotes).

It can be concluded that in comparison with other bacterial species *B. subtilis 168* is highly unusual in several regards: (i) It possesses an extremely high number of LysM motif containing proteins (18 LysM motif proteins) and individual LysM motifs (30 LysM motifs). (ii) It holds three LysM-rich proteins with more than two LysM motifs (LytE, CwlS, LytF). And (iii) it is member of the *Bacillaceae* family which has on average the most LysM motifs per strain (median 15.5). Although no correlation was found between the Gram-strain and the frequency of LysM motifs within a strain, the high number of LysM motifs in *B. subtilis* might be related to its natural habitat. Some evidence was found that environmental bacteria have on average more LysM motifs than pathogens. It is further plausible that bacteria that grow associated with the roots of plants and fungi such as *B. subtilis 168* (Vlamakis *et al.*, 2013), use LysM motif proteins to bind to the polysaccharides found in this environment. Although studies have shown that LysM motifs are unlikely to bind to cellulose (the main component of the cell wall of plants) (Akcapinar *et al.*, 2015; Mesnage *et al.*, 2014), they bind to chitin and chitin-like compounds (that make up the cell wall of fungi) (Akcapinar *et al.*, 2015; Buist *et al.*, 2008). Therefore, possessing a multitude of LysM motif proteins could be an adaptation of *Bacillus subtilis* to growth in this environment.

#### 6.3.3.4 Duplicated LysM motifs

The amino acid sequence data of all 38,990 LysM motifs present in the compiled dataset was used to screen for LysM motifs with identical amino acid sequence. Of 38.990 listed LysM motifs the amino acid sequence of 46 LysM motifs occurred more than once, with 107 instances in total. All instances of duplicated motifs are listed in Table 8.2 (appendix, page 234). The majority of these duplicated motifs occurred twice (40 motifs). Three motifs were duplicated three times each (within the proteins A8MI50, A8UCR7 and E6LIK0). One motif was duplicated five times (within the protein U5S7Z8), one motif was duplicated six times (within the protein D4G4L0) and one motif was duplicated seven times (within the protein F4BMM5). The instances of duplicated LysM motifs all occurred within one protein. No two proteins were found that shared

identical amino acid sequence for a LysM motif. Therefore, of a total of 26.388 LysM motif containing proteins, 46 proteins possessed two or more LysM motifs with identical amino acid sequence which represents a share of 0.17 %.

Furthermore, in a majority of cases (97 out of 107) the duplicated LysM motifs were in juxtaposition with each other, as illustrated in Figure 6.7 (a). Only four proteins contained duplicated LysM motifs that were separated by another LysM motif (with different sequence). In the proteins A0A0E1YD61, F8C3Z4, J6DEL2 and J6NW46 respective, a duplicated LysM motif is present on positon 1 and 3 (see Figure 6.7 b / c). Although in some instances this could indicate that all three LysM motifs (at position one, two and three) originally possessed the same amino acid sequence and that the amino acid sequence of the centric LysM motifs was rendered through a recent mutation event, as seems likely in the case of the protein J6DEL2 (see Figure 6.7 b) the protein A0A0E1YD61 does not fit this pattern (see Figure 6.7 c). A0A0E1YD61 possesses six LysM motifs in total, of which two (at position 1 and 3) share the same amino acid sequence. However, the intermediate LysM motif at position 2 does not possess great sequence similarity to either one of the LysM motifs with which it is flanked, but instead is closely related to the LysM motif on position 4. Therefore, it seems possible that LysM motif duplication events can also skip one or more existing motifs.

It was further noted that LysM motif duplication events take place in proteins with few as well as with many LysM motifs. However, it seems that proteins with many LysM motifs are more likely to contain several instances of a duplicated motif. For example as noted earlier the occurrence of proteins with nine, ten, elven or twelve LysM motifs is extremely rare (< 0.1 %) and only one instance of a protein with eleven LysM motifs and two instances of a protein with twelve LysM motifs were identified in this study (see Figure 6.2 c, page 192). However, all three of these proteins possess duplicated LysM motifs, which would amount to an extraordinary coincidence, if assumed that LysM motif duplication events occurred at a random rate.

Only one protein (with more than 2 LysM motifs in total) was identified in which all LysM motifs share the same amino acid sequence. This is the protein D4G4L0 of the strain *Bacillus subtilis subsp. natto BEST195* in which all six LysM motifs have identical amino acid sequence (see Figure 6.7 d). Interestingly these six LysM motifs are also highly similar (although not identical) on a nucleotide level. This may indicate that LysM motif duplication events can create several repeats of a single motif at once, leading to the rapid expansion of a motif.

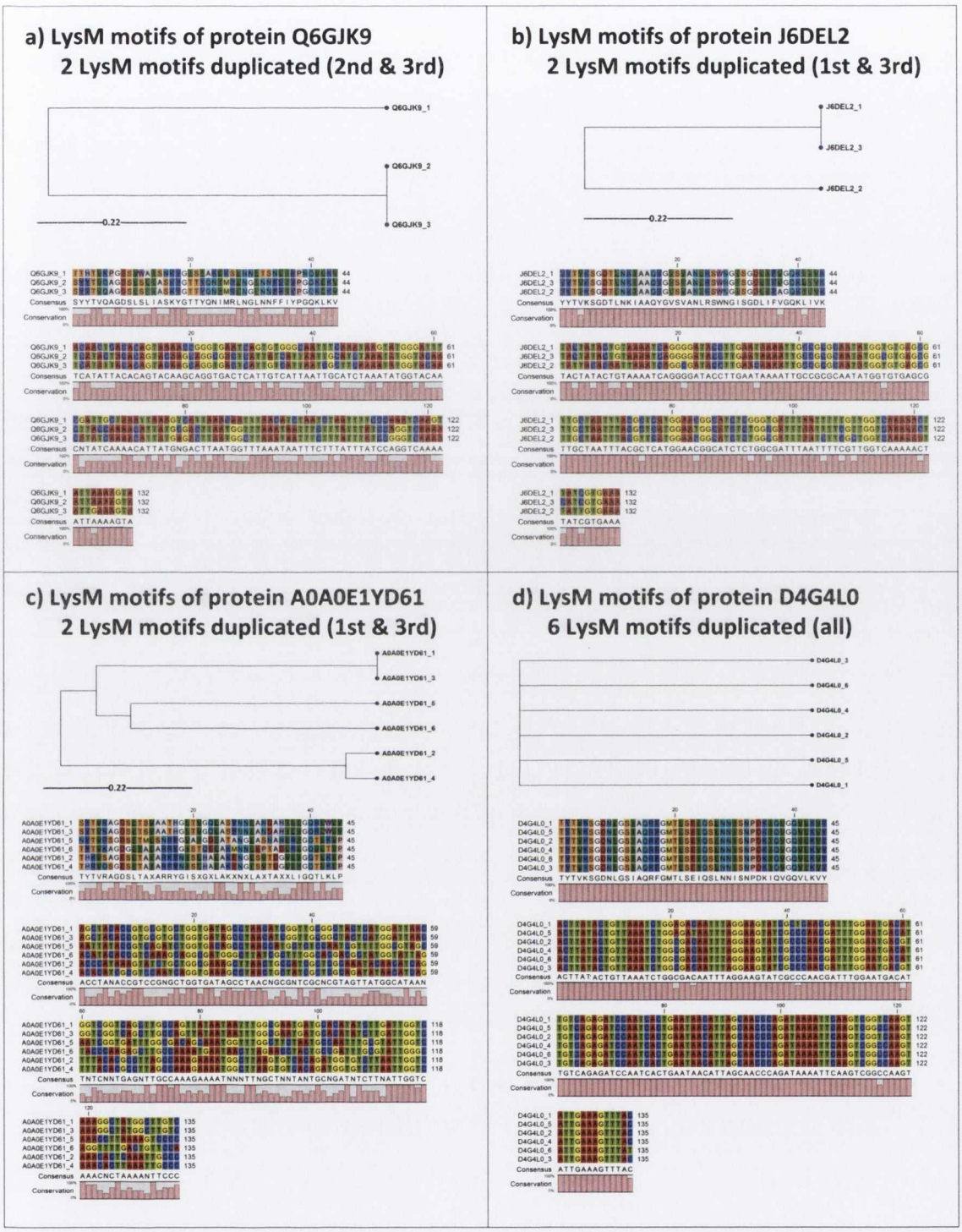



**Figure 6.7: Four examples of proteins that contain LysM motifs with identical amino acid sequence.** The amino acid and nucleotide sequence alignments (constructed with Clustal Omega) of all LysM motifs of the respective protein are shown as well as a phylogenetic tree that was derived from the amino acid sequence (Distance measure = Jukes-Cantor). **a)** Protein Q6GJK9 (N-acetylmuramoyl-L-alanine amidase sle1) of *Staphylococcus aureus* (*strain MRSA252*), 2 motifs of 3 motifs in total have identical amino acid sequence (motif 2 and 3). **b)** Protein J6DEL2 (Putative gametolysin) of *Enterococcus faecalis* (*strain ERV65*), 2 motifs of 3 motifs in total have identical amino acid sequence (motif 1 and 3. **c)** Uncharacterised protein A0A0E1YD61 of *Moraxella catarrhalis* (*strain CO72*), 2 motifs of 6 motifs in total have identical amino acid sequence (motif 1 and 3). **d)** Uncharacterised protein D4G4L0 of *Bacillus subtilis subsp. natto* (*strain BEST195*), 6 motifs of 6 motifs in total have identical amino acid sequence. Motifs are numbered starting from the $NH_2$-terminal motif.

Further evidence in favour of this explanation is brought forward by Bjorklund *et al.* (2006) who investigated the expansion of a multitude of protein motifs in eukaryotic and bacterial proteins and found that duplications of several protein motifs at a time is common. Bjorklund *et al.* also suggested that expansion can occur in distinct units of domains (such as 2 units at a time) which possibly explains the observed sequence similarity of LysM motif 1 and 3 and LysM motif 2 and 4 in the protein A0A0E1YD61. Assuming that the ancestral protein to A0A0E1YD61 originally possessed two LysM motifs it seems likely that these two motifs were duplicates as a unit and the copy inserted behind the two original motifs, leading to the observed sequence similarities.

Interestingly the majority of duplicated LysM motifs were present within the Phylum *Firmicutes* (79.4 %), especially in *Enterococcus faecalis*. This is a much higher percentage than could be expected, provided that duplicated motifs occurred in all phyla at the same rate (LysM motifs of the *Firmicutes* phylum represent 32.8 % of all investigated bacterial LysM motifs), suggesting that LysM motif expansion is either a more recent or more frequent event in species of the *Firmicutes* phylum.

It can be concluded that motif duplication is one option in which LysM motif protein amplify the number of their LysM motifs. Given the interesting observation that all six LysM motifs of the protein D4G4L0 share identical amino acid sequence and that the only identified proteins with more than ten LysM motifs also possess duplicated LysM motifs, it seems that LysM motif duplication can create several repeats in one duplication event, leading rapidly to LysM-rich proteins. The fact that no instance of identical LysM motifs occurring in different proteins was observed indicates that LysM motif containing proteins that possess similar LysM motifs (such as CwlS, LytE and LytF of *B. subtilis*) are likely to have a shared common ancestor.

## 6.4 An evolutionary model for the development of CwlS, LytE and LytF

### 6.4.1 Objective

It was observed earlier that the three LysM motifs of LytE are functionally distinct and that the first LysM motif (N-terminal) seems to confer a higher peptidoglycan-binding activity to a variant LytE protein with only one LysM motif than the second or the third LysM motif of LytE (Chapter 5). Furthermore, cluster analysis of all LysM motifs of *B. subtilis* yielded the interesting observation that although the 12 LysM motifs of the three LysM motif containing D,L-endopeptidases (CwlS, LytE and LytF) have a high

sequence similarity, the first LysM motifs of each enzyme were more similar to each other than to any of the consecutive LysM motifs within the respective enzyme (see Figure 6.1, page 188).

This raises interesting questions about the evolution of the LysM motif configuration in these three enzymes. Since they also have highly similar enzymatic domains (for an alignment see Figure 1.5, page 26), it seems likely that the three enzymes evolved through gene duplication and that the ancestral enzyme contained at least one LysM motif. However, the manner in which these three closely related enzymes evolved to have different numbers of LysM motifs is unknown. Therefore, this study attempts to unravel the evolutionary relationship between the CwlS, LytE and LytF LysM motif containing D,L-endopeptidases of *B. subtilis 168*. The focus of the study is to investigate the evolutionary mechanism by which they came to have different numbers of LysM motifs. This question was addressed by identifying orthologs of CwlS, LytE and LytF in other bacterial species that are closely related to *B. subtilis 168* and by comparing their LysM motifs through cluster analysis. A model for the evolutionary development of these three D,L-endopeptidases in *B. subtilis* is presented.

### 6.4.2 Approach

To identify orthologs to CwlS, LytE and LytF in other bacterial species, only closely related strains to *B. subtilis 168* were included in the analysis. UniProt (UniProt, 2015) lists the taxonomic lineage of the stain *B. subtilis 168* (scientific name: *Bacillus subtilis subsp. subtilis*) as shown in Figure 6.8. This taxonomy classification is maintained at the NCBI (National Centre for Biotechnology Information (Sayers *et al.*, 2009)), which states that its taxonomic classification is mainly built from sequence data information.

| | |
|---|---|
| › SUPERKINGDOM: | Bacteria |
| › PHYLUM: | Firmicutes |
| › CLASS: | Bacilli |
| › ORDER: | Bacillales |
| › FAMILY: | Bacillaceae |
| ›GENUS: | Bacillus |
| › SPECIES GROUP: | Bacillus subtilis group |
| › SPECIES: | Bacillus subtilis |
| › SUBSPECIES: | Bacillus subtilis subsp. subtilis |

**Figure 6.8: Full taxonomic lineage of *B. subtilis 168***

For the purpose of this study, all species listed in the same species group as *Bacillus subtilis 168* were considered to be closely related. Twenty strains, which are listed in Table 6.2, are included in this group. Figure 6.15 (b) (page 220) presents a basic phylogenetic tree, which was obtained from the NCBI taxonomy database and illustrates the relationship between these strains.

**Table 6.2: The 20 strains of the *Bacillus subtilis* species group**

| **Strains of the *Bacillus subtilis* species group** | |
|---|---|
| • *Bacillus amyloliquefaciens DSM 7* | • *Bacillus subtilis BEST7613* |
| • *Bacillus amyloliquefaciens Y2* | • *Bacillus subtilis subsp. inaquosorum KCTC 13429* |
| • *Bacillus atrophaeus 1942* | • *Bacillus subtilis subsp. natto BEST195* |
| • *Bacillus licheniformis CG-B52* | • *Bacillus subtilis subsp. niger* |
| • *Bacillus licheniformis DSM 13* | • *Bacillus subtilis subsp. spizizenii TU-B-10* |
| • *Bacillus licheniformis S 16* | • *Bacillus subtilis subsp. spizizenii W23* |
| • *Bacillus licheniformis WX-02* | • *Bacillus subtilis subsp. subtilis SC-8* |
| • *Bacillus paralicheniformis ATCC 9945a* | • *Bacillus subtilis XF-1* |
| • *Bacillus sonorensis L12* | • *Bacillus velezensis FZB42* |
| • *Bacillus subtilis 168* | • *Bacillus velezensis UCMB5033* |

All LysM motif proteins (with N-terminal LysM motifs) of these twenty strains were identified with the help of the dataset constructed in section 6.3.2. (a total of 161 proteins). Every strain had on average 16 LysM motif proteins with 27 LysM motifs, but exact numbers varied between strains. The 161 proteins were analysed by cluster analysis to determine orthologs of the CwlS, LytE and LytF D,L-endopeptidases of *B. subtilis 168* by the level of identity shown by their catalytic domains.

Two complementary methods of cluster analysis were conducted, a phylogenetic tree and a metric multi-dimensional cluster analysis. Both methods required a amino acid sequence alignment as input, which was then used to determine similarities between sequences. Therefore, the same multiple sequence alignment was used for both types of analysis. The amino acid sequence alignment was achieved with the ClustalOmega algorithm (Sievers *et al.*, 2011). BLOSUM80 was used as a substitution matrix, based on the fact that BLOSUM80 is optimized for the comparison of more related proteins and is based on local alignments. The default settings were adapted for all other parameters including the gap opening and extension penalty.

The so generated multiple sequence alignment was then deployed to construct a phylogenetic tree and to conduct a multi-dimensional cluster analysis. Both methods initially compute a distance matrix from the multiple sequence alignment. The distance matrix is then used to create the phylogenetic tree or is visualized in a plane by metric multidimensional scaling. The main difference between these two methods is that the tree methods follows a hierarchical, binary approach to the data and compares two clusters at a time. On the other hand, the space-method allows a multi-variant analysis of the data and considers all sequences at once. This method projects the sequences as dots on a plane in which the visualized distances best approximate the original distances of the distance matrix. The phylogenetic tree was constructed with the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) algorithm. The metric multidimensional scaling was conducted with the bios2mds package (Pele *et al.*, 2012) in R. Following the identification of the orthologs the individual LysM motif complements of all orthologs were compared to each other by cluster analysis. Subsequently a potential model for the evolutionary development of the three D,L-endopeptidases was developed.

### 6.4.3 Results

#### 6.4.3.1 Identification of orthologs to CwlS, LytE and LytF by their catalytic domain

The first step of the analysis was to identify orthologs to CwlS, LytE and LytF of *B. subtilis 168* among the 161 LysM motif containing proteins of the twenty strains of the *Bacillus subtilis* species group.

Figure 6.9 shows the results of the analysis that led to the identification of orthologs by their catalytic domain. It is evident that orthologs to CwlS, LytE and LytF form a separate cluster in the space-approach (Figure 6.9 a) as well as in the tree-based approach (Figure 6.9 c). The bootstrapping values (between 60 and 100) indicate a medium to high reproducibility of the tree. The tree and space-based analysis were then repeated on the group of proteins presumed to be orthologous (i.e. non-orthologs excluded). In the second analysis both metric multidimensional scaling (Figure 6.9 b) and the phylogenetic tree (Figure 6.9 d) showed a clear distinction between the three orthologs. All identified orthologs are listed in Table 8.3 (appendix, page 236). Interestingly, although an ortholog to LytE and LytF was identified in every strain, six out of twenty strains did not possess a CwlS ortholog. The catalytic domains of this enzyme cohort exhibited 80 – 85% identity at the amino acid level.

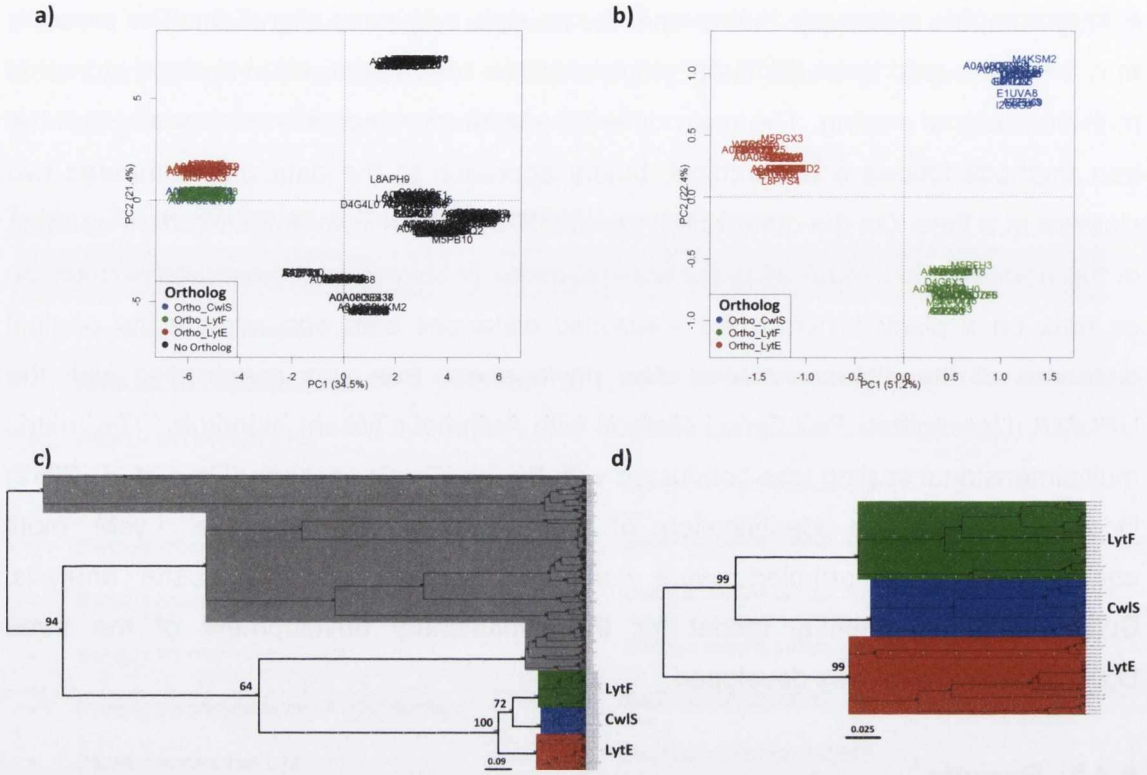



**Figure 6.9: Identification of orthologs to LytE, CwlS and LytF in the *B. subtilis* strain group by amino acid sequence analysis of their catalytic domains.** The amino acid sequence of the catalytic domains of all LysM motif proteins with N-terminal localized LysM motif among proteins of the *B. subtilis* strain group were aligned (Algorithm = Clustal Omega, Substitution Matrix = "BLOSUM80) and the alignment visualized by **a) / b) Metric multidimensional scaling.** The catalytic domains are mapped in a plane formed by the first (PC1) and second (PC2) component of the sequence space by MDS analysis with the distances based on the difference score. There is no unit for the principal component axis, but the percentage given between brackets indicates the percent of variance explained by this component and thus the relative weight of this component. **c) / d) Phylogenetic tree.** The phylogenetic tree was obtained with the UPGMA algorithm. Distance measures are given in Jukes-Cantor units. Bootstrap values are shown (1000 Replicates). The colours correspond to three orthologs (Red = LytE ortholog, Green = LytF ortholog, Blue = CwlS ortholog). Black and Grey indicate proteins that were identified not to represent an ortholog to LytE, LytF or CwlS.

Furthermore, alignments showed that this enzyme cohort shows 78 – 80% identity when the whole protein sequence (LysM and catalytic domains) was considered. These data also suggest that the catalytic domains of CwlS and LytF orthologs are more similar to each other than to the catalytic domain of the LytE ortholog. These results support the view that the enzymes of each group listed in Table 8.3 (appendix, page 236) are true orthologs of the LytE, LytF and CwlS enzymes.

### 6.4.3.2 The LysM motifs complement of CwlS, LytE and LytF orthologs

The complement of LysM motifs present in the cohort of CwlS, LytE and LytF orthologs were identified and the number of LysM motif that each protein possesses is listed in Table 8.3 (appendix, page 236). It is indeed interesting that some of the constituent proteins of each group differed in the number of LysM motifs (Figure 6.10). The 20 orthologous LytE proteins have either two or three LysM motifs. Of the 14 CwlS orthologs identified, 13 had four LysM motifs while the protein encoded in strain *Bacillus subtilis XF-1* had only three LysM motifs. The complement of 20 LytF orthologs identified, had either four or five LysM motifs.

The 20 strains under investigation could be categorized into 4 groups (A – D) on the basis of D,L-endopeptidase composition and LysM motif distribution as shown in Table 6.3 (see page 211).

The subdivision of the twenty strains into these groups correlates to some extent with the basic phylogenetic tree presented in Figure 6.15 (b) (see page 220) that exemplifies the relationship between those strains. This is especially true for members of group C which form a separate cluster that holds the two *Bacillus subtilis subsp. spizizenii* strains as well as the two *Bacillus subtilis subsp. subtilis* strains and *Bacillus subtilis subsp. inaquosorum* KCTC 13429, *Bacillus subtilis subsp. natto* BEST195 and *Bacillus subtilis* BEST7613. However, this cluster also contains on member of group B (*Bacillus subtilis subsp. niger*) and the only strain that has a CwlS ortholog with three LysM motifs (*Bacillus subtilis XF-1*, Group D).

Furthermore, all four *Bacillus licheniformis* strains, which are part of group A, form a separate cluster. Two small clusters each hold the two *Bacillus velezensis* strains and the two *Bacillus amyloliquefaciens* strains (all part of group B).

Therefore, it seems that the grouping system devised on the basis of D,L-endopeptidase composition and LysM motif distribution reflects to some extent the evolutionary relationship between those strains.

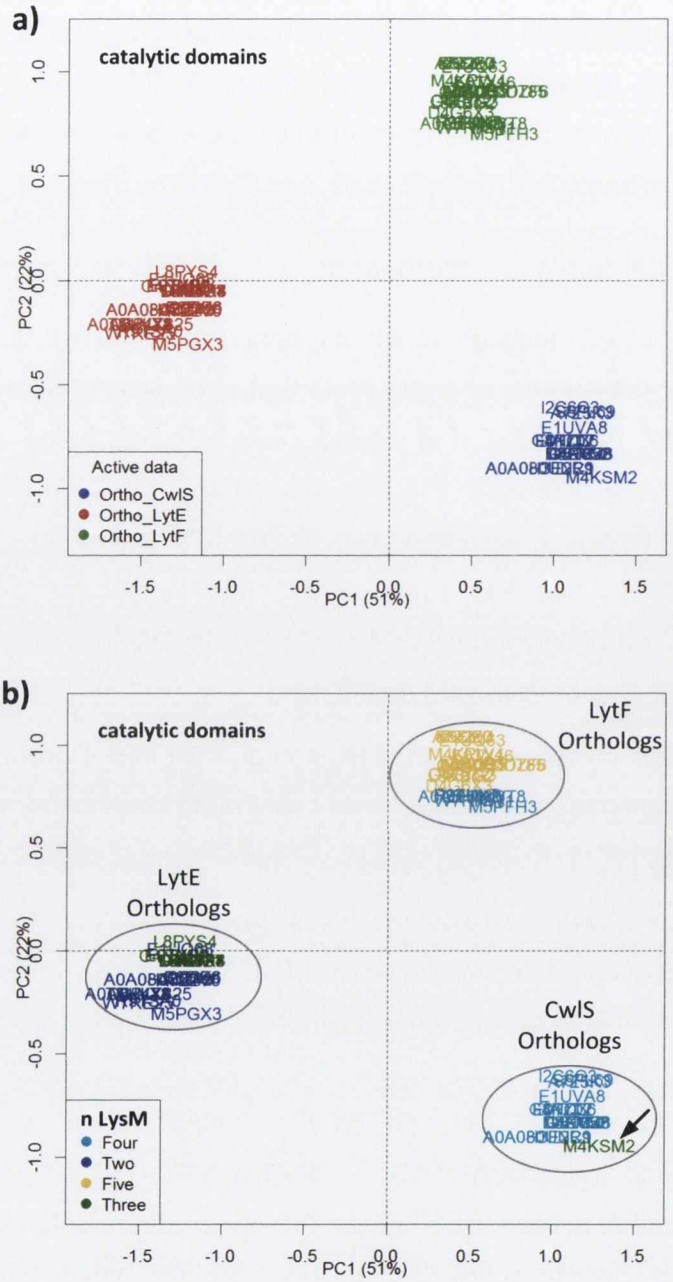



**Figure 6.10: Metric multidimensional scaling of the catalytic domains of the LytE, CwlS and LytF orthologs.** The amino acid sequences of the catalytic domains of all orthologs were aligned (Algorithm = Clustal Omega, Substitution Matrix = "BLOSUM80) and the alignment visualized by metric multidimensional scaling. The catalytic domains are mapped in a plane formed by the first (PC1) and second (PC2) component of the sequence space by MDS analysis with the distances based on the difference score. There is no unit for the principal component axis, but the percentage given between brackets indicates the percent of variance explained by this component and thus the relative weight of this component. Three clusters are visible in **(a)** that represent the three orthologs (LytE orthologs, CwlS orthologs and LytF orthologs). The colours in **(b)** highlight the total number of LysM motifs (n LysM) that each ortholog protein possesses (Blue = Two, Green = Three, Turquoise = Four, Yellow = Five). Interestingly the protein M4KSM2 (indicated by arrow) is the only CwlS ortholog with three LysM motifs.

**Table 6.3: Categorisation of strains into 4 groups (A, B, C, D) on the basis of D,L-endopeptidase composition and LysM motif distribution**

| D,L-endopeptidase composition | Strains |
|---|---|
| **Group A** | |
| **one LytE-type enzyme (2 LysM)**<br>**no CwlS-type enzyme**<br>**one LytF-type enzyme (4 LysM)** | • *Bacillus licheniformis DSM 13*<br>• *Bacillus licheniformis WX-02*<br>• *Bacillus licheniformis CG-B52*<br>• *Bacillus paralicheniformis ATCC 9945a*<br>• *Bacillus licheniformis S 16*<br>• *Bacillus sonorensis L12* |
| **Group B** | |
| **one LytE-type enzyme (2 LysM)**<br>**one CwlS-type enzyme (4 LysM)**<br>**one LytF-type enzyme (5 LysM)** | • *Bacillus amyloliquefaciens DSM 7*<br>• *Bacillus amyloliquefaciens Y2*<br>• *Bacillus atrophaeus 1942*<br>• *Bacillus velezensis FZB42*<br>• *Bacillus velezensis UCMB5033*<br>• *Bacillus subtilis subsp. niger* |
| **Group C** | |
| **one LytE-type enzyme (3 LysM),**<br>**one CwlS-type enzyme (4 LysM)**<br>**one LytF-type enzyme (5 LysM)** | • *Bacillus subtilis 168*<br>• *Bacillus subtilis BEST7613*<br>• *Bacillus subtilis subsp. inaquosorum KCTC 13429*<br>• *Bacillus subtilis subsp. natto BEST195*<br>• *Bacillus subtilis subsp. spizizenii W23*<br>• *Bacillus subtilis subsp. spizizenii TU-B-10*<br>• *Bacillus subtilis subsp. subtilis SC-8* |
| **Group D** | |
| **one LytE-type enzyme (3 LysM)**<br>**one CwlS-type enzyme (3 LysM)**<br>**one LytF-type enzyme (5 LysM)** | • *Bacillus subtilis XF-1* |

### 6.4.3.3 Cluster analysis of individual LysM motifs of all CwlS, LytE and LytF orthologs

There are 197 LysM motifs contained in the complement of LytE-type (20), LytF-type (20) and CwlS-type (14) enzymes encoded in the 20 *Bacillus* strains listed in Table 8.3 (appendix, page 236). The relationships between these 197 LysM motifs was determined by cluster analysis as described in section 6.4.3.1 The resultant phylogenetic tree is presented in Figure 6.11 and the metric multidimensional scaling is presented in Figure 6.12.

Both types of analysis show that individual LysM motifs cluster according to their position within the protein and not according to enzyme identity (i.e. LytE, LytF and CwlS). This is consistent with the result of the more restricted analysis that was performed with only the 12 LysM motifs of the CwlS, LytE and LytF enzymes of *B. subtilis* strain 168 (see Figure 6.1, page 188). The phylogenetic tree (Figure 6.11) therefore has three main clusters: LysM motifs at position one, LysM motifs at positions two and three and LysM motifs at positions four and five. Interestingly the cluster of LysM motifs at position two & tree and four & five are more similar to each other than to the cluster of LysM motifs at position one. A bootstrap value of > 68 supports the veracity of the observation. It is evident that there are sub-clusters within each of the main clusters (i.e. clusters LysM_1, LysM_2 & LysM_3 and LysM4_ & LysM_5) that group according to enzyme type. For example there are three subgroups within Group 1 clade (LysM_1), one for each enzyme type, within the cluster of LysM_1 motifs. Thus, the LysM_1 motifs of each of the CwlS orthologs are more similar to each other, than to the LysM_1 motifs of LytE or LytF. Furthermore, it is apparent that the LysM_1 motifs of LytF and CwlS are more similar to each other than either is to LysM_1 of LytE.

The subgroups within the Group 2 clade (LysM_2 and LysM_3) are less well defined, as shown by significantly lower bootstrap values. However, accepting the distinctions as presented in Figure 6.11, it is clear that LysM_2 of LytF and CwlS are closely related to each other while LysM_2 of LytE is most closely related to LysM_3 of LytE and more closely related to LysM_3 of CwlS and LytF than to the LysM_2 domains of LytF and CwlS.

The subgroups within the Group 3 clade (LysM_4 and LysM_5) are also less well defined and show a distinction between the LysM_4 of CwlS and the majority of the LysM_4 of LytF. However, some of the LysM_4 of LytF are more similar to the LysM_4 of CwlS as well as to the LysM_5 of LytF.

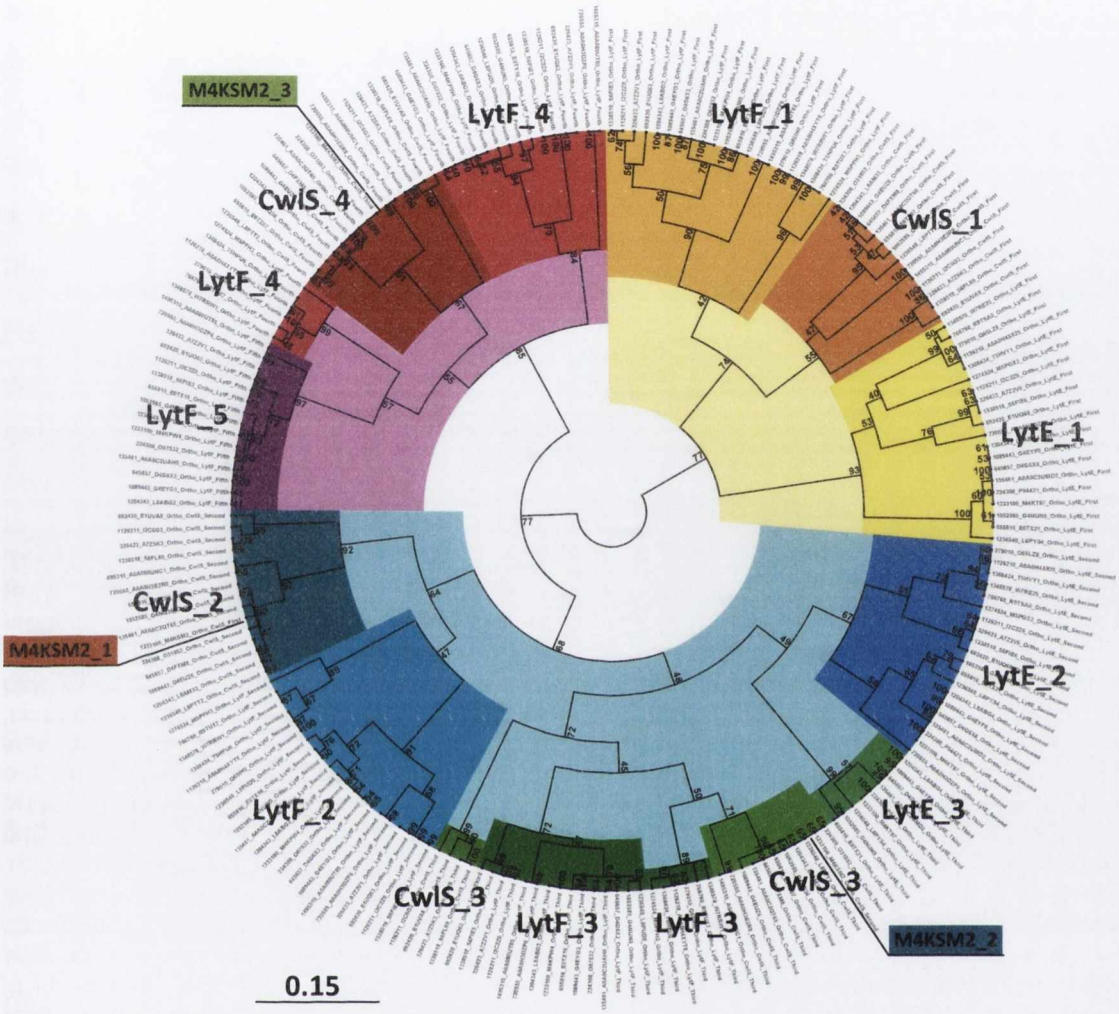



**Figure 6.11: Circular phylogenetic tree of the LysM motifs of CwlS, LytE and LytF orthologs.** The amino acid sequence of the LysM motifs of the CwlS, LytE and LytF orthologs of the *B. subtilis* species group were aligned (Algorithm = Clustal Omega, Substitution Matrix = "BLOSUM80) and the alignment visualized in a phylogenetic tree. The phylogenetic tree was obtained with the UPGMA algorithm. Distance measures are given in Jukes-Cantor units. Bootstrap values are shown (1000 Replicates). Colours indicate clusters of LysM motifs that correlate to the position of a LysM motif within the respective proteins and to the type of ortholog (shades of yellow = first position, shades of blue = second position, shades of green = third position, shades of red = fourth position, purple = fifth position). Clusters are further labelled manually (e.g. LytE 2 – cluster with the second LysM motifs of the LytE orthologs). Interestingly the LysM motifs of the protein M4KSM2 (indicated) do not cluster together with the other LysM motifs of the CwlS orthologs.

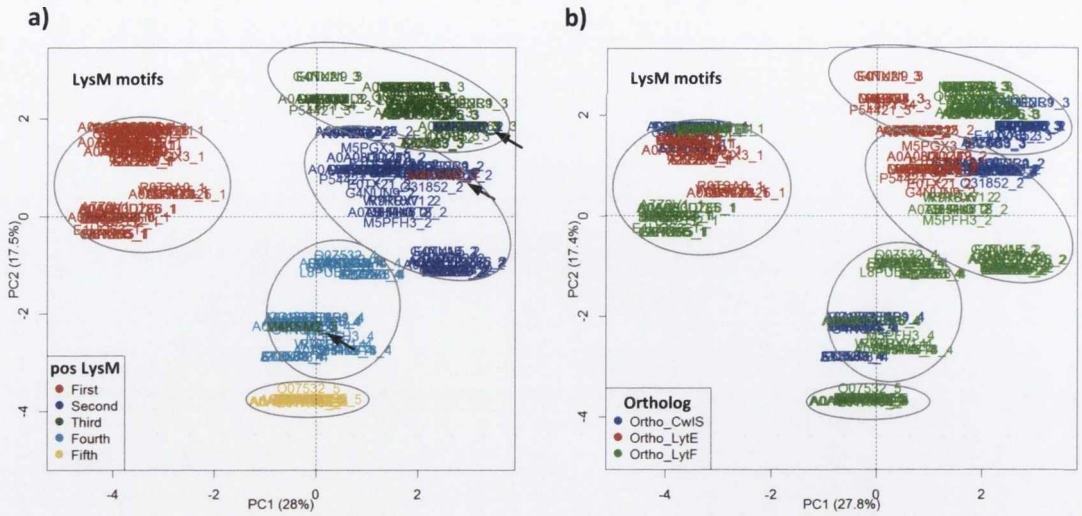



**Figure 6.12: Metric multidimensional scaling of the LysM motifs of LytE, CwlS and LytF orthologs.** The amino acid sequence of the LysM motifs of all orthologs were aligned (Algorithm = Clustal Omega, Substitution Matrix = "BLOSUM80) and the alignment visualized by metric multidimensional scaling. The LysM motifs are mapped in a plane formed by the first (PC1) and second (PC2) component of the sequence space by MDS analysis with the distances based on the difference score. There is no unit for the principal component axis, but the percentage given between brackets indicates the percent of variance explained by this component and thus the relative weight of this component. Both diagrams (**a** and **b**) present the same data, however the colouring highlights different characteristics of the LysM motifs. LysM motifs in diagram **a)** are grouped (i.e. coloured) by the position of the LysM motif within the respective protein (Red = First, Blue = Second, Green = Third, Turquoise = Forth, Yellow = Fifth). Numbering starts with the $NH_2$-terminal LysM motifs within a protein (first position). Five clusters are indicated that correlate with the positon of the LysM motif. Interestingly the three LysM motifs of the protein M4KSM2 (indicated by arrows) are the only LysM motifs that do not cluster according to their positions. e.g. the N-terminal LysM motif of M4KSM2 (M4KSM2_1) is localized in the cluster of LysM motifs on the second position. LysM motifs in diagram **b)** are grouped (i.e. coloured) by the type of ortholog to which the protein that possess the LysM motif belongs (Red = LytE ortholog, Green = LytF ortholog, Blue = CwlS ortholog).

However, the sub-division into groups of orthologs is less distinct than the clustering according to position. This is especially obvious in the metric multi-dimensional scaling approach (Figure 6.12). Due to the fact that in this approach all sequences are compared to each other at once, clusters are less obvious and more difficult to identify. However, just as in the phylogenetic tree, LysM motifs at position one are remarkably distinct from all LysM motifs at a different position. However, clusters for LysM motifs at a different position are slightly more difficult to identify. Though in general just as in the phylogenetic tree, LysM motifs at positions 2 & 3 and 4 & 5 were close to each other.

Remarkably there is one protein, whose LysM motifs clustered in an uncharacteristic way. This protein is M4KSM2 of strain *Bacillus subtilis XF-1*, which had stood out before by being the only CwlS ortholog with three instead of four LysM motifs. As is obvious from both types of analysis (phylogenetic tree and metric multidimensional scaling) the N-terminal LysM motifs of M4KSM2 clusters with the second LysM motifs of the CwlS orthologs, its second LysM motif clusters with the third LysM motifs of the CwlS orthologs and its third LysM motif clusters with the fourth LysM motifs of the CwlS orthologs. Thus, each LysM motif seems more similar to a LysM motif at one higher position.

### 6.4.4 Discussion

This study has identified orthologs to the three D,L-endopeptidases (CwlS, LytF and LytE) of *B. subtilis 168* in 19 closely related species. Interestingly although all strains possessed one ortholog to LytE and LytF, only fourteen out of twenty strains held a CwlS ortholog. Surprisingly these orthologs also varied in the number of their LysM motifs. In *B. subtilis 168*, LytE has three LysM motifs, CwlS has four LysM motifs and LytF has five LysM motifs. However, a fraction of LytE orthologs possessed two instead of three LysM motifs and the majority of LytF orthologs had only four LysM motifs. Furthermore, although most CwlS orthologs had four LysM motifs, one possessed only three LysM motifs. Depending on the D,L endopeptidase composition and LysM motif distribution of each strain four groups could be defined (A, B, C, D).

Cluster analysis of the LysM motifs of the orthologs resulted in the observation that all but three LysM motifs clustered according to their position within the respective protein and that the LysM motifs at the first position were most different of all LysM motifs at a subsequent position. A possible explanation for this phenomenon is the close association of the N-terminal LysM motifs with the signal peptide, which might exhibit an evolutionary pressure on the first LysM motif to preserve its amino acid sequence. On the other hand, given that previous work has shown that one LysM motif is

sufficient to allow the LytE enzyme to fulfil the D,L-endopeptidase requirement in *B. subtilis 168*, the subsequent LysM motifs may be of secondary importance allowing the alteration of their nucleotide sequence by mutation.

The three LysM motifs that did not cluster according to their position belonged to the protein M4KSM2 of strain *Bacillus subtilis XF-1*, which had stood out before by being the only CwlS ortholog with three instead of four LysM motifs. Both types of analysis (phylogenetic tree and metric multidimensional scaling) indicate that M4KSM2 originally possessed four LysM motifs (as do the other CwlS-type enzymes), but experienced the loss of the first LysM motif, which resulted in each consecutive LysM motifs to take the position of the previous LysM motif in the cluster analyis.

The high sequence similarity of the catalytic domains of the orthologs to CwlS, LytE and LytF as well as the cluster analysis of their LysM motifs suggest an evolutionary model in which a common ancestor underwent gene and LysM motif duplication events, which led to the development of the three orthologs and their different LysM motifs. In accordance with the result of the cluster analysis two evolutionary models were developed. The first model (Figure 6.13) attempts to identify how the three paralogs evolved and how orthologs in different species acquired different number of LysM motifs. The second model (Figure 6.15 a, page 220) builds on the first and is a suggestion as to how the members of the four groups (A, B, C, D) in this study are related to each other.

The first model (Figure 6.13) was developed in several logical steps, as outlined below, and strives to deliver a plausible and parsimonious explanation as to how a combination of gene and motif duplication events led to the development of the three different paralogous enzymes with different number of LysM motifs. Gene duplication events are well established (Elliott *et al.*, 2013; Reams and Roth, 2015) and protein motif duplication are thought to arise either through tandem duplication within a gene (Andrade *et al.*, 2001) or through DNA slippage (Djian, 1998).

1. On the basis of the high sequence similarity of the catalytic domains of CwlS, LytE and LytF (see Figure 1.5, page 26), it is assumed that all three paralogs have a common ancestor.
2. Following the observation that LysM motif duplication can only occur within a gene, not by gene conversion (Table 8.2, appendix, page 234), it can further be deduced that the common ancestor possessed one or more LysM motifs.

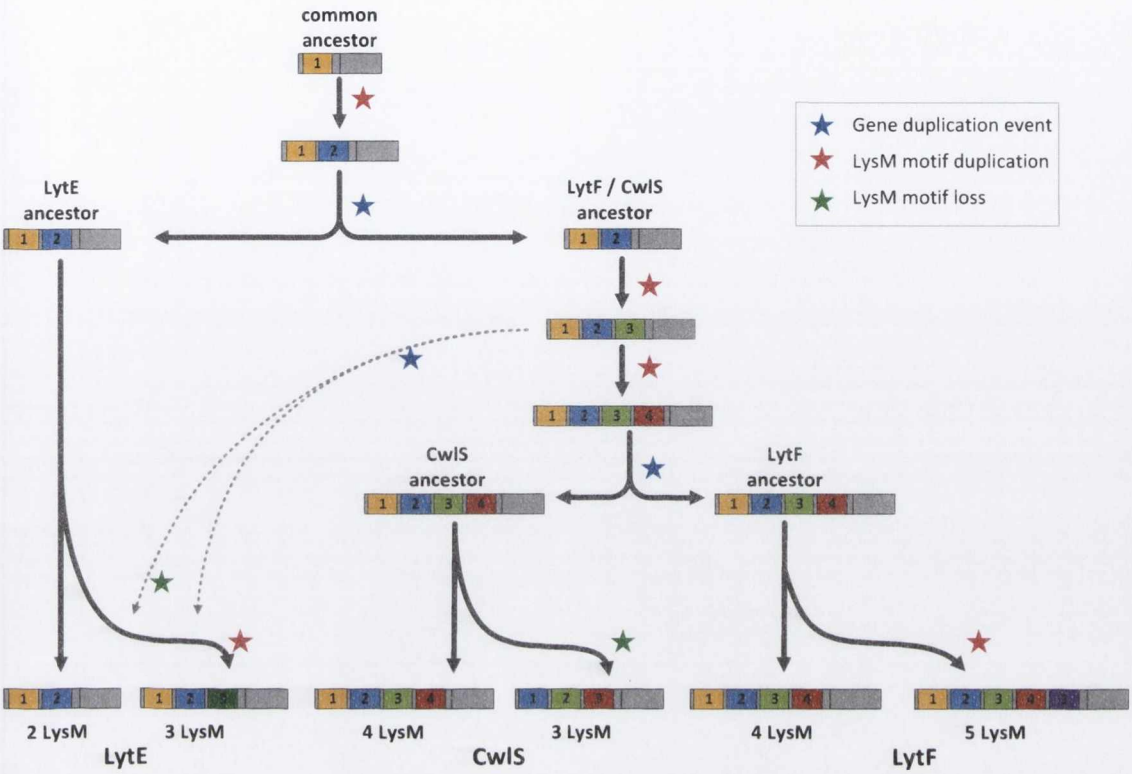



**Figure 6.13: A putative model of the evolutionary development of the CwlS, LytE and LytF D,L-endopeptidases and their LysM motifs.** The proposed model assumes a stable inheritance of genes. The common ancestor is a gene that encodes for a D,L-endopeptidase type protein with one LysM motif. Through a combination of LysM motif (red star) and gene duplication (blue star) events, the model arrives at a set of genes with LytE orthologs (having either two or three LysM motifs), CwlS orthologs (four or three LysM motifs) and LytF orthologs (with either four or five LysM motifs). While the main focus is on the gain of motifs and complete genes by duplication events, the dashed arrow explores an alternative scenario that involves the loss of a LysM motif (green star).

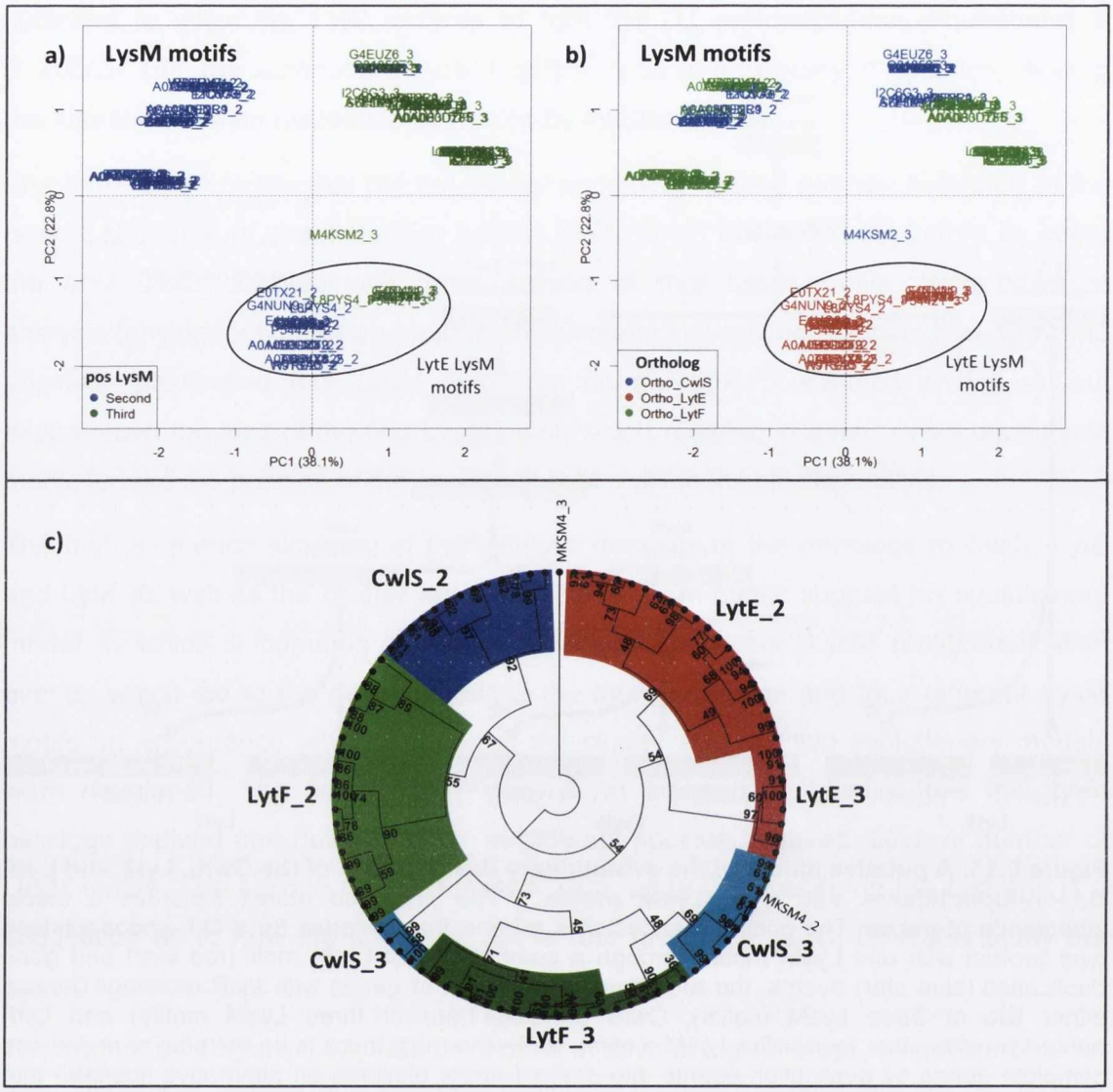



**Figure 6.14: Cluster analysis of the second and third LysM motifs of CwlS, LytE and LytF orthologs.** The amino acid sequence of the second and third LysM motifs of CwlS, LytE and LytF orthologs were aligned (Algorithm = Clustal Omega, Substitution Matrix = "BLOSUM80) and the alignment visualized by metric multidimensional scaling. The LysM motifs are mapped in **(a)** and **(b)** in a plane formed by the first (PC1) and second (PC2) component of the sequence space by MDS analysis with the distances based on the difference score. There is no unit for the principal component axis, but the percentage given between brackets indicates the percent of variance explained by this component and thus the relative weight of this component. Both diagrams (a) and (b) present the same data, however the colouring highlights different characteristics of the LysM motifs. LysM motifs in diagram **(a)** are grouped (i.e. coloured) by the position of the LysM motif within the respective protein (Blue = Second, Green = Third). Numbering starts with the $NH_2$-terminal LysM motifs within a protein (first position). One cluster is indicated that contains the LytE LysM motifs. LysM motifs in diagram **(b)** are grouped (i.e. coloured) by the type of ortholog to which the protein that possess the LysM motif belongs (Red = LytE ortholog, Green = LytF ortholog, Blue = CwlS ortholog). **(c)** The alignment is further visualized in a phylogenetic tree. The phylogenetic tree was obtained with the UPGMA algorithm. Distance measures are given in Jukes-Cantor units. Bootstrap values are shown (1000 Replicates). Colours indicate clusters of LysM motifs that correlate to the position of a LysM motif within the respective proteins and to the type of ortholog (shades of red = LytE, shades of green = LytF, shades of blue = CwlS). Clusters are further labelled manually (e.g. LytE 2 – cluster with the second LysM motifs of the LytE orthologs). Interestingly the LysM motifs of the protein M4KSM2 (indicated) do not cluster together with the other LysM motifs of the CwlS orthologs.

3. It was shown by cluster analysis that the catalytic domain of CwlS and LytF are more similar to each other than to the catalytic domain of LytE (see Figure 6.9. page 208). Therefore, it is assumed that the common ancestor underwent a gene duplication event and that one copy became the LytE ancestor gene and the other the ancestor of both CwlS and LytF. A gene duplication event is further supported by the fact that the *lytE* and *lytF* open reading frames are localized comparably close to each other on the *B. subtilis 168* genome (BSU09420 and BSU09370). However, CwlS is localized further away (BSU19410). This leads to the question as to when the ancestor gene underwent the gene duplication event and how many LysM motifs were present in the enzyme at this time.
4. Since most LytE orthologs have two LysM motifs and none of the CwlS or LytF orthologs have less than three LysM motifs, it is assumed that the ancestor also possessed two LysM motifs. However, given the evidence for the loss of a LysM motif in protein M4KSM2 another possibility is that the ancestor underwent gene duplication, when it possessed three LysM motifs and that one LysM motif was lost in the LytE orthologs of members of group A and B (that have a LytE ortholog with two LysM motifs) (see dashed arrow in Figure 6.13, page 217). In that case, one would expect to find that the third LysM motifs of LytE orthologs are more similar to the third LysM motifs of CwlS and LytF orthologs, than to the second LysM motifs of the LytE orthologs. Yet when comparing the second and third LysM motifs of all orthologs (see Figure 6.14) it becomes apparent that the third LysM motifs of LytE orthologs are indeed more similar to the second LysM motif of LytE orthologs, than to the third LysM motifs of CwlS and LytF orthologs. Hence it is assumed that the ancestral gene possessed two LysM motifs when it underwent gene duplication and that the *lytE* gene acquired another LysM motif independently through motif duplication in LytE orthologs of group C and D.
5. Similarly the fact that all CwlS orthologs possess four LysM motifs (with the exception of M4KSM2) and the greater similarity between the LysM motifs of CwlS and LytF suggests that the CwlS / LytF ancestor gene acquired four LysM motifs before it underwent a further gene duplication event. This is supported by the fact that members of group A do not possess a CwlS orthologs, but that their LytF ortholog has four LysM motifs. Members of group B and C are therefore strains in which the CwlS / LytF ancestral gene underwent gene duplication and in which one copy became CwlS and one copy evolved to become LytF (after another LysM motif duplication event).

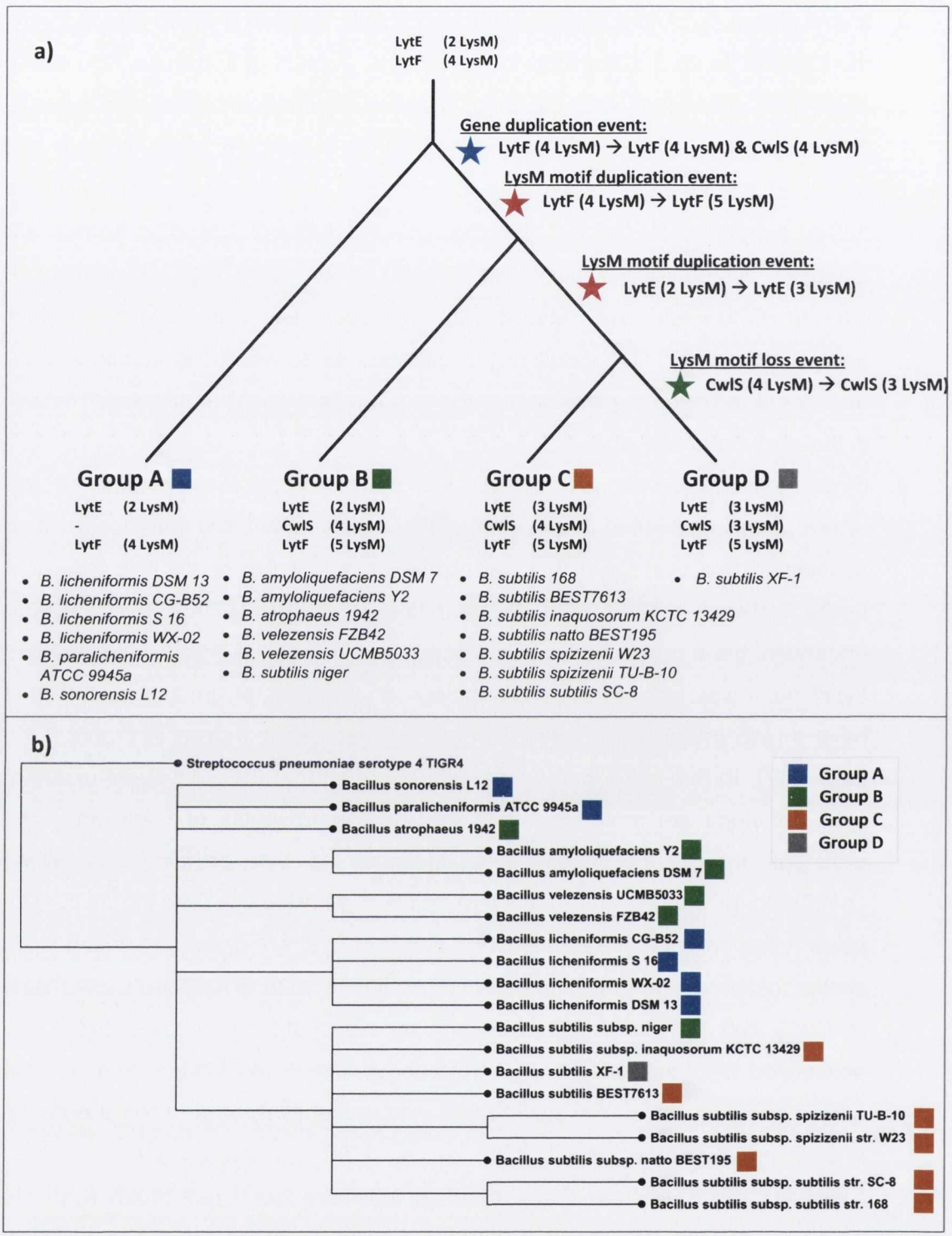



**Figure 6.15: Proposed model for the evolutionary development of groups with different sets of paralogs and basic phylogenetic tree of all twenty strains of the *B. subtilis* strain group.** **a)** The top panel describes the composition of paralogs that each group possesses (identified by their names and in brackets the number of LysM motifs the evolutionary relationship between the four groups, and the strains that make up each group. The model assumes a stable inheritance of genes and describes how through a combination of LysM motif (red star) and gene (blue star) duplication events as well as loss of LysM motifs (green star) each group gains a different set of paralogs. **b)** The phylogenetic tree is based on the classification of the NCBI taxonomy database (Sayers *et al.*, 2009), which is derived from a diverse array of phylogenetic resources. *Streptococcus pneumoniae serotype 4 TIGR4* is included as outgroup. Colours indicate the paralogs groups.

This model was further applied to unravel the evolutionary relationship between members of the four different groups. A respective 4-step proposal is visualized in Figure 6.15 (a) and can be described as following.

1. It is suggested that all groups originated from group A. Members of group A possess one LytE ortholog (2 LysM), no CwlS ortholog and one LytF ortholog (4 LysM).
2. A gene duplication event of the LytF ortholog, followed by a LysM motif duplication event, led to the creation of group B. Members of group B, similarly to group A, have a LytE ortholog with 2 LysM motifs, but a CwlS ortholog with 4 LysM motifs and a LytF ortholog with 5 LysM motifs.
3. Thereafter another LysM motif duplication event created group C, which is essentially the same as group B, but whose members have a LytE ortholog with 3 not 2 LysM motifs.
4. Group D is a derivative of group C, in which one LysM motif of the CwlS ortholog is lost, leading to the following setup: one LytE ortholog (3 LysM), one CwlS ortholog (3 LysM) and one LytF ortholog (5 LysM).

A key feature of this model is the expansion of LysM motif repeats. Although LysM motif loss is a real possibility (as seen in the case of the CwlS-type enzyme M4KSM2) several arguments support a model with expanding LysM motifs.

First of all, this study has shown that a majority (85 %) of LysM motif proteins possesses one or two LysM motifs and that a minority of proteins has (< 2 %) five or more LysM motifs (see Figure 6.2, page 192). Therefore, it seems more likely to assume that the common ancestor also possessed one or two LysM motifs. Secondly, other studies have shown that protein evolution tends to generate longer and more complex proteins and that expansion of proteins occurs more frequently than their truncation (Kummerfeld and Teichmann, 2005; Snel *et al.*, 2000). And thirdly the fact that the second and third LysM motifs of LytE orthologs are more similar to each other than to the third LysM motifs of CwlS and LytF orthologs seems to indicate that LytE acquired a third LysM motif independently of CwlS and LytF (Figure 6.14, page 218). Finally the fact that, as observed earlier in this study, the sub-division of the strains into the here described four group correlates with a basic phylogenetic tree pictured in Figure 6.15 (b) gives further support to this model.

Therefore, it can be concluded that LysM motif expansion (or reduction) can occur independently in paralogs and that the number of LysM motif within a protein is a poor criterion for identifying an ortholog. This observation is supported by Looman *et al.*

(2002), who demonstrated that the number of zinc finger motifs expanded independently in KRAB zinc finger protein orthologs.

# 7 Summary

The cell wall of *B. subtilis* is a multi-layered structure located outside the cell composed of peptidoglycan and an anionic glycopolymer, poly(glycerol phosphate), in approximately equal amounts. It is a rigid structure that determines cell shape, resists turgor pressure, provides a microclimate between the cell and the external *milieu* and acts as a platform for exposure of cellular proteins. However, it is also a dynamic structure that undergoes continual synthesis and turnover to accommodate the remodelling required for cell growth, cell division and sporulation.

This remodelling is carried out in *B. subtilis* by a battery of 35 autolytic enzymes that collectively have the capability to cleave all of the chemical bonds in peptidoglycan. Interestingly there is very significant redundancy in the enzymatic capability and specificity of these 35 autolysins in *B. subtilis*. For example, there are seven D,L-endopeptidases (CwlO, CwlS, LytE, LytF, PgdS, YkfC, YqgT) each with the potential to cleave the chemical bond between the D-glutamate and *meso*-diaminopimelate located in the stem peptides of peptidoglycan. It is clear that the expression and activity of these enzymes must be tightly regulated, as an unregulated expression of each has the capability to cause cell lysis. Thus, regulation of autolysin expression is complex and with varied and diverse mechanisms being identified.

Although none of the D,L-endopeptidases in *B. subtilis* is essential for viability, the *lytE* and *cwlO* genes are synthetically lethal, indicating that they perform an overlapping essential function. An interesting feature of CwlO and LytE is that these proteins bind to different components of the cell envelope and by completely different protein domains. CwlO interacts with the membrane embedded FtsEX complex through its N-terminal coiled-coil domain while LytE binds to peptidoglycan through three N-terminal LysM motifs. The synthetic lethality of CwlO and LytE presents the opportunity to investigate experimentally the functional features of a D,L-endopeptidase necessary to fulfil the requirement for viability. To achieve this, an experimental system was established in which genes encoding different D,L-endopeptidases and variants of LytE were introduced at the *cwlO* locus (replacing the *cwlO* gene) and expressed from the *cwlO* promoter. The capability of these D,L-endopeptidases and LytE variants encoded by the genes at the *cwlO* locus to fulfil the D,L-endopeptidase requirement for viability was then tested by attempting to delete the endogenous *lytE* gene. Successful deletion of the endogenous *lytE* gene signifies that the D,L-endopeptidase / LytE variant expressed from the *cwlO* promoter can fulfil the requirement for viability while an

inability to delete *lytE* signifies that they cannot. A particular focus of this study was to examine the three peptidoglycan-binding LysM motifs of LytE to establish how many are needed for LytE function and the extent to which they are functionally equivalent.

The following was established:

**(i) LytE expression is upregulated in the absence of CwlO.**

It was demonstrated that the levels of *lytE* transcript and LytE protein increase significantly when the *cwlO* gene is deleted. The mechanism by which this is achieved was reported by Salzberg *et al.* (2013). Expression of *cwlO* is directed by the WalRK TCS which is responsive to cell growth, while expression of *lytE* is under the control of the SigIRsgI regulatory system, which is proposed to be responsive to cell wall stress. Thus, the absence of the CwlO D,L-endopeptidase activity leads to cell wall stress that is sensed presumably by the membrane embedded RsgI antisigma factor. This results in SigI activation and increased expression of LytE. Thus, the WalRK and SigIRsgI regulators form a regulatory network that functions to ensure that there is an adequate level of D,L-endopeptidase enzyme present under normal growth (CwlO) and cell wall stress (LytE) conditions.

**(ii) LytE can fulfil the D,L-endopeptidase requirement for viability when expressed from the *cwlO* promoter.**

It was of interest to establish if LytE can fulfil the D,L-endopeptidase requirement for viability when expressed from the *cwlO* promoter in the absence of the endogenous *lytE* gene. It was found that the strain LU21 ($P_{cwlOleader}$*lytE* Δ*cwlO* Δ*lytE*) was viable, signifying that changing the regulation of LytE expression from its natural promoter to the *cwlO* promoter is compatible with viability and indeed normal growth. LytE expression was higher when driven by the *cwlO* promoter than by the endogenous *lytE* promoter (un-induced or induced). LytE expression (stabilized transcript and increased protein level) was further increased by deletion of the *cwlO* leader sequence. Although neither growth nor cell morphology were affected when LytE was expressed from the *cwlO* promoter when the leader sequence was included, removal of the *cwlO* leader sequence (resulting in very high expression levels of LytE) led to minor changes in cell morphology and growth inhibition, indicating that such an elevated level of autolytic activity is detrimental to the cell.

**(iii) Expression of LytF can fulfil the D,L-endopeptidase requirement for viability.**

The experimental system which was generated allowed us to establish if other D,L-endopeptidases encoded in the *B. subtilis* genome could support viability when expressed from the *cwlO* promoter. We chose to examine two D,L-endopeptidases that

differed from LytE only in the number of LysM motifs located in the N-terminal region of the protein i.e. LytF (5 LysM motifs) and CwlS (4 LysM motifs) and one enzyme with an uncharacterised N-terminal domain (YqgT). Our results showed that a viable strain LU116 ($P_{cwlOleader}$*lytF* Δ*lytF* Δ*cwlO* Δ*lytE*) could be constructed in which LytF was expressed from the *cwlO* promoter and the *cwlO* and *lytE* genes were deleted. The fact that LytF is unable to fulfil the D,L-endopeptidase requirement for viability when expressed from the *lytF* promoter, but can do so when expressed from the *cwlO* promoter, indicates that complementation requires either greater than normal LytF protein levels and / or LytF expression at a different time point throughout the growth cycle. However, although a viable strain could be constructed, severe growth and morphology defects were observed, which were further amplified in the strain without *cwlO* leader sequence. Since LytF expression is increased by deletion of the leader sequence, exacerbation of the aberrant phenotype suggests that the level of LytF in this strain may be too high for normal growth. The observation that LytF was mainly localized to cell septa and poles in strain LU151 ($P_{cwlOleader}$*lytF-HA* Δ*lytF* Δ*cwlO*) also indicates that localization of LytF may be suboptimal for normal growth. Thus, these results indicate that the level of D,L-endopeptidase expression and the subcellular localization are critical for normal growth. It is important to remember in this context that the CwlO enzyme is active only when associated with the FtsEX complex embedded in the membrane. In contrast, the activity of LytE or LytF does not have such a mechanism to control the activity of the enzyme produced. Thus, activity levels of LytF will be critical to the viability of the cell.

**(iv) CwlS and YqgT cannot fulfil the D,L-endopeptidase requirement for viability.**

Expression of neither CwlS nor YqgT was able to fulfil the D,L-endopeptidase requirement for cell viability. While this result may seem somewhat surprising in that the CwlS enzyme had four LysM motifs at the N-terminus of the protein, the enzyme appeared to be located almost exclusively to cell septa and poles. This indicates that enzyme correct localization of autolysins with LysM motifs may require association with other proteins for normal cellular function. Of course it is also possible that the specific activity of the CwlS enzyme may not be sufficient to support cell viability. It was not possible to determine YqgT localization definitively and it is unclear if / how it binds to peptidoglycan given that it does not possess LysM motifs. However, it is clear that it does remain cell-associated at high levels by a mechanism that remains to be established. Thus, it is possible that YqgT fails to complement for CwlO, because it does not bind to the sites within the cell cylinder at which D,L-endopeptidase enzymatic activity is required.

**(v) A LytE variant with any single LytE LysM motif can fulfil the D,L-endopeptidase requirement for viability.**

The experimental system devised allowed us to establish the requirement for LytE binding to peptidoglycan by generating LytE variants with (i) different numbers of LysM motifs and (ii) the LysM motifs permuted. These variant LytE proteins could then be tested for their ability to fulfil the D,L-endopeptidase requirement for viability when expressed from the *cwlO* promoter. It was somewhat surprising that all variant LytE proteins with any single LysM motif (of LytE) were able to fulfil the D,L-endopeptidase requirement for viability. However, viability was not supported by a LytE variant without LysM motif. Thus, although LytE has three LysM motifs, any of these domains present in single copy in the protein was sufficient to support viability. Furthermore, there was impairment in growth or morphology when strains were grown in rich or minimal media.

However, strains in which the D,L-endopeptidase requirement was fulfilled by a variant LytE with a single LysM motif exhibited a higher sensitivity to salt stress. This defect was more pronounced in LytE variants with a single copy of the second (LysM2) or third (LysM3) LysM motifs. This increased effectiveness of the first LysM motif (LysM1) may be due to the fact that it has evolved to be more effective in the position adjacent to the signal sequence. The effectiveness of the LytE with single LysM motifs was also diminished when they were expressed from the weaker LytE promoter. This indicates a deficiency in cell association of LytE variants with reduced numbers of LysM motifs. This was verified by western blot analysis which showed that less LytE protein associated with the cell, when the variant LytE proteins possessed only a single LysM motif. Therefore, increased salt sensitivity in these strains might be caused by a lack of D,L-endopeptidase enzymatic activity as a result of a diminished binding capacity of the variant LytE protein.

Cluster analysis of the LysM motifs of the LytE, CwlS and LytF proteins revealed that (i) the motifs at similar positions were more closely related to each other than to LysM motifs at different positions and (ii) that LysM motifs at the first position were the most distinct. It is therefore possible that the first LysM motif of LytE has acquired (or retained) some characteristics that make it more suitable for interaction with the signal peptide or give it a greater binding capacity resulting in the variant LytE protein LytE[L1cat] being superior to LytE[L2cat] or LytE[L3cat].

**(vi) *In vitro* analysis of variant LytE proteins.**

Selected LytE variants were chosen for analysis of their properties *in vitro*. It was found that variant proteins with a single LysM motif bind to peptidoglycan with less affinity that those with multiple LysM motifs. It was further found that LytE proteins bind less to

peptidoglycan as one increases the concentration of NaCl. This supports the conclusion that the reduced viability of cells expressing LytE variants with fewer LysM motifs is caused by a decrease in LytE binding to the cell wall. Our analysis also showed that the enzymatic activity of LytE variants did not vary significantly with variance in the number of LysM motifs. This is in contrast to Wong *et al.* (2014), who showed a dependence of enzymatic activity on LysM motif composition, but in accordance with Steen *et al.* (2003). Since all variant LytE proteins showed similar enzymatic activity *in vitro,* it was concluded that the observed stress sensitivity in strains in which the D,L-endopeptidase requirement is fulfilled by a variant LytE protein with a single LysM motifs is caused by the reduced capability of the protein to bind to the cell wall, which in turn leads to a lack of D,L-endopeptidase enzymatic activity at the required sites within the cell envelope.

**(vii) A heterologous LytE protein containing the single LysM motif of the XlyB amidase from the defective prophage PBSX can fulfil the D,L-endopeptidase requirement for viability.**

The ability of heterologous D,L-endopeptidase enzymes composed of the LytE catalytic domain and LysM1 (i.e. adjacent to the signal sequence) domains from CwlS (D,L-endopeptidase from *B. subtilis*), LytF (D,L-endopeptidase from *B. subtilis*), XlyB (N-acetylmuramoyl-L-alanine amidase from the PBSX defective prophage of *B. subtilis*) and Aaa (adhesin / autolysin from *S. aureus*) to fulfil the D,L-endopeptidase requirement for viability was tested. Our results show that only the single LysM motif of XlyB can fulfil this requirement, when the heterologous variant LytE was expressed from the *cwlO* promoter. It is somewhat surprising that the single LysM motif of XlyB was able to complement for the LysM motifs of LytE, while the first LysM motif of neither CwlS nor LytF were able to do so, given that they share close sequence similarity to the first LysM motif of LytE. However, it was shown *in vitro* that purified LytE[$L1^{XlyB}cat^{LytE}$] heterologous protein bound to bacterial cell walls to a great extent than did the LytE[$L1^{LytF}cat^{LytE}$] protein. This indicates that the ability of variant LytE proteins to fulfil the D,L-endopeptidase requirement is dependent on their binding affinity for binding to peptidoglycan. Moreover it suggests that here is a threshold binding affinity below which complementation is not achieved.

**(viii) Bioinformatic analysis of bacterial LysM motifs.**

A bioinformatic analysis was performed on the collection of LysM motifs in bacterial proteins in the UniProt database. There were approximately 38,000 bacterial LysM motifs in the database. This analysis showed that greater than 90 % of proteins contained on average just 1 or 2 LysM motifs. Moreover >85 % of bacterial strains

have between 1 and 6 LysM motif containing proteins with a total complement of 1 to 9 LysM motifs. *B. subtilis* subspecies *subtilis* (WT168) is quite atypical in that its genome encodes 18 proteins with LysM motifs and has a total complement of 30 LysM motifs encoded in the genome. In fact, the bacterial family *Bacillaceae* (to which *B. subtilis* belongs) have the most LysM motifs with a median of 15.5 LysM motifs per strain. There is evidence that *Bacillus* species are common in the rhizosphere, associated with plant roots (Choudhary and Johri, 2009). The higher than usual incidence of LysM containing proteins may be an adaptation for the interaction of *Bacilli* with plant root components or for interaction with other bacterial inhabitants of this ecological niche.

**(ix) Duplication of LysM motifs within proteins.**

Forty-five proteins were identified that have two or more LysM motifs with identical amino acid sequences. However, identical LysM motifs (at the amino acids level) were not identified between any two proteins. This indicates that duplication of LysM motifs occurs within proteins. This finding indicates that the evolutionary relationship between orthologous proteins possessing LysM motifs can be investigated by comparing LysM motif sequences.

Thus, the evolutionary development of LytE, CwlS and LytF D,L-endopeptidases was investigated through an analysis of the number and sequence of their LysM motifs. Orthologous LytE, CwlS and LytF D,L-endopeptidases were identified in 20 closely related species of *B. subtilis* by amino acid sequence similarity of their catalytic domains. Interestingly the number of LysM motifs orthologs varied in their number of LysM motifs demonstrating that the total number of LysM motif within an enzyme is a poor indicator for the identification of orthologs of these D,L-endopeptidases. A cluster analysis of the LysM motifs within these orthologs confirmed that the LysM motif adjacent to the signal sequence differed significantly from LysM2 and LysM3 – the latter two being similar to each other. This analysis confirms our experimental findings, showing that LysM1 is distinct from the subsequent LysM motifs.

The proposed model describes how orthologs developed through a combination of gene duplication and LysM motif duplication events. Interestingly although the general trend is the development of larger more complex proteins with more LysM motifs evidence for the loss of a LysM motif was also found.

In summary, it is clear that expression of D,L-endopeptidase activity is required for normal growth and viability of *B. subtilis* as reported by Bisicchia *et al.* (2007). Under normal growth conditions, this activity is provided by CwlO and by LytE under stress conditions. Our results show that of three tested other D,L-endopeptidases encoded in the *B. subtilis* genome – LytF, CwlS and YqgT – the latter two cannot complement

CwlO / LytE activity, even when expressed from the *cwlO* promoter. LytF gives partial complementation, but the cells are not robust. It is clear that D,L-endopeptidases must have many other features to be able to fulfil this requirement for viability. Among those requirements are the level of enzyme expressed, the stability of the protein when expressed and the localization of the protein within the cell envelope. It is somewhat surprising that LytE variants with a single LysM motif are viable, suggesting that the peptidoglycan-binding affinity of normal LytE protein is excessive. However, our experiments were performed under laboratory conditions. It is also clear that not all LysM motifs within a protein are equal and the experimental and bioinformatics analysis of the three LysM motifs of LytE indicate that within this enzyme the first LysM motif is distinct. This may be due to some interaction or coevolution with the adjacent signal peptide, which resulted in the N-terminal LysM motif preserving or developing certain attributes that the subsequent LysM motifs do not possess. An evolutionary development study showed that LytE, CwlS and LytF share a common ancestor and that LysM motif repeats expanded independently in the different orthologs.

Overall this study offers further insights in the D,L-endopeptidase requirement of *B. subtilis*. To our knowledge this is the first work that has demonstrated that the D,L-endopeptidase requirement of *B. subtilis* can be fulfilled by a D,L-endopeptidase other than LytE or CwlO and the first study that has investigated the importance and impact of modifying LysM motifs on the functionality of an autolytic enzyme *in vivo*. Furthermore, few studies so far have investigated bacterial LysM motifs on a large scale and the bioinformatics analysis of around 38,000 bacterial LysM motifs conducted in this study led to some interesting observations, such as that although bacterial proteins can contain up to 12 LysM motifs the great majority does not possess more than two LysM motifs. In addition evidence was collected which supports that, although the loss of a LysM motif can occur, LysM motif duplication events (within proteins) occur more frequently. Also a first model was developed describing the relationship between LytE, CwlS and LytF.

However, future work could collect further experimental validation in favour of these results. For example it remains unknown why LytF, but not CwlS was able to complement for CwlO, given the great sequence similarity of their catalytic domains (also localization is a likely suspect). Likewise it is puzzling why the LysM motif of XlyB, but not of LytF or CwlS (again given their great sequence similarity) was able to complement for the LytE LysM motifs. Although it was shown that the individual LysM motifs are not of equal relevance for the functionality of the LytE enzyme it remains unknown what makes the first LysM motifs superior over the subsequent motifs. Future bioinformatics studies could be refined by including whole proteome

information, thereby allowing for the possibility to investigate differences between bacterial species that possess and that do not possess LysM motif proteins. Evolutionary studies could further investigate at which rate LysM motifs expand and why this expansion seems to occur more frequently in species of the *Firmicutes* phylum.

This study also yielded some interesting experimental observations that could be explored in more detail in independent projects. For example, it was found that null mutants in *yqgT* exhibit an affected cell morphology with distinct bulges and thickenings, which is considerable amplified when combined with a disruption of the *lytE* gene. However despite its clear importance for cell wall metabolism *yqgT* and its regulation and localization remain largely unknown.

Another interesting finding of this study was that LytE null mutants exhibit a much reduced autolysis, indicating the great impact of LytE enzymatic activity on cell wall degradation. However purified LytE enzyme was more efficient in degrading purified cell walls than in lysing live cells perhaps due to some unknown interaction of the LytE enzyme with other proteins or cell wall components.

Investigating this and other aspects of autolytic enzymes and cell wall metabolism not only deepens our understanding of these processes, but also has the potential to be utilized for the development of new antibiotics and for the optimization of bacterial strains for enzyme and protein production. Furthermore an in-depth understanding of LysM motifs is likely to be of importance for the development of future binding assays and cell surface displays.

# 8 Appendix

## 8.1 Figures

**Table 8.1: Summary statistic of the frequency distribution of LysM motifs among different bacterial families.**

| Taxonomic family | Total number of | | | | LysM motifs / strain group* | | |
|---|---|---|---|---|---|---|---|
| | strains | strain groups* | LysM motifs | LysM motif proteins | median | mean | standard deviation |
| *Bacillaceae* | 315 | 135 | 4150 | 2756 | **15.5** | **16.43** | 8.31 |
| *Peptococcaceae* | 29 | 26 | 369 | 283 | **12.0** | **12.61** | 4.92 |
| *Enterococcaceae* | 121 | 32 | 1500 | 885 | **11.0** | **10.06** | 3.20 |
| *Moraxellaceae* | 218 | 107 | 2508 | 985 | **11.0** | **11.18** | 2.06 |
| *Thermoanaerobacteraceae* | 22 | 18 | 232 | 123 | **11.0** | **10.22** | 3.67 |
| *Borreliaceae* | 22 | 16 | 188 | 83 | **9.0** | **8.80** | 0.60 |
| *Paenibacillaceae* | 71 | 58 | 704 | 544 | **9.0** | **9.63** | 3.37 |
| *Vibrionaceae* | 105 | 69 | 979 | 572 | **9.0** | **9.19** | 1.96 |
| *Listeriaceae* | 32 | 17 | 267 | 143 | **8.7** | **8.92** | 2.48 |
| *Bacteroidaceae* | 148 | 89 | 1134 | 702 | **8.0** | **7.48** | 1.23 |
| *Flavobacteriaceae* | 111 | 101 | 943 | 420 | **8.0** | **8.66** | 3.87 |
| *Porphyromonadaceae* | 69 | 46 | 503 | 267 | **8.0** | **7.30** | 2.21 |
| *Pseudoalteromonadaceae* | 25 | 22 | 218 | 136 | **8.0** | **8.64** | 1.29 |
| *Spirochaetaceae* | 43 | 26 | 348 | 206 | **8.0** | **7.48** | 2.57 |
| *Burkholderiaceae* | 128 | 79 | 927 | 736 | **7.0** | **6.58** | 3.19 |
| *Shewanellaceae* | 25 | 21 | 196 | 135 | **7.0** | **7.62** | 2.13 |
| *Enterobacteriaceae* | 655 | 176 | 4152 | 3399 | **6.5** | **6.16** | 1.98 |
| *Alteromonadaceae* | 41 | 35 | 254 | 176 | **6.0** | **6.25** | 1.12 |
| *Leptospiraceae* | 82 | 28 | 451 | 354 | **6.0** | **5.95** | 0.93 |
| *Neisseriaceae* | 60 | 26 | 332 | 212 | **6.0** | **5.69** | 0.85 |
| *Streptomycetaceae* | 96 | 93 | 559 | 546 | **6.0** | **5.80** | 1.88 |
| *Micrococcaceae* | 41 | 34 | 272 | 175 | **5.5** | **5.99** | 3.70 |
| *Leuconostocaceae* | 25 | 16 | 147 | 95 | **5.3** | **5.31** | 2.48 |

| Taxonomic family | Total number of | | | | LysM motifs / strain group* | | |
|---|---|---|---|---|---|---|---|
| | strains | strain groups* | LysM motifs | LysM motif proteins | median | mean | standard deviation |
| *Clostridiaceae* | 187 | 148 | 1030 | 699 | **5.0** | **5.27** | 3.72 |
| *Lactobacillaceae* | 170 | 62 | 984 | 734 | **5.0** | **5.99** | 4.36 |
| *Prevotellaceae* | 94 | 76 | 450 | 338 | **5.0** | **4.95** | 1.36 |
| *Pseudomonadaceae* | 196 | 115 | 1070 | 791 | **5.0** | **5.18** | 1.40 |
| *Xanthomonadaceae* | 83 | 38 | 387 | 386 | **5.0** | **4.91** | 1.19 |
| *Alcaligenaceae* | 33 | 22 | 150 | 118 | **4.9** | **4.26** | 1.78 |
| *Chlamydiaceae* | 25 | 13 | 94 | 69 | **4.0** | **3.98** | 1.29 |
| *Microbacteriaceae* | 29 | 25 | 104 | 59 | **4.0** | **3.56** | 1.16 |
| *Rhizobiaceae* | 68 | 39 | 279 | 214 | **4.0** | **3.98** | 0.78 |
| *Lachnospiraceae* | 93 | 77 | 394 | 350 | **3.7** | **4.07** | 2.68 |
| *Phyllobacteriaceae* | 59 | 58 | 206 | 151 | **3.5** | **3.51** | 0.93 |
| *Bradyrhizobiaceae* | 39 | 30 | 113 | 77 | **3.0** | **3.03** | 1.16 |
| *Comamonadaceae* | 47 | 35 | 175 | 170 | **3.0** | **3.22** | 1.53 |
| *Desulfovibrionaceae* | 28 | 22 | 193 | 84 | **3.0** | **7.02** | 6.97 |
| *Eubacteriaceae* | 30 | 30 | 92 | 82 | **3.0** | **3.07** | 1.78 |
| *Pasteurellaceae* | 107 | 34 | 337 | 253 | **3.0** | **3.19** | 0.69 |
| *Pseudonocardiaceae* | 29 | 25 | 96 | 86 | **3.0** | **3.24** | 1.83 |
| *Rhodobacteraceae* | 95 | 89 | 326 | 237 | **3.0** | **3.46** | 0.98 |
| *Rhodospirillaceae* | 23 | 22 | 67 | 48 | **3.0** | **2.89** | 0.84 |
| *Ruminococcaceae* | 37 | 35 | 189 | 138 | **3.0** | **5.06** | 5.21 |
| *Prochlorococcaceae* | 27 | 6 | 81 | 35 | **3.0** | **3.32** | 1.87 |
| *Sphingomonadaceae* | 50 | 46 | 137 | 92 | **2.8** | **2.69** | 1.51 |
| *Brucellaceae* | 31 | 18 | 65 | 65 | **2.0** | **2.13** | 0.50 |
| *Campylobacteraceae* | 63 | 28 | 100 | 67 | **2.0** | **1.70** | 0.72 |
| *Fusobacteriaceae* | 42 | 18 | 98 | 55 | **2.0** | **2.82** | 1.30 |
| *Helicobacteraceae* | 245 | 30 | 268 | 246 | **2.0** | **1.77** | 0.77 |
| *Nocardiaceae* | 25 | 15 | 79 | 61 | **2.0** | **2.75** | 1.47 |
| *Peptostreptococcaceae* | 26 | 9 | 78 | 70 | **2.0** | **2.17** | 1.01 |
| *Propionibacteriaceae* | 32 | 19 | 70 | 38 | **2.0** | **2.06** | 0.62 |
| *Staphylococcaceae* | 51 | 29 | 144 | 100 | **2.0** | **2.74** | 2.23 |
| *Streptococcaceae* | 268 | 86 | 764 | 643 | **2.0** | **2.60** | 1.28 |

| Taxonomic family | Total number of | | | | LysM motifs / strain group* | | |
|---|---|---|---|---|---|---|---|
| | strains | strain groups* | LysM motifs | LysM motif proteins | median | mean | standard deviation |
| *Actinomycetaceae* | 43 | 34 | 66 | 62 | **1.0** | **1.60** | 0.92 |
| *Bartonellaceae* | 31 | 24 | 41 | 41 | **1.0** | **1.28** | 0.72 |
| *Bifidobacteriaceae* | 108 | 36 | 157 | 135 | **1.0** | **1.60** | 0.96 |
| *Caulobacteraceae* | 21 | 18 | 48 | 28 | **1.0** | **1.93** | 1.58 |
| *Gordoniaceae* | 24 | 21 | 35 | 30 | **1.0** | **1.52** | 1.25 |
| *Mycobacteriaceae* | 71 | 38 | 143 | 100 | **1.0** | **1.98** | 1.53 |
| *Veillonellaceae* | 62 | 53 | 91 | 85 | **1.0** | **1.48** | 0.84 |
| ***sum:*** | **5176** | **2793** | **30534** | **20970** | | - | - |

* The calculated mean and median number of LysM motifs per strain was corrected for the cumulative occurrence of closely related strains that were pooled in 'strain groups' (e.g. *B. cereus* AH187 and *B. cereus* AH820 both belong to the strain group *B. cereus*).

**Table 8.2: List of 45 proteins containing duplicated LysM motifs**

| Species | Protein accession number | Protein function | Total number of LysM motifs | Position of duplicated LysM motifs* |
|---|---|---|---|---|
| *Chloracidobacterium thermophilum B* | G2LL77 | Muramidase | 2 | 1/2 |
| *Cellulomonas bogoriensis DSM 16987* | A0A0A0BZ36 | Lytic transglycosylase | 4 | 3/4 |
| *Cellulomonas flavigena DSM 20109* | D5UFG1 | Lytic transglycosylase | 4 | 2/3 |
| *Mycobacterium abscessus subsp. bolletii 50594* | R4UKV4 | Uncharacterised protein | 4 | 2/3 |
| *Bacillus cellulosilyticus ATCC 21833* | E6TQJ0 | Cell wall hydrolase SleB | 4 | 3/4 |
| *Bacillus coagulans 2-6* | F7Z796 | NLP/P60 protein | 2 | 1/2 |
| *Bacillus subtilis subsp. natto BEST195* | D4G4L0 | Uncharacterised protein | 6 | 1/2/3/4/5/6 |
| *Carnobacterium sp. 17-4* | F4BMM5 | Gamma-D-glutamate-meso-diaminopimelate muropeptidase | 12 | 3/4/5/7/8/9/10 |
| *Carnobacterium sp. AT7* | A8UCR7 | Autolysin | 10 | 2/3/4 |
| *Carnobacterium sp. WN1359* | U5S7Z8 | Autolysin | 11 | 2/3/5/8/9 |
| *Granulicatella elegans ATCC 700633* | D0BJS7 | Uncharacterised protein | 6 | 2/3 |
| *Alkaliphilus oremlandii OhILAs* | A8MI50 | Uncharacterised protein | 12 | 5/6/7 |
| *Enterococcus faecalis ATCC 700802* | P37710 | Autolysin | 6 | 1/2 |
| *Enterococcus faecalis EnGen0065* | A0A0M1Z8Y5 | Autolysin | 6 | 1/2 |
| *Enterococcus faecalis EnGen0354* | A0A0M2AV31 | Autolysin | 6 | 1/2 |
| *Enterococcus faecalis EnGen0359* | R3H3W0 | Autolysin | 6 | 1/2 |
| *Enterococcus faecalis EnGen0426* | A0A0M2CKR6 | Autolysin | 6 | 1/2 |
| *Enterococcus faecalis ERV129* | J6C6L6 | Putative gametolysin | 3 | 1/2 |
| *Enterococcus faecalis ERV25* | J6NW46 | Putative gametolysin | 3 | 1/3 |
| *Enterococcus faecalis ERV31* | J6NUC2 | Putative gametolysin | 6 | 1/2 |
| *Enterococcus faecalis ERV37* | J5E9I2 | Putative gametolysin | 6 | 1/2 |
| *Enterococcus faecalis ERV41* | J6DNQ5 | Putative gametolysin | 6 | 1/2 |
| *Enterococcus faecalis ERV63* | J6DQZ4 | Putative gametolysin | 6 | 1/2 |
| *Enterococcus faecalis ERV65* | J6DEL2 | Putative gametolysin | 3 | 1/3 |
| *Enterococcus faecalis ERV72* | J6QF25 | Putative gametolysin | 3 | 1/2 |
| *Enterococcus faecalis ERV73* | J6FV19 | Acetylglucosaminidase | 2 | 1/2 |
| *Enterococcus faecalis Merz96* | C7WW74 | Hydrolase | 6 | 1/2 |
| *Enterococcus faecalis PF3* | V7ZQ24 | Uncharacterised protein | 6 | 1/2 |
| *Enterococcus faecalis TX0309A* | E6GV35 | Uncharacterised protein | 6 | 1/2 |
| *Enterococcus faecalis WKS-26-18-2* | S4F7J2 | Putative gametolysin | 3 | 1/2 |
| *Enterococcus faecium SB2C-2* | S4EPJ0 | Putative gametolysin | 2 | 1/2 |
| *Enterococcus italicus DSM 15952* | E6LIK0 | Uncharacterised protein | 5 | 1/2/3 |
| *Oenococcus oeni ATCC BAA-1163* | A0NHR5 | Uncharacterised protein | 3 | 1/2 |

| Species | Protein accession number | Protein function | Total number of LysM motifs | Position of duplicated LysM motifs* |
|---|---|---|---|---|
| *Oenococcus oeni AWRIB429* | D3LA14 | Uncharacterised protein | 3 | 1/2 |
| *Listeria grayi DSM 20601* | D7UVA2 | Uncharacterised protein | 5 | 3/4 |
| *Listeria innocua ATCC 33091* | H1GFS3 | Gametolysin | 4 | 2/3 |
| *Listeria innocua serovar 6a CLIP 11262* | Q927F1 | Lin2838 protein | 4 | 2/3 |
| *Oscillibacter valericigenes DSM 18026* | G4L0F3 | Uncharacterised protein | 7 | 4/5 |
| *Staphylococcus aureus MRSA252* | Q6GJK9 | N-acetylmuramoyl-L-alanine amidase sle1 | 3 | 2/3 |
| *Acinetobacter baumannii 1295743* | A0A009HNH1 | Spore coat assembly SafA domain protein | 10 | 6/7 |
| *Moraxella catarrhalis CO72* | A0A0E1YD61 | Uncharacterised protein | 6 | 1/3 |
| *Neisseria meningitidis alpha14* | C6S7U8 | Putative membrane peptidase | 2 | 1/2 |
| *Neisseria meningitidis LNP27256* | A0A0C1PX08 | Lipoprotein NlpD | 2 | 1/2 |
| *Haemophilus parainfluenzae T3T1* | E1W6U8 | Uncharacterised protein | 3 | 1/2 |
| *Thermodesulfobacterium geofontis OPF15* | F8C3Z4 | Hydrolase | 3 | 1/3 |

* Positions are numbered beginning with the LysM motif proximal to the N-terminus

**Table 8.3: List of the identified orthologs to LytE, CwlS and LytF**

| | Total number of | | Protein accession number – number of LysM motifs | | | Group* |
|---|---|---|---|---|---|---|
| **Strain** | **LysM motif proteins** | **LysM motifs** | **LytE** | **CwlS** | **LytF** | |
| ***Bacillus subtilis 168*** | 18 | 30 | P54421 – 3 LysM motifs | O31852 – 4 LysM motifs | O07532 – 5 LysM motifs | C |
| ***Bacillus subtilis BEST7613*** | 19 | 32 | L8ABG4 – 3 LysM motifs | L8AM33 – 4 LysM motifs | L8ABG2 – 5 LysM motifs | C |
| ***Bacillus subtilis subsp. inaquosorum KCTC 13429*** | 18 | 30 | L8PYS4 – 3 LysM motifs | L8PTT2 – 4 LysM motifs | L8PUD5 – 5 LysM motifs | C |
| ***Bacillus subtilis subsp. natto BEST195*** | 19 | 36 | D4G6X8 – 3 LysM motifs | D4FXM8 – 4 LysM motifs | D4G6X3 – 5 LysM motifs | C |
| ***Bacillus subtilis subsp. spizizenii W23*** | 15 | 27 | E0TX21 – 3 LysM motifs | E0TZ07 – 4 LysM motifs | E0TX16 – 5 LysM motifs | C |
| ***Bacillus subtilis subsp. spizizenii TU-B-10*** | 16 | 27 | G4NUN9 – 3 LysM motifs | G4NQ36 – 4 LysM motifs | G4NUN5 – 5 LysM motifs | C |
| ***Bacillus subtilis subsp. subtilis SC-8*** | 16 | 28 | G4EYF8 – 3 LysM motifs | G4EUZ6 – 4 LysM motifs | G4EYG3 – 5 LysM motifs | C |
| ***Bacillus subtilis XF-1*** | 16 | 27 | M4KT97 – 3 LysM motifs | M4KSM2 – 3 LysM motifs | M4KPW4 – 5 LysM motifs | D |
| ***Bacillus amyloliquefaciens DSM 7*** | 18 | 29 | E1UQ68 – 2 LysM motifs | E1UVA8 – 4 LysM motifs | E1UQ63 – 5 LysM motifs | B |
| ***Bacillus amyloliquefaciens Y2*** | 14 | 25 | I2C2Z6 – 2 LysM motifs | I2C6G3 – 4 LysM motifs | I2C2Z0 – 5 LysM motifs | B |
| ***Bacillus atrophaeus 1942*** | 18 | 29 | A0A0H3DZF9 – 2 LysM motifs | A0A0H3E2R9 – 4 LysM motifs | A0A0H3DZF6 – 5 LysM motifs | B |
| ***Bacillus velezensis FZB42*** | 14 | 25 | A7Z2V6 – 2 LysM motifs | A7Z5K3 – 4 LysM motifs | A7Z2V1 – 5 LysM motifs | B |
| ***Bacillus velezensis UCMB5033*** | 18 | 29 | S6FIE6 – 2 LysM motifs | S6FL69 – 4 LysM motifs | S6FIE3 – 5 LysM motifs | B |
| ***Bacillus subtilis subsp. niger*** | 18 | 29 | A0A080UQ42 – 2 LysM motifs | A0A080UNC1 – 4 LysM motifs | A0A080UT85 – 5 LysM motifs | B |
| ***Bacillus licheniformis DSM 13*** | 15 | 25 | Q65LZ8 – 2 LysM motifs | none | Q65IH0 – 4 LysM motifs | A |

| Strain | Total number of | | Protein accession number – number of LysM motifs | | | Group* |
|---|---|---|---|---|---|---|
| | LysM motif proteins | LysM motifs | LytE | CwlS | LytF | |
| *Bacillus licheniformis CG-B52* | 13 | 22 | T5HVY1 – 2 LysM motifs | none | T5HPU6 – 4 LysM motifs | A |
| *Bacillus licheniformis S 16* | 13 | 21 | W7RE25 – 2 LysM motifs | none | W7RRW1 – 4 LysM motifs | A |
| *Bacillus licheniformis WX-02* | 15 | 25 | A0A0H4X825 – 2 LysM motifs | none | A0A0H4XYT8 – 4 LysM motifs | A |
| *Bacillus paralicheniformis ATCC 9945a* | 13 | 22 | R9TSA0 – 2 LysM motifs | none | R9TUX7 – 4 LysM motifs | A |
| *Bacillus sonorensis L12* | 15 | 25 | M5PGX3 – 2 LysM motifs | none | M5PFH3 – 4 LysM motifs | A |
| *sum* | 321 | 543 | 20 | 14 | 20 | - |
| *median* | 16.00 | 27.00 | | | | |
| *mean* | 16.05 | 27.15 | | | | |
| *standard deviation* | 1.76 | 2.77 | | | | |

* Strains are assigned to groups (A, B, C, D) that describe the set of CwlS, LytE and LytF D,L-endopeptidases present in the respective organism.

## 8.2 Sequences

### 8.2.1 Nucleotide and amino acid sequence of LytE

```
1    M  K  K  Q  I  I  T  A  T  T  A  V  V  L  G  S  T  L  F  A
1    atgAAAAAGCAAATCATTACAGCTACGACAGCAGTTGTTTTAGGATCGACGTTATTTGCA

21   G  A  A  S  A  Q  S  I  K  V  K  K  G  D  T  L  W  D  L  S
61   GGAGCGGCATCTGCACAAAGCATTAAGGTGAAAAAAGGCGACACGTTATGGGATCTTTCA

41   R  K  Y  D  T  T  I  S  K  I  K  S  E  N  H  L  R  S  D  I
121  AGAAAATACGACACAACGATCAGTAAAATTAAATCAGAGAACCACCTTCGTTCAGACATT

61   I  Y  V  G  Q  T  L  S  I  N  G  K  S  T  S  S  K  S  S  S
181  ATTTATGTGGGACAAACTTTATCGATTAACGGCAAATCTACAAGTTCAAAAAGCAGCAGT

81   S  S  S  S  S  S  T  Y  K  V  K  S  G  D  S  L  W  K  I  S
241  TCTTCTTCCTCTTCTTCTACATACAAAGTAAAGAGCGGGGACAGCCTTTGGAAAATTTCA

101  K  K  Y  G  M  T  I  N  E  L  K  K  L  N  G  L  K  S  D  L
301  AAAAAATACGGCATGACAATCAATGAACTGAAGAAGCTGAATGGCTTAAAATCAGATTTG

121  L  R  V  G  Q  V  L  K  L  K  G  S  T  S  S  S  S  S  S  S
361  CTTCGTGTTGGACAAGTCCTGAAACTGAAAGGTTCAACTAGTTCAAGCAGCTCCAGCTCA

141  S  K  V  S  S  S  S  T  S  T  Y  K  V  K  S  G  D  S  L  S
421  TCAAAAGTGTCATCGTCTTCAACTTCTACTTATAAAGTGAAGAGCGGAGACAGCCTTTCT

161  K  I  A  S  K  Y  G  T  T  V  S  K  L  K  S  L  N  G  L  K
481  AAAATTGCGAGCAAATACGGCACTACGGTTAGCAAATTAAAAAGCTTAAACGGCTTAAAA

181  S  D  V  I  Y  V  N  Q  V  L  K  V  K  G  T  S  T  S  S  S
541  TCAGATGTAATCTATGTCAACCAAGTATTGAAGGTGAAAGGAACAAGCACAAGCAGCTCA

201  K  P  A  S  S  S  S  S  S  S  S  K  T  S  S  T  S  L  N  V
601  AAACCTGCTTCATCTTCATCGTCTTCAAGCAGCAAAACGTCATCTACATCACTTAATGTG

221  S  K  L  V  S  D  A  K  A  L  V  G  T  P  Y  K  W  G  G  T
661  AGCAAGCTGGTTTCTGATGCAAAAGCGTTAGTCGGAACGCCATATAAATGGGGCGGAACG

241  T  T  S  G  F  D  C  S  G  F  I  W  Y  V  L  N  K  Q  T  S
721  ACAACTTCAGGCTTTGACTGCAGCGGATTCATTTGGTACGTACTGAATAAACAAACAAGT

261  V  G  R  T  S  T  A  G  Y  W  S  S  M  K  S  I  A  S  P  S
781  GTGGGCAGAACAAGCACTGCAGGATACTGGAGTTCTATGAAGAGCATTGCCAGCCCGTCT

281  V  G  D  F  V  F  F  T  T  Y  K  S  G  P  S  H  M  G  I  Y
841  GTTGGTGATTTCGTCTTCTTCACAACATATAAATCCGGCCCTTCTCACATGGGGATTTAC

301  I  G  N  N  S  F  I  H  A  G  S  D  G  V  Q  I  S  S  L  N
901  ATTGGAAACAACAGTTTCATTCATGCAGGATCTGACGGCGTACAAATCAGCAGCCTGAAC

321  N  S  Y  W  K  P  R  Y  L  G  A  K  R  F  *
961  AACAGCTACTGGAAGCCTCGTTACCTCGGTGCGAAAAGATTCtaa
```

**Figure 8.1: Nucleotide and amino acid sequences of the wild-type LytE.** Sequences are highlighted to identify the signal sequence (green), LysM motifs (orange), and the catalytic domain (red).

### 8.2.2 Amino acid sequences of overexpressed and purified variant LytE proteins with 6xHis-tag

**>LytE[L123cat]$^{WT168}$**

MSIKVKKGDTLWDLSRKYDTTISKIKSENHLRSDIIYVGQTLSINGKSTSSKSSSSSSSS
STYKVKSGDSLWKISKKYGMTINELKKLNGLKSDLLRVGQVLKLKGSTSSSSSSSSKV
SSSSTSTYKVKSGDSLSKIASKYGTTVSKLKSLNGLKSDVIYVNQVLKVKGTSTSSSKP
ASSSSSSSSKTSSTSLNVSKLVSDAKALVGTPYKWGGTTTSGFDCSGFIWYVLNKQT
SVGRTSTAGYWSSMKSIASPSVGDFVFFTTYKSGPSHMGIYIGNNSFIHAGSDGVQIS
SLNNSYWKPRYLGAKRFLEHHHHHH

**>LytE[L123cat]$^{R}$**

MSIKVKKGDTLWDLSRKYDTTISKIKSENHLRSDIIYVGQTLSINGKSTSSKSSSSSSSS
STYKVKSGDSLWKISKKYGMTINELKKLNGLKSDLLRVGQVLKLKGKSTSSKSSSSSS
SSSTYKVKSGDSLSKIASKYGTTVSKLKSLNGLKSDVIYVNQVLKVKGTSTSSSKPASS
SSSSSSKTSSTSLNVSKLVSDAKALVGTPYKWGGTTTSGFDCSGFIWYVLNKQTSVG
RTSTAGYWSSMKSIASPSVGDFVFFTTYKSGPSHMGIYIGNNSFIHAGSDGVQISSLN
NSYWKPRYLGAKRFLEHHHHHH

**>LytE[L31cat]**

MTYKVKSGDSLSKIASKYGTTVSKLKSLNGLKSDVIYVNQVLKVKGKSTSSKSSSSSS
SSSSIKVKKGDTLWDLSRKYDTTISKIKSENHLRSDIIYVGQTLSINGTSTSSSKPASSS
SSSSSKTSSTSLNVSKLVSDAKALVGTPYKWGGTTTSGFDCSGFIWYVLNKQTSVGR
TSTAGYWSSMKSIASPSVGDFVFFTTYKSGPSHMGIYIGNNSFIHAGSDGVQISSLNN
SYWKPRYLGAKRFLEHHHHHH

**>LytE[L13cat]**

MSIKVKKGDTLWDLSRKYDTTISKIKSENHLRSDIIYVGQTLSINGKSTSSKSSSSSSSS
STYKVKSGDSLSKIASKYGTTVSKLKSLNGLKSDVIYVNQVLKVKGTSTSSSKPASSSS
SSSSKTSSTSLNVSKLVSDAKALVGTPYKWGGTTTSGFDCSGFIWYVLNKQTSVGRT
STAGYWSSMKSIASPSVGDFVFFTTYKSGPSHMGIYIGNNSFIHAGSDGVQISSLNNS
YWKPRYLGAKRFLEHHHHHH

**>LytE[L1cat]**

MSIKVKKGDTLWDLSRKYDTTISKIKSENHLRSDIIYVGQTLSINGTSTSSSKPASSSSS
SSSKTSSTSLNVSKLVSDAKALVGTPYKWGGTTTSGFDCSGFIWYVLNKQTSVGRTS
TAGYWSSMKSIASPSVGDFVFFTTYKSGPSHMGIYIGNNSFIHAGSDGVQISSLNNSY
WKPRYLGAKRFLEHHHHHH

**>LytE[L3cat]**

MTYKVKSGDSLSKIASKYGTTVSKLKSLNGLKSDVIYVNQVLKVKGTSTSSSKPASSS
SSSSSKTSSTSLNVSKLVSDAKALVGTPYKWGGTTTSGFDCSGFIWYVLNKQTSVGR
TSTAGYWSSMKSIASPSVGDFVFFTTYKSGPSHMGIYIGNNSFIHAGSDGVQISSLNN
SYWKPRYLGAKRFLEHHHHHH

**>LytE[cat]**

MGTSTSSSKPASSSSSSSSKTSSTSLNVSKLVSDAKALVGTPYKWGGTTTSGFDCSG
FIWYVLNKQTSVGRTSTAGYWSSMKSIASPSVGDFVFFTTYKSGPSHMGIYIGNNSFI
HAGSDGVQISSLNNSYWKPRYLGAKRFLEHHHHHH

**>LytE[L1$^{LytF}$cat$^{LytE}$]**

MIKVKSGDSLWKLAQTYNTSVAALTSANHLSTTVLSIGQTLTIPGTSTSSSKPASSSSS
SSSKTSSTSLNVSKLVSDAKALVGTPYKWGGTTTSGFDCSGFIWYVLNKQTSVGRTS
TAGYWSSMKSIASPSVGDFVFFTTYKSGPSHMGIYIGNNSFIHAGSDGVQISSLNNSY
WKPRYLGAKRFLEHHHHHH

**>LytE[L1$^{XlyB}$cat$^{LytE}$]**

MYHVKKGDTLSGIAASHGASVKTLQSINHITDPNHIKIGQVIKLPGTSTSSSKPASSSSS
SSSKTSSTSLNVSKLVSDAKALVGTPYKWGGTTTSGFDCSGFIWYVLNKQTSVGRTS
TAGYWSSMKSIASPSVGDFVFFTTYKSGPSHMGIYIGNNSFIHAGSDGVQISSLNNSY
WKPRYLGAKRFLEHHHHHH

## 8.3 Strains used in this study

**Table 8.4: List of *B. subtilis* strains used in this study**

| Strain | Genotype | Source |
|---|---|---|
| **CK1** | *trpC2 ΔphoR::kan$^r$ phoP::phoR erm$^r$* | unpublished, Eric Botella |
| **CK6** | *trpC2 ΔphoR::kan$^r$ phoP::phoR [Met249Lys] erm$^r$* | Botella *et al.*, 2014 |
| **EB35** | *trpC2 ΔcwlO::kan$^r$* | unpublished, Eric Botella |
| **EB156** | *trpC2 ΔxylR ΔxylA ΔxylB $P_{xyl}$tagO spc$^r$* | Botella *et al.*, 2014 |
| **DN3000** | *trpC2 ΔcwlOleader* | Noone *et al.*, 2014 |
| **DN3061** | *trpC2 ΔlytE::kan$^r$* | unpublished, David Noone |
| **JL11** | *trpC2 ΔcwlO::$P_{cwlOleader}$lytE [L1$^{cwlS}$cat$^{lytE}$] kan$^r$* | pJL1 → WT168<br>unpublished, Jennifer Lennon |
| **JL13** | *trpC2 ΔcwlO::$P_{cwlOleader}$lytE [L1$^{lytF}$cat$^{lytE}$] kan$^r$* | pJL2 → WT168<br>unpublished, Jennifer Lennon |
| **JL15** | *trpC2 ΔcwlO::$P_{cwlOleader}$lytE [L1$^{xlyB}$cat$^{lytE}$] kan$^r$* | pJL4 → WT168<br>unpublished, Jennifer Lennon |
| **JL17** | *trpC2 ΔcwlO::$P_{cwlOleader}$lytE [L1$^{aaa}$cat$^{lytE}$] kan$^r$* | pJL5 → WT168<br>unpublished, Jennifer Lennon |
| **JL19** | *trpC2 ΔcwlO::$P_{cwlOleader}$lytE [L1$^{xlyB}$cat$^{lytE}$] kan$^r$ ΔlytE::cm$^r$* | pJL4 → LS089<br>unpublished, Jennifer Lennon |
| **JL21** | *trpC2 ΔcwlO::$P_{cwlOleader}$lytE-HA [L1$^{lytF}$cat$^{lytE}$] kan$^r$ spc$^r$* | pLU46 → JL13<br>unpublished, Jennifer Lennon |
| **JL22** | *trpC2 ΔcwlO::$P_{cwlOleader}$lytE-HA [L1$^{xlyB}$cat$^{lytE}$] kan$^r$ spc$^r$* | pLU46 → JL15<br>unpublished, Jennifer Lennon |
| **JL24** | *trpC2 ΔcwlO::$P_{cwlOleader}$lytE-HA [L1$^{xlyB}$cat$^{lytE}$] kan$^r$ spc$^r$ ΔlytE::cm$^r$* | pLU46 → JL19<br>unpublished, Jennifer Lennon |
| **LS089** | *trpC2 ΔlytE::cm$^r$* | Salzberg *et al.* (2013) |
| **LS103** | *trpC2 ΔcwlO::cm$^r$* | Salzberg *et al.* (2013) |
| **LU1** | *trpC2 ΔlytE::spc$^r$* | LFH → WT168 |
| **LU17** | *trpC2 ΔcwlO::$P_{cwlOleader}$lytE erm$^r$* | LFH → WT168 |
| **LU21** | *trpC2 ΔcwlO::$P_{cwlOleader}$lytE erm$^r$ ΔlytE::spc$^r$* | LFH → LU17 |
| **LU23** | *trpC2 ΔcwlO::$P_{cwlO}$lytE erm$^r$* | LFH → WT168 |
| **LU25** | *trpC2 ΔcwlO::$P_{cwlO}$lytE erm$^r$ ΔlytE::spc$^r$* | LFH → LU23 |
| **LU45** | *trpC2 lytE-HA kan$^r$* | pLU2 → WT168 |
| **LU47** | *trpC2 lytE-HA kan$^r$ ΔcwlO::cm$^r$* | pLU2 → LS103 |

| Strain | Genotype | Source |
|---|---|---|
| **LU61** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE-HA kan$^r$ erm$^r$ ΔlytE::spc$^r$* | pLU2 → LU21 |
| **LU63** | *trpC2 ΔcwlO::P$_{cwlO}$lytE-HA kan$^r$ erm$^r$ ΔlytE::spc$^r$* | pLU2 → LU25 |
| **LU72** | *trpC2 ΔcwlS* | LFH *dif* → WT168 |
| **LU74** | *trpC2 ΔlytF* | LFH *dif* → WT168 |
| **LU76** | *trpC2 ΔyqgT* | LFH *dif* → WT168 |
| **LU90** | *trpC2 ΔcwlS ΔcwlO::kan$^r$* | LFH → LU72 |
| **LU92** | *trpC2 ΔcwlS ΔlytE::kan$^r$* | LFH → LU72 |
| **LU96** | *trpC2 ΔyqgT ΔcwlO::kan$^r$* | LFH → LU76 |
| **LU98** | *trpC2 ΔyqgT ΔlytE::kan$^r$* | LFH → LU76 |
| **LU100** | *trpC2 ΔlytF ΔcwlO::kan$^r$* | LFH → LU74 |
| **LU102** | *trpC2 ΔlytF ΔlytE::kan$^r$* | LFH → LU74 |
| **LU106** | *trpC2 ΔcwlO::P$_{cwlOleader}$cwlS erm$^r$ ΔcwlS* | LFH → LU72 |
| **LU111** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytF erm$^r$ ΔlytF* | LFH → LU74 |
| **LU112** | *trpC2 ΔcwlO::P$_{cwlOleader}$yqgT erm$^r$ ΔyqgT* | LFH → LU76 |
| **LU116** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytF erm$^r$ ΔlytE::spc$^r$ ΔlytF* | LFH → LU111 |
| **LU123** | *trpC2 ΔcwlO::P$_{cwlO}$cwlS erm$^r$ ΔcwlS* | LFH → LU72 |
| **LU125** | trpC2 *ΔcwlO::P$_{cwlO}$lytF erm$^r$ ΔlytF* | LFH → LU74 |
| **LU134** | trpC2 *ΔcwlO::P$_{cwlO}$lytF erm$^r$ ΔlytE*::spc$^r$ *ΔlytF* | LFH → LU125 |
| **LU145** | trpC2 *cwlS*-HA *kan$^r$* | pLU7 → WT168 |
| **LU147** | trpC2 *lytF*-HA *kan$^r$* | pLU6 → WT168 |
| **LU149** | trpC2 *yqgT*-HA *kan$^r$* | pLU8 → WT168 |
| **LU153** | trpC2 *ΔcwlO::P$_{cwlOleader}$lytF*-HA *erm$^r$ kan$^r$ ΔlytE*::spc$^r$ *ΔlytF* | pLU6 → LU116 |
| **LU157** | trpC2 *ΔcwlO::P$_{cwlO}$lytF*-HA *erm$^r$ kan$^r$ ΔlytE*::spc$^r$ *ΔlytF* | pLU6 → LU134 |
| **LU159** | trpC2 *ΔcwlO::P$_{cwlOleader}$cwlS*-HA *erm$^r$ kan$^r$ ΔcwlS* | pLU7 → LU106 |
| **LU161** | trpC2 *ΔcwlO::P$_{cwlO}$cwlS*-HA *erm$^r$ kan$^r$ ΔcwlS* | pLU7 → LU123 |
| **LU163** | trpC2 *ΔcwlO::P$_{cwlOleader}$yqgT*-HA *erm$^r$ kan$^r$ ΔyqgT* | pLU8 → LU112 |
| **LU205** | trpC2 *ΔcwlO::P$_{cwlOleader}$lytE [L11cat] kan$^r$ ΔlytE*::cm$^r$ | LFH → LU239 |
| **LU207** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L12cat] kan$^r$* | Lin. pLU15 → WT168 |
| **LU209** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L12cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU207 |
| **LU212** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L13cat] kan$^r$* | Lin. pLU14 → WT168 |
| **LU213** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L13cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU212 |
| **LU216** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L21cat] kan$^r$* | Lin. pLU17 → WT168 |
| **LU217** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L21cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU216 |

| Strain | Genotype | Source |
|---|---|---|
| LU220 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L22cat] kan$^r$* | Lin. pLU18 → WT168 |
| LU222 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L22cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU220 |
| LU223 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L23cat] kan$^r$* | Lin. pLU19 → WT168 |
| LU226 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L23cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU223 |
| LU227 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L31cat] kan$^r$* | Lin. pLU20 → WT168 |
| LU229 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L31cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU227 |
| LU232 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L32cat] kan$^r$* | Lin. pLU21 → WT168 |
| LU233 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L32cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU232 |
| LU235 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L33cat] kan$^r$* | Lin. pLU22 → WT168 |
| LU238 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L33cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU235 |
| LU239 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L11cat] kan$^r$* | Lin. pLU16 → WT168 |
| LU261 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L1cat] kan$^r$* | Lin. pLU23 → WT168 |
| LU262 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L1cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU261 |
| LU265 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L2cat] kan$^r$* | Lin. pLU24 → WT168 |
| LU266 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L2cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU265 |
| LU269 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L3cat] kan$^r$* | Lin. pLU25 → WT168 |
| LU270 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L3cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU269 |
| LU272 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L311cat] kan$^r$* | Lin. pLU26 → 168 |
| LU273 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L311cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU272 |
| LU274 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L3331cat] kan$^r$* | Lin. pLU32 → 168 |
| LU275 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L3331cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU274 |
| LU276 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L331cat] kan$^r$* | Lin. pLU27 → WT168 |
| LU277 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L331cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU276 |
| LU278 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L33331cat] kan$^r$* | Lin. pLU33 → WT168 |
| LU279 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L33331cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU278 |
| LU280 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L233cat] kan$^r$* | Lin. pLU28 → WT168 |
| LU281 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L233cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU280 |
| LU282 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L2333cat] kan$^r$* | Lin. pLU34 → WT168 |
| LU283 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L2333cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU282 |
| LU286 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L222cat] kan$^r$* | Lin. pLU30 → WT168 |
| LU287 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L222cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU286 |
| LU288 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L212cat] kan$^r$* | Lin. pLU31 → WT168 |
| LU289 | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L212cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU288 |

| Strain | Genotype | Source |
|---|---|---|
| **LU312** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L213cat] kan$^r$* | Lin. pLU29 → WT168 |
| **LU313** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L213cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU312 |
| **LU314** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [cat] kan$^r$* | Lin. pLU37 → WT168 |
| **LU315** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L113cat] kan$^r$* | Lin. pLU38 → WT168 |
| **LU316** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L113cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU316 |
| **LU318** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L123cat]$^R$ kan$^r$* | Lin. pLU40 → WT168 |
| **LU320** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L123cat]$^R$ kan$^r$ ΔlytE::cm$^r$* | LFH → LU320 |
| **LU322** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L1333cat] kan$^r$* | Lin. pLU41 → WT168 |
| **LU323** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L1333cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU322 |
| **LU324** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L11111cat] kan$^r$* | Lin. pLU42 → WT168 |
| **LU325** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L11111cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU324 |
| **LU326** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L1221cat] kan$^r$* | Lin. pLU43 → WT168 |
| **LU327** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L1221cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU326 |
| **LU328** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L12221cat] kan$^r$* | Lin. pLU44 → 168 |
| **LU329** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE [L12221cat] kan$^r$ ΔlytE::cm$^r$* | LFH → LU328 |
| **LU331** | *trpC2 ΔlytE::spc$^r$* | LFH → WT168 |
| **LU335** | *trpC2 ΔlytE::P$_{lytE}$lytE [L123cat]$^R$ erm$^r$* | LFH → WT168 |
| **LU337** | *trpC2 ΔlytE::P$_{lytE}$lytE [L1cat] erm$^r$* | LFH → WT168 |
| **LU339** | *trpC2 ΔlytE::P$_{lytE}$lytE [L2cat] erm$^r$* | LFH → WT168 |
| **LU341** | *trpC2 ΔlytE::P$_{lytE}$lytE [L3cat] erm$^r$* | LFH → WT168 |
| **LU343** | *trpC2 ΔlytE::P$_{lytE}$lytE [L123cat]$^R$ erm$^r$ ΔcwlO::kan$^r$* | LFH → LU335 |
| **LU350** | *trpC2 ΔlytE::P$_{lytE}$lytE [L1cat] erm$^r$ ΔcwlO::kan$^r$* | LFH → LU337 |
| **LU352** | *trpC2 ΔlytE::P$_{lytE}$lytE [L2cat] erm$^r$ ΔcwlO::kan$^r$* | LFH → LU339 |
| **LU354** | *trpC2 ΔlytE::P$_{lytE}$lytE [L3cat] erm$^r$ ΔcwlO::kan$^r$* | LFH → LU341 |
| **LU372** | *trpC2 ΔlytE::P$_{lytE}$lytE [L31cat] erm$^r$* | LFH → WT168 |
| **LU374** | *trpC2 ΔlytE::P$_{lytE}$lytE [L331cat] erm$^r$* | LFH → WT168 |
| **LU383** | *trpC2 ΔlytE::P$_{lytE}$lytE [L31cat] erm$^r$ ΔcwlO::kan$^r$* | LFH → LU372 |
| **LU385** | *trpC2 ΔlytE::P$_{lytE}$lytE [L331cat] erm$^r$ ΔcwlO::kan$^r$* | LFH → LU374 |
| **LU398** | *trpC2 ΔlytE::P$_{lytE}$lytE [cat] erm$^r$* | LFH → WT168 |
| **LU399** | *trpC2 ΔlytE::P$_{lytE}$lytE [L11cat] erm$^r$* | LFH → WT168 |
| **LU401** | *trpC2 ΔlytE::P$_{lytE}$lytE [L21cat] erm$^r$* | LFH → WT168 |
| **LU403** | *trpC2 ΔlytE::P$_{lytE}$lytE [L13cat] erm$^r$* | LFH → WT168 |
| **LU405** | *trpC2 ΔlytE::P$_{lytE}$lytE [L213cat] erm$^r$* | LFH → WT168 |

| Strain | Genotype | Source |
|---|---|---|
| **LU407** | *trpC2 ΔlytE::P$_{lytE}$lytE [L11cat] erm$^r$ ΔcwlO::kan$^r$* | LFH → LU399 |
| **LU409** | *trpC2 ΔlytE::P$_{lytE}$lytE [L21cat] erm$^r$ ΔcwlO::kan$^r$* | LFH → LU401 |
| **LU411** | *trpC2 ΔlytE::P$_{lytE}$lytE [L13cat] erm$^r$ ΔcwlO::kan$^r$* | LFH → LU403 |
| **LU413** | *trpC2 ΔlytE::P$_{lytE}$lytE [L213cat] erm$^r$ ΔcwlO::kan$^r$* | LFH → LU405 |
| **LU414** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE-HA [L11cat] kan$^r$ spc$^r$ ΔlytE::cm$^r$* | pLU46 → LU205 |
| **LU415** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE-HA [L13cat] kan$^r$ spc$^r$ ΔlytE::cm$^r$* | pLU46 → LU213 |
| **LU416** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE-HA [L21cat] kan$^r$ spc$^r$ ΔlytE::cm$^r$* | pLU46 → LU217 |
| **LU417** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE-HA [L31cat] kan$^r$ spc$^r$ ΔlytE::cm$^r$* | pLU46 → LU229 |
| **LU418** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE-HA [L1cat] kan$^r$ spc$^r$ ΔlytE::cm$^r$* | pLU46 → LU262 |
| **LU419** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE-HA [L2cat] kan$^r$ spc$^r$ ΔlytE::cm$^r$* | pLU46 → LU266 |
| **LU421** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE-HA [L123cat]$^R$ kan$^r$ spc$^r$ ΔlytE::cm$^r$* | pLU46 → LU320 |
| **LU422** | *trpC2 ΔlytE::P$_{lytE}$lytE-HA [L123cat]$^R$ erm$^r$ spc$^r$ ΔcwlO::kan$^r$* | pLU46 → LU343 |
| **LU424** | *trpC2 ΔlytE::P$_{lytE}$lytE-HA [L331cat] erm$^r$ spc$^r$ ΔcwlO::kan$^r$* | pLU46 → LU385 |
| **LU425** | *trpC2 ΔlytE::P$_{lytE}$lytE-HA [L11cat] erm$^r$ spc$^r$ ΔcwlO::kan$^r$* | pLU46 → LU407 |
| **LU427** | *trpC2 ΔlytE::P$_{lytE}$lytE-HA [L13cat] erm$^r$ spc$^r$ ΔcwlO::kan$^r$* | pLU46 → LU411 |
| **LU432** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE-HA [L213cat] kan$^r$ spc$^r$ ΔlytE::cm$^r$* | pLU46 → LU313 |
| **LU435** | *trpC2 ΔlytE::P$_{lytE}$lytE-HA [L1cat] erm$^r$ spc$^r$ ΔcwlO::kan$^r$* | pLU46 → LU350 |
| **LU440** | *trpC2 ΔcwlO::P$_{cwlOleader}$lytE-HA [L3cat] kan$^r$ spc$^r$ ΔlytE::cm$^r$* | pLU46 → LU270 |
| **LU442** | *trpC2 ΔlytE::P$_{lytE}$lytE-HA [L2cat] erm$^r$ spc$^r$ ΔcwlO::kan$^r$* | pLU46 → LU352 |
| **LU443** | *trpC2 ΔlytE::P$_{lytE}$lytE-HA [L213cat] erm$^r$ spc$^r$ ΔcwlO::kan$^r$* | pLU46 → LU413 |
| **WT168** | *trpC2* | Laboratory stock, Kunst *et al.*, 1997 |

[*cm$^r$*] chloramphenicol resistance gene; [*erm$^r$*] erythromycin resistance gene; [*kan$^r$*] kanamycin resistance gene; [*spc$^r$*] spectinomycin resistance gene; [P$_{lytE}$] *lytE* promoter; [P$_{cwlOleader}$] *cwlO* promoter including *cwlO* leader sequence; [P$_{cwlO}$] *cwlO* promoter excluding *cwlO* leader sequence; [Δ] deletion; [::] replacement; [HA] HA-epitope tag; [*lytE[LXcat]*] *lytE* open reading frame with permuted LysM motifs; [*lytE[cat]*] *lytE* open reading frame without LysM motifs; [LFH → WT168] transformation of WT168 with a long flanking homology fragment that integrates by double cross-over; [LFH *dif* → WT168] transformation of WT168 with a long flanking homology fragment with *dif*-sites that integrates by double cross-over; [pLUX → WT168] transformation of WT168 with a plasmid that integrates over single cross-over; [Lin. pLUX → WT168] transformation of WT168 with a linearized plasmid that integrates by double cross-over

**Table 8.5: List of *E. coli* strains used in this study**

| Strain | Genotype | Source |
|---|---|---|
| **BL21(DE3)** | *F ompT hsdSB* ($r_B$ $m_B$) *gal dcm* (*DE3*) | Studier and Moffatt, 1986 |
| **TG1** | *supE hsdΔ5 thi Δ(laC-proAB) F´[traD36 proAB$^+$ lacI$^q$ lacZΔM15]* | Sambrook *et al.*, 1989 |
| **TP611** | *recBC hsdRM cya610 pcn* | Glaser *et al.*, 1993 |

## 8.4 Plasmids used in this study

**Table 8.6: List of plasmids used in this study**

| Plasmid | Genotype | Source |
|---|---|---|
| **pDG780** | plasmid with *kan*$^r$ cassette | Guerout-Fleury *et al.*, 1995 |
| **pDG646** | plasmid with *erm*$^r$ cassette | Guerout-Fleury *et al.*, 1995 |
| **pDG1727** | plasmid with *spc*$^r$ cassette | Guerout-Fleury *et al.*, 1995 |
| **pDG1661** | plasmid with *cm*$^r$ cassette | Guerout-Fleury *et al.*, 1995 |
| **pETLytE** | pET21b+ derivative, protein expression vector for expression of the wild-type LytE protein | unpublished, David Noone |
| **pJL1** | pLU10 derivative enabling integration of variant *lytE [L1*$^{cwlS}$*cat*$^{lytE}$*]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | unpublished, Jennifer Lennon |
| **pJL2** | pLU10 derivative enabling integration of variant *lytE [L1*$^{lytF}$*cat*$^{lytE}$*]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | unpublished, Jennifer Lennon |
| **pJL4** | pLU10 derivative enabling integration of variant *lytE [L1*$^{xlyB}$*cat*$^{lytE}$*]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | unpublished, Jennifer Lennon |
| **pJL5** | pLU10 derivative enabling integration of variant *lytE [L1*$^{aaa}$*cat*$^{lytE}$*]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | unpublished, Jennifer Lennon |
| **pLU2** | pDG780 derivative enabling C-terminal HA-tagging of *lytE* over single cross-over, *kan*$^r$ cassette | This work |
| **pLU9** | pDG780 derivative with upstream homology to the *cwlO* locus (including *cwlO* promoter and *cwlO* leader sequence) | This work |
| **pLU10** | pLU9 derivative with up- and downstream homology to the *cwlO* locus (including *cwlO* promoter and *cwlO* leader sequence) | This work |
| **pLU14** | pLU10 derivative enabling integration of variant *lytE [L13cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU15** | pLU10 derivative enabling integration of variant *lytE [L12cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU16** | pLU10 derivative enabling integration of variant *lytE [L11cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU17** | pLU10 derivative enabling integration of variant *lytE [L21cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU18** | pLU10 derivative enabling integration of variant *lytE [L22cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU19** | pLU10 derivative enabling integration of variant *lytE [L23cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |

| Plasmid | Genotype | Source |
|---|---|---|
| **pLU20** | pLU10 derivative enabling integration of variant *lytE [L31cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU21** | pLU10 derivative enabling integration of variant *lytE [L32cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU22** | pLU10 derivative enabling integration of variant *lytE [L33cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU23** | pLU10 derivative enabling integration of variant *lytE [L1cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU24** | pLU10 derivative enabling integration of variant *lytE [L2cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU25** | pLU10 derivative enabling integration of variant *lytE [L3cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU26** | pLU20 derivative enabling integration of variant *lytE [L311cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU27** | pLU20 derivative enabling integration of variant *lytE [L331cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU28** | pLU19 derivative enabling integration of variant *lytE [L233cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU29** | pLU19 derivative enabling integration of variant *lytE [L213cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU30** | pLU18derivative enabling integration of variant *lytE [L222cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU31** | pLU18 derivative enabling integration of variant *lytE [L212cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU32** | pLU20 derivative enabling integration of variant *lytE [L3331cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU33** | pLU20 derivative enabling integration of variant *lytE [L33331cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU34** | pLU19 derivative enabling integration of variant *lytE [L2333cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU37** | pLU10 derivative enabling integration of variant *lytE [cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU38** | pLU14 derivative enabling integration of variant *lytE [L113cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU39** | pLU14 derivative enabling integration of variant *lytE [L1113cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU40** | pLU14 derivative enabling integration of variant *lytE [L123cat]*$^R$ at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU41** | pLU14 derivative enabling integration of variant *lytE [L1333cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |

| Plasmid | Genotype | Source |
|---|---|---|
| **pLU42** | pLU16 derivative enabling integration of variant *lytE [L11111cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU43** | pLU16 derivative enabling integration of variant *lytE [L1221cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU44** | pLU16 derivative enabling integration of variant *lytE [L12221cat]* at the *cwlO* locus by double cross-over, *kan*$^r$ cassette | This work |
| **pLU46** | pLU2 derivative enabling C-terminal HA-tagging of *lytE* over single cross-over, *spc*$^r$ cassette | This work |
| **pLU47** | pET21b+ derivative, protein expression vector for expression of the variant LytE [L123cat]$^R$ | This work |
| **pLU48** | pET21b+ derivative, protein expression vector for expression of the variant LytE [L31cat] | This work |
| **pLU49** | pET21b+ derivative, protein expression vector for expression of the variant LytE [L13cat] | This work |
| **pLU50** | pET21b+ derivative, protein expression vector for expression of the variant LytE [L1cat] | This work |
| **pLU51** | pET21b+ derivative, protein expression vector for expression of the variant LytE [L3cat] | This work |
| **pLU52** | pET21b+ derivative, protein expression vector for expression of the variant LytE [cat] | This work |
| **pLU53** | pET21b+ derivative, protein expression vector for expression of the variant LytE [L1$^{LytF}$cat$^{LytE}$] | This work |
| **pLU54** | pET21b+ derivative, protein expression vector for expression of the variant LytE [L1$^{XlyB}$cat$^{LytE}$] | This work |

[*cm*$^r$] chloramphenicol resistance gene; [*erm*$^r$] erythromycin resistance gene; [*kan*$^r$] kanamycin resistance gene; [*spc*$^r$] spectinomycin resistance gene

## 8.5 Primers and oligonucleotides used in this study

**Table 8.7: List of primers used in this study**

| Primer name | Sequence* |
|---|---|
| 16S_Bsu_F | GTCAGCTCGTGTCGTGAGATG |
| 16S_Bsu_R | CAACTGAATGCTGGCAACTAAGATC |
| Aaa_LysM_Fwd | cgcggatccacgttatttgcaggagcggcatctgcacaaGTAAAACCGGGTGAATCAG |
| Aaa_LysM_Rev | atgtagatgacgttttgctgcttgaagacgatgaagatgaagcaggttttgagctgcttgtgcttgttcc TACTTTTAGTACTTGGTTTG |
| Catalytic_LytE_Fwd_Spac | ttcaagcagcaaaacgtcatctacatcacttaatgtgagcaagctggtttctgatgcaaaagcgttagtc GGAACGCCATATAAATGGGG |
| Catalytic_LytE_Rev_Ery | attatgtcttttgcgcagtcggcGAAAAAAACCCGTTCCAATGAACGG |
| Catalytic_LytE_Rev_XmaI | tccccccgggGAAAAAAACCCGTTCCAATGAACGG |
| CwlO_DF_LysM | ccgctcgagATTAACATATACCCTTTTCGAATA |
| CwlO_DR_LysM | cggggtaccCCTTCTTCATTCGTAATGCCCAGGT |
| CwlO_UF | CCTTCATCTGTGGAGCAG |
| CwlO_UF_LysM | aaggaaaaaagcggccgcATCATTATATAAAACCGGCTT |
| CwlO_UF_SmaI | gcatcccgggCCTTCATCTGTGGAGCAG |
| CwlO_UR | cgtagctgtaatgatttgctttttcatTTTTTATATCCTCCCTTTTACTGTGAACC |
| CwlO_UR_CwlS | aagccggctacaatcttctttttcatTTTTTATATCCTCCCTTTTACTGTGAA |
| CwlO_UR_LysM | taacgtggatcctaaaacaactgctgtcgtagctgtaatgatttgctttttcat TTTTTATATCCTCCCTTTTAC |
| CwlO_UR_LytF | gccctgctgctaatttctttttcatTTTTTATATCCTCCCTTTTACTGTGAA |
| CwlO_UR_Yqgt | gtttcaccgtaataaacgccatcatTTTTTATATCCTCCCTTTTACTGTGAA |
| CwlS_Do_Fwd | tatattatgtaaactaatatgcaagAAAATGAGTATGACAAAAAGCC |
| CwlS_Do_Rev | TCTGGGGAAATAGGCTGCTCT |
| CwlS_Fwd | ATGAAAAAGAAGATTGTAGCCGGC |
| CwlS_LysM_Fwd | cgcggatccacgttatttgcaggagcggcatctgcacaaATAAAAGTAAAAAGCGGTGA |
| CwlS_LysM_Rev | atgtagatgacgttttgctgcttgaagacgatgaagatgaagcaggttttgagctgcttgtgcttgttcc AGGAACCTTAAGACTTTGTC |
| Cwls_Rev_Ery | tcgattatgtcttttgcgcagtcggcTAAACAAAAAAGCCTATAAAAGGC |
| CwlS_Tag_F_BamHI | cgggatccGTCGGAGTGCCTTACCGATGGGGC |
| CwlS_Tag_R_HA_EcoRI | ggaattctaaacaaaaaagcctataaaaggctttttgtcatactcatttttattatgcataatccg gaacatcatacggataAAAATAACTTCTTGCGCCCAAGTA |
| CwlS_Up_Fwd | CGGCAAGCATATTAAAACAAGTGAG |
| CwlS_Up_Rev | tatattctaggaagtaggtattcacTTCATTCAACCTCCTAAACTG |
| Exp_Cat_Rev | ccgctcgagGAATCTTTTCGCACCGAGGTAACG |
| Exp_L1_Fwd | ggaattccatatgAGCATTAAGGTGAAAAAAGGCGACA |
| Exp_L2_Fwd | ggaattccatatgACATACAAAGTAAAGAGCGGGGACA |
| Exp_L3_Fwd | ggaattccatatgACTTATAAAGTGAAGAGCGGAGACA |
| Exp_LytF_Fwd | ggaattccatatgATTAAGGTCAAAAGCGGAGATTCTTT |
| Exp_Spac_Fwd | ggaattccatatgGGAACAAGCACAAGCAGCTCAAAAC |
| Exp_XlyB_Fwd | ggaattccatatgTATCATGTAAAAAAAGGTGACACACTG |
| L1_Fwd_Link | ggcaaatctactagttcaaaaagcagcagttcttcttcctcttcttct AGCATTAAGGTGAAAAAAGGC |
| L1_Fwd_Sig_BamHI | cgcggatccacgttatttgcaggagcggcatctgcacaa AGCATTAAGGTGAAAAAAGGC |
| L1_Rev_Link | agaagaagaggaagaagaactgctgcttttgaactagtagatttgcc GTTAATCGATAAAGTTTGTCCC |
| L1_Rev_Spac | atgtagatgacgttttgctgcttgaagacgatgaagatgaagcaggttttgagctgcttgtgcttgttcc GTTAATCGATAAAGTTTGTC |
| L2_Fwd_Link | ggcaaatctactagttcaaaaagcagcagttcttcttcctcttcttct ACATACAAAGTAAAGAGCGGG |

| Primer name | Sequence* |
|---|---|
| L2_Fwd_Sig_BamHI | cgcggatccacgttatttgcaggagcggcatctgcacaa ACATACAAAGTAAAGAGCGGG |
| L2_Rev_Link | agaagaagaggaagaagaactgctgctttttgaactagtagatttgcc TTTCAGTTTCAGGACTTGTCC |
| L2_Rev_Spac | atgtagatgacgttttgctgcttgaagacgatgaagatgaagcaggttttgagctgcttgtgcttgttcc TTTCAGTTTCAGGACTTGTC |
| L3_Fwd_Link | ggcaaatctactagttcaaaaagcagcagttcttcttcctcttcttct ACTTATAAAGTGAAGAGCGGA |
| L3_Fwd_Sig_BamHI | cgcggatccacgttatttgcaggagcggcatctgcacaa ACTTATAAAGTGAAGAGCGGA |
| L3_Rev_Link | agaagaagaggaagaagaactgctgctttttgaactagtagatttgcc TTTCACCTTCAATACTTGGTTG |
| L3_Rev_Spac | atgtagatgacgttttgctgcttgaagacgatgaagatgaagcaggttttgagctgcttgtgcttgttcc TTTCACCTTCAATACTTGGT |
| LytE_DF | ATGAAAAAGCAAATCATTACAGCTACG |
| LytE_Do_Fwd_Ery | ggttgccagagttaaaggatcATTAGACGGAGCACAGGCGTTTT |
| LytE_Do_Fwd_Spec | caataaacccttgccctcgctacgTACTGGAAGCCTCGTTACCTCGG |
| LytE_Do_Fwd_Spec | caataaacccttgccctcgctacgATTAGACGGAGCACAGGCGTTTT |
| LytE_Do_Rev | GTATTTGCCGTATGCGCTCAGGC |
| LytE_DR_Ery | attatgtcttttgcgcagtcggcGAAAAAAACCCGTTCCAATGAA |
| LytE_Probe_F | TATCGATTAACGGCAAATCTACAAGTT |
| LytE_Probe_T7_R | taatacgactcactatagggGCAGTGCTTGTTCTGCCCACACTT |
| LytE_qPCR_F2_ | GGAGGAAAATATGAAAAAGCAAATCA |
| LytE_qPCR_R2 | TCGCTCTCATACGTATCAACACTCT |
| LytE_Spacer_Fwd_Sig | cgcggatccacgttatttgcaggagcggcatctgcacaaGGAACAAGCACAAGCAGC |
| LytE_Tag_F_BamHI | cgggatccTTAGTCGGAACGCCATATAAATGG |
| LytE_Tag_R_HA_EcoRI | ggaattcgaaaaaaacccgttccaatgaacgggttttctctaaaaattattatgcata atccggaacatcatacggataGAATCTTTTCGCACCGAGGTAACG |
| LytE_Up_Fwd | AACGGTAGGCTCTTATGTAAGGC |
| LytE_Up_Fwd_2 | AACGGTAGGCTCTTATGTAAGGC |
| LytE_Up_Rev_Spec | cgttacgttattagcgagccagtcCCTAAAACAACTGCTGTCGTAGC |
| LytF_Do_Fwd | tatattatgtaaactaatatgcaagTATTAGAGCCTAAAGACGTTGATGC |
| LytF_Do_Rev | CCGGGATTTGTTTATGTGAGAA |
| LytF_Fwd | ATGAAAAAGAAATTAGCAGCAGGG |
| LytF_LysM_Fwd | cgcggatccacgttatttgcaggagcggcatctgcacaaATTAAGGTCAAAAGCGGAG |
| LytF_LysM_Rev | atgtagatgacgttttgctgcttgaagacgatgaagatgaagcaggttttgagctgcttgtgcttgttcc TGGAATAGTAAGCGTCTGG |
| LytF_qPCR_Fwd | TGTCGGAGATTTCGTTTTCTTTTC |
| LytF_qPCR_fwd | TGTCGGAGATTTCGTTTTCTTTTC |
| LytF_qPCR_Rev | CCCGTTGCCTAGGTAAATTCC |
| LytF_qPCR_rev | CCCGTTGCCTAGGTAAATTCC |
| LytF_Rev_Ery | tcgattatgtcttttgcgcagtcggcGCATCAACGTCTTTAGGCTCTAATA |
| LytF_Tag_F_BamHI | cgggatccCTTGGGGTTCCGTATCGCTGGGGA |
| LytF_Tag_R_HA_EcoRI | ggaattcgcatcaacgtctttaggctctaataaaaaactgtacagccgttaaggccgcacagtttctg tttttattatgcataatccggaacatcatacggataGAAATATCGTTTTGCACCGAGATA |
| LytF_Up_Fwd | CTTTTGGTTTATCCTTCGGCCTT |
| LytF_Up_Rev | tatattctaggaagtaggtattcacCTTAGCTCTCCTTTTTCCCCC |
| MreBH_qPCR_F | GCAAAACCCGCTTGAATCTG |
| MreBH_qPCR_R | TGCCTTTTGCAGCTTATCAATC |
| OE93_Fwd_Ery | ggttgccagagttaaaggatcGAGCAATCCTTACTGGAAAGCAGC |
| OE94_Rev | ATCGACAAGCGTTACGTTACCGAC |
| SigI_Chip_Fwd | AAATTCTCTGCTCGCATTTGC |
| SigI_Chip_Rev | GGAAAGCTCGGCTTCAATCAG |
| Signal_LytE_Fwd | ATGAAAAAGCAAATCATTACAGCTACG |

| Primer name | Sequence* |
|---|---|
| Signal_LytE_Rev | TTGTGCAGATGCCGCTCCTGCAA |
| Signal_LytE_Rev_Spec | cgttacgttattagcgagccagtcTTGTGCAGATGCCGCTCCTGCAA |
| SigX_qPCR_F | GCGCTTCTCGGAAAGCAAAG |
| SigX_qPCR_R | AAAAGCTCCATATGTTTCCGAAGTAC |
| Spacer_Check_Rev | GATGTAGATGACGTTTTGC |
| Spec_Fwd_EcorI | ccggaattccggCTAATAACGTAACGTGACTGGCA |
| Spec_Rev_SalI | ttccgcggccgctatggccgacgtcgacgcgtAGGGTTTATTGTTTTCTAAAATCTG |
| XlyB_LyM_Fwd | cgcggatccacgttatttgcaggagcggcatctgcacaaTATCATGTAAAAAAAGGTGAC |
| XlyB_LysM_Rev | atgtagatgacgttttgctgcttgaagacgatgaagatgaagcaggttttgagctgcttgtgcttgttcc AGGCAGTTTGATCACCTG |
| YocH_northern_F | GCAGCAAACACAGCTACTGAGAAT |
| YocH_northern_R | GCTCTTTTTGCACAGCCTCTT |
| YqgT_Do_Fwd | tatattatgtaaactaatatgcaagACTGAAACTTTTTCTTTTTTCTGC |
| YqgT_Do_Rev | TTGGAGCGGCCATTTCGCTGT |
| YqgT_Fwd | ATGATGGCGTTTATTACGGTG |
| YqgT_Rev_Ery | tcgattatgtcttttgcgcagtcggcTCTCATATTTGCTTCGGCAGAA |
| YqgT_Tag_F_BamHI | cgggatccACGGAGAATCCGCCTGACAGACTG |
| YqgT_Tag_R_HA_EcoRI | ggaattctctcatatttgcttcggcagaaaaaagaaaaagtttcagttattatgcataatccgg aacatcatacggataCTCATGAAAAAATAACGCCCTCCA |
| YqgT_Up_Fwd | TTACTCGAGGCTGATTACACTG |
| YqgT_Up_Rev | tatattctaggaagtaggtattcacTCATCCCCCTTTTTTGTTACTATA |
| YvcE_3qPCR_Orf*2 | GGCGGATCTCAAGGATACATAGATG |
| YvcE_5qPCR_Orf*2 | GTCTGCGTCCACAATAGATGAAAC |

* lower case = 5´ overhang, upper case = annealing sequence

# 9 References

**Adams, D.W. and Errington, J.** (2009) Bacterial cell division: assembly, maintenance and disassembly of the Z ring. *Nature Reviews Microbiology*, 7, 642–653.

**Afkham, A., Vlamakis, H., Losick, R. and Kolter, R.** (2007) Thinking About *Bacillus subtilis* as a Multicellular Organism. *Current Opinion in Microbiology*, 10, 638–643.

**Akcapinar, G.B., Kappel, L., Sezerman, O.U. and Seidl-Seiboth, V.** (2015) Molecular diversity of LysM carbohydrate-binding motifs in fungi. *Current Genetics*, 61, 103–113.

**Albertini, A., Amati, G., Andersen, K.K., Arnaud, M., Ashikaga, S., Aymerich, S., Bessieres, P., Boland, F., Brignell, S.C., Bron, S., Bunai, K., Chapuis, J., Christiansen, L.C., Danchin, A., Débarbouille, M., Dervyn, E., Deuerling, E., Devine, K.M., Devine, S.K., Dreesen, O., Ehrlich, S.D., Errington, J., Eschevins, C., Fillinger, S., Foster, S.J., Fujita, Y., Fukushima, T., Galizzi, A., Gardan, R., Haga, K., Harwood, C.R., Hecker, M., Hosoya, D., Hullo, M.F., Imamura, D., Ishikawa, S., Ishimaru, M., Ishio, I., Kakeshita, H., Karamata, D., Kasahara, Y., Kawamura, F., Kobayashi, K., Koga, K., Koski, P., Kuwana, R., Le Coq, D., Masson, A., Mauël, C., Meima, R., Mellado, R.P., Moir, A., Moriya, S., Nagakawa, E., Nakai, S., Nanamiya, H., Nygaard, P., Ogasawara, N., Ogura, M., Ohanan, T., O'Reilly, M., O'Rourke, M., Pooley, H.M., Pragai, Z., Rapoport, G., Rawlins, J.P., Rivas, L., Rivolta, C., Sadaie, A., Sadaie, Y., Sarvas, M., Sato, T., Saxild, H.H., Scanlan, E., Schumann, W., Seegers, J.F., Sekiguchi, J., Sekowska, A., Séror, S.J., Simon, M., Stragier, P., Studer, R., Takamatsu, H., Takeuchi, M., Tanaka, T., Thomaides, H.B., Vagner, V., van Dijl, J.M., Watabe, K., Wipat, A., Yamamoto, H., Yamamoto, M., Yamamoto, Y., Yamane, K., Yata, K., Yoshida, K., Yoshikawa, H. and Zuber, U.** (2003) Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci.*, 100, 4678–4683.

**Alex, L.A. and Simon, M.I.** (1994) Protein histidine kinases and signal transduction in prokaryotes and eukaryotes. *Trends in Genetics*, 10, 133–138.

**Anagnostopoulus, C. and Spizizen, J.** (1960) Requirements for Transformation in *Bacillus subtilis*. *Journal of Bacteriology*, 81, 741–746.

**Andrade, M.A., Perez-Iratxeta, C. and Ponting, C.P.** (2001) Protein repeats: structures, functions, and evolution. *Journal of Structural Biology*, 134, 117–131.

**Andre, G., Leenhouts, K., Hols, P. and Dufrene, Y.F.** (2008) Detection and Localization of Single LysM-Peptidoglycan Interactions. *Journal of Bacteriology*, 190, 7079–7086.

**Antolín-Llovera, M., Ried, M.K., Binder, A. and Parniske, M.** (2012) Receptor Kinase Signaling Pathways in Plant-Microbe Interactions. *Annual Review of Phytopathology*, 50, 451–473.

**Apic, G., Gough, J. and Teichmann, S.A.** (2001) Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *Journal of Molecular Biology*, 310, 311–325.

**Ara, K., Ozaki, K., Nakamura, K., Yamane, K., Sekiguchi, J. and Ogasawara, N.** (2007) *Bacillus* minimum genome factory: effective utilization of microbial genome information. *Biotechnology and Applied Biochemistry*, 46, 169–178.

**Asai, K., Baik, S.H., Kasahara, Y., Moriya, S. and Ogasawara, N.** (2000) Regulation of the transport system for C4-dicarboxylic acids in *Bacillus subtilis*. *Microbiology*, 146, 263–271.

**Asai, K., Ootsuji, T., Obata, K., Matsumoto, T., Fujita, Y. and Sadaie, Y.** (2007) Regulatory role of RsgI in *sigI* expression in *Bacillus subtilis*. *Microbiology*, 153, 92–101.

**Atilano, M.L., Pereira, P.M., Vaz, F., Catalao, M.J., Reed, P., Grilo, I.R., Sobral, R.G., Ligoxygakis, P., Pinho, M.G. and Filipe, S.R.** (2014) Bacterial autolysins trim cell surface peptidoglycan to prevent detection by the *Drosophila* innate immune system. *eLife*, 3, 1–23.

**Attwood, P.V.** (2013) Histidine kinases from bacteria to humans. *Biochemical Society Transactions*, 41, 1023–1028.

**Barbe, V., Cruveiller, S., Kunst, F., Lenoble, P., Meurice, G., Sekowska, A., Vallenet, D., Wang, T., Moszer, I., Medigue, C. and Danchin, A.** (2009) From a consortium sequence to a unified sequence: the *Bacillus subtilis 168* reference genome a decade later. *Microbiology*, 155, 1758–1775.

**Bateman, A. and Bycroft, M.** (2000) The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). *Journal of Molecular Biology*, 299, 1113–1119.

**BCC Research** (2014) *Global Markets for Enzymes in Industrial Applications:* BCC Research.

**BCC Research** (2015) *Microbial Products: Technologies, Applications and Global Markets:* BCC Research.

**Bick, M.J., Lamour, V., Rajashankar, K.R., Gordiyenko, Y., Robinson, C.V. and Darst, S.A.** (2009) How to switch off a histidine kinase: crystal structure of *Geobacillus stearothermophilus* KinB with the inhibitor Sda. *Journal of Molecular Biology*, 386, 163–177.

**Bijlsma, J.J. and Groisman, E.A.** (2003) Making informed decisions: regulatory interactions between two-component systems. *Trends in Microbiology*, 11, 359–366.

**Biller, S.J., Wayne, K.J., Winkler, M.E. and Burkholder, W.F.** (2011) The Putative Hydrolase YycJ (WalJ) Affects the Coordination of Cell Division with DNA Replication in *Bacillus subtilis* and May Play a Conserved Role in Cell Wall Metabolism. *Journal of Bacteriology*, 193, 896–908.

**Bisicchia, P., Lioliou, E., Noone, D., Salzberg, L.I., Botella, E., Hübner, S. and Devine, K.M.** (2010) Peptidoglycan metabolism is controlled by the WalRK (YycFG) and PhoPR two-component systems in phosphate-limited *Bacillus subtilis* cells. *Molecular Microbiology*, 75, 972–989.

**Bisicchia, P., Noone, D., Lioliou, E., Howell, A., Quigley, S., Jensen, T., Jarmer, H. and Devine, K.M.** (2007) The essential YycFG two-component system controls cell wall metabolism in *Bacillus subtilis*. *Molecular Microbiology*, 65, 180–200.

**Bjorklund, A.K., Ekman, D. and Elofsson, A.** (2006) Expansion of protein domain repeats. *PLoS Computational Biology*, 2, 1–12.

**Blencke, H.-M., Homuth, G., Ludwig, H., Mäder, U., Hecker, M. and Stülke, J.** (2003) Transcriptional profiling of gene expression in response to glucose in *Bacillus subtilis*: regulation of the central metabolic pathways. *Metabolic Engineering*, 5, 133–149.

**Bloor, A.E. and Cranenburgh, R.M.** (2006) An Efficient Method of Selectable Marker Gene Excision by Xer Recombination for Gene Replacement in Bacterial Chromosomes. *Applied and Environmental Microbiology*, 72, 2520–2525.

**Bodenhofer, U., Bonatesta, E., Horejs-Kainrath, C. and Hochreiter, S.** (2015) msa: an R package for multiple sequence alignment. *Bioinformatics*, 31, 3997–3999.

**Bosma, T., Kanninga, R., Neef, J., Audouy, S.A.L., van Roosmalen, M.L., Steen, A., Buist, G., Kok, J., Kuipers, O.P., Robillard, G. and Leenhouts, K.** (2006) Novel surface display system for proteins on non-genetically modified gram-positive bacteria. *Applied and Environmental Microbiology*, 72, 880–889.

**Botella, E., Devine, S.K., Hubner, S., Salzberg, L.I., Gale, R.T., Brown, E.D., Link, H., Sauer, U., Codée, J.D., Noone, D. and Devine, K.M.** (2014) PhoR autokinase

activity is controlled by an intermediate in wall teichoic acid metabolism that is sensed by the intracellular PAS domain during the PhoPR-mediated phosphate limitation response of *Bacillus subtilis*. *Molecular Microbiology*, 94, 1242–1259.

**Boudreau, M.A., Fisher, J.F. and Mobashery, S.** (2012) Messenger functions of the bacterial cell wall-derived muropeptides. *Biochemistry*, 51, 2974–2990.

**Bourret, R.B.** (2010) Receiver domain structure and function in response regulator proteins. *Current Opinion in Microbiology*, 13, 142–149.

**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*. *Journal of Bacteriology*, 175, 7931–7937.

**Breukink, E. and Kruijff, B. de** (2006) Lipid II as a target for antibiotics. *Nature Reviews. Drug discovery*, 5, 321–332.

**Bridier, A., Sanchez-Vizuete, P., Guilbaud, M., Piard, J.-C., Naitali, M. and Briandet, R.** (2015) Biofilm-associated persistence of food-borne pathogens. *Food Microbiology*, 45, 167–178.

**Brown, S., Santa Maria, J.P. and Walker, S.** (2013) Wall Teichoic Acids of Gram-Positive Bacteria. *Annual Review of Microbiology*, 67, 313–336.

**Buescher, J.M., Liebermeister, W., Jules, M., Uhr, M., Muntel, J., Botella, E., Hessling, B., Kleijn, R.J., Le Chat, L., Lecointe, F., Mader, U., Nicolas, P., Piersma, S., Rugheimer, F., Becher, D., Bessieres, P., Bidnenko, E., Denham, E.L., Dervyn, E., Devine, K.M., Doherty, G., Drulhe, S., Felicori, L., Fogg, M.J., Goelzer, A., Hansen, A., Harwood, C.R., Hecker, M., Hubner, S., Hultschig, C., Jarmer, H., Klipp, E., Leduc, A., Lewis, P., Molina, F., Noirot, P., Peres, S., Pigeonneau, N., Pohl, S., Rasmussen, S., Rinn, B., Schaffer, M., Schnidder, J., Schwikowski, B., van Dijl, J.M., Veiga, P., Walsh, S., Wilkinson, A.J., Stelling, J., Aymerich, S. and Sauer, U.** (2012) Global Network Reorganization During Dynamic Adaptations of *Bacillus subtilis* Metabolism. *Science*, 335, 1099–1103.

**Buist, G., Kok, J., Leenhouts, K.J., Dabrowska, M., Venema, G. and Haandrikman, A.J.** (1995) Molecular cloning and nucleotide sequence of the gene encoding the major peptidoglycan hydrolase of *Lactococcus lactis*, a muramidase needed for cell separation. *Journal of Bacteriology*, 177, 1554–1563.

**Buist, G., Steen, A., Kok, J. and Kuipers, O.P.** (2008) LysM, a widely distributed protein motif for binding to (peptido)glycans. *Molecular Microbiology*, 68, 838–847.

**Cai, S.J. and Inouye, M.** (2002) EnvZ-OmpR interaction and osmoregulation in *Escherichia coli*. *The Journal of Biological Chemistry*, 277, 24155–24161.

**Cai, X.-H., Zhang, Q., Shi, S.-Y. and Ding, D.-F.** (2005) Searching for Potential Drug Targets in Two-component and Phosphorelay Signal-transduction Systems using Three-dimensional Cluster Analysis. *Acta Biochimica et Biophysica Sinica*, 37, 293–302.

**Cairns, L.S., Hobley, L. and Stanley-Wall, N.R.** (2014) Biofilm formation by *Bacillus subtilis*: new insights into regulatory strategies and assembly mechanisms. *Molecular Microbiology*, 93, 587–598.

**Candela, T. and Fouet, A.** (2006) Poly-gamma-glutamate in bacteria. *Molecular Microbiology*, 60, 1091–1098.

**Cao, M., Wang, T., Ye, R.W. and Helmann, J.D.** (2002) Antibiotics that inhibit cell wall biosynthesis induce expression of the *Bacillus subtilis* Sigma W and Sigma M regulons. *Molecular Microbiology*, 45, 1267–1276.

**Carballido-Lopez, R., Formstone, A., Li, Y., Ehrlich, D.S., Noirot, P. and Errington, J.** (2006) Actin homolog MreBH governs cell morphogenesis by localization of the cell wall hydrolases LytE. *Developmental Cell*, 11, 399–409.

**Carballido-Lopez, R. and Formstone, A.** (2007) Shape determination in *Bacillus subtilis*. *Current Opinion in Microbiology*, 10, 611–616.

**Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., Golland, P. and Sabatini, D.M.** (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7, 1–11.

**Casino, P., Rubio, V. and Marina, A.** (2009) Structural Insight into Partner Specificity and Phosphoryl Transfer in Two-Component Signal Transduction. *Cell*, 139, 325–336.

**Chai, Y., Norman, T., Kolter, R. and Losick, R.** (2010) An epigenetic switch governing daughter cell separation in *Bacillus subtilis*. *Genes & Development*, 24, 754–765.

**Chauhan, N. and Calderone, R.** (2008) Two-Component Signal Transduction Proteins as Potential Drug Targets in Medically Important Fungi. *Infection and Immunity*, 76, 4795–4803.

**Chen, C.-L., Wu, S.-C., Tjia, W.M., Wang, C., Lohka, M.J. and Wong, S.-L.** (2008) Development of a LytE-based high-density surface display system in *Bacillus subtilis*. *Microbial Biotechnology*, 1, 177–190.

**Chen, I. and Dubnau, D.** (2004) DNA uptake during bacterial transformation. *Nature Reviews Microbiology*, 2, 241–249.

**Cheung, J. and Hendrickson, W.A.** (2010) Sensor domains of two-component regulatory systems. *Current Opinion in Microbiology*, 13, 116–123.

**Choudhary, D.K. and Johri, B.N.** (2009) Interactions of *Bacillus spp.* and plants - with special reference to induced systemic resistance (ISR). *Microbiological Research*, 164, 493–513.

**Chumsakul, O., Takahashi, H., Oshima, T., Hishimoto, T., Kanaya, S., Ogasawara, N. and Ishikawa, S.** (2011) Genome-wide binding profiles of the *Bacillus subtilis* transition state regulator AbrB and its homolog Abh reveals their interactive role in transcriptional regulation. *Nucleic Acids Research*, 39, 414–428.

**Claverys, J.-P., Prudhomme, M. and Martin, B.** (2006) Induction of Competence Regulons as a General Response to Stress in Gram-Positive Bacteria. *Annual Review of Microbiology*, 60, 451–475.

**Commichau, F.M., Pietack, N. and Stülke, J.** (2013) Essential genes in *Bacillus subtilis*: a re-evaluation after ten years. *Molecular BioSystems*, 9, 1068–1075.

**Couldwell, D.L. and Lewis, D.A.** (2015) *Mycoplasma genitalium* infection: current treatment options, therapeutic failure, and resistance-associated mutations. *Infection and Drug Resistance*, 8, 147–161.

**Cullen, P.A., Haake, D.A. and Adler, B.** (2004) Outer membrane proteins of pathogenic *spirochetes*. *FEMS Microbiol Rev*, 28, 291–318.

**Daley, G.M., Russell, D.B., Tabrizi, S.N. and McBride, J.** (2014) *Mycoplasma genitalium*: a review. *International journal of STD & AIDS*, 25, 475–487.

**Darmon, E., Noone, D., Masson, A., Bron, S., Kuipers, O.P., Devine, K.M. and Dijl, J.M.v.** (2002) A Novel Class of Heat and Secretion Stress-Responsive Genes Is Controlled by the Autoregulated CssRS Two-Component System of *Bacillus subtilis*. *Journal of Bacteriology*, 184, 5661–5671.

**den Blaauwen, T., de Pedro, M.A., Nguyen-Disteche, M. and Ayala, J.A.** (2008) Morphogenesis of rod-shaped sacculi. *FEMS Microbiology Reviews*, 32, 321–344.

**Devine, K.M.** (2012) Bacterial L-forms on tap: an improved methodology to generate *Bacillus subtilis* Lforms heralds a new era of research. *Molecular Microbiology*, 83, 10–13.

**Djian, P.** (1998) Evolution of Simple Repeats in DNA and Their Relation to Human Disease. *Cell*, 94, 155–160.

**Dominguez-Cuevas, P., Porcelli, I., Daniel, R. and Errington, J.** (2013) Differentiated roles for MreB-actin isologues and autolytic enzymes in bacterial morphogenesis. *Molecular Microbiology*, 89, 1084–1098.

**Dörr, T., Cava, F., Lam, H., Davis, B.M. and Waldor, M.K.** (2013) Substrate specificity of an elongation-specific peptidoglycan endopeptidase and its implications for cell wall architecture and growth of *Vibrio cholerae*. *Molecular Microbiology*, 89, 949–962.

**Dowds, B.C.A., O`Kane, C., Gormley, E., McConnel, 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, Fla: CRC Press, pp. 138–154.

**Dramsi, S., Magnet, S., Davison, S. and Arthur, M.** (2008) Covalent attachment of proteins to peptidoglycan. *FEMS Microbiology Reviews*, 32, 307–320.

**Dubnau, D. and Losick, R.** (2006) Bistability in bacteria. *Molecular Microbiology*, 61, 564–572.

**Dubnau, D.** (2010) Swim or chill: lifestyles of a bacillus. *Genes & Development*, 24, 735–737.

**Dubrac, S., Bisicchia, P., Devine, K.M. and Msdaek, T.** (2008) A matter of life and death: cell wall homeostasis and the WalKR (YycGF) essential signal transduction pathway. *Molecular Microbiology*, 70, 1307–1322.

**Dubrac, S. and Msadek, T.** (2008) Tearing Down the Wall: Peptidoglycan Metabolism and the WalK/WalR (YycG/YycF) Essential Two-Component System. *Advances in Experimental Medicine and Biology*, 631, 214–228.

**Earl, A.M., Losick, R. and Kolter, R.** (2008) Ecology and genomics of *Bacillus subtilis*. *Trends in Microbiology*, 16, 269–275.

**Eckert, C., Lecerf, M., Dubost, L., Arthur, M. and Mesnage, S.** (2006) Functional analysis of AtlA, the major N-acetylglucosaminidase of *Enterococcus faecalis*. *Journal of Bacteriology*, 188, 8513–8519.

**Egan, A.J.F. and Vollmer, W.** (2013) The physiology of bacterial cell division. *Annals of the New York Academy of Sciences*, 1277, 8–28.

**Eijlander, R.T., Jong, A. de, Krawczyk, A.O., Holsappel, S. and Kuipers, O.P.** (2013) SporeWeb: an interactive journey through the complete sporulation cycle of *Bacillus subtilis*. *Nucleic Acids Research*, 42, 1–7.

**Ekman, D., Bjorklund, A.K., Frey-Skott, J. and Elofsson, A.** (2005) Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. *Journal of Molecular Biology*, 348, 231–243.

**Elliott, K.T., Cuff, L.E. and Neidle, E.L.** (2013) Copy number change: evolving views on gene amplification. *Future Microbiology*, 8, 887–899.

**Elwell, C., Mirrashidi, K. and Engel, J.** (2016) *Chlamydia* cell biology and pathogenesis. *Nature Reviews. Microbiology*, 14, 385–400.

**Ercan, D. and Demirci, A.** (2015) Current and future trends for biofilm reactors for fermentation processes. *Critical Reviews in Biotechnology*, 35, 1–14.

**Fabret, C., Feher, V.A. and Hoch, J.A.** (1999) Two-Component Signal Transduction in *Bacillus subtilis*: How One Organism Sees Its World. *Journal of Bacteriology*, 181, 1975–1983.

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

**Fenton, M., Ross, P., McAuliffe, O., O'Mahony, J. and Coffey, A.** (2010) Recombinant bacteriophage lysins as antibacterials. *Bioengineered Bugs*, 1, 9–16.

**Foster, S.J.** (1994) The role and regulation of cell wall structural dynamics during differentiation of endospore-forming bacteria. *Journal of Applied Bacteriology*, 76, 25–39.

**Frankel, M.B. and Schneewind, O.** (2012) Determinants of Murein Hydrolase Targeting to Cross-wall of *Staphylococcus aureus* Peptidoglycan. *Journal of Biological Chemistry*, 287, 10460–10471.

**Frirdich, E. and Gaynor, E.C.** (2013) Peptidoglycan hydrolases, bacterial shape, and pathogenesis. *Current Opinion in Microbiology*, 16, 767–778.

**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.

**Fukushima, T., Afkham, A., Kurosawa, S.-i., Tanabe, T., Yamamoto, H. and Sekiguchi, J.** (2006) A New D,L-Endopeptidase Gene Product, YojL (Renamed CwlS), Plays a Role in Cell Separation with LytE and LytF in *Bacillus subtilis*. *Journal of Bacteriology*, 188, 5541–5550.

**Fukushima, T., Furihata, I., Emmins, R., Daniel, R., Hoch, J.A. and Szurmant, H.** (2011) A role for the essential YycG sensor histidine kinase in sensing cell division. *Molecular Microbiology*, 79, 503–522.

**Galperin, M.Y.** (2005) A census of membrane-bound and intracellular signal transduction proteins in bacteria: Bacterial IQ, extroverts and introverts. *BMC Microbiology*, 5, 1–19.

**Galperin, M.Y.** (2010) Diversity of structure and function of response regulator output domains. *Current Opinion in Microbiology*, 13, 150–159.

**Gerstmans, H., Rodriguez-Rubio, L., Lavigne, R. and Briers, Y.** (2016) From endolysins to Artilysin(R)s: novel enzyme-based approaches to kill drug-resistant bacteria. *Biochemical Society Transactions*, 44, 123–128.

**Ghuysen, J.M. and Hakenbeck, R.** (1994) *Bacterial Cell Wall. Microbial peptidoglycan (murein) hydrolases:* Elsevier Science.

**Glaser, P., Kunst, F., Arnaud, M., Coudart, M.-P., Gonzales, W., Hullo, M.-F., Ionescu, M., Lubochinsicy, B., Marcelino, L., Moszer, I., Presecan, E., Santana, M., Schneider, E., Schwelzer, J., Vertes, A., Rapoport, G. and Danchin, A.** (1993) *Bacillus subtilis* genome project. Cloning and sequencing of the 97 kb region from 325° to 333°. *Molecular Microbiology*, 10, 371–384.

**Goodwell, E.W., Lopez, R. and Tomasz, A.** (1976) Suppression of lytic effect of beta lactams on *Escherichia coli* and other bacteria. *Proc. Natl. Acad. Sci.*, 73, 3293–3297.

**Gotoh, Y., Eguchi, Y., Watanabe, T., Okamoto, S., Doi, A. and Utsumi, R.** (2010) Two-component signal transduction as potential drug targets in pathogenic bacteria. *Current Opinion in Microbiology*, 13, 232–239.

**Goulian, M.** (2010) Two-component signaling circuit structure and properties. *Current Opinion in Microbiology*, 2, 184–189.

**Graumann, P.** (2007) *Bacillus cellular and molecular biology.* Norfolk: Caister Academic Press.

**Greeff, C., Roux, M., Mundy, J. and Petersen, M.** (2012) Receptor-like kinase complexes in plant innate immunity. *Frontiers in Plant Science*, 3, 1–7.

**Groban, E.S., Clarke, E.J., Salis, H.M., Miller, S.M. and Voigt, C.A.** (2009) Kinetic buffering of cross talk between bacterial two-component sensors. *Journal of Molecular Biology*, 390, 380–393.

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

**Gunka, K., Stannek, L., Care, R.A. and Commichau, F.M.** (2013) Selection-driven accumulation of suppressor mutants in *Bacillus subtilis*: the apparent high mutation frequency of the cryptic *gudB* gene and the rapid clonal expansion of *gudB*(+) suppressors are due to growth under selection. *PLoS ONE*, 8, 1–15.

**Gurung, N., Ray, S., Bose, S. and Rai, V.** (2013) A broader view: microbial enzymes and their relevance in industries, medicine, and beyond. *BioMed research international*, 2013, 1–18.

**Gust, A.A., Willmann, R., Desaki, Y., Grabherr, H.M. and Nürnberger, T.** (2012) Plant LysM proteins: modules mediating symbiosis and immunity. *Trends in Plant Science*, 17, 495–502.

**Guttenplan, S.B., Shaw, S. and Kearns, D.B.** (2013) The cell biology of peritrichous flagella in *Bacillus subtilis*. *Molecular Microbiology*, 87, 211–229.

**Hartig, E. and Jahn, D.** (2012) Regulation of the anaerobic metabolism in *Bacillus subtilis*. *Advances in Microbial Physiology*, 61, 195–216.

**Harwood, C.R.** (1992) *Bacillus subtilis* and its relatives: molecular biological and industrial workhorses. *Trends in Biotechnology*, 10, 247–256.

**Harwood, C.R. and Cranenburgh, R.** (2008) *Bacillus* protein secretion: an unfolding story. *Trends in Microbiology*, 16, 73–79.

**Hashimoto, M., Ooiwa, S. and Sekiguchi, J.** (2012) Synthetic Lethality of the *lytE cwlO* Genotype in *Bacillus subtilis* Is Caused by Lack of D,L-Endopeptidase Activity at the Lateral Cell Wall. *Journal of Bacteriology*, 194, 796–803.

**Hayhurst, E.J., Kailas, L., Hobbs, J.K. and Foster, S.J.** (2008) Cell wall peptidoglycan architecture in *Bacillus subtilis*. *Proc. Natl. Acad. Sci.*, 105, 14603–14608.

**Hecker, M., Pané-Farré, J. and Uwe, V.** (2007) SigB-Dependent General Stress Response in *Bacillus subtilis* and Related Gram-Positive Bacteria. *Annual Review of Microbiology*, 61, 215–236.

**Higgins, D. and Dworkin, J.** (2012) Recent progress in *Bacillus subtilis* sporulation. *FEMS Microbiology Reviews*, 36, 131–148.

**Hirose, I., Sano, K., Shioda, I., Kumano, M., Nakamura, K. and Yamane, K.** (2000) Proteome analysis of *Bacillus subtilis* extracellular proteins: a two-dimensional protein electrophoretic study. *Microbiology*, 146, 65–75.

**Hoch, J.A.** (2000) Two-component and phosphorelay signal transduction. *Current Opinion in Microbiology*, 3, 165–170.

**Hoch, J.A. and Varughese, K.** (2001) Keeping Signals Straight in Phosphorelay Signal Transduction. *Journal of Bacteriology*, 183, 4941–4949.

**Hoiczyk, E. and Hansel, A.** (2000) Cyanobacterial Cell Walls. News from an Unusual Prokaryotic Envelope. *Journal of Bacteriology*, 182, 1191–1199.

**Hong, H.A., Huang, J.-M., Khaneja, R., Hiep, L.V., Urdaci, M.C. and Cutting, S.M.** (2008) The safety of *Bacillus subtilis* and *Bacillus indicus* as food probiotics. *Journal of Applied Microbiology*, 105, 510–520.

**Hong, H.A., Khaneja, R., Tam, N.M.K., Cazzato, A., Tan, S., Urdaci, M., Brisson, A., Gasbarrini, A., Barnes, I. and Cutting, S.M.** (2009) *Bacillus subtilis* isolated from the human gastrointestinal tract. *Research in Microbiology*, 160, 134–143.

**Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. and Pease, L.R.** (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, 77, 61–68.

**Howell, A., Dubrac, S., Andersen, K.K., Noone, D., Fert, J., Msadek, T. and Devine, K.M.** (2003) Genes controlled by the essential YycG YycF twocomponent system of *Bacillus subtilis* revealed through a novel hybrid regulator approach. *Molecular Microbiology*, 49, 1639–1655.

**Howell, A., Dubrac, S., Noone, D., Varughese, K.I. and Devine, K.M.** (2006) Interactions between the YycFG and PhoPR two-component systems in *Bacillus subtilis*: the PhoR kinase phosphorylates the non-cognate YycF response regulator upon phosphate limitation. *Molecular Microbiology*, 59, 1199–1215.

**Hu, S., Kong, J., Kong, W., Guo, T. and Ji, M.** (2010) Characterization of a novel LysM domain from *Lactobacillus fermentum* bacteriophage endolysin and its use as

an anchor to display heterologous proteins on the surfaces of lactic acid bacteria. *Applied and Environmental Microbiology*, 76, 2410–2418.

**Huang, K.C., Mukhopadhyay, R., Wen, B., Gitai, Z. and Wingreen, N.S.** (2008) Cell shape and cell-wall organization in Gram-negative bacteria. *Proc. Natl. Acad. Sci.*, 105, 19282–19287.

**Huang, W.-Z., Wang, J.-J., Chen, H.-J., Chen, J.-T. and Shaw, G.-C.** (2013) The heat-inducible essential response regulator WalR positively regulates transcription of *sigI*, *mreBH* and *lytE* in *Bacillus subtilis* under heat stress. *Research in Microbiology*, 164, 998–1008.

**Huynh, T.N. and Stewart, V.** (2011) Negative control in two-component signal transduction by transmitter phosphatase activity. *Molecular Microbiology*, 82, 275–286.

**Ishikawa, S., Hara, Y., Ohnishi, R. and Sekiguchi, J.** (1998) Regulation of a New Cell Wall Hydrolase Gene, *cwlF*, which affects Cell Separation in *Bacillus subtilis*. *Journal of Bacteriology*, 1890, 2549–2555.

**Jakobs, M. and Meinhardt, F.** (2015) What renders *Bacilli* genetically competent? A gaze beyond the model organism. *Applied Microbiology and Biotechnology*, 99, 1557–1570.

**Jernigan, K.K. and Bordenstein, S.R.** (2015) Tandem-repeat protein domains across the tree of life. *PeerJ*, 3, 1–16.

**Jiang, M., Shao, W., Perego, M. and Hoch, J.A.** (2000) Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. *Molecular Microbiology*, 38, 535–542.

**Johnson, J.W., Fisher, J.F. and Mobashery, S.** (2013) Bacterial cell-wall recycling. *Annals of the New York Academy of Sciences*, 1277, 54–75.

**Jordan, S., Hutchings, M.I. and Mascher, T.** (2008) Cell envelope stress response n Gram-positive bacteria. *FEMS Microbiology Reviews*, 32, 107–146.

**Jung, K., Fried, L., Behr, S. and Heermann, R.** (2012) Histidine kinases and response regulators in networks. *Current Opinion in Microbiology*, 15, 118–124.

**Kajimura, J., Fujiwara, T., Yamada, S., Suzawa, Y., Nishida, T., Oyamada, Y., Hayashi, I., Yamagishi, J.-i., Komatsuzawa, H. and Sugai, M.** (2005) Identification and molecular characterization of an N-acetylmuramyl-L-alanine amidase Sle1 involved in cell separation of *Staphylococcus aureus*. *Molecular Microbiology*, 58, 1087–1101.

**Kanamaru, K., Aiba, H., Mizushimas, S. and Mizuno, T.** (1989) Signal Transduction and Osmoregulation in *Escherichia coli*. *Journal of Biological Chemistry*, 264, 21633–21637.

**Kearns, D.B., Chu, F., Branda, S.S., Kolter, R. and Losick, R.** (2005) A master regulator for biofilm formation by *Bacillus subtilis*. *Molecular Microbiology*, 55, 739–749.

**Kearns, D.B. and Losick, R.** (2003) Swarming motility in undomesticated *Bacillus subtilis*. *Molecular Microbiology*, 49, 581–590.

**Keener, J. and Kustu, S.** (1988) Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins NTRB and NTRC of enteric bacteria: roles of the conserved amino-terminal domain of NTRC. *Proc. Natl. Acad. Sci.*, 85, 4976–4980.

**Keep, N.H., Ward, J.M., Cohen-Gonsaud, M. and Henderson, B.** (2006) Wake up! Peptidoglycan lysis and bacterial non-growth states. *Trends in Microbiology*, 14, 271–276.

**Kesel, S., Mader, A., Hofler, C., Mascher, T. and Leisner, M.** (2013) Immediate and heterogeneous response of the LiaFSR two-component system of *Bacillus subtilis* to the peptide antibiotic bacitracin. *PLoS ONE*, 8, 1–12.

**Kiil, K., Ferchaud, J.B., David, C., Binnewies, T.T., Wu, H., Sicheritz-Pontén, T., Willenbrock, H. and Ussery, D.W.** (2005) Genome update: distribution of two-component transduction systems in 250 bacterial genomes. *Microbiology*, 151, 3447-52.

**King, J. and Laemmli, U.K.** (1971) Polypeptides of the tail of the bacteriophage T4. *Journal of Molecular Biology*, 62, 465–477.

**Kiriyama, Y., Yazawa, K., Tanaka, T., Yoshikawa, R., Yamane, H., Hashimoto, M., Sekiguchi, J. and Yamamoto, H.** (2014) Localization and expression of the *Bacillus subtilis* DL-endopeptidase LytF are influenced by mutations in LTA synthases and glycolipid anchor synthetic enzymes. *Microbiology*, 160, 2639–2649.

**Kruijff, B. de, van Dam, V. and Breukink, E.** (2008) Lipid II: A central component in bacterial cell wall synthesis and a target for antibiotics. *From Lipidomics to Human Health*, 79, 117–121.

**Kummerfeld, S.K. and Teichmann, S.A.** (2005) Relative rates of gene fusion and fission in multi-domain proteins. *Trends in Genetics*, 21, 25–30.

**Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessières, 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. and Danchin, A.** (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature*, 390, 249–267.

**Laddomada, F., Miyachiro, M.M. and Dessen, A.** (2016) Structural Insights into Protein-Protein Interactions Involved in Bacterial Cell Wall Biogenesis. *Antibiotics*, 5, 1–13.

**Lamprecht, M.R., Sabatini, D.M. and Carpenter, A.E.** (2007) CellProfiler: free, versatile software for automated biological image analysis. *BioTechniques*, 42, 71–75.

**Laub, M.T. and Goulian, M.** (2007) Specificity in Two-Component Signal Transduction Pathways. *Annual Review of Genetics*, 41, 121–145.

**Lechat, P., Hummel, L., Rousseau, S. and Moszer, I.** (2007) GenoList: an integrated environment for comparative analysis of microbial genomes. *Nucleic Acids Research*, 36, 469–474.

**Leclerc, D. and Asselin, A.** (1989) Detection of bacterial cell wall hydrolases after denaturing polyacrylamide gel electrophoresis. *Canadian Journal of Microbiology*, 35, 749–753.

**Lee, S.Y., Cho, H.S., Pelton, J.G., Yan, D., Berry, E.A. and Wemmer, D.E.** (2001) Crystal structure of activated CheY. Comparison with other activated receiver domains. *The Journal of Biological Chemistry*, 276, 16425–16431.

**Lee, T.K. and Huang, K.C.** (2013) The role of hydrolases in bacterial cell-wall growth. *Current Opinion in Microbiology*, 16, 760–766.

**Livak, K.J. and Schmittgen, T.D.** (2001) Analysis of Relative Gen Expression Data Using Real-Time Quantitative PCR and the 2- CT Method. *Methods*, 25, 402–408.

**Lois, A.F., Weinstein, M., Ditta, G.S. and Helinskii, D.R.** (1993) Autophosphorylation and Phosphatase Activities of the Oxygensensing Protein FixL of *Rhizobium meliloti* Are Coordinately Regulated by Oxygen. *Journal of Biological Chemistry*, 268, 4370–4375.

**Looman, C., Abrink, M., Mark, C. and Hellman, L.** (2002) KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. *Molecular Biology and Evolution*, 19, 510–521.

**Lopez, D., Vlamakis, H., Losick, R. and Kolter, R.** (2009) Cannibalism enhances biofilm development in *Bacillus subtilis*. *Molecular Microbiology*, 74, 609–618.

**Lukat, G.S., McCleary, W.R., Stock, A.M. and Stock, J.B.** (1992) Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors. *Proc. Natl. Acad. Sci.*, 89, 718–722.

**Luo, Y. and Helmann, J.D.** (2012) Analysis of the role of *Bacillus subtilis* σM in β-lactam resistance reveals an essential role for c-di-AMP in peptidoglycan homeostasis. *Molecular Microbiology*, 83, 623–639.

**MacFarlane, S.A. and Merrick, M.** (1985) The nucleotide sequence of the nitrogen regulation gene *ntrB* and the *glnA-ntrBC* intergenic region of *Klebsiella pneumoniae*. *Nucleic Acids Research*, 13, 7591–7606.

**Macheboeuf, P., Contreras-Martel, C., Job, V., Dideberg, O. and Dessen, A.** (2006) Penicillin binding proteins: key players in bacterial cell cycle and drug resistance processes. *FEMS Microbiology Reviews*, 30, 673–691.

**Margot, P., Pagni, M. and Karamata, D.** (1999) *Bacillus subtilis 168* gene *lytF* encodes a 7-Dglutamate-meso-diaminopimelate muropeptidase expressed by the alternative vegetative sigma factor, SigD. *Microbiology*, 145, 57–65.

**Margot, P., Wahlen, M., Gholamhuseinian, A., Piggot, P. and Karamata, D.** (1998) The *lytE* Gene of *Bacillus subtilis 168* Encodes a Cell Wall Hydrolase. *Journal of Bacteriology*, 180, 749–752.

**Marina, A., Mott, C., Auyzenberg, A., Hendrickson, W.A. and Waldburger, C.D.** (2001) Structural and mutational analysis of the PhoQ histidine kinase catalytic domain. Insight into the reaction mechanism. *The Journal of Biological Chemistry*, 276, 41182–41190.

**Martin, P.K., Li, T., Sun, D., Biek, D.P. and Schmid, M.B.** (1999) Role in Cell Permeability of an Essential Two-Component System in *Staphylococcus aureus*. *Journal of Bacteriology*, 181, 3666–3673.

**Mascher, T., Helmann, J.D. and Unden, G.** (2006) Stimulus Perception in Bacterial Signal-Transducing Histidine Kinases. *Microbiology and Molecular Biology*, 70, 910–938.

**McCleary, W.R., Stock, J.B. and Ninfa, A.J.** (1993) Is acetyl phosphate a global signal in *Escherichia coli*? *Journal of Bacteriology*, 175, 2793–2798.

**Meisner, J., Llopis, P.M., Sham, L.-T., Garner, E., Bernhardt, T.G. and Rudner, D.Z.** (2013) FtsEX is required for CwlO peptidoglycan hydrolase activity during cell wall elongation in *Bacillus subtilis*. *Molecular Microbiology*, 89, 1069–1083.

**Mesnage, S., Dellarole, M., Baxter, N.J., Rouget, J.-B., Dimitrov, J.D., Wang, N., Fujimoto, Y., Hounslow, A.M., Lacroix-Desmazes, S., Fukase, K., Foster, S.J. and Williamson, M.P.** (2014) Molecular basis for bacterial peptidoglycan recognition by LysM domains. *Nature Communications*, 5, 1–11.

**Michna, R.H., Commichau, F.M., Todter, D., Zschiedrich, C.P. and Stulke, J.** (2013) SubtiWiki-a database for the model organism *Bacillus subtilis* that links pathway, interaction and expression information. *Nucleic Acids Research*, 42, 1–7.

**Mielich-Suss, B. and Lopez, D.** (2015) Molecular mechanisms involved in *Bacillus subtilis* biofilm formation. *Environmental Microbiology*, 17, 555–565.

**Miller, D.A., Suen, G., Clements, K.D. and Angert, E.R.** (2012) The genomic basis for the evolution of a novel form of cellular reproduction in the bacterium Epulopiscium. *BMC Genomics*, 13, 1–18.

**Minzuno, T.** (1998) His-Asp Phosphotransfer Signal Transduction. *Journal of Biochemistry*, 123, 555–563.

**Miyashiro, T. and Goulian, M.** (2008) High stimulus unmasks positive feedback in an autoregulated bacterial signaling circuit. *Proc. Natl. Acad. Sci.*, 105, 17457–17462.

**Mizuno, M., Masuda, S., Takemaru, K., Hosono, S., Sato, T., Takeuchi, M. and Kobayashi, Y.** (1996) Systematic sequencing of the 283 kb 210 degrees-232 degrees region of the *Bacillus subtilis* genome containing the skin element and many sporulation genes. *Microbiology*, 142, 3103–3111.

**Molle, V., Fujita, M., Jensen, S.T., Eichenberger, P., González-Pastor, J.E., Liu, J.S. and Losick, R.** (2003) The Spo0A regulon of *Bacillus subtilis*. *Molecular Microbiology*, 50, 1683–1701.

**Molle, V., Nakaura, Y., Shivers, R.P., Yamaguchi, H., Losick, R., Fujita, Y. and Sonenshein, A.L.** (2003) Additional targets of the *Bacillus subtilis* global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. *Journal of Bacteriology*, 185, 1911–1922.

**Moszer, I.** (1998) The complete genome of *Bacillus subtilis*: from sequence annotation to data management and analysis. *FEBS Letters*, 430, 28–36.

**Moszer, I., Jones, L.M., Moreira, S., Fabry, C. and Danchin, A.** (2002) SubtiList: the reference database for the *Bacillus subtilis* genome. *Nucleic Acids Research*, 30, 62–65.

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

**Müller, J.P., An, Z., Merad, T., Hancock, I.C. and Harwood, C.R.** (1997) Influence of *Bacillus subtilis phoR* on cell wall anionic polymers. *Microbiology*, 143, 947–956.

**Navarre, W.W. and Schneewind, O.** (1999) Surface Proteins of Gram-Positive Bacteria and Mechanisms of Their Targeting to the Cell Wall Envelope. *Microbiology and Molecular Biology Reviews*, 63, 174–229.

**Nicolas, P., Mader, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M., Aymerich, S., Becher, D., Bisicchia, P., Botella, E., Delumeau, O., Doherty, G., Denham, E.L., Fogg, M.J., Fromion, V., Goelzer, A., Hansen, A., Hartig, E., Harwood, C.R., Homuth, G., Jarmer, H., Jules, M., Klipp, E., Le Chat, L., Lecointe, F., Lewis, P., Liebermeister, W., March, A., Mars, R.A.T., Nannapaneni, P., Noone, D., Pohl, S., Rinn, B., Rugheimer, F., Sappa, P.K., Samson, F., Schaffer, M., Schwikowski, B., Steil, L., Stulke, J., Wiegert, T., Devine, K.M., Wilkinson, A.J., van Dijl, J.M., Hecker, M., Volker, U., Bessieres, P. and Noirot, P.** (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science*, 335, 1103–1106.

**Nijman, S.M.B.** (2011) Synthetic lethality: general principles, utility and detection using genetic screens in human cells. *FEBS Letters*, 585, 1–6.

**Ninfa, A.J., Ninfa, E.G., Lupas, A.N., Stock, A., Magasanik, B. and Stock, J.** (1988) Crosstalk between bacterial chemotaxis signal transduction proteins and regulators of transcription of the Ntr regulon: evidence that nitrogen assimilation and chemotaxis are controlled by a common phosphotransfer mechanism. *Proc. Natl. Acad. Sci.*, 85, 5492–5496.

**Noone, D., Botella, E., butler, C., Hansen, A., Jende, I. and Devine, K.M.** (2012) Signal Perception by the Secretion Stress-Responsive CssRS Two-Component System in *Bacillus subtilis*. *Journal of Bacteriology*, 194, 1800–1814.

**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. *Journal of Bacteriology*, 182, 1592–1599.

**Noone, D., Salzberg, L.I., Botella, E., Basell, K., Becher, D., Antelmann, H. and Devine, K.M.** (2014) A Highly Unstable Transcript Makes CwlO D,L-Endopeptidase Expression Responsive to Growth Conditions in *Bacillus subtilis*. *Journal of Bacteriology*, 196, 237–247.

**Ohki, R., Giyanto, A., Tateno, K., Masuyama, W., Moriya, S., Kobayashi, K. and Ogasawara, N.** (2003) The BceRS two-component regulatory system induces expression of the bacitracin transporter, BceAB, in *Bacillus subtilis*. *Molecular Microbiology*, 49, 1135–1144.

**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*. *Journal of Bacteriology*, 181, 3178–3184.

**Ohnuma, T., Onaga, S., Murata, K., Taira, T. and Katoh, E.** (2008) LysM domains from Pteris ryukyuensis chitinase-A: a stability study and characterization of the chitin-binding site. *The Journal of Biological Chemistry*, 283, 5178–5187.

**Okano, K., Zhang, Q., Kimura, S., Narita, J., Tanaka, T., Fukuda, H. and Kondo, A.** (2008) System using tandem repeats of the cA peptidoglycan-binding domain from *Lactococcus lactis* for display of both N- and C-terminal fusions on cell surfaces of lactic acid bacteria. *Applied and Environmental Microbiology*, 74, 1117–1123.

**Oldroyd, G.E.D.** (2013) Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews. Microbiology*, 11, 252–263.

**Ortet, P., Whitworth, D.E., Santaella, C., Achouak, W. and Barakat, M.** (2015) P2CS: updates of the prokaryotic two-component systems database. *Nucleic Acids Research*, 43, 536–541.

**Park, J.T. and Uehara, T.** (2008) How Bacteria Consume Their Own Exoskeletons (Turnover and Recycling of Cell Wall Peptidoglycan). *Microbiology and Molecular Biology Reviews*, 72, 211–227.

**Parkinson, J.S. and Kofoid, E.C.** (1992) Communication Modules in Bacterial Signaling Proteins. *Annual Review of Genetics*, 26, 71–112.

**Pazy, Y., Wollish, A.C., Thomas, S.A., Miller, P.J., Collins, E.J., Bourret, R.B. and Silversmith, R.E.** (2009) Matching Biochemical Reaction Kinetics to the Timescales of Life: Structural Determinants That Influence the Autodephosphorylation Rate of Response Regulator Proteins. *Journal of Molecular Biology*, 392, 1205–1220.

**Pele, J., Becu, J.-M., Abdi, H. and Chabbert, M.** (2012) Bios2mds: an R package for comparing orthologous protein families by metric multidimensional scaling. *BMC Bioinformatics*, 13, 1–7.

**Percy, M.G. and Grundling, A.** (2014) Lipoteichoic acid synthesis and function in gram-positive bacteria. *Annual Review of Microbiology*, 68, 81–100.

**Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E.** (2004) UCSF Chimera. A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25, 1605–1612.

**Pinho, M.G., Kjos, M. and Veening, J.-W.** (2013) How to get (a)round: mechanisms controlling growth and division of coccoid bacteria. *Nature Reviews. Microbiology*, 11, 601–614.

**Podgornaia, A.I. and Laub, M.T.** (2013) Determinants of specificity in two-component signal transduction. *Current Opinion in Microbiology*, 16, 156–162.

**Ponting, C.P., Aravind, L., Schultz, J., Bork, P. and Koonin, E.V.** (1999) Eukaryotic Signalling Domain Homologues in Archaea and Bacteria. Ancient Ancestry and Horizontal Gene Transfer. *Journal of Molecular Biology*, 289, 729–745.

**R Core Team** (2015) R: A language and environment for statistical computing, https://www.R-project.org/.

**Reams, A.B. and Roth, J.R.** (2015) Mechanisms of gene duplication and amplification. *Cold Spring Harbor Perspectives in Biology*, 7, 1–26.

**Reith, J. and Mayer, C.** (2011) Peptidoglycan turnover and recycling in Gram-positive bacteria. *Applied Microbiology and Biotechnology*, 92, 1–11.

**Reshef, D., Itzhaki, Z. and Schueler-Furman, O.** (2010) Increased sequence conservation of domain repeats in prokaryotic proteins. *Trends in Genetics*, 26, 383–387.

**Roilides, E., Simitsopoulou, M., Katragkou, A. and Walsh, T.J.** (2015) How Biofilms Evade Host Defenses. *Microbiology Spectrum*, 3, 1–10.

**Rowland, S.L., Burkholder, W.F., Cunningham, K.A., Maciejewski, M.W., Grossman, A.D. and King, G.F.** (2004) Structure and mechanism of action of Sda, an inhibitor of the histidine kinases that regulate initiation of sporulation in *Bacillus subtilis*. *Molecular Cell*, 13, 689–701.

**Russo, F.D. and Silhavy, T.J.** (1993) The essential tension: opposed reactions in bacterial two-component regulatory systems. *Trends in Microbiology*, 1, 306–310.

**Saita, E., Abriata, L.A., Tsai, Y.T., Trajtenberg, F., Lemmin, T., Buschiazzo, A., Dal Peraro, M., Mendoza, D. de and Albanesi, D.** (2015) A coiled coil switch mediates cold sensing by the thermosensory protein DesK. *Molecular Microbiology*, 98, 258–271.

**Salzberg, L.I., Botella, E., Hokamp, K., Antelmann, H., Maaß, S., Becher, D., Noone, D., Devine, K.M. and Silhavy, T.J.** (2015) Genome-Wide Analysis of Phosphorylated PhoP Binding to Chromosomal DNA Reveals Several Novel Features of the PhoPR-Mediated Phosphate Limitation Response in *Bacillus subtilis*. *Journal of Bacteriology*, 197, 1492–1506.

**Salzberg, L.I. and Helmann, J.D.** (2007) An antibiotic-inducible cell wall associated protein that protects *Bacillus subtilis* from autolysis. *Journal of Bacteriology*, 189, 4671–4680.

**Salzberg, L.I., Powell, L., Hokamp, K., Botella, E., Noone, D. and Devine, K.M.** (2013) The WalRK (YycFG) and σ I RsgI regulators cooperate to control CwlO and LytE expression in exponentially growing and stressed *Bacillus subtilis* cells. *Molecular Microbiology*, 87, 180–195.

**Sambrook, J., Fritsch, E.F. and Maniatis, T.** (1989) *Molecular cloning. a laboratory manual.* 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

**Santelli, E., Liddington, R.C., Mohan, M.A., Hoch, J.A. and Szurmant, H.** (2007) The Crystal Structure of *Bacillus subtilis* YycI Reveals a Common-fold for Two Members of an Unusual Class of Sensor Histidine Kinase Regulatory Proteins. *Journal of Bacteriology*, 189, 3290–3295.

**Satomura, T., Shimura, D., Asai, K., Sadaie, Y., Hirooka, K. and Fujita, Y.** (2005) Enhancement of glutamine utilization in *Bacillus subtilis* through the GlnK-GlnL two-component regulatory system. *Journal of Bacteriology*, 187, 4813–4821.

**Sauvage, E., Kerff, F., Terrak, M., Ayala, J.A. and Charlier, P.** (2008) The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS*, 32, 234–258.

**Sayers, E.W., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L.Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D.J., Madden, T.L., Maglott, D.R., Miller, V., Mizrachi, I., Ostell, J., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusova, T.A., Wagner, L., Yaschenko, E. and Ye, J.** (2009)

Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 37, 5–15.

**Schallmey, M., Singh, A. and Ward, O.P.** (2004) Developments in the use of *Bacillus* species for industrial production. *Canadian Journal of Microbiology*, 50, 1–17.

**Scheffers, D.-J. and Pinho, M.G.** (2005) Bacterial Cell Wall Synthesis: New Insights from Localization Studies. *Microbiology and Molecular Biology Reviews*, 69, 585–607.

**Scheffers, D.-J. and Tol, M.B.** (2015) LipidII: Just Another Brick in the Wall? *PLoS Pathogens*, 11, 1–12.

**Schneewind, O. and Missiakas, D.M.** (2012) Protein secretion and surface display in Gram-positive bacteria. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367, 1123–1139.

**Schneider, C.A., Rasband, W.S. and Eliceiri, K.W.** (2012) NIH Image to ImageJ. 25 years of image analysis. *Nature Methods*, 9, 671–675.

**Schneider, T. and Sahl, H.-G.** (2010) An oldie, but a goodie - cell wall biosynthesis as antibiotic target pathway. *International Journal of Medical Microbiology*, 300, 161–169.

**Schreiber, M., Res, I. and Matter, A.** (2009) Protein kinases as antibacterial targets. *Current Opinion in Cell Biology*, 21, 325–330.

**Scoffone, V., Dondi, D., Biino, G., Borghese, G., Pasini, D., Galizzi, A. and Calvio, C.** (2013) Knockout of *pgdS* and *ggt* genes improves gamma-PGA yield in *B. subtilis*. *Biotechnology and Bioengineering*, 110, 2006–2012.

**Serizawa, M., Kodama, K., Yamamoto, H., Kobayashi, K., Ogasawara, N. and Sekiguchi, J.** (2014) Functional Analysis of the YvrGHb Two-Component System of *Bacillus subtilis*. Identification of the Regulated Genes by DNA Microarray and Northern Blot Analyses. *Bioscience, Biotechnology and Biochemestry*, 69, 2155–2169.

**Serizawa, M., Yamamoto, H., Yamaguchi, H., Fujita, Y., Kobayashi, K., Ogasawara, N. and Sekiguchi, J.** (2004) Systematic analysis of SigD-regulated genes in *Bacillus subtilis* by DNA microarray and Northern analysis analyses. *Gene*, 329, 125–136.

**Setlow, P.** (2006) Spores of *Bacillus subtilis*: their resistance to and killing by radiation, heat and chemicals. *Journal of Applied Microbiology*, 101, 514–525.

**Sévin, D.C., Stählin, J.N., Pollak, G.R., Kuehne, A., Sauer, U. and Shi, H.** (2016) Global Metabolic Responses to Salt Stress in Fifteen Species. *PLoS ONE*, 11, 1–21.

**Shah, I.M., Laaberki, M.-H., Popham, D.L. and Dworkin, J.** (2008) A Eukaryotic-like Ser/Thr Kinase Signals Bacteria to Exit Dormancy in Response to Peptidoglycan Fragments. *Cell*, 135, 486–496.

**Shao, X., Jiang, M., Yu, Z., Cai, H. and Li, L.** (2009) Surface display of heterologous proteins in *Bacillus thuringiensis* using a peptidoglycan hydrolase anchor. *Microbial Cell Factories*, 8, 1–17.

**Shao, X., Ni, H., Lu, T., Jiang, M., Li, H., Huang, X. and Li, L.** (2012) An improved system for the surface immobilisation of proteins on *Bacillus thuringiensis* vegetative cells and spores through a new spore cortex-lytic enzyme anchor. *New Biotechnology*, 29, 302–310.

**Sierro, N., Makita, Y., Hoon, M. de and Nakai, K.** (2008) DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. *Nucleic Acids Research*, 36, 93–96.

**Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D. and Higgins, D.G.** (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7, 1–6.

**Silhavy, T.J., Kahne, D. and Walker, S.** (2010) The Bacterial Cell Envelope. *Cold Spring Harbor Perspectives in Biology*, 2, 1–17.

**Simsek, O., Sabanoglu, S., Con, A.H., Karasu, N., Akcelik, M. and Saris, P.E.J.** (2013) Immobilization of nisin producer *Lactococcus lactis* strains to chitin with surface-displayed chitin-binding domain. *Applied Microbiology and Biotechnology*, 97, 4577–4587.

**Singh, S.K., SaiSree, L., Amrutha, R.N. and Reddy, M.** (2012) Three redundant murein endopeptidases catalyse an essential cleavage step in peptidoglycan synthesis of *Escherichia coli* K12. *Molecular Microbiology*, 86, 1036–1051.

**Siryaporn, A. and Goulian, M.** (2008) Cross-talk suppression between the CpxA-CpxR and EnvZ-OmpR two-component systems in *E. coli*. *Molecular Microbiology*, 70, 494–506.

**Skerker, J.M., Perchuk, B.S., Siryaporn, A., Lubin, E.A., Ashenberg, O., Goulian, M. and Laub, M.T.** (2008) Rewiring the Specificity of Two-Component Signal Transduction Systems. *Cell*, 133, 1043–1054.

**Skerker, J.M., Prasol, M.S., Perchuk, B.S., Biondi, E.G. and Laub, M.T.** (2005) Two-Component Signal Transduction Pathways Regulating Growth and Cell Cycle Progression in a Bacterium: A System-Level Analysis. *PLoS Biology*, 3, 1770–1788.

**Smith, T.J., Blackman, S.A. and Foster, S.J.** (2000) Autolysins of *Bacillus subtilis*: multiple enzymes with multiple functions. *Microbiology*, 146, 249–262.

**Snel, B., Bork, P. and Huynen, M.** (2000) Genome evolution. Gene fusion versus gene fission. *Trends in Genetics*, 16, 9–11.

**Sonenshein, A.L., Hoch, J.A. and Losick, R.** (1993) *Bacillus subtilis and other gram-positive bacteria. Biochemistry, physiology, and molecular genetics.* Washington, D.C: American Society for Microbiology.

**Song, Y., Peisach, D., Pioszak, A.A., Xu, Z. and Ninfa, A.J.** (2004) Crystal structure of the C-terminal domain of the two-component system transmitter protein nitrogen regulator II (NRII; NtrB), regulator of nitrogen assimilation in *Escherichia coli*. *Biochemistry*, 43, 6670–6678.

**Spaink, H.P.** (2004) Specific recognition of bacteria by plant LysM domain receptor kinases. *Trends in Microbiology*, 12, 201–204.

**Staron, A., Finkeisen, D.E. and Mascher, T.** (2011) Peptide antibiotic sensing and detoxification modules of *Bacillus subtilis*. *Antimicrobial Agents and Chemotherapy*, 55, 515–525.

**Staron, A., Sofia, H.J., Dietrich, S., Ulrich, L.E., Liesegang, H. and Mascher, T.** (2009) The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) sigma factor protein family. *Molecular Microbiology*, 74, 557–581.

**Steen, A., Buist, G., Horsburgh, G.J., Venema, G., Kuipers, O.P., Foster, S.J. and Kok, J.** (2005) AcmA of *Lactococcus lactis* is an N-acetylglucosaminidase with an optimal number of LysM domains for proper functioning. *The FEBS Journal*, 272, 2854–2868.

**Steen, A., Buist, G., Leenhouts, K.J., El Khattabi, M., Grijpstra, F., Zomer, A.L., Venema, G., Kuipers, O.P. and Kok, J.** (2003) Cell Wall Attachment of a Widely Distributed Peptidoglycan Binding Domain Is Hindered by Cell Wall Constituents. *Journal of Biological Chemistry*, 278, 23874–23881.

**Steen, A., Palumbo, E., Deghorain, M., Cocconcelli, P.S., Delcour, J., Kuipers, O.P., Kok, J., Buist, G. and Hols, P.** (2005) Autolysis of *Lactococcus lactis* Is

Increased upon d-Alanine Depletion of Peptidoglycan and Lipoteichoic Acids. *Journal of Bacteriology*, 187, 114–124.

**Stephenson, K.** (2000) The Mechanism of Action of Inhibitors of Bacterial Two-component Signal Transduction Systems. *Journal of Biological Chemistry*, 275, 38900–38904.

**Stephenson, K. and Hoch, J.A.** (2002) Evolution of signalling in the sporulation phosphorelay. *Molecular Microbiology*, 46, 297–304.

**Stephenson, K. and Hoch, J.A.** (2004) Developing inhibitors to selectively target two-component and phosphorelay signal transduction systems of pathogenic microorganisms. *Current Medicinal Chemistry*, 11, 765–773.

**Stock, A.M., Chen, T., Welsh, D. and Stock, J.** (1988) CheA protein, a central regulator of bacterial chemotaxis, belongs to a family of proteins that control gene expression in response to changing environmental conditions. *Proc. Natl. Acad. Sci.*, 85, 1403–1407.

**Stock, A.M., Robinson, V.L. and Goudreau, P.N.** (2000) Two-Component Signal Transduction. *Annual Review of Biochemistry*, 69, 183–215.

**Studier, W.F. and Moffatt, B.A.** (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, 189, 113–130.

**Suzuki, T. and Tahara, Y.** (2003) Characterization of the *Bacillus subtilis ywtD* gene, whose product is involved in gamma-polyglutamic acid degradation. *Journal of Bacteriology*, 185, 2379–2382.

**Swanson, R.V., Alex, L.A. and Simon, M.I.** (1994) Histidine and aspartate phosphorylation: two-component systems and the limits of homology. *Trends in Biochemical Sciences*, 19, 485–490.

**Szurmant, H., Fukushima, T. and Hoch, J.A.** (2007) The Essential YycFG Two-Component System of *Bacillus subtilis*. In *Two-Component Signaling Systems, Part A:* Elsevier, pp. 396–417.

**Szurmant, H., Mohan, M.A., Imus, P.M. and Hoch, J.A.** (2007) YycH and YycI Interact To Regulate the Essential YycFG Two-Component System in *Bacillus subtilis*. *Journal of Bacteriology*, 189, 3280–3289.

**Szurmant, H., Nelson, K., Kim, E.-J., Perego, M. and Hoch, J.A.** (2005) YycH regulates the activity of the essential YycFG two-component system in *Bacillus subtilis*. *Journal of Bacteriology*, 187, 5419–5426.

**Szweda, P., Schielmann, M., Kotlowski, R., Gorczyca, G., Zalewska, M. and Milewski, S.** (2012) Peptidoglycan hydrolases-potential weapons against *Staphylococcus aureus*. *Applied Microbiology and Biotechnology*, 96, 1157–1174.

**Tan, I.S. and Ramamurthi, K.S.** (2014) Spore formation in *Bacillus subtilis*. *Environmental Microbiology Reports*, 6, 212–225.

**Tanaka, K., Henry, C.S., Zinner, J.F., Jolivet, E., Cohoon, M.P., Xia, F., Bidnenko, V., Ehrlich, S.D., Stevens, R.L. and Noirot, P.** (2013) Building the repertoire of dispensable chromosome regions in *Bacillus subtilis* entails major refinement of cognate large-scale metabolic model. *Nucleic Acids Research*, 41, 687–699.

**Tanaka, K., Kobayashi, K. and Ogasawara, N.** (2003) The *Bacillus subtilis* YufLM two-component system regulates the expression of the malate transporters MaeN (YufR) and YflS, and is essential for utilization of malate in minimal medium. *Microbiology*, 149, 2317–2329.

**Thomas, S.A., Brewster, J.A. and Bourret, R.B.** (2008) Two variable active site residues modulate response regulator phosphoryl group stability. *Molecular Microbiology*, 69, 453–465.

**Thomson, J.A., ed.** (1988) *Recombinant DNA and bacterial fermentation.* Boca Raton, Fla: CRC Press.

**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. *Microbiology and Molecular Biology Reviews*, 64, 515–547.

**Tseng, C.-L., Chen, J.-T., Lin, J.-H., Huang, W.-Z. and Shaw, G.-C.** (2011) Genetic evidence for involvement of the alternative sigma factor SigI in controlling expression of the cell wall hydrolase gene *lytE* and contribution of LytE to heat survival of *Bacillus subtilis*. *Archives of Microbiology*, 193, 677–685.

**Tseng, C.-L. and Shaw, G.-C.** (2008) Genetic Evidence for the Actin Homolog Gene *mreBH* and the Bacitracin Resistance Gene *bcrC* as Targets of the Alternative Sigma Factor SigI of *Bacillus subtilis*. *Journal of Bacteriology*, 190, 1561–1567.

**Tu, L., He, Y., Shan, C. and Wu, Z.** (2016) Preparation of Microencapsulated *Bacillus subtilis* SL-13 Seed Coating Agents and Their Effects on the Growth of Cotton Seedlings. *BioMed Research International*, 2016, 1–7.

**Tzeng, Y. and Hoch, J.A.** (1997) Molecular Recognition in Signal Transduction: The Interaction Surfaces of the Spo0F Response Regulator with its Cognate

Phosphorelay Proteins Revealed by Alanine Scanning Mutagenesis. *Journal of Molecular Biology*, 272, 200–212.

**Uehara, T. and Bernhardt, T.G.** (2011) More than just lysins: peptidoglycan hydrolases tailor the cell wall. *Current Opinion in Microbiology*, 14, 698–703.

**Ulrich, L.E. and Zhulin, I.B.** (2010) The MiST2 database: a comprehensive genomics resource on microbial signal transduction. *Nucleic Acids Research*, 38, 401–407.

**UniProt, C.** (2015) UniProt: a hub for protein information. *Nucleic Acids Research*, 43, 204–212.

**van Dijl, J.M. and Hecker, M.** (2013) *Bacillus subtilis*: from soil bacterium to super-secreting cell factory. *Microbial Cell Factories*, 12, 1–16.

**van Heijenoort, J.** (2007) Lipid intermediates in the biosynthesis of bacterial peptidoglycan. *Microbiology and Molecular Biology Reviews*, 71, 620–635.

**Veiga, P., Bulbarela-Sampieri, C., Furlan, S., Maisons, A., Chapot-Chartier, M.-P., Erkelenz, M., Mervelet, P., Noirot, P., Frees, D., Kuipers, O.P., Kok, J., Gruss, A., Buist, G. and Kulakauskas, S.** (2007) SpxB regulates O-acetylation-dependent resistance of *Lactococcus lactis* peptidoglycan to hydrolysis. *The Journal of Biological Chemistry*, 282, 19342–19354.

**Visweswaran, G.R.R., Dijkstra, B.W. and Kok, J.** (2011) Murein and pseudomurein cell wall binding domains of bacteria and archaea—a comparative view. *Applied Microbiology and Biotechnology*, 92, 921–928.

**Visweswaran, G.R.R., Leenhouts, K., van Roosmalen, M., Kok, J. and Buist, G.** (2014) Exploiting the peptidoglycan-binding motif, LysM, for medical and industrial applications. *Applied Microbiology and Biotechnology*, 98, 4331–4345.

**Vlamakis, H., Chai, Y., Beauregard, P., Losick, R. and Kolter, R.** (2013) Sticking together: building a biofilm the *Bacillus subtilis* way. *Nature Reviews Microbiology*, 11, 157–168.

**Voigt, B., Antelmann, H., Albrecht, D., Ehrenreich, A., Maurer, K.-H., Evers, S., Gottschalk, G., van Dijl, J.M., Schweder, T. and Hecker, M.** (2009) Cell Physiology and Protein Secretion of *Bacillus licheniformis* Compared to *Bacillus subtilis*. *Journal of Molecular Microbiology and Biotechnology*, 16, 53–68.

**Vollmer, W., Blanot, D. and de Pedro, M.A.** (2008) Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*, 32, 149–167.

**Vollmer, W., Joris, B., Charlier, P. and Foster, S.J.** (2008) Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiology Reviews*, 32, 256–286.

**Vollmer, W.** (2012) Bacterial growth does require peptidoglycan hydrolases. *Molecular Microbiology*, 86, 1031–1035.

**Wach, A.** (1996) PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in *S. cerevisiae*. *Yeast*, 12, 259–265.

**Wanner, B.L.** (1992) Is Cross Regulation by Phosphorylation of Two-Component Response Regulator Proteins Important in Bacteria? *Journal of Bacteriology*, 174, 2053–2058.

**Wassmann, M., Moeller, R., Rabbow, E., Panitz, C., Horneck, G., Reitz, G., Douki, T., Cadet, J., Stan-Lotter, H., Cockell, C.S. and Rettberg, P.** (2012) Survival of spores of the UV-resistant *Bacillus subtilis* strain MW01 after exposure to low-earth orbit and simulated martian conditions: data from the space experiment ADAPT on EXPOSE-E. *Astrobiology*, 12, 498–507.

**Wickham, H.** (2009) *Ggplot2. Elegant graphics for data analysis.* Dordrecht, New York: Springer.

**Wickham, H. and Francois, R.** (2015) dplyr: A Grammar of Data Manipulation, http://CRAN.R-project.org/package=dplyr.

**Wickham, H.** (2016) tidyr: Easily Tidy Data with `spread` and `gather` Functions, http://CRAN.R-project.org/package=tidyr.

**Wittekind, M. and Schuch, R.** (2016) Cell wall hydrolases and antibiotics: exploiting synergy to create efficacious new antimicrobial treatments. *Current Opinion in Microbiology*, 33, 18–24.

**Wolfe, A.J.** (2005) The acetate switch. *Microbiology and Molecular Biology Reviews*, 69, 12–50.

**Wong, J.E.M.M., Alsarraf, H.M.A.B., Kaspersen, J.D., Pedersen, J.S., Stougaard, J., Thirup, S. and Blaise, M.** (2014) Cooperative binding of LysM domains determines the carbohydrate affinity of a bacterial endopeptidase protein. *The FEBS Journal*, 281, 1196–1208.

**Wuichet, K., Cantwell, B.J. and Zhulin, I.B.** (2010) Evolution and phyletic distribution of two-component signal transduction systems. *Current Opinion in Microbiology*, 13, 219–225.

**Xu, Q., Abdubek, P., Astakhova, T., Axelrod, H.L., Bakolitsa, C., Cai, X., Carlton, D., Chen, C., Chiu, H.J., Chiu, M., Clayton, T., Das, D., Deller, M.C., Duan, L., Ellrott, K., Farr, C.L., Feuerhelm, J., Grant, J.C., Grzechnik, A., Han, G.W., Jaroszewski, L., Jin, K.K., Klock, H.E., Knuth, M.W., Kozbial, P., Krishna, S.S., Kumar, A., Lam, W.W., Marciano, D., Miller, M.D., Morse, A.T., Nigoghossian, E., Nopakun, A., Okach, L., Puckett, C., Reyes, R., Tien, H.J., Trame, C.B., van den Bedem, H., Weekes, D., Wooten, T., Yeh, A., Hodgson, K.O., Wooley, J., Elsliger, M.A., Deacon, A.M., Godzik, A., Lesley, S.A. and Wilson, I.A.** (2010) Structure of the γ-D-glutamyl-L-diamino acid endopeptidase YkfC from *Bacillus cereus* in complex with L-Ala-γ-D-Glu: insights into substrate recognition by NlpC/P60 cysteine peptidases. *Acta crystallographica. Section F, Structural Biology and Crystallization Communications*, 66, 1354–1364.

**Xu, W., Huang, M., Zhang, Y., Yi, X., Dong, W., Gao, X. and Jia, C.** (2011) Novel surface display system for heterogonous proteins on *Lactobacillus plantarum*. *Letters in Applied Microbiology*, 53, 641–648.

**Yamaguchi, H., Furuhata, K., Fukushima, T., Yamamoto, H. and Sekiguchi, J.** (2004) Characterization of a New *Bacillus subtilis* Peptidoglycan Hydrolase Gene, *yvcE* (Named *cwlO*), and the Enzymatic Properties of Its Encoded Protein. *Journal of Bioscience and Bioengineering*, 98, 174–181.

**Yamamoto, H., Hashimoto, M., Higashitsuji, Y., Harada, H., Hariyama, N., Takahashi, L., Iwashita, T., Ooiwa, S. and Sekiguchi, J.** (2008) Post-translational control of vegetative cell separation enzymes through a direct interaction with specific inhibitor IseA in *Bacillus subtilis*. *Molecular Microbiology*, 70, 168–182.

**Yamamoto, H., Miyake, Y., Hisaoka, M., Kurosawa, S.-i. and Sekiguchi, J.** (2008) The major and minor wall teichoic acids prevent the sidewall localization of vegetative DL-endopeptidase LytF in *Bacillus subtilis*. *Molecular Microbiology*, 70, 297–310.

**Yamamoto, H., Murata, M. and Sekiguchi, J.** (2000) The CitST two-component system regulates the expression of the Mg-citrate transporter in *Bacillus subtilis*. *Molecular Microbiology*, 37, 898–912.

**Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y.** (2014) The I-TASSER Suite. Protein structure and function prediction. *Nature Methods*, 12, 7–8.

**Young, K.D.** (2006) The selective value of bacterial shape. *Microbiology and Molecular Biology Reviews*, 70, 660–703.

**Yu, A.C.S., Loo, J.F.C., Yu, S., Kong, S.K. and Chan, T.-F.** (2014) Monitoring bacterial growth using tunable resistive pulse sensing with a pore-based technique. *Applied Microbiology and Biotechnology*, 98, 855–862.

**Zahrt, T.C. and Deretic, V.** (2000) An essential two-component signal transduction system in *Mycobacterium tuberculosis*. *Journal of Bacteriology*, 182, 3832–3838.

**Zapf, J., Sen, U., Madhusudan, A., Hoch, J.A. and Varughese, K.I.** (2000) A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. *Structure*, 8, 851–862.

**Zhang, S., Li, X., Wang, X., Li, Z. and He, J.** (2016) The two-component signal transduction system YvcPQ regulates the bacterial resistance to bacitracin in *Bacillus thuringiensis*. *Archives of Microbiology*, 198, 773–784.

**Zhang, X.-C., Cannon, S.B. and Stacey, G.** (2009) Evolutionary genomics of LysM genes in land plants. *BMC Evolutionary Biology*, 9, 1–13.

**Zuber, U., Drezewiecki, K. and Hecker, M.** (2001) Putative Sigma Factor SigI (YkoZ) of *Bacillus subtilis* Is Induced by Heat Shock Factor SigI (YkoZ) of B Induced by Heat Shock. *Journal of Bacteriology*, 183, 1472–1475.