Re-allocation of cellular resources for the production of heterologous proteins in Bacillus subtilis

Marwa Zaarour

université PARIS-SACLAY

NNT : 2019SACLS213

Thèse de doctorat

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir

Ré-allocation des ressources cellulaires pour la production de protéines hétérologues chez *Bacillus subtilis*

Thèse de doctorat de l'Université Paris-Saclay préparée à l'Université Paris Sud

Ecole doctorale n°577 : Structure et Dynamique des Systèmes Vivants (SDSV)

Spécialité de doctorat: Sciences de la vie et de la santé

Thèse présentée et soutenue à Jouy-en-Josas, le 18 Juillet 2019, par

Mme Marwa Zaarour

Composition du Jury:

Frédérique Braun Maître de conférences, Université Paris VII – EGM-IBPC, Paris	**Rapportrice**
Eric Guédon Directeur de Recherche, INRA – STLO, Rennes	**Rapporteur**
Nicolas Bayan Professeur, Université Paris XI – I2BC, Gif-sur-Yvette	**Président**
Isabelle Martin-Verstraete Professeure, Université Paris VII – Institut Pasteur, Paris	**Examinatrice**
Matthieu Jules Professeur, AgroParisTech – Micalis, Jouy-en-Josas	**Directeur de thèse**
Anne Goelzer Ingénieur de Recherche, INRA – MaIAGE, Jouy-en-Josas	**Invitée**
Vincent Sauveplane Maître de conférences, AgroParisTech – Micalis, Jouy-en-Josas	**Invité**

université PARIS-SACLAY | ÉCOLE DOCTORALE Structure et dynamique des systèmes vivants (SDSV)

Title: Re-allocation of cellular resources for the production of heterologous proteins in *Bacillus subtilis*

Keywords: Gratuitous protein, cell resources, recombinant protein production, resource allocation, *Bacillus subtilis.*

Abstract: Recombinant protein production in microorganisms is of great interest for the production of biopharmaceuticals, therapeutics and industrial enzymes. However, recombinant protein production has always shown a harmful effect on the microorganism cell physiology when excessively produced. Cell resources (*i.e.* metabolites, energy, molecular machinery, cytosolic space, *etc*.) are used to produce the host's proteins and the overproduced gratuitous protein. As a result, this unnatural extra load typically leads to slower growth and lower protein yields, a phenomenon known as 'burden'. This burden comes from the fact that the recombinant protein has no benefit for the microorganism, and that it only uses cell resources at the expense of the production of the endogenous essential proteins. In my PhD project, the issues were (1) to decipher the consequences of gratuitous protein overproduction on the cell physiology, (2) to identify the limiting type of resources, and (3) to overcome this limitation to improve protein production. To address the first issue (1), we analyzed growth rates, production of several proteins of interest, and genome-wide proteomes of *Bacillus subtilis* strains overproducing various levels of reporter proteins. The reporter proteins were chosen so that they were easily quantifiable by fluorescence and β-galactosidase activity assays (*i.e.* GFP, mKate2, LacZ, *etc*.). To obtain the various levels of expression, we built synthetic sequences made of the assembly of various constitutive and inducible promoters and translation initiation regions (TIR, RBS). Hence, we showed that higher was the amount (and size) of the protein produced, lower were the rates of growth and higher were the cell sizes. For instance, the growth rate decreased down by over 20% when GFP was overproduced above 5% of the total soluble protein amount according to both biochemical and fluorescence assays. To further identify the limiting type of resources (2), we performed a relative protein quantification on the strains overproducing GFP at different levels. Hence, we showed that some non-essential proteins were less abundant in the strains overproducing GFP. We next targeted the reporter proteins for degradation using a synthetic tool previously engineered in *B. subtilis*, so that amino acids can be recycled back to the pool of cell resources. Degrading the reporter gratuitous protein should also relieve the constraint on the cytosolic density by liberating intracellular space. With a degradation of 50-60% of GFP and mKate2, we observed a 50% restoration of the growth rate. This result together with the proteome analysis suggested that the amount of amino acids (and consequently their utilization in protein synthesis) was the main limiting type of resources. To overcome this limitation and improve protein production (3), we aimed at exploring a synthetic, amino acid recycling system based on the above mentioned degradation system. We decided to improve the targeted degradation system by overproducing the *E. coli* and *B. subtilis* ClpXP proteases together with an *E. coli* adaptor protein SspB. This tool may allow to target proteins for degradation in order to save resources and improve the production of a protein of interest. We showed that the overproduction of either ClpXP or SspB/ ClpXP were sufficient to allow a complete degradation of the proteins produced low and intermediate levels, and up to 50% of degradation of the proteins highly produced. As ClpXP is a protease involved in stress responses, we aimed to know whether the overproduction of ClpXP may have negative consequences on the cell physiology. We therefore performed relative protein quantification on a strain overproducing ClpXP. The results showed that ClpXP overproduction causes a global reorganization on the proteome without affecting the growth rate of the cell.

Titre: Ré-allocation des ressources cellulaires pour la production de protéines hétérologues chez *Bacillus subtilis*

Mots clés: Protéine gratuite, ressources cellulaires, surproduction de protéines, *Bacillus subtilis.*

Résumé: La synthèse de protéines recombinantes chez les microorganismes est d'un intérêt majeur pour la production de produits biopharmaceutiques, thérapeutiques et enzymatiques industriels. Cependant, la surproduction de protéines à un effet néfaste sur la physiologie cellulaire. Les ressources cellulaires (métabolites, énergie, machinerie moléculaire, espace cytosolique, *etc*.) sont en effet partagées entre les protéines de l'hôte et la protéine "gratuite". Cette surcharge non naturelle entraîne une croissance plus lente et des rendements en protéines plus faibles, un phénomène connu sous le nom de "burden". Dans mon projet de doctorat, il s'agissait (1) de déchiffrer les conséquences de la surproduction de protéines gratuites sur la physiologie cellulaire, (2) d'identifier le type de ressources limitantes, et (3) de surmonter cette limitation pour améliorer la production de protéines. Afin de déchiffrer les conséquences de la surproduction de protéines (1), nous avons analysé le taux de croissance, la production de protéines d'intérêt et le protéome de souches de *Bacillus subtilis* surproduisant divers niveaux de protéines rapportrices. Les protéines rapportrices ont été choisies de manière à être facilement quantifiables par fluorescence et par des tests d'activité (*i.e.* GFP, mKate2, LacZ, *etc*.). Pour obtenir les différents niveaux d'expression, nous avons construit des séquences synthétiques par assemblage de promoteurs constitutifs et inductibles et de régions d'initiation de traduction (TIR, RBS) variés. Nous avons ainsi montré que plus la quantité (et la taille) de la protéine produite était élevée, plus les taux de croissance étaient faibles et plus la taille des cellules était élevée. Par exemple, le taux de croissance a diminué de plus de 20 % lorsque la GFP était surproduite à plus de 5 % de la quantité totale de protéines solubles, selon des quantifications biochimiques et de fluorescence. Pour identifier le type de ressources limitantes (2), nous avons effectué une quantification relative des protéines sur les souches surproductrices de GFP et montré que certaines protéines non essentielles étaient moins abondantes dans ces souches. Nous avons ensuite dégradé spécifiquement les protéines rapportrices à l'aide d'un outil de biologie de synthèse précédemment mis au point pour *B. subtilis*, afin que les acides aminés puissent être recyclés dans le pool de ressources cellulaires. Avec une dégradation de 50-60% de GFP et mKate2, nous avons observé une restauration de 50% du taux de croissance. Ces résultats suggèrent que la quantité d'acides aminés (et par conséquent leur utilisation dans la synthèse des protéines) est le principal type de ressources limitantes. Pour améliorer la production de protéines (3), nous avons cherché à développer un système synthétique de recyclage des acides aminés basé sur le système de dégradation mentionné ci-dessus en surproduisant les protéases d'*E. coli* et *B. subtilis* (ClpXP) avec une protéine adaptatrice (SspB) d'*E. coli*. Cet outil pourrait permettre de dégrader spécifiquement des protéines non essentielles pour économiser des ressources cellulaires. Nous avons montré que la surproduction de ClpXP ou de SspB/ClpXP était suffisante pour permettre une dégradation complète des protéines produites à des niveaux bas et intermédiaires, et jusqu'à 50% des protéines fortement produites. Comme ClpXP est une protéase impliquée dans la réponse au stress, nous avons cherché à savoir si la surproduction de ClpXP pouvait avoir des conséquences négatives sur la physiologie cellulaire. Une quantification relative des protéines sur une souche surproductrice de ClpXP a montré que la surproduction de ClpXP provoque une réorganisation globale du protéome sans toutefois affecter le taux de croissance de la cellule.

Acknowledgements

This PhD project was realized through a collaboration between SyBER group (Systems Biology for bacterial Engineering and Redesign) which forms a part of the GCBS team (Genetic Controls in Bacterial Systems) in MICALIS (unité de Microbiologie de l'Alimentation au service de la Santé) and BioSys team in MaIAGE (unité de Mathématiques et Informatique Appliquées du Génome à l'Environnement) at INRA-Jouy-en-Josas. The project was part of an innovative training network (H2020-MSCA-ITN-2014/"ProteinFactory") coordinated by Prof. Jan Maarten van Dijl, and it was funded by Marie Skłodowska-Curie Actions (MSCA)**.**I would like to thank all the members who contributed to the success of this thesis including the departments of MICA (Microbiologie et Chaine Alimentaire) and MIA (Mathématiques et Informatique Appliquées) for the extra funds.

I would like to thank my director Matthieu Jules, and my co-director Vincent Fromion for accepting me in their teams. Matthieu, thank you for your precious advices and for all what I have learned with you. Vincent, thank you for the help and the valuable discussions that we had.

My PhD project wouldn't have been achieved without the dedication of my supervisors Vincent Sauveplane, and Anne Goelzer. Vincent, you were always available since the beginning of the project, you led me on the right way and you were always ready to help. Anne, you were always willing to help too and you gave me a lot of support all along the 3 years.

I would like to thank as well the reviewers Frédérique Braun and Eric Guedon for reviewing my PhD thesis. I thank as well the examinators Isabelle Martin-Verstraete and Nicolas Bayan for examining the project.

The ITN group with whom I shared a great experience, it was a great opportunity to be part of this consortium. I would like to thank Prof. Jan Maarten van Dijl for coordinating the consortium and for all his advices concerning my project. I would like to thank Dr. Tjeerd van Rij for the nice collaboration we had in DSM-Netherlands, where I learned new techniques that helped me a lot. I would like to thank as well Prof. Dorte Becher and Minia Antelo for the collaboration we had at the University of Greifswald-Germany.

I would like to deeply thank GCBS team and especially SyBER group for the unforgettable experience I had during the past 3 years. Olivier (thank you for the advices and the fruitful discussions), Stephen (who will shout at me by "WHAT DO YOU WANT??!!, I will miss your funny jokes), Anne-Gaëlle (thank you for your motivational support), Etienne (thank you for your precious advices), Magali (I will miss our discussions, you were my supportive colleague), Rahma (my sweetheart colleague), Teddy (who brings the laughter to the team), Gabriela (my companion during the whole period who shared me the good and the bad days), Cecile (the source of positivity in the team), Claire (already missing our 'pause thé'), Simon (keep it up and stay motivated as you are), Irène, Cédric, Lydia, Narimane, Michèle, Aurélie (thank you for your support), Vlademir and Elena (thank you both for your help and support), Dominique, and Philippe.

I thank as well ProCeD team for helping me when I needed. Thanks for all MICALIS members, the PhD students and the Post-docs, and especially the prepa members who were always helpful.

To my friends, I can't tell how much I appreciate having you during the whole period. You were my support in different ways: Elie Saikali, Fatima Al-Reda, Dima El-Khechen, Mustafa Al-Reda, Mohammad Ayoub, Mohammad Tarhini, Hussein Wehbi, Sokna Bazzi, Zeinab and Marwa Koumeiha, Najwa, Myriam Salameh, Mayla and Yasmine.

Lastly, I deeply thank my parents for your endless love and support. All what I have accomplished today is because of you. I am so grateful for my sister Faten, my brother in law Bassam, and my brothers Ahmad and Tarek for being by my side, giving me the strength and motivation, and ofcourse my lovely nieces Sarah and Rina who bring us the joy, love, and the positive spirit.

Content

1 *Bacillus subtilis* use in Biotechnology 16

1.1 *Bacillus subtilis* use for protein production 16

1.1.1 The market of protein production 16

1.1.2 Recombinant protein production in microorganisms 17

1.1.3 *Bacillus subtilis:* characteristics and ecology 17

1.1.4 *B. subtilis:* importance in protein production 18

1.1.5 Advances to improve recombinant protein production 19

2 Resource allocation principles 24

2.1 Introduction to resource allocation principles 24

2.2 The growth rate impact on gene expression 25

2.3 Phenomenological resource allocation models 29

2.4 Genome-wide scale resource allocation models 34

2.5 What are the molecular mechanisms that drive resource allocation? 40

2.5.1 Metabolic pathways regulation 41

2.5.2 ppGpp stringent response 43

2.5.3 The CtsR heat shock response 44

2.6 Bottlenecks of protein production in Bacteria 46

2.6.1 Effect on the growth rate 46

2.6.2 Effect on the protein folding 48

2.6.3 Effect on the cell size 48

2.7 Explanations around the gratuitous protein effect 51

2.7.1 Ribosomes constraint 51

2.7.2 RNA Polymerase constraint 52

2.7.3 Translation and transcription processes 52

2.7.4 Quality control system constraint 52

3 Quality control system 57

3.1 Overview on the quality control system 57
3.2 ATP dependent proteases 58
3.2.1 Family and Structure 58
3.2.2 Members of the AAA+ proteases and their targets 59
3.2.3 Substrate recognition 60
3.2.4 Adaptor proteins 61
3.3 Regulation of proteolysis 63
3.4 Targeted protein degradation in synthetic biology 67
4 Problem Overview 71
5 Gratuitous Protein Overproduction Alters *B. subtilis* Cell Physiology 77
5.1 Introduction 77
5.2 Design and characterization of suitable expression cassettes for overproduction of gratuitous proteins 78
5.2.1 Construction of a set of inducible and well-controlled promoters 78
5.2.2 Selection of the gratuitous proteins to be overproduced 82
5.3 Overproduction of a gratuitous protein and consequences on the *B. subtilis* cell physiology 87
5.3.1 Overproduction of a gratuitous protein in *B. subtilis* reduces the rate of growth 87
5.3.2 Growth rate is altered when the GFP production reaches a certain threshold ... 91
5.3.3 Gratuitous protein overproduction affects the bacterium cell size 94
5.4 *B. subtilis* cells adapt to a huge overproduction of a gratuitous protein by alleviating constraints on the use of amino acids 97
5.4.1 GFP overproduction in *B. subtilis* results in reduced production of other proteins 97
5.4.2 Degradation of the overproduced gratuitous protein restores the growth rate . 113
5.5 Discussion 119

5.5.1 Demonstration of the gratuitous protein over production impacts on *B. subtilis* 119

5.5.2 Overproduction of a heterologous protein leads to a reduction in non-essential endogenous proteins of the NRPS family 120

5.5.3 The amount of the gratuitous protein with respect to thetotal soluble protein determine the impact on the growth rate 121

5.5.4 Relation of cell size and high protein production..... 123

5.5.5 Amino acids are the limiting resource in a limited cytosolic density 124

6 Development of a synthetic biology tool to finely tune gene expression 132

6.1 Introduction 132

6.2 Characterization of the targeted degradation system in the presence of SspB 133

6.3 Over-producing SspB, ClpX, and ClpPin a single operon 135

6.3.1 Design of the ClpXP-based degradation system 135

6.3.2 Characterization of the system with different promoters 138

6.4 Consequences of ClpXP over production on the proteome..... 141

6.4.1 General description 142

6.4.2 Major differences in the ClpXP overproduction strain 144

6.5 Discussion..... 148

6.5.1 Targeted protein degradation improvement 148

6.5.2 ClpXP overproduction reorganizes the proteome 148

7 Conclusion and Perspectives..... 153

8 Materials and Methods 159

8.1 Molecular Biology..... 159

8.1.1 Methods for strain construction..... 159

8.1.2 Polymerase Chain Reaction 160

8.1.3 DNA fragment purification 161

8.2 Bacterial transformation methodology 161

8.2.1 *E. coli*specific methodologies 161

8.2.2 *Bacillus subtilis* specific methodologies 161
8.3 Bacterial strain construction 162
8.3.1 The set of inducible promoters 162
8.3.2 Strains with reporter genes 164
8.3.3 LacI-expressing strain 165
8.3.4 SspB-expressing strain 165
8.3.5 *sspB-clpXP* operon 166
8.4 Growth media 172
8.5 Live Cell Array 173
8.5.1 Culture preparation and data acquisition 173
8.5.2 Data treatment 173
8.6 Miller Assay for beta-galactosidase activity 176
8.7 Protein Gels and Western Blotting 177
8.7.1 Cell culture 177
8.7.2 Cell lysis 177
8.7.3 Protein gels and sample preparation 178
8.8 Flow cytometry 180
8.9 Proteomics 180
8.9.1 Relative protein quantification performed at the University of Greifswald-Germany 180
8.9.2 Relative protein quantification performed at PAPPSO platform at INRA 182
9 Résumé Détaillé en Français 194
9.1 Introduction 194
9.2 Résultats 195
9.2.1 Le choix des promoteurs 196
9.2.2 Le choix de protéine 197
9.2.3 Surproduction de protéines et conséquences sur la physiologie cellulaire de *B. subtilis* 199

9.2.4 Le taux de croissance est modifié lorsque la production de GFP atteint un certain seuil 203

9.2.5 La surproduction de *B. subtilis* par les GFP entraîne une réduction de la production d'autres protéines 204

9.2.6 La dégradation de la protéine gratuite surproduite rétablit le taux de croissance 205

9.2.7 Développement d'un outil de biologie synthétique pour affiner l'expression des gènes 207

9.3 Conclusion 208

References 210

LIST OF FIGURES

Figure 2.1 Growth curve of *E. coli* in synthetic medium as a function of the glucose concentration (Monod 1949). 24
Figure 2.2 Representative graph of the diauxic shift of *E. coli* as documented by Jacques Monod 25
Figure 2.3 Cell mass of Salmonella typhimurium as a function of its growth rate 26
Figure 2.4 The impact of the different parameters on the gene expression 27
Figure 2.5 Representation of the growth rate-dependent parameters affecting the gene expression 28
Figure 2.6 GFP abundance and translation efficiency decreases with increasing growth rate 29
Figure 2.7 Self-replicator system consisting of 4 enzymes and a membrane 30
Figure 2.8 Experimental data of shifts in metabolism in *L. lactis* 31
Figure 2.9 Interactions between a synthetic circuit and a host cell 32
Figure 2.10 The effect of unnecessary protein in the cell 33
Figure 2.11 Proteome partitioning models with three and four fractions 33
Figure 2.12 The representation of the basis of FBA model. 35
Figure 2.13A scheme of the cell showing the constraints imposed by RBA model (Goelzer et al. 2011) 37
Figure 2.14 RBA predictions versus experimental measurements in different media 38
Figure 2.15 Summary of the dataset for RBA calibration and validation in each condition: growth rate, number of measured proteins and their allocation (in fraction of total measured protein cost) between the main cellular functions and uptake and excretion rates of nutrients 39
Figure 2.16 Global and local regulations of a metabolic pathway 42
Figure 2.17 The mechanism of (p)ppGpp stringent response 44
Figure 2.18 Heat shock response in *B. subtilis* 45
Figure 2.19 RBA simulation to predict the AmyE excretion flux when including and neglecting the chaperon cost. 47
Figure 2.20 Gratuitous protein effect shown on *E. coli* 47
Figure 2.21 Scheme of the cell's response to high protein production 49
Figure 2.22 UDP-glc availability affects the localization of the cell division inhibitors 51
Figure 2.23 Illustration of the tradeoff between the different proteins produced in the cell 53
Figure 3.1A scheme for the AAA^+ superfamily 59
Figure 3.2 Representation of the mechanism of degradation 61
Figure 3.3 The process of ATP-dependent proteolysis of a substrate recognition by means of the adaptor protein 63

Figure 3.4 Sigma B regulation in stress conditions..... 65
Figure 3.5 CtsR regulation under different environmental conditions..... 66
Figure 3.6 SsrA sequence amino acid sequence recognized in *E. coli*..... 68
Figure 4.1 A representation of the resource allocation to the cell tasks..... 72
Figure 5.1 Gene dosage is influenced by location on the chromosome..... 78
Figure 5.2 Growth curves followed by LCA for the wild type versus the strains carrying the synthetic inducible promoters upstream *gfp* in CHG medium..... 81
Figure 5.3 GFP abundance variation with the synthetic promoters (s*P*)..... 82
Figure 5.4 GFP-LacZ protein is less abundant in the cell..... 83
Figure 5.5 Growth rate is not affected with the synthetic promoters..... 83
Figure 5.6 Microscopic images showing the GFP fluorescence..... 84
Figure 5.7 Miller Assay for β-Galactosidase activity..... 86
Figure 5.8 Growth rate of LacZ strain and a control strain..... 86
Figure 5.9 Comparison between P_{veg} and P_{hs} promoters fused to *gfp* and *lacZ* in a rich medium 87
Figure 5.10 Growth rate reduction shown on *B. subtilis* upon expression of *lacZ* gene under the control of P_{veg} promoter..... 88
Figure 5.11 Slower growth rate for all the strains carrying gratuitous genes in different defined media..... 89
Figure 5.12 Growth rate variation with respect to different levels of GFP in different defined media..... 91
Figure 5.13 The heterologous protein quantification with respect to the total soluble proteins extract..... 93
Figure 5.14 GFP abundance in *E. coli* and *B. subtilis*..... 94
Figure 5.15 Microscope images for *B. subtilis* strains over-producing gratuitous proteins..... 95
Figure 5.16 Flow cytometry data..... 96
Figure 5.17 Growth curves of the wild type and GFP overproducing strains in S medium..... 98
Figure 5.18 GFP abundance in the strains over producing GFP by one and two copies of the expression cassette P_{veg} *sfgfp*..... 99
Figure 5.19 Pie chart representation of the number of differentially produced proteins in GFP overproducing strains depending on the ratio between the level of one protein in the wild-type strain compared with the GFP overproducing strains..... 101
Figure 5.20 Heat map representation of the 298 most affected proteins in the wild type and GFP producing strains. On the right, zoom on the 10 most affected proteins and their respective metabolic functions in *B. subtilis*..... 103
Figure 5.21 Correspondence between the normalized concentration and the normalized mass in the control strain (MZ72). Large proteins have a significant impact in terms of protein cost with a low concentration such as srfAB (401 kDA) or srfAA (402kDa). Small proteins such

as the elongation factor Ef-Tu (tfuA gene - 43kDa) with a concentration ten times higher than srfAA has a cost in the order of magnitude of that of srfAA 105
Figure 5.22 Correspondence between the normalized concentration and the normalized mass in the one copy GFP strain (MZ139) 106
Figure 5.23 Correspondence between the normalized concentration and the normalized mass in the two copies GFP strain (MZ140) 107
Figure 5.24 Protein mass repartition between the one GFP copy strain (MZ139) and the control strain (MZ72) 108
Figure 5.25 Protein mass repartition between the two GFP copies strain (MZ140) and the control strain (MZ72) 109
Figure 5.26Influence of GFP overexpression on the production of proteins encoded by the *pps* operon responsible for plipastatin production in *B. subtilis* 110
Figure 5.27 Influence of GFP overexpression on the production of proteins encoded by the *dhb* operon responsible for enterobactin production in *B. subtilis* 111
Figure 5.28 Influence of GFP overexpression on the production of proteins encoded by the *srfA* operon responsible for surfactin production in *B. subtilis* 112
Figure 5.29 Variation of GFP abundance between the untagged and tagged GFP. 113
Figure 5.30Addition of tetracycline on exponentially growing cells reveals the quick degradation of the GFP^{ALGG} as compared to the unmodified GFP 115
Figure 5.31 Growth rate restoration upon protein degradation 117
Figure 5.32Fluorescence measurement by flow cytometry 118
Figure 5.33 Cell pole localization of protein aggregates. 120
Figure 5.34 GFP abundance decreases with increasing growth rate 122
Figure 5.35 Representation of applying a targeted degradation system on the gratuitous protein 126
Figure 5.36 Beta-galactosidase activity of LacZ before and after degradation 127
Figure 6.1 GFP abundance of the tagged GFP measured in different media in the presence and absence of the adaptor protein 134
Figure 6.2 Western Blot image showing the sfGFP with the different ssrA* 135
Figure 6.3 Operon design for the degradation system 136
Figure 6.4 LCA measurements for the different operons 138
Figure 6.5 Different levels of GFP^{ALGG} production and degradation with and without the induction of *sspB clpX clpP* 139
Figure 6.6 Protein degradation at different levels of GFP and mKate abundances 140
Figure 6.7 Comparable growth rate between the wild type and the P_{veg}*clpXP* strain. 142
Figure 6.8 Changes in the protein abundance in the wild type and the ClpXP mutant strains . 143

Figure 6.9 Number of proteins (grouped in categories) differentially expressed between the wild type and ClpXP mutant..... 144
Figure 6.10 High abundance of ribosomal proteins in the ClpXP mutant..... 147
Figure 6.11 Low abundance in the motility proteins in ClpXP mutant 147
Figure 8.1 Gibson Assembly method 160
Figure 8.2 Representation of the IPTG-inducible promoter Phs..... 163
Figure 8.3 Illustration of the synthetic inducible promoters..... 163
Figure 8.4 Illustration of the *amyE* and *nprEloci* in *B. subtilis* chromosome..... 165
Figure 8.5 The genetic construct scheme of LacI-strain 166
Figure 8.6 The LCA 96-well microtiterplate preparation 175
Figure 8.7 Beta-galactosidase enzymatic reaction..... 176
Figure 8.8 Schematic drawing representing the SDS-PAGE gel setup 179
Figure 8.9 Scheme showing the transfer procedure setup. 180
Figure 8.10 SDS-PAGE gels for the soluble protein extracts which were treated by PAPPSO . 182
Figure 9.1 Variation de l'abondance de GFP avec les promoteurs synthétiques (sP)..... 197
Figure 9.2 La protéine GFP-LacZ est moins abondante dans la cellule que la GFP native..... 198
Figure 9.3 Taux de croissance d'une souche produisant LacZ et d'une souche témoin (Control) 199
Figure 9.4 Réduction du taux de croissance sur *B. subtilis* lors de l'expression du gène *lacZ* sous le contrôle du promoteur P_{veg}..... 200
Figure 9.5 Ralentissement de la croissance de toutes les souches porteuses de gènes gratuits dans différents milieux définis. 201
Figure 9.6 Données de cytométrie en flux 202
Figure 9.7 Quantification hétérologue des protéines par rapport à l'extrait total de protéines solubles..... 203
Figure 9.8 Représentation sur heat map des 298 protéines les plus affectées dans le type sauvage et les souches productrices de GFP. A droite, zoom sur les 10 protéines les plus affectées et leurs fonctions métaboliques respectives chez *B. subtilis* 204
Figure 9.9 Restauration du taux de croissance lors de la dégradation des protéines..... 206
Figure 9.10 Nombre de protéines (regroupées en catégories) exprimées différemment entre le type sauvage et le mutant ClpXP 208

LIST OF TABLES

Table 1-1 List of different industrial enzymes and the industry (sector) they are used in. Adapted from (Kumar et al. 2014)..... 16
Table 5-1 The constitutive promoters and the TIR elements constituting each of the synthetic inducible promoters..... 80
Table 5-2 The mean and the standard deviation (std) of the size of each of the strains in the gated zone..... 97
Table 6-1 Design of the various *sspB clpX clpP* operons..... 137
Table 6-2 Percentage of GFP degradation with SspB and the various *sspB clpX clpP* operons. 137
Table 6-3 Similarities between the DNA microarray analysis in the case of translational inhibition and protein abundances in the case of ClpXP overproduction..... 145
Table 6-4 Abundance of the cell division proteins in the ClpXP overproducing strain..... 146
Table 8-1 The sequences of the IPTG-inducible synthetic promoters (*s*P)..... 164
Table 8-2 The sequences of the genes composing the 4 different operons..... 168
Table 8-3 Strains used in this work..... 183
Table 8-4 The primers used for the strain constructions..... 186

CHAPTER 1

Bacillus subtilis use in Biotechnology

1 *Bacillus subtilis* use in Biotechnology

1.1 *Bacillus subtilis* use for protein production

1.1.1 The market of protein production

According to "Markets and Markets"; a global market research company, the enzyme market was evaluated at 4.61 billion USD in 2016 (https://www.marketsandmarkets.com/PressReleases/industrial-enzymes.asp). It includes the use of different kinds and categories of enzymes in many types of markets. In more details, enzymes such as amylases, lipases, proteases, cellulases, and phytases of plant, animal and microorganisms origins are used mainly in the food and beverages, animal feed, biofuels, detergents, pharmaceuticals, and textile industries [Table 1-1]. The main key producers of industrial enzymes are BASF SE (Germany), E.I. du Pont de Nemours and Company (U.S.), Associated British Foods plc (U.K.), Koninklijke DSM N.V. (Netherlands), Novozymes A/S (Denmark), and others. Moreover, the industrial enzyme market is estimated to grow at a CAGR (Compound annual growth rate) of 5.8% and expected to reach a value of 6.30 billion USD by 2022. This rise in the industrial enzyme market is a key factor for the scientific challenge in transforming microorganisms to cell factories for protein production.

Sector	Enzymes
Laundry detergents	Alkaline Protease, Alkaline Lipase, Alkaline Cellulase, Alkaline Amylase
Pulp and paper industry	Cellulase, Xylanase, α-amylase, Lipase, Ligninase, Laccase, Mannanases
Leather industry	Alkaline and acid Protease, Alkaline and acid Lipase
Starch and sugar industry	Glucose isomerase, Glycosyltransferase, Dextranase, α-amylase, Glucoamylase, Xylanase, Pullulnase
Baking industry	α-amylase, Xylanase, Lipase, Protease, Pentosanase, Oxidoreductase
Dairy industry	Chymosin, lyzozyme, Lipases, β-galactosidase, Lactase
Brewing industry	Amyloglucosidase, Proteae, Pentosanase, Xylanase
Animal feed industry	Phytase, Xylanase, β-Glucanase

Table 1-1 List of different industrial enzymes and the industry (sector) they are used in. Adapted from (Kumar et al. 2014).

1.1.2 Recombinant protein production in microorganisms

The high demand on commercial enzymes and pharmaceuticals encouraged scientists to think about ways to obtain biological tools with high protein production and secretion. The production of metabolites and pharmaceuticals can be achieved by using traditional processes such as chemical synthesis or by catalytic enzymes. Chemical synthesis has drawbacks such as low catalytic efficiency, need for high temperature, low pH, and high pressure (Adrio & Demain 2014). In that way, biological enzymes are more helpful due to their stability and specific activity. The production of enzymes in microorganisms progressed in 1980s-1990s. Before that, they were extracted from plant and animal sources. However, they were obtained in low yields and sold in high prices. The discovery of microbial enzymes made it easier as microbial cells can grow much faster than animal or plant cells (Demain & Vaishnav 2011). In addition, microbes are simpler organisms for genetic manipulation so they form potential hosts to produce proteins. Some microbial enzymes highly used in industry are proteases, amylases, lipases, lactases, xylanases, and others. During the 1990s, the recombinant DNA technology progressed and further promoted the use of microbes in industry. So nowadays, proteins from plants and animal sources are produced in microbes thanks to genetic, metabolic, and protein engineering. Some of the proteins and pharmaceuticals produced by recombinant production are human insulin, human growth hormone, albumin, plant phytase, *etc*.

1.1.3 *Bacillus subtilis:* characteristics and ecology

Bacillus subtilis is a rod shaped Gram-positive bacterium. Gram-positive (G^{+ve}) bacteria have a single membrane at the internal face of a thick peptidoglycan, unlike Gram-negative (G^{-ve}) bacteria, which have an inner and outer membrane, and a gel-like periplasmic space in between.

B. subtilis lives on organic sources such as substrates of plant origins and on decaying material such as dead roots (Siala et al. 1974). For instance, growing *Arabidopsis thaliana* in a sterile soil with *B. subtilis* resulted in a biofilm formation on the root surfaces of the plant (Zou et al. 2013). In addition, *B. subtilis* is considered as a rhizobacterium, *i.e.* it promotes plant growth by helping in nutrient uptake and in defense against fungal attacks (Emmert J. & Handelsman 1999; Asaka & Shoda 1996; Zou et al. 2013; Kumar et al. 2012). Its biocontrol activity was clearly shown in (Zou et al. 2013) where the formation of protective and antibacterial biofilms of *B. subtilis* was observed against the pathogenicity of *P. syringae.* For

this reason, *B. subtilis* and its relatives are also used as bio pesticides (Das & Mukherjee, 2006; Rodgers, 1993; Thakore, 2006).

Moreover, *B. subtilis* can form resistant dormant endospores in environmental stress conditions or when facing nutritional deprivation (Earl et al. 2010). The spores can be dispersed by wind, so they can go long distances. As *B. subtilis* lives in the soil, it can be ingested by animals, and was found to act as a probiotic in helping to maintain healthy bacterial communities in the body (Hong et al. 2005).

B. subtilis is non-pathogenic and so, it is considered as GRAS (Genetically Recognized as Safe) by the US Food and Drug Administration. *B. subtilis* has been commercially used since a long time due to its secreted enzymes such as amylases, lipases, and proteases, its antipathogenic properties, and its use in Natto food production. *B. subtilis* is genetically amenable, easy to manipulate, and it is naturally competent (Solomon & Grossman 1996). Due to these beneficial properties, this model organism has been extensively studied for a long time. The interesting characteristics of this model organism encouraged researchers to sequence its genome (Kunst & al. 1997), and to study its transcriptomes (Jürgen et al. 2001; Nicolas et al. 2012), proteomes (Eymann et al. 2004), and metabolomes (Soga et al. 2003) across various environmental conditions.

1.1.4 *B. subtilis:* importance in protein production

Bacillus species and *Escherichia coli* are the most commonly used prokaryotes for the industrial production of recombinant proteins (Westers et al. 2004). *Bacillus* sp. contribute to around 60% of theavailable commercial enzymes, while *E. coli* is mostly used for the industrial production of pharmaceutical proteins (Westers et al. 2004). The *bacillus* biochemistry and physiology have been studied in details. Gram positive bacteria have good secretion systems, in contrast to Gram negative bacteria which accumulate proteins in their periplasmic space (Schallmey et al. 2004). *Bacillus* secretes many of its proteins at high concentrations to the external medium, which provides a more cost effective means of producing proteins than extraction from the cytoplasm or periplasm. As a consequence, *Bacillus* sp are highly used for the industrial enzyme production for their high fermentation properties, high production yields 20-25 g/L, and of course for the absence of toxic by-products (van Dijl & Hecker 2013). In contrary, *E. coli* and more generally G^{-ve} bacteria have lipopolysaccharides (LPS); that is referred to be as endotoxins. LPS are present in the outer membrane of their cell wall and they add complications to detoxify the secreted products. The

absence of LPS in G^{+ve} bacteria allows to obtain proteins with less processing requirements for co-purification of contaminants. Thanks to the abovementioned positive points, *B. subtilis* is mainly used in the production of industrial enzymes originating from *bacilli* (Palva 1982; Pohl et al. 2013). On the other hand, multicopy plasmids are not known to be stable in *B. subtilis* (Bron et al. 1991; Ehrlich et al. 1991; Leonhardt & Alonso 1991).This drawback explains that the production of pharmaceutical and therapeutic proteins takes place in *E. coli,* which has the ability to maintain multicopy plasmids.

1.1.5 Advances to improve recombinant protein production

Recombinant protein production is usually increased by (i) overexpressing the gene of interest using strong promoters and translation initiation regions, and/or (ii) by inserting in the host the gene of interest in multicopy, either by multiple integrations into the genome, or by using a multicopy plasmid. Some work was done on obtaining strong synthetic promoters in *B. subtilis* and other microorganisms. The design of a dual promoter P_{hpaII}-P_{amyQ} resulted in very high extracellular production of α and β-CTGase and pullulanase that are highly used in food, cosmetics, pharmaceuticals, and chemical industries (Zhang et al. 2017). Other engineered synthetic promoters are based on native promoters such as the library that was created by random mutations on P_{srfA} (Han et al. 2017), or by randomizing the P_{veg} promoter and the TSS (Transcription Start Site) (Guiziou et al. 2016). These approaches and many others aimed to obtain high expression levels of recombinant proteins and to finely tune gene expression.

Beside the importance of the expression level, there is the importance of the expression systems that are suitable for each kind of protein. Mainly the post-translational modifications of the protein determine the host organism to be chosen. The expression systems which are highly used in recombinant protein production are cell cultures of bacteria, yeasts, molds, mammals, plants, or insects (Demain & Vaishnav 2011). Most of the recombinant proteins that are commercially available are produced in bacteria such as *E. coli* and *B. subtilis* due to their fast growth, ease in manipulation, and well-studied metabolism. Yeasts are used to assemble large fragments and to produce proteins that require more complex folding. By contrast, mammalian cell cultures are used to produce mammalian proteins that requires complex post-translational modifications. For example, if the protein is glycosylated then it is favored to be produced in an eukaryotic system. If it possesses disulfide bonds, then it can be produced in a prokaryotic system (Overton 2014). Thus, the choice of

the host organism is a key issue to obtain recombinant proteins accurately folded and in high amount.

Summary:

The enzyme market is projected to grow up to 6.30 billion USD by 2022. Various enzymes are used in different industrial sectors. The high demand of enzymes encouraged scientists to use microorganisms due to their ease in manipulation and their quick growth. Attempts were made to enhance protein production, both quantitatively and qualitatively. *B. subtilis* is highly used in the industry because of its well-known biochemistry and physiology. It is a rod-shaped Gram-positive bacterium that lives in the soil, it is non-pathogenic and so it is considered as GRAS. *Bacillus* has an efficient secretion system. It secretes many of its proteins at high levels into the external medium. Nevertheless, high protein production is facing difficulties due to secretion stress. The interest of achieving high yield of secreted protein pushed scientists to explore the secretion stress. But until today the consequence of high production level on the cell physiology and on the growth rate of *B. subtilis* have not been documented yet.

CHAPTER 2.

Resource allocation principles

2 Resource allocation principles

2.1 Introduction to resource allocation principles

Bacterial growth and adaptation to the environment has been intensively studied since more than 60 years. Theoretical and experimental investigations on the relation between the medium and the growth rate in exponential phase started with Jacques Monod in the 1940s (Monod 1949). During the exponential phase, the cells are in steady-state growth, *i.e.* the growth rate is constant and all the relative concentration of enzymes and metabolites are constant. Monod showed that the growth of the cells in exponential growth can vary with respect to the extracellular substrate concentration [Figure 2.1], as a Michaelis-Menten like relation:

$$\mu = \mu_{max} \frac{s}{K_s + s}$$

μ stands for the growth rate, s stands for the nutrient concentration, μ_{max} is the rate limit for increasing concentrations of s, and K_s is the concentration of the nutrient at which the rate is half the maximum.

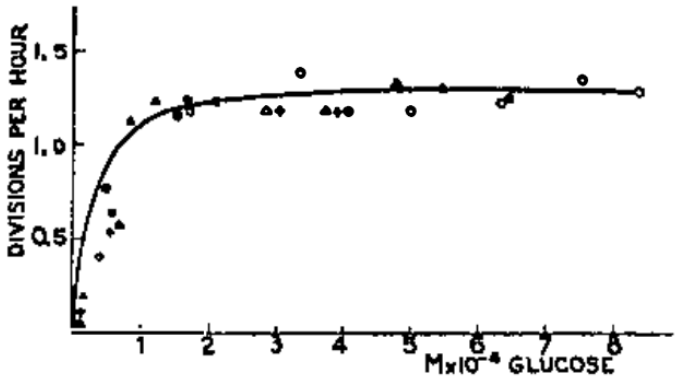

Figure 2.1 Growth curve of *E. coli* in synthetic medium as a function of the glucose concentration (Monod 1949).

The exponential growth varies with the substrate concentration as a Michealis Menten equation, where the cell is considered as an enyme.

Moreover, the growth rate can vary when different carbon sources are present in the medium. Monod showed that the growth profile of *E. coli* with glucose and sorbitol, consisted of two consecutive exponential phases, separated by a lag phase. The two exponential phases appeared to have different growth rates. This phenomenon was called the diauxic shift. *E. coli* grows by consuming glucose first then it passes through a lag phase, as an adaptation phase, allowing to produce the enzymes needed to metabolize the second carbohydrate, and finally

continues its growth by consuming the second sugar but at a slower growth rate [Figure 2.2]. This mechanism shows that microorganisms have the tendency to equilibrate the use of metabolites in order to serve the needs and the metabolic capabilities. They have regulatory mechanisms to control the expression of genes needed to uptake and metabolize the available carbon source. Therefore, the cell economizes the use of its resources and regulates the expression of the necessary genes to escape the exhaustion of its capabilities.

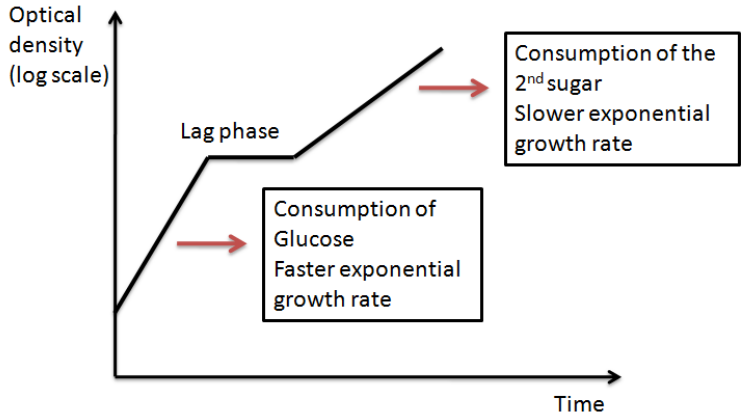

Figure 2.2 Representative graph of the diauxic shift of *E. coli* as documented by Jacques Monod

When *E. coli* is grown on glucose and a different carbohydrate, it consumes glucose first, then it passes through an adaptation phase to be ready to consume the second sugar, then it resumes its growth at a slower rate.

2.2 The growth rate impact on gene expression

In response to external influences such as medium composition or environmental stress, metabolites are distributed to feed the cell needs. As a consequence, gene expression is adapted which leads to an adaptation of the macromolecular composition. These changes are reflected by the growth rate of microorganisms. Replication, transcription and translation are dependent on the growth rate through the RNA polymerases, ribosomes amount, and the gene copy number (Bremer & Dennis 1996). The first findings about the cellular macromolecular compositionin relation to the growth rate date back 60 years ago (Schaechter & Kjeldgaard, 1958). Schaechter and coworkers showed that at a given temperature the DNA, RNA, protein amounts and cell size are dependent on the growth rate of *Salmonella typhimurium*. As shown on Figure 2.3, the cell mass increases with increasing growth rate. The same increase applies on the RNA and DNA. However, the relative proportions of these parameters change with the growth rate, so that they can be described as exponential functions of the growth rate. Moreover, they showed that regardless of the medium composition, each medium resulting in

identical growth rate produces identical physiological states; mainly cell size and composition.

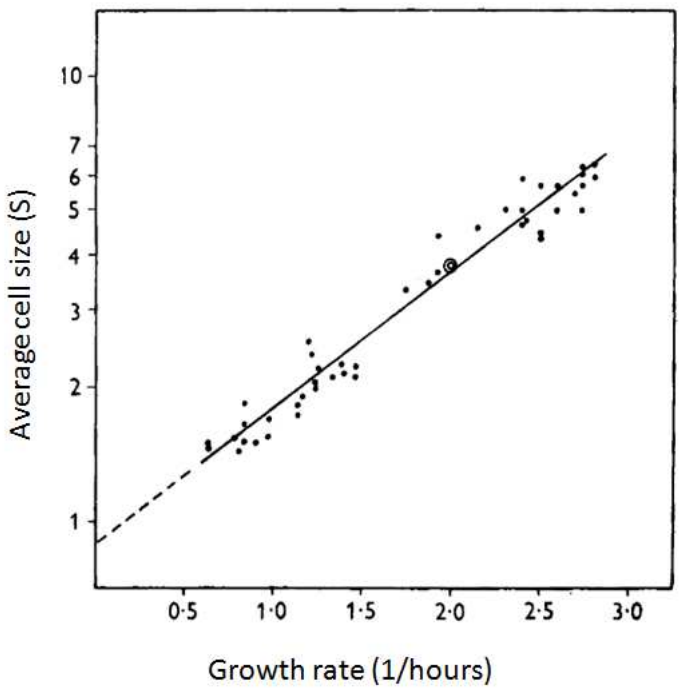

Figure 2.3 Cell mass of Salmonella typhimurium as a function of its growth rate

The average cell size indicating the cell mass, increases with increasing growth rate (Schaechter & Kjeldgaard 1958)

The "Copenhagen school" with Maaloe and Kjeldgaard gave experimental determination of the biomass composition in *E.coli.* They found that the macro-molecular composition is dependent on the growth rate. A mathematical interpretation of the empirical relationships of the Copenhagen was achieved by Cooper et al. (Cooper 1968), who derived an equation to evaluate the amount of DNA per average cell during the exponential phase as a function of the time required to replicate the chromosome, the time between termination of replication and the next division, and the culture doubling time. Later in 1982, a team progressed on obtaining mathematical equations for the DNA, RNA, and protein amount of bacteria as a function of growth rate (Churchward et al. 1982). Other experimental studies were achieved to enhance the understanding of the cell composition and its change with the growth rate by introducing new concepts such as the average mass per cell (Donachie 1968), the relation between RNA and protein (Schleif 1967).

These relations between the macromolecular composition and growth rate were revisited by Marr (1991), who built a mechanistic model taking into account ribosome

synthesis, translation, and the regulation by the alarmone, ppGpp, while considering that the cell density was constant (Marr 1991). Using this model, he was able to recover the known changes in the abundances of the ribosomal and the non-ribosomal proteins with respect to the growth rate. In such a model, the growth rate results from a trade-off between the availability of charged tRNAs and the level of protein synthesis. The 'Marr' model already implicitly contained the underlying mechanisms governing resource allocation that were to be revealed 20 years later (see section 2.3, 2.4). Finally, Bremer and Dennis unified the different findings that were published in a review (Bremer & Dennis 1996; Dennis & Bremer 2008).

The global effects dependent on the growth rate and on gene expression were revisited in (Klumpp S et al. 2009). Klumpp et al. (2009) investigated the dependence of the protein and mRNA expression levels of a constitutively expressed gene on the growth rate. By doing so, Klumpp and coworkers revisited the work of Bremer & Dennis (Bremer & Dennis 1996). The gene expression was modeled using 6 growth rate-dependent parameters, the gene copy number in the cell, the transcription rate per copy of the gene, the mRNA degradation rate, the translation rate per mRNA, the protein degradation rate, and the cell volume. The figure below [Figure 2.4] illustrates the contribution of these parameters to the protein concentration with respect to the growth rate.

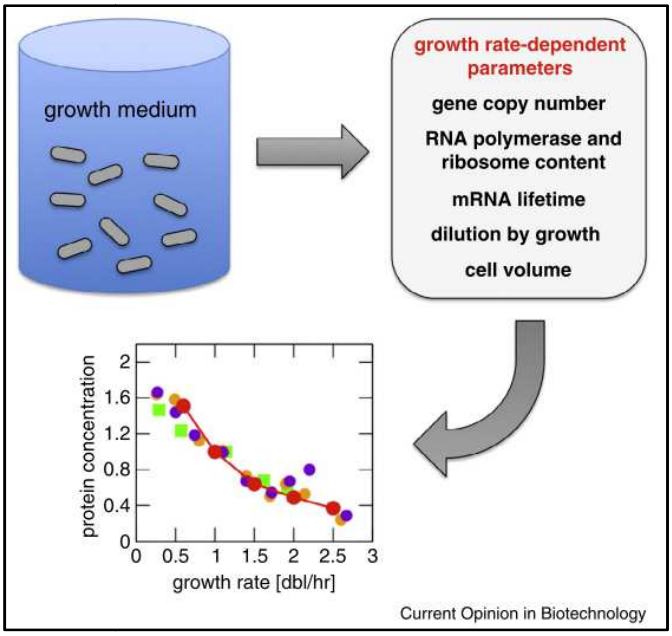

Figure 2.4 The impact of the different parameters on the gene expression

The growth medium impacts the growth rate of the cell, which in turn influences 6 growth-rate dependent parameters. The different listed parameters were used to model the gene expression. The protein concentration decreases with increasing growth rate. Taken from (Klumpp & Hwa 2014).

The variations of the 6 parameters with respect to the growth rate are presented in [Figure 2.5]. We can see that with increasing growth rate, all of the transcription rate, the gene copy number, the protein dilution, and the cell volume increase, while the mRNA degradation rate and the translation rate remain constant. Klumpp et al. (2009) suggested that the increase in the dilution rate explains the decrease in protein abundance at higher growth rate. However, the dependency on the growth rate is more complex. Indeed, Michealis-Menten like relationships were derived with *free* ribosomes [Figure 2.6] (Borkowski, Goelzer, et al. 2016), and *free* polymerases (Gerosa et al. 2013). These model suggests that the drop in protein concentration at higher growth rate is not only due to an increased dilution rate but also to a decrease in translation efficiency because of less available free ribosomes (Borkowski, Goelzer, et al. 2016).

Therefore, the effect of the growth rate is pleiotropic. It has a role not only in the macromolecular composition, but also in the functioning of molecular machines such as ribosomes, and RNA polymerases. These concepts were then revisited within the framework of the resource allocation principle since 2009.

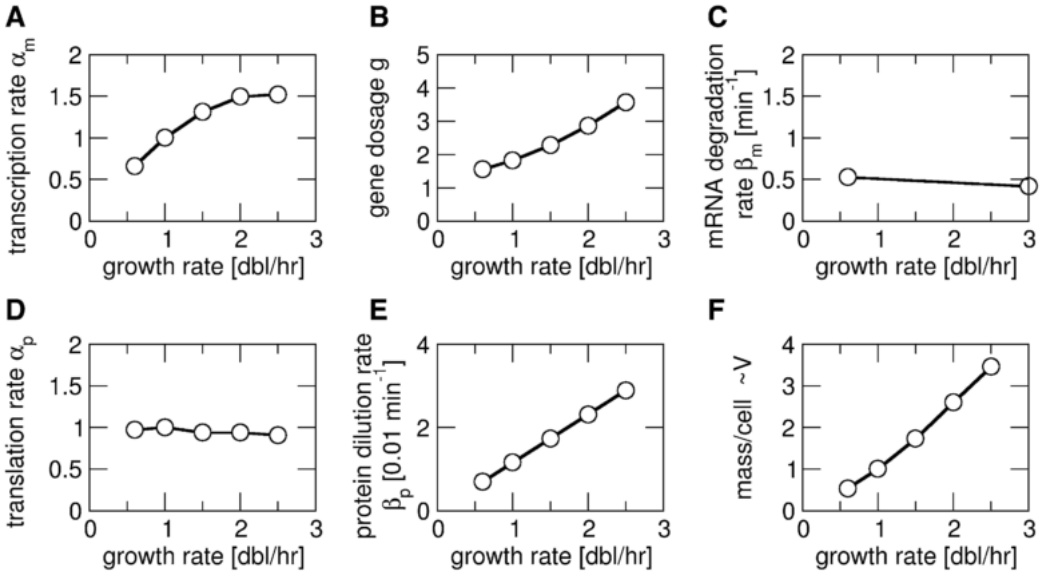

Figure 2.5 Representation of the growth rate-dependent parameters affecting the gene expression

(A) The transcription rate per gene. (B) The gene dosage or copy number per cell. (C) The mRNA degradation rate. (D) The translation rate per mRNA molecule. (E) The protein dilution rate due to growth. (F) The cell mass that was used to measure the cell volume. All the mentioned parameters are presented as a function of growth rate (Klumpp S et al. 2009).

A. **B.**

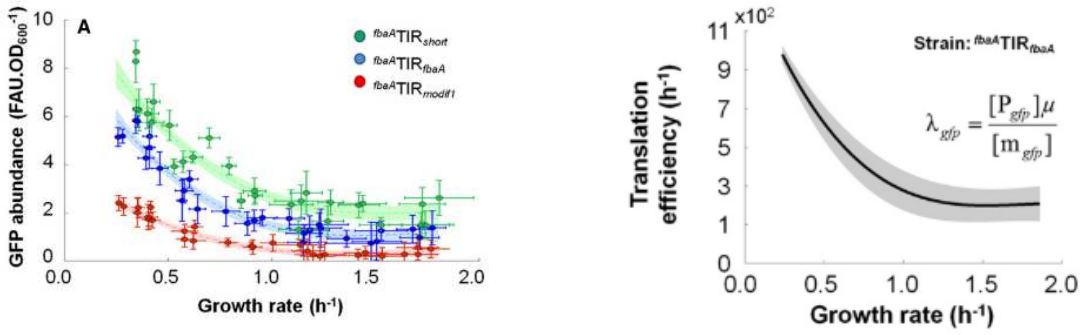

Figure 2.6 GFP abundance and translation efficiency decreases with increasing growth rate

(A) The *gfp* was expressed under different promoters, and the GFP abundance corresponding to protein production was measured at different growth rates. GFP abundance decreases with increasing growth rate. (B) Translation efficiency for one construct as a function of the growth rate. The translation efficiency decreases with the growth rate which explains the decrease in the GFP abundance in A. (Borkowski, Goelzer, et al. 2016).

Summary:

Bacterial growth was first studied by Monod when he theoretically and experimentally investigated the relation between the medium and the growth rate. He suggested the first growth law which gives the cell's growth rate during exponential phase. Then, studies on *Salmonella typhimurium* by Scheachter suggested that the cell size, the DNA and RNA contents change proportionally with the growth rate. Later on, the Copenhagen school gave experimental determination of the biomass composition in *E.coli.* They found that macro-molecular components are dependent on the growth rate. Mathematical interpretations were performed in different studies to find the DNA, RNA, and protein content as a function of growth rate in bacteria during the exponential phase. All these findings were revisited with Bremer & Dennis to give detailed explanation of the chemical composition and its change with the growth rate. Recent models were built by the late 2000s to interpret the global effects dependent on the growth rate and on individual gene expression.

2.3 Phenomenological resource allocation models

The cell's growth rate results from the trade-off in the cellular resources towards the metabolic network, the translational apparatus, and the housekeeping proteins (Scott et al.

2010). Starting from the substrate uptake, it requires transporters to be produced according to the concentration of the extracellular substrate. Then, depending on the carbon source, specific metabolic enzymes are required. Molenaar et al. (2009) built a non-linear constraint-based model that predicted a shift in the growth strategy depending on the available nutrients (Molenaar et al. 2009). The cell was considered as a self-replicator made of 3 enzymes, one ribosome and one structural component in the form of lipid membrane [Figure 2.7]. Thus, the ribosome synthesizes both the ribosome and the other proteins from a given metabolic precursor. Hence, Molenaar et al. (2009) explained the experimentally observed gradual shifts in growth strategies when growing on low/high substrate concentrations. Experimental evidence of the gradual shift in the growth strategy of *L. lactis* is presented in [Figure 2.8]. The tendency to shift from metabolically to catalytically efficient metabolisms with increasing substrate concentration is the result of optimizing the cellular economy for the growth. The same behavior of metabolic shift to inefficient metabolism is observed when producing a recombinant protein in *E. coli*. The model can simulate the consequences of recombinant protein production when introducing a fixed level of a useless protein, which has no function except occupying a volume and a fraction of the ribosome's capacity. When Molenaar et al. (2009) looked at a fixed growth rate and a low substrate concentration where the efficient metabolism is usually used, the useless protein production resulted in an apparent shift to inefficient metabolism. Indeed, the recombinant protein production takes place at the expense of other proteins, so it decreased the ribosome's capacity/availability.

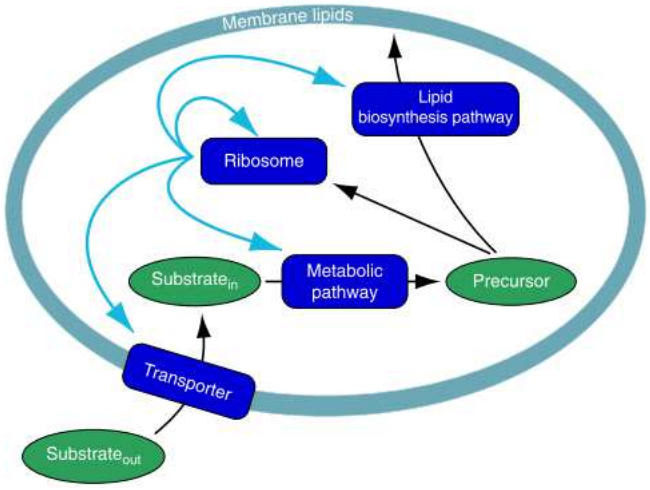

Figure 2.7 Self-replicator system consisting of 4 enzymes and a membrane

The ribosome synthesizes the metabolic enzymes, the lipid biosynthesis enzymes, the transporter proteins, and itself (Molenaar et al. 2009).

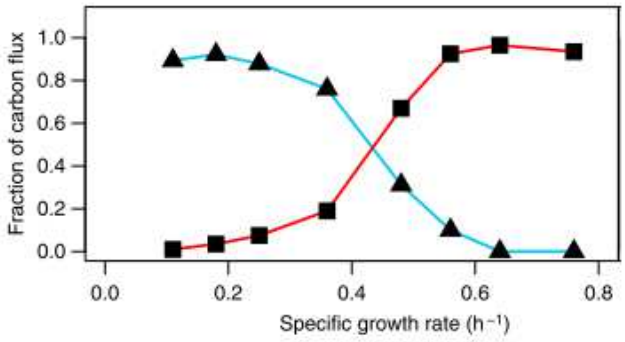

Figure 2.8 Experimental data of shifts in metabolism in *L. lactis*

L. lactis shifts from the metabolically efficient mixed-acid fermentation at low growth rates to lactic acid fermentation at high growth rates (Molenaar et al. 2009)

Another model emerged that coupled gene expression with growth rate from the perspective of resource allocation (Weiße et al. 2015). The model of Weiße and coworkers is also a coarse-grained model and integrates 3 tradeoffs that can be considered as universal, such that they have been experienced in all living organisms. The 3 trade-offs are the finite levels of cellular energy, ribosomes, and proteome (cell mass). The model implements the tradeoffs by considering two core processes: gene expression and nutrient import and metabolism. In comparison to the previous model of Molenaar et al. (2009), Weiße and coworkers (2015) added the transcription process and the competition of mRNA binding to *free* ribosomes. The model of Weiße and coworkers (2015) can thus predict the effects of synthetic circuits insertion in host cells. As a test case, they introduced a repressilator in the cellular chassis. The repressilator is composed of 3 repressive genes [Figure 2.9 A] (Elowitz & Leibler 2000). The model predicted a sigmoidal decrease in growth with increasing induction of the genes from the synthetic circuitry (*i.e.* the repressilator). At low induction levels, the expression of the synthetic circuit is produced at the expense of housekeeping proteins and ribosomes. Weiße et al. (2015) stated that the cell can compensate for this load and the consequent reduction of energy levels through transcriptional regulation and repartitioning of the proteome. However, at a stronger induction, the competition of the mRNAs of the synthetic circuit for *free* ribosomes fully inhibits the production of the host enzymes needed for nutrient transport and metabolism. Consequently, the decrease in the host proteins leads to a drop in the growth rate [Figure 2.9 B].

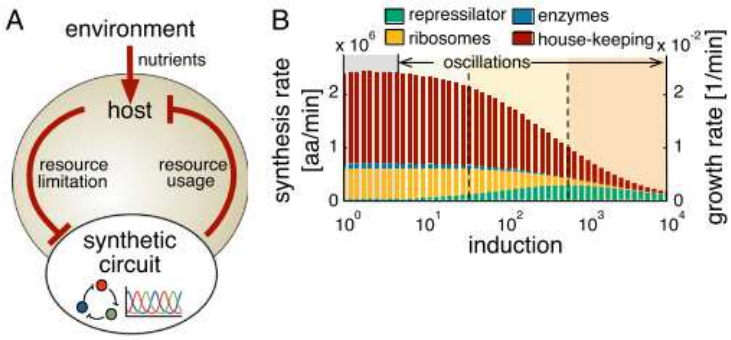

Figure 2.9 Interactions between a synthetic circuit and a host cell

(A) A representation of the interaction between the environment, the host cell, and the synthetic circuit represented by a repressilator. (B) The growth rate and the resources decrease with increasing inductions of the synthetic circuit.

A second coarse-grained model was proposed, which is based on proteome partitioning into growth rate-independent fraction that includes housekeeping proteins (Q-class), and a growth rate-dependent fraction that includes the translation apparatus, such as ribosomal and other translation proteins (R-class), and metabolic proteins such as transporters, catabolic, and anabolic enzymes, (P-class) (Scott et al. 2010a). Depending on the growth rate, the allocation of the fractions changes, so that if one fraction is increased, it is at the expense of the other fraction. An efficient resource allocation requires that the abundance of P- and R-classes are adjusted so that the rate of nutrient influx provided by P matches the rate of protein synthesis achievable by R. In addition, this model allowed to predict the allocation of the fractions when a recombinant protein was considered [Figure 2.10]. The effect of unnecessary protein production decreased the fraction allocated towards the P- and R- sectors leading to a decrease in the growth rate. Later on, the same group extended the model by adding another fraction. They split the ribosomal proteins sector into ribosomal fraction and a fraction of T proteins, which is related to the growth rate-dependent translation speed [Figure 2.11]. Both fractions increased and decreased together. The increase in the translational speed added a cost on the cell to produce more proteins (*i.e.* tRNA synthetases and elongation factors). Therefore, Scott and coworkers revealed that the allocation of the proteome fraction as a function of the growth rate and the medium mainly depends on a constraint related to the ribosome in determining the cell physiology.

The above-mentioned models illustrate the resource allocation principle by constraints imposed on the finite proteome. The models showed different aspects of resource allocation

and firmly helped to conclude that resource allocation to different cell processes is the cornerstone of the limitation of the growth rate (Goelzer & Fromion 2011). Protein and ribosome synthesis, translation capacity, and cellular energy were the main constraints imposed on the cell to manage the resource allocation (Molenaar et al. 2009; Weiße et al. 2015; Scott et al. 2010). However, these models remained at a macro-molecular scale. Therefore, novel models were later proposed to investigate resource allocation genome-wide.

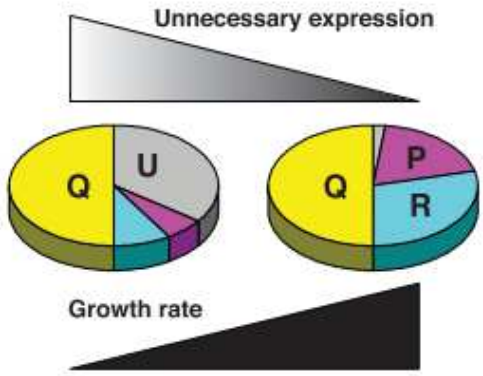

Figure 2.10 The effect of unnecessary protein in the cell

The proteome is divided between the housekeeping proteins (Q), ribosomal proteins (R), other proteins (P), and unnecessary protein (U). When the unnecessary protein takes a big part of the proteome, it decreases the resource allocation towards the ribosomal proteins leading to growth rate decrease. (Scott et al. 2010)

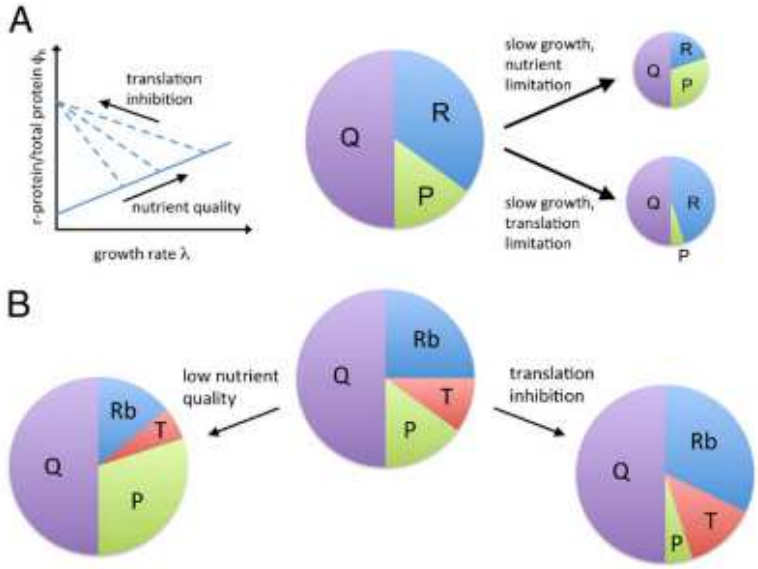

Figure 2.11 Proteome partitioning models with three and four fractions

(A) three fractions of the proteome presented in (Scott et al. 2010), it consists of the growth rate-independent Q fraction including the housekeeping proteins, the two growth rate-dependent R fraction including the ribosomal and translational proteins, and P fraction including the metabolic proteins. The

R fraction has a positive correlation with the growth rate and with translational inhibition. (B) four fractions of the proteome presented in (Scott et al. 2014), it splits the R-fraction into ribosomal protein fraction Rb and translation speed-affiliated proteins T.

Summary:

We presented here three phenomenological models that include the principle of resource allocation between a few cell functions. The self-replicator model by Molenaar and co-workers is the simplest one, considering that the cell contains one enzyme (the ribosome), that is responsible for synthesizing itself and all the other enzymes (metabolic and transporters). The model predicted a gradual shift in growth as the medium changes from low to high substrate concentration. The same kind of behavior is observed with a useless protein. This shift is the result of optimizing the cellular economy for the growth rate. Then, Weiße and co-workers came with a more detailed model, where they introduced the parameters of transcription. They considered 3 cellular traeoffs: finite levels cellular energy, finite levels of ribosomes, and finite levels of proteome (cell mass). They implemented them into two core proteins: the gene expression and the nutrient uptake and metabolism. They suggested that adding a synthehtic circuit in the cell results in a decrease in the growth rate because of the competition of the mRNAs for the ribosomes. Lastly, the third model for Scott and coworkers, is similar in partitioning the proteome into 3 sectors: the growth rate independent sector contains the housekeeping proteins (Q), and the growth rate dependent proteins containing the ribosomal proteins (R) and the metabolic proteins (P). They suggested that the useless protein leads to a decrease in the R- and P- sectors which results in a growth rate decrease.

2.4 Genome-wide scale resource allocation models

In parallel to phenomenological macro-molecular models, emerged more sophisticated genome-scale models, through the introduction into the models of a larger set of cellular entities such as mRNAs, tRNAs, ribosome amount, chaperones, metabolic fluxes, *etc*. The basics of the genome-scale models refers to the flux balance analysis model (FBA), where they introduced the constraints on the metabolic fluxes through the mass conservation at steady-state (Orth et al. 2010; Varma & Palsson 1994). Actually, the rate of change of metabolite concentrations in the metabolic network is given by

$$\frac{\boldsymbol{dx(t)}}{\boldsymbol{dt}} = \boldsymbol{Sv(t)}$$

where **x(t)** represents the vector of size **m**of metabolite concentrations, and **v(t)** is the vector of metabolic fluxes of size **n** associated to the biochemical reactions that consume and produce metabolites. **S** is the stoichiometry matrix of size **m** x **n**, in which each coefficient (s_{ij}) give the stoichiometry of the i-th metabolite in the j-th reaction. At steady-state, the metabolite concentrations and the fluxes are constant, which leads to a set of equalities $\boldsymbol{Sv = 0}$**.** In addition, FBA, and more generally, constraint-based models enable to set inequality constraints on some fluxes. Typically, inequality constraints impose the flux distribution boundaries on the network by ($\boldsymbol{\alpha_i \leq v_i \leq \beta_i}$) which define solution in an allowable space. In practice, they are also used to model thermodynamic constraints, where the flux is nonnegative or irreversible reactions ($v_i \geq 0$), and be positive or negative for reversible reactions.

Finally, FBA predicts the optimal metabolic flux distribution that maximizes or minimizes a given criterion, such as maximizing the growth rate at steady state. This is modelled by the objective function $\mathbf{Z = C^T v}$**,** where **C** is the vector that represents how much each reaction contributes to the objective function [Figure 2.12]. Finally, the FBA problem is:

$$\begin{array}{ll} maximizes & C^T v \\ v \in \mathbb{R}^n & Sv = 0 \\ & \alpha_i \leq v_i \leq \beta_i \end{array}$$

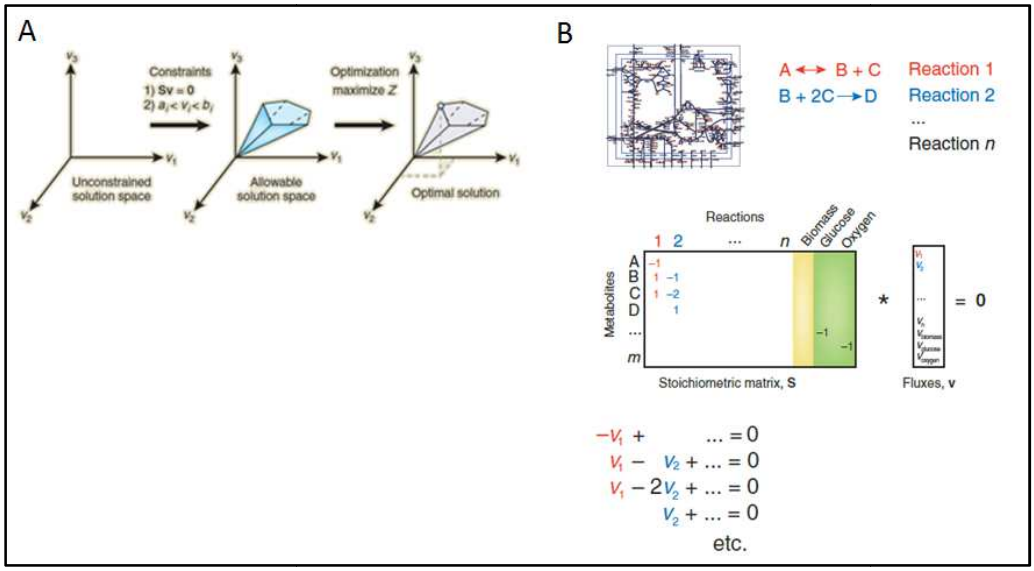

Figure 2.12 The representation of the basis of FBA model.

(A) The constraints imposed on the system by ($\boldsymbol{\alpha_i} \leq \boldsymbol{v_i} \leq \boldsymbol{\beta_i}$) which define solution in an allowable space. An objective function Z is set to be maximized through optimizing the metabolic flux distribution. (B) the stoichometric matrix (S) of size (m x n) representing the metabolites (m) and the reactions (n) multiplied by all the fluxes (Orth et al. 2010)

FBA models succeeded in predicting fluxes within the metabolic network and also the growth rate for cells grown on different growth media or using different carbon sources, or even when increasing the flux towards the production of a metabolite of interest. However, FBA does not take into account the cost of the proteins produced by the cells, knowing that the protein production cost is the most important cost according to the resource allocation principle. In addition, FBA does not consider any constraint of space inside the cell (cytosolic cell volume/size), which is not reasonable as the enzymes that catalyze the metabolic reactions also compete for the cytoplasmic space (Beg et al. 2007; Vazquez et al. 2008). Based on this idea the authors of (Beg et al. 2007) improved the FBA method by adding a constraint on the limited cytosolic space allocated to the enzymes.

Since 2009, the Resource Balance Analysis (RBA) framework has emerged as an extension of FBA to introduce and formalize more constraints at the genome-scale level between the cellular processes in *B. subtilis.* The RBA method mathematically formalizes the interactions and allocation of resources between the cellular processes (Goelzer et al. 2011; Goelzer & Fromion 2011; Goelzer et al. 2015; Goelzer & Fromion 2009). All these relationships take the form of linear growth-rate dependent equalities and inequalities, for cells growing in exponential phase at a rate μ,

(I) the metabolic network has to produce all metabolic precursors necessary for biomass production;

(II) the capacity of all molecular machines such asenzymes and transporters for the metabolic network, ribosomes and chaperones for macromolecular processes must be sufficient to ensure their function, *i.e.* to catalyze chemical conversions at a sufficient rate;

(III) the intracellular density of compartments and the occupancy of membranes are limited

(IV) mass conservation is satisfied for all molecule types.

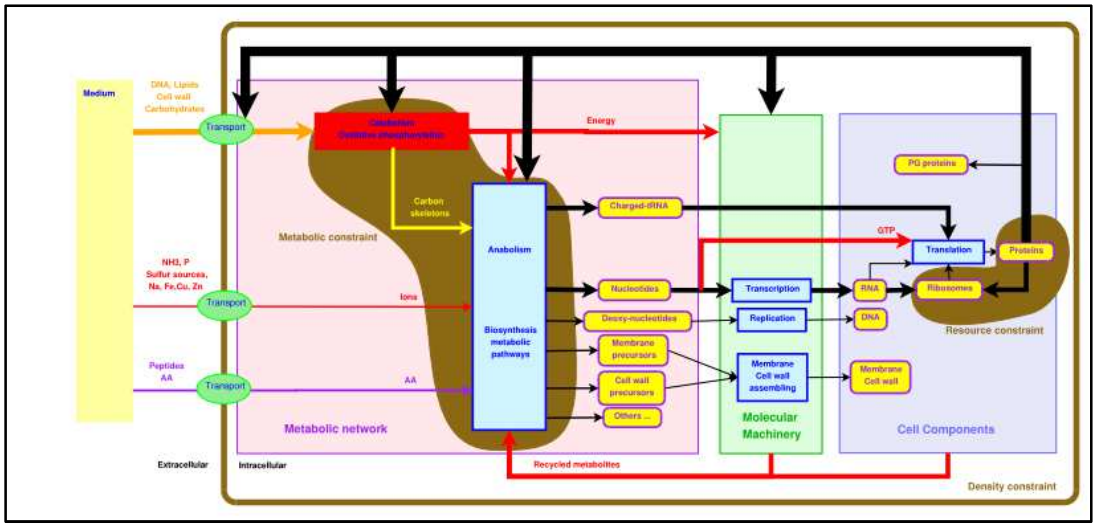

Figure 2.13A scheme of the cell showing the constraints imposed by RBA model (Goelzer et al. 2011)

The metabolic constraint where the metabolic network has to produce all the precursors for different cell processes such as transcription, translation, and membrane and cell wall assembly. The cell processes require molecular machinery, energy and metabolites which are considered as cell resources. The constraint of these resources is that they have to be available to ensure the proper cellular functions for cell growth. The molecular machines and the metabolic resources occupy a cytosolic space of limited cell density, thus forming the density constraint.

Taken together, the equalities and inequalities define, at a given rate μ, a feasibility linear programming (LP) problem that can be efficiently solved. Actually, the equalities and inequalities define the set of all possible phenotypes of the cell at given growth rate. Parsimonious resource allocation between cellular processes is modeled mathematically by optimizing the maximal cell growth, and computed by solving a series of such LP feasibility problems for different growth rate values (Goelzer et al. 2011; Goelzer & Fromion 2011; Goelzer et al. 2015). For a given medium, solving an RBA optimization problem predicts the maximal possible growth rate, the corresponding reaction fluxes (including the substrate up-take and by-product secretion rates) and the abundances of molecular machines. However, other criteria could be maximized, such as the maximal secretion rate of a heterologous protein at given growth rate (see section 2.6.1).

The cell is in this case considered as a whole interconnected system made of "subsystem" (Goelzer et al. 2011), where each subsystem has a function that contributes to the different cell tasks, and at the same time it necessitates different form of resources to fulfill this job for achieving a maximum growth rate. The accuracy of RBA predictions requires an accurate estimation of the protein abundance in the cell and the protein efficiency such as the catalytic activity of each enzyme, or translation rates. To do so, the growth rate and protein

costs were measured in different growth media. The catalytic activity was considered for enzymes as the apparent catalytic rate $\boldsymbol{k_E}$ that is calculated from the flux $\boldsymbol{v_i}$ and the enzyme abundance $\boldsymbol{E_i}$ by the equation $|v_i| = k_{E_i} E_i$. Considering the absolute protein abundances and the metabolic fluxes through the substrate uptake and secretion rates the apparent catalytic rate can be estimated [Figure 2.15]. Using the estimated parameters, RBA predictions were of high accuracy when compared to experimental measurements [Figure 2.14].

In complement with RBA, there exist two other constraint-based modeling methods that integrated the notion of resource allocation: CAFBA (Mori et al. 2016) and ME-Model (O'Brien et al. 2013). In CAFBA, the authors introduced global constraints on the fluxes that encode for relative adjustment of proteome sectors at different growth rate of *E. coli*. The proteome was partitioned based on the previous phenomenological models (Scott et al. 2010a), but dividing the P-sector into biosynthetic enzymes (E-sector) and proteins devoted to carbon intake and transport (C-sector). By adding these constraints, they formalized the interplay between growth and the expression of genes related to the metabolism in order to make quantitative predictions. Therefore, CAFBA is a genome-scale metabolic model that embed metabolic pathways and an aggregated complex reaction for biomass composition.

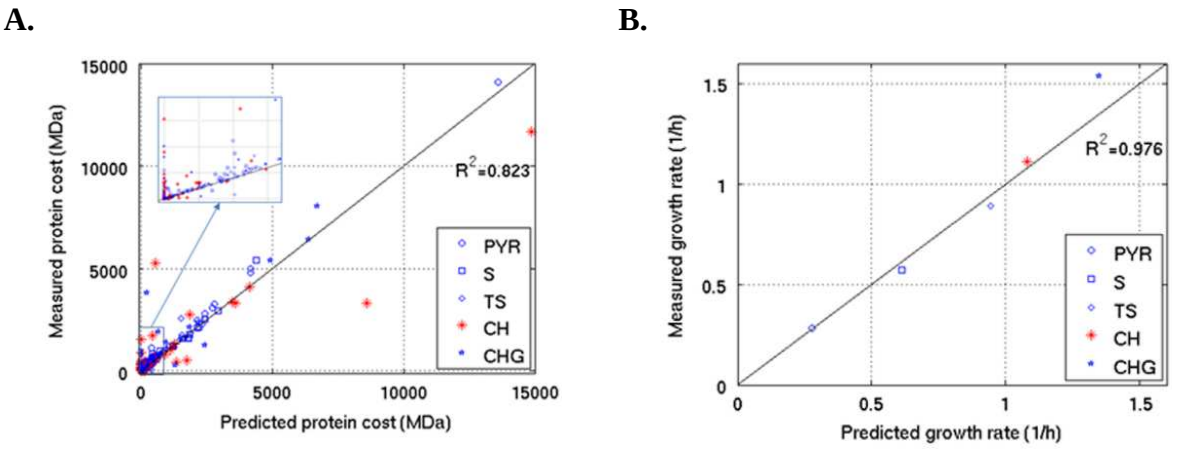

Figure 2.14 RBA predictions versus experimental measurements in different media

(A) Predicted protein cost by RBA agrees with the experimentally measured protein cost for cellular processes in different growth media (CHG, CH, TS, S, PYR) ranging from very rich to very poor medium respectively. (B) Experimentally measured growth rate agrees as well with the predicted growth rate (h^{-1}).

The ME-model (O'Brien et al. 2013) is a metabolic and gene expression genome-scale model, which aims at computing the optimal cellular state for growth in a given steady-state

environment. At a given nutrient composition, the model predicts the cell's maximum growth rate, the substrate uptake and by-product secretion rates, metabolic fluxes, and gene product expression levels. The constraints imposed by the model are mainly at the level of the proteome and at the uptake of nutrients. The ME model predicts the proteome segments allocation and regulation in response to different carbon sources to provide the optimal growth rate (Brien et al. 2016).

The models presented in this chapter led to a better understanding of the resource allocation principle in microorganisms. All of the models agreed on the principle and the outcome, but they differed in their complexity. From the FBA, the simplest model, to RBA the most complex model, the framework is progressing to understand resource allocation in microorganisms and to give better predictions for metabolic engineering and synthetic biology applications.

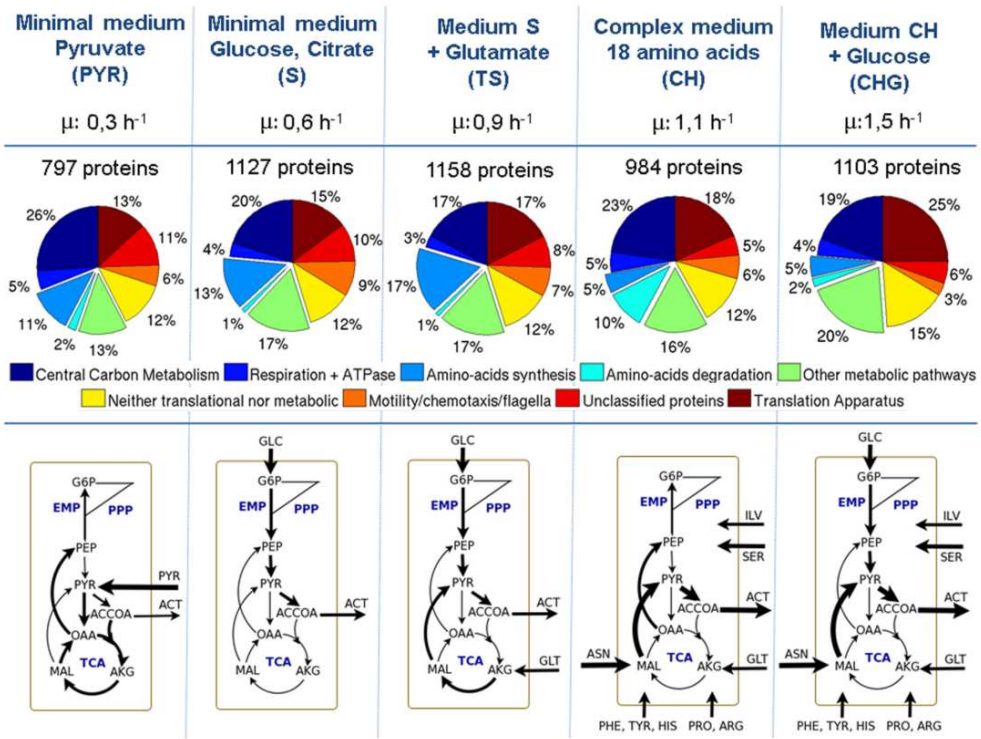

Figure 2.15 Summary of the dataset for RBA calibration and validation in each condition: growth rate, number of measured proteins and their allocation (in fraction of total measured protein cost) between the main cellular functions and uptake and excretion rates of nutrients

The protein cost of one biological process is determined as the sum of the costs of each individual proteins involved in that process (e.g. number of protein copies per cell multiplied by mass (Da)). For flux panel, the thickness of arrows is proportional to the absolute flux value. Abbreviations are PPP Pentose Phosphate Pathway, TCA Tricarboxylic acid cycle, EMP Embden–Meyerhof pathway, PYR pyruvate, ACT acetate, AKG 2-oxoglutarate, MAL Malate, OAA oxaloacetate, PEP phosphoenolpyruvate, G6P glucose-6-phosphate, GLC glucose, ACCOA acetyl-coA, GLT glutamate, ASN asparagine, SER serine, ILV isoleucine, valine, leucine, PHE phenylalanine, TYR tyrosine, HIS histidine, PRO proline, ARG arginine (Goelzer et al. 2015).

Summary

Genome-wide resource allocation models emerged from previous coarse-grain models. By introducing more cellular entities, the models progressed to give a better understanding of the resource allocation. Starting by FBA, the basic model that provided a stoichiometry matrix of reactions and introduced constraints on the metabolic fluxes at steady state by the mass conservation. The FBA drawback was that it does not take into account neither the cost of the proteins, nor their occupied intracellular space. Improvements were provided by adding the constraints on the cell density and on the proteome. Then, in 2009, RBA model for *B. subtilis* emerged providing a detailed refinement of the cell by imposing constraints on all cell processes. It predicts for a given medium, the maximal possible growth rate, the corresponding reaction fluxes, the molecular machines abundances. In complement with RBA, emerged two constraint-based models: CAFBA model and ME-model. CAFBA introduced global constraints on the fluxes that encodes for relative adjustment of proteome at different growth rate. The ME-model is a metabolic and gene expression genome-scale model and aims at computing the optimal cellular state for growth in a given steady state environment. They predict the allocation of the proteome sectors with different carbon sources for an optimal growth rate. They impose the constraints mainly at the metabolic fluxes, the substrate uptake and the proteome.

2.5 What are the molecular mechanisms that drive resource allocation?

The configuration of the cell, or in other words the phenotype of the cell is obtained through the metabolic and genetic regulatory network. The initial point in the cell system is its response to the environment such as the nutrients, the aeration, and the physiological parameters. These conditions positively or negatively influence signal transduction pathways, which results in transcriptional activation or inhibition of genes whose products are metabolic proteins or quality control proteins and others. Overall, the result is an interrelated system driving the resource allocation to respond to what is needed, so that the system does not exhaust. In the following subsections, we briefly review the main molecular regulatory mechanisms that control the gene expression of the main cellular processes that are relevant for this Ph.D.

2.5.1 Metabolic pathways regulation

In response to environmental changes, global genetic regulations link the metabolic network with the genetic network (Janga et al. 2006; Goelzer et al. 2008; Buescher et al. 2012). This occurs through the metabolites pool, which circulates through the biochemical reactions forming the metabolic network, such as the central carbon metabolism, aerobic and anaerobic respiration, fermentation, amino acids, nucleotides, and fatty acids metabolism. Genetic regulations sense the level of metabolites (their effectors), and then turn on/off genes whose products form part of the metabolic network. Figure 2.16 illustrates the most common regulation of anabolic synthesis pathway. A transcription factor (TF_2) senses the level of the last metabolite of the pathway, and adjusts the level of the enzymes of the pathway accordingly. The (TF_2) can further be regulated by global genetic regulations such as (TF_1) in response to environmental/physiological conditions. A reconstruction of the link between the metabolic and the genetic regulatory network for *B. subtilis* (Goelzer et al. 2008), and *E. coli* (Janga et al. 2006) have been reported.

When growing on a complex medium composed of several carbon sources, bacteria tend to select a preferred one, such as glucose, and repress the uptake of the other ones (Monod 1949). This phenomenon is called the catabolite repression, and the molecular mechanisms driving the catabolic repression are called the carbon catabolite repression (CCR) system (Stülke & Hillen 1999). In *E. coli*, the phosphotransferase system (PTS) is responsible for the transport and phosphorylation of carbohydrates. If we take the example of the *lac* operon (contains genes responsible for the lactose metabolism), in the absence of glucose, the glucose-specific permease ($EIIA^{Glc}$) stimulates adenylate cyclase which in turn modulates the concentration of cAMP (cyclic adenosine monophosphate). cAMP interacts with CAP (catabolite activator protein) to bind to the DNA and allow the transcription of the *lac* operon. However, the presence of glucose leads to lower concentration of cAMP. As a result, lactose permease is inhibited, so that the *lac* operon transcription is repressed (Bru & Titgemeyer 2002; Siegal 2015). In *B. subtilis*, the CCR system involves the catabolite control protein CcpA which binds DNA to the CRE operator (catabolite responsive element) (Warner & Lolkema 2003; Fujita 2009). The regulation is carried out through an HPr Kinase that phosphorylates the HPr and/or Crh at a serine residue to produce P-Ser-HPr and/or P-Ser-Crh. The P-Ser-HPr and/or P-Ser-Crh form a complex with CcpA, which allows CcpA to bind to CRE binding site (Deutscher et al. 1994; Galinier et al. 1999). The P-Ser-HPr is able to stop the induction of catabolic genes and operons by phosphorylating inducer enzymes,

transcriptional activators, or anti-terminators (Stulke & Arnaud 1998; Stülke & Hillen 1999; Bru & Titgemeyer 2002).

Moreover, mechanisms of gene regulation occurs when the bacteria choose to switch to nitrate respiration or fermentation in the absence of oxygen. There are signal transduction pathways that positively regulate the transcription of the genes whose products play a role in nitrate respiration. For instance, ResD and ResE are proteins required for aerobic and anaerobic respiration. When ResD is phosphorylated by the ResE kinase, it turns on the transcription of the *fnr* gene. Then, FNR activates the regulatory pathway for the anaerobic respiration by up-regulating the transcription of the operon *narGHJI*. This later codes for the nitrite reductase required for nitrate respiration (Sun et al. 1996).

These examples illustrate how regulatory pathways determined by growth conditions influence gene expression to fulfill the cellular requirements without wasting energy and resources for unnecessary products.

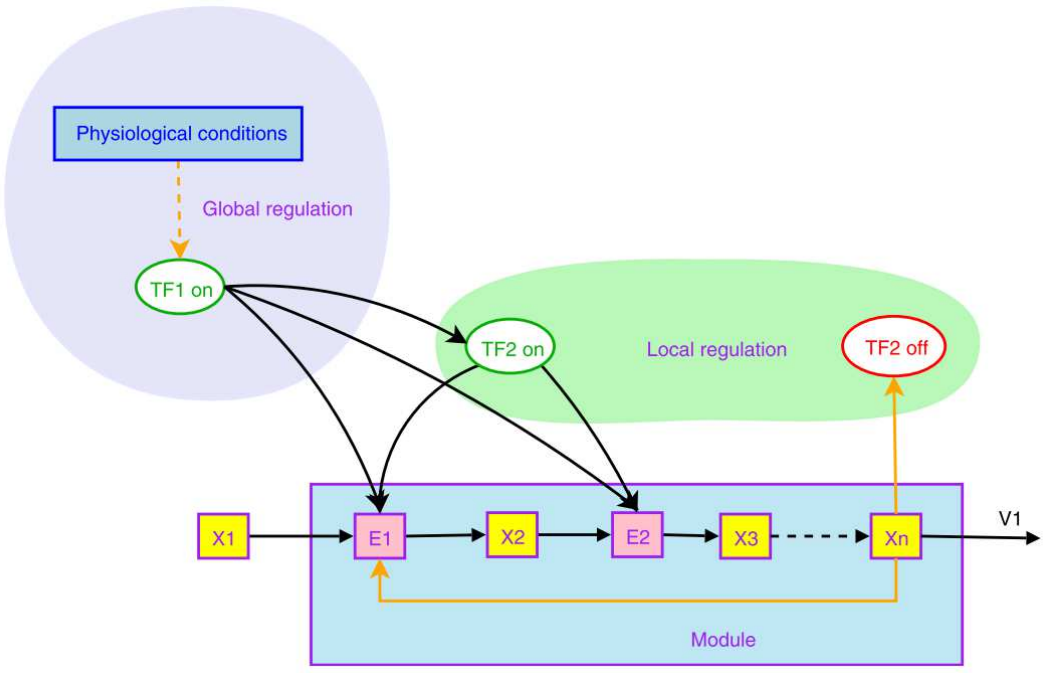

Figure 2.16 Global and local regulations of a metabolic pathway

Yellow boxes represent metabolite pools, pink boxes represent the enzymes, the transcription factors (TF) are on or off, meaning that it is able to bind to DNA or not. The local transcription factor (TF2) is sensitive to an intermediate metabolite (Xn), and modulates the synthesis of enzyme(s) involved in the pathway. The global regulator TF1, sensitive to another signal, can modulate (i) the synthesis of intermediate enzymes; (ii) the synthesis of the local transcription factor TF2, or (iii) both.

2.5.2 ppGpp stringent response

When the cell faces nutrient deprivation, bacteria adjust gene expression to maintain growth and enhance survival. The adaptation to nutritional stress is mediated by the accumulation of the alarmone guanosine 5', 3' bipyrophosphate and pentaphosphate (ppGpp and pppGpp respectively, which will be collectively referred to (p)ppGpp). This phenomenon is called the stringent response. The accumulation of (p)ppGpp influences the cell physiology through regulating transcription and metabolism to favor cell adaptation. A drop occurs in protein synthesis, DNA replication, cell wall and lipid synthesis. However, transcription for stress response, amino acid biosynthesis, and all that favor cell survival are activated (Potrykus & Cashel 2008; Liu et al. 2015).

In *Bacillus subtilis* and other Gram positive bacteria, (p)ppGpp is synthesized by the RelA enzyme and two small alarmone synthetases (SAS) named YjbM (RelP) and YwaC (RelQ) (Nanamiya et al. 2008; Atkinson et al. 2011). RelA has a (p)ppGpp-synthetase domain and a hydrolase domain (Wendrich & Marahiel 1997). It senses the deprivation of amino acids by the presence of uncharged tRNAs in the A site of ribosomes (Wendrich et al. 2002) [Figure 2.17]. Then, (p)ppGpp regulates transcription of rRNA to decrease ribosomes synthesis. This regulation does not occur by direct interaction with the RNA polymerase as in *E. coli*, but through a drop in the concentration in GTP since (p)ppGpp synthesis of GTP (Kriel et al. 2012). As a consequence, the ratio of GTP/ATP decreases. This decrease modulates the expression of genes controlled by promoters sensitive to the concentration of the initiating nucleotide (Gourse & Krasny 2004). When the (p)ppGpp concentration is increasing, the genes beginning with a 'G' such as rRNA are down-regulated, while the genes beginning by an 'A' are up-regulated (Gourse & Krasny 2004). The drop in GTP levels occurs not only during amino acid starvation but also during fatty acid starvation (Pulschen et al. 2017). Moreover, low levels of GTP leads to the release of the repressor CodY, a GTP-binding protein (Handke et al. 2008). CodY regulates about 200 genes, involved in the adaptation of bacteria to poor media. During exponential phase, CodY is an active repressor (Slack et al. 1993; Stack et al. 1995). It loses its activity, when the cell enters the stationary phase so that all the genes required for adaptation to nutrient limitation are expressed. When GTP level decreases, it results in the upregulation of the CodY-targeted genes which are involved in cell survival during stationary phase such as amino acid biosynthesis, extracellular proteases production, nutrient transport, sporulation, competence, enzyme secretion, and antibiotic production (Geiger & Wolz 2014; Potrykus & Cashel 2008; Barbieri et al. 2015;

Molle et al. 2003). Conversely, when the level of amino acids becomes high, (p)ppGpp is degraded by a hydrolase SpoT (in *E. coli*) and RelA (in *B. subtilis*). As a result, the RNAP then transcribes the genes necessary for bacterial growth (Potrykus & Cashel 2008; Dalebroux & Swanson 2012).

The stringent response was mostly studied during nutrient starvation, as an adaptive mechanism to face nutrient exhaustion. However, the production of (p)ppGpp occurs in all physiological conditions. Actually, the stringent response is the main molecular mechanism that adjusts the ribosome activity to the metabolic network (Kriel et al. 2012; Kriel et al. 2014). Therefore, it is one of the main molecular mechanisms that coordinate resource allocation between the cell processes.

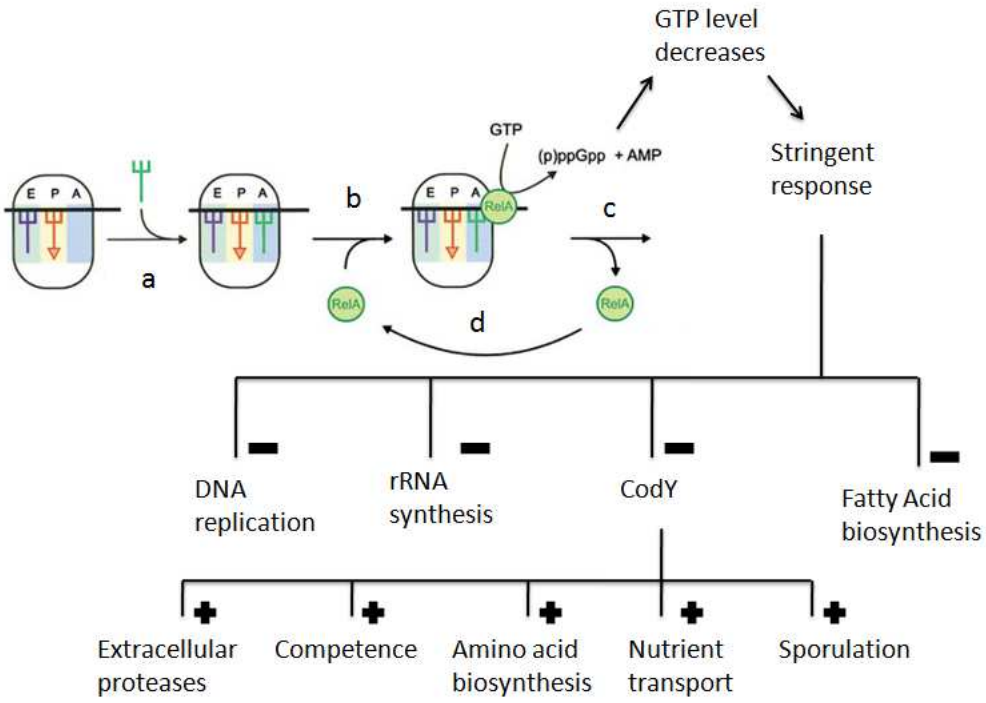

Figure 2.17 The mechanism of (p)ppGpp stringent response

(a) Amino acid starvation generates large pools of uncharged tRNAs. (b) RelA detects a blocked ribosome with uncharged tRNA in the A site (c) RelA mediates the conversion of GTP to (p)ppGpp in the presence of the uncharged tRNA at the A site. (d) RelA continues to the next blocked ribosome, and the synthesis of (p)ppGpp is repeated. When GTP level decreases, it turns on stringent response. The DNA replication, rRNA synthesis and fatty acid biosynthesis decrease, the CodY repressor is released so it leads to the up regulation of amino acid biosynthesis, nutrient transport, sporulation, competence, and extracellular proteases. Adapted from (Wendrich et al. 2002)

2.5.3 The CtsR heat shock response

The third example of molecular mechanisms is the regulation of the quality control system in response to stressful conditions. The CtsR repressor in *B. subtilis* and other Gram

positive bacteria, controls the expression of genes involved in the heat shock stress (Kruger et al. 1994). CtsR (Class three stress repressor) acts on the genes coding for class III heat shock proteins which are the *clpC*, *clpE*,and *clpP* operons (Kruger et al. 1996; Kruger et al. 1997). The *ctsR* gene forms the first gene in the *clpC* tetracistronic operon. It is regulated by the sigma factor σ^B(Kruger et al. 1996). Moreover, it is modulated by two proteins, the MscA and McsB (modulator of CtsR repressor) [Figure 2.18]. At low temperature, CtsR represses expression of genes encoding the proteases involved in the heat shock response, *clpC*, *clpE* and *clpP*. However, upon a temperature upshift, CtsR is released from the DNA binding site. Then, McsB inactivates CtsR by phosphorylation through its kinase activity. Finally, the inactive CtsR is degraded by ClpCP (Kruger et al. 2001) and ClpXP (Derre et al. 2000). In *E. coli*, a similar mechanism occurs but with the alternative sigma factor σ^{32}. Upon a temperature rise, σ^{32} is kept at high concentration to regulate positively the *clp* genes. In normal states, σ^{32} is negatively modulated by the DnaK chaperone and targeted for degradation by the FtsH protease (Tomoyasu et al. 2000).

The regulation network of gene expression in response to the environment is highly tuned. Proteases are modulated in the cell when required to get rid of the misfolded proteins upon stressful conditions. Therefore, we can emphasize again on the cell ability to tune the resource distribution to what is crucial and important to economize protein production and maximize growth rate.

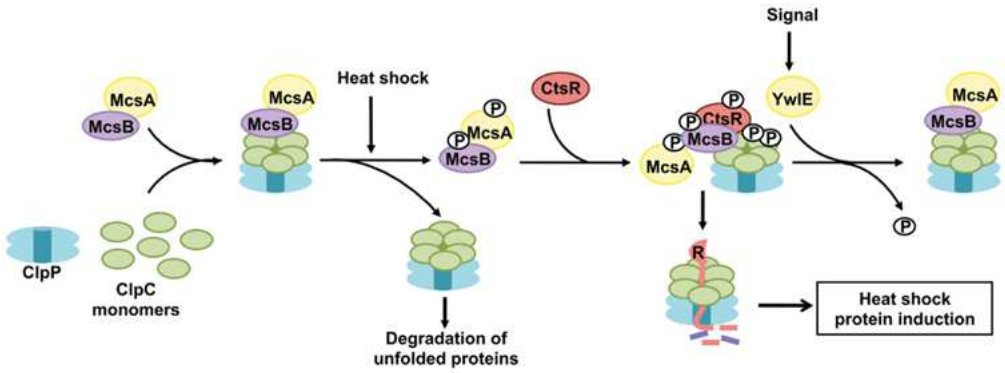

Figure 2.18 Heat shock response in *B. subtilis*

McsA and McsB are adaptor proteins for ClpC, and they modulate CtsR. During heat shock ClpC releases McsA and McsB to degrade misfolded proteins. McsA stimulates the kinase activity of McsB. McsB-P interacts with CtsR leading to its degradation by ClpCP. Through unknown signal, YwlE dephosphorylates McsB to deactivate its action. (Battesti et al. 2017)

Summary:

The phenotype of the cell is obtained through the metabolic and genetic regulatory network. These molecular mechanisms drive resource allocation in the cell in response to environmental conditions. The first example is the selection of a preferred sugar from the medium. In the presence of more than one carbon source, regulatory pathways take place to produce the required proteins for the sugar uptake and metabolism. The carbon catabolite repression system (CCR) represses the uptake and the utilization of a second available sugar. Second, molecular mechanisms take place when bacteria switch from respiration to fermentation. Third, the stringent response is an adaptive mechanism to protect bacteria from nutrient starvation. Transcription and translation are regulated in response to the metabolic network through various pathways. Lastly, the CtsR response regulates the transcription of proteases. They are modulated in the cell when required to get rid of the misfolded proteins upon stressful conditions. Therefore, several different molecular mechanismsdrive the cell to tune the resource distribution to what is crucial and important to economize protein production and maximize growth rate.

2.6 Bottlenecks of protein production in Bacteria

2.6.1 Effect on the growth rate

The tradeoff in resource allocation to the cell processes is of high importance to maintain normal cell growth. However, the attempts in engineering strains to overexpress recombinant proteins cause a disruption in the cell's physiological state. The over-production of a gratuitous protein is not contributing to the cell's growth; in contrast, it occupies an intracellular space and takes resources at the expense of the proteins needed for growth. Consequently, it leads to a burden on the cell and to a growth rate decrease that was experimentally proved. Mathematical models (as those presented in the former chapters) are capable to predict the consequence of the resource management at a given condition. For example, RBA can predict the secretion rate of an overexpressed protein to be decreased with increasing growth rate. In [Figure 2.19], RBA was simulated to compute the secretion rate of the AmyE amylase in *B. subtilis* while including and neglecting the activity of chaperones. What is obvious is that the secretion decreases when including the cost of the chaperones.

Therefore, this illustrates the ability of the bacterium to tune its resources for achieving a balanced distribution that serves all the necessary functions.

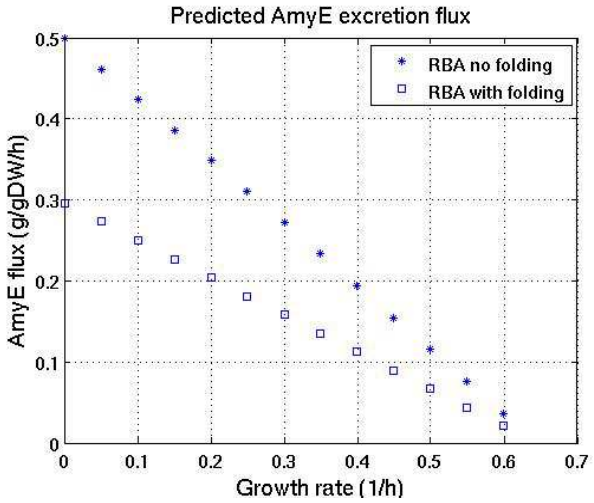

Figure 2.19 RBA simulation to predict the AmyE excretion flux when including and neglecting the chaperon cost.

When neglecting the chaperon cost, the AmyE flux is predicted to be higher than that when including the chaperon cost.

From an experimental point of view, the attempts on studying the cell physiology in response to high protein production go back to many years of studies (Bertrand & Lenski 1989; Dong et al. 1995). Dong et al. showed growth inhibition when 30% of normal proteins in *E. coli* were replaced by a gratuitous protein [Figure 2.20] (Dong et al. 1995). More recently, the effect was shown on baking yeast by integrating 20 copies of the *gfp* and *mCherry* genes into the chromosome (Kafri, Metzl-Raz, et al. 2016). However, in *B. subtilis* the effect of protein overproduction was not shown yet, as researchers in this community focused more on the secretion bottlenecks.

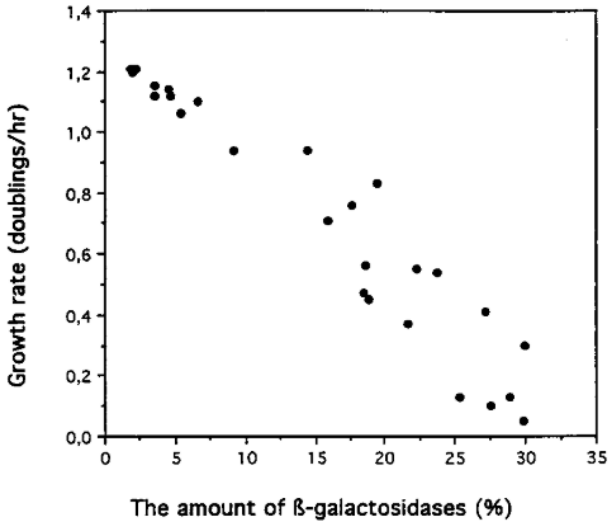

Figure 2.20 Gratuitous protein effect shown on *E. coli*

The growth rate reduction in *E. coli* when the amount of LacZ protein expression reached 30% (Dong et al. 1995).

2.6.2 Effect on the protein folding

Overexpression of a heterologous protein is often accompanied by the formation of protein aggregates. This means that the produced protein becomes insoluble and loses its function, so it forms what is called inclusion bodies (Georgiou & Valax 1996). In *E. coli* for example, inclusion bodies are formed in the cytosol due to changes in the original environment of the produced protein such as pH, osmolarity, cofactors, and improper folding (Carrio & Villaverde 2002; Hartley & Kane 1986). The same challenge is faced in *B. subtilis* due to a limited amount in chaperones (Wu et al. 1998), so that Wu et al. (1998) worked on increasing the production of chaperones to avoid protein aggregates. The chaperones overproduction resulted in less protein aggregates and higher protein secretion. In general, eukaryotic or mammalian proteins need post-translational modifications that can not be performed in bacteria (Villaverde & Carrió 2003). Hence, the absence of post-translational modifications leads as well to misfolded and aggregated proteins, which consequently upregulates the stress-related cell responses.

2.6.3 Effect on the cell size

The effect of gratuitous protein does not only appear on the growth rate, but also on the cell size as well. The limitation of resources is accompanied by the limitation of intracellular cell density. This later is known to be constant in *E. coli* and *B. subtilis* (Marr 1991; Vazquez et al. 2008; Goelzer et al. 2011; Kubitschek 1984; Kubitschek et al. 1983). It is the relation between the mass and the volume, so when the macromolecular mass increases the volume increases as well to keep a constant intracellular density to allow the stable diffusion of the cellular components within the cytosol. The relation between the cell size and protein production was observed on *E. coli* (Basan et al. 2015) and on *S. cerevisiae* (Kafri, Metzl-Raz, et al. 2016). However, they did not demonstrate an experimental explanation for the reason behind the cell size increase. In *E. coli*, Basan et al. (2015) correlated the change in size with the growth rate when overproducing LacZ. The authors also characterized the dry cell mass and the change in the macromolecular components of the cell with the growth rate. Then, they built the threshold initiation model based on the idea that the amount of proteins required for cell division (X-fraction) has to reach a fixed threshold for the cell division to occur. Thus, the production of an additional protein (LacZ) results in a decrease in the X- and G- protein fractions. Consequently, the cell has the chance to grow longer until the X-fraction

reaches the required threshold [Figure 2.21]. In the budding yeast, Kafri et al (2016) observed an increase of 40% in cell size when they added 20 copies of *mCherry* in the genome. Kafri and coworkers explained that the increase in volume is consistent with the estimated fraction of added proteins, and so the increase in size is attributed to the extra proteins produced.

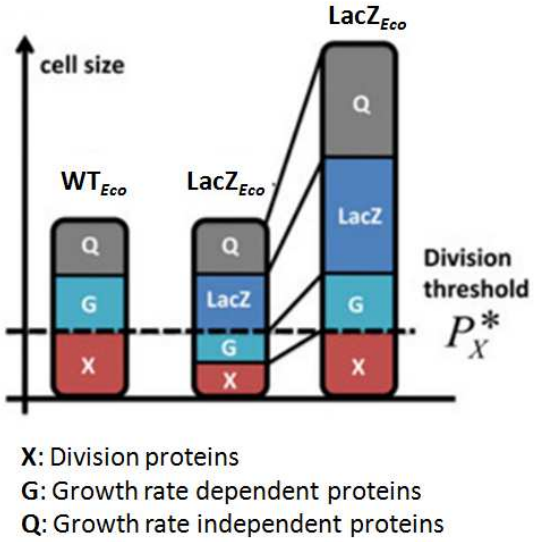

Figure 2.21 Scheme of the cell's response to high protein production

The threshold initiation model of cell size control proposed by (Basan et al. 2015). Q is the housekeeping proteins, G is the protein fraction that is growth rate dependent, and X is the protein fraction responsible for cell division. When LacZ is overproduced, it decreases the fractions of the cell division proteins (X), thus the cell size keeps on growing until X reaches the quantity required for cell division.

Cell division is mediated by the polymerization of the FtsZ ring at the midcell (Bi & Lutkenhaus 1991). FtsZ requires GTPase activity and is accompanied by regulatory proteins, which help the ring assembly in a controlled manner. Positive regulators aid in the assembly and stabilization of the FtsZ ring such as ZapA in *B. subtilis* and *E. coli*, and ZipA in *E. coli*. However, negative regulators prevent the ring assembly in wrong sites and maintain the dynamic assembly in the midcell such as EzrA in *B. subtilis* and the Min system in *B. subtilis* and *E. coli* (Anderson et al. 2004; Levin PA et al. 1999; Huang & Durand-heredia 2013). In addition, cellular levels of FtsZ were found to remain constant during the whole cell cycle in both *B. subtilis* and *E. coli* and were growth-rate independent, but the frequency of the assembly dynamics changed with the growth rate (Huang & Durand-heredia 2013; Weart & Levin 2003). Thus, it is not the FtsZ concentration that determines the polymerization of the ring, but it is rather cell cycle signals, which induces the ring formation, or released inhibitors

from FtsZ. For instance, down below are two molecular mechanisms presented that drive the cell division control, the response to the glucose availability in the medium, and the response to nutrient availability through the ppGpp stringent regulator.

The presence of UDP-glucose in the medium plays a role in cell division and thus in the determination of cell size. UDP-glucose is a metabolic product generated from glucose-6-phosphate in the glycolysis. It is the substrate (metabolic signal) of glycosyltransferase UgtP in *B. subtilis* and OpgH in *E. coli*. Binding to their substrate, they interact with FtsZ thus inhibiting the formation of FtsZ ring and result in elongated cells. However, in the lack of UDP-glucose (nutrient poor medium), the UgtP forms clusters away from the division ring formation so FtsZ is activated. Therefore, the cell divides at smaller size. Levin et al. (2016) suggested that the cell size is coordinated with the growth rate through nutrient dependent changes (Levin et al. 2016; Weart et al. 2007) [Figure 2.22].

The mediators to the stringent response that is the (p)ppGpp accumulation at low growth rate lead to the decrease in GTP levels, which in turn inhibits the transcription from rRNA promoters (Gourse & Kra 2004). A (p)ppGpp increase leads to the decrease of lipid synthesis by the inhibition of fabHDG in *E.coli*, and the ppGpp role here is to control the lipid synthesis in relation to the cytoplasmic metabolic state in response to nutritional availability to maintain the cell envelope integrity (Vadia et al. 2018).

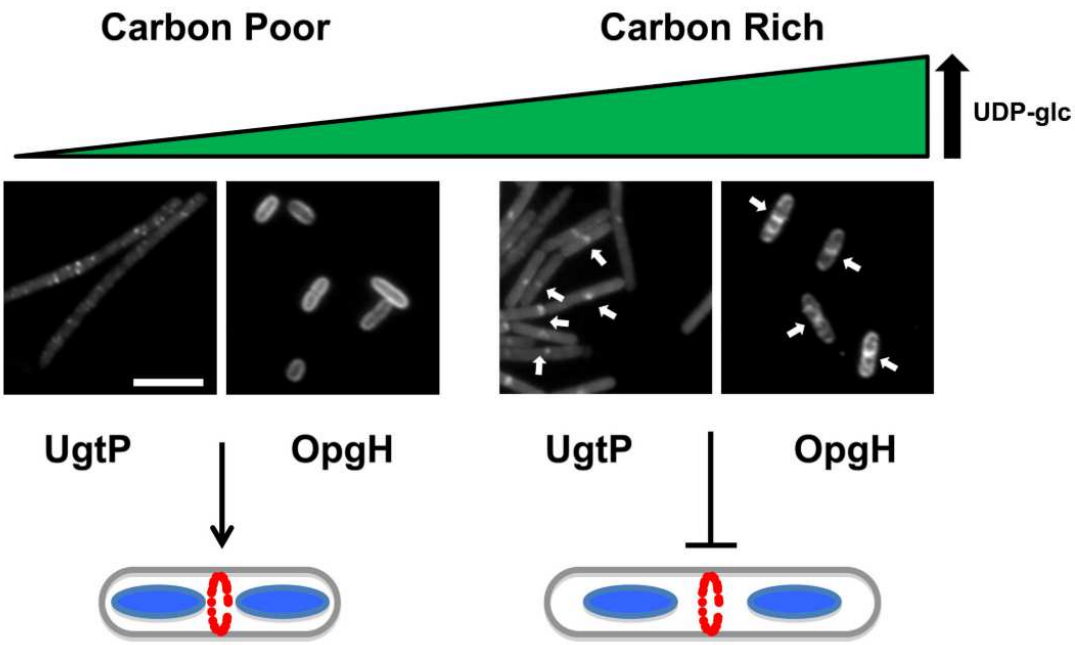

Figure 2.22 UDP-glc availability affects the localization of the cell division inhibitors

In nutrient poor medium when UDP-glc is at low level, the enzymes UgtP (of *B. subtilis*) is distributed in the cytoplasm at low concentrations, and the OpgH (of *E. coli*) is distributed at the cell periphery, allowing FtsZ ring formation and cell division. In nutrient rich medium, when UDP-glc is at high amounts, the enzymes concentrate at the mid of the cell inhibiting the FtsZ ring formation. (Levin et al. 2016)

2.7 Explanations around the gratuitous protein effect

The cost of recombinant protein production is intensively studied in microorganisms. There are many challenges and bottlenecks that were shown and the investigations to find out the explanations of these bottlenecks were presented in different studies. The main and general description of the cost of high protein production is a result of the competition on the resources between the essential proteins and the useless over-expressed protein. The regulation of gene expression in bacteria shows that microorganisms tend to optimize the expression in response to the environment and their needs. Right now, the focus will be on the main assumptions that were given on the limitations causing the consequences of the gratuitous protein production.

2.7.1 Ribosomes constraint

Ribosomes directly affect protein synthesis so they are considered as indispensable resources for the cell. The regulation of their synthesis in response to the growth conditions is a part of the mechanisms that economize the resources. Phenomenological models suggested

that the over-produced protein leads to decreased levels of ribosomal protein fractions and of the growth rate-dependent protein fractions (Weiße et al. 2015; Scott et al. 2010).

2.7.2 RNA Polymerase constraint

Ribosomes and RNA polymerase are crucial for gene expression. In addition, their activity and amount are growth rate-dependent. Therefore, it was suggested that the limitation in RNA polymerase forms a limitation when over-producing a gratuitous protein. When heterologous genes are added in multicopy plasmids or in multiple integrations in the chromosomes, a competition of RNA polymerases occurs. As a result, it happens to be a limiting resource (Gyorgy et al. 2015).

2.7.3 Translation and transcription processes

Some studies considered that the protein burden is caused by the process of producing it. Transcription of the heterologous gene results in nucleotides and RNA polymerase wasting. On the other hand, translation limits *free* ribosomes, tRNAs, and wastes amino acids to produce a useless protein. Therefore, the process of production by itself is more costly than the product (Stoebel et al. 2008). Each of these processes may become a limitation depending on the growth conditions (Kafri, Metzl-Raz, et al. 2016). Therefore, the relative importance of these processes is altered according to the growth medium quality.

2.7.4 Quality control system constraint

The quality control system composed of proteases and chaperones was suggested to form a limitation for gratuitous protein production (Cookson et al. 2011; Daraba & Alma 2018). The limitation in the proteases and chaperones results in the accumulation of misfolded and aggregated proteins. This results in the induction of the stress response. When the cell faces starvation, the malfunctional proteins queue for ClpXP protease (Cookson et al. 2011). The accumulation of misfolded proteins in this case causes the ClpXP to be charged and limited, which leads to the buildup of sigma factor σ^S that is targeted by ClpXP in *E. coli*. The accumulation of σ^S then, leads to the stress response. The limitation of chaperones as well proved to be a bottleneck for high protein production. In some studies they increased the production of chaperones to aid in the folding of heterologous proteins; mainly eukaryotic proteins. In *E. coli*, the co-expression of GroEL/GroES resulted in better production of rhIFN-γ protein; a human interferon, and even a better growth due to the enhanced protein folding and reduction in toxicity (Yan et al. 2012). In *B. subtilis*, Yan and coworkers co-expressed intracellular chaperones moderately for not causing a growth burden (GroEL, GroES, DnaK).

It resulted in a 2-fold decreased protein aggregates (Wu et al. 1998). The authors of Wu et al. co-expressed as well the extracytoplasmic chaperons PrsA. They noticed an increase in a soluble protein in the cytosolic and secreted fractions.

These studies indicate that there is not only one 'cost' in overproducing gratuitous protein, but a combination of all the above-mentioned bottlenecks. The availability of all the macromolecules (RNA polymerase, ribosomes, metabolic enzymes, chaperons *etc*.) in sufficient amounts to do their proper function is essential to overcome a growth burden. Adding to that the importance of the availability of metabolites, energy precursors, and all kind of cell resources that have to supply the macromolecules for their proper function. The tradeoff in the resource allocation to serve cell processes is the key to have a normal growth. When a heterologous protein is added, the resource supply is re-allocated to produce it besides all the other essential proteins [Figure 2.23]. The burden occurs when the machinery (translation, transcription, quality control system, *etc*.) is saturated leading to the accumulation of proteins in the cytosol.

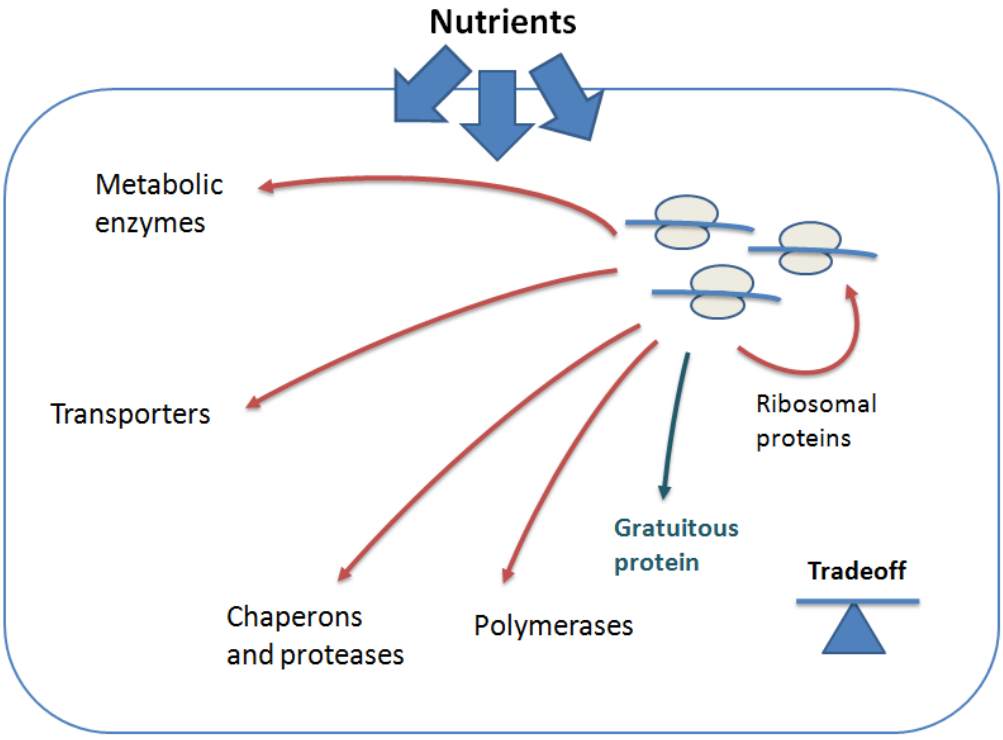

Figure 2.23 Illustration of the tradeoff between the different proteins produced in the cell

Each protein requires cellular resources and has a task to fulfill so that the cell grows at a normal growth rate.

Summary:

The over-production of a gratuitous protein results in a competition with the essential proteins for the limited cell resources. As a result, it disrupts the cell's physiology and results in a burden. The consequences on the cell can be shown on the growth rate, on the protein folding, and on the cell size. Different studies investigated on knowing the bottleneck of gratuitous protein. Some of which utilized mathematical interpretations, and others were experimental. The suggested bottlenecks were the limitation in resources like ribosomes, RNA polymerases, proteases and chaperons, or more generally in the process of protein production.

CHAPTER 3.

Quality control system

3 Quality control system

In this chapter, the focus will be on the chaperone-protease system in *B. subtilis* that is known as the quality control system. Its role in the cell occurs during translation (co-translationally) or during the post-translational processes of protein production. It ensures that the protein will be well folded and functional, otherwise it will be sent for proteolysis. In addition, it plays an important role in the stress response of the cell. The protein quality control system forms a part of the cell machinery resources. Their role in maintaining the proteins in good structure is required. In this chapter, we will present an overview of the proteases and chaperones in the cell. How are they regulated to equilibrate the cell resources? What are their targets? And, how can they be used to uncover the limitations of recombinant protein production?

3.1 Overview on the quality control system

During a normal growth, the proteases and chaperones are present moderately in the cytoplasm. They target the mutated, non-functional, misfolded and, truncated proteins. Some said that during early stationary phase the cellular levels of some proteases increases more than the exponential phase; mainly ClpAP in *E. coli* (Farrell et al. 2005). The same thing was also shown in *B. subtilis.* Under non-stress conditions the proteins were stable during exponential phase, then the degradation rate increased upon entry into the stationary phase (Kock et al. 2004). But in general, when the cell faces stressful conditions such as a temperature rise, it efficiently deals with the protein aggregates and protein folding by synthesizing molecular chaperons and proteases through regulated mechanisms. Indeed, *B. subtilis* is a soil bacterium well adapted to environmental changes. It developed systems to cope with environmental stresses differently than *E. coli.* For instance, while *B. subtilis* can survive temperature over 50°C, *E. coli* cannot. This means that although they have much of similarities, the quality control system and its regulation in *B. subtilis* differs from that of *E. coli* (Vdlker et al. 1994).

B. subtilis and *E. coli* share different classes of chaperons such as GroEL/GroES, DnaK/DnaJ/GrpE, trigger factor, Hsp90/HtpG, and proteases such as the AAA^+ Clp ATPases/Hsp100 family. This family has an unfoldase activity and a protease activity. The unfoldase activity is carried on by the Clp ATPases, while the protease activity is carried on by the protein ClpP, common to all Clp ATPases. Each member of the AAA^+ Clp ATPases/Hsp100 family selects its targets with high specificity through the Clp ATPases.

Summary:

Proteases and chaperons form the quality control system in the cell. They are responsible to target misfolded and mal functional proteins. In addition to their role in stress conditions, they are produced in regulated mechanisms to target protein aggregates. *B. subtilis* and *E. coli* share many molecular chaperons and proteases such as GroEL/GroES, DnaK/DnaJ/GrpE, trigger factor, Hsp90/HtpG, and proteases such as the AAA^+ Clp ATPases/Hsp100 family. However, the quality control system and its regulation in *B. subtilis* differ than *E. coli.*

3.2 ATP dependent proteases

3.2.1 Family and Structure

ATP dependent proteasesbelongs to the AAA+ superfamily (ATPase associated with diverse cellular activities), whose function requires ATP. They carry different cellular regulations through proteolysis such as targeting transcription factors. Thus, they have a role in the cell cycle regulation, cellular development and adaptation (Sauer et al. 2004; Frees et al. 2007; Gottesman 2003). One family of the AAA^+ superfamily is the Hsp100/Clp proteases. Each member of this family is a complex of an ATPase chaperone component (AAA domain or the NBD domain for nucleotide binding domain) whose role is to unfold the protein and a protease component responsible for degrading the protein into small peptides of (6-15 amino acids) (Gottesman 1996). The AAA domains such as in ClpX, ClpC, and ClpE are hexameric rings with a hole in the center to allow the translocation and the unfolding of the protein by using ATP. They have several conserved motifs including those responsible for ATP binding and hydrolysis. Moreover, they are divided into two classes, class I contains two highly conserved AAA core domain and includes ClpE, ClpC (in *B. subtilis*), ClpA, and ClpB (in *E. coli*), and class II contains only one conserved AAA core domain and includes ClpX and ClpY. In addition, there is a less conserved N-terminal domain (N-domain). The N-terminal domain modulates the binding with the substrates or the adaptor proteins and thus provides the specificity of the Hsp100 chaperons and proteases [reviewed in (Dougan et al. 2002)]. Moreover, the active sites of the ClpP complex are situated inwards, thus avoiding the degradation of unwanted proteins (Molière & Turgay 2009; Baker & Sauer 2006). Besides, there are other members of the AAA+ proteases which have both activities, the unfoldase and

the protease, in the same polypeptide chain or in one hexameric subunit such as Lon and FtsH proteases (Baker & Sauer 2006).

3.2.2 Members of the AAA+ proteases and their targets

In *B. subtilis* there are seven ATP dependent proteases, ClpCP, ClpEP, ClpXP, ClpYQ, Lon, and FtsH (Chan et al. 2014) [Figure 3.1]. The Hsp100/Clp are implicated in the heat shock response in *B. subtilis* and others. Each member of the Clp ATPases recognize specific targets and thus has a specific role. In what follows, a summary of the different members of the AAA+ family and their targets is presented.

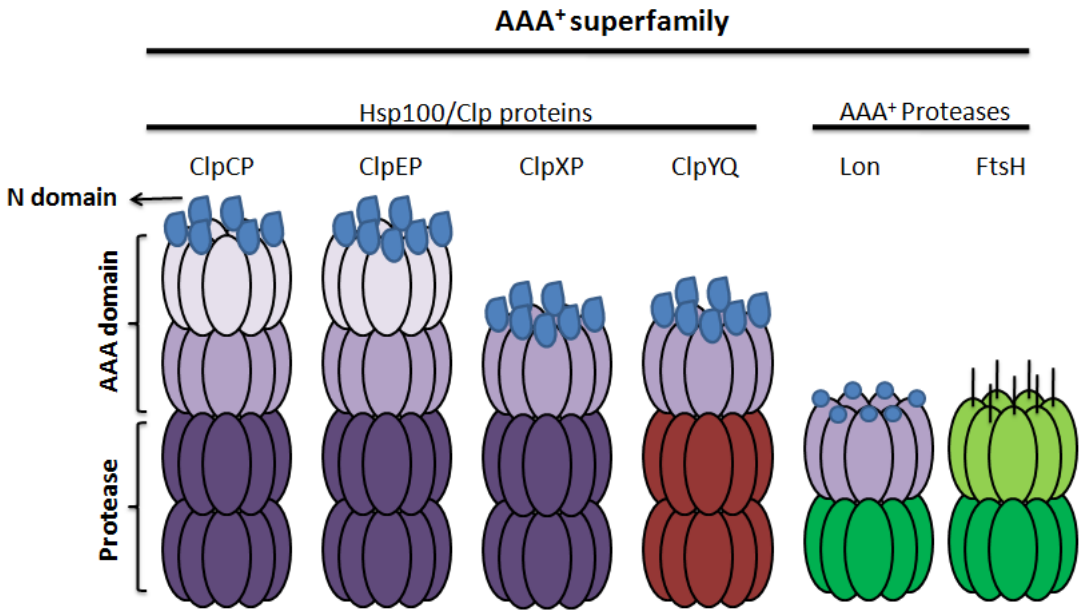

Figure 3.1A scheme for the AAA$^+$ superfamily

It shows the members of the 2 subfamilies (Hsp100/Clp and AAA$^+$ proteases) and their structure. The protease domain is the bottom part, the AAA domain is the upper part which could be 2 core domains (class I) or 1 core domain (class II), and the N domain which modulates the substrate or adaptor protein binding. Adapted from (Elsholz et al. 2017)

- ClpCP is involved in the controlled degradation of regulatory proteins such as the transcription factor ComK that is required for the expression of competence genes (Kursad Turgay et al. 1998), the CtsR which is a heat shock repressor (Kruger et al. 2001; Kirstein et al. 2007), the SpoIIAB an anti-anti sigma factor involved in sporulation(Pan et al. 2001), and the SlrR a regulator of biofilm formation (Yunrong Chai, Roberto Kolter 2010). In addition, ClpCP targets misfolded proteins so that its

absence results in their accumulation (Krüger et al. 2000). ClpCP degrades enzymes in the central carbon metabolism in response to poor growth conditions (Gerth et al. 2008).

- ClpEP is highly involved in the heat shock response, so that *clpE* mutation affects ClpP expression and the regulation by CtsR (Miethke et al. 2006).
- ClpXP is involved in heat, thiol and oxidative stress. It targets Spx for degradation which is a global transcriptional regulator that serves in thiol homeostasis and competence (Chan et al. 2014), and it inhibits swimming motility upon exposure to heat and oxidative stress (Molière et al. 2016). ClpX interacts with the adaptor protein YjbH to degrade Spx in normal conditions. In contrary, under stress conditions YjbH form aggregates and becomes inactive which leads to the increase of Spx levels (Engman & Wachenfeldt 2015). In addition ClpXP ensures proper spore envelop formation. When the cell has defects in the spore envelop maturation, the adaptor protein CmpA interacts with ClpXP to degrade the coat morphogenetic protein SpoIVA leading to spore lysis (Irene S. Tan 2016). Finally, ClpXP is involved as well in non-stress conditions such as degrading truncated proteins [more details in section 3.4].
- ClpYQ contains the protease activity in the ClpQ part and the ATPase activity in ClpY part. ClpYQ regulates multicellular development. Its deletion caused robust and early biofilm formation, impaired swarming and swimming, but was not important for the heat shock response (Yu et al. 2018).
- Lon protease is implicated in the regulation of swarming and swimming depending on the surface. When the surface is liquid, Lon protease prevents cell swarming by degrading SwrA, the master activator of flagella biosynthesis. However, Lon activity is inhibited when the surface is solid, and thus SwrA accumulation allows the synthesis of the flagella (Mukherjee et al. 2015).
- FtsH is a cytoplasmic protease responsible for the membrane protein quality, cell division heat shock response, and biofilm formation and sporulation (Deuerling et al. 1997).

3.2.3 Substrate recognition

Protein degradation is achieved by a diverse set of peptide signals which act as degradation tags (degrons) allowing the recognition and engagement of the substrate to the ATPase [Figure 3.2], or by attaching the substrate to the ATPase, thus increasing the

substrate's concentration and the probability of degradation (Sauer et al. 2004). The degrons can be located on the N-terminal of the substrate such as the N degron, the C-terminal such as the ssrA tag, or in an internal region (Kirstein et al. 2009). The location of the peptide signals can be the clue to allow a specific degradation by the specific ATPase. For instance, the ssrA tag; a well-known peptide signal in *E. coli* and *B. subtilis*, is added to the C-terminus of the substrate to be degraded by ClpXP in *B. subtilis* and ClpXP/ClpAP in *E. coli*. Another example is the N-degron generated on the amino terminus of the substrate is targeted by ClpAP (Dougan et al. 2010).

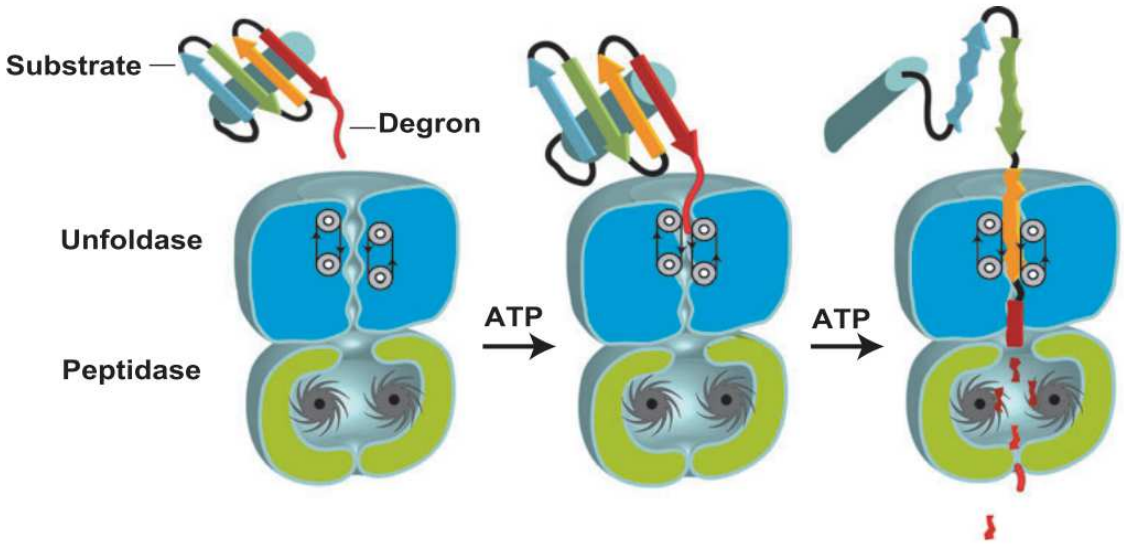

Figure 3.2 Representation of the mechanism of degradation

The substrate is recognized by a degron (peptide signal), then it passes through the unfolding process in the ATPase part, and then it is degraded by the protease part in an ATP-dependent manner [Taken from (Sauer & Baker 2011)]

3.2.4 Adaptor proteins

The specificity of degradation is mediated in some cases through adaptor proteins besides the recognition tags [Figure 3.3]. The function of the adaptor protein is to aid the ATPases to recognize their wide spectrum of different substrates. They are relatively small proteins (Dougan et al. 2002). They recognize binding regions on the substrate's recognition tag and they have a binding site to anchor on the ATPases. In *B. subtilis* there are 5 known adaptor proteins; MecA,YpbH and McsB which interact with ClpC, YjbH which interacts with ClpX, and SmiA which interacts with Lon (Mukherjee et al. 2015; Kirstein et al. 2009).

- MecA functions in the competence development in *B. subtilis.* In addition, it plays a role in targeting misfolded and aggregated proteins for degradation. In non-competent

states, MecA recognizes ComK; the master competent regulator,and mediates its degradation by ClpC (Schlothauer et al. 2003). When the cell develops competence state, ComS is synthesized which then binds to MecA, thus inhibiting the interaction with ComK. The release of ComK from the proteolysis complex contributes to its accumulation, thus turning on the expression of the competence-related genes (Kursad Turgay et al. 1998).

- YpbH is a paralogue of MecA, which suggests that they may have similar interactions with ClpC. It contributes to the protein quality control by enabling the degradation of misfolded and aggregated proteins by ClpC. YpbH has a role in competence and sporulation, but it does not contribute to the degradation of ComK or ComS (Persuh et al. 2002).
- McsB plays a role in the regulation of class III heat shock genes. It aids ClpCP to degrade the CtsR repressor. McsB is a protein kinase, which can auto-phosphorylate and is activated by McsA. During heat shock, activated (phosphorylated) McsB-P phosphorylates CtsR. Then, the inactive CtsR-P is degraded by ClpCP, and thus leading to the up regulation of the heat shock response genes (Elsholz AKW et al. 2011).
- YjbH is an adaptor protein that interacts with ClpXP. It recognizes the transcriptional regulator Spx and facilitate its degradation under normal conditions. However, under stress conditions, a small protein named YirB directly interacts with YjbH, thus inhibiting its action. This results in Spx stabilization and protection from degradation (Kommineni et al. 2011).
- CmpA is an adaptor that interacts with ClpXP to degrade SpoIVA. This later is required for the assembly of the spore envelop (Mckenney et al. 2010). When the spore envelop misassembles, SpoIVA is targeted for degradation. Therefore, CmpA forms part of the quality control system that ensures the assembly of proper spore envelope (Irene S. Tan 2016).
- SmiA is a swarmer inhibitor, which helps Lon protease to regulate the switch between swarming and swimming. When the cell is in a liquid environment, Lon protease degrades SwrA; the master activator of flagella biosynthesis, by the presence of SmiA (Mukherjee et al. 2015).

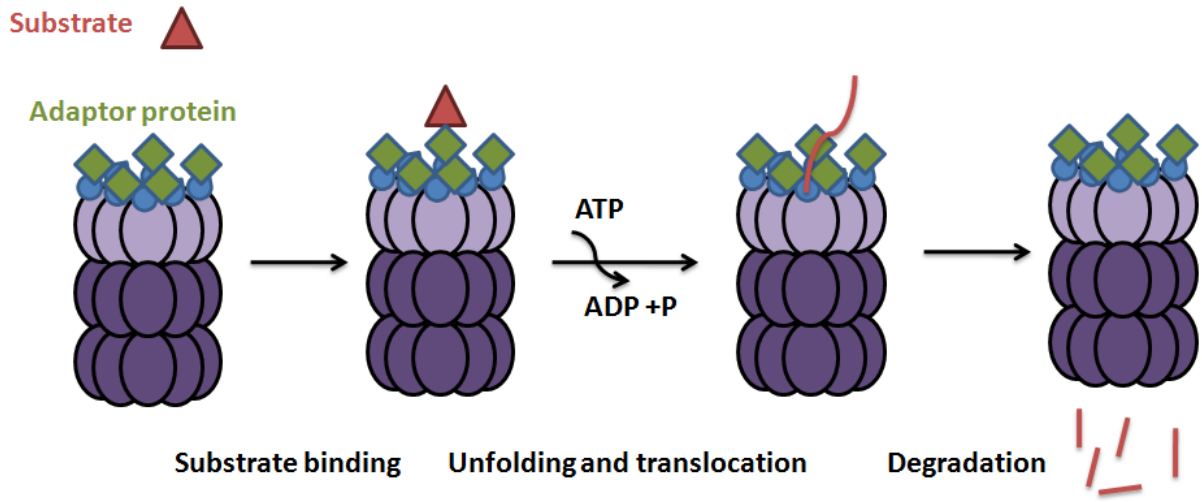

Figure 3.3 The process of ATP-dependent proteolysis of a substrate recognition by means of the adaptor protein

Summary:

ATP dependent proteases are referred to the AAA+ superfamily. They contain the Hsp100/Clp proteases and the AAA+ proteases. The first has a chaperon subunit (the Clp ATPases) which recognizes the protein and is responsible for protein unfolding. It has also a protease subunit responsible for degrading the protein into small peptides. However, the AAA+ proteases (FtsH and Lon) have both roles in one polypeptide. In *B. subtilis* there are seven ATP dependent proteases, ClpCP, ClpEP, ClpXP, ClpYQ, Lon, and FtsH. They have different roles in cell cycle regulation, cellular development and adaptation. In addition they are very specific in their substrates' recognition. Their specificity is helped by peptide signals present on the C- or N- terminus of the targeted proteins, and by different adaptors specific for each protease.

3.3 Regulation of proteolysis

The expression of the protein quality control system in *B. subtilis* costs resources (proteins, energy, cytosolic space, etc.). Like for other cellular processes, regulatory mechanisms are there to produce the appropriate level of the protein quality control system in agreement with the needs of the cell, in response to environmental conditions. It is important to point that the regulation in *B. subtilis* differs from that of *E. coli*. In this later organism, all

the heat shock genes are controlled by the sigma factor σ^{32} encoded by the *rpoH* gene which upregulate the expression of the related genes at a temperature up shift between 30°C to 42°C [reviewed in (Tomoyasu et al. 2000)]. However, *B. subtilis*, which can tolerate temperature up to 50°C, has different regulatory pathways and mechanisms controlling several classes of heat shock proteins. The most known classes will be presented in the following paragraphs.

The first class of the heat shock genes is composed of *groE* (*groEL*, and*groES)* and *dnaK* operon (*dnaK, dnaJ*, and *grpE*). They are under the control of the vegetative promoter σ^{A} (σ^{43}) and a 9-base pair inverted repeat CIRCE separated by a highly conserved 9-base pair spacer (Zuber & Schumann 1994). CIRCE is situated between the transcriptional and the translational sites in both operons. The first gene of the *dnaK* operon is the *hrcA*, which codes for the transcriptional repressor HrcA. In normal conditions, HrcA interacts with the CIRCE element, thus inhibiting the expression of both operons. Upon a temperature rise, HrcA changes its conformation then dissociates from the two CRICE elements. Consequently, the transcription of both operons turns on (Schulz & Schumann 1996). HrcA becomes active again when interacting with GroESL chaperons. When denatured proteins are refolded or degraded, more free GroESL become available to refold HrcA (Mogk et al. 1997).

The second class is under the control of sigma factor σ^{B} and it is known as the general stress response. The σ^{B}regulon contains the heat shock genes and additional genes related to other kind of stresses such as salt, oxidation, acid, starvation stresses, *etc*. It is said that the σ^{B} regulon is the largest among the different classes, consisting of 127 members (Price & Fawcett 2001). Under normal conditions, σ^{B} is regulated by an anti-sigma factor encoded by the *rsbW* gene so that it cannot interact with the RNA polymerase. RsbW has a kinase activity. It phosphorylates through its kinase activity an anti-anti-sigma factor RsbV to inactivate it, so that RsbW stays active and capable to inhibit σ^{B}. In response to an environmental stress factor, the RsbU phosphatase is activated, so it removes the phosphate from RsbV. Then, RsbV becomes free to attack the RsbW-σ^{B} complex. Finally, σ^{B} is released and can upregulate its corresponding regulon [Figure 3.4] [reviewed in (Schumann 2003)].

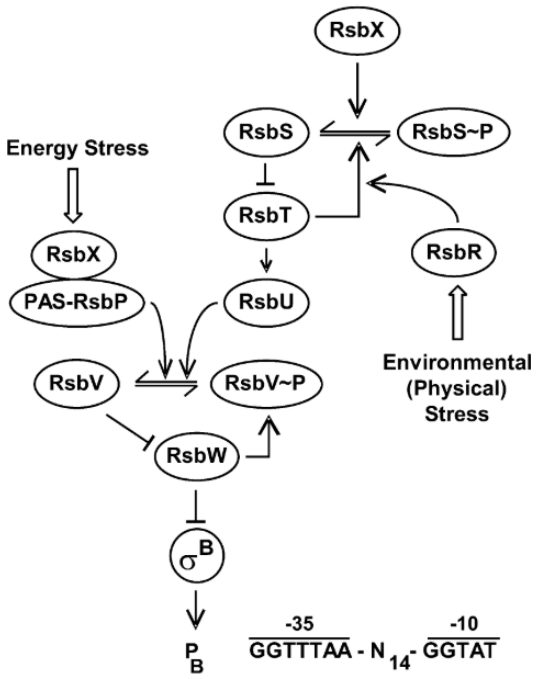

Figure 3.4 Sigma B regulation in stress conditions

Under normal conditions, most σ^B molecules are bound to anti-sigma factor RsbW. RsbW is a kinase that phosphorylates the anti-antisigma factor RsbV to keep it inactive. Energy stress leads to the dephosphorylation of the anti- antisigma factor RsbV by the RsbP phosphatase, whereas environmental (physical) stress activates the RsbU phosphatase. Dephosphorylated RsbV targets the RsbW- σ^B complex. Then thesigma factor is released. Finally, σ^B transcribes all the genes of the σ^B regulon.(Schumann 2003)

The third class of heat shock genes is the CtsR regulon. CtsR regulates the expression of some of the Hsp100/Clp proteins, as discussed previously: the *clpC* operon (*ctsR, mcsA, mcsB, clpC*) and the monocistronic *clpP* and *clpE* operons [Figure 3.5]. In addition to the abovementioned genes, the authors of (Hecker 1996) proposed that *clpX* is classified in class III since it is transcribed from a σ^A-dependent promoter and it has neither a CIRCE element nor a σ^B-dependent promoter. The CtsR is thus the repressor that inhibits the expression of the protease complexes. McsA and McsB are modulators of CtsR activity. The 3 operons are preceded by an upstream σ^B and downstream σ^A dependent promoters. The CtsR has a DNA binding site domain which interacts with the transcriptional start site of the operons, thus preventing the RNA polymerase binding to the σ^A dependent promoter site (Kruger et al. 1996; Kruger et al. 1997). CtsR has a conserved tetraglycine loop that senses temperature change. When the temperature increases, the conformation of the loop changes, and thus the DNA binding is impaired. As a result, the genes controlled by CtsR are induced (Elsholz et al. 2010; Fuhrmann 2009). In response to heat stress as well, McsB detaches from ClpC, then becomes active as a protein kinase (Kirstein et al. 2005; Alexander K W Elsholz et al. 2011).

Consequently, McsB phosphorylates CtsR to inactivate it (Elsholz AKW et al. 2011). Then, CtsR is targeted for degradation by ClpCP [Figure 3.5] (Derre et al. 2000; Kruger et al. 2001; Derre et al. 1999). Moreover, McsB is inactivated by the phosphatase YwlE. This phosphatase has a cysteine residue that is sensitive to oxidative damage. In this case, the YwlE loses its phosphatase activity (Fuhrmann et al. 2017). In addition, McsA, which usually attaches to McsB in normal conditions, becomes oxidated. As a consequence, it releases McsB. Consequently, McsB regains its kinase activity resulting in CtsR inactivation and degradation. In parallel, native CtsR is also targeted by the ClpXP protease. This means that CtsR must be maintained at a required level to keep the basal expression level of the operons.

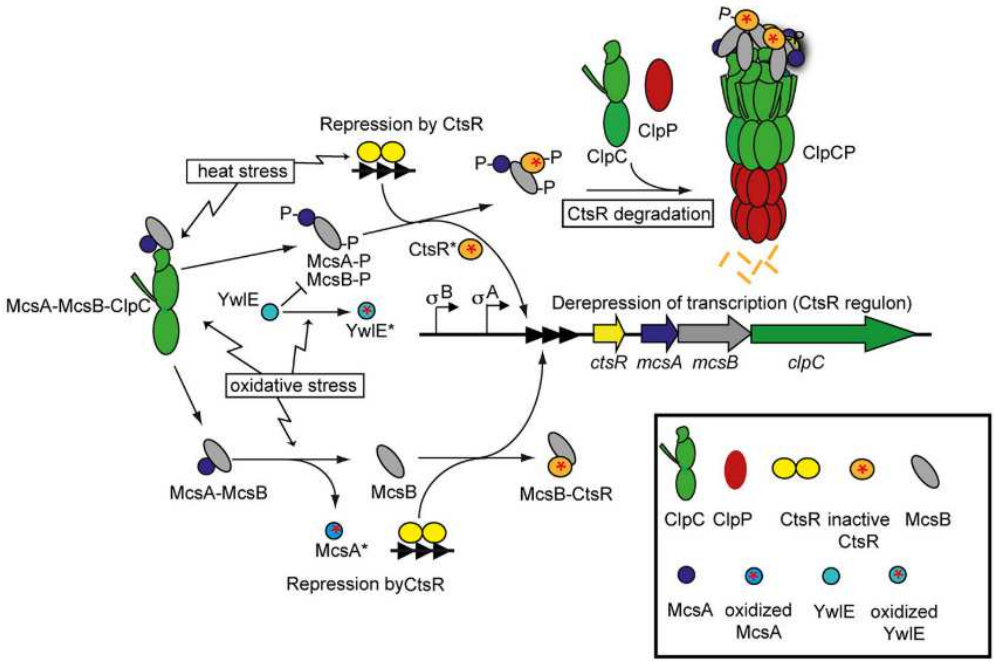

Figure 3.5 CtsR regulation under different environmental conditions

In normal conditions, McsB is dephosphorylated by YwlE, and bind to ClpC. Upon heat exposure, CtsR change conformation and is released from DNA. McsB is released in response to heat, and regains its kinase activity. McsB-P phosphorylate CtsR. Then, CtsR is targeted to degradation by ClpCP. During thiol-reactive stress conditions, YwlE is inactivated by oxidation (YwlE*). McsA is also oxidized (McsA*), to it is released from McsB. Then, McsB becomes active and free to inactivated CtsR (Elsholz et al. 2017).

What is evident in the preceding information is the high level of regulation and specificity in proteolysis. There are classes of chaperone-protease complexes which share conserved motifs for ATP binding and hydrolysis, and mechanisms of regulation. Nevertheless, each chaperone-protease complex has its own targets and adaptor proteins determined by specific domains. The targeted degradation deduced from the knowledge about the system made it

applicable in synthetic biology as tools to finely tune the degradation of specific heterologous or endogenous proteins.

Summary:

In response to environmental conditions, regulatory pathways are developed to express the necessary genes and turn off others. There are different regulatory pathways and mechanisms controlling several classes of heat shock proteins. The first class of the heat shock genes is composed of *groE* (*groEL,* and *groES)* and *dnaK* operon (*dnaK, dnaJ*, and *grpE*). They are under the control of the vegetative promoter σ^A and HrcA that binds the inverted repeat CIRCE. The second class is the largest consisting of 127 members and it is under the control of sigma factor σ^B. The third class of heat shock genes is the CtsR regulon. It controls some of the Hsp100/Clp proteins, the ClpE, ClpX and ClpP.

3.4 Targeted protein degradation in synthetic biology

One role of the ClpXP protease is to target truncated proteins. This mechanism happens in *E. coli* and *B. subtilis*. When a ribosome reaches the 3' end of mRNA without finding a stop codon a tmRNA (transfer messenger RNA) charged with alanine at its 3' end enters the acceptor position of the ribosome. The alanine residue is transferred to the nascent polypeptide chain. Then, the ribosome switches to the tmRNA to continue translation. The tmRNA is encoded by the *ssrA* gene (small stable RNA A) (Karzai et al. 2000; Wiegert & Schumann 2001). The ssrA tag which consists of 11 amino acids in *E. coli* (AANDENYALAA) and 15 amino acids in *B. subtilis* (AGKTNSFNQNVALAA) is added to the C-terminus of the truncated protein. While the ssrA-tagged protein is degraded by ClpXP in *B. subtilis*, it needs an adaptor protein named SspB in *E. coli* that tethers it to ClpXP [Figure 3.6]. The tagged substrate can also be degraded by ClpAP without the means of SspB in *E. coli* (Flynn et al. 2001).

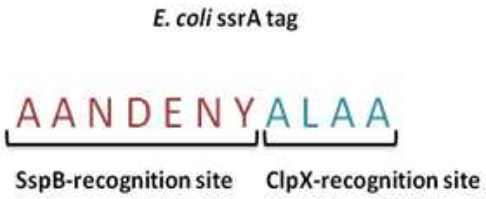

Figure 3.6 SsrA sequence amino acid sequence recognized in *E. coli*
It shows the SspB binding site (N-terminal) and the ClpX binding site (C-terminal)

The ssrA-tagged system was used to develop synthetic tools to target proteins for degradation in a controlled way (Cameron & Collins 2014; Griffith & Grossman 2008; McGinness et al. 2006; Guiziou et al. 2016). Tools like that can provide means to tune a protein of interest without interfering with the transcriptional or translational processes. It allows to understand a protein function for instance, or to design a genetic circuit, or even to establish a metabolic model.

Summary:

Researchers took the insights of the ssrA-targeted protein degradation used by bacteria in order to make use of it in synthetic biology. The ssrA tag is added to truncated proteins which are recognized by the ClpXP. Different studies engineered ssrA tags to obtain controlled systems to use as biological tools. In our project we aimed to use an ssrA-targeted protein degradation to understand and to point to the resource limitations in *B. subtilis.*

CHAPTER 4.

Problem Overview

4 Problem Overview

Improving protein production in *B. subtilis* is needed to fulfill the high demand on industrial enzymes in the market in an easy and inexpensive way. However, the overproduction of a heterologous protein leads to a growth rate decrease in microorganisms (Dong et al. 1995; Kafri, Metzl-Raz, et al. 2016). In addition, heterologous protein production is accompanied by lower secreted yields than the expected amounts (20-25 g/L) in *Bacillus* species (Ploss et al. 2016). In *B. subtilis*, the bottlenecks were identified at the secretion level, which is known as the secretion stress. The attempts in understanding the bottlenecks are focused on studying the secretion machinery, the membrane protein complexes and interactions, and the cell signaling pathways to obtain the super secreting strains. But until today, the consequences of high production level on the cell physiology and the growth rate of *B. subtilis* have not been documented yet. When the cell is overwhelmed with synthetic circuits, it is expected that to a certain level this system would disturb the cell by consuming all forms of resources (machines, energy, metabolites, *etc*.). Therefore, in my PhD project we were interested in (1) understanding the cell physiology, to investigate the effects of protein overproduction on the cell physiology and on different cell processes, (2) pinpoint the bottlenecks and the type of limiting resource, (3) and to overcome this limitation to improve protein production.

First, the project aimed to investigate the effect of an overproduced non-secreted gratuitous protein on *B. subtilis* physiology. To do that, we focused on choosing the expression system, the right protein, and the integration *locus* on the genome. First, high gene expression requires strong promoters and TIR (*Translation Initiation Region*) to ensure high levels of mRNA and recruitment of ribosomes. Second, the gratuitous proteins were preferred to be stable and costly in their amino acids resources. Reporter proteins such as GFP, mKate2 and LacZ were used since they are stable and can be followed during their production. Third, the genetic constructs were integrated in *B. subtilis* genome to ensure a stable expression. The gene dosage was also considered to ensure high gene expression. After constructing the strains, we used Live Cell Array (LCA) to follow the protein abundance by measuring the fluorescence and growth. LCA is a high-throughput technology that allows to monitor the optical density and protein fluorescence using a 96-well microplate. The measurements were done in different growth conditions in chemically defined media. The data collected from the measurements were treated by Matlab to compute the growth rates and the protein abundance. The results will be presented in the **chapter** 5.

Second, the project aimed to understand the bottlenecks of gratuitous protein production. The re-allocation of resources to produce the gratuitous protein deprives the cell to feed the production of essential endogenous proteins required for the cell's physiology and growth [Figure 4.1]. Cell resources can be metabolites such as amino acids or energy, cell machineries such as ribosomes, chaperones, or polymerases that are present in a limited cytosolic density. Different assumptions were published around the main resource causing the growth rate decrease in the presence of a gratuitous protein. Hence, in order to investigate more in detail the reallocation of resources, we performed relative protein quantification for strains overproducing different levels of GFP. Further analysis to investigate the limitations was done by the means of a targeted degradation system previously engineered in *B. subtilis* using ClpXP protease (Griffith & Grossman 2008; Guiziou et al. 2016). Targeting the gratuitous protein for degradation is a way to recycle amino acids to the pool of resources and to release the constraint on the constant cytosolic density. The growth of the cell was monitored while degrading the gratuitous protein. If the growth rate was restored, it would help us to conclude that the amino acids recycling helped the cell to restore the growth rate. Therefore, we used the targeted degradation system against the overproduced protein. Next, we monitored the growth rate of the cell during protein degradation. The results will be presented in the **chapter** 5.

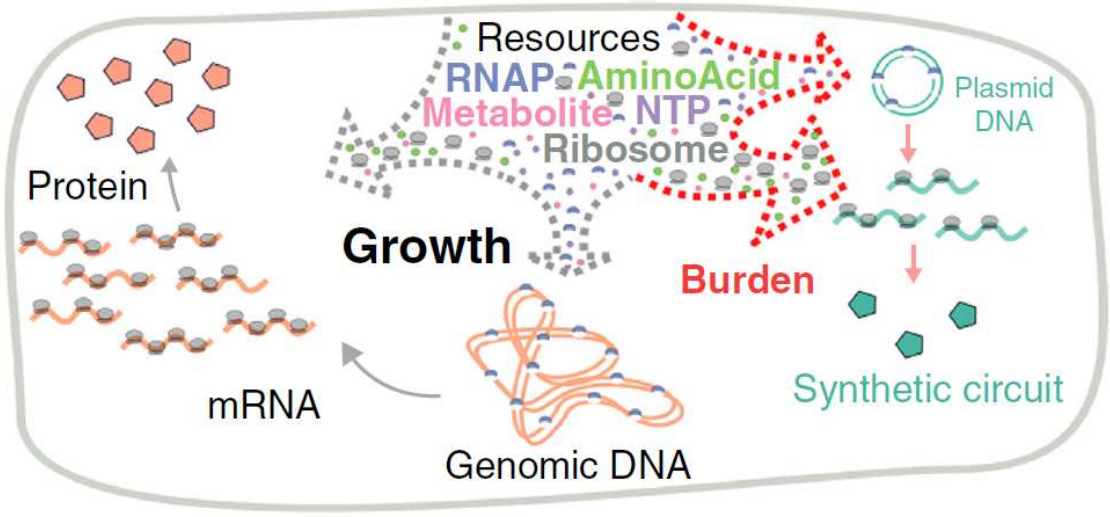

Figure 4.1 A representation of the resource allocation to the cell tasks

The expression of the synthetic circuit requires resources that are taken on the expense of the endogenous genes, and therefore it results in a cell burden (Borkowski, Ceroni, et al. 2016).

Third, to overcome the amino acid limitation, we decided to improve the degradation system in order to reach higher levels of degradation. For this purpose, we overproduced ClpXP protease from *E. coli* and *B. subtilis* together with the adaptor protein SspB from *E. coli*. This tool will allow us to target proteins for degradation, and thus to save resources, in order to favor the production of our protein of interest. Moreover, we aimed to investigate if the overproduction of ClpXP had negative consequences on the cell physiology, so we did relative protein quantification for a wild type and an overproduced ClpXP. The results will be presented in the **chapter 6.**

CHAPTER5.

Gratuitous Protein Overproduction Alters *B. subtilis* Cell Physiology

5 Gratuitous Protein Overproduction Alters *B. subtilis* Cell Physiology

5.1 Introduction

The first part of my PhD project is dedicated to the analysis of the consequences of high protein production on *B. subtilis*. My investigations focused on the gene expression level and on the cell physiology consequences. The most likely hypothesis at present is that the over-expression of a heterologous gene is costly for the cell in terms of cell resources and cytosolic space occupation. Consequently, a decrease in the cell growth rate is expected. Recent studies on *B. subtilis* showed that this microorganisms can unexpectedly tolerate the production of very high amounts of protein without showing any effect on the growth rate of cells grown on rich LB medium (Sauer et al. 2018). In the present work, we aimed at revisiting these results and firmly conclude on the effect of protein overproduction on the cell physiology using various growth media of defined composition (from poor to rich media). For that purpose, we built mutants of *B. subtilis* strains expressing heterologous genes, and the mutant growth and gene expression were monitored in different growth media.

The level of gene overexpression was controlled by three different means:

1. The expression system (*i.e.* the 'strength' of the genetic sequences controlling transcription and translation),
2. The gene of interest (*i.e.* short *versus* long coding sequences),
3. The location of the gene of interest on the circular chromosome (*i.e.* close to or far from the origin of replication).

First, we built a small set of inducible promoters to control gene expression at different levels. Many efforts have been made in *B. subtilis* to build synthetic expression systems enhancing protein production. Some of which make use of inducible promoters such as the P_{xyl}, the xylose-inducible promoter (Kim et al. 1996), the P_{sacB}, the sucrose-inducible promoter (Steinmetz et al. 1985), and the $P_{hyperspank}$ (P_{hs}) the IPTG (Isopropyl β-D-1-thiogalactopyranoside)-inducible promoter. We decided to use a non-metabolized inducer such as IPTG (analogous to allolactose), as metabolic inducers such as xylose or sucrose are also used as energy sources by the cell. Addition of such a 'metabolic' inducer would alter the growth rate in synthetic media and interfere with our conclusions.

Second, we used reporter proteins that can be easily monitored. Genes coding for proteins such as GFP, LacZ, and mKate2 were integrated in single or two copies into the

genome of *B. subtilis.* The use of proteins of different size and amino acid composition should allow us to deeply revisit the effect of protein overproduction on the cell physiology.

Third, the gene integration *locus* on the chromosome matters with respect to the copy number (*i.e.* gene dosage) as a function of the growth rate. In *B. subtilis* and *E. coli*, it was shown that genes located near the origin of replication exhibit high levels of expression at high growth rates (Sousa et al. 1997; Bryant et al. 2014; Sauer et al. 2016). The genes near the origin are replicated earlier than those close to the terminus, and in fast growing cells more than one replication fork takes place, meaning that a replication round starts before the previous one finishes. Hence, the gene dosage increases at regions close to the replication origin (Sousa et al. 1997; Couturier & Rocha 2006). Gene dosage was also observed and quantified in *B. subtilis*, for which it was shown that the expression of a heterologous gene can differ up to 5-folds depending on its location on the chromosome [Figure 5.1] (Sauer et al. 2016).

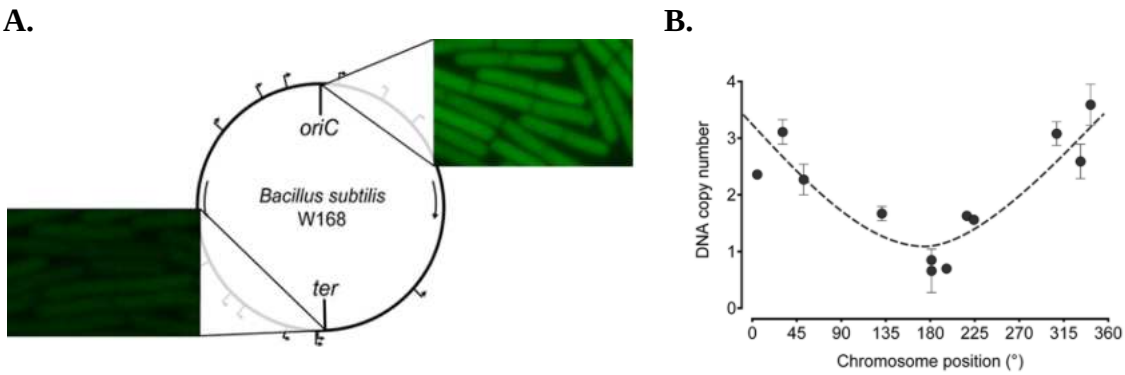

Figure 5.1 Gene dosage is influenced by location on the chromosome

(A) GFP is brighter when the gene is integrated at a close locus to the origin *oriC* than that integrated close to the terminator *ter*. (B) The DNA copy number measured by qPCR decreases with increasing distance from the origin of replication. (Sauer et al. 2016)

5.2 Design and characterization of suitable expression cassettes for overproduction of gratuitous proteins

5.2.1 Construction of a set of inducible and well-controlled promoters

A set of inducible, well-controlled promoters was required in order to widen the range of gene expression. Such a range of gene expression will allow us to determine the level of expression at which the growth rate decreases in *B. subtilis*. Besides, we thought that using a controlled

production through an inducible promoter would limit, during the process of strain construction, the occurrence of unexpectedgenetic adaptations (such as point mutations) that may be fixed in the population to overcome the cost of producing useless (*i.e.* gratuitous) proteins (Sleight & Sauro 2013).

We built a set of inducible promoters based on the architecture of the IPTG-inducible P_{hs}. IPTG induces expression by binding to the repressor LacI from *E. coli*, thus preventing LacI from binding to the *lacO* operator DNA sequences (present in the P_{hs}), close to the binding region of the RNA polymerase. Therefore, the use of an IPTG-inducible promoter first requires the insertion of the *lacI* gene in the *B. subtilis* genome.

P_{hs} was associated in tandem with constitutive promoters (with the P_{hs} downstream of a constitutive promoter such as $P_{const}P_{hs}$) in order to increase the level of expression as compared to the P_{hs} alone in the presence of IPTG. The use of different constitutive promoters will allow to obtain a range of strength [see section 8.3.1 for more details]. The constitutive promoters chosen are natural promoters controlling *B. subtilis* genes. The P_{ylxM} is the promoter of the putative signal recognition particle (SRP) component encoding gene. It was selected based on transcriptomic analysis (Nicolas et al. 2012). It allows relatively constitutive gene expression across growth conditions. P_{fbaA} is the promoter of the fructose 1-6 biphosphate aldolase encoding gene (Ludwig et al. 2001), it leads to a constitutive transcription level across environmental growth conditions (Nicolas et al. 2012;Borkowski, Goelzer, et al. 2016). Lastly, the P_{rrnJP2} is the second promoter element controlling the expression of the *rrnJ* gene coding for ribosomal RNA (Samarrai et al. 2011). The *rrn* genes are usually controlled by tandem promoters named P_1 and P_2 and the transcription is enhanced with increasing growth rate (Wellington & Spiegelmane 1993).

The different promoters alone or in tandem were coupled with variable Translation Initiation Regions (TIRs). The TIRs were selected from previous studies as well. The $^{gtlX}TIR_{gtlX}$ is from the work of (Sauveplane et al., unpublished) and corresponds to the 5'UTR of the *B. subtilis gltX* gene. The $^{fbaA}TIR_{short}$ is a modified form of the natural TIR of *fbaA*, and the $^{hs}TIR_{fbaA}$ is a modified form of the TIR of the artificial P_{hs} and of the natural TIR of *fbaA* (Borkowski, Goelzer, et al. 2016). A summary of the promoters and TIRs is presented in [Table 5-1].

Table 5-1 The constitutive promoters and the TIR elements constituting each of the synthetic inducible promoters

Name	Promoter	TIR
sP_1	P_{ylxM}	$^{gtlX}TIR_{gtlX}$
sP_2	P_{fbaA}	$^{fbaA}TIR_{short}$
sP_3	P_{ylxM}	$^{hs}TIR_{fbaA}$
sP_4	P_{rrnJP2}	$^{gtlX}TIR_{gtlX}$
sP_5	-	$^{gtlX}TIR_{gtlX}$

The inducible synthetic promoters were fused to a *gfp* gene. The genetic constructs were then integrated in the chromosome by single cross-over at a *locus* close to the origin of replication to maintain high gene dosage. The locus was chosen based on the work of (Nicolas et al. 2012) and (Sauveplane *et al*., unpublished).

The growth of the strains and the fluorescence of the GFP were followed by Live Cell Array (LCA). It is a high-throughput methodology which allows to monitor the growth of *B. subtilis* cells and the expression of a gene of interest in 96-microtiter plate reader (Botella et al. 2010; Buescher et al. 2012). To evaluate the strength of the different synthetic genetic constructs, measurements were performed in rich medium. The growth curves obtained for the strains grown in triplicates with and without induction by IPTG. The growth of the wildtype and each of P_{hs}*gfp*, sP_1*gfp*, sP_2*gfp*, are plotted in [Figure 5.2 A, B, and C respectively].

The result in [Figure 5.3] shows a range of GFP production. Moreover, without IPTG induction there is no GFP production, which means that the promoters were well controlled. We expected that the action of two tandem promoters would result in higher expression levels than the synthetic P_{hs}. However, the results showed that P_{hs} and sP_4 (P_{rrnJP2} $^{gtlX}TIR_{gtlX}$ *gfp*) are the strongest promoters among the synthetic promoters. Knowing that the synthetic TIR of P_{hs} was not used in this set, this suggests that if we would have used it with other promoters we may have obtained higher expression levels. Besides, the sP_4 is made of the P_{rrnJ} derived from the natural promoter of the *rrnJ* gene encoding rRNA, which is a very strong promoter in *B. subtilis* (Samarrai et al. 2011). This is consistent with the result of high GFP production when using this promoter next to P_{hs}.

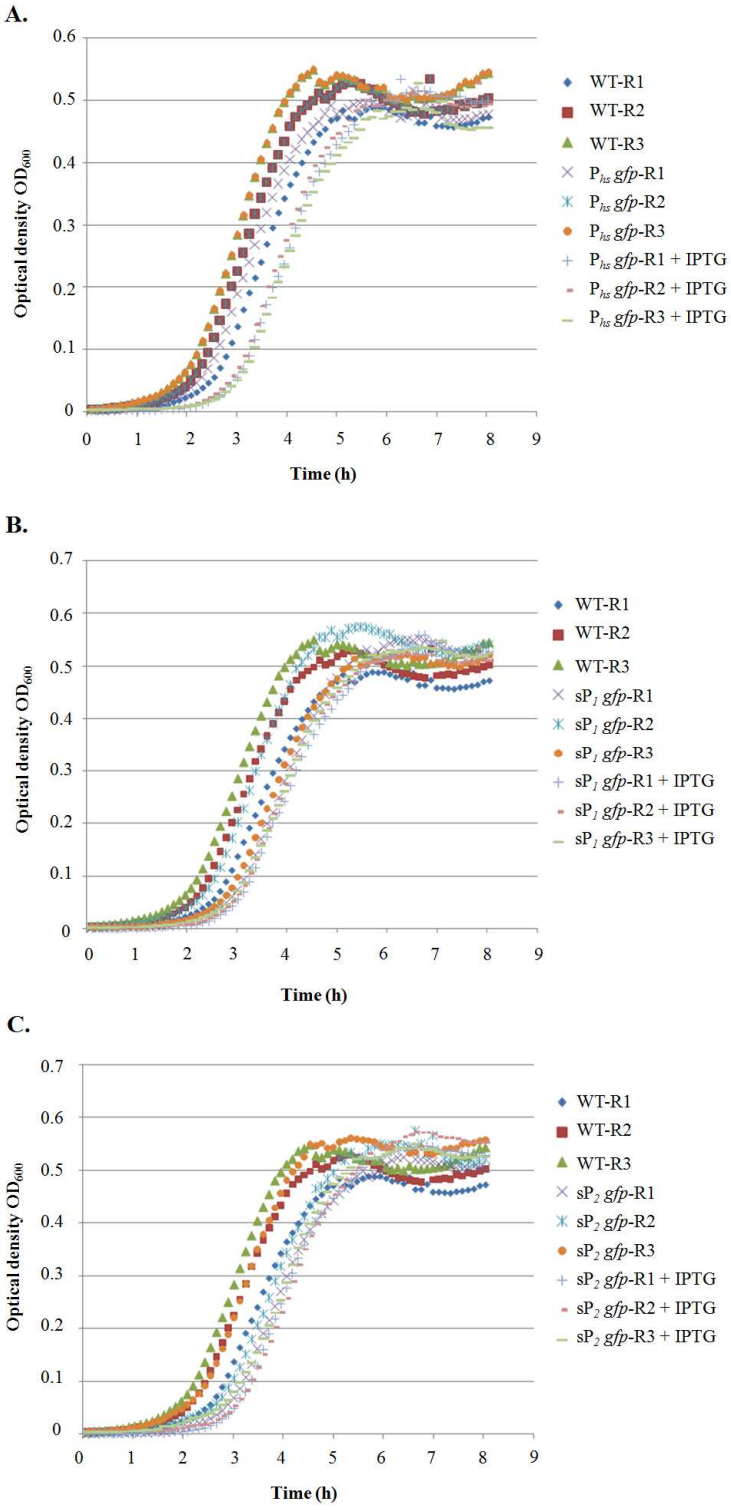

Figure 5.2 Growth curves followed by LCA for the wild type versus the strains carrying the synthetic inducible promoters upstream *gfp* in CHG medium

(A) Growth curves for the wild type vs P_{hs}*gfp*, (B) wild type vs sP_1*gfp*, (C) wild type vs. sP_2*gfp*. The strains were grown in triplicates, with and without the induction by IPTG (1 mM).

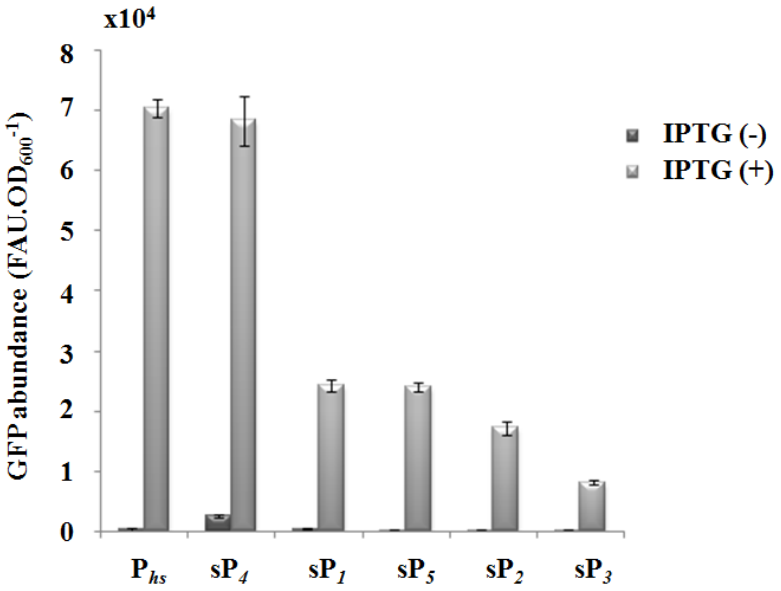

Figure 5.3 GFP abundance variation with the synthetic promoters (*sP*)

LCA measurements in rich medium CHG with IPTG induction (1mM) shows the GFP abundance when expressed under the control of the synthetic promoters (*sP*) and the P_{hs}.

5.2.2 Selection of the gratuitous proteins to be overproduced

First, we decided to build a fusion protein between GFP and β-Galactosidase from *E. coli*. This fusion should produce a costly protein in terms of amino acids. In addition, its expression should be easily monitored by fluorescence. The fusion protein was built by adding a linker of 12 amino acids (ASGGGGSGGGGS) to facilitate the folding of both proteins. The resulting chimeric protein was of 140 kDa. The *gfp-lacZ* fusion was assembled downstream the abovementioned inducible promoters. The genetic constructs were obtained in *E. coli*, verified by sequencing and finally transformed into *B. subtilis*. Growth and gene expression were monitored by LCA. The measurements were performed in defined media for the strains carrying the GFP alone or the GFP-LacZ fusion. The results showed a decrease in fluorescence in the GFP-LacZ-carrying strains as compared to the GFP-carrying strains [Figure 5.4 A]. However, the growth rate of the strains producing the GFP or the GFP-LacZ were similar to that of the control strain grown in the very same media. The growth rate is shown in [Figure 5.5] for some of the strains expressing *gfp* or *gfp-lacZ* under the control of the strongest P_{hs} and two of the synthetic promoters *s*P_4 ($P_{rrjP2}{}^{gtlX}TIR_{gtlX}$ *gfp*) and *s*P_5 ($P_{hs}{}^{gtlX}TIR_{gtlX}$ *gfp*), of high and moderate strength, respectively. The strains presented in [Figure 5.5] were grown in the defined medium, S, with a fullLacI derepression (*i.e.* 1 mM

IPTG). However, no significant growth rate decrease as compared to the control strain was observed. Similar results were observed with the other promoters whatever the growth medium being used.

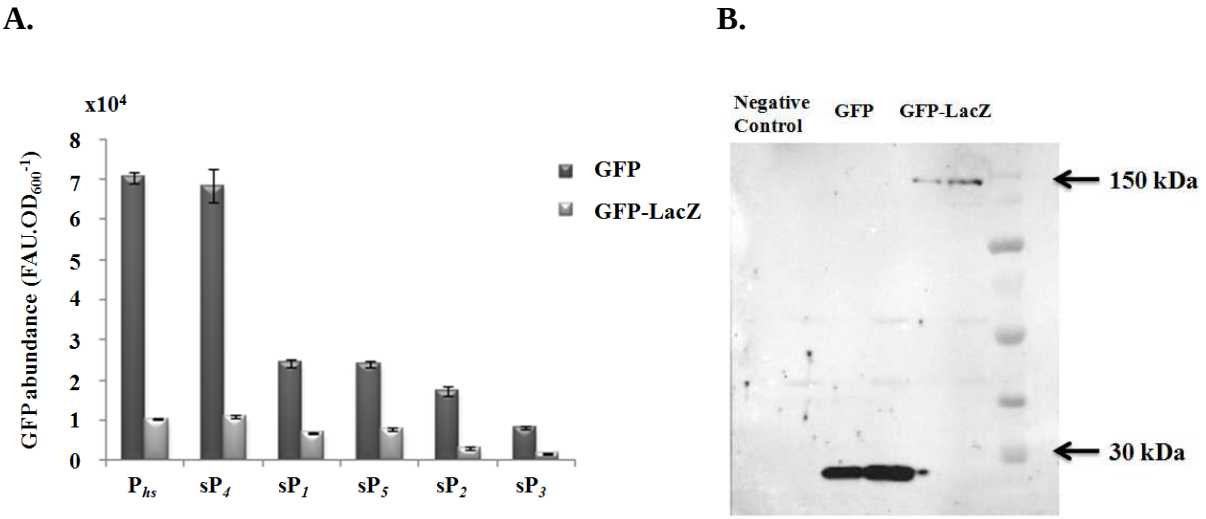

Figure 5.4 GFP-LacZ protein is less abundant in the cell
(A) LCA measurements for GFP abundance (Fluorescence AU.OD_{600}^{-1}) in strains expressing GFP or GFP-LacZ under the control of inducible promoters by IPTG 1mM. There is a clear low fluorescence in the GFP-LacZ carrying strains. (B) Western blot against GFP of the strains expressing GFP or GFP-LacZ controlled by P_{hs}. The cell extract was loaded two times in different amounts for each of the strains. The bands of GFP in GFP-LacZ are very light compared to GFP.

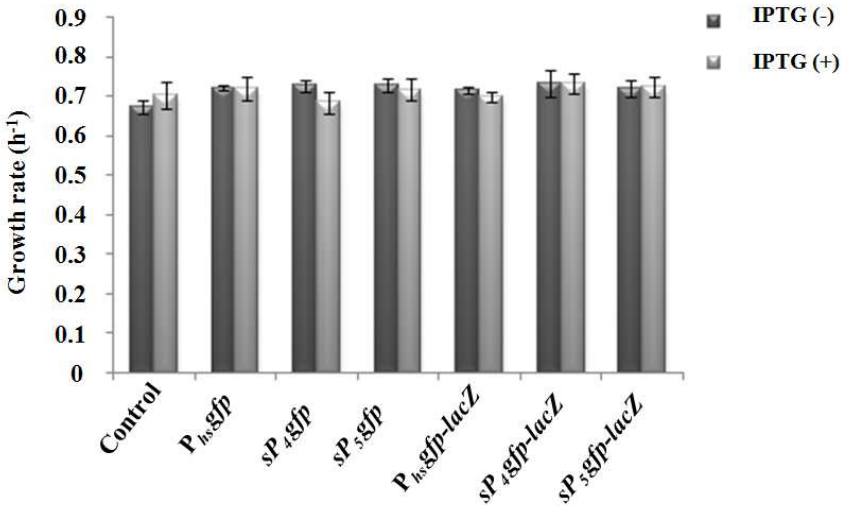

Figure 5.5 Growth rate is not affected with the synthetic promoters

When expressing *gfp* or *gfp-lacZ* fusion under the control of P_{hs} or the synthetic promoters no significant decrease in the growth rate is observed in S medium with maximum IPTG induction (1mM).

To check whether the 'quite low' fluorescence level observed on the GFP-LacZ fusion strains [Figure 5.4 A] was due to a low abundance in proteins or to a low fluorescence activity because of folding issues of the chimera protein, we quantified by western blot the GFP produced by the strains producing GFP and GFP-LacZ under P_{hs} [Figure 5.4 B]. The western blot indicated that the quantity of GFP in the produced fusion protein was much lower than that of the GFP alone, which is consistent with the LCA results. The lower abundance of the GFP-LacZ fusion protein than the GFP alone might be due to protein folding issues since LacZ is a tetramer larger than GFP, which may affect the proper folding of GFP. Therefore, to check whether there are misfolding problems in the GFP-LacZ fusion strains, we performed microscopic images for the GFP and GFP-LacZ fusion strains. As shown on [Figure 5.6 A], when GFP alone is produced it shows a homogeneous fluorescence in the cells. However, when GFP is fused to LacZ, it shows clear aggregate-like structures at the cell poles which could mean that low fluorescence is due to improper protein folding [Figure 5.6 B].

Due to the very low level of protein detected in the cells, the fusion protein was obviously not a good candidate to study the impact of protein overproduction on the cell physiology. Indeed, there was not any negative effect on the cell at the level of growth. We thus decided not to overproduce chimera to investigate further the consequences of protein overproduction on the cell physiology.

A. **B.**

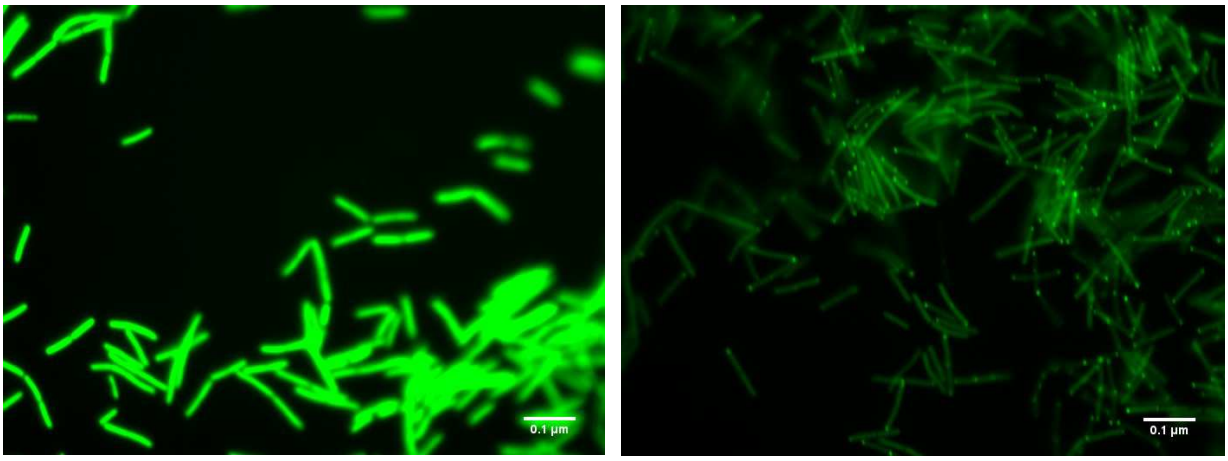

Figure 5.6 Microscopic images showing the GFP fluorescence
(A) Homogenous fluorescence of GFP in GFP-producing strains. (B) Aggregate-like structures of the GFP-LacZ fusion protein appear at the cell poles.

Beta-galactosidase is a large protein (112 kDa) and it was shown to cause a growth rate defect on *E. coli* when massively produced (Dong et al. 1995). Therefore, we constructed new strains by integrating one or two copies of *lacZ* at 2 *loci* in the *B. subtilis* chromosome. The first *locus* is *amyE*, a non-essential gene, close to the origin of replication and known as a usual integration *locus* in *B. subtilis*. The second *locus* is *nprE*, a non-essential gene as well, located near the terminus of replication. Two integrative plasmids each carrying the homology regions for either *amyE* or *nprE* were built to integrate the gratuitous reporter genes using the StarGate cloning method; an efficient and time saving technique in molecular biology (see section8.1.1.2). This methodology was acquired in DSM in the Netherlands and we made use of their in-house plasmids. Two copies of *gfp* or *lacZ*, or one copy of each of *gfp* and *lacZ* were constructed. The abovementioned set of synthetic promoters was used to control the expression of the reporter genes.

Miller Assay was performed to estimate the LacZ activity in the strains carrying single or two copies of *lacZ*. The results of the assay is shown in [Figure 5.7] where LacZ(2x) has nearly twice higher activity than the LacZ producing strain. Cell growth was monitored using LCA in rich defined medium. The effect on the growth rate in rich medium was not significant even in fully derepressed growth conditions (1 mM IPTG), although IPTG was either added at the beginning of the experiment or injected during the exponential growth. In poor medium, the 'expected' growth defect was neither stronger nor significant (M9 Pyruvate which normally gives rise to a growth rate of about 0.3 h^{-1} for a wild type strain) [Figure 5.8]. We concluded that the level of expression of *lacZ*, when present in two copies and expressed under the control of P_{hs} in the presence of 1 mM IPTG, was not high enough to cause a growth rate decrease in *B. subtilis*.

Alternatively, this result may indicate that there exists in *B. subtilis* a compensation mechanism to maintain a normal growth rate when overproducing a gratuitous protein.

The level of production of the gratuitous protein can be further increased by various means. For instance, this can be achieved by increasing the gene copy number through a multi-copy plasmid or by increasing the production strength using stronger promoters. Since multi-copy plasmids are not highly stable in *B. subtilis* (Bron et al. 1991; Wong 1995), we decided to build new expression cassettes using a stronger promoter/TIR DNA sequence.

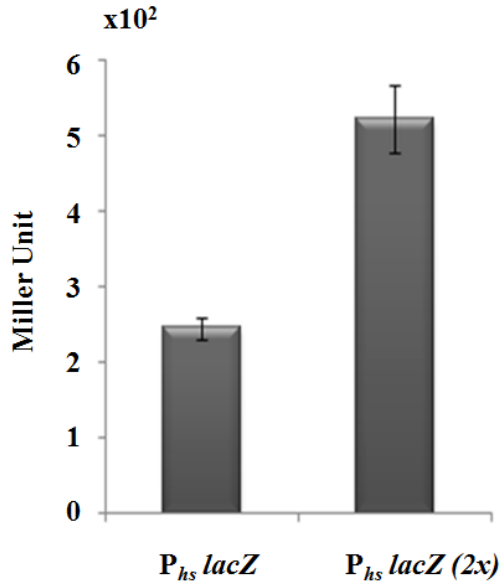

Figure 5.7 Miller Assay for β-Galactosidase activity

The assay was performed on the strains expressing one *lacZ* copy (P_{hs}*lacZ*) and two *lacZ* copies (P_{hs}*lacZ* (2x)).

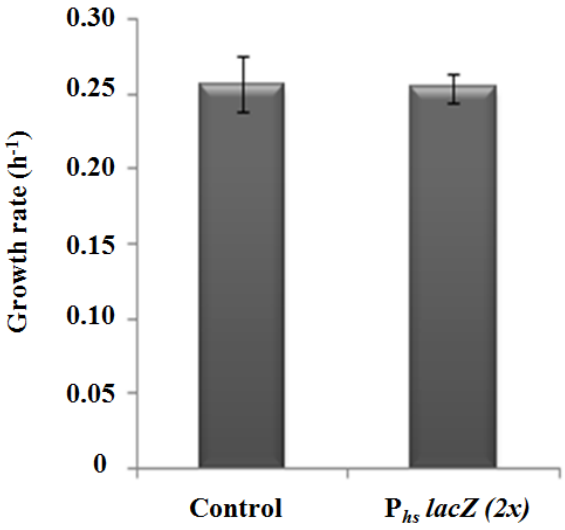

Figure 5.8 Growth rate of LacZ strain and a control strain

Double integration of *lacZ* showed a similar growth rate to the control strain in poor defined medium (M9P).

5.3 Overproduction of a gratuitous protein and consequences on the *B. subtilis* cell physiology

5.3.1 Overproduction of a gratuitous protein in *B. subtilis* reduces the rate of growth

The promoter used in our previous strategies was one of the strongest inducible promoters, yet we did not observe any growth defect as expected. Therefore, we decided to use the constitutive P_{veg}, the strongest native constitutive promoter in *B. subtilis* known so far (Guiziou et al. 2016). The transcription of P_{veg} is almost ten times higher than that of the P_{hs} in derepressed conditions as shown in [Figure 5.9]. The strains used in [Figure 5.9A] carry the *gfpmut3* gene (Cormack et al. 1996) under the control of P_{hs} and the *sfgfp* gene (superfolder *gfp*) (Pedelacq et al. 2006) under the control of P_{veg}. The sfGFP folds faster and is brighter than the GFPmut3 variant. Similarly, the overexpression of *lacZ* (as reporter gene) showed a 10-fold increase when using P_{veg} instead of P_{hs} [Figure 5.9B].

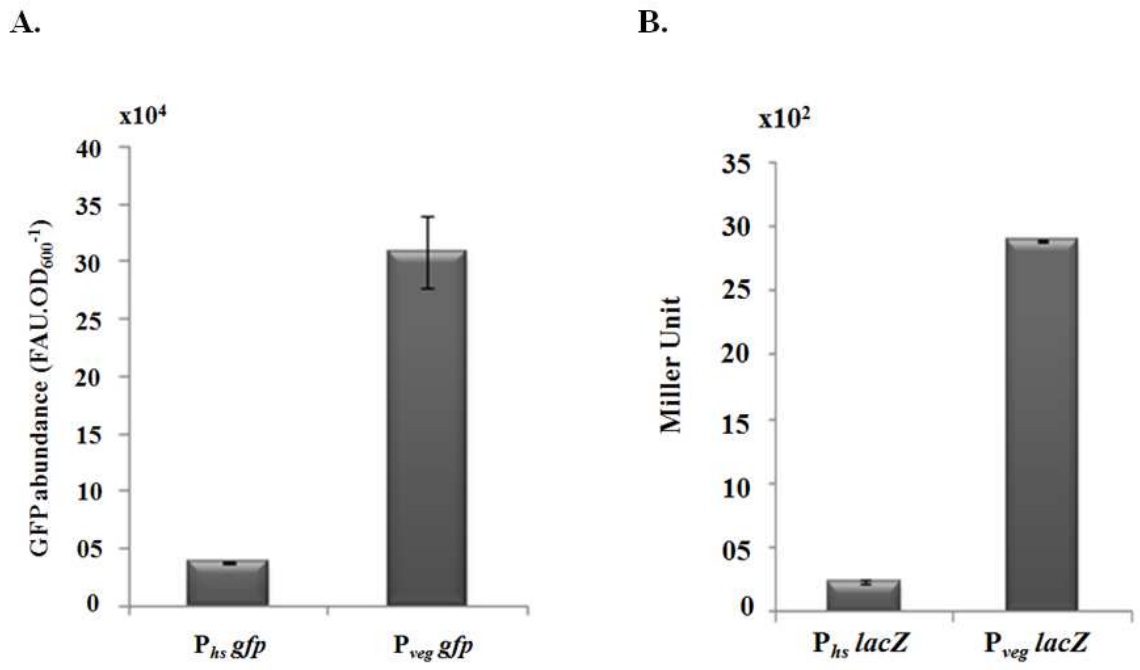

Figure 5.9 Comparison between P_{veg} and P_{hs} promoters fused to *gfp* and *lacZ* in a rich medium

(A) LCA measurements for strains carrying P_{veg} *gfp* or P_{hs} *gfp*. (B) Miller Assay performed for strains carrying P_{veg} *lacZ* or P_{hs} *lacZ*.

A strain carrying P_{veg}*lacZ* in the *amyE locus* was built (see the Materials and Methods section) and LCA measurements were performed to monitor cell growth. First, the measurements were performed in S medium (a minimal medium with glucose as carbon

source, which allows the *B. subtilis* wild type strain to reach a growth rate of about 0.7 h^{-1}). The growth rate of the P_{veg} *lacZ* carrying strain decreased by ~15% [Figure 5.10].

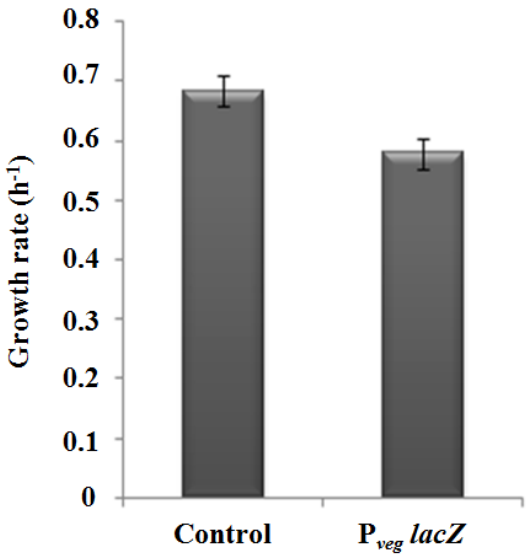

Figure 5.10 Growth rate reduction shown on *B. subtilis* upon expression of *lacZ* gene under the control of P_{veg} promoter

In poor defined medium (S), P_{veg} *lacZ* carrying strain had a 15% growth rate decrease compared to a control strain.

In order to confirm this result, the *sfgfp*, *mkate2*, and *lacZ* genes were placed under the control of P_{veg} and integrated into the *B. subtilis* chromosome in either one or two copies. The first copies were integrated into the *amyE* locus, and the second copies (if present) were integrated in the *nprE locus*. Growth was monitored using LCA in cells grown in poor (S), intermediate (SX that is S + asparagine, malate, and glutamate), and rich (CHG) media. These media give rise for a wild type strain to growth rates of approximately 0.7 h^{-1}, 1.0 h^{-1}, and 2.0 h^{-1}, respectively. The growth rate decrease previously observed in the P_{veg}*lacZ* carrying strain was also observed in cells overproducing any of the three different gratuitous proteins from both single and double gene copies. The results are presented on [Figure 5.11], which shows the growth rates (in duplicates if LCA measurements) of the strains carrying the gratuitous genes (*gfp*, *mkate*, and *lacZ*) in different media. The growth defect was strong in S and SX media and increased with the double gene copy integrations (indicated by 2x). On the other hand, when the strains were grown in rich medium (CHG), the growth rate decrease was not as important as in the defined minimal media, S and SX. It was more significant with the strains overproducing the gratuitous protein from two gene copies.

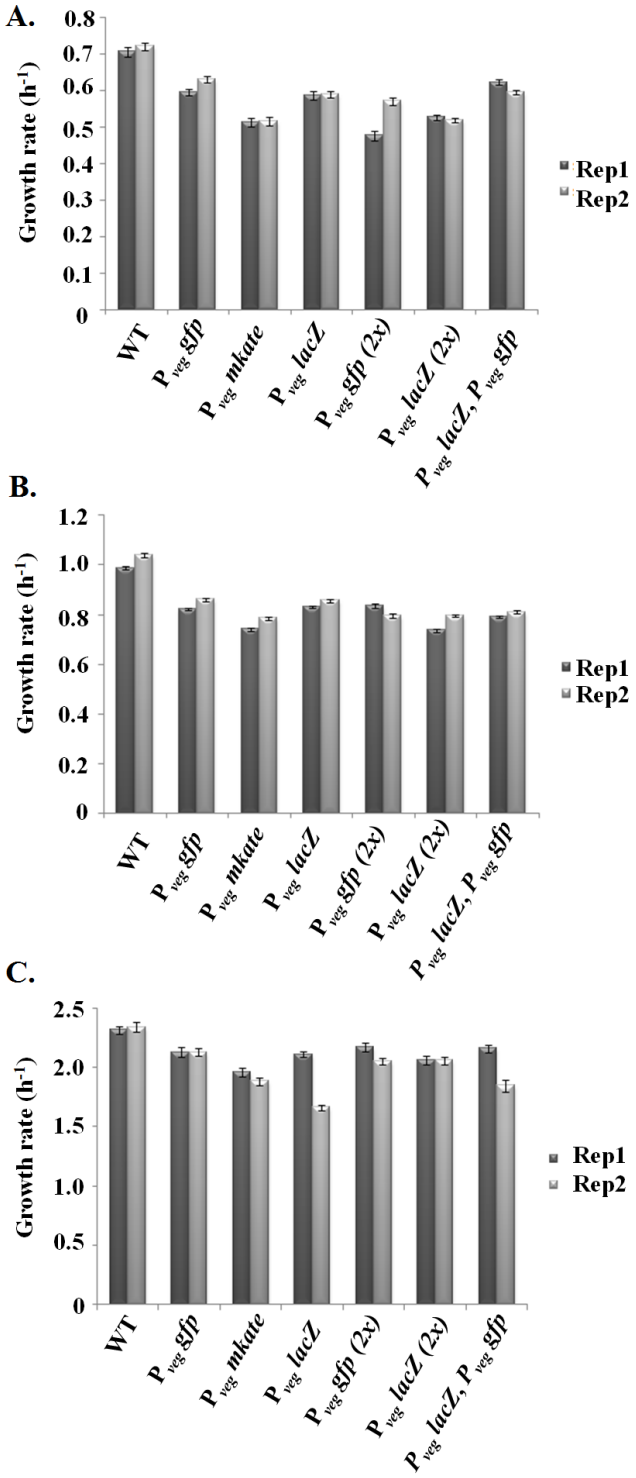

Figure 5.11 Slower growth rate for all the strains carrying gratuitous genes in different defined media

The graphs show duplicates of the growth rate for the strains carrying single or double copies of *gfp*, *mkate*, and *lacZ* grown in (A) poor defined medium (with glucose) S, (B) SX medium (S + malate, glutamate, asparagine), (C) rich defined medium CHG.

After showing a growth rate decrease with high protein expression, we were interested in quantifying the variation of the growth rate with respect to different levels of protein expression. This would give us information on the tolerance of *B. subtilis* to protein overproduction, by showing the minimum level of protein production resulting in growth rate decrease. Therefore, we built new expression cassettes based on a library of promoters selected from the work of Guiziou S. et al. (Guiziou et al. 2016). LCA measurements were performed to monitor the growth rate and the GFP fluorescence in CHG, SX, and S media. The results are shown in [Figure 5.12]; the graphs represent the variation of the growth rate as a function of the GFP fluorescence level (measured as fluorescence/OD_{600} in arbitrary unit). A sharp decrease in growth rate in S medium, a smoother decrease in SX medium, and then a constant variation in CHG medium were observed. This means that with lower amount of proteins produced, the cell slowers its growth rate in poor media more than in richer media. Nevertheless, this suggests that we should be able to observe a growth rate decrease in rich medium under the condition that the protein production level is increased.

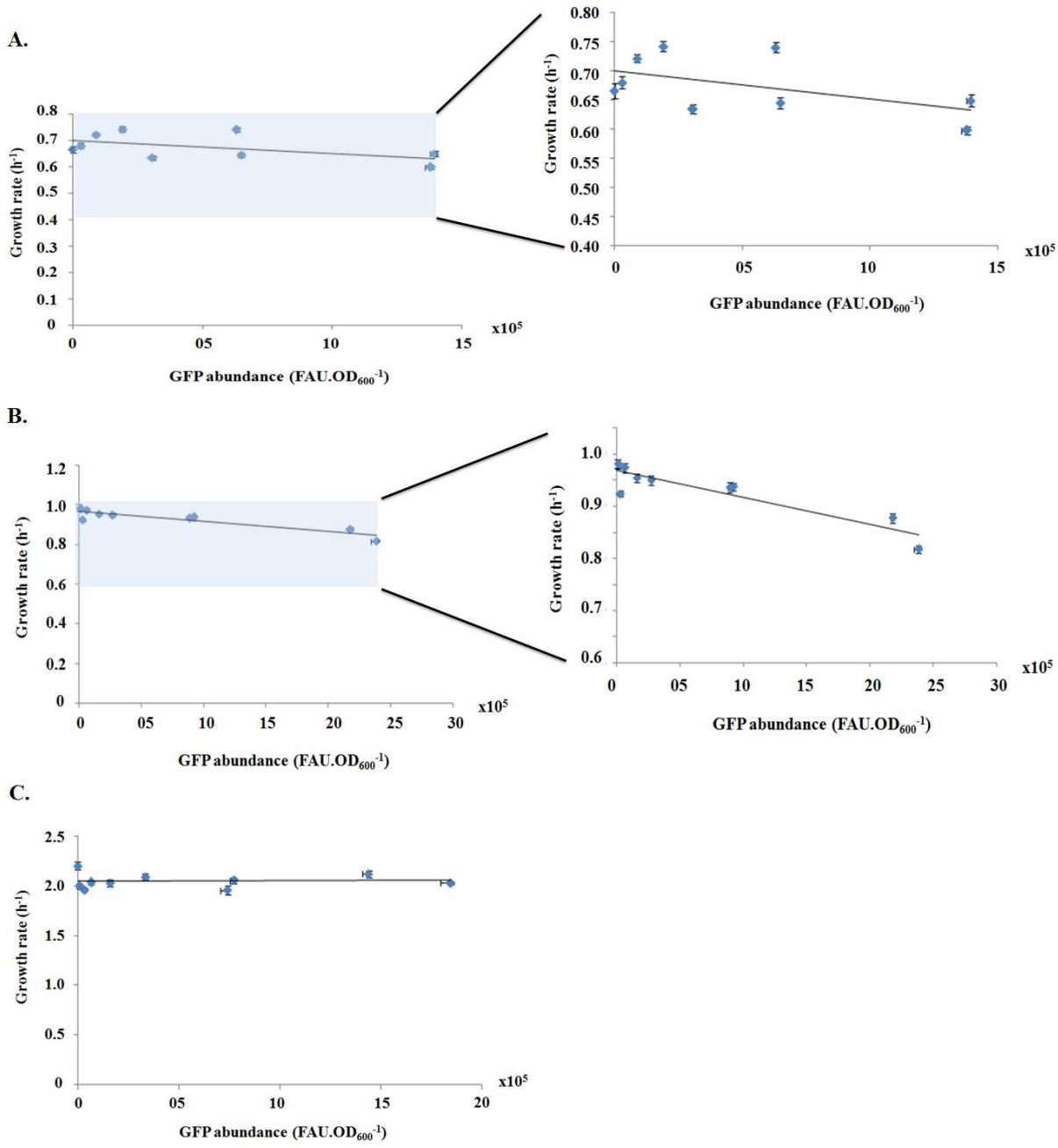

Figure 5.12 Growth rate variation with respect to different levels of GFP in different defined media (A) In medium S the growth rate (h^{-1}) drops at relatively low values of $FAU.OD_{600}^{-1}$, (B) in the intermediate SX medium the growth rate shows a smoother decrease compared to A, and (C) in CHG medium the variation is constant.

5.3.2 Growth rate is altered when the GFP production reaches a certain threshold

After showing a burden on the growth rate, we were interested in finding the quantity of proteins that contributed to the burden. For that purpose, we first did an SDS-PAGE gel to

check the heterologous protein abundance with respect to the total soluble cell proteins. The strains overproducing GFP, LacZ, and LacZ with GFP, were grown in S medium until the exponential phase. They were harvested and treated to break the cells and extract the cytosolic proteins. The protein extracts were measured by the Bradford assay to be able to adjust the amounts when loading them on the SDS-PAGE gel. The cell extracts were loaded on SDS-PAGE gel and revealed by coomassie staining [Figure 5.13A]. GFP, LacZ and LacZ with GFP clearly appear to be abundant in comparison to the cytosolic proteins [the GFP^{ALGG} in Figure 5.13 will be explained later in 5.4.2.3]. In order to estimate the percentage of the heterologous proteins in the lanes, we did an estimation based on the signal intensity using ImageLab software. The calculated percentages for the heterologous proteins are as follows by respecting the order on the SDS-PAGE [Figure 5.13 A]: (2) GFP_{Bsu} 9.07%, , (3) $GFP^{ALGG}{}_{Bsu}$ 6.01%, (4)$LacZ_{Bsu}$ 6.92%, (5) GFP_{Bsu} 6.94% and $LacZ_{Bsu}$ 5.49%, (7) GFP_{Eco} 7.56%.

For a stronger validation of the protein amount that was really produced, we performed western blots against GFP, and one representative result is shown on [Figure 5.13C]. From a total cell extract of ~6.75 µg in 20 µL total volume, GFP was estimated to be 0.66 µg based on the slope calculated from the intensities of bands corresponding to the GFP scale of known amounts [Figure 5.13 B, C]. As a result, the GFP represented ~9.77% of the total extracted proteins, which is close to the values calculated based on the signal intensity. Moreover, in [Figure 5.13 D] the cells on LB agar plates carrying P_{veg}*mkate* (pink) and P_{veg}*gfp* (green) show the fluorescence of the proteins by the naked-eye.

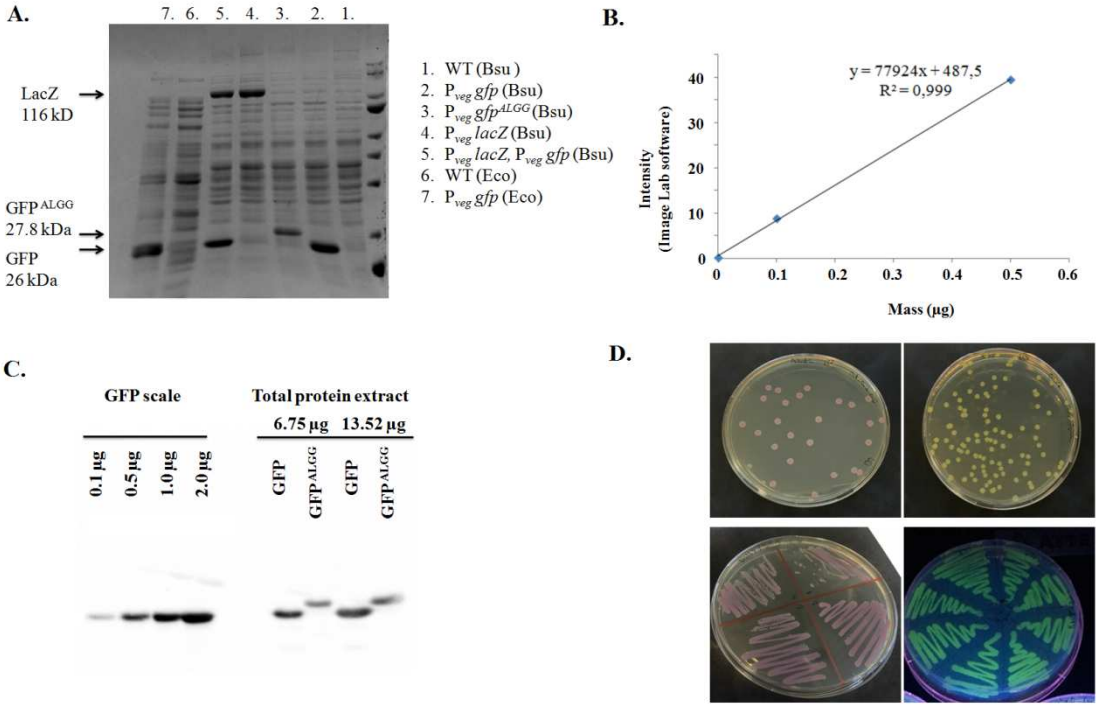

Figure 5.13 The heterologous protein quantification with respect to the total soluble proteins extract

(A) A coomassie stained SDS-PAGE gel that shows the total soluble proteins extract for the wild type and each of the strains producing GFP, LacZ, and LacZ, GFP. (B) The graph representing the intensity of the bands measured by ImageLab software versus protein mass, which was used to quantify the GFP in the western blot (C) A western blot against GFP showing a GFP scale of known amounts, and the GFP produced in a total soluble proteins extract of 6.75 µg and 13.52µg. (D) The fluorescence of the proteins appears on the cells producing mKate2 and GFP which shows the intensity of production.

In *E. coli*, 30% of gratuitous protein caused the cell to cease down (Dong et al. 1995). In their work, the protein was produced using a multi-copy plasmid. Obtaining 30% of gratuitous protein in *B. subtilis* was not reachable in our case. However, we performed an experiment to compare the two species producing GFP. In order to estimate and compare the growth rate of *E. coli* and *B. subtilis* with respect to protein production, we used a multi-copy plasmid carrying P_{veg}*gfp* in *E. coli* and an integrated copy in the *B. subtilis* chromosome. As shown on [Figure 5.14A], the GFP abundance in *E. coli*was 3.5 times lower than that in *B. subtilis*. However, in *E. coli* it led to a growth rate decrease of around 16% in CHG medium, while a 3.5 times higher GFP abundance in *B. subtilis* only led to a 10% growth rate reduction in CHG medium. In order to confirm that there is more GFP produced in *B. subtilis* than in *E. coli*, we ran a western blot and estimated the amount of GFP in both species by referring to a known concentration of a purified GFP (already purified in the lab). The western blot showed that the GFP in *E. coli* was ~1.5 times less abundant than in *B. subtilis* [Figure 5.14B].

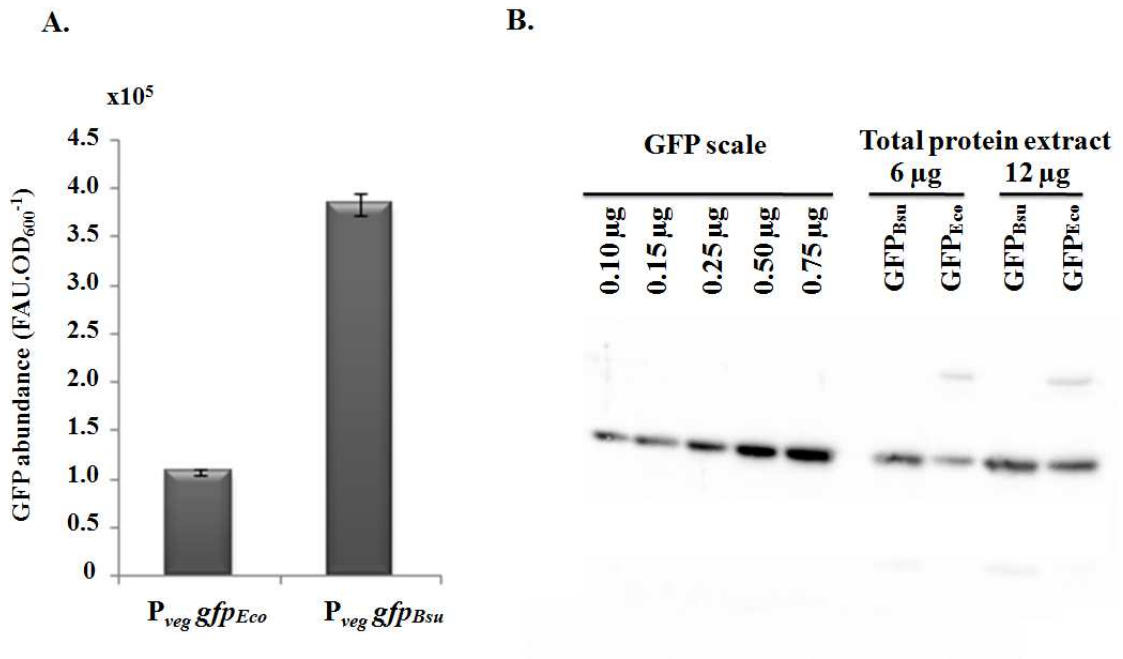

Figure 5.14 GFP abundance in *E. coli* and *B. subtilis*

(A) LCA measurements in CHG medium, showing the GFP abundance in *E. coli* and *B. subtilis*. (B) Western blot against GFP produced in *B. subtilis* and *E. coli* from a total cytosolic protein extract of 6 and 12 µg and a scale of known GFP amounts.

These results suggest that the growth rate of *E. coli* is more affected than that of *B. subtilis* for a given amount of gratuitous protein. How does *B. subtilis* as compared to *E. coli* to compensate for the overproduction of a gratuitous protein is an issue that will be addressed in the [section5.4] of the manuscript.

5.3.3 Gratuitous protein overproduction affects the bacterium cell size

Protein overproduction in *B. subtilis* does not only slower growth but it also results in increased cell size. *B. subtilis* strains overproducing the gratuitous proteins were cultured in S medium and harvested during exponential growth (OD_{600}=0.2). As shown in [Figure 5.15], microscopic analysis revealed a slight increase in the cell size of the GFP and LacZ overproducing strains as compared to the WT. The WT cells (A) looked smaller than the cells overproducing GFP (B), and LacZ (C), [the GFP^{ALGG} in Figure 5.15 will be explained later in 5.4.2.4].

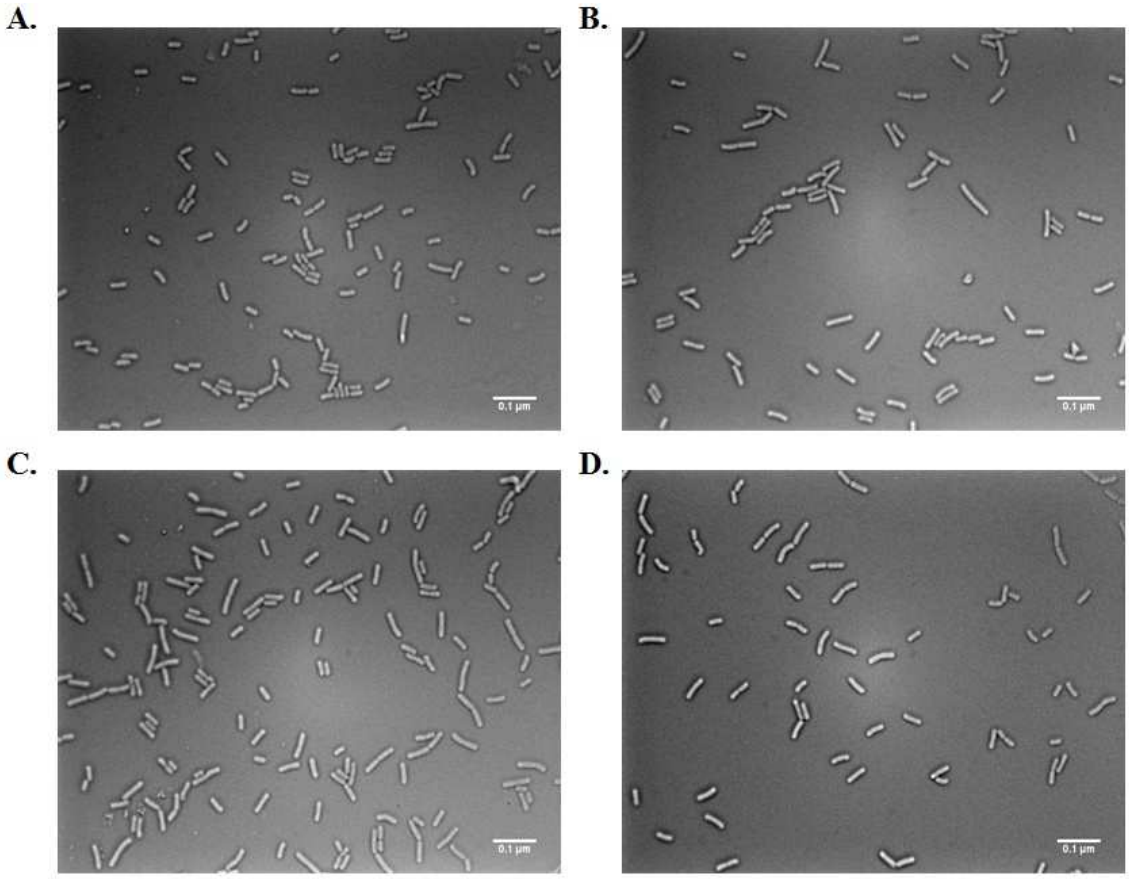

Figure 5.15 Microscope images for *B. subtilis* strains over-producing gratuitous proteins
(A) Wild type strain, (B) GFP-producing strain, (C) GFP^{ALGG}-producing strain, (D) LacZ-producing strain.

In order to confirm the cell size increase caused by overproducing a heterologous protein, we performed flow cytometry analysis. Flow cytometry allows to analyze a cell population at the single-cell level and to obtain statistical analysis about the size and fluorescence of cell populations. As above, cells were grown in S medium, harvested at OD_{600} 0.2 and treated for cell fixation before flow cytometry measurements. A scatter plot for the Fsc (Forward Scatter; proportional to the cell size) *versus* the SSC (Side Scatter; related to the 'granulation') for each of the wild type strain, and GFP and LacZ overproducing strains are presented in [Figure 5.16 A, B, D]. Within these populations, around 50,000 cells were selected in an identical ellipse used to gate the data. The gated cells were plotted as a histogram showing the size distribution of all the strains [Figure 5.16 E].

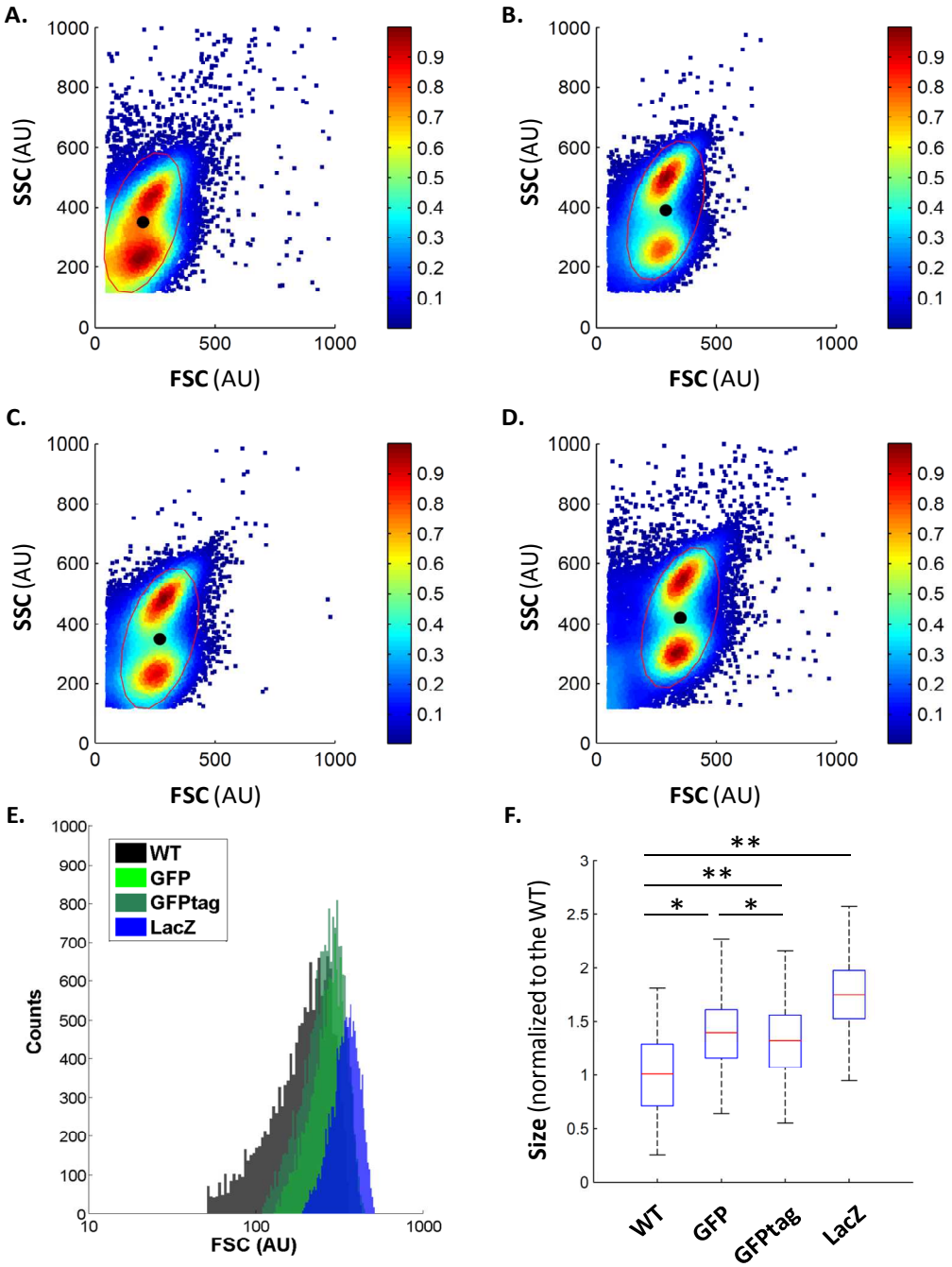

Figure 5.16 Flow cytometry data

A scatter plot for the Fsc (Forward Scatter) versus the SSC (Side Scatter) for each of the (A) wild type strain, GFP (B), GFP^{ALGG} (C) and LacZ (D) producing strains. The ellipse is the gated treated cells, which is identical in all of the plots. (E) The gated cells are plotted in a histogram showing the size distribution of all the strains. (F) The Fsc values of the mutant strains presented in boxplot are normalized to the Fsc value of the wild type.

These results indicate that the cell size of the LacZ and GFP overproducing strains are bigger than that of the wild type. The Fsc values of the mutants were normalized to that of the wild type [Figure 5.16 F]. The Fsc mean values and the standard deviation are presented [Table

5-2]. Note that the computed *t* test (*i.e.* student test) *p*-values were zero, which indicated that the mean, variance and distribution of the datasetswere with extremely high significance (<0.001) from unrelated populations. In other words, the cell sizes were significantly different between the wild type and overproducing strains.

The results from the previous chapter showed that the growth rate of the strains overproducing LacZ decreased more than that of the strains overproducing the GFP. Similarly, the effect of protein overproduction on the cell size also turned out to be dependent on the 'cost' of the overproduced protein.

Table 5-2 The mean and the standard deviation (std) of the size of each of the strains in the gated zone

	Fsc	
Data zone (elipse)	mean	std
WT	198.94	73.11
GFP	276.14	62.88
GFP^{ALGG}	261.94	66.54
LacZ	347.38	63.85

5.4 *B. subtilis* cells adapt to a huge overproduction of a gratuitous protein by alleviating constraints on the use of amino acids

The next question to address was why as compared to the current knowledge in *E. coli* is the growth rate defect only observed after overproducing so much of a gratuitous protein? In order to address this issue, we ran proteomic analysis in GFP overproducing strains.

5.4.1 GFP overproduction in *B. subtilis* results in reduced production of other proteins

Proteomics analysis were carried out on three strains expressing either one copy of the expression cassette "P$_{veg}$*sfgfp*" (strain MZ139) or two copies of this expression cassette (strain MZ140), and a wild type strain (control strain MZ72). Cells were grown in S medium until it reached an optical density (OD_{600nm}) of 0.23 meaning that the cells were harvested in exponential phase. The growth curves of the triplicates of each strain are presented in [Figure 5.17]. Soluble proteins were analyzed by proteomics based on three biological replicates for each strain. Statistical analysis were performed by Dr. C. Henry using the PAPPSO software X!TandemPipeline (Langella et al. 2017).

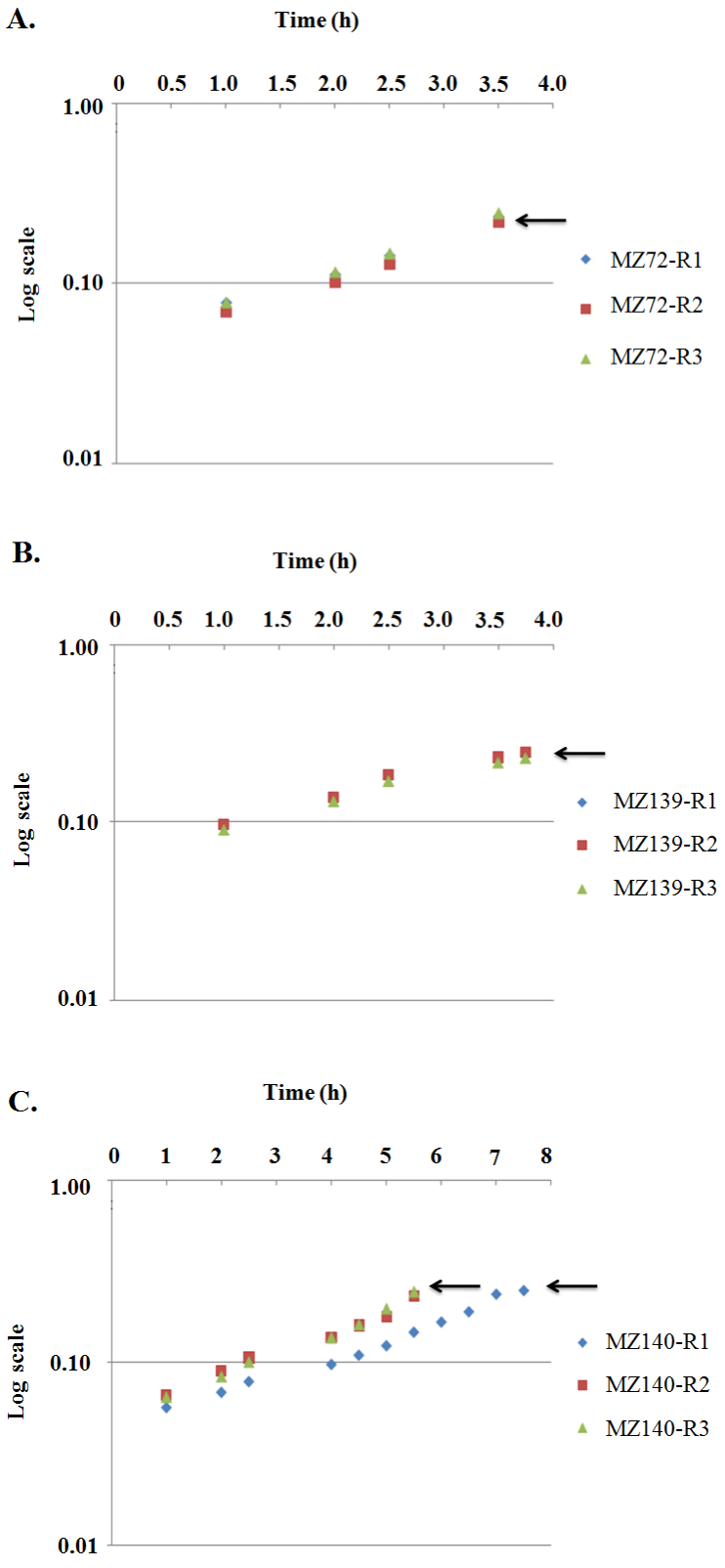

Figure 5.17 Growth curves of the wild type and GFP overproducing strains in S medium

The cells were harvested around OD_{600} 0.23 as indicated by the arrows. (A) MZ72 the control strain, (B) MZ139 the strain with one *gfp* copy, (C) MZ140 the strain with two *gfp* copies. The average growth rates of each of the strains were 0.454 h^{-1} (MZ72), 0.364 h^{-1} (MZ139), 0.307h^{-1} (MZ140).

5.4.1.1 Detection and relative quantification of GFP

The proteomics analysis confirmed an increase in GFP production between strains transformed by one and two copies of the expression cassette. Indeed, the NSAF (Normalized Spectral Abundance Factor) indicates a proportion of GFP ranging from 4 to 6% of the total soluble proteins detected and quantified by the mass spectrometer [Figure 5.18 B]. According to this quantification and in agreement with measured fluorescence [Figure 5.18 A], double integration of the same GFP overexpression cassette did not result in the same increase in GFP production by the strains.

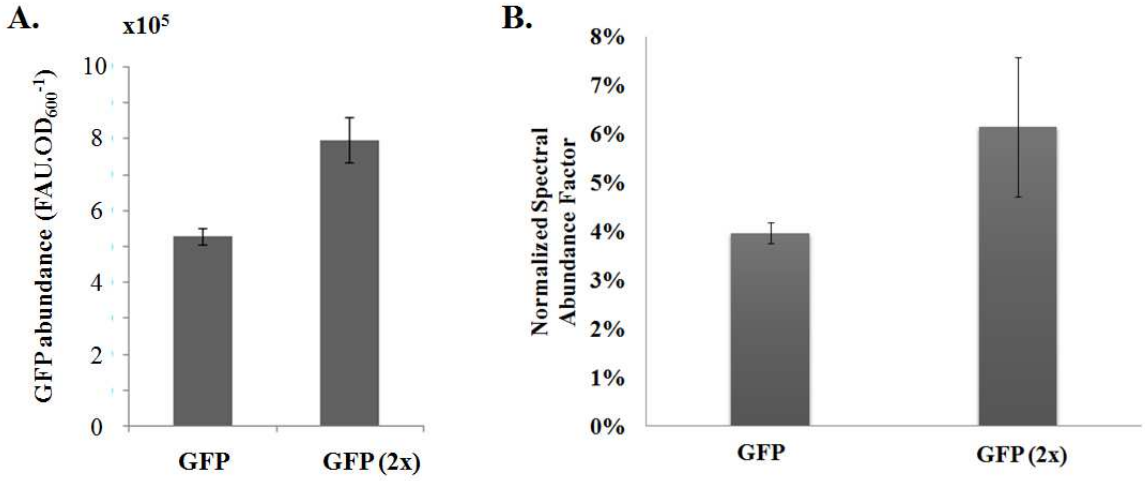

Figure 5.18 GFP abundance in the strains over producing GFP by one and two copies of the expression cassette P_{veg} *sfgfp*.

(A) The GFP fluorescence measured by LCA. (B) The normalized spectral abundance factor for the GFP in relation to the total soluble proteins detected by the mass spectrometer.

In another growth condition, such a result [Figure 5.18] could have been linked to the integration *locus*. The second copy of the cassette is integrated close to the terminus of replication. This integration locus is supposed to lower the gene dosage when cells grow rapidly in comparison with a locus closer to the origin of replication. However, the impact of the *locus* of integration with regard to the gene dosage is substantial when cells grow in a rich medium. Indeed, DNA replication needs to compensate the speed of division of the cells by increasing their number of DNA replication forks resulting in a bias in gene dosage. Such an

effect is thus not expected to be as strong in the minimal S medium that allows a cell division every hours.

One hypothesis to explain the different production of GFP in these strains is linked to resource allocation. Strains carrying a double copy of the GFP expression cassette displayed a decreased growth rate as compared to ”single copy” or to control strains. This decrease is linked to a non-optimal distribution of resources between endogenous cell processes and the heterologous process. The difficulty to increase the quantity of GFP produced by increasing the cassette copy number somehow indicates that the level of production of a protein does not increase linearly with the gene dosage in this growth condition.

5.4.1.2 Global impact on the proteome of *B. subtilis*

Regarding the total soluble protein content, interestingly, the major effect of overproducing a gratuitous protein in *B. subtilis* is a decrease in the quantity of many proteins in the overproducing strains as compared with their respective amount in the control strain.

To analyze the global impact on the protein production, we compared the mean spectral count of the biological replicates for a given protein in an overproducing strain and compared it with that of the control strain. That way we were able to select 399 proteins that are repressed in the “GFP one copy” strain (MZ139) and 423 proteins that are repressed in the “GFP two copies” strain (MZ140). Although the term 'repressed' does not seem to be semantically accurate in our case, as we do not know how the cell was able to turn off the production of several proteins, in this subchapter we will make use of the term 'repressed'. The repressed proteins were categorized in the pie chart [Figure 5.19] according to the folds of repression compared to the wild type strain. Looking at the two pie charts (Figure 5.19, panels A and B), we can observe that the most repressed proteins (at least 4 folds) increased from 1 to 11 when having an additional copy of *gfp*. Globally, the intensity of repression of the selected soluble proteins increased together with the increase in heterologous protein production. Indeed, comparing the slices of the most intense repression factors between the “one copy” and “two copies” strains indicates a possible shift towards higher repression when GFP production increases.

In order to analyse the effect of increasing the GFP level on the protein repression, we globally compared the distribution of the proteins from one slice of the pie chart for the one-copy strain into the pie chart of the two-copies strain [Figure 5.19C]. As an example, the first pie chart of this panel indicates that proteins that were downregulated 1 to 1.5 times in the

one-copy strain show an increase in repression in the two-copies strain. Indeed most of these proteins are repressed more than 1.5 times in the two copies strain. The other pie charts from this panel shows that a few proteins were less repressed in the two-copies strain than in the one-copy. However, the general behavior of the repression show that the more a strain expressed the heterologous protein, the more these selected proteins were repressed.

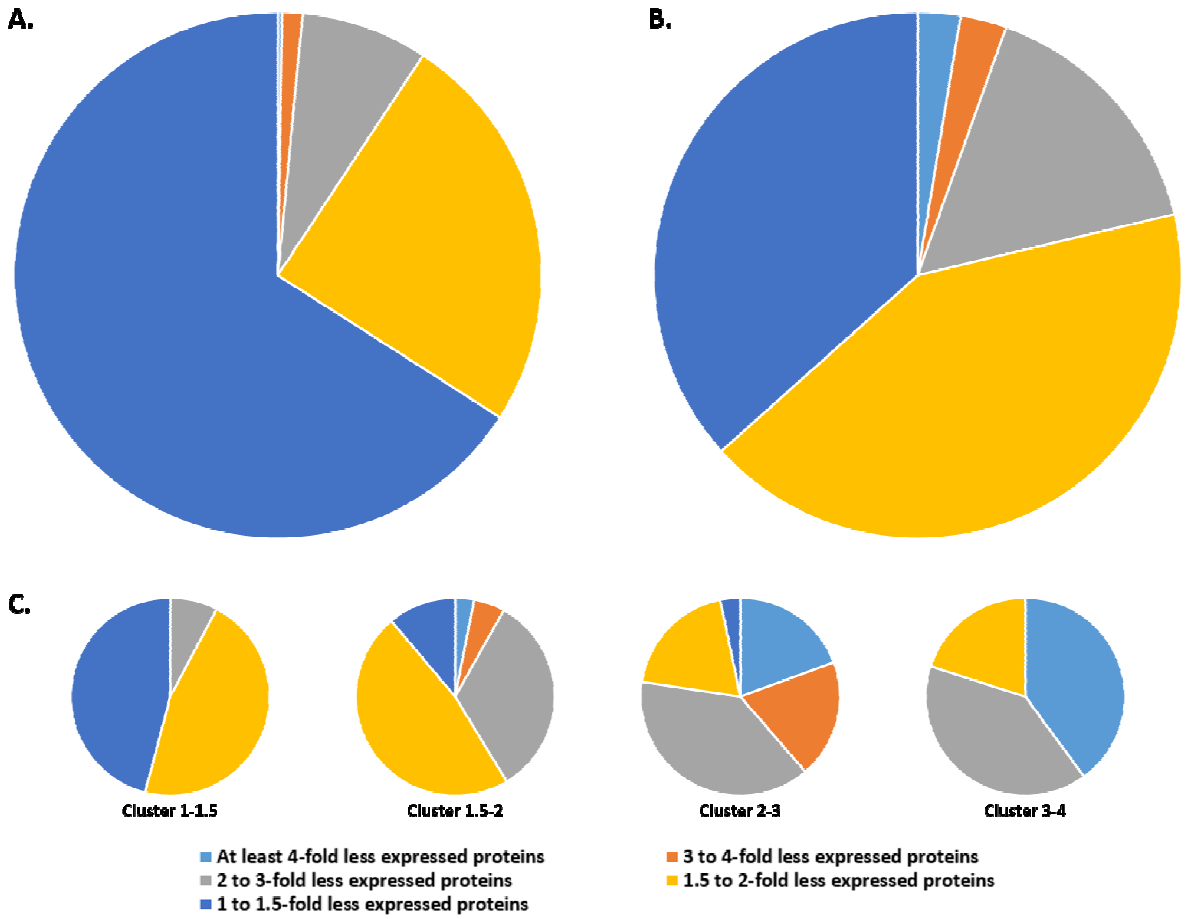

Figure 5.19 Pie chart representation of the number of differentially produced proteins in GFP overproducing strains depending on the ratio between the level of one protein in the wild-type strain compared with the GFP overproducing strains

(A) Repression in strain carrying one copy of P_{veg}-*sfgfp*(strain MZ139) (B) Repression in strain carrying two copies of P_{veg}-*sfgfp* (strain MZ140) (C) Level of repression of proteins in the strain MZ140 clustered by their level of repression in the strain MZ139

This general repression behavior indicates a potential economy of the cell resources to maintain a rate of growth as close as possible to that of the control strain. We analyzed that the underlying mechanisms of this 'possible economy' in more details to understand what part of the metabolism is potentially affected by the overexpression.

5.4.1.3 Overexpression of GFP impacts the production of non-ribosomal peptide synthetases (NRPS)

A list of the most differentially produced proteins in the overexpressing strains *versus* the control strain has been generated based on the spectral counts data given by the PAPPSO platform. In order to keep the most accurate data, we decided to focus on the 300 most detected proteins based on the total number of spectral counts across all samples. This protein list has been plotted in a heat map using the $\log_{10}$ ratios of the mean spectral counts for the biological replicates of a given strain divided by the mean of the spectral counts for all the samples analyzed [Figure 5.20 – left side]. In this figure, the list of the proteins has been sorted by the decreasing order of the $\log_{10}$ ratios for the strain carrying two copies of the *gfp* expression cassette.

Generally looking at the heat map, we observe that many proteins are less produced in the GFP over producing strains compared to the wild type strain. Moreover, we can observe the difference in the proteins distribution between the GFP strains. The upper part of the heat map reveals the most affected proteins when overexpressing a GFP. We made a zoom on the 10 most 'repressed' protein in the two copies strain (MZ140) [Figure 5.20 – right side]. Interestingly among this list of ten proteins, eight belong to enzymes involved in the synthesis of non-ribosomal peptides such as plipastatin synthesis, surfactin synthesis and enterobactin (also named bacillibactin) synthesis. The genes coding these non-ribosomal peptide synthetases are organized in operons. Detecting proteins from the same operons that are differentially expressed in the same way between the various strains confirmed the significance of the proteomics analysis.

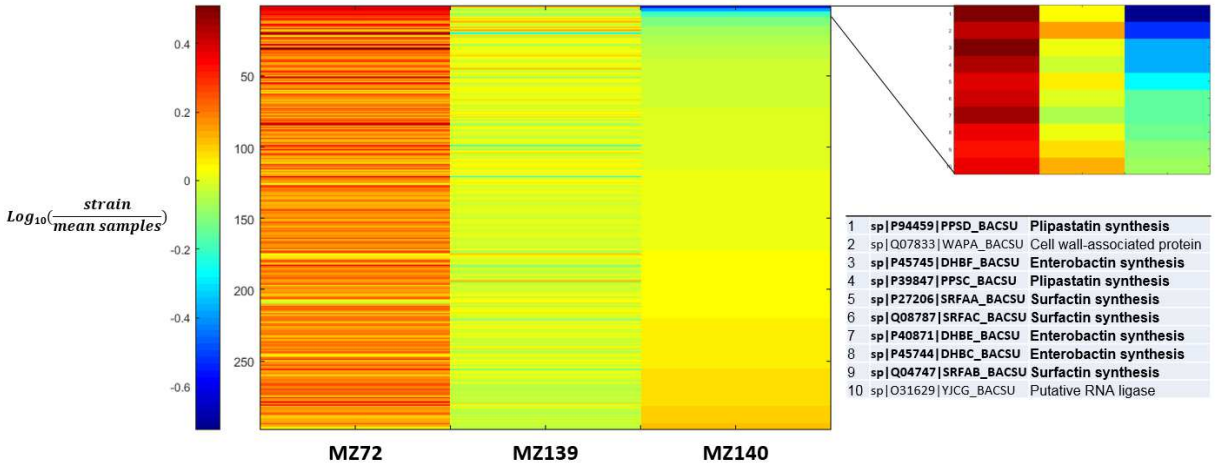

Figure 5.20 Heat map representation of the 298 most affected proteins in the wild type and GFP producing strains. On the right, zoom on the 10 most affected proteins and their respective metabolic functions in *B. subtilis*

Spectral counts for the proteins of the analyzed strains were compared to the mean of spectral counts in all the strains and $\log_{10}$-ratios were used for visualization. MZ72 Control strain, MZ139 P_{veg} *sfgfp* one copy, and MZ140 P_{veg} *sfgfp* two copies.

NRPS (non-ribosomal peptide synthetases) are very large proteins. Indeed, among the four proteins coded by the *srfA* operon, two proteins have a molecular weight higher than 400 kDa, four out of the five proteins coded by the *pps* operon are bigger than 280 kDa and the biggest protein encoded by the *dhb* operon has a molecular weight of 263 kDa. The compounds produced by enzymes encoded by these operons are small non-ribosomal peptides that have shown various biological activities as for example antibiotics, iron transport and active surfactant (Caulier et al. 2019).

Proteins encoded by the NRPS operons are not the most represented proteins in terms of protein number. However, the size of these proteins makes of it heavy proteins in terms of resources for the cells in these growth conditions as revealed in [Figure 5.21].

As shown on [Figure 5.21], [Figure 5.22] and [Figure 5.23], they represent the mass of the proteins as a percentage of the total mass with respect to their concentration as a percentage of the total concentration. The relative respective cost of SrfAA and SrfAB for example decreases together in mass with an increase in GFP production. While Ef-Tu decreases in concentration with an increase in GFP production. This behavior highlights that proteins may have the same mass, but small proteins (Ef-Tu) are present in high concentrations compared to big proteins (SrfAA or SrfAB) are present in low concentration. This indicates that the cell

manages its resources according to the cost of each protein when overproducing a heterologous protein.

We then clustered the detected proteins based on the metabolic subsystems in which they are involved. Results presented in [Figure 5.24] and [Figure 5.26] show that the major effect of heterologous protein expression result in a global decrease of the productionof native protein. The most affected metabolic pathway is related to antibiotic synthesis (due to surfactin synthetase and plipastatin synthetase downregulation). However, these figures also highlight a potential overexpression of other metabolic pathways obviously linked to the heterologous protein production, with for example, an increase in the proportion of ribosomal proteins with respect to the overall protein content of the control strain.

Looking at the zoom on the heat map presented in [Figure 5.20], it is obvious that the level of production of the 10 most affected proteins decreases with an increasing level of production of the heterologous protein sfGFP. We analyzed this decrease in more detail by making the boxplot presented in [Figure 5.26, Figure 5.27 and Figure 5.28]. The upper part of these three figures represents the genetic organization of the operon. The lower part shows relative quantity of each of the detected proteins based on the spectral counts obtained by proteomics.

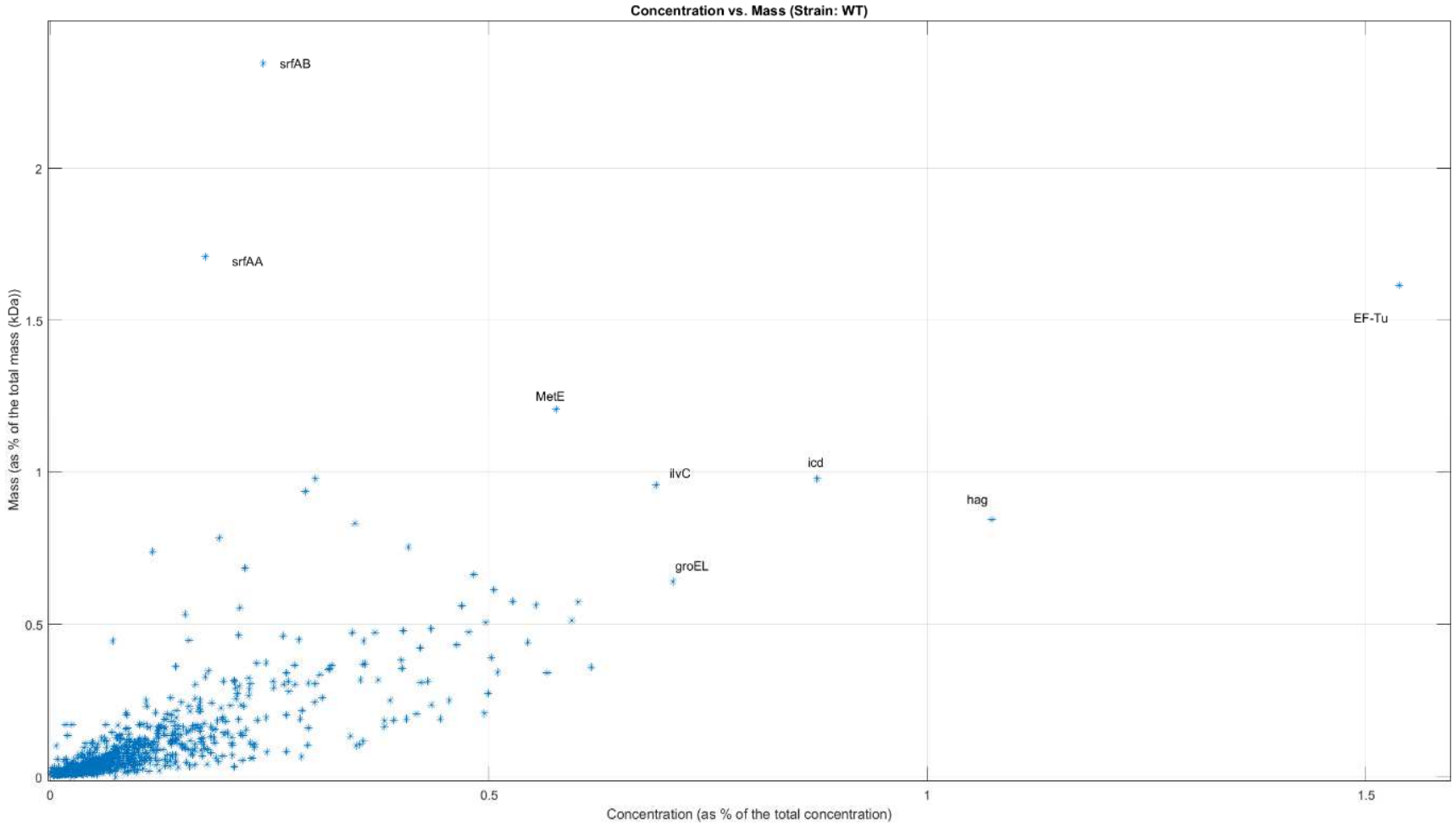

Figure 5.21 Correspondence between the normalized concentration and the normalized mass in the control strain (MZ72). Large proteins have a significant impact in terms of protein cost with a low concentration such as srfAB (401 kDA) or srfAA (402kDa). Small proteins such as the elongation factor Ef-Tu (tfuA gene - 43kDa) with a concentration ten times higher than srfAA has a cost in the order of magnitude of that of srfAA

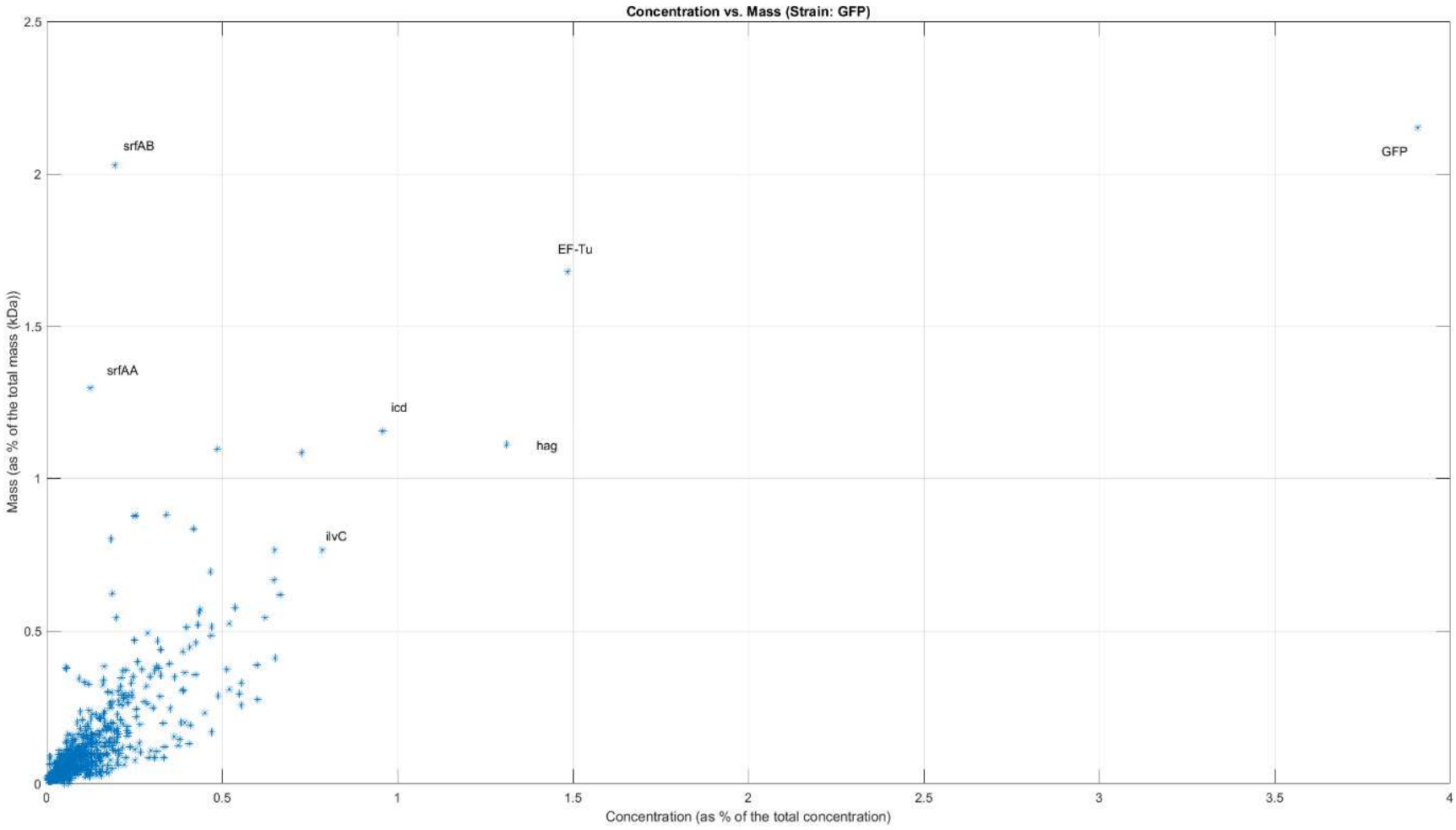

Figure 5.22 Correspondence between the normalized concentration and the normalized mass in the one copy GFP strain (MZ139)

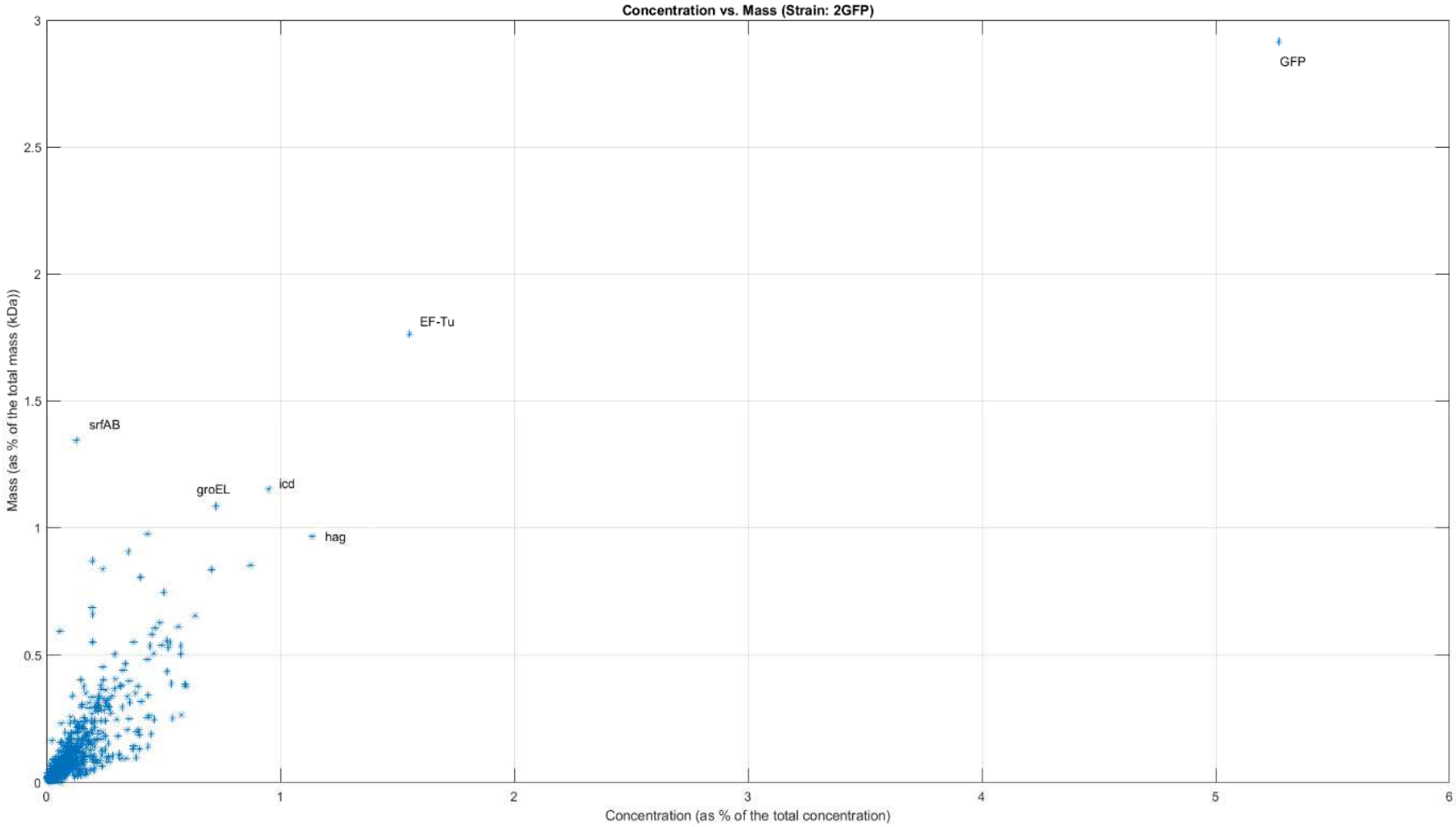

Figure 5.23 Correspondence between the normalized concentration and the normalized mass in the two copies GFP strain (MZ140)

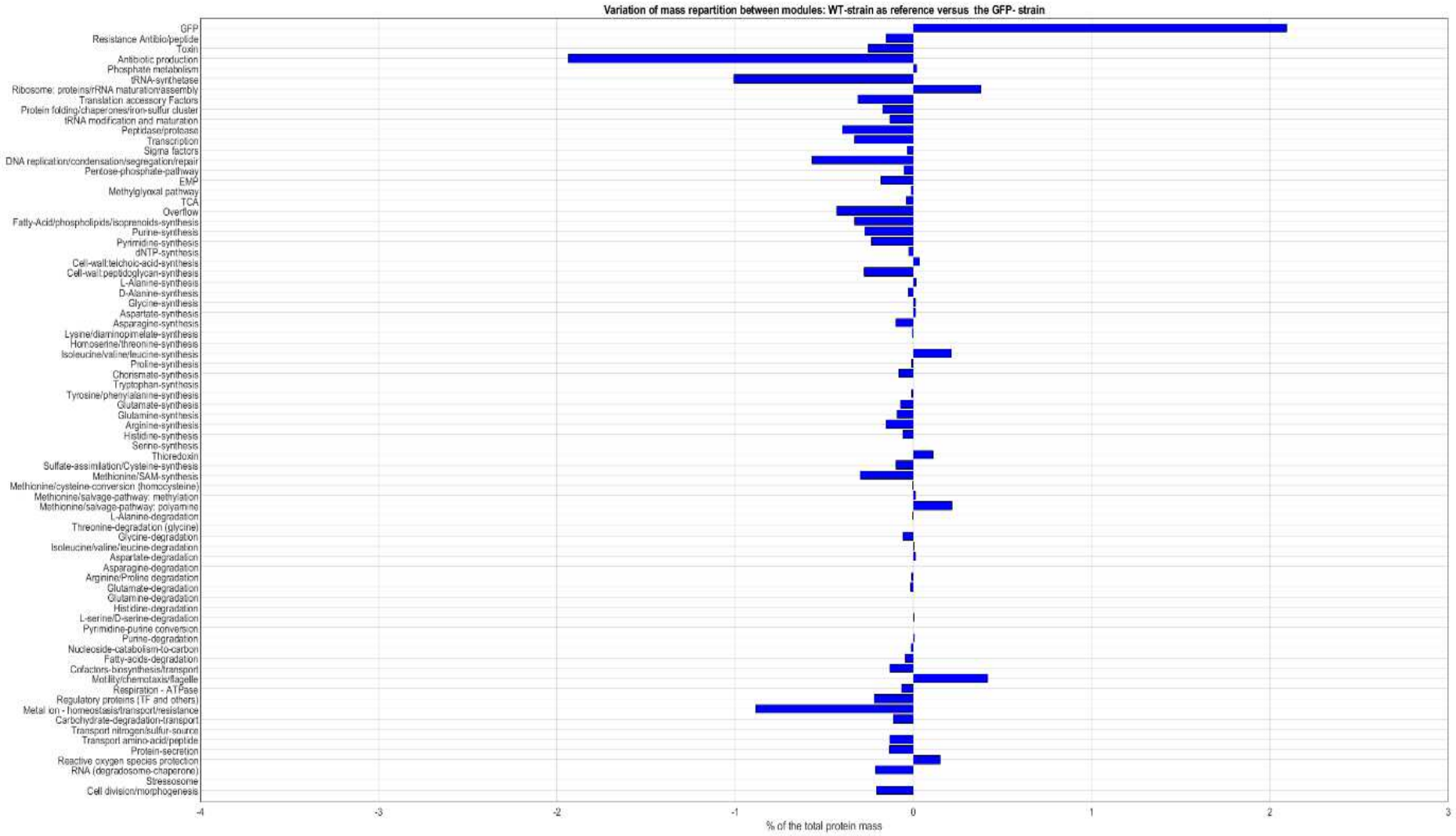

Figure 5.24 Protein mass repartition between the one GFP copy strain (MZ139) and the control strain (MZ72)

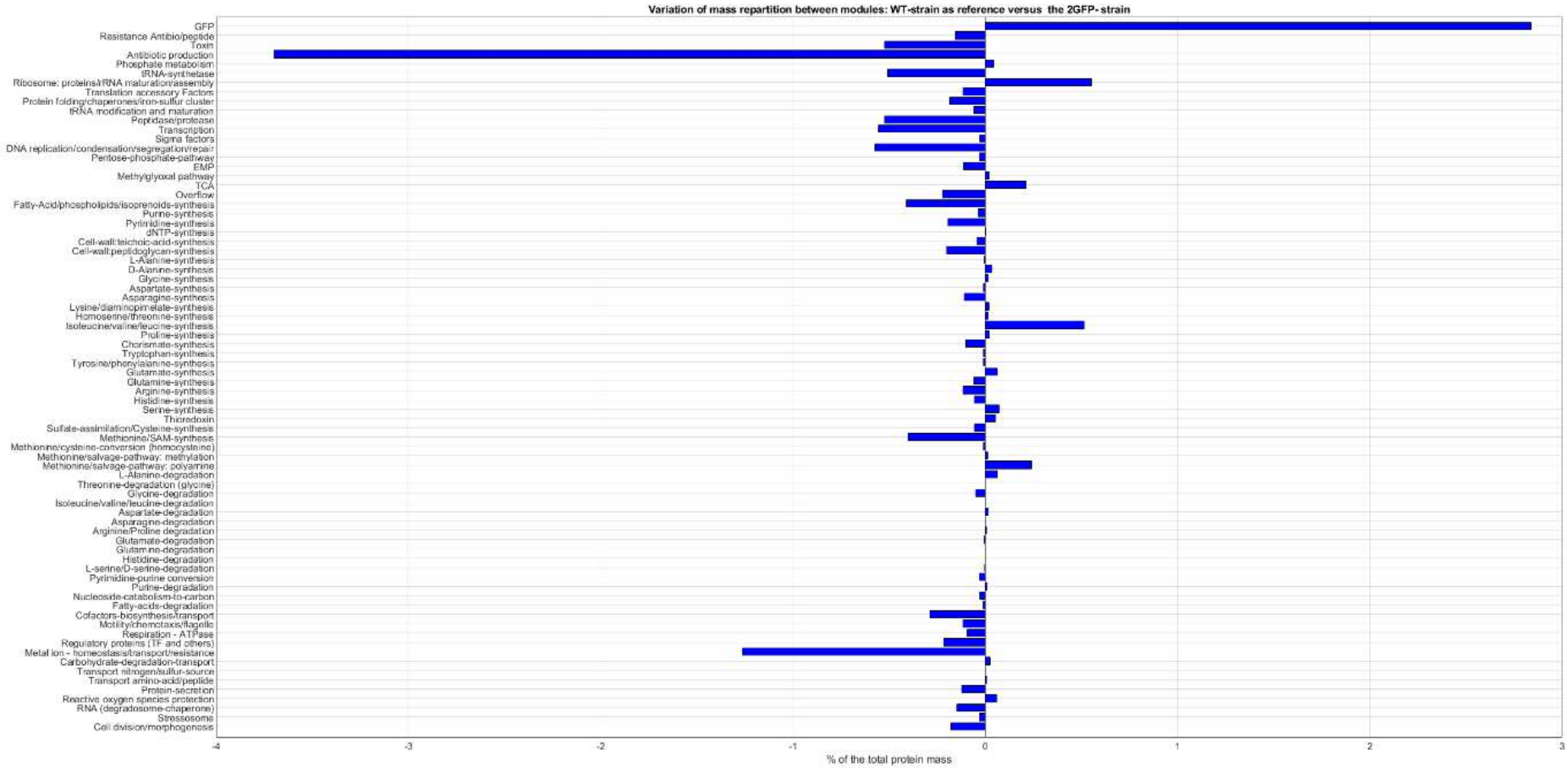

Figure 5.25 Protein mass repartition between the two GFP copies strain (MZ140) and the control strain (MZ72)

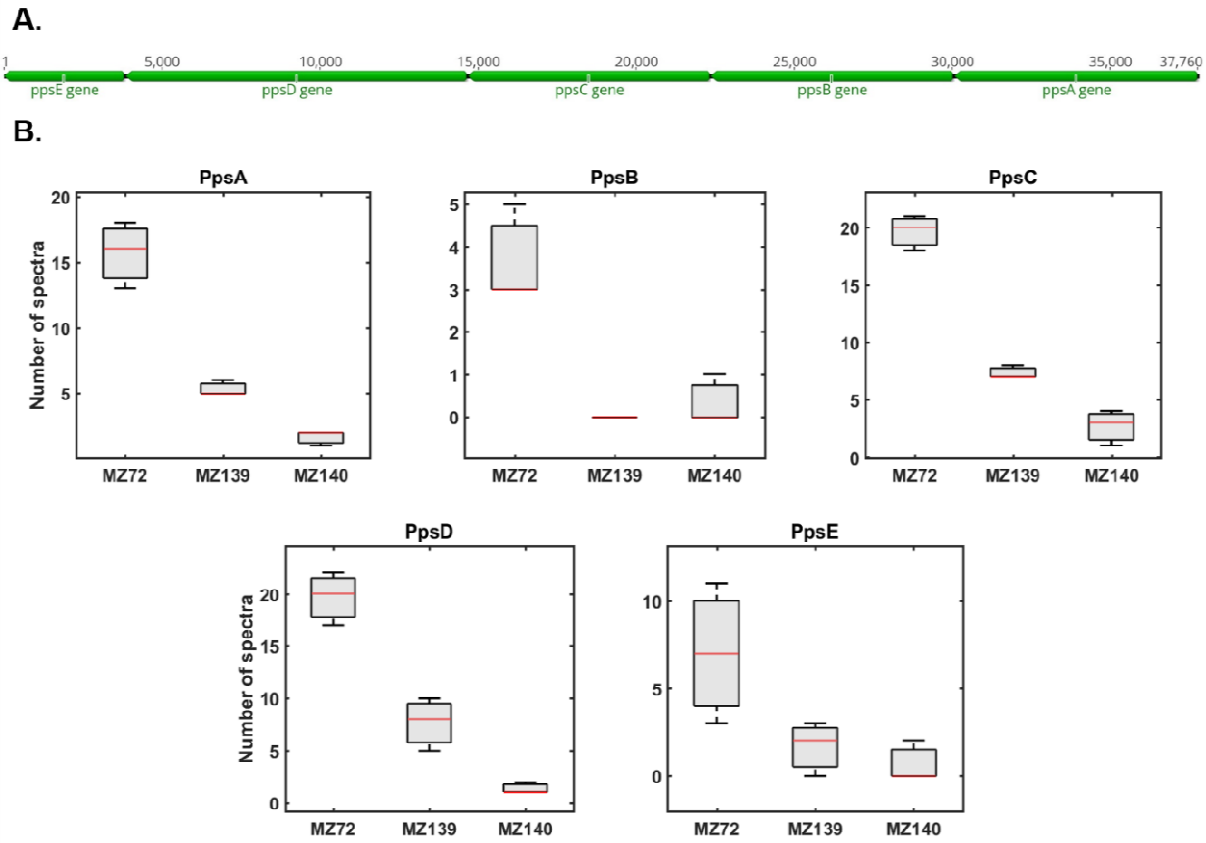

Figure 5.26 Influence of GFP overexpression on the production of proteins encoded by the *pps* operon responsible for plipastatin production in *B. subtilis*

A.Genetic organization of the operon in *B. subtilis* chromosome. B. Boxplot representation of the spectral counts for each protein encoded by the *pps* operon in MZ72 (control strain), MZ139 (one copy of the GFP expression cassette) and MZ140 (two copies of the GFP expression cassette)

Figure 5.26 represents the data for the plipastatin synthetase complex. This complex is encoded by an operon of approximately 38 kb divided in 5 open reading frames. All the proteins of this complex have been detected in the control strain and it is clear that four out of five proteins from this operon decreased with increasing GFP production. Looking at PpsB, the very low number of spectral counts makes it very difficult to detect in the GFP producing strains. Thus, evolution of the PpsB level between the two strains MZ139 and MZ140 is not significant since we are in the background of the technology.

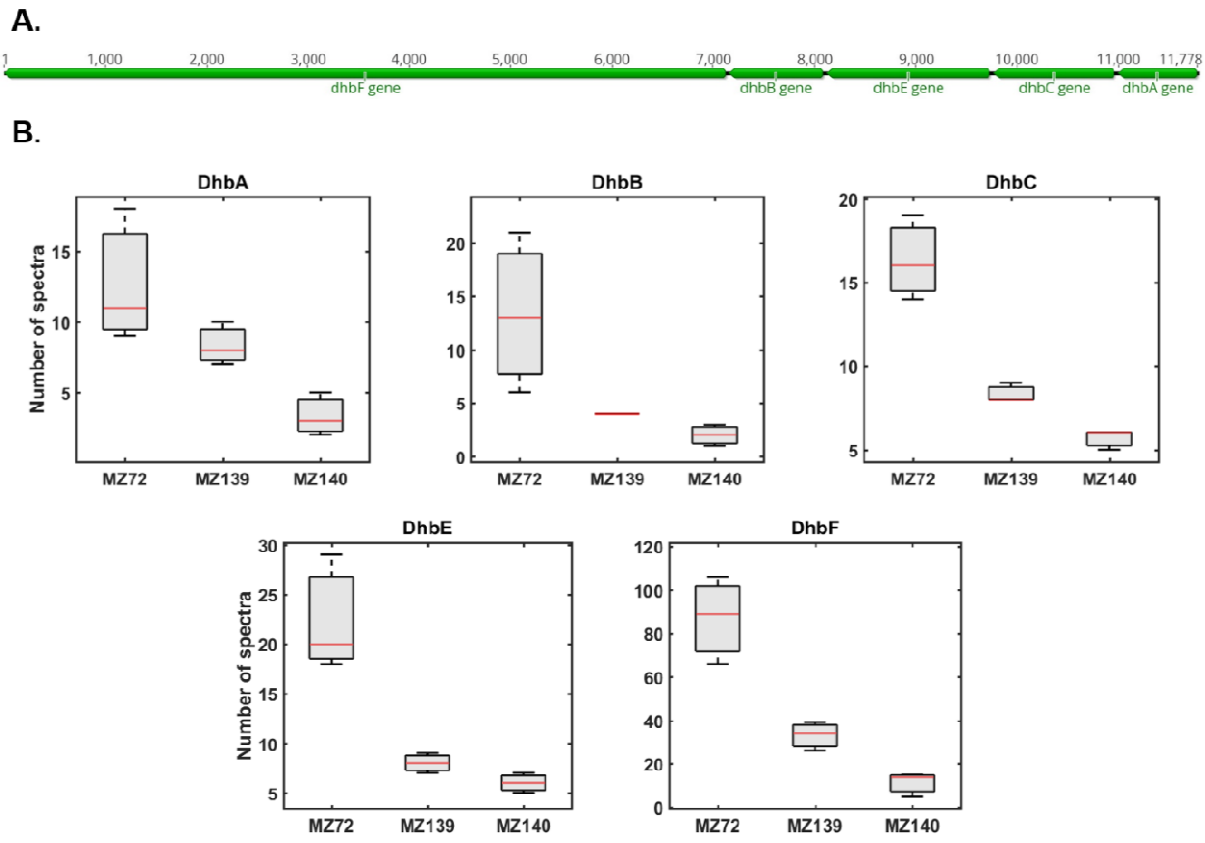

Figure 5.27 Influence of GFP overexpression on the production of proteins encoded by the *dhb* operon responsible for enterobactin production in *B. subtilis*

(A) Genetic organization of the operon in *B. subtilis* chromosome. (B) Boxplot representation of the spectral counts for each protein encoded by the *dhb* operon in MZ72 (control strain), MZ139 (one copy of the GFP expression cassette) and MZ140 (two copies of the GFP expression cassette)

Figure 5.27 represents the data for the enterobactin (or bacillibactin) synthetase complex. This complex is encoded by an operon of approximately 11 kb divided in 5 open reading frames. All the proteins of this complex have been detected in the control strain and as for the proteins of the plipastatin complex synthetase, the protein relative quantity decreased with increasing GFP production.

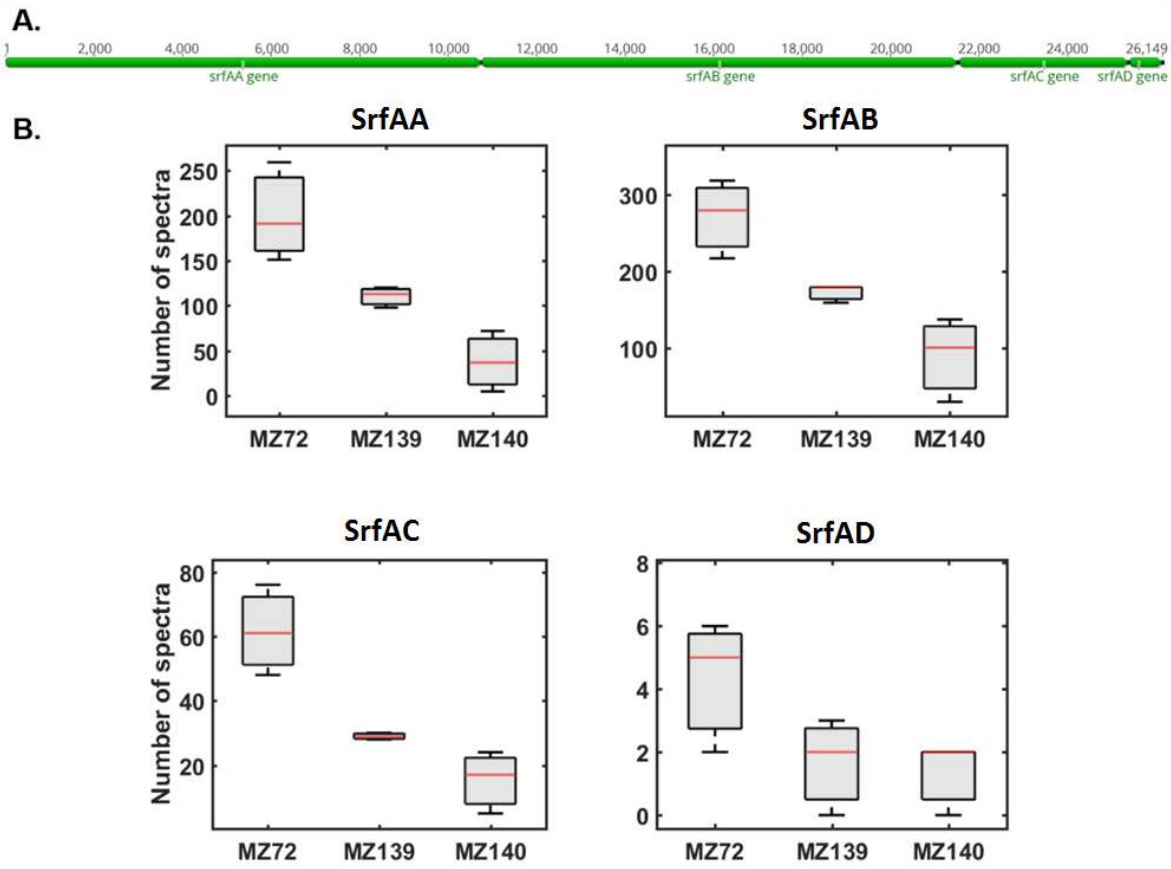

Figure 5.28 Influence of GFP overexpression on the production of proteins encoded by the *srfA* operon responsible for surfactin production in *B. subtilis*

(A) Genetic organization of the operon in *B. subtilis* chromosome. (B) Boxplot representation of the spectral counts for each protein encoded by the *srfa* operon in MZ72 (control strain), MZ139 (one copy of the GFP expression cassette) and MZ140 (two copies of the GFP expression cassette)

Figure 5.28 represents the data for the surfactin synthetase complex. This complex is encoded by an operon of approximately 26 kb divided in 4 open reading frames. Again, all the proteins of this complex have been detected in the control strain and as for the proteins of the two others NRPs operons, the protein relative quantity decreased with increasing GFP production.

The fact that the production of these proteins were reduced when a gratuitous protein was overproduced may indicate that the pool of amino acids was limiting in terms of resource allocation to feed the production of the essential and non-essential proteins including our protein of interest. To address this issue we decided to tag the GFP protein to be degraded and allow the recycling of the amino acids used during GFP synthesis.

5.4.2 Degradation of the overproduced gratuitous protein restores the growth rate

In *E. coli* and *B. subtilis*, when a ribosome reaches the 3' end of an mRNA without reaching a stop codon, the tRNA encoded by the *ssrA* gene adds a tag called the ssrA tag. In *B. subtilis*, the ClpXP protease recognizes the ssrA tagged protein and proceed to its degradation. We decided to make use of this property and generate a ClpXP-based degradation system to specifically degrade the overproduced gratuitous protein and release amino acids into the cell without relieving the burden on the ribosome load.

5.4.2.1 Design, construction and characterization of GFP-SsrA variants

Based on previously published results (Griffith & Grossman 2008; Guiziou et al. 2016), we built strains producing GFP tagged at their C-terminal with five candidate ssrA tags (ALGG, DDAS, ADAN, ADCS, AASV) in the sequence (AANDENYSENY or AGKTNSFNQNV recognized by the ClpXP protease; the use of these two sequences will be explained in the [section 6.1]**.** The synthetic cassettes carrying the P_{veg} *sfgfp-ssrA** were integrated in the *amyE locus* (ssrA* stands for the various above-mentioned ssrA tags) [from now on we will use the term GFP instead of sfGFP]. The growth and fluorescence of GFP of the different strains were monitored by LCA. As shown on [Figure 5.29], we observed that the GFP abundance of any of the GFP-ssrA* variants was lower than that of the GFP.

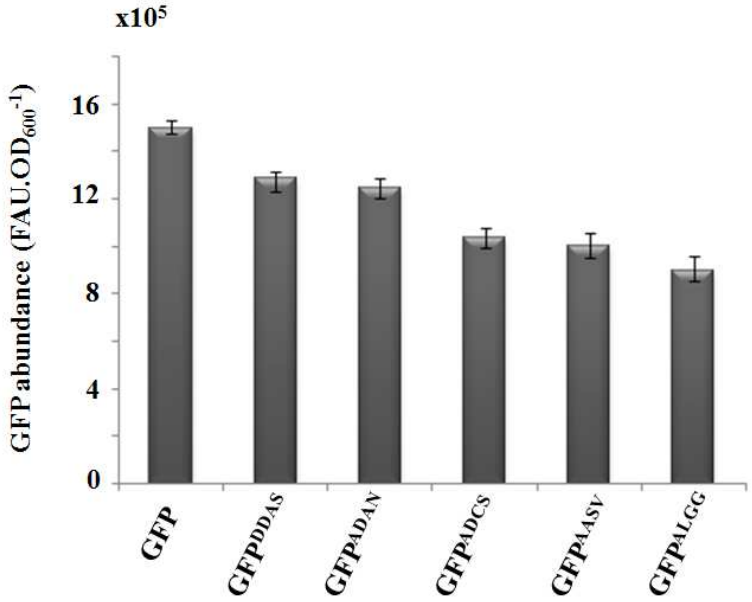

Figure 5.29 Variation of GFP abundance between the untagged and tagged GFP.

Five engineered ssrA tags whose last 4 aminoacids are DDAS, ADAN, ADCS, AASV, and ALGG, recognized by ClpXP protease were fused to the C-terminal of GFP.

The GFP abundance was estimated by the level of fluorescence, assuming that the transcription and translation processes, as well as the mRNA stability, were not altered by the C-terminal ssrA tag addition. However as it is a critical assumption for the current work, we checked whether the lower GFP-emitted fluorescence was related to the degradation of the GFP-ssrA variants by monitoring fluorescence upon the addition of a translation inhibitor on exponentially growing cells. Growth and translation inhibition was carried out using tetracycline. If the GFP-ssrA variants are less stable than the GFP (which is considered to be highly stable; half-life over 10h), we expect to observe a quick drop in fluorescence after tetracycline addition. This would indicate that the C-terminal tag is well recognized by ClpX.On the contrary, if the GFP and GFP-ssrA variants are not degraded, we would observe a constant fluorescence following tetracycline addition. The consequence would be that thedifference in the fluorescence level observed on [Figure 5.29] were not due to the GFP-ssrA degradation but to an altered efficiency of GFP production (transcription and/or mRNA stability and/or translation) by the C-terminal tag.

We selected one variant of GFP-ssrA (ALGG, later on referred to as GFP^{ALGG}), which showed the lowest level of fluorescence in S medium grown cells and tested three different concentrations of tetracycline (1, 10 and 20 $\mu g.mL^{-1}$). Although the three different concentrations of tetracycline gave rise to similar results concerning the degradation of the GFP and GFP^{ALGG} variant, we will focus our analysis on the use of 10 $\mu g.mL^{-1}$, since the use of 1 $\mu g.mL^{-1}$ of tetracycline turned out to be too low as cells started to regrow after a 2 h growth arrest and the use of 20 $\mu g.mL^{-1}$ of tetracycline turned out to be too high as cells lysed quickly after tetracycline addition. When we added 10 $\mu g.mL^{-1}$ of tetracycline on S medium exponentially growing cells, cell growth ceasedrapidly but cells remained intact [Figure 5.30A]. As shown on [Figure 5.30 B], the fluorescence in the GFP strain remained constant, while it quickly decreased in the GFP^{ALGG} strain. This result indicated that the GFP^{ALGG}variant is degraded by the active proteases.As shown on [Figure 5.30 C], following tetracycline addition the specific production rate of GFP decreased down to zero in one hour, indicating that it took about one hour to complete translation and folding of the remaining GFP. On the contrary, the specific production rate of GFP^{ALGG} became negative in less than 20 minutes, indicating a rapid degradation of the protein. The degradation of GFP^{ALGG}went on for about 2 h following drug addition in a nonlinear manner and then stopped. Based on the fluorescence quantified during the exponential growth (before tetracycline injection) and assuming the absence of degradation of the GFP, we estimated by applying the mass

conservation law the turnover rate of GFP^{ALGG} at *approx.* $0.59 \pm 0.08\ h^{-1}$. Upon tetracycline addition, this turnover rate seemed to have quickly decreased down to 0.15 ± 0.08 h-1 for about two hours before the GFP^{ALGG} degradation ceased.The GFP^{ALGG} degradation arrest is consistent with the fact that ClpXP is an ATP-dependent protease, which requires the constant production of ATP to remain fully active.

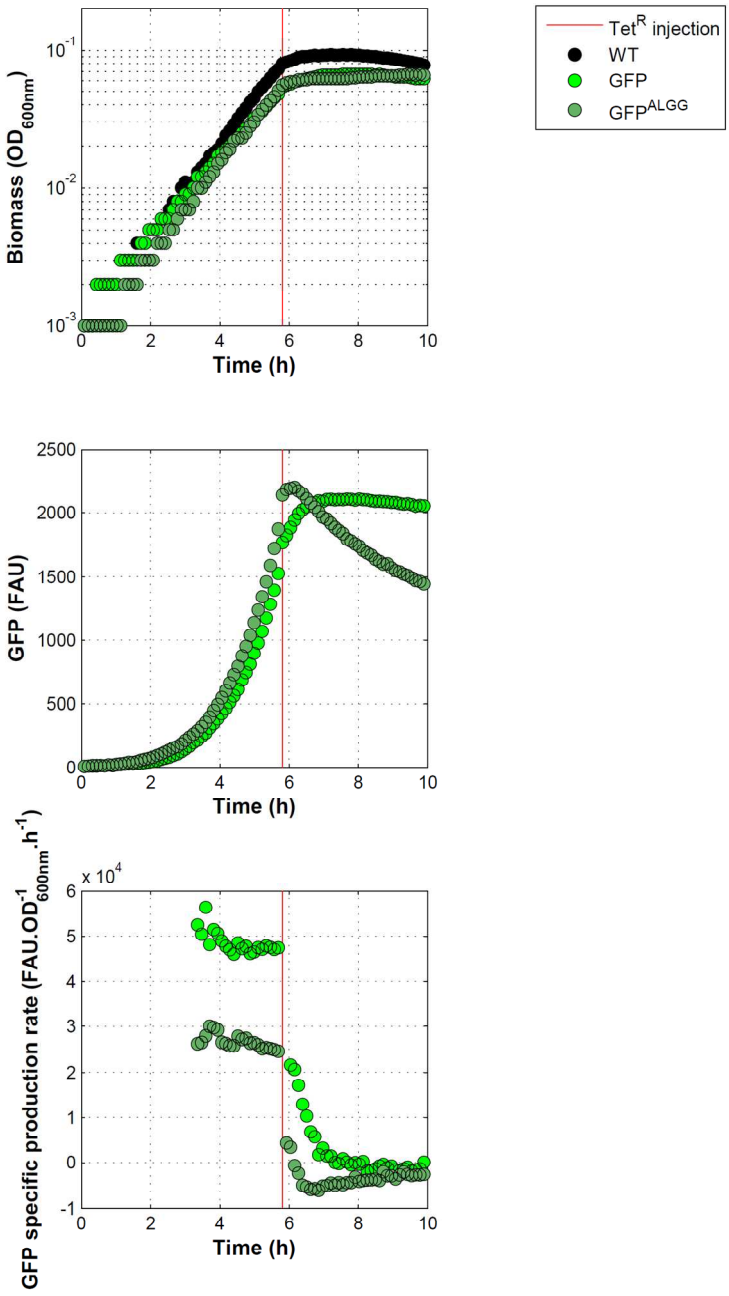

Figure 5.30Addition of tetracycline on exponentially growing cells reveals the quick degradation of the GFP^{ALGG} as compared to the unmodified GFP

(A) A representative growth curve of WT, GFP and GFP^{ALGG} strains (each in 6 replicates). The growth ceased after 10 $\mu g.mL^{-1}$tetracycline injection. (B) A fluorescence curve for the two strains showed that after injection the fluorescence of the GFP in the GFP strain remained constant, while it decreases in the GFP^{ALGG} strain. (C) The specific production rate of GFP was computed for the GFP and GFP^{ALGG}

strains, and showed that within less than half an hour the GFP^{ALGG} was significantly degraded (<0), while the GFP was stable for hours (=0).

5.4.2.2 Degradation of the synthesized gratuitous protein restores the growth rate

In order to specifically degrade the overproduced gratuitous protein and release amino acids into the cell without relieving the burden on the ribosome load, we tested the ClpXP-based degradation systemon both GFP and mKate2. The growth rate was monitored in different defined media (S, SX, and CHG) [Figure 5.31 A, B, C respectively]. When specifically targeting GFP^{ALGG}, we observed a growth rate restoration mainly in S and SX media. The growth rate restauration was of about 50%. When degrading the $mKate2^{ALGG}$ a growth rate restoration was significantly observed in the S, SX, and CHG media. We concluded that the degradation of the newly synthetized gratuitous proteins permitted the recycling of amino acids to be reused by the cell, so that it allowed growth rate increase. Hence, the resulting growth rate restoration indicated that the burden caused on the cell by the gratuitous protein was mainly due to the limitation in amino acids.

5.4.2.3 The mass quantification of GFP^{ALGG} is half of that of GFP

A western blot against GFP was performed to estimate the quantities of GFP and GFP^{ALGG} in the extracted proteins. To do that, a scale of quantified purified GFP was prepared to estimate the amount of the loaded GFP and GFP^{ALGG} [Figure 5.13 B]. From a total cell extract of ~6.75µg in 20 µL total volume, the GFP amount was estimated to be of ~0.66 µg while the GFP^{ALGG} amount was of 0.32 µg. As a result, the GFP represented 9.77% of the total extracted proteins, while the GFP^{ALGG} represented 4.82%. This result agrees with the LCA measurements for the GFP abundance, where that of GFP^{ALGG} was more or less half that of the GFP.

This quantification allows us to confirm that with partial degradation of the GFP, we were able to observe a partial restoration of the growth rate. This indicates that the recycled amino acids were enough to increase the growth rate, even if partially restored. Together with the proteomics data, which gave the hint that some non-essential proteins of big size were downregulated when a gratuitous protein was overproduced, we confirmed that amino acids represent an important part/factor in cell resources.

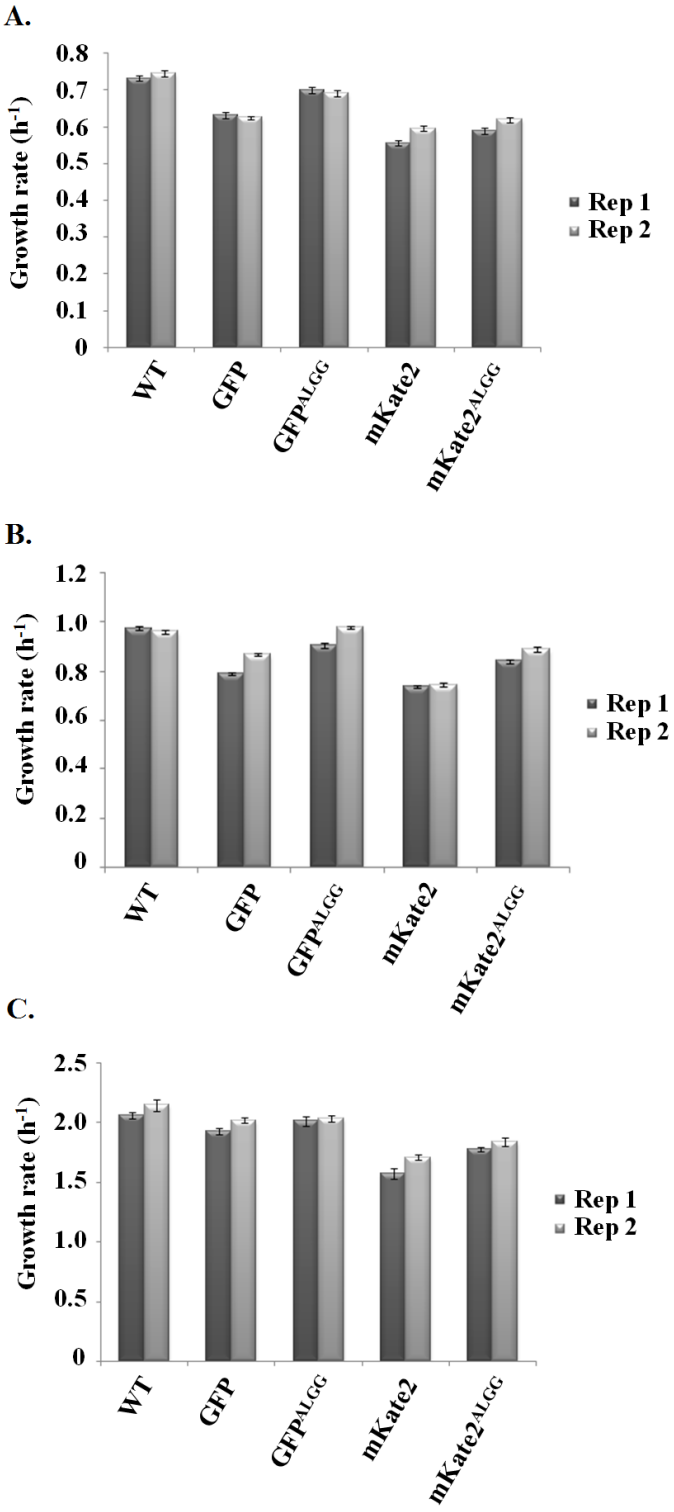

Figure 5.31 Growth rate restoration upon protein degradation

The graphs show duplicates of LCA measurements where the growth rate was followed in different media for different gratuitous proteins to know whether the degradation of the proteins will change the growth rate. (A) S medium, (B) SX medium, (C) CHG medium.

5.4.2.4 Degrading the gratuitous protein changes the cell size

With the partial degradation of the GFP^{ALGG} [Figure 5.29], we checked whether the cell size changed back to normal. The flow cytometry data [Figure 5.16 E - F] showed a decrease of the cell size towards the wild type cell size accompanied by the decrease in GFP abundance [Figure 5.32]. The proteomics data analysis showed that one of the proteins involved in the cytokinesis ring assembly, FtsZ, is decreased by 1.45 times in the GFP producing strain. This decrease might be one of the reasons that delayed cell division and resulted in higher cell size when overproducing the gratuitous protein. Therefore, the released of amino acids after degrading the overproduced protein reduced the inhibition of the synthesis of FtsZ proteins. Thus, cell division occurs earlier due to molecular mechanisms and resource rerouting towards the production of the required proteins for cell division

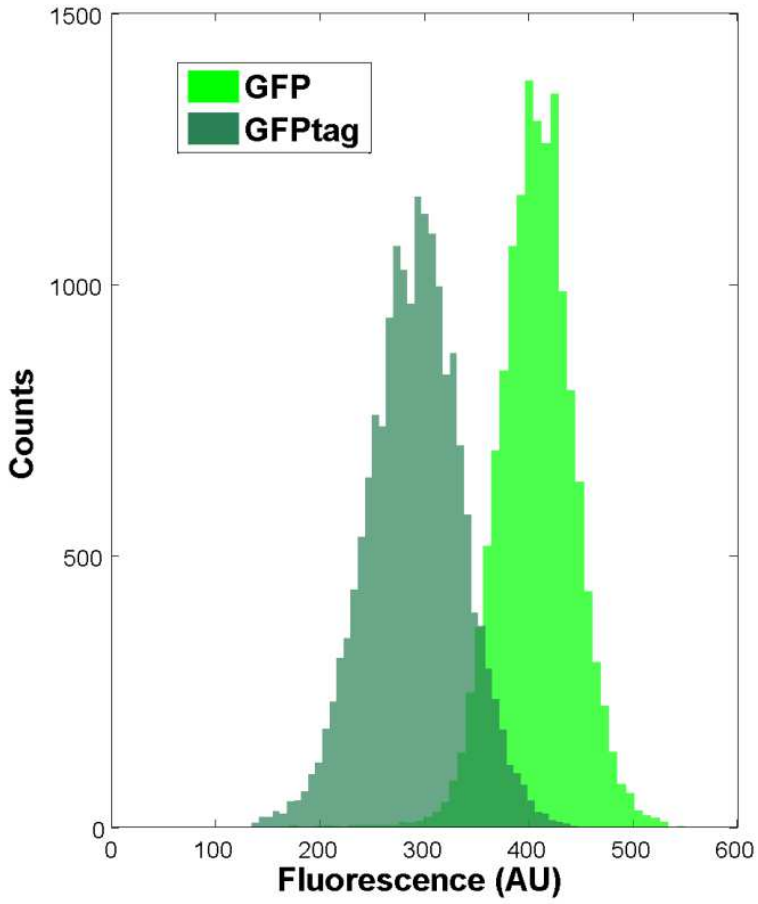

Figure 5.32Fluorescence measurement by flow cytometry

The GFP and GFP^{ALGG} fluorescence measured by flow cytometry showing lower GFP^{ALGG} fluorescence.

5.5 Discussion

5.5.1 Demonstration of the gratuitous protein over production impacts on *B. subtilis*

Engineering cell factories for protein production faces some difficulties imposed on the physiology of microorganisms. Bacteria and yeast share similar effects including cell fitness, and cell size. However, the effect of gratuitous protein production on *B. subtilis* physiology was still unknown or neglected. We found that *B. subtilis* is like other microbes, *i.e.* impacted by high protein production. *B. subtilis* responds to protein overproduction by a slower growth rate, and a bigger cell size.

In order to investigate the consequences of high protein production, we focused choosing the right expression system, the right protein, and the locus of gene integration on the genome. We had different trials on each of the mentioned points until we observed the impacts on the cell.

First, when we constructed a GFP-LacZ fusion protein, we encountered the problem of protein aggregates without showing an effect on the cell fitness. Misfolded proteins diffuse freely in the cytoplasm and tend to stick together by exposing their hydrophobic regions [Figure 5.33 A]. As the aggregates grow, they are excluded from the nucleoid due to nucleoid occlusion (Laloux & Jacobs-Wagner 2014). This mechanism occurs because the formation of big protein aggregates is energetically more favorable away from a bulky polymer such as the nucleotide [Figure 5.33 B]. As a result, the aggregates diffuse towards the cell poles. In addition, it was revealed in *B. subtilis*, that the proteolysis machinery including ClpP with its ATPases and the aggregated proteins form clusters at the polar regions of the cell (Kirstein et al. 2008). Similar observations were shown in *E. coli*, when the formation of aggregates after protein misfolding were shown to localize at the cell poles (Laloux & Jacobs-Wagner 2014; Winkler et al. 2010) where the quality control machinery (Hsp100/Clp proteases) localize.

Second, the inducible promoters did not help to obtain a production level that causes an impact even with double integrations of the same gene in the chromosome. In *S. cerevisiae*for instance, the authors of (Kafri, Metzl-raz, et al. 2016) showed a growth rate decrease with more than 18 copies of *gfp* and *mCherry* genes. Indeed, the facilities to integrate many copies in *B. subtilis* are not available and it is possible to face gene recombination problems.

Later, by using the P_{veg} promoter which is 6-fold stronger than P_{hs}, the growth rate decreased in different media and was confirmed with 3 reporter proteins. For instance, two copies of P_{veg}*lacZ* decreased the growth rate by about 30% in poor medium with glucose (S) and by about 13% in rich medium (CHG).

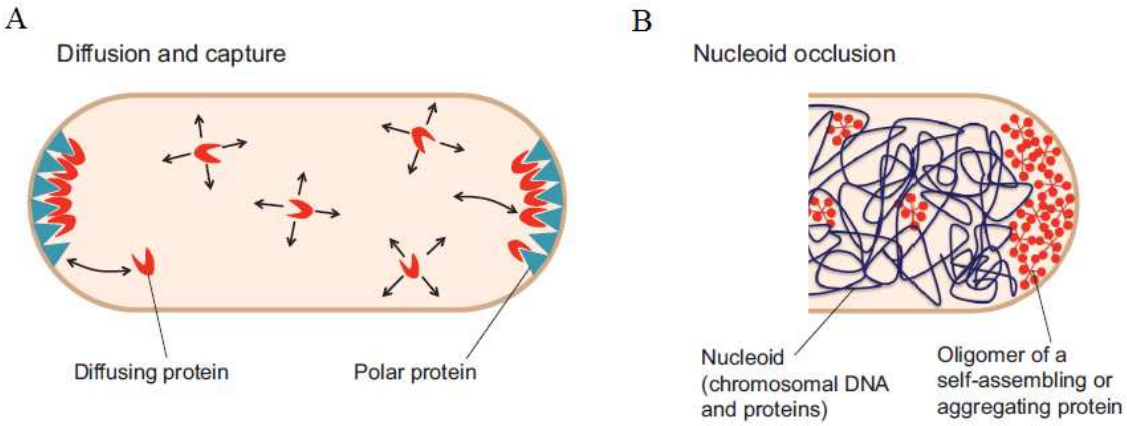

Figure 5.33 Cell pole localization of protein aggregates.

(A) A Protein diffuses in the cytoplasm until it meets a polar protein with which it has a binding affinity. (B) Large protein aggregates are excluded to the cell poles away from the bulky nucleotide in a mechanism called nucleotide occlusion.

5.5.2 Overproduction of a heterologous protein leads to a reduction in non-essential endogenous proteins of the NRPS family

Considered as non-essential, NRPS operons encode enzymes from metabolic pathways that will finally be useful for adaptation of the cells to their environment (competition with other microorganisms, biofilm formation (Fan et al. 2017) or optimization of iron uptake for example). These adaptation capacities are probably not essential for optimizing cell fitness in lab culture conditions. That may allow the cell to save resources when producing other proteins (heterologous or native) by decreasing the production of these heavy (in term of mass and cost) proteins.

Moreover, non-ribosomal peptides synthesis consumes metabolites that represents a cost for the cell. For example, surfactin is a lipopeptide made of seven amino acids and a β-hydroxylated fatty acid with a length of 13 to 16 carbons. Plipastatin is also a lipopeptide composed of 10 amino acids and of a β-hydroxylated fatty acid with a length of 14 to 18 carbons. Down-regulation of non-ribosomal peptide synthetase can thus be a strategy for *B. subtilis* to save energy by limiting both the cost of the synthesis of enzymes and the use of

their substrates (mainly amino acids or amino acids precursors) for production of the non-ribosomal peptides. However, if such a regulation of the synthesis of NRPS exists in *B. subtilis*, the strains used in this study derived from the 168 strain, which have been reported to not produce surfactin or plipastatin. Indeed, a mutation causing a frame-shift in the *sfp* gene encoding the phosphopantetheine transferase Sfp prevent the production of surfactin in this genetic background. In the same way, a substitution in the -10 box in the sequence of the *degQ* promoter prevent its expression which is required for plipastatin production (Gao et al. 2017). The impossibility of producing these NRPs in our strains makes it impossible to safeguard resources related to the production of these peptides. Thus if cells have 'coded' a mechanism to reroute resources by decreasing NRPS synthesis, in a 168 derived strain, the main effect will come from saving the resources by a decrease in the synthesis of these specific enzymes.

5.5.3 The amount of the gratuitous protein with respect to thetotal soluble protein determine the impact on the growth rate

Our results show a direct relation between protein production and the impact on the growth rate depending on the medium. The higher the level of protein production is, the higher the impact is as the medium tends to be poorer. This result agrees with the result of (Malakar & Venkatesh 2012) where it was shown a higher burden when increasing the production level of the LacZ protein in *E.coli*, or with less carbon source in the growth medium. However, Malakar and Venkatesh (2012) considered that the burden was caused by the cost of protein production and by the cost of IPTG uptake through the proton pumps. In our case, the gratuitous protein encoding genes were expressed constitutively, which allows us to conclude that it is the protein production itself that caused the burden.

Recent studies on *B. subtilis* showed that there is no effect on the growth rate by producing massively a heterologous protein in LB medium (Sauer et al. 2018). They designed a large library of synthetic expression modules, and showed that there is no impact on the cell fitness even with 13x stronger expression than that of P_{veg}. However, more analysis about the quantity of protein produced must be performed in order to understand why they did not show any impact on the growth rate. First, according to their P_{veg} variant, the translation initiation region is different than the one we used and therefore the amount of produced protein might be less. Second, as it is known, the efficiency of translation decreases with increasing growth rate (Gerosa et al. 2013; Borkowski, Goelzer, et al. 2016). Our results are in agreement with this fact [Figure 5.34], so that as the medium tends to be richer (the growth rate tends to be

higher), the GFP abundance decreases. Quantification of GFP by western blot showed that the GFP ratio to the total cytosolic proteins is of 3% in CHG, whereas it is of 9% in S medium. Knowing that the results obtained in LB medium with a higher growth rate for a wild type strain than CHG, it is possible that the quantity of protein produced in ratio with the total soluble proteins is not high enough to cause a growth rate decrease.

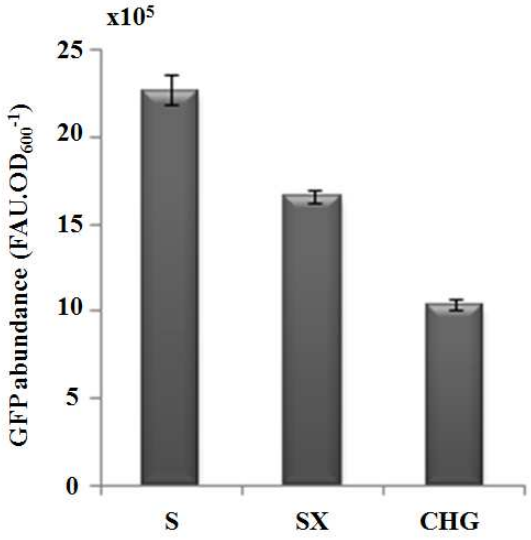

Figure 5.34 GFP abundance decreases with increasing growth rate

As a function of the growth rate represented by the growth medium, the GFP abundance clearly decreases when the medium tends to be richer.

Different quantification methodologies were used in the current work to quantify the GFP abundance and they all showed similar results. First, an estimation by signal intensity on an SDS-PAGE coomassie stained gel was applied on *B. subtilis* strains producing GFP, LacZ, and together LacZ and GFP, and an *E. coli* strain producing GFP. The intensity of the band is related to the mass of the proteins. The result showed that:

- GFP is 1.5% more produced in *B. subtilis* than in *E. coli.*
- GFP represents 9.07% of the soluble proteins.
- LacZ represents 6.92% of the soluble protein, which is 2.15% less than that of GFP, which could mean since LacZ is a bigger protein than GFP, its production is limited not only by the amino acids, but also by the cytosolic occupancy more than GFP.
- GFP (6.94%) and LacZ (5.49%) when present together are less present as compared to when they are produced alone which can be referred to the cell strategy in saving resources.

Second, a quantification made by western blot against GFP, which is based on a GFP scale of known amounts, showed that the GFP abundance is 9.77%. This result confirmed the previous estimation performed by signal intensity.

Third, we used the relative protein quantification done by PAPPSO platform to estimate GFP quantity to be ~5%Taken all together these results confirmed that between 5-10% of gratuitous protein relative to the total soluble protein abundance leads to a burden on the growth rate in *B. subtilis*.

5.5.4 Relation of cell size and high protein production

Cell size is a parameter correlated with the cell's growth rate, thereby it changes with the bacterial physiological conditions. Schaechter et al. (1958) were the first to demonstrate and discuss the increase of cell size and cell composition with increasing growth rate (Schaechter & Kjeldgaard 1958). This observation was explored throughout the years to understand the principles of the growth law and the correlation between the nutrient availability, the molecular mechanisms that control the cell size and composition, and the growth rate. On the other hand, and in contrast to the growth law, the cell size behavior towards high protein production is the inverse state.

We have reported that when forcing cells to produce high amounts of proteins, it resulted in larger cells. The relation between the bacterial cell size and gratuitous protein overproduction is common in microorganisms. As in *B. subtilis*, the same relation applies on *E. coli* and baking yeast (Basan et al. 2015; Kafri, Metzl-Raz, et al. 2016). Normally, the relation of cell size is proportional to growth rate depending on the medium nutritional quality (Schaechter & Kjeldgaard 1958). Therefore, we always expect that in a rich medium both the growth rate, and the cell size increase, and *vice versa*. However, when a gratuitous protein is overproduced, the growth rate decreases, but the cell size increases. The result was confirmed with two proteins, GFP and LacZ. We showed that the effect on the cell size depends on the protein size and cost, so that the LacZ-producing cells were bigger than GFP-producing cells. But, what is the explanation for this contradiction?

This contradictory observation was correlated with the proteome allocation (Scott et al. 2014; Weiße et al. 2015; Scott et al. 2010). When a gratuitous protein is added to the finite proteome, it changes the proteome allocation, thereby decreasing the ribosomal proteins and the metabolic proteins and enzymes. Researchers made use of the proteome allocation

principle to explain the relation of high protein production and cell size such as the authors of (Basan et al. 2015) who explained it by their threshold model this relation. Based on the proteome partitioning, they assigned a fraction for the proteins required for cell division (P_X). When an added protein takes place in the proteome, it decreases the fraction P_X so it leads to increasing cell size until P_X reaches the threshold required for cell division. Moreover, a coarse-grained theory of bacterial physiology was presented by (Bertaux et al. 2016), where they unified the proteome allocation with the structural model of division control. These two models are examples on recent provided theoretical explanations, which both agree on the decrease of structural proteins related to cell division in the presence of a gratuitous protein. An experimental validation of the threshold hypothesis was recently reported (Si et al. 2019). In this work, they concluded that cell division timing was reasoned to dynamics in proteins and precursors required for division. A threshold of accumulation of division proteins and balanced biosynthesis to maintain proportionality between their production and cell volume are required for cell size homeostasis. Although they agreed with the threshold hypothesis, more investigation on the interference of gratuitous protein production with this hypothesis are missing. How does it affect the division proteins experimentally?

According to the threshold theory, reaching the required threshold is slowed down in the presence of a useless protein, thereby leading to cell elongation [Figure 2.21]. This explanation could be true but it is quite general since the process of cell division is controlled through different means. Cell division is regulated through different cellular signals. If the only reason is the insufficient levels of proteins responsible for cell division, then the level of FtsZ ring inhibitors would not be enough as well to control the proper assembly and we would have expected very small or malformed cells. The interference of high protein production with cell size means that some 'regulations' lead to cell volume expansion in response to increase in cellular mass.

5.5.5 Amino acids are the limiting resource in a limited cytosolic density

Growth rate decrease caused by protein over-production is a result of a lack in cellular resources. The assumptions on the main limitations that stand in front of high protein production include cell processes such as translation, transcription, and protein folding. The translation process is costly by using amino acids, ribosomes, and tRNAs for a useless protein. Transcription also uses additional NTPs and RNA polymerases. Both processes were the major bottlenecks that were said to control the growth rate rather than the protein itself being produced (Stoebel et al. 2008). Moreover, the bottleneck was attributed to the growth

medium conditions that alter the relative importance of the transcription and translation processes (Kafri, Metzl-Raz, et al. 2016). For instance, when the yeast was grown in a medium with low phosphate, the limitation was attributed to transcription initiation. Whereas in a medium with low nitrogen it was attributed to translation elongation. In addition, phenomenological models reported the main reason of resource limitations to be the finite proteome allocation (Scott et al. 2014; Weiße et al. 2015; Scott et al. 2010). The proteome allocation considers a finite proteome divided between growth rate-independent housekeeping proteins, and growth rate dependent ribosomal proteins and other enzymes. Any additional gratuitous protein takes part in the finite proteome, thus decreasing the resource allocation towards the synthesis of the ribosomal proteins and other enzymes. Consequently, this proteome re-allocation leads to growth rate decrease. Other studies narrowed their assumptions to cellular components such as competing to RNA polymerase (Gyorgy et al. 2015), the proteases (Cookson et al. 2011), and molecular chaperons (Daraba & Alma 2018).

The constraint on heterologous protein production does not seem to reside in only one compartment or one process, it could be a combination of many cell resources limitations. However, the real answer to that is not established yet. We believe that the limitation is the combination of all the above mentioned reasons, but is there a higher impact of one limitation than another? In order to investigate on the limitations, we used an engineered targeted protein degradation tool in *B. subtilis* (Griffith & Grossman 2008; Guiziou et al. 2016). If the gratuitous protein is targeted for degradation by the proteases, it will recycle amino acids back to the cell and it will liberate intracellular space. In this case, what would be the effect on the growth rate? If a growth rate restoration occurs, this would suggest a limitation in the amino acids and a higher occupation of the intracellular space are causing the growth rate decrease [Figure 5.35]. However, if there was no restoration, it would suggest more that the machinery of protein synthesis are saturated.

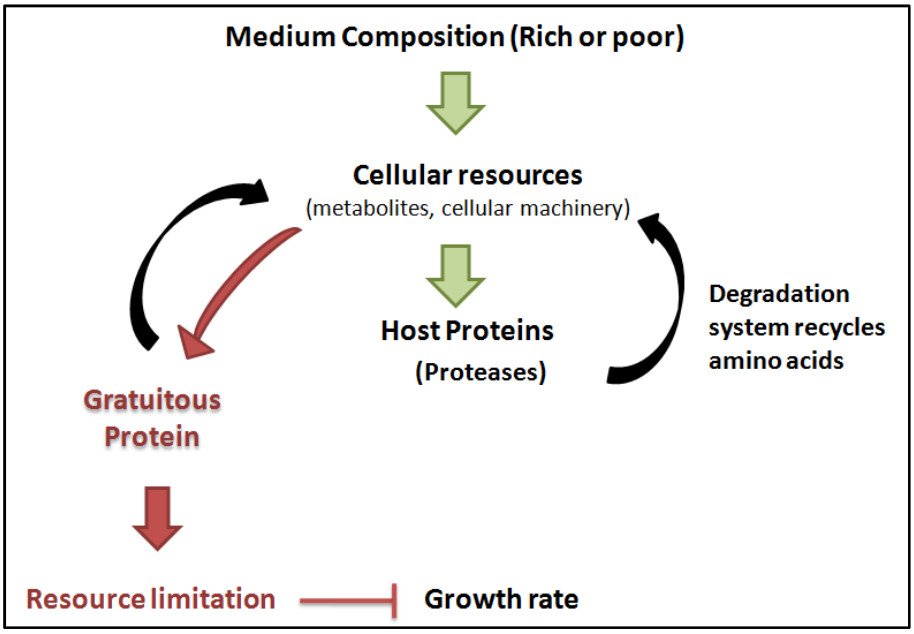

Figure 5.35 Representation of applying a targeted degradation system on the gratuitous protein

By using a targeted degradation system to degrade the gratuitous protein, we had the chance to see a growth rate restoration during protein degradation. Knowing that the ClpXP protease consumes energy in the form of ATP during degradation, nevertheless we observed a significant growth rate restoration. This means that amino acid recycling and the liberated space were more important than the ATP consumption. Actually, depending on the substrate, the ATP consumption by the protease can reach more than 500 molecules to denature a protein (Kenniston et al. 2003; Baker & Sauer 2006). Therefore, if the competition for the proteases is a bottleneck, we would not had a growth rate restoration. Hence, the proteases can be eliminated from being a main bottleneck to cause the burden of high protein production in *B. subtilis*.

The amino acid importance in protein production presents a major cell resource to form a building block of all the cell proteins. Referring to our results, the higher impact on the growth rate was observed in a poor medium (S). In this medium, the growth decrease ranged between 15-30% depending on the level and size of the expressed protein. In such medium that lacks amino acids, the cell is pushed to synthesize the necessary amino acids required to produce all the proteins. As a result, the synthesis of aminoacids exhaust the cell resulting in a low growth rate compared to richer media, and a greater effect when adding a gratuitous protein. When the gratuitous protein is degraded, the resulting growth rate restoration can be explained by the recycled amino acids which saves resources by synthesizing amino acids by

the cell. Thereby, saving enzymes and energy. This allows us to deduce that the transcription is not the main limitation of gratuitous protein production.

The amino acids were not considered as a limitation in a study made on *E. coli* that addressed a similar issue related to resource limitation (Stoebel et al. 2008). Stoebel et al. (2008) used the targeted degradation system against LacZ and monitored the growth rate. LacZ activity decreased by 38.9%, but it had no effect on the cell fitness. Therefore, they assumed that amino acids are not a limiting cell resource. Knowing that LacZ is a big and stable protein, it might not be easy to degrade it which is clear from their result that the activity decreased by less than 50%. In addition, the amino acids recycling resulting from the LacZ degradation might not be enough to allow growth rate restoration. In our case, besides GFP and mKate, we targeted LacZ for degradation. As a result, we obtained similar decrease in activity (38%) in LB medium [Figure 5.36]. However, in poor medium the growth rate was not restored upon its degradation. Therefore, their result agrees with our result. In contrary to LacZ, looking at the result of GFP and mKate degradation, a significant growth rate restoration occurs. Thereby, assuring that amino acid recycling might play a role in forming a resource limitation for high protein production. Hence, this suggests that the released amino acids from LacZ were not the only resources that limited the growth rate. However, the cytosolic occupancy of LacZ even after its degradation, was still a limiting resource

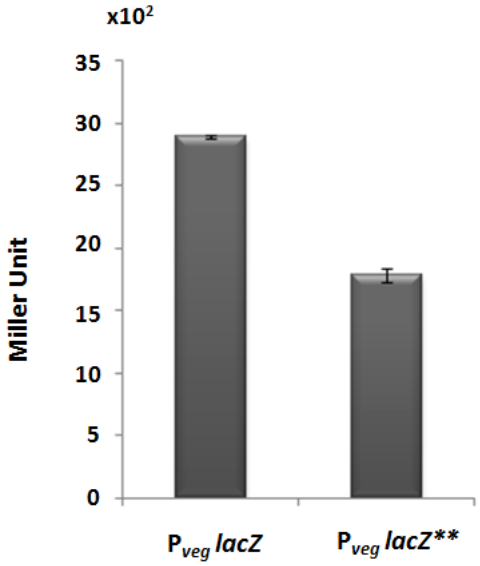

Figure 5.36 Beta-galactosidase activity of LacZ before and after degradation

Growing the strains in LB, and harvesting them during exponential phase, the activity decreased by 38% upon degrading LacZ.

Protein occupancy in the cytosol forms an important limitation for high protein production. This can be explained by the constraint of constant cell density, which pushes the mass and the volume to change proportionally. The increase in volume in response to increased protein production requires the synthesis of phospholipids and proteins to be incorporated to the membrane. Therefore, the allocation of proteins and phospholipids to the membrane requires additional resources.

CHAPTER 6

Development of a synthetic biology tool to finely tune gene expression

CHAPTER 6. Development of a synthetic biology tool to finely tune gene expression

6 Development of a synthetic biology tool to finely tune gene expression

6.1 Introduction

In *E. coli*, the recognition of the ssrA tag by ClpXP is dependent on an adaptor protein called SspB. It brings the targeted protein to ClpXP, in contrary to *B. subtilis* where the protease recognizes the tag without the need of a SspB homologue. Griffith and coworkers (2008) engineered *E .coli* ssrA tags (AANDENYSENY**XXXX**) fused to the C-terminus of a protein so that the degradation of the tagged protein by ClpXP proteases is dependent on the adaptor protein "SspB" from *E. coli*. As there is no SspB homologue in *B. subtilis*, they assumed that the system was well controlled and that the tagged protein was only degraded when *sspB* from *E. coli* was heterologously expressed (Griffith & Grossman 2008). They altered the SsrA tag of *E. coli* by changing the last four amino acids, which are normally recognized by ClpX. The ssrA tag of *E. coli* consists of the SspB recognition site (AANDENY) and the ClpX recognition site (ALAA). In addition, they added a linker of 4 amino acids (SENY) between the SspB and the ClpX recognition sites that is assumed to enhance the degradation of tagged proteins in *E. coli* (McGinness et al. 2006). Hence, they constructed *B. subtilis* strains producing GFP, tagged with 19 versions of ssrA with various ClpX recognition site (GFP^{ssrA}). They cloned *sspB* under the control of an IPTG-inducible promoter and integrated one copy of the expression cassette in the chromosome of *B. subtilis*. They quantified the stability of GFP in the absence of induction and its degradation during one hour following IPTG induction. Out of the 19 versions, they selected a single tag (ALGG), which allowed (1) the tagged GFP (GFP^{ALGG}) to accumulate to a level comparable to that of the unmodified GFP in the absence of induction of SspB and, (2) a complete degradation of GFP^{ALGG} following SspB induction.

In the previous chapter, we have presented the characterization of the strains carrying a tagged *gfp*. We showed that the degradation of GFP was incomplete in our growth conditions. In this chapter, we characterized the system with the use of the adaptor protein SspB, and we aimed to explore whether we can improve the system efficiency by overproducing the protease ClpXP. Such a system may be used to overcome the limitations in amino acids in order to improve the production of a protein of interest by targeting other endogenous proteins.

6.2 Characterization of the targeted degradation system in the presence of SspB

We aimed here to investigate the efficiency of the degradation system by using SspB. P_{veg}*gfp*^{ssrA*} was integrated in the *amyE locus* (ssrA* corresponds to the ssrA-like tags that we used) and the P_{hs}*sspB* was integrated in the *nprE locus*. In addition, we used a tag (AGKTNSFNQNVAASV) that allows partial degradation of GFP in *B. subtilis* without the need of any adaptor protein as a control (Guiziou et al. 2016).

The fluorescence of GFP was assayed by LCA in different media to monitor the degradation in different growth conditions. The GFP abundance [Figure 6.1] of any of the GFP^{ssrA*} across growth conditions tested so far, was lower than the abundance of the untagged GFP. The presence of SspB did not result in higher degradation levels even with maximum LacI derepression (1 mM IPTG). According to Griffith *et al.* (2008), the GFP^{ssrA*}was stable in the absence of SspB. However, we did not observe similar results. In order to verify if the decrease in the fluorescence is related to a lower brightness of the tagged GFP rather than a degradation, western blot were performed. The strains were grown until stationary phase when the degradation is supposed to be higher than inthe exponential growth phase (Farrell et al. 2005). Then, the soluble proteins were extracted, then they were loaded on the SDS-PAGE gel by adjusting the amount of the added samples. The western blot image [Figure 6.2] shows that the bands of the GFP are more intense than the bands of the GFP^{ssrA*} variants whether the SspB protein is present or not.

The engineered targeted degradation system presented by Griffith *et al.* (2008) is indeed fully effective on only moderate levels of production of tagged proteins. The reason is that they used the very same IPTG-inducible promoter to express the gene of interest and the *sspB* gene. Hence, they were able to observe a complete degradation of the tagged protein. Together with the results presented in the previous chapter, it suggests that, SspB or the ClpXP protease may be limiting for a complete degradation of highly produced proteins. We aimed at improving the degradation system in order to increase the level of degradation, as being able to finely tune protein levels might serve several biotech applications.

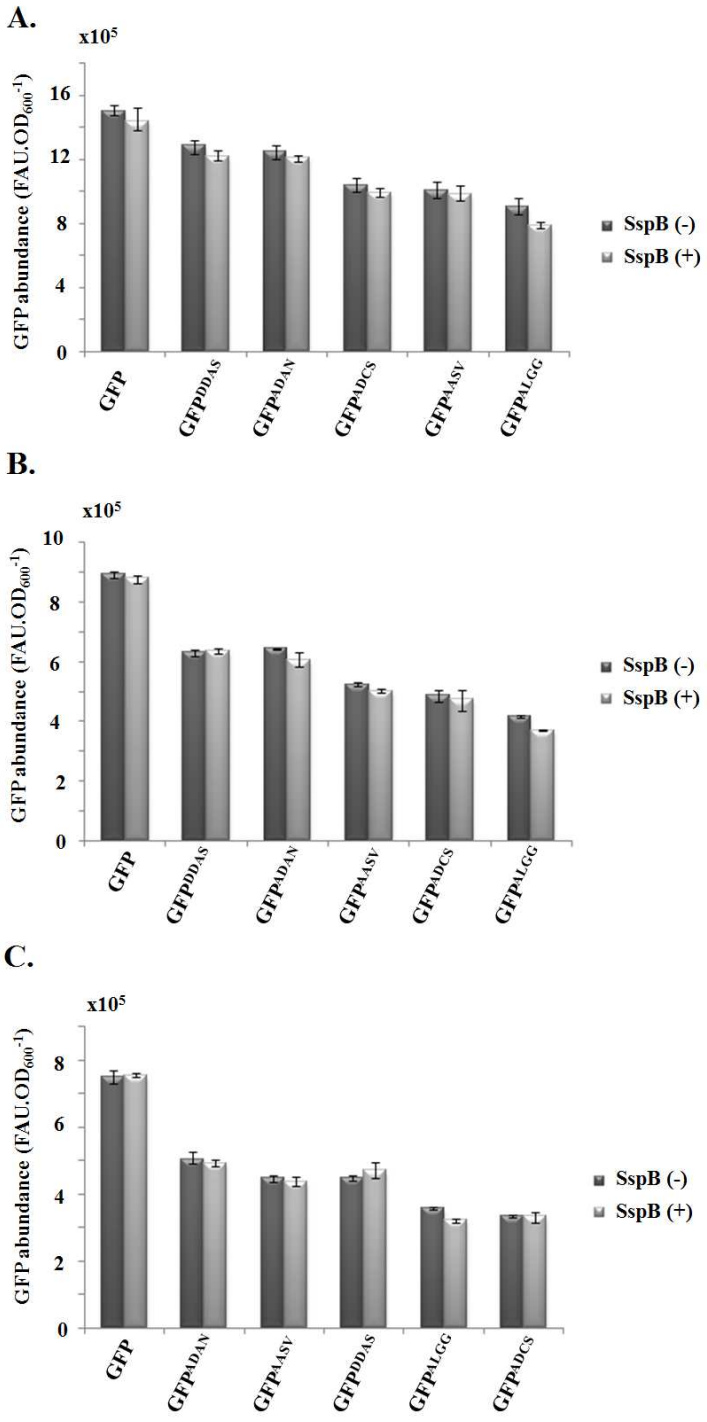

Figure 6.1 GFP abundance of the tagged GFP measured in different media in the presence and absence of the adaptor protein

(A) S medium, (B) SX medium, (C) CHG medium. Maximum derepression of *sspB* by 1mM IPTG.

Figure 6.2 Western Blot image showing the sfGFP with the different ssrA*

From the left to the right, untagged sfGFP without *sspB*, untagged GFP without inducing *sspB*, untagged GFP with inducing *sspB*. sfGFP-ALGG without *sspB*, sfGFP-ALGG without inducting *sspB*, sfGFP-ALGG with inducing *sspB*, etc.

6.3 Over-producing SspB, ClpX, and ClpPin a single operon

6.3.1 Design of the ClpXP-based degradation system

In order to increase the level of degradation of a protein of interest, we decided to build a strain in which the productions of both the ClpXP protease and SspB were increased. Our hypothesis was that this strategy may circumvent the putative limiting amount of ClpXP and consequently lead to a complete degradation of the GFP^{ssrA} protein when highly produced.

We designed various operonic structures consisting of the *sspB*, *clpX* and *clpP* downstream the P_{hs} and strong, synthetic RBSs [Figure 6.3, Table 6-1,Table 8-2]. The design of the operons was inspired from naturally existing operonic structures of *B. subtilis*. We selected two different operons, the *dhbACE* and the *aroFBH* operons because:

1. when comparing transcriptomic and proteomic data obtained in (Buescher et al. 2012; Goelzer et al. 2015; Nicolas et al. 2012), it turned out that the translation of these

genes/mRNAs led in each operon to proteins with comparable concentrations across growth conditions (thus excluding any post-translational process; Figure 6.3).

2. they possess two different translational structures.The genes *dhbA,dhbC,* and *dhbE* are separated by intergenic regions containing each an RBS [Figure 6.3 A]. By contrast, the *aroF*, *aroB* and *aroH* genes overlap and only possess one RBS is present upstream *aroF* on the mRNA [Figure 6.3 B].

Moreover, for each design, we made use of either the natural *B. subtilis clpX* and *clpP* genes or the heterologous *E. coli clpX* and *clpP* genes. The use of proteases from both microorganisms allows us to know whether SspB has a better affinity to either one of the ClpX proteins (*i.e. E. coli* or *B. subtilis*), thus leading to a higher degradation rate.

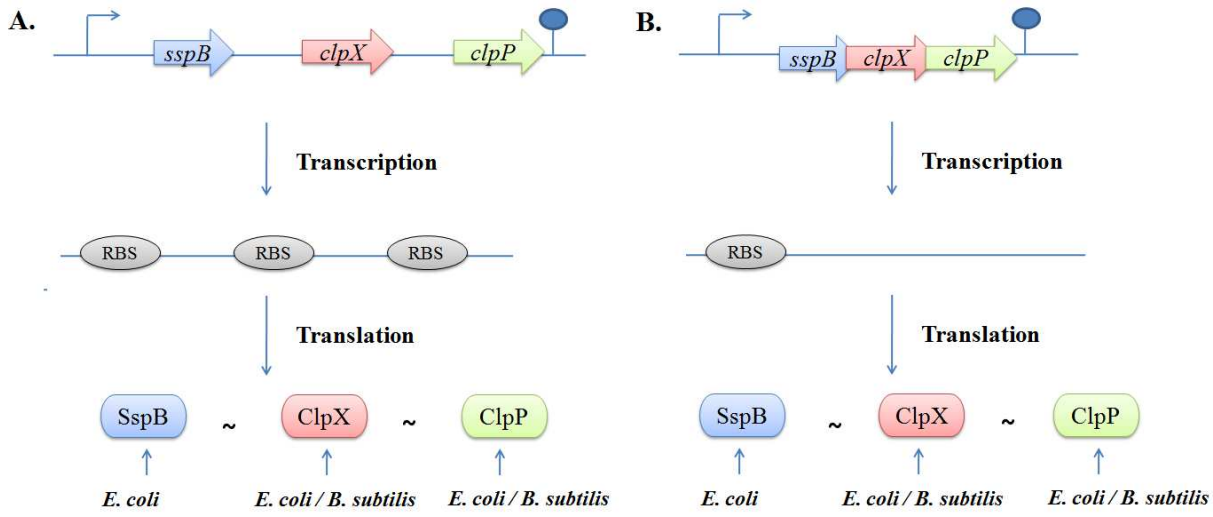

Figure 6.3 Operon design for the degradation system

Under the control of P_{hs}, all of the *sspB, clpX*, and *clpP* are designed in an operon based on the natural *aroFBH* and *dhbACE* operons of *B. subtilis*. Both designs differ by the fact that *dhbACE* has intergenic regions including RBS, while the aroFBH does not carry any intergenic regions. ClpX and ClpP are products of either the endogenous *B. subtilis* genes or the heterologous *E. coli* genes.

Table 6-1 Design of the various *sspB clpX clpP* operons

Operon name	Natural operonic structure	Structure between genes	*clpXP* origin is from
Int-clpXP$_{Bsu}$	*dhbACE*	Intergenic region exist	*B. subtilis*
Int-clpXP$_{Eco}$	*dhbACE*	Intergenic region exist	*E. coli*
Ov-clpXP$_{Bsu}$	*aroFBH*	Genes overlap	*B. subtilis*
Ov-clpXP$_{Eco}$	*aroFBH*	Genes overlap	*E. coli*

The constructs were integrated in the *nprE locus*, while the tagged *gfp* (*gfp*^{ALGG}) was integrated into the *amyE locus*. Then, LCA measurements of the strains expressing the P$_{veg}$*gfp*^{ALGG} and each of the 4 operons were performed in different media (S and SX) in order to assess the degradation profile in different conditions. The results in [Figure 6.4A and B] show a comparison in the GFP abundance before and after degradation by either SspB or by the SspB and ClpXP, in two chemically defined media (SX and S). GFP^{ALGG} degradation with the various operons were close to the degradation when SspB was produced alone in both media [Table 6-2]. In addition, the operon designs which have an intergenic region with a RBS element (Int-clpXP$_{Bsu}$ and Int-clpXP$_{Eco}$) resulted in a better degradation (between 18 and 27%) than that of the designs with overlapping genes (Ov-clpXP$_{Bsu}$ and Ov-clpXP$_{Eco}$). Hence, with the P$_{veg}$ promoter the GFP^{ALGG} was not completely degraded even when ClpXP protease was overproduced. Therefore, for the following characterization, we decided to focus on only one construct (Int-clpXP$_{Bsu}$) and explore a range of GFP^{ALGG} production levels.

Table 6-2 Percentage of GFP degradation with SspB and the various *sspB clpX clpP* operons

	SspB	Int-clpXP$_{Bsu}$	Int-clpXP$_{Eco}$	Ov-clpXP$_{Bsu}$	Ov-clpXP$_{Eco}$
SX medium	16%	18%	23%	18%	14%
S medium	24%	22%	27%	13%	22%

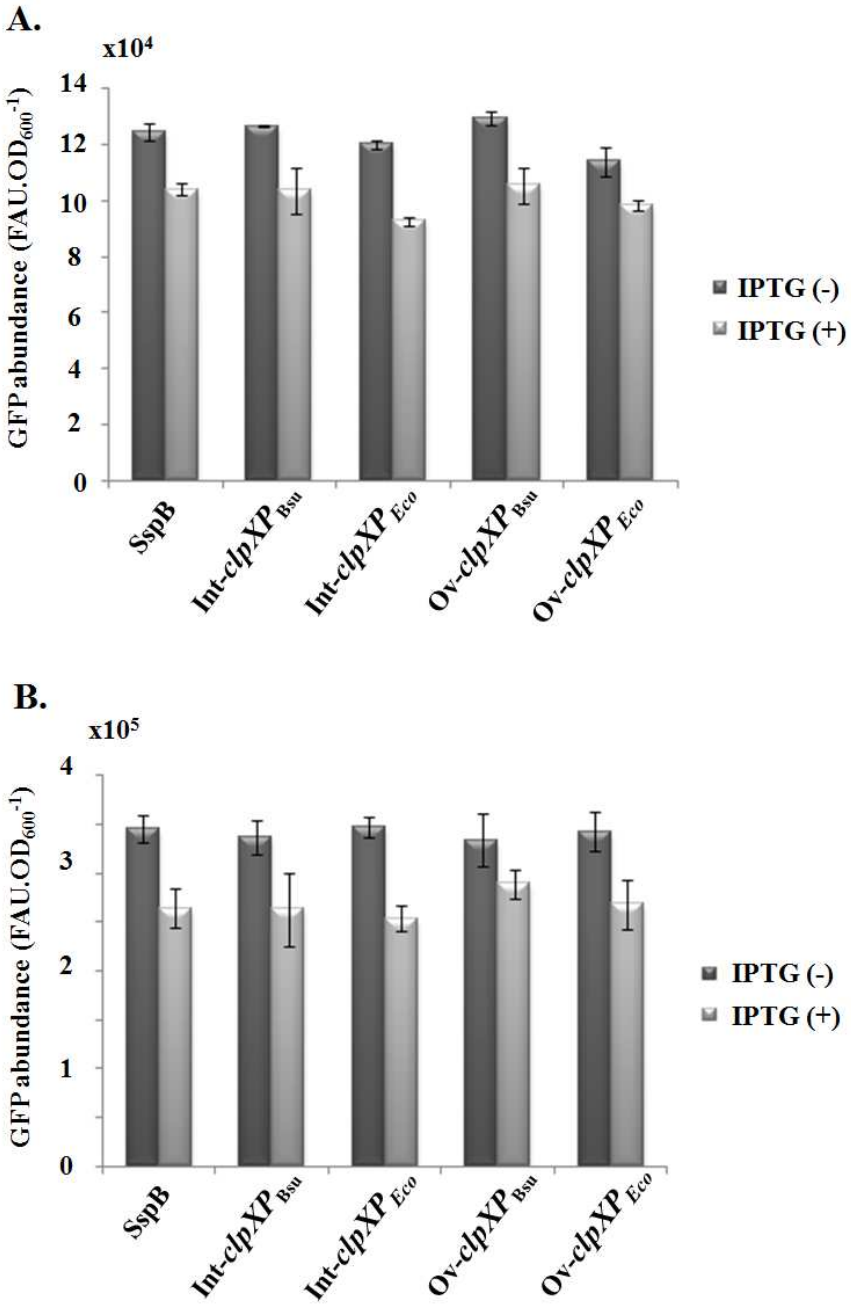

Figure 6.4 LCA measurements for the different operons

The GFP abundance is presented for the strains containing P_{veg}*gfp-ssrA*^{ALGG} and either P_{hs}*sspB* or P_{hs} *sspB clpX clpP*,with and without the induction by IPTG in SX medium (A) and S medium (B).

6.3.2 Characterization of the system with different promoters

In order to characterize the efficiency of the ClpXP-based degradation system throughout a range of protein concentrations, we used various promoters previously published

(Guiziou et al. 2016). The genetic constructs contained each of the promoters fused upstream of *gfp* and *gfp*^{ALGG} in *amyE locus*, and P_{hs} *sspB clpX clpP* (Int-clpXP$_{Bsu}$) in the *nprE locus*. In order to characterize the efficiency of the ClpXP-based degradation system across growth conditions, cells were grown in either a defined rich medium (CHG) or a defined poor medium with glucose (S). Cells were grown in the presence of 1 mM IPTG to fully induce the P_{hs} promoter. The profile of GFP^{ALGG}degradation presented in [Figure 6.5 A and B], showed an enhanced degradation at intermediate to low levels of production of GFP. In addition, the degradation was more efficent in S medium than in CHG medium, maybe due to more *free* proteases in S medium.

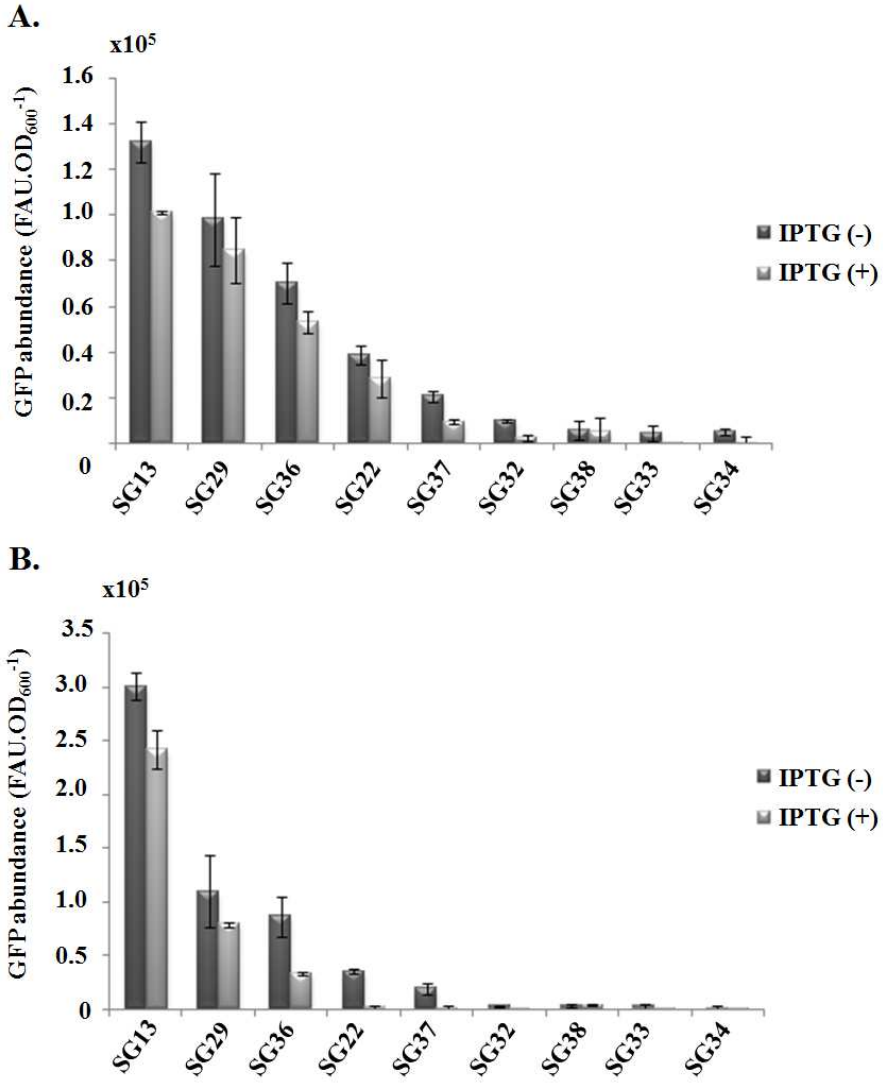

Figure 6.5 Different levels of GFP^{ALGG} production and degradation with and without the induction of *sspB clpX clpP*

(A) The degradation profile in CHG medium, (B) and in S medium show a full degradation at intermediate level of GFP production.

We next aimed at evaluating the efficiency of degradation of a synthetic system based on the *sspB clpX clpP* operon *versus* sspB alone. In addition, we monitored the degradation of mKate2 to generalize the results observed with GFP. We used some of the strains carrying the promoters from [Figure 6.5], then we performed LCA measurements to follow the protein abundance when inducing P_{hs}*sspB* or P_{hs}*sspB clpX clpP*. [Figure 6.6 A and B] shows the reporter protein abundance of *gfp*^{ALGG} and *mkate*^{ALGG} in CHG and S media respectively. Besides [Figure 6.6 C and D] shows their degradation in S medium. As a result, the new degradation system (the selected *sspB clpX clpP* operon: Int-clpXP$_{Bsu}$) did not improve the GFP^{ALGG}, nor the mKate^{ALGG} degradation as compared to strains expressing only *sspB*. Altogether these results suggest that SspB is the bottleneck in the degradation process, especially when ClpXP are not highly overproduced (as it is the case here with the use of the P_{hs} promoter).

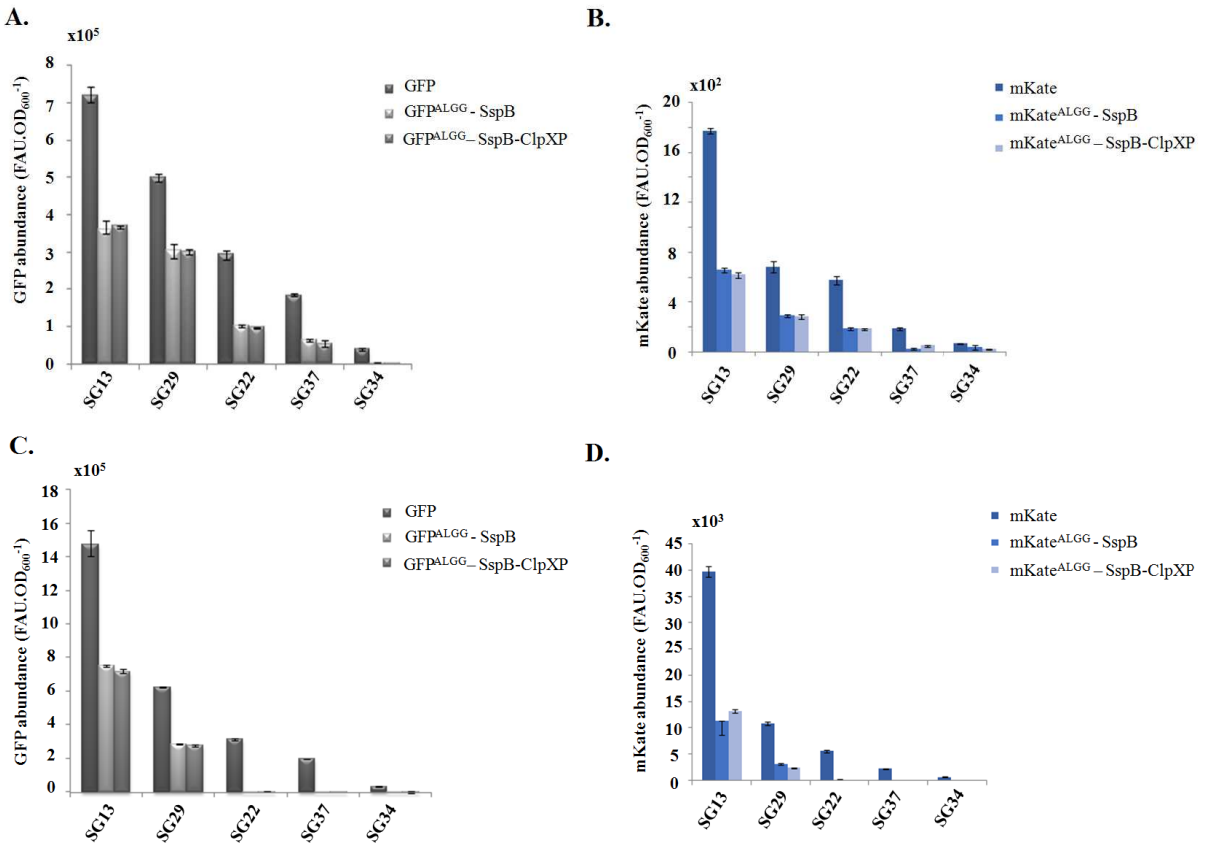

Figure 6.6 Protein degradation at different levels of GFP and mKate abundances

With a range of GFP production using a library of promoters, the level of degradation of GFP and mKate is enhanced more in S medium than in CHG medium. (A) GFP^{ALGG} degradation in CHG medium, (B) mKate^{ALGG} degradation in CHG medium, (C) GFP^{ALGG} degradation in CHG medium n S medium, (D) mKate^{ALGG} degradation in S medium.

6.4 Consequences of ClpXP over production on the proteome

As we aimed to setup an amino acids recycling synthetic tool based on the use of ClpXP, we wondered whether ClpXP may have off-targets, especially if overproduced (together with sspB). Therefore, to investigate the consequences of an overproduction of ClpXP on the cell physiology, we constructed a P_{veg} *clpX clpP* operon using the *clpX clpP* genes from *B. subtilis* and the strongest promoter known so far in *B. subtilis*, P_{veg}. Note that we used the P_{hs} in the previous chapter, however we aimed at exploring a much higher level of *clpX clpP* expression hoping that this may allow to increase the degradation system efficiency and target more proteins. The P_{veg} *clpX clpP* strain and a wild type strain were grown in CHG and S media, each in 6 biological replicates in Erlenmeyer flasks. The growth rates were not significantly different between the P_{veg} *clpX clpP* strain and the wild type strain for each medium [Figure 6.7]. Then, the cells were harvested during the exponential phase to be treated for relative protein quantification in the University of Greifswald-Germany in Prof. Dorte Becher's lab (see Material and methods in section 8.9). Moreover, the computational and statistical treatments were performed by the Perseus software (Tyanova *et al.* 2016; Cox *et al.* 2014).

Here we provide a prelimary analysis of the dataset, to evaluate the impact of ClpXP over-expression on the total soluble proteome. In S medium, only 5 proteins showed significant differences out of the 886 detected proteins[1], where ClpP showed the highest increase 5.4 times). In CHG medium, 782 proteins were detected. Moreover, 496 proteins showed a significant difference in abundances between the P_{veg} *clpX clpP* strain over the wild-type strain, and ClpX and ClpP showed the highest increase (6.8 times). We present in the next section an analysis of the proteomics dataset of cells grown in the CHG medium.

[1]In this experiment, the statistical treatments used in the Perseus software were stringent, which might explain the few number of statistically significant ratios.

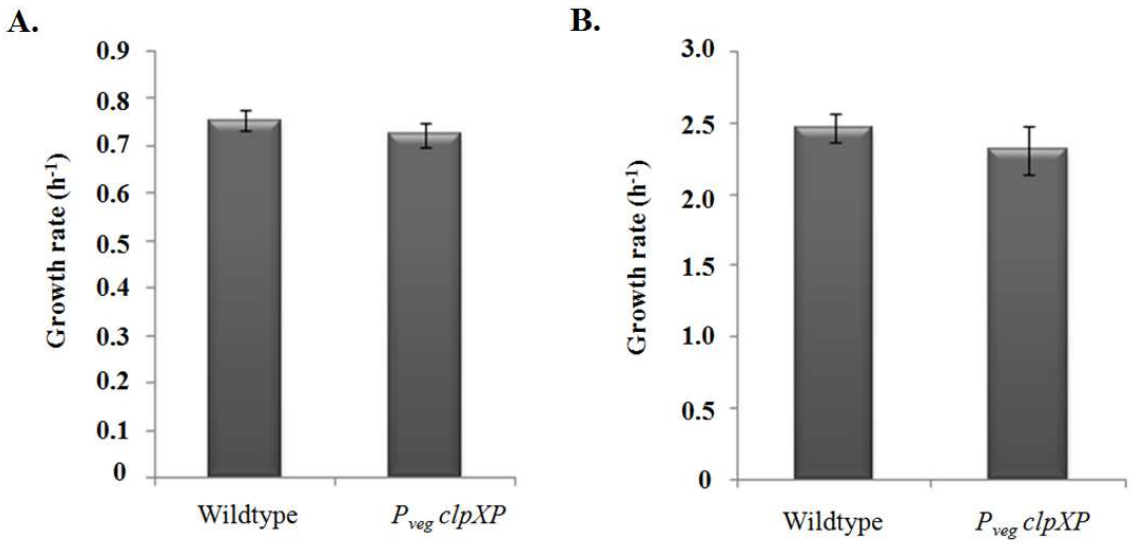

Figure 6.7 Comparable growth rate between the wild type and the P$_{veg}$*clpXP* strain.
(A) in poor medium S, (B) in rich medium CHG.

6.4.1 General description

Generally looking at the number of detected proteins, there were more proteins in the wild-type strain than in the P$_{veg}$ *clpX clpP* strain [Figure 6.8]. We observed 316 proteins down-regulated and 181 proteins up-regulated in the P$_{veg}$ *clpX clpP* strain as compared to the control strain. On [Figure 6.9] are the categories of proteins for which the abundance changed between the two strains. In addition to ClpX and ClpP, most of the proteins that were up-regulated are related to amino acid biosynthetic proteins, nucleotides synthesis or t-RNA synthetases. Proteins that were down-regulated are mostly involved in motility and chemotaxis (flagella synthesis), or controlled by the sigma factor σ_D, (*i.e.* the sigma factor that activates flagella synthesis). The remaining categories of proteins slightly varied. Some proteins involved in the general stress response and oxidative stress, or other proteases slightly vary, but no other clear trend emerged.

Since the level of ClpXP increases in the P$_{veg}$ *clpX clpP* strain, we checked the variation of known direct targets of ClpXP (see section 3.2.2), and in particular of Spx which regulates many genes of the heat, oxidative, and general stress responses. SpoIVA, another ClpXP target, is present only during sporulation and not in exponential growth. However, Spx is not

detected in our dataset[2], but we checked if the genes under the direct control of Spx show some variations.

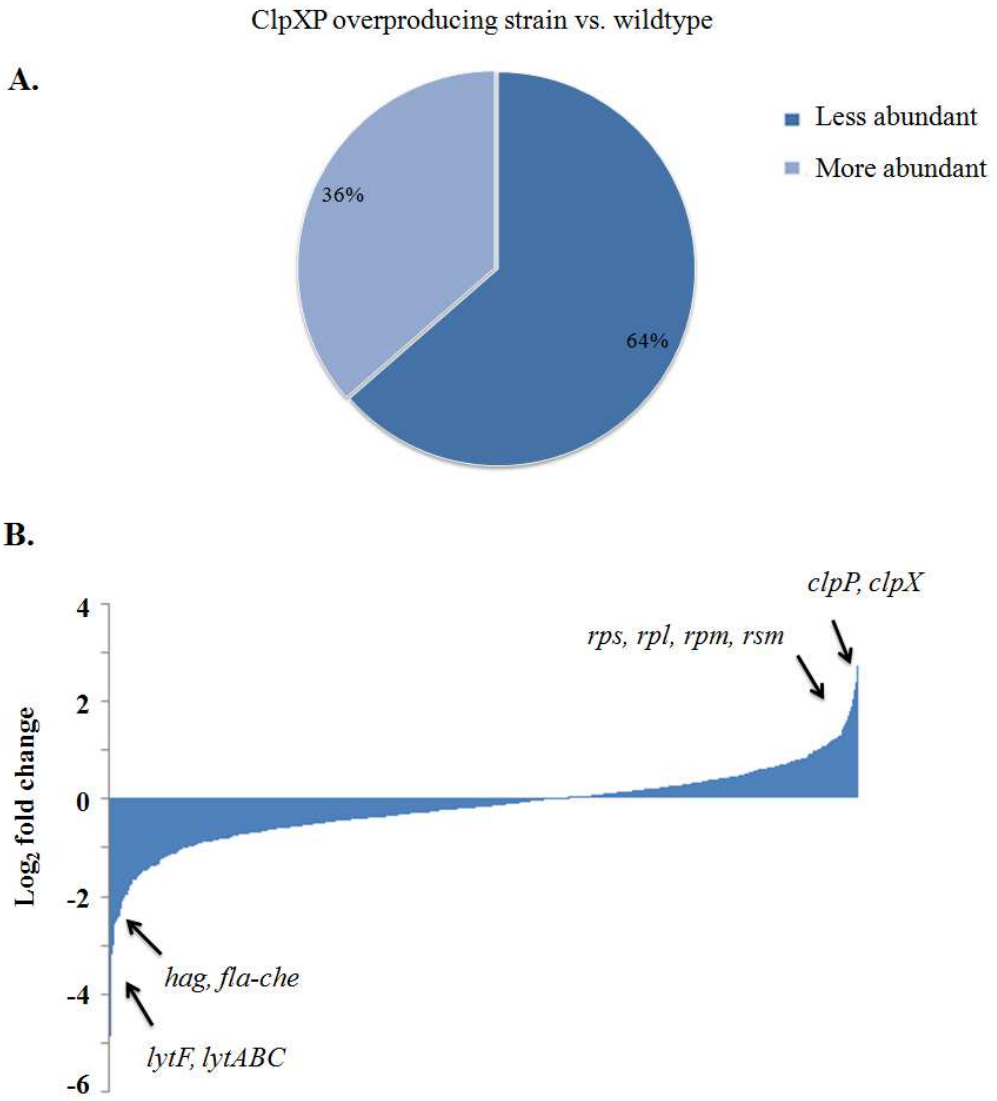

Figure 6.8 Changes in the protein abundance in the wild type and the ClpXP mutant strains

(A) General representation of the detected proteins abundance percentages in the ClpXP overproducing strain versus wild type strain. (B) Differentially regulated genes (log_2 fold change). The highest changes are indicated by arrows (*lytF*, *lytABC* code for autolysin, *hag* and *fla-che* codes for motility and chemotaxis genes, *rps*, *rpl*, *rpm*, *rsm* code for ribosomal proteins.

[2]This was also the case in the proteomics dataset of (Goelzer et al. 2015). The method for protein quantification used in Greifswald cannot detect low abundant and small size proteins. Spx is a small protein (~15kDa). So the undetection of Spx is not surprising.

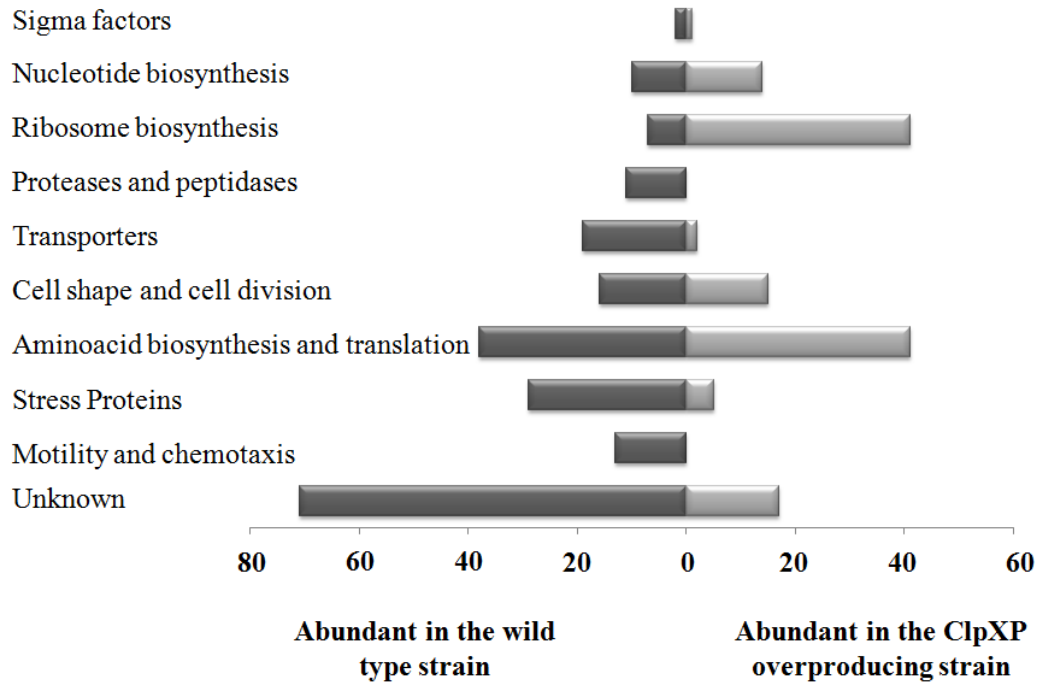

Figure 6.9 Number of proteins (grouped in categories) differentially expressed between the wild type and ClpXP mutant

6.4.2 Major differences in the ClpXP overproduction strain

First of all, the major proteins that are more abundant when overproducing ClpXP are some ribosomal proteins and some other translation proteins. Many of them are related to stringent response (see section 2.5.2). Besides, the proteomics data agrees with some of the results of the DNA microarray analysis performed on *B. subtilis* after adding sub inhibitory concentrations of antibiotics acting as a translation inhibitor [Table 6-3] (Lin et al. 2005). In addition, more genes encoding translation elongation factors and initiation factors increased in the ClpXP strain such as *greA*, and *efp*, which decrease, *tufA*, *tsf*, *fusA*, and *infA*.

Ribosomal proteins that were overproduced are presented in [Figure 6.10]. They are more abundant than in the wild-type strain by a range of 2.4 $\log_2$-fold down to 0.57 $\log_2$-fold. Accessory translation factors such as IF1, IF3, EF-Tu, EF-Ts, and EF-G also showed a significant increase by a log2-fold of 1.7, 0.82, 0.42, 0.62, and 0.93 respectively.

Table 6-3 Similarities between the DNA microarray analysis in the case of translational inhibition and protein abundances in the case of ClpXP overproduction.

The data are based on the work of (Lin et *al*. 2005). The abundance of the proteins in the ClpXP strain agrees very much with the DNA microarray data.

Function and Gene name	**Protein name**	**Up/Down regulated** (Lin et al. 2005)	**Abundance in ClpXP strain**
Protein synthesis			
infC	Initiation factor IF-3	Upregulated	High
rplA	Ribosomal protein L1	Upregulated	High
rplR	Ribosomal protein L18	Down regulated	Low
rplU	Ribosomal protein L3	Upregulated	Low
rpsK	Ribosomal protein S11	Down regulated then upregulated after 60 min	High
Transport			
yclQ	Putative iron-siderophore ABC transporter	Down regulated	Low
pyrP	Uracil permease	Upregulated	High
ycgN	1-pyrroline-5-carboxylate dehydrogenase	Upregulated	High
Amino acid metabolism			
pepT	Peptidase T	Downregulated	Low
rocA	pyrroline-5-carboxylate dehydrogenase	Downregulated	Low
rocD	Ornithine aminotransferase	Down regulated	Low
Purine and Pyrimidine Biosynthesis			
pyrG	CTP synthetase	Upregulated	High
pyrD	Dihydroorotate dehydrogenase	Upregulated	High
pyrF	Orotidine 5'-phosphate decarboxylase	Upregulated	High
pyrP	Uracil permease	Upregulated	High
Heatshock proteins			
clpX	ATP-dependent Clp protease ATP-binding subunit	Upregulated	High (overproduced)
clpP	ATP-dependent Clp protease proteolytic subunit	Upregulated	High (overproduced)
nadE	NH3-dependent NAD^+synthetase	Downregulated	Low

Since the growth rate did not change significantly between the two strains, the stringent response should not be the main mechanism responsible for this adaptation. However, recently in (Schäfer et al. 2019), Schäfer and colleagues showed a relation between *clpX* and r-protein. They showed that the deletion of *clpX* resulted in a downregulation of the

r-proteins during heat shock, and that this downregulation was further mediated by Spx. Spx was shown to downregulate promoters that initiate transcription of rRNA and, to a lower extent, promoters of ribosomal proteins, and this by the oxidized and reduced forms of Spx. This suggests, as expected, that Spx is less abundant in the P_{veg} *clpX clpP* strain than in the WT strain, and that the lower abundance of Spx could be responsible of the adaptation of the translation apparatus in our experiments.

On the other hand, the major proteins that decreased in the P_{veg} *clpX clpP* strain in comparison with the WT strain are the motility proteins. They decreased by 3.2- to 1.1- log2-fold [Figure 6.11]. Molière and colleagues (2016) showed that Spx plays a role in the regulation of motility. Spx indirectly and negatively regulates motility genes (Molière et al. 2016). If the level of Spx is decreasing, as expected by the overexpression of ClpXP and suggested by the readaptation of the translation apparatus, then the motility genes should increase. Actually, motility is a highly integrated life style, which is strongly controlled by an interplay of different regulatory mechanisms such as Spo0A, CodY, SwrA, DegU, ComK, AbrB, sigD, etc., and could be the cause of the observed behavior of motility genes.

Table 6-4 Abundance of the cell division proteins in the ClpXP overproducing strain

Gene name	**Protein and function**	**Abundance in ClpXP strain**	**Significant**
erzA	Negative regulator of cell division	Low	Yes
ftsZ	Cell division initiation	High	No
ftsE	Cell division ABC transporter	Low	Yes
ftsH	Cell division and general stress protein	High	Yes
ftsX	Cell division ABC transporter	High	Yes
ftsA	Essential for FtsZ ring assembly	Low	Yes
sepF	Over lapping activity with FtsA, Required for proper septal morphology	Low	Yes

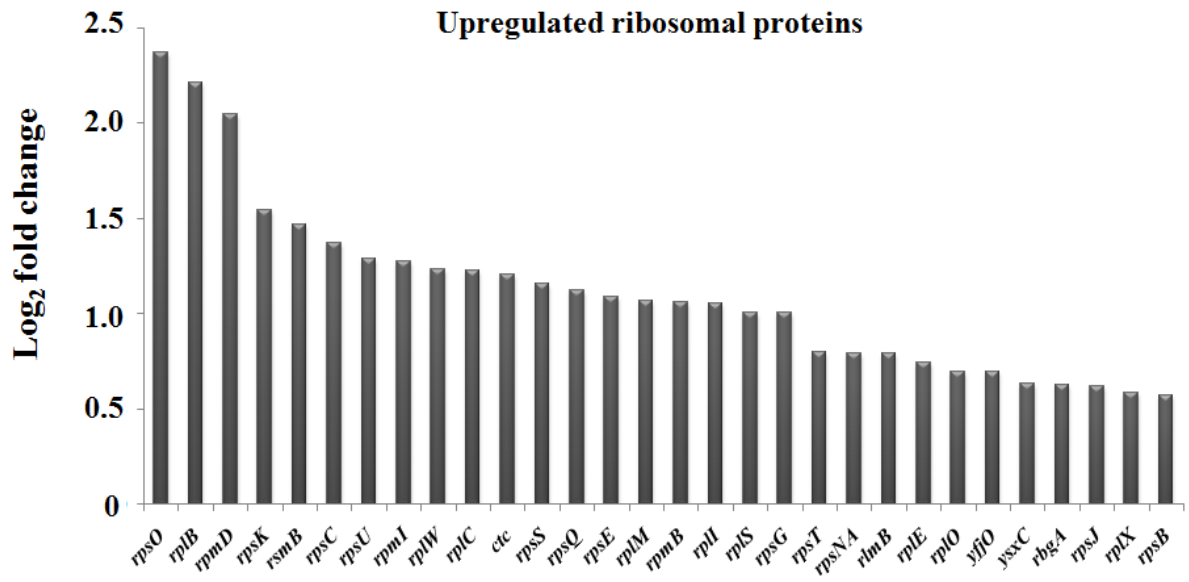

Figure 6.10 High abundance of ribosomal proteins in the ClpXP mutant

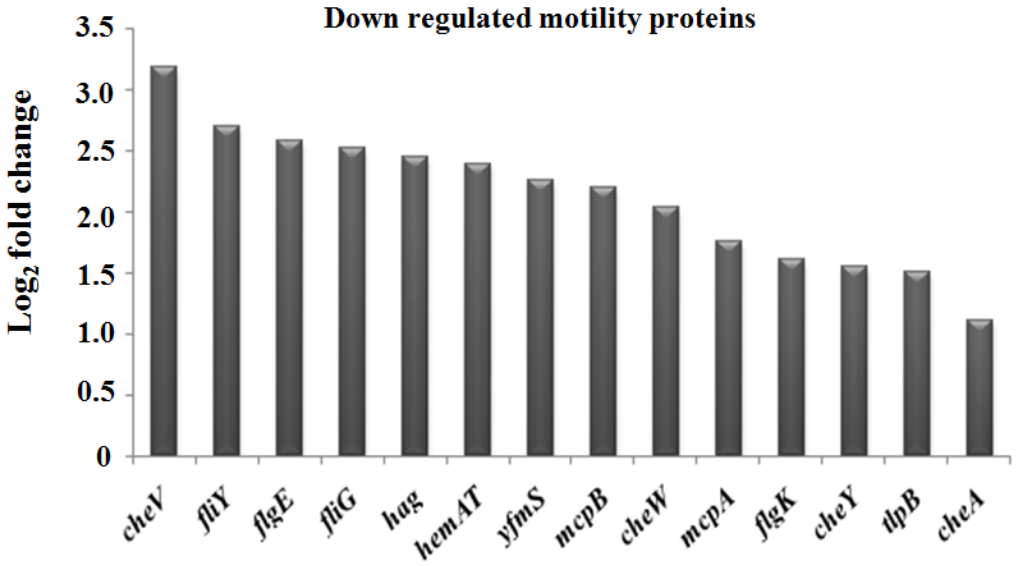

Figure 6.11 Low abundance in the motility proteins in ClpXP mutant

6.5 Discussion

6.5.1 Targeted protein degradation improvement

The ClpXP-degradation system we developed was compared to that of (Griffith & Grossman 2008). First, we analyzed the 4 best tags which showed a complete degradation in the work of Griffith *et al*. (2008). We used the LCA to give precise measurements of the GFP or mKate2 abundance estimated by dividing the fluorescence by the OD_{600}. After that we quantified by western blot to confirm LCA results. Surprisingly, we found that the tag destabilized the protein to a certain level without the need for the adaptor protein. Therefore, if the purpose of such a system was to use it to finely tune protein production, it would not be the best choice. But, since our purpose was to monitor the growth rate of the cells during the full experiment, it was not crucial to have a well-controlled system.

Second, the system was limiting when the quantity of produced protein was too high. When using the P_{veg} promoter for the gene of interest, that is much stronger than P_{hs}, the system could not degrade the GFP^{ALGG} and $mKate2^{ALGG}$ completely. In order to improve this limitation, we decided to overproduce ClpXP together with SspB thinking that ClpXP might be limiting. So, we constructed some operons carrying the *sspB*, *clpX*, and *clpP* genes under the control of P_{hs}. However, unfortunately we did not observe any improvement in the reporter protein degradation. Therefore, we have to keep in mind that we needed to use a promoter comparable to that used for the *sspB* in order to have a full degradation of our protein of interest. Griffith *et al* (2008) used P_{hs} upstream of *sspB*, and an equivalent to P_{hs} upstream *gfp*. But in fact P_{veg} is a very strong promoter, that needs an equivalent promoter (but inducible if possible) to get a complete degradation of the GFP^{ALGG} and $mKate2^{ALGG}$. To obtain such a promoter, we thought that controlling P_{veg} with LacI by adding the *lacO* regions upstream and downstream P_{veg}would be a way to control P_{veg} and to have a very strong inducible promoter. This strategy may be tested in the future.

6.5.2 ClpXP overproduction reorganizes the proteome

ClpXP overproduction caused 64% of the proteins to be less abundant than in the wild type [Figure 6.8A]. This indicated that a ClpXP overproduction caused a major reorganization in the proteome. Many proteins of unknown function were decreased. Nevertheless, in our analysis we focused on relating the known targets of ClpXP to the data that we obtained.

The majority of the proteins that were more produced in the P_{veg}*clpXP* strain are the ribosomal proteins, translation proteins, and others related to the stringent response. We have related these observations to two findings. (1) The profile of protein abundances where similar to the DNA microarray study made on *B. subtilis* in the presence of translation inhibitors (Lin *et al.* 2005). (2) The Spx, a main target of ClpXP, downregulates ribosomal proteins in a similar pattern to the stringent response (Schäfer *et al.* 2019). Concerning the finding (1) a translation inhibition leads to an increase in ribosome abundance because as the translation slows down, more amino acids become available which consequently leads to the release of ppGpp from inhibiting ribosome synthesis (Scott et al. 2010). Concerning the finding (2), the overproduction of ClpXP causes more degradation of Spx. Consequently, in its absence all the regulatory pathways that are related to the stringent response will be shut down. Taking the two findings together, we can propose that ClpXP has an indirect role in ensuring 'normal' ribosome abundance in the cell.

Motility proteins are globally regulated through different regulators such as Spo0A, CodY, SwrA, DegU, ComK, AbrB. In stress conditions, Spx plays a major role in suppressing motility through indirect way (Molière *et al.* 2016). The overproduction of ClpXP is expected to rather enhance motility genes as it targets Spx. However, our results showed a downregulation of motility proteins when ClpXP was overproduced, which suggests that ClpXP has a role in regulating what activates motility too.

CHAPTER 7

Conclusion and Perspectives

7 Conclusion and Perspectives

My PhD project have brought a new perspective to understand the bottlenecks of obtaining microbial cell factories for protein production. We came out with new results and conclusions on the behavior of *B. subtilis* when overproducing a heterologous protein. *B. subtilis* has been a case study to investigate the secretion stress that it faces when subjected to protein overproduction. However, we did not know what are the consequences at the production level before being secreted. Does the growth rate change as it is shown on other microorganisms, and as predicted by mathematical models? If yes, what causes the growth rate decrease? What is the resource type that is the most important? And, how do we overcome the resource limitation to enhance protein production?

The answers to these questions are presented in this PhD project. We now know that *B. subtilis* slowers its growth rate when it is exposed to gratuitous protein overproduction. We investigated the growth rate reduction in different chemically defined media, and by using different reporter proteins (GFP, LacZ, and mKate2). LacZ resulted in a higher growth rate decrease because of its size and structure. Indeed, GFP and mKate2 are of similar size, but mKate2 resulted in a higher growth rate decrease. This might be due to the amino acid composition of the protein and/or might be that the mRNA abundance between these two genetic constructs differed. Therefore, we propose to perform a qPCR in order to know the level of mRNA for each strains carrying the *gfp* and the *mkate2* genes. If the *mKate2* mRNA abundance is higher, then it may explain the experimentally observed higher growth rate reduction.

The use of different media allowed us to know the behavior of *B. subtilis* in different growth conditions, and especially to identify the level of protein to be produced that lead to significant growth rate reduction in each medium. Growth rate reduction starts to be more significant as the medium tends to be poorer, and of course with more than one gene copy integration (in the chromosome). A single P_{veg}*gfp* integrated copy in the genome resulted in less than 10% growth rate reduction in the richest medium (CHG), and 15% in the minimal medium with glucose (S). Besides, the single P_{veg}*gfp* copy was estimated to present 3% of the total soluble proteins in CHG medium, while it was estimated to present around 10% in S medium. This observation agrees with the fact that the efficiency of translation increases in poorer media because of more available ribosomes (Borkowski, Goelzer, et al. 2016). Therefore, GFP was more produced in S medium, and thus resulted in higher growth rate reduction. Nevertheless, expressing the cassette P_{veg}*gfp* from a multicopy plasmid in *E. coli*

resulted in a higher growth rate reduction in CHG medium than in *B. subtilis*. This observation raised the question about a compensation strategy followed by *B. subtilis* to re-allocate its cellular resources to maintain a maximal growth rate. To answer this question we performed relative protein quantification with the strains overproducing GFP at different levels against a wild type strain.

We gave a detailed analysis of the proteome data for the relative protein quantification where we showed a general decrease in the abundance of the endogenous proteins in the GFP overproducing strains. A general decrease in the proteome is predicted to happen in the presence of a useless heterologous protein (Scott et al. 2014; Scott et al. 2010). When entering in more details on the type of proteins that changes the most, we found that big-sized, non-essential proteins related to the non-ribosomal peptide synthetases became less abundant with increasing GFP production. The size of these proteins causes them to be costly proteins in terms of resources for the cell, rather than being costly because of their number. These findings allowed us to deduce a relationship between the amino acids utilization to produce the essential and non-essential proteins including the gratuitous protein, the cytosolic occupancy of the proteins, and the growth rate reduction. This hypothesis was proved by applying a synthetic tool of targeted protein degradation against the gratuitous proteins (GFP, mKate2, and LacZ). The result showed a growth rate restoration when degrading GFP and mKate2, which indicated that the released amino acids were reused by the cell. On the other hand, and in agreement with (Stoebel et al. 2008) the degradation of LacZ did not result in a growth rate restoration even after 38% decrease in activity. Hence, the recycled amino acids from LacZ degradation were not enough to increase the growth rate, which suggested that a second bottleneck is still contributing to the burden. When looking at the percentage (by mass) of GFP and LacZ with respect to the total soluble protein in their corresponding strains, we found that GFP represented ~ 9 % while LacZ represented ~ 7% [section 5.3.2]. Knowing that LacZ is bigger than GFP by 4.4 times, it means that the amino acids are not the only limitation in this case, but also the cytosolic occupancy by LacZ prevented the growth rate restoration. Alternatively, it may be that $LacZ^{ALGG}$ is not fully degraded as we showed it by western for the GFP^{ALGG} but degraded enough to lose its activity. To answer this question we should perform a western blot and a coomassie gel to verify whether LacZ is fully degraded or not. Therefore, depending on the protein size and its compaction, the type of the limiting resource can be either the amino acid composition, or the cytosolic occupancy, or both. In addition, it is not the load on the ribosomes, chaperones, proteases, or the lack of RNA

polymerases that are the limiting resources when overproducing a gratuitous protein (Scott et al. 2010; Scott et al. 2014; Weiße et al. 2015; Cookson et al. 2011; Daraba & Alma 2018; Gyorgy et al. 2015). However, we cannot neglect their importance as machineries required to be available for all the proteins including the gratuitous protein, but they are not the major cause as it was discussed by the above-mentioned authors. To further investigate on the limitations across growth conditions, we suggest to do proteomics analysis for the strains carrying two copies of the heterologous reporter genes in CHG growth condition. Knowing that the amino acids availability is not a constraint in rich media, would amino acids still be a limitation when proteins are overproduced?

The targeted degradation system can be used as a tool to degrade some costly proteins in order to save cell resources and favor the production of our protein of interest. On the other hand, knocking out genes would also help to save resources and increase the growth rate. However, if the proteins to target are essential, then applying the degradation system may help. In order to assess how efficient this strategy would be, we propose to target costly proteins and quantify the gratuitous protein. Besides, knocking out the NRPS would help to increase the protein production.

For further improvement of the degradation system to obtain a complete targeted protein degradation, we propose that an equivalent (in strength) promoter to P_{veg} must be designed to be strong and controlled. To do so, we can engineer a set of P_{veg} promoter carrying *lacO* elements to be regulated by LacI. Having such a promoter would allow to express *sspB* or *sspB clpX clpP* at high levels.

Finally, our findings pave the way for further studies on the overproduction of secreted proteins. Combining the information on the secretion stress and the overproduction of the secretion machinery to favor gratuitous protein production and secretion will open new perspectives to improve microbial cell factories.

CHAPTER 8

Materials and Methods

8 Materials and Methods

8.1 Molecular Biology

8.1.1 Methods for strain construction

The Gibson Assembly method, StarGate cloning and classical cloning by restriction enzymes were the methods mainly used to construct the strains used in the project.

8.1.1.1 Gibson Assembly

The Gibson Assembly® method was set up by Daniel Gibson (Gibson et al. 2009) [Figure 8.1].The principle of Gibson Assembly® is to *in vitro* assemble multiple DNA fragments each sharing 15-25 base pair overlap with the adjacent fragment. The desired fragments are mixed with three enzymes, (1) an exonuclease, (2) a DNA polymerase and (3) a DNA ligase,inan appropriate buffer.

1) The exonuclease chews back the 5' end creates a 3' overhang to allow the annealing with the adjacent fragment.
2) The polymerase fills the gaps within the annealed fragments.
3) The DNA ligase joins the nicks in the DNA fragments after the gaps were filled in.

The mix is then incubated at 50°C for 15-60 minutes depending on the number of fragments. The resulting product is transformed into*E. coli* Mach1T1 strain. The NEB Gibson Assembly® Cloning Kit, and the NEBuilder® HiFi DNA Assembly Cloning Kit which is also based on Gibson Assembly®, were used according to the supplier's recommendations.

8.1.1.2 StarGate Cloning

The StarGate® Cloning (IBA GmbH) is a direct transfer cloning for rapid assembly and transformation. The protocol starts by the amplification ofthe promoter/gene of interest using primers with a 5' extension containing a sequence recognized by type II restriction enzymes. Type II restriction enzymes (*Bsm*BI is in this work) cut the DNA at defined positions, usually outside the recognition sequence, and in an oriented manner. An identical restriction site is contained in the acceptor plasmid. The StarGate reaction consists of mixing the PCR fragments (4-8 nM), the acceptor vector (4 nM), 1 µL T4 DNA ligase (400 U/µL)(New England Biolabs #M0202S), 1 µL *Bsm*BI (10 U/µL) (New England Biolabs R0580S), 5 µL T4 DNA ligase Buffer (5x), and up to 50 µL MilliQ water. Then, the mix is incubated for 3 hours at 30°C (optional: additional 1 hour at 37°C to increase the efficiency). Finally, 25 µL of the resulting reaction is used to transform 75 µL of *B. subtilis* competent cells.

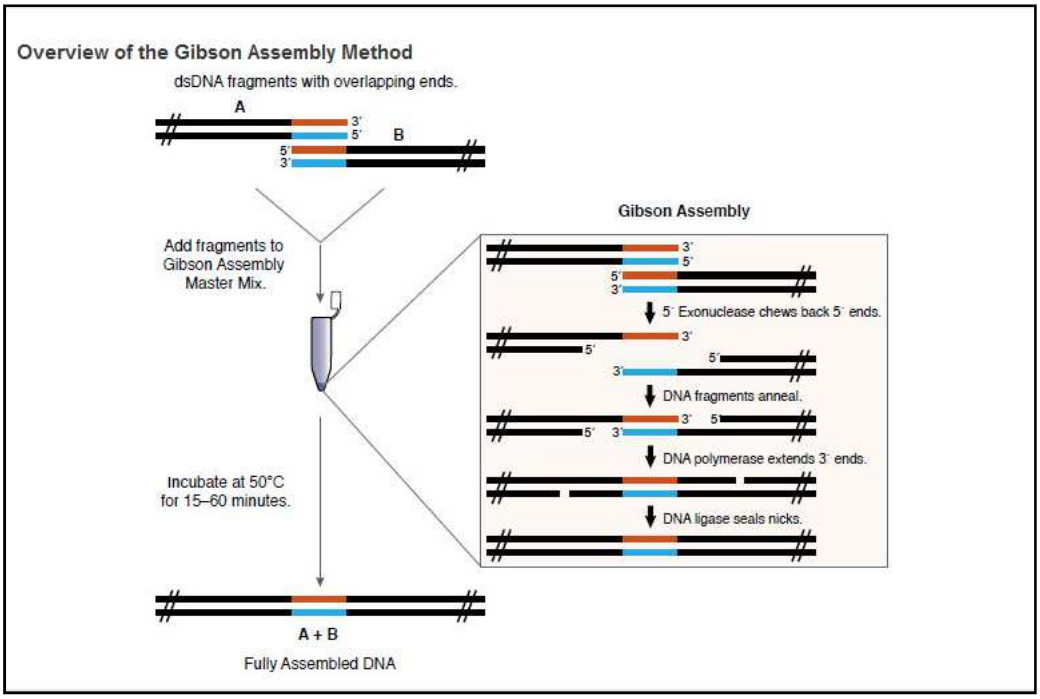

Figure 8.1 Gibson Assembly method

In one reaction containing the fragments to be assembled each sharing an overlap with its adjacent fragment, with 3 enzymes: the exonuclease that chews back the 5' end creating the 3' overhang, the DNA polymerase that extends the 3'ends, and the DNA ligase, which seals the nicks. The mix is incubated at 50°C up to 1 hour and then the resulting product is transformed in E. coli competent cells.(from NEBuilder Gibson Assembly® cloning kit Product Information).

8.1.1.3 Classical Cloning using restriction enzymes

The restriction enzymes were chosen and used according to the desired design and supplier recommendations.

The digested DNA fragments were ligated with T4 DNA ligase (New England Biolabs) according to the supplier recommendations.

8.1.2 Polymerase Chain Reaction

DNA fragments were amplified through successive cycles of synthesis in a total volume of 50 µL. Different DNA polymerases were used depending on the purpose of amplification. ThePhusion® High Fidelity DNA polymerase, the Q5® High Fidelity DNA polymerase(#M0491L), theThermoScientific™Extensor High Fidelity PCR Master Mix, and the Invitrogen™Platinum™SuperFi™ Green PCR Master Mix wereusedtoamplify sequences forwhichaveryaccurate amplification isrequired. The Dream TaqGreen PCR Master Mix (Ref. K1081) and TaKaRa Ex Taq (Cat# RR001) were used to perform colony PCR for screening purposes. All the enzymes were used according to the supplier recommendations.

8.1.3 DNA fragment purification

The PCR-amplified DNA fragments and the DNA sequences / fragments digested by restriction enzymes were purified from agarose gel using either the QIAquick® Gel Extraction Kit (QIAGEN) or Wizard® SV Gel (PROMEGA), or directly purified on columnusing either the QIAquick® PCR Purification Kit (QIAGEN) or PCR Clean up system (PROMEGA).

8.2 Bacterial transformation methodology

8.2.1 *E. coli*specific methodologies

8.2.1.1 Competent cells preparation

The *E. coli* Mach1-T1 competent cells were prepared based on the Rubidium Chloride competent cells protocol. This protocol was adapted from (Green & Rogers 2013).

8.2.1.2 *E. coli* Transformation

E. coli transformation was performed by heat shock. The 'heat shock' protocol starts by thawing the Mach1-T1 competent cells on ice. 100 µL of cells are then mixed with the desired plasmid (20-200 ng), and incubated on ice for 30 minutes. Then, the mix is heat-shocked at 42°C for 1 minute and then returned on ice for 5 minutes. 500 µL LB is added to the mix and incubated at 37°C for 1 hour under constant shaking. The mix is finally plated onto LB-ampicillin (100 µg/µL) plates and incubated at 37°C overnight.

8.2.1.3 Plasmid Extraction

The plasmids were extracted from *E. coli* strains using the QIAprep® Spin Miniprep Kit (QIAGEN). The protocol was performed according to the suppliers instructions.

8.2.2 *Bacillus subtilis* specific methodologies

8.2.2.1 *B. subtilis* transformation

B. subtilis cells (BSB1) used in this work (Nicolas et al. 2012; Buescher et al. 2012); prototrophic derivative of 168 *trpC2B. subtilis*) become competent upon entry in stationary phase when cells are metabolically less active and smaller in size (Sadaie & Kada 1983). To obtain competent cells, an overnight LB culture is diluted in a minimal medium (MG1) to OD_{600}= 0.1, and incubated at 37°C under constant shaking until cells reach OD_{600}= 0.4. The MG1 culture is then diluted 1:10 in a deficient medium (MG2) and incubated for about 2 hours. At this stage, the cells become ready to take up DNA found in the medium. This

exogenous DNA can be used as a nutritional source or bring genetic information to the bacterium (acquisition of plasmids or insertion of DNA fragments into bacterial DNA). The MG2 culture is centrifuged at 2500 g for 5 minutes, and 90% of the supernatant is discarded. The cells then are resuspended in the remaining medium. Finally, the obtained cells are mixed with DNA (*i.e.* 100 µL cells + 3-5 µL of plasmid). Usually 10^6 transformants are obtained per µg of DNA (Ehrlich 1978). After incubation at 37°C for around 30 minutes, the cells are plated onto LB agar plates plus the appropriate antibiotic.

8.2.2.2 Genomic DNA extraction

The genomic DNA of *B. subtilis* was extracted using the GenElute Bacterial Genomic DNA (Sigma- Aldrich) kit.The extraction was performed following the supplier's instructions.

8.3 Bacterial strain construction

8.3.1 The set of inducible promoters

The set of inducible promoters were built by fusing constitutive promoters to the $P_{hypersank}$(P_{hs}). The P_{hs}contains the *lacO* regions (in bold below) recognized by the LacI repressor

(**AAATGTGAGCACTCACAATTC**ATTTTGCAAAAGTTGTTGACTTTATCTACAAGG TGTGGCATAATGTGTGT**AATTGTGAGCGGATAACAATT**AAGC) and amplified from the pIC665plasmid (Lab collection). In the absence of IPTG, the inhibitor protein LacI binds to two operator regions (*lacO*) to prevent the binding of the RNA polymerase, thereby acting on the upstream (constitutive) promoter via a roadblock repression [Figure 8.2].Besides the different constitutive promoters that give the variability in strength, we used different translation initiation regions (TIR) from Borkowski et al. (2016). An illustration of the promoters design and the TIR sequences are presented in [Figure 8.3,Table 8-1]. The inducible synthetic promoters were fused to the*gfp* gene. They were assembled in an integrative plasmid after creating cloning regions for homologous recombination by classical ligation.

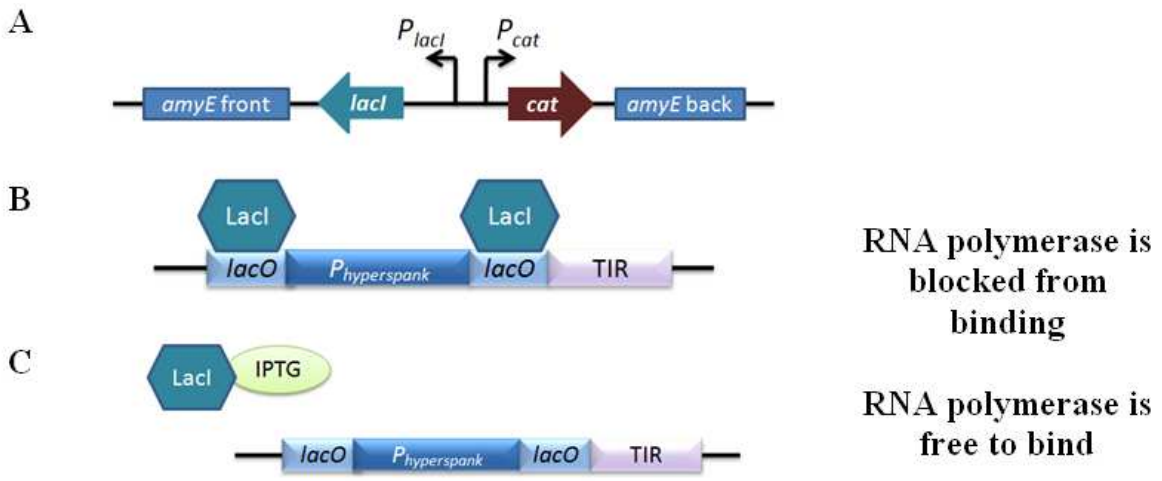

Figure 8.2Representation of the IPTG-inducible promoter Phs

(A) Genetic construct of lacI from E. coli integrated in amyE in B. subtilis (BSOB1 strain from the thesis of Olivier Borkowski), (B) LacI binds to lacO regions, thereby blocking the transcription, (C) IPTG binds with LacI, thus freeing the promoter for RNA polymerase.

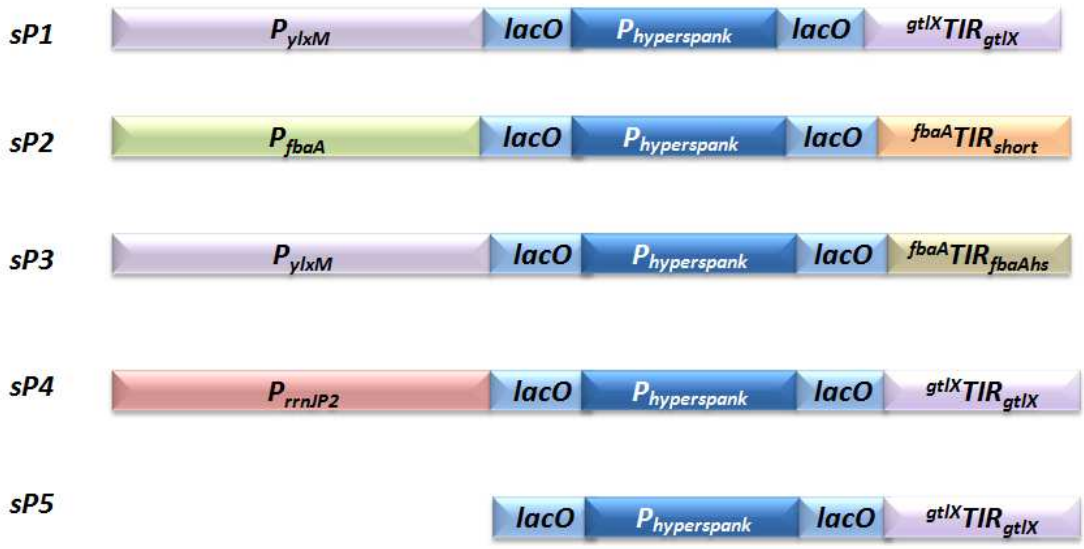

Figure 8.3 Illustration of the synthetic inducible promoters

Fusion between different constitutive promoters with the P_{hs}, in addition to different translation initiation regions.

Table 8-1 The sequences of the IPTG-inducible synthetic promoters (sP)

Each of the constitutive promoters and the TIR elements are presented. The underlined part is the +1/+8 element after the P_{hs}.

Name	Promoter	Sequence	TIR	Sequence
sP*1*	P_{ylxM}	ACAAGATAAAAACTTGACAG TGTCATTAAAACCGTGTAAA CTAAGTTATC	$^{gtlX}TIR_{gtlX}$	<u>GTAAAGGG</u>TTGCCTGAG GCCATACATGACATGAA AGGAAGTATTTGAAA
sP*2*	P_{fbaA}	AATCATGTCATTATGTTGCC GATTTGTCGAAAAGTTGGTA TCCTAGTTAT	$^{fbaA}TIR_{short}$	<u>GGAGAAAA</u>AGCGATCT GAGTATTTACATATGAC AGCAATATATGGGTCAT GCTAGGGTGGAAAGCTT TTTTCGCTAGAAGACAA TCAGGCTACAGGTGGGA AGGAGGAACTACT
sP*3*	P_{ylxM}	ACAAGATAAAAACTTGACAG TGTCATTAAAACCGTGTAAA CTAAGTTATC	$^{hs}TIR_{short}$	<u>GTAAAGGG</u>AGCGGATA ACAATTGGTGGGAAGGA GGACATTCGAC
sP*4*	P_{rrnJP2}	CTTCAAAAAAAGTTATTGAC TTCACTGAGTCAACGAGTTA TAATAATAAA	$^{gtlX}TIR_{gtlX}$	<u>GTCGCTTG</u>TTGCCTGAG GCCATACATGACATGAA AGGAAGTATTTGAAA
sP*5*	-	-	$^{gtlX}TIR_{gtlX}$	<u>GTAAAGGG</u>TTGCCTGAG GCCATACATGACATGAA AGGAAGTATTTGAAA

8.3.2 Strains with reporter genes

The strains expressing the reporter genes under the control of P_{hs}, the synthetic inducible promoters or P_{veg} were constructed using either the classical restriction/ligation, Gibson assembly, or StarGate® cloningmethods. This later methodology was acquired at DSM-Netherlands with their in-house plasmids during a two-month internship.

The $P_{hs}$was amplified by the corresponding primers from pIC665. The *gfp-lacZ* fusion was placed downstream the synthetic inducible promoters and integrated in a region close to the origin of replication by single crossing over (the very same region used to integrate *gfp* mentioned in the previous section).The P_{veg}*sfgfp*(superfolder GFP, later on referred to as P_{veg}*gfp*) was PCR amplified from the genomic DNA of SG13 strain (Guiziou et al. 2016) with primers that contained a 5' extension with the *Bsm*BI restriction site. A similar approach was performed to build the P_{veg}*mkate*and P_{veg}*lacZ*reporter strains. The *mKate* gene and *lacZ*(*spoVG-lacZ*) were PCR amplified from the pSG15 (Guiziou et al. 2016) and pDG1661,

respectively, using primers containing the *Bsm*BI restriction site. When the abovementioned reporter genes were fused to a ssrA tag, the reverse primer additionally contained the tag with the *Bsm*BI restriction site.

The amplified fragments were purified, then the ligation reaction was performed by StarGate® cloning using the right plasmid to integrate by double crossing over in *amyE* or the *nprE loci* [Figure 8.4]. The P_{veg} genetic constructs were transformed into the BSB1 strain, and the P_{hs} genetic constructs were transformed into a BSB1 derivative strain expressing the *lacI* gene (see [section 8.3.3]).

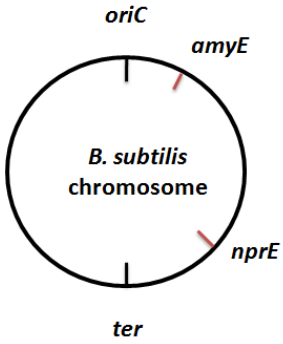

Figure 8.4 Illustration of the *amyE* and *nprEloci* in *B. subtilis* chromosome

8.3.3 LacI-expressing strain

The BSB1 derivative strain expressing the *lacI* gene was built by using two plasmids, the pDRDG2 (Borkowski, Goelzer, et al. 2016a) that containsthe *lacI* gene and the pAGP18 (kind gift from Dr. Anne-Gaëlle Planson) that contains the *sacA* homology region (as integration platform). Both plasmids were digested by the *Bam*HI and *Xma*I restriction enzymes according to the supplier's recommendations [Figure 8.5]. Fragments were purified and ligated by the T4 DNA ligase. The resulting ligation reaction was transformed into *E.coli* Mach1-T1 competent cells. Restriction analysis was performed on the purified plasmids to select the good assembly, which was then transformed into BSB1 to be integrated in the *sacAlocus*.

8.3.4 SspB-expressing strain

The *sspB* gene was amplified from the genome of *E. coli* DH5α using primers with 5' extension containing the *Bsm*BI restriction site. The P_{hs}promoter was PCR amplified as mentioned previously in [section8.3.2]. After purification of the amplified fragments, they

were ligated using the StarGate® cloning and transformed into the LacI-expressing strain[section 8.3.3] to be integrated by double cross over at the *nprElocus*.

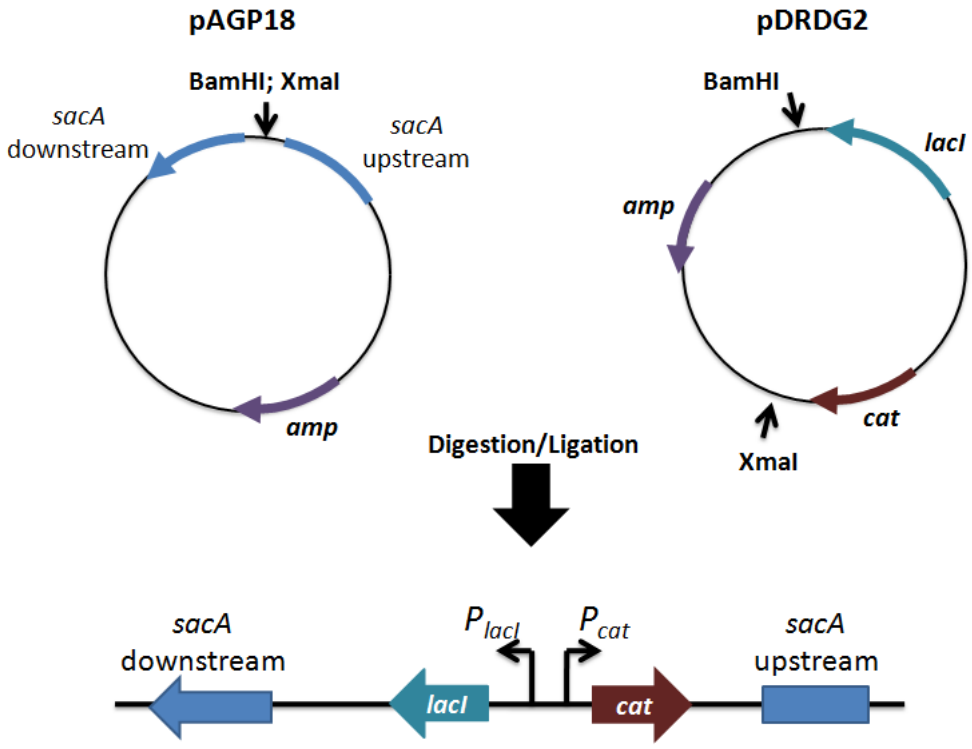

Figure 8.5 The genetic construct scheme of LacI-strain

The pDRDG2 containing the lacI gene and the chloramphenicol resistance marker was digested by BamHI and XmaI. The vector pAGP18 containing the *sacA* homology region was digested with the same restriction enzymes to open the plasmid. Then the ligation between the insert (*lacI* with Chloramphenicol) and the pAGP18 vector allowed us to obtain the genetic construct ready to integrate in the BSB1 genome in *sacAlocus*.

8.3.5 *sspB-clpXP* operon

The synthetic operon was designed to be assembled and controlled by the P_{hs} promoter. Two different design of the *sspB clpX clpP* operon were proposed. The first design resembled that of the native design of the *dhbA dhbC dhbE* operon where we made use of the very same intergenic regions (TIR sequences). The second design resembled that of the *aroF aroB aroH*operon, which does not exhibit RBS/TIR-like sequences following the second and third genes of the operon (without TIR sequences in between) [Table 8-2]. In the case of the *aroF aroB aroH*operon, it is believed that translation of the second and third genes occurs following that of the first gene without disassembly of the ribosome at the stop codon of *aroF*. These designs were based on a comparison between transcriptomic data (Nicolas et al. 2012) and proteomic data (Goelzer et al. 2015), which revealed that these two operonic structures

led to stoichiometric translation of each gene within the operon. The *sspB* gene was amplified from *E. coli* DH5α and the *clpX* and *clpP* genes were amplified from both *E. coli* DH5α and *B. subtilis* BSB1 strains. After PCR amplification, the fragments and vector were digested by *Dpn*I to digest the methylated plasmids to avoid false positive colonies. Following digestion, the fragments and vector were purified, then they were assembled using the NEBuilder® HiFi DNA Assembly protocol which is based on the Gibson assembly® method. The resulting mix was transformed into *E. coli*. After screening the colonies to obtain the correct assembly by PCR on colony, plasmid extraction was performed, plasmids were sequenced, and then transformed in the *B. subtilis* LacI-expressing strain [section 8.3.3] to be integrated in the *nprElocus* by double crossing over.

Table 8-2 The sequences of the genes composing the 4 different operons

Each gene is represented by color, sspB (blue), clpX (red), clpP (green). The intergenic regions for the dhbACE design are in bold, and the overlapping regions in the aroFBH are in underlined bold. sspB gene is from E. coli DH5α, while clpX and clpP genes are from E. coli DH5α and B. subtilis BSB1.

Name	
dhbACE_Bsu	ATGGATTTGTCACAGCTAACACCACGTCGTCCCTATCTGCTGCGTGCATTCTATGAGTGGTTGCTGGATAACCAGCTCACGCC GCACCTGGTGGTGGATGTGACGCTCCCTGGCGTGCAGGTTCCTATGGAATATGCGCGTGACGGGCAAATCGTACTCAACATTG CGCCGCGTGCTGTCGGCAATCTGGAACTGGCGAATGATGAGGTGCGCTTTAACGCGCGCTTTGGTGGCATTCCGCGTCAGGTT TCTGTGCCGCTGGCTGCCGTGCTGGCTATCTACGCCCGTGAAAATGGCGCAGGCACGATGTTTGAGCCTGAAGCTGCCTACGA TGAAGATACCAGCATCATGAATGATGAAGAGGCATCGGCAGACAACGAAACCGTTATGTCGGTTATTGATGGCGACAAGCCA GATCACGATGATGACACTCATCCTGACGATGAACCTCCGCAGCCACCACGCGGTGGTCGACCGGCATTACGCGTTGTGAAGT AACTGAATTTAAAGGAGGTTCATGAGAATGTTTAAATTTAACGAGGAAAAAGGACAATTAAAATGCTCGTTCTGTGGAAAAA CACAAGATCAGGTTCGTAAGCTTGTAGCTGGACCTGGTGTATATATATGTGACGAATGTATCGAACTCTGCACTGAAATTGTA GAGGAAGAACTCGGAACAGAAGAAGAAGTAGAATTTAAAGACGTACCAAAGCCTCAGGAAATTCGCGAAATTTTGAATGAA TATGTCATCGGCCAAGATCAAGCGAAGAAATCACTTGCTGTTGCTGTGTATAACCACTATAAGCGCATTAACTCCAACAGCAA AGTTGATGATGTTGAGCTTTCAAAAAGTAATATTTCATTAATCGGCCCTACGGGAAGCGGTAAAACCCTTCTGGCGCAAACAT TGGCTCGCATTTTAAATGTGCCGTTTGCGATTGCGGACGCTACATCTTTGACTGAAGCTGGATACGTGGGTGAAGATGTAGAG AATATTCTCTTGAAGCTCATCCAAGCTGCTGATTATGATGTGGAAAAAGCCGAAAAAGGCATTATCTATATTGATGAAATCGA TAAAGTGGCTAGAAAATCTGAAAACCCGTCTATCACACGTGATGTGTCAGGTGAGGGTGTACAGCAGGCATTGCTTAAAATT CTTGAGGGTACAGTGGCAAGCGTACCGCCTCAAGGTGGACGTAAGCATCCTCATCAAGAATTCATTCAAATTGACACAACAA ATATTTTGTTCATTTGCGGCGGAGCTTTCGATGGTATCGAACAAATCATCAAACGCCGTTTAGGCCAAAAAGTGATTGGATTC GGTGCCGACAATAAAGCTGCTGATCTTGAGAAAGAAGATCTTCTTTCAAAAGTGCTTCCAGAAGATTTGCTCCGTTTCGGGTT AATTCCTGAATTTATCGGACGTCTTCCGGTTATTGCAAGCCTTGAAAAGCTTGACGAAGAAGCATTGGTTGCGATCTTAACAA AACCGAAAAACGCTCTTGTTAAGCAATTCAAGAAAATGCTTGAGCTTGACAACGTTGAGCTTGAGTTTGAAGAAGAAGCGCT TTCTGAAATTGCTAAAAAAGCAATTGAACGCAAGACTGGAGCACGCGGACTCCGTTCTATCATTGAAGGCATCATGCTTGATG TGATGTTTGAGCTGCCTTCTCGTGATGACATTGAAAAATGTGTAATCACCGGGGCAACTGTGACACACGGAGAGCCTCCTCGC CTCTTATTAAAAGACGGCACTGAGGTAAGCCAAGATAAAACATCTGCATAAACAGAATGAATTGCTGGAGGGATACAACATG AATTTAATACCTACAGTCATTGAACAAACGAACCGCGGGGAAAGAGCGTATGACATTTATTCTCGTCTATTAAAGGATCGTAT CATCATGCTTGGATCTGCGATTGATGACAACGTTGCGAACTCCATCGTGTCACAGCTTTTATTCTTAGCAGCAGAAGACCCTG AAAAAGAAATTTCTCTTTACATCAACAGCCCGGGCGGCTCTATTACAGCCGGTATGGCGATCTATGATACCATGCAGTTTATT AAGCCGAAGGTATCTACAATTTGTATCGGTATGGCTGCGTCAATGGGCGCGTTCCTGCTTGCAGCCGGCGAAAAAGGCAAAC

	GCTATGCGCTTCCAAACAGTGAAGTCATGATTCACCAGCCTCTTGGCGGTGCGCAAGGTCAAGCGACAGAAATTGAAATTGC TGCGAAACGCATTCTCTTGCTTCGCGACAAATTAAACAAAGTCCTAGCTGAACGTACTGGCCAGCCGCTTGAAGTGATCGAAC GCGACACAGACCGTGATAACTTCAAGTCTGCTGAAGAAGCGCTTGAATACGGCCTGATTGACAAAATTTTGACTCACACAGA AGACAAAAAGTAA
dhbACE_Eco	ATGGATTTGTCACAGCTAACACCACGTCGTCCCTATCTGCTGCGTGCATTCTATGAGTGGTTGCTGGATAACCAGCTCACGCC GCACCTGGTGGTGGATGTGACGCTCCCTGGCGTGCAGGTTCCTATGGAATATGCGCGTGACGGGCAAATCGTACTCAACATTG CGCCGCGTGCTGTCGGCAATCTGGAACTGGCGAATGATGAGGTGCGCTTTAACGCGCGCTTTGGTGGCATTCCGCGTCAGGTT TCTGTGCCGCTGGCTGCCGTGCTGGCTATCTACGCCCGTGAAAATGGCGCAGGCACGATGTTTGAGCCTGAAGCTGCCTACGA TGAAGATACCAGCATCATGAATGATGAAGAGGCATCGGCAGACAACGAAACCGTTATGTCGGTTATTGATGGCGACAAGCCA GATCACGATGATGACACTCATCCTGACGATGAACCTCCGCAGCCACCACGCGGTGGTCGACCGGCATTACGCGTTGTGAAGT AA**CTGAATTTAAAGGAGGTTCATGAGA**ATGACAGATAAACGCAAAGATGGCTCAGGCAAATTGCTGTATTGCTCTTTTTGC GGCAAAAGCCAGCATGAAGTGCGCAAGCTGATTGCCGGTCCATCCGTGTATATCTGCGACGAATGTGTTGATTTATGTAACGA CATCATTCGCGAAGAGATTAAAGAAGTTGCACCGCATCGTGAACGCAGTGCGCTACCGACGCCGCATGAAATTCGCAACCAC CTGGACGATTACGTTATCGGCCAGGAACAGGCGAAAAAAGTGCTGGCGGTCGCGGTATACAACCATTACAAACGTCTGCGCA ACGGCGATACCAGCAATGGCGTCGAGTTGGGCAAAAGTAACATTCTGCTGATCGGTCCGACCGGTTCCGGTAAAACGCTGCT GGCTGAAACGCTGGCGCGCCTGCTGGATGTTCCGTTCACCATGGCCGACGCGACTACACTGACCGAAGCCGGTTATGTGGGT GAAGACGTTGAAAACATCATTCAGAAGCTGTTGCAGAAATGCGACTACGATGTCCAGAAAGCACAGCGTGGTATTGTCTACA TCGATGAAATCGACAAGATTTCTCGTAAGTCAGACAACCCGTCCATTACCCGAGACGTTTCCGGTGAAGGCGTACAGCAGGC ACTGTTGAAACTGATCGAAGGTACGGTAGCTGCTGTTCCACCGCAAGGTGGGCGTAAACATCCGCAGCAGGAATTCTTGCAG GTTGATACCTCTAAGATCCTGTTTATTTGTGGCGGTGCGTTTGCCGGTCTGGATAAAGTGATTTCCCACCGTGTAGAAACCGGC TCCGGCATTGGTTTTGGCGCGACGGTAAAAGCGAAGTCCGACAAAGCAAGCGAAGGCGAGCTGCTGGCGCAGGTTGAACCG GAAGATCTGATCAAGTTTGGTCTTATCCCTGAGTTTATTGGTCGTCTGCCGGTTGTCGCAACGTTGAATGAACTGAGCGAAGA AGCTCTGATTCAGATCCTCAAAGAGCCGAAAAACGCCCTGACCAAGCAGTATCAGGCGCTGTTTAATCTGGAAGGCGTGGAT CTGGAATTCCGTGACGAGGCGCTGGATGCTATCGCTAAGAAAGCGATGGCGCGTAAAACCGGTGCCCGTGGCCTGCGTTCCA TCGTAGAAGCCGCACTGCTCGATACCATGTACGATCTGCCGTCCATGGAAGACGTCGAAAAAGTGGTTATCGACGAGTCGGT AATTGATGGTCAAAGCAAACCGTTGCTGATTTATGGCAAGCCGGAAGCGCAACAGGCATCTGGTGAATAA**ACAGAATGAAT** **TGCTGGAGGGATACAAC**ATGTCATACAGCGGCGAACGAGATAACTTTGCACCCCATATGGCGCTGGTGCCGATGGTCATTG AACAGACCTCACGCGGTGAGCGCTCTTTTGATATCTATTCTCGTCTACTTAAGGAACGCGTCATTTTTCTGACTGGCCAGGTTG AAGACCACATGGCTAACCTGATTGTGGCGCAGATGCTGTTCCTGGAAGCGGAAAACCCAGAAAAAGATATCTATCTGTACAT TAACTCCCCAGGCGGGGTGATCACTGCCGGGATGTCTATCTATGACACCATGCAGTTTATCAAGCCTGATGTCAGCACCATCT GTATGGGCCAGGCGGCCTCGATGGGCGCTTTCTTGCTGACCGCAGGGGCAAAAGGTAAACGTTTTTGCCTGCCGAATTCGCGC GTGATGATTCACCAACCGTTGGGCGGCTACCAGGGCCAGGCGACCGATATCGAAATTCATGCCCGTGAAATTCTGAAAGTTA AAGGGCGCATGAATGAACTTATGGCGCTTCATACGGGTCAATCATTAGAACAGATTGAACGTGATACCGAGCGCGATCGCTT CCTTTCCGCCCCTGAAGCGGTGGAATACGGTCTGGTCGATTCGATTCTGACCCATCGTAATTGA
aroFBH_Bsu	ATGGATTTGTCACAGCTAACACCACGTCGTCCCTATCTGCTGCGTGCATTCTATGAGTGGTTGCTGGATAACCAGCTCACGCC

	GCACCTGGTGGTGGATGTGACGCTCCCTGGCGTGCAGGTTCCTATGGAATATGCGCGTGACGGGCAAATCGTACTCAACATTG CGCCGCGTGCTGTCGGCAATCTGGAACTGGCGAATGATGAGGTGCGCTTTAACGCGCGCTTTGGTGGCATTCCGCGTCAGGTT TCTGTGCCGCTGGCTGCCGTGCTGGCTATCTACGCCCGTGAAAATGGCGCAGGCACGATGTTTGAGCCTGAAGCTGCCTACGA TGAAGATACCAGCATCATGAATGATGAAGAGGCATCGGCAGACAACGAAACCGTTATGTCGGTTATTGATGGCGACAAGCCA GATCACGATGATGACACTCATCCTGACGATGAACCTCCGCAGCCACCACGCGGTGGTCGACCGGCATTACGCGTTGTGAAG**T** **AATG**TTTAAATTTAACGAGGAAAAAGGACAATTAAAATGCTCGTTCTGTGGAAAAACACAAGATCAGGTTCGTAAGCTTGTA GCTGGACCTGGTGTATATATATGTGACGAATGTATCGAACTCTGCACTGAAATTGTAGAGGAAGAACTCGGAACAGAAGAAG AAGTAGAATTTAAAGACGTACCAAAGCCTCAGGAAATTCGCGAAATTTTGAATGAATATGTCATCGGCCAAGATCAAGCGAA GAAATCACTTGCTGTTGCTGTGTATAACCACTATAAGCGCATTAACTCCAACAGCAAAGTTGATGATGTTGAGCTTTCAAAAA GTAATATTTCATTAATCGGCCCTACGGGAAGCGGTAAAACCCTTCTGGCGCAAACATTGGCTCGCATTTTAAATGTGCCGTTT GCGATTGCGGACGCTACATCTTTGACTGAAGCTGGATACGTGGGTGAAGATGTAGAGAATATTCTCTTGAAGCTCATCCAAGC TGCTGATTATGATGTGGAAAAAGCCGAAAAAGGCATTATCTATATTGATGAAATCGATAAAGTGGCTAGAAAATCTGAAAAC CCGTCTATCACACGTGATGTGTCAGGTGAGGGTGTACAGCAGGCATTGCTTAAAATTCTTGAGGGTACAGTGGCAAGCGTACC GCCTCAAGGTGGACGTAAGCATCCTCATCAAGAATTCATTCAAATTGACACAACAAATATTTTGTTCATTTGCGGCGGAGCTT TCGATGGTATCGAACAAATCATCAAACGCCGTTTAGGCCAAAAAGTGATTGGATTCGGTGCCGACAATAAAGCTGCTGATCTT GAGAAAGAAGATCTTCTTTCAAAAGTGCTTCCAGAAGATTTGCTCCGTTTCGGGTTAATTCCTGAATTTATCGGACGTCTTCCG GTTATTGCAAGCCTTGAAAAGCTTGACGAAGAAGCATTGGTTGCGATCTTAACAAAACCGAAAAACGCTCTTGTTAAGCAATT CAAGAAAATGCTTGAGCTTGACAACGTTGAGCTTGAGTTTGAAGAAGAAGCGCTTTCTGAAATTGCTAAAAAAGCAATTGAA CGCAAGACTGGAGCACGCGGACTCCGTTCTATCATTGAAGGCATCATGCTTGATGTGATGTTTGAGCTGCCTTCTCGTGATGA CATTGAAAAATGTGTAATCACCGGGGCAACTGTGACACACGGAGAGCCTCCTCGCCTCTTATTAAAAGACGGCACTGAGGTA AGCCAAGATAAAACATCTGCA**TGATG**AATTTAATACCTACAGTCATTGAACAAACGAACCGCGGGGAAAGAGCGTATGACA TTTATTCTCGTCTATTAAAGGATCGTATCATCATGCTTGGATCTGCGATTGATGACAACGTTGCGAACTCCATCGTGTCACAGC TTTTATTCTTAGCAGCAGAAGACCCTGAAAAAGAAATTTCTCTTTACATCAACAGCCCGGGCGGCTCTATTACAGCCGGTATG GCGATCTATGATACCATGCAGTTTATTAAGCCGAAGGTATCTACAATTTGTATCGGTATGGCTGCGTCAATGGGCGCGTTCCT GCTTGCAGCCGGCGAAAAAGGCAAACGCTATGCGCTTCCAAACAGTGAAGTCATGATTCACCAGCCTCTTGGCGGTGCGCAA GGTCAAGCGACAGAAATTGAAATTGCTGCGAAACGCATTCTCTTGCTTCGCGACAAATTAAACAAAGTCCTAGCTGAACGTA CTGGCCAGCCGCTTGAAGTGATCGAACGCGACACAGACCGTGATAACTTCAAGTCTGCTGAAGAAGCGCTTGAATACGGCCT GATTGACAAAATTTTGACTCACACAGAAGACAAAAAGTAA
aroFBH_Eco	ATGGATTTGTCACAGCTAACACCACGTCGTCCCTATCTGCTGCGTGCATTCTATGAGTGGTTGCTGGATAACCAGCTCACGCC GCACCTGGTGGTGGATGTGACGCTCCCTGGCGTGCAGGTTCCTATGGAATATGCGCGTGACGGGCAAATCGTACTCAACATTG CGCCGCGTGCTGTCGGCAATCTGGAACTGGCGAATGATGAGGTGCGCTTTAACGCGCGCTTTGGTGGCATTCCGCGTCAGGTT TCTGTGCCGCTGGCTGCCGTGCTGGCTATCTACGCCCGTGAAAATGGCGCAGGCACGATGTTTGAGCCTGAAGCTGCCTACGA TGAAGATACCAGCATCATGAATGATGAAGAGGCATCGGCAGACAACGAAACCGTTATGTCGGTTATTGATGGCGACAAGCCA GATCACGATGATGACACTCATCCTGACGATGAACCTCCGCAGCCACCACGCGGTGGTCGACCGGCATTACGCGTTGTGAAG**T** **A****ATG**ACAGATAAACGCAAAGATGGCTCAGGCAAATTGCTGTATTGCTCTTTTTGCGGCAAAAGCCAGCATGAAGTGCGCAAG

	CTGATTGCCGGTCCATCCGTGTATATCTGCGACGAATGTGTTGATTTATGTAACGACATCATTCGCGAAGAGATTAAAGAAGT TGCACCGCATCGTGAACGCAGTGCGCTACCGACGCCGCATGAAATTCGCAACCACCTGGACGATTACGTTATCGGCCAGGAA CAGGCGAAAAAAGTGCTGGCGGTCGCGGTATACAACCATTACAAACGTCTGCGCAACGGCGATACCAGCAATGGCGTCGAGT TGGGCAAAAGTAACATTCTGCTGATCGGTCCGACCGGTTCCGGTAAAACGCTGCTGGCTGAAACGCTGGCGCGCCTGCTGGA TGTTCCGTTCACCATGGCCGACGCGACTACACTGACCGAAGCCGGTTATGTGGGTGAAGACGTTGAAAACATCATTCAGAAG CTGTTGCAGAAATGCGACTACGATGTCCAGAAAGCACAGCGTGGTATTGTCTACATCGATGAAATCGACAAGATTTCTCGTAA GTCAGACAACCCGTCCATTACCCGAGACGTTTCCGGTGAAGGCGTACAGCAGGCACTGTTGAAACTGATCGAAGGTACGGTA GCTGCTGTTCCACCGCAAGGTGGGCGTAAACATCCGCAGCAGGAATTCTTGCAGGTTGATACCTCTAAGATCCTGTTTATTTG TGGCGGTGCGTTTGCCGGTCTGGATAAAGTGATTTCCCACCGTGTAGAAACCGGCTCCGGCATTGGTTTTGGCGCGACGGTAA AAGCGAAGTCCGACAAAGCAAGCGAAGGCGAGCTGCTGGCGCAGGTTGAACCGGAAGATCTGATCAAGTTTGGTCTTATCCC TGAGTTTATTGGTCGTCTGCCGGTTGTCGCAACGTTGAATGAACTGAGCGAAGAAGCTCTGATTCAGATCCTCAAAGAGCCGA AAAACGCCCTGACCAAGCAGTATCAGGCGCTGTTTAATCTGGAAGGCGTGGATCTGGAATTCCGTGACGAGGCGCTGGATGC TATCGCTAAGAAAGCGATGGCGCGTAAAACCGGTGCCCGTGGCCTGCGTTCCATCGTAGAAGCCGCACTGCTCGATACCATG TACGATCTGCCGTCCATGGAAGACGTCGAAAAAGTGGTTATCGACGAGTCGGTAATTGATGGTCAAAGCAAACCGTTGCTGA TTTATGGCAAGCCGGAAGCGCAACAGGCATCTGGTGAA**TGATG**TCATACAGCGGCGAACGAGATAACTTTGCACCCCATATG GCGCTGGTGCCGATGGTCATTGAACAGACCTCACGCGGTGAGCGCTCTTTTGATATCTATTCTCGTCTACTTAAGGAACGCGT CATTTTTCTGACTGGCCAGGTTGAAGACCACATGGCTAACCTGATTGTGGCGCAGATGCTGTTCCTGGAAGCGGAAAACCCAG AAAAAGATATCTATCTGTACATTAACTCCCCAGGCGGGGTGATCACTGCCGGGATGTCTATCTATGACACCATGCAGTTTATC AAGCCTGATGTCAGCACCATCTGTATGGGCCAGGCGGCCTCGATGGGCGCTTTCTTGCTGACCGCAGGGGCAAAAGGTAAAC GTTTTTGCCTGCCGAATTCGCGCGTGATGATTCACCAACCGTTGGGCGGCTACCAGGGCCAGGCGACCGATATCGAAATTCAT GCCCGTGAAATTCTGAAAGTTAAAGGGCGCATGAATGAACTTATGGCGCTTCATACGGGTCAATCATTAGAACAGATTGAAC GTGATACCGAGCGCGATCGCTTCCTTTCCGCCCCTGAAGCGGTGGAATACGGTCTGGTCGATTCGATTCTGACCCATCGTAAT TGA

8.4 Growth media

For the transformation of *B. subtilis*, MG1 and MG2 were prepared according to the following description; all the solutions were prepared and filter-sterilized separately.

MG1 is composed of $(NH4)_2SO_4$(2 g/L), C_6H_5Na3O7 (1 g/L), $K_2HPO_4 \cdot 3H_2O$ (14 g/L), KH_2PO_4(6 g/L), glucose(5 g/L), $MgSO_4 \cdot 7H_2O$ (40 mg/L),casein hydrolysate (2.50 mg/L) and yeast extract (10 mg/L);MG2 is composed of$(NH_4)_2SO_4$(2 g/L), $C_6H_5Na_3O_7$(1 g/L),$K_2HPO_4 \cdot 3H_2O$ (14 g/L),KH_2PO_4(6 g/L), glucose(5g/L), $MgSO_4 \cdot 7H_2O$ (100 mg/L), caseinhydrolysate (1.25 mg/L), yeast extract (2.50 mg/L) and $Ca(NO_3)_2$(0.75 mg/L).

Growth media used to study the growth rate and protein production were the:

- CH medium (Partridge & Errington 1993)

Casein hydrolysate (10 g/L), L-glutamic acid-Na-H_2O (4 g/L), L-alanine (1.3 g/L), L-asparagine (1.5 g/L), KH_2PO_4 (0.01 M), NH_4Cl (0.025 M), Na_2SO_4 (0.01 g/L), NH_4NO_3 (0.01 g/L), $FeCl_3$-6H2O (0.05 mM) ,C6H8O7-H2O (0.1 mM), $MgSO_4$-7H_2O (0.4 mM), $CaCl_2$-2H_2O (0.1 mM), $MnSO_4$-H2O (0.03 g/L).

- CHG medium is a CH with 0.5% (w/v) glucose.
- S medium is a minimal medium with 0.5% w/v glucose (Sharpe et al. 1998).

$(NH_4)_2SO_4$ (2 g/L), K_2HPO_4 (14 g/L), KH_2PO_4 (6 g/L), Sodium citrate-2H_2O (1 g/L), $MgSO_4$-7H_2O (0.2 g/L), $MnSO_4$-H_2O (0,001 g/L), Glucose (5 g/L), $FeCl_3$-6H2O 13.5 mg/L and $C_6H_8O_7$-H_2O 21 mg/L

- SX is composed of S medium with 0.15% arginine, 0.4% glutamate, 0.4% malate (w/v).
- M9 minimal medium (Chubukov et al. 2013) with 0.5% of pyruvate.
  - M9 medium (5X): Na_2HPO_4.2H_2O (42.5 g/L), KH_2PO_4 (15 g/L), NH_4Cl (5 g/L), NaCl (2.5 g/L)
  - Trace salts (100X): $MnCl_2$.4H_2O (100 mg/L), $ZnCl_2$ (170 mg/L), $CuCl_2$.2H_2O (43 mg/L), $CoCl_2$.6H_2O (60 mg/L), Na_2MoO_4.2H_2O (60 mg/L).
  - M9Pyruvate: $CaCl_2$ (0.1mM), trace salts (1X), $MgSO_4$ (1 mM), $FeCl_3$-6H2O (13.5 mg/L), $C_6H_8O_7$-H_2O (21 mg/L), sodium pyruvate (5 g/L), and M9 medium (1X)

All the stock solutions were prepared and filter-sterilized separately. The antibiotic concentration added for selection in the liquid media or agar plates were as follows: for *B. subtilis* spectinomycin (100 µg/mL), kanamycin (8 µg/mL), chloramphenicol (5 µg/mL). And for *E. coli* ampicillin (100µg/mL).

8.5 Live Cell Array

Live Cell Array is a high-throughput methodology, which allows to follow the growth of bacterial cells and gene expression in 96-well microtiter plates.

8.5.1 Culture preparation and data acquisition

The cell culture is prepared in a 96-well microtiterplate (CELLSTAR®, Greiner Bio-one) [Figure 8.6]. It starts by preparing an overnight pre-culture in LB with the appropriate antibiotic. On the day of the experiment, a first LB pre-culture is prepared by 20-fold dilution from the overnight LB pre-culture. The LB pre-culture is grown until OD_{600} of 0.3-0.4, then it is diluted in the desired medium (20-fold dilution in poor media, 100-fold dilution in rich media) to adapt to the new medium before the final culture. The second pre-culture is grown until mid-exponential phase and then it is diluted 100-fold for the final culture in the same medium. During all the experiment the culture is incubated at 37°C with continuous shaking in the Synergy ^{TM} 2 multimode microtiterplate reader (Biotek®) while measuring every 10 minutes the OD_{600}, and the fluorescence of excitation (485/20 nm) and emission (528/20 nm) for GFP, and of excitation (590/20 nm) and emission (635/20 nm) for mKate2.

8.5.2 Data treatment

The LCA data treatment was previously explained in the work of O. Borkowski et al. (Borkowski, Goelzer, et al. 2016a) which is based on the work of (Botella et al. 2010; Aïchaoui et al. 2012; Buescher 2012).

	1	2	3	4	5	6	7	8	9	10	11	12
A	X	Fluo 10nM	Fluo 1nM	X	X	X	X	X	X	X	X	X
B	X	Strain 1 Replicate 1	Strain 2 Replicate 1	Strain 3 Replicate 1	Strain 4 Replicate 1	Strain 5 Replicate 1	Strain 6 Replicate 1	Strain 7 Replicate 1	Strain 8 Replicate 1	Strain 9 Replicate 1	Strain 10 Replicate 1	X
C	X	Strain 1 Replicate 2	Strain2 Replicate 2	Strain 3 Replicate 2	Strain 4 Replicate 2	Strain 5 Replicate 2	Strain 6 Replicate 2	Strain 7 Replicate 2	Strain 8 Replicate 2	Strain 9 Replicate 2	Strain 10 Replicate 2	X
D	X	Strain 1 Replicate 3	Strain 2 Replicate 3	Strain 3 Replicate 3	Strain 4 Replicate 3	Strain 5 Replicate 3	Strain 6 Replicate 3	Strain 7 Replicate 3	Strain 8 Replicate 3	Strain 9 Replicate 3	Strain 10 Replicate 3	X
E	X	Strain 1 Replicate 4	Strain 2 Replicate 4	Strain 3 Replicate 4	Strain 4 Replicate 4	Strain 5 Replicate 4	Strain 6 Replicate 4	Strain 7 Replicate 4	Strain 8 Replicate 4	Strain 9 Replicate 4	Strain 10 Replicate 4	X
F	X	Strain 1 Replicate 5	Strain 2 Replicate 5	Strain 3 Replicate 5	Strain 4 Replicate 5	Strain 5 Replicate 5	Strain 6 Replicate 5	Strain 7 Replicate 5	Strain 8 Replicate 5	Strain 9 Replicate 5	Strain 10 Replicate 5	X
G	X	Strain 1 Replicate 6	Strain 2 Replicate 6	Strain 3 Replicate 6	Strain 4 Replicate 6	Strain 5 Replicate 6	Strain 6 Replicate 6	Strain 7 Replicate 6	Strain 8 Replicate 6	Strain 9 Replicate 6	Strain 10 Replicate 6	X
H	X	X	X	X	X	X	X	X	X	Fluo 1nM	Fluo 10nM	X

Figure 8.6 The LCA 96-well microtiterplate preparation

The exterior wells contain only sterile medium to prevent the evaporation from the inner wells where the strains are inoculated. The fluorescein of different concentrations are added in 4 wells in order to correct day-to-day possible variability in LCA experiments.

8.6 Miller Assay for beta-galactosidase activity

Beta-Galactosidase enzyme catalyzes the hydrolysis of lactose into glucose and galactose to be used as carbon and energy resources in the cell. In molecular biology, it is used as a reporter gene because it is very stable and resistant to proteolytic degradation in the celland it can be measured for its enzymatic activity. The β-Gal activity protocol; the so-called Miller Assay, was published by Jeffery Miller in 1972. Instead of lactose, the substrate used is the synthetic compound o-nitrophenyl-β-D-galactoside (ONPG) which gives o-Nitrophenol (ONP); a yellow product, when hydrolyzed by β-gal [Figure 8.7]. The activity of the enzyme corresponds to the rate of production of ONP (yellow color) measured by spectrometer at λ_{max} = 420nm

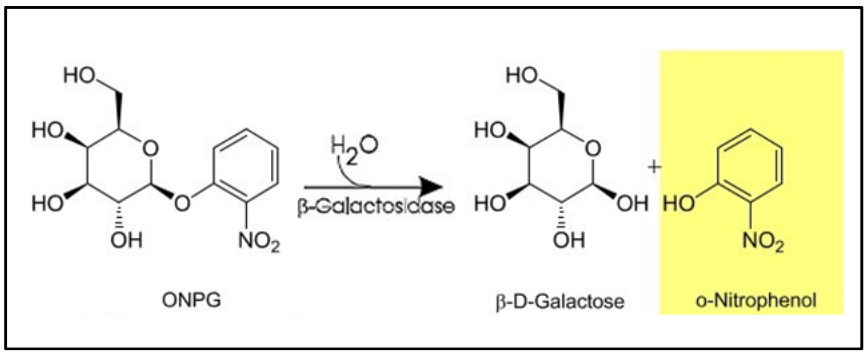

Figure 8.7 Beta-galactosidase enzymatic reaction

β-Galactosidase utilizes ONPG as a substrate to produce β-D-galactose and o-Nitrophenol, which gives the yellow color.

The strains expressing LacZ were evaluated for their β-galactosidase activity by growing them in LB medium until cells reached OD_{600}= 1. Then, 1 mL was harvested and centrifuged at 10,000 rpm for 5 minutes. The cell pellet can be stocked at -20°C if needed.

For the cell lysis, 400 µL of Z buffer (0.06M $Na_2HPO_4.7H_2O$, 0.04M $NaH_2PO_4.H_2O$,0.01M KCl, 0.001M $MgSO_4.7H_2O$, 10^{-3}M DTT, adjust at pH 7, then filter and store at room temperature) with 2 µL of lysis solution (10mg/mL lysozyme, 1.25 mg/mL DNase) were used to resuspend the cells. Then they were incubated at 37°C for 20 minutes. The tubes were centrifuged at 12000 g for 5 minutes at 4°C and the supernatant are then stored on ice.

Bradford assay was performedto quantify proteins. A protein of known concentration (usually BSA) is used to do the calibration curve. 8 dilutions were prepared (0, 10, 20, 30, 40, 50, 70, 80 µg/mL) which were measured at OD_{590} by adding 200µL Bradford solution to 800µL of

BSA samples. The extracted samples were measured at OD_{590} by adding 10µL to 790µL water and 200µL Bradford solution. The graph OD_{590} as a function of concentration was plotted for the BSA values. Then, the equation of the straight line is obtained y = mx; where y is the absorbance, m is the slope, x is the concentration of the protein. Then, to calculate the concentration of each extracted sample, x = sample absorbance/the slope of the BSA calibration curve, then multiplied by the dilution rate.

To perform the assay in a 96 well microplate, the microplate reader must be prepared to be at 28°C, and the absorbance must be set at λ_{max} = 420nm to read the absorbance of the ONP.

20µL of sample is added to 60µLof Z buffer. The reaction starts when 20µL of 4mg/mL ONPG is added (t = 0). When the yellow color starts to appear, the reaction is stopped by adding 50µL of 1M $NaCO_3$ and the time of the reaction is noted.

$$1\ Miller\ Unit = \frac{Abs_{420} * V_{reaction}}{Protein\ concentration\ * t * V_{sample} * 0.00486}$$

Abs_{420} is the absorbance of o-nitrophenol; $V_{reaction}$ is the total volume of the reaction (mL); the protein concentration (mg/mL) is calculated from Bradford assay; t is the time of the reaction in minutes; V_{sample} is the volume of the extract sample added in the reaction; the 0.00486 is the molar extinction coefficient of the o-nitrophenol.

8.7 Protein Gels and Western Blotting

8.7.1 Cell culture

Western Blots were performed to quantify GFP. Strains were grown in the desired medium (LB or CH) until cells reached the exponential phase or later (depending on the purpose of the experiment). They were harvested by taking 1 mL, centrifuged at the 8,000 g for 5 minutes and then stored at -20°C.

8.7.2 Cell lysis

Pellets were resuspended by a lysis buffer (400µL of 50mM Tris pH7.5 with 0.2M NaCl) and then sonicated. Sonication is a sound energy applied through an ultrasonic probe to agitate particles in the sample using a high ultrasonic frequency (greater than 20 kHz). This leads to cell disruption and cell content release. Each sample was sonicated for 15 seconds and repeated around 5 times. During all this time, the samples must stay on ice. Less turbidity was noticed in the samples with each turn of sonication, which indicates if they are well lysed.

When finished with cell lysis, the tubes were centrifuged at a highestspeed for30 minutes. Then, the supernatant (cell extract) was transferred to clean Eppendorf tubes and stored at -20°C or kept on ice for the next steps.

8.7.3 Protein gels and sample preparation

The protein SDS-PAGE (SDS-Polyacrylamide gel) gels are made of two parts: first, the resolving gel makes 90% of the total gel and the stacking gel that is poured on top of the first gel [Figure 8.8]. The stacking differs by the pH of the Tris-HCl buffer and it is lower in the polyacrylamide concentration. The purpose is to pack the loaded proteins so that they reach the resolving gel well aligned and thus they migrate together at the same time. The resolving gel is more concentrated in polyacrylamide, which help the proteins to migrate differently depending on their molecular weight. The SDS contained in the gel helps to give the negative charge to the proteins. For the resolving gel preparation, the 10% acrylamide final concentration contains: 3.75 mL acrylamide 29/1 (40%), 3.75 mL Tris-HCl Buffer 4X pH8.8, 7.5 mL H2O, 90µL APS (10%), 14µL Temed. The stacking gel of 5% acrylamide final contains: 625µL acrylamide 29/1 (40%), 1.25 mL Tris-HCl Buffer 4X pH6.8, 3.1 mL H2O, 50µL APS (10%), 8µL Temed.

The samples concentration were measured by Bradford assay, then the volumes are adjusted to load the same amount of proteins on the gel. The loading buffer was added to the samples, and boiled at 95°C for 5 minutes to denaturate the proteins prior to loading. The generator was set at 200 mV.

If the purpose of the experiment was to only prepare an SDS-PAGE stained gel, then gels were stained by coomassie staining buffer. However, if the purpose was to proceed to a western blot procedure we did not stain. The stacking gel part was cut, then it was prepared for the transfer procedure that is explained below.

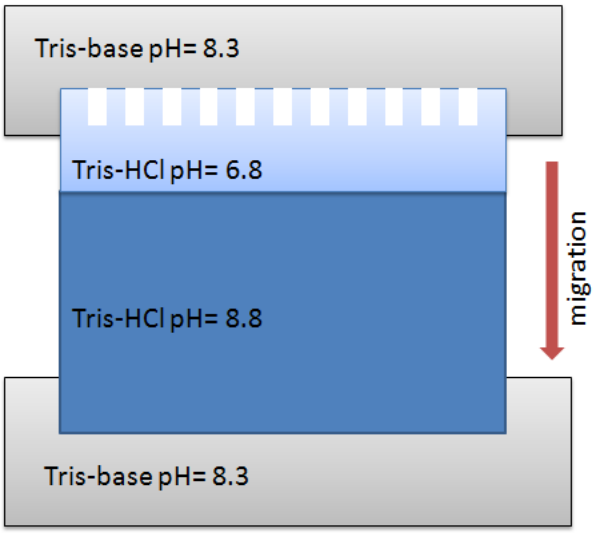

Figure 8.8 Schematic drawing representing the SDS-PAGE gel setup

The SDS-PAGE gel is the blue part; in the bottom is the resolving gel, on top is the stacking gel, and the migration buffer (Tris-base pH=8.3) is poured in the tank (grey color). The migration starts from the anode towards the cathode (the direction of the red arrow).

Transfer:the gel was incubated for 10 minutes in the transfer buffer (25mM Tris base, 190mM glycine, 10% ethanol, 0.05% SDS). During this time, the membranes (PDVF) were cut (9cm/7cm) and soaked for 30 seconds in ethanol 100%. Then they were rinsed by the transfer buffer. The watman papers were cut with the same dimensions (9cm/7cm; ~20 papers in total) and then soaked with the transfer buffer. When everything was ready, 10 papers of watman were put on the transfer electrode, then the membrane was placed on top of the watman papers, then the gel was placed on top of the membrane, and finally the remaining 10 watman papers were placed on top of the gel. The sandwich form was well soaked with transfer buffer to prevent being dried. Any bubbles must be avoided by pressing well on the watman papers. Lastly, the negative electrode was added on top of the sandwich and it was set at 36mA for overnight transfer. A scheme of the setup is represented in [Figure 8.9].

After transfer:the membrane was incubated in a blocking buffer: TBST (10mM Tris pH8, 150mM NaCl, 0.05% tween 20; the pH is adjusted to pH8 with HCl) with 5% milk. The incubation must be of at least 1 hour at ambient temperature. Then, the membrane was incubated for 3 hours in anti-GFP antibodies, which were diluted in TBST buffer with milk (5%). Then, the membrane was rinsed with TBST buffer (5x, 15 minutes incubation each time) to remove any unbound antibodies. After rinsing, the secondary antibody was diluted in the same buffer (TBST with milk) and incubated of around 2 hours at ambient

temperature.The membrane was rinsed again many times. Finally, the ECL kit of western blot from Pharmacia was used to relieve the membrane according to the supplier's instructions.

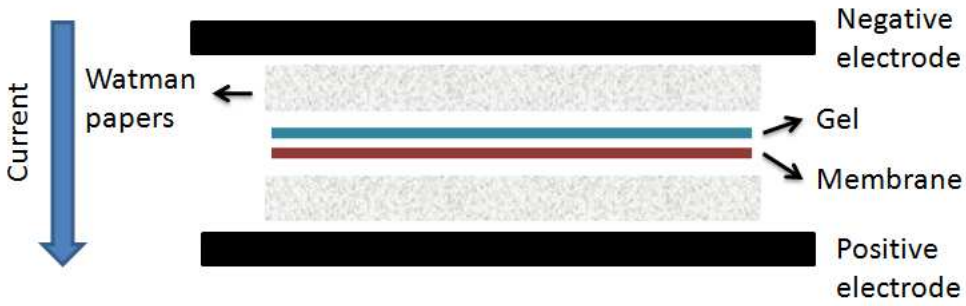

Figure 8.9Scheme showing the transfer procedure setup.

In between the piles of watman papers the gel is placed on top of the membrane so that the current passing from the negative electrode towards the positive electrode allows the transfer of the proteins from the gel to the membrane.

8.8 Flow cytometry

After cell growth, the samples were harvested for cell fixation. Cell fixation starts by resuspending the cells in 100µL of 4% formaldehyde in prepared in PBS (v/v). Then, they were incubated for 7 minutes at room temperature and centrifuged at highest speed on a bench top centrifuge for 30 seconds. The pellet was resuspended in 1mL PBS. After that, the samples were centrifuged again at a 12000 g for 30 seconds. Finally, they were resuspended in 100µL GTE buffer (20 mM Tris-HCl pH8, 10 mM EDTA, 50 mM Glucose).The flow cytometry procedure was based on the protocol used in (Verplaetse et al. 2015).

8.9 Proteomics

8.9.1 Relative protein quantification performed at the University of Greifswald-Germany

The relative quantification of the cytosolic proteins of the over-expressed ClpXP strains was performed in the University of Greifswald (Dr. Dörte Becher lab)- Germany.

The strains were grown in rich (CHG condition) and poor (S condition) media, they were harvested at OD_{600} = 0.6.

Cell wash: The cell pellet was resuspended in 500µL TE buffer (50mM Tris pH8 with 10mM EDTA) and centrifuged at a maximum speed for 5 minutes. This step was repeated. Finally, the cell pellet of the samples was resuspended in 200µL of TE buffer.

Cell disruption: The samples were transferred to tubes containing glass beads for cell lysis. The cell disruption was done through cycles of 3x (30 seconds each with cooling on ice for 5 minutes) in a FastPrep tissue homogenizer at 6.5 m/s. Then the tubes were centrifuged at a maximum speed for 5 minutes. This later step was repeated twice (15 minutes in the last repeat). Finally, the lysed samples were transferred to new tubes for the determination of protein concentration (Bradford Assay).

In solution digest for protein samples:30µg proteins of each sample were prepared in Eppendorf tubes with RapiGest (1/5 of the total digestion volume). RapiGest is a reagent, which solubilizes proteins making them more exposed to digestion. 50mM TEAB (tetraethylammonium bromide) and 0.071 mg/µL TCEP (Tris 2-carboxyethyl-phosphine hydrochloride) were then added, andthe samples were incubated at 60°C for 45 minutes.The TCEP helps the proteins to be more exposed to trypsin. Then, 0.092 mg/µL IAA (Iodoacetamide) was added and incubated for 15 minutes in the dark at room temperature. During the incubation time, 10x trypsin was activated for 15 minutes at 37°C. Then, 1.25µL of the activated trypsin was added to the samples, which were incubated after that for 5 hours at 37°C. The reaction was stopped by adding TFA (Trifluoroacetic acid) (0.1% final concentration). An incubation at 37°C for 45 minutes was followed then the tubes were centrifuged for 10 minutes at 4°C (maximum speed).

Zip-Dip purification: This is a purification method done to desalt the proteins using Millipore µC18 pipette tips. It is performed in successive steps of wetting and equilibrating the tips, peptide binding, washing, and finally eluting the purified peptides in glass vials to make them ready for the mass spectrometer measurements. It starts by wetting the tips with the 10µL wetting solution (700µL acetonitrile in 300µL milliQH_2O); the supernatant is discarded. This step is done twice. The second step continues by equilibrating the tips by 10µL (30µL acetonitrile, 1µL acetic acid, 970µL miliQ H_2O); the supernatant is discarded. This step is also done twice. Then, the sample is pipetted 10 times up and down, then the supernatant is returned to the to the Eppendorf tube. Here the peptides are bound. After that, the tips are washed by (**Buffer A**; 1µL acetic acid and 999µLmilliQ H_2O) and thesupernatant is discarded.This step is done twice. Finally, the purified samples are eluted by (600µL acetonitrile, 1µL acetic acid, and 400µL milliQ H_2O) by pipetting 5 times up and down. The eluted sample is placed in the glass vials.

Finally, the samples are concentrated by adding 15µL milliQ H_2O. Then the glass vials are placed in a vacuum centrifuge to concentrate the sample to ~5µL and to remove the acetonitrile. The volume is adjusted to 10µL with Buffer A.

The samples then will be ready to be measured by mass spectrometry. The relative quantification raw data was generated by MaxQuant which is an integrated suite of algorithms developed for high resolution, quantitative MS data (Cox & Mann 2008).

8.9.2 Relative protein quantification performed at PAPPSO platform at INRA

First the strains were grown in S medium in triplicates each until OD_{600} 0.23. From a 40 mL culture, 20 mL were harvested and centrifuged at 8000 g for 30 minutes. Then the collected pellet was lysed by sonication [explained in 8.7.2]. Next, the extracted soluble proteins were prepared to be loaded on SDS-Protein gels. The loaded extracts were run through the gel and stopped once they entered the running gel. After the gel was stained and washed [Figure 8.10 each lane was cut into small pieces in order to be treated for in-gel digestion. The later steps were performed by Dr. Celine Henry at PAPPSO.

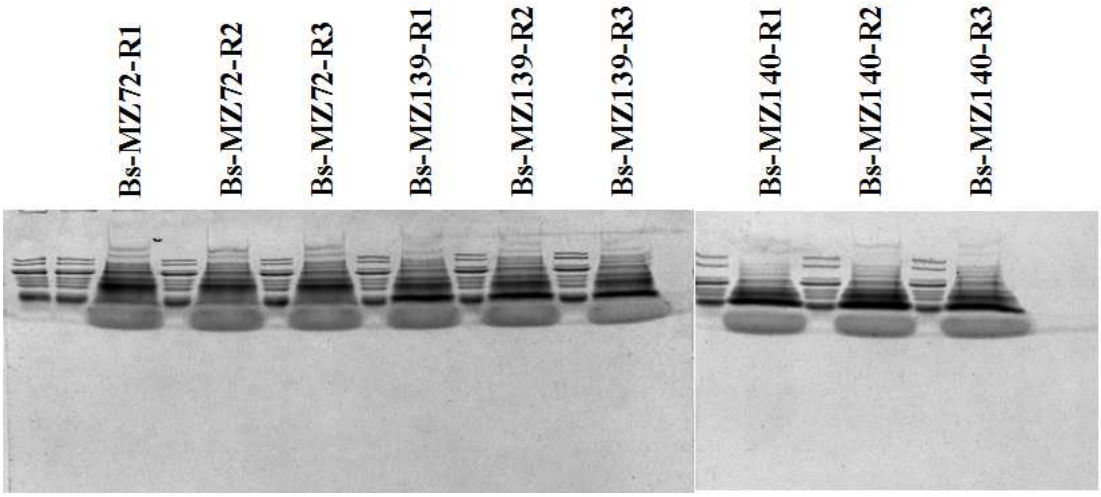

Figure 8.10 SDS-PAGE gels for the soluble protein extracts which were treated by PAPPSO

Soluble protein extracts migrated on the gel until they entered the running gel. This image is taken before being cut

Table 8-3 Strains used in this work

Name	**Relevant Genotype**	**Strain construction / plasmid → Strain[1]**
BSB1	Wild type	
Bs-OB01	Δ*amyE::lacI/cm*	(Borkowski, Goelzer, et al. 2016b)
Bs-MZ07	*Locus 1*:: P_{hs} *gfp/spec*, Δ*amyE::lacI/cm*	pMZ01 → Bs-OB01
Bs-MZ08	*Locus 1*::sP_4 *gfp/spec*, Δ*amyE::lacI/cm*	pMZ05 → Bs-OB01
Bs-MZ09	*Locus 1*::sP_5 *gfp/spec*, Δ*amyE::lacI/cm*	pMZ06 → Bs-OB01
Bs-MZ10	*Locus 1*::P_{hs} *gfp-lacZ/spec*, Δ*amyE::lacI/cm*	pMZ07 → Bs-OB01
BsMZ11	*Locus 1*::sP_4 *gfp-lacZ/spec*, Δ*amyE::lacI/cm*	pMZ11 → Bs-OB01
BsMZ12	*Locus 1*::sP_5 *gfp-lacZ/spec*, Δ*amyE::lacI/cm*	pMZ12 → Bs-OB01
Bs-MZ14	Δ*sacA::lacI/cm*	pMZ13 → BSB1
Bs-MZ19	Δ*amyE*:: P_{hs} *gfp/spec*, Δ*sacA::lacI/cm*	pCS75[2] (StarGate cloning) → Bs-MZ14
Bs-MZ21	Δ*amyE*:: P_{hs} *lacZ/spe*, Δ*sacA::lacI/cm*	pCS75[2] (StarGate cloning) → Bs-MZ14
Bs-MZ29	Δ*amyE*:: P_{hs} *gfp/spec*, Δ*nprE::* P_{hs} *gfp/kan*, Δ*sacA::lacI/cm*	pCS72[2] StarGate cloning → Bs-MZ19
Bs-MZ31	Δ*amyE*:: P_{hs} *lacZ/spec*, *Δnpr*E:: P_{hs} *lacZ/kan*, Δ*sacA::lacI/cm*	pCS72[2] StarGate cloning → Bs-MZ21
Bs-MZ38	ΔamyE::spec, Δ*sacA::lacI/cm*	pCS75[2] → Bs-MZ14
Bs-MZ43	*ΔnprE::* P_{hs} *gfp/kan*, Δ*sacA::lacI/cm*	pCS72[2] (StarGate cloning) → BsMZ14
Bs-MZ44	Δ*amyE*:: P_{veg}*sfgfp*-ALGG, Δ*sacA::lacI/cm*	pCS75[2] (StarGate cloning) → Bs-MZ14
Bs-MZ45	Δ*amyE*:: P_{veg}*sfgfp*-DDAS, Δ*sacA::lacI/cm*	pCS75[2] (StarGate cloning) → Bs-MZ14
Bs-MZ46	Δ*amyE*:: P_{veg}*sfgfp*-ADCS, Δ*sacA::lacI/cm*	pCS75[2] (StarGate cloning) → Bs-MZ14
Bs-MZ47	Δ*amyE*:: P_{veg}*sfgfp*-ADAN, Δ*sacA::lacI/cm*	pCS75[2] (StarGate cloning) → Bs-MZ14
Bs-MZ48	Δ*amyE*:: P_{veg}*sfgfp*-AASV, Δ*sacA::lacI/cm*	pCS75[2] (StarGate cloning) → Bs-MZ14
Bs-MZ49	Δ*amyE*:: P_{veg}*sfgfp*, Δ*sacA::lacI/cm*	pCS75[2] (StarGate cloning) → Bs-MZ14
Bs-MZ51	Δ*sacA::lacI/cm*, Δ*nprE::kan*	pCS72[2] → Bs-MZ14
Bs-MZ52	Δ*sacA::lacI/cm*, Δ*nprE::sspB/kan*	pCS72[2] (StarGate cloning) → Bs-MZ14
Bs-MZ54	Δ*amyE*:: P_{veg}*sfgfp*-ALGG, Δ*sacA::lacI/cm*, Δ*nprE::sspB/kan*	gDNA of MZ44 → Bs-MZ52
Bs-MZ55	Δ*amyE*:: P_{veg}*sfgfp*-DDAS, Δ*sacA::lacI/cm*, Δ*nprE::sspB/kan*	gDNA of MZ45 → Bs-MZ52
Bs-MZ56	Δ*amyE*:: P_{veg}*sfgfp*-ADCS, Δ*sacA::lacI/cm*, Δ*nprE::sspB/kan*	gDNA of MZ46 → Bs-MZ52
Bs-MZ57	Δ*amyE*:: P_{veg}*sfgfp*-ADAN, Δ*sacA::lacI/cm*, Δ*nprE::sspB/kan*	gDNA of MZ47 → Bs-MZ52
Bs-MZ58	Δ*amyE*:: P_{veg}*sfgfp*-AASV, Δ*sacA::lacI/cm*, Δ*nprE::sspB/kan*	gDNA of MZ48 → Bs_MZ52
Bs-MZ59	Δ*amyE*:: P_{veg}*sfgfp*, Δ*sacA::lacI/cm*, Δ*nprE::sspB/kan*	gDNA of MZ49 → Bs-MZ52

Bs-MZ60	Δ*amyE*:: *spec, ⊿nprE::kan*	pCS72[2] → Bs-MZ38
Bs-MZ61	Δ*amyE*:: P_{hs}*gfp/spec*, Δ*sacA::lacI/cm*, Δ*nprE::kan*	pCS75[2] (StarGate cloning)→ Bs-MZ51
Bs-MZ66	Δ*amyE*:: P_{veg}*lacZ/spec*	pCS75[2] (StarGate cloning)→ BSB1
Bs-MZ67	Δ*amyE*:: P_{veg}*lacZ*^{ALGG}*/spec*, Δ*sacA::lacI/cm*, Δ*nprE::kan*	pCS75[2] (StarGate cloning)→ Bs-MZ51
Bs-MZ69	Δ*amyE*:: P_{veg}*lacZ*^{ALGG}*/spec*, Δ*sacA::lacI/cm*, Δ*nprE::sspB/kan*	pCS75[2] (StarGate cloning)→ Bs-MZ52
Bs-MZ71	Δ*amyE*:: P_{veg}*lacZ*^{ALGG}*/spec*	pCS75[2] (StarGate cloning)→ BSB1
Bs-MZ72	Δ*amyE*::*spec*	pCS75[2](StarGate cloning)→ BSB1
Bs-MZ73	Δ*sacA::lacI/cm*, Δ*nprE::* P_{hs}^{int}*sspB clpX clpP*$_{Bsu}$*/kan*	pMZ14→ Bs-MZ14
Bs-MZ74	Δ*sacA::lacI/cm*, Δ*nprE::* P_{hs}^{int}*sspB clpX clpP*$_{Eco}$*/kan*	pMZ15→ Bs-MZ14
Bs-MZ75	Δ*sacA::lacI/cm*, Δ*nprE::* P_{hs}^{Ov}*sspB clpX clpP*$_{Bsu}$*/kan*	pMZ16→ Bs-MZ14
Bs-MZ76	Δ*sacA::lacI/cm*, Δ*nprE::* P_{hs}^{Ov}*sspB clpX clpP*$_{Eco}$*/kan*	pMZ17→ Bs-MZ14
Bs-MZ77	Δ*amyE*:: P_{veg}*gfp*^{ALGG}*/spec*,Δ*sacA::lacI/cm*, Δ*nprE::*$^{int}P_{hs}$*sspB clpX clpP*$_{Bsu}$*/kan*	gDNA of BsMZ44 → Bs-MZ73
Bs-MZ78	Δ*amyE*:: P_{veg}*gfp*^{ALGG}*/spec*, Δ*sacA::lacI/cm*, Δ*nprE::*$^{int}P_{hs}$*sspB clpX clpP*$_{Eco}$*/kan*	gDNA of BsMZ44 → Bs-MZ74
Bs-MZ79	Δ*amyE*:: P_{veg}*gfp*^{ALGG}*/spec*, Δ*sacA::lacI/cm*, Δ*nprE::*$^{Ov}P_{hs}$*sspB clpX clpP*$_{Bsu}$*/kan*	gDNA of BsMZ44 → Bs-MZ75
Bs-MZ80	Δ*amyE*:: P_{veg}*gfp*^{ALGG}*/spec*, Δ*sacA::lacI/cm*, Δ*nprE::*$^{Ov}P_{hs}$*sspB clpX clpP*$_{Eco}$*/kan*	gDNA of BsMZ44 → Bs-MZ76
Bs-MZ113	Δ*amyE*:: P_{veg}*mkate/spec*	pCS75[2] (StarGate cloning) → BSB1
Bs-MZ114	Δ*amyE*:: P_{SG22}*mkate/spec*	pCS75[2] (StarGate cloning) → BSB1
Bs-MZ115	Δ*amyE*:: P_{SG29}*mkate/spec*	pCS75[2] (StarGate cloning) → BSB1
Bs-MZ116	Δ*amyE*:: P_{SG34}*mkate/spec*	pCS75[2] (StarGate cloning) → BSB1
Bs-MZ117	Δ*amyE*:: P_{SG37}*mkate/spec*	pCS75[2] (StarGate cloning) → BSB1
Bs-MZ118	Δ*amyE*:: P_{SG38}*mkate/spec*	pCS75[2] (StarGate cloning) → BSB1
Bs-MZ137	Δ*nprE*:: P_{veg}*lacZ/kan*	pCS72[2] (StarGate cloning) → BSB1
Bs-MZ139	Δ*amyE*:: P_{veg}*gfp/ spec*	pCS75[2] (StarGate cloning) → BSB1
Bs-MZ140	Δ*amyE*:: P_{veg}*gfp/ spec*, Δ*nprE*:: P_{veg}*gfp/kan*	pCS72[2] (StarGate cloning) → Bs-MZ139
Bs-MZ141	Δ*amyE*:: P_{veg}*lacZ/spec*, Δ*nprE*:: P_{veg}*lacZ/kan*	pCS72[2] (StarGate cloning) → Bs-MZ66
Bs-MZ142	Δ*amyE*:: P_{veg}*lacZ/spec*, Δ*nprE*:: P_{veg}*gfp/kan*	pCS72[2] (StarGate cloning) → Bs-MZ137
Bs-MZ143	Δ*amyE*::*spec*, Δ*nprE*::*kan*	pCS72[2] → BsMZ72
Bs-MZ144	Δ*nprE*:: P_{veg}^{Int}*clpX clpP/kan*	pCS72[2]→ BSB1
Bs-MZ145	Δ*nprE*:: P_{veg}*sspB/kan*	pCS72[2]→ BSB1

Bs-MZ150	Δ*amyE*:: P_{SG22}*gfp*^{ALGG}*/spec*,Δ*sacA::lacI/cm*, Δ*nprE::*$^{int}P_{hs}$*sspB clpX clpP*$_{Bsu}$*/kan*	pCS75[2] (StarGate cloning) → Bs-MZ73
Bs-MZ151	Δ*amyE*:: P_{SG29}*gfp*^{ALGG}*/spec*,Δ*sacA::lacI/cm*, Δ*nprE::*$^{int}P_{hs}$*sspB clpX clpP*$_{Bsu}$*/kan*	pCS75[2] (StarGate cloning) → Bs-MZ73
Bs-MZ152	Δ*amyE*:: P_{SG34}*gfp*^{ALGG}*/spec*,Δ*sacA::lacI/cm*, Δ*nprE::*$^{int}P_{hs}$*sspB clpX clpP*$_{Bsu}$*/kan*	pCS75[2] (StarGate cloning) → Bs-MZ73
Bs-MZ153	Δ*amyE*:: P_{SG36}*gfp*^{ALGG}*/spec*,Δ*sacA::lacI/cm*, Δ*nprE::*$^{int}P_{hs}$*sspB clpX clpP*$_{Bsu}$*/kan*	pCS75[2] (StarGate cloning) → Bs-MZ73
Bs-MZ154	Δ*amyE*:: P_{SG37}*gfp*^{ALGG}*/spec*,Δ*sacA::lacI/cm*, Δ*nprE::*$^{int}P_{hs}$*sspB clpX clpP*$_{Bsu}$*/kan*	pCS75[2] (StarGate cloning) → Bs-MZ73
Bs-MZ155	Δ*amyE*:: P_{SG38}*gfp*^{ALGG}*/spec*,Δ*sacA::lacI/cm*, Δ*nprE::*$^{int}P_{hs}$*sspB clpX clpP*$_{Bsu}$*/kan*	pCS75[2] (StarGate cloning) → Bs-MZ73
Bs-MZ157	Δ*amyE*:: P_{veg}*mkate*^{ALGG}*/spec*,Δ*sacA::lacI/cm*, Δ*nprE::*$^{int}P_{hs}$*sspB clpX clpP*$_{Bsu}$*/kan*	pCS75[2] (StarGate cloning) → Bs-MZ73
Bs-MZ158	Δ*amyE*:: P_{SG22}*mkate*^{ALGG}*/spec*,Δ*sacA::lacI/cm*,Δ*nprE::*$^{int}P_{hs}$*sspB clpX clpP*$_{Bsu}$*/kan*	pCS75[2] (StarGate cloning) → Bs-MZ73
Bs-MZ159	Δ*amyE*:: P_{SG29}*mkate*^{ALGG}*/spec*,Δ*sacA::lacI/cm*,Δ*nprE::*$^{int}P_{hs}$*sspB clpX clpP*$_{Bsu}$*/kan*	pCS75[2] (StarGate cloning) → Bs-MZ73
Bs-MZ160	Δ*amyE*:: P_{SG34}*mkate*^{ALGG}*/spec*,Δ*sacA::lacI/cm*,Δ*nprE::*$^{int}P_{hs}$*sspB clpX clpP*$_{Bsu}$*/kan*	pCS75[2] (StarGate cloning) → Bs-MZ73
Bs-MZ161	Δ*amyE*:: P_{SG37}*mkate*^{ALGG}*/spec*,Δ*sacA::lacI/cm*,Δ*nprE::*$^{int}P_{hs}$*sspB clpX clpP*$_{Bsu}$*/kan*	pCS75[2] (StarGate cloning) → Bs-MZ73
Bs-MZ162	Δ*amyE*:: P_{SG38}*mkate*^{ALGG}*/spec*,Δ*sacA::lacI/cm*,Δ*nprE::*$^{int}P_{hs}$*sspB clpX clpP*$_{Bsu}$*/kan*	pCS75[2] (StarGate cloning) → Bs-MZ73

- *Locus 1* coordinates: 213017 → 213757
- [1] Plasmids transformed (→) in the indicated strains.
- [2] pCS75 and pCS72 are the inhouse plasmids from DSM-Netherlands used to construct the strains by StarGate cloning.

Table 8-4 The primers used for the strain constructions.

Plasmid/ construction	**Amplified sequence**	**Forward primer**	**Reverse primer**
VS-P1/ backbone plasmid for *Locus 1* integartion	Amplification on PL1S03 All plasmid downstream *gfp* to add BglII and XmaI	VS-31 CCCCACTAGTGGATCCATGCGTAAAGGAGAAGAACTTTTCACTG	VS-32 CCCCGGATCCACTAGTGCTGGGAAAGCCCGCGGTAAAAGTCGAC
VS-P2/ backbone plasmid for *Locus 1* integartion	All plasmid upstream *gfp* to add BamHI and SpeI restriction	VS-33 GGGGCCCGGGAGATCTTAAGCTTAATTAGCTGAGCTTGGACTCCTG	VS-34 GGGGAGATCTCCCGGGTTTGTATAGTTCATCCATGCCATGTG
*lacZ*fussion to *gfp*	Amplification on *lacZ* to add the linker sequence (pdg1661)	Pr-MZ05 GCATGGATGAACTATACAAAGCTAGCGGCGGCGGCGGCTCAGGCGGCGGCGGCTCAATGGAAGTTACTGACGTAAGATTAC	Pr-MZ06 CCAAGCTCAGCTAATTAAGCTTATTTTTGACACCAGACCAAC
pMZ-01, pMZ-07/ addition of P_{hs} promoter	Amplification of P_{hs}	Pr-MZ12-Phs CGTCGACTTTTACCGCGGGCTTTCCCAGCAAATGTGAGCACTCACAATTC	Pr-MZ13-Phs CCAGTGAAAAGTTCTTCTCCTTTACGCATAGTAGTTCCTCCTTATGTAAG
StarGate cloning (Bs-MZ19, Bs-MZ29, Bs-MZ43)	• Amplification of *gfp* • Amplification of P_{hs}	• Pr-MZ40 AGCGCGTCTCCTATGCGTAAAGGAGAAGAAC • Pr-MZ34 AGCGCGTCTCCCCGCAAATGTGAGCACTCACAATTC	• Pr-MZ41 AGCGCGTCTCCTTATTTGTATAGTTCATCCATGCC • Pr-MZ35 AGCGCGTCTCCCATAAGTAGTTCCTCCTTATGTAAG
StarGate cloning (Bs-MZ21, Bs-MZ31)	• Amplification of *lacZ* • Amplification of P_{hs}	• Pr-MZ36 AGCGCGTCTCCTATGGAAGTTACTGACGTAAG • Pr-MZ34 AGCGCGTCTCCCCGCAAATGTGAGCACTCACAATTC	Pr-MZ38 AGCGCGTCTCCTTATTTTTGACACCAGACCAACTG • Pr-MZ35 AGCGCGTCTCCCATAAGTAGTTCCTCCTTATGTAAG
StarGate cloning for Bs-MZ44	Amplification of P_{veg}*sfgfp*to obtain	VS-34	Pr-MZ55

	P_{veg} *gfp*^{ALGG}	AGCGCGTCTCCCCGCAATTTTG TCAAAATAATTTTATTGACAAC G	AGCGCGTCTCCTTATTAGCCGC CAAGAGCATAATTTTCGCTATA ATTTTCATCATTTGCAGCTTTAT AAAGTTCGTCCATACCGTGAG
StarGate cloning for Bs-MZ45	Amplification of P_{veg}*sfgfp*to obtain P_{veg} *gfp*^{DDAS}	VS-34 AGCGCGTCTCCCCGCAATTTTG TCAAAATAATTTTATTGACAAC G	Pr-MZ56 AGCGCGTCTCCTTATTATGAAG CATCATCATAATTTTCGCTATA ATTTTCATCATTTGCAGCTTTAT AAAGTTCGTCCATACCGTGAG
StarGate cloning for Bs-MZ46	Amplification of P_{veg}*sfgfp*to obtain P_{veg} *gfp*^{ADCS}	VS-34 AGCGCGTCTCCCCGCAATTTTG TCAAAATAATTTTATTGACAAC G	Pr-MZ57 AGCGCGTCTCCTTATTATGAGC AATCAGCATAATTTTCGCTATA ATTTTCATCATTTGCAGCTTTAT AAAGTTCGTCCATACCGTGAG
StarGate cloning for Bs-MZ47	Amplification of P_{veg}*sfgfp*to obtain P_{veg}*gfp*^{ADAN}	VS-34 AGCGCGTCTCCCCGCAATTTTG TCAAAATAATTTTATTGACAAC G	Pr-MZ58 AGCGCGTCTCCTTATTAATTAG CATCAGCATAATTTTCGCTATA ATTTTCATCATTTGCAGCTTTAT AAAGTTCGTCCATACCGTGAG
StarGate cloning for Bs-MZ48	Amplification of P_{veg}*sfgfp*to obtain P_{veg} *gfp*^{AASV}	VS-34 AGCGCGTCTCCCCGCAATTTTG TCAAAATAATTTTATTGACAAC G	Pr-MZ59 AGCGCGTCTCCTTATTAAACTG ATGCAGCAACATTTTGATTAAA TGAATTTGTTTTGCCTGCTTTAT AAAGTTCGTCCATACCGTGAG
StarGate cloning for Bs-MZ49	Amplification of P_{veg}*sfgfp*	VS-34 AGCGCGTCTCCCCGCAATTTTG TCAAAATAATTTTATTGACAAC G	Pr-MZ60 AGCGCGTCTCCTTATTATTTAT AAAGTTCGTCCATACCGTG
StarGate cloning for Bs-MZ52	• Amplification of *sspB* • Amplification of P_{hs}	• Pr-MZ61 AGCGCGTCTCCTATGGATTTGT CACAGCTAACAC • Pr-MZ34 AGCGCGTCTCCCCGCAAATGTG AGCACTCACAATTC	• Pr-MZ62 AGCGCGTCTCCTTATTACTTCA CAACGCGTAATG • Pr-MZ35 AGCGCGTCTCCCATAAGTAGTT CCTCCTTATGTAAG

StarGate cloning for Bs-MZ66	• Amplification of P_{veg} • Amplification of *lacZ*	• VS-43 AGCGCGTCTCCCCGCAATTTTGTCAAAATAATTTTATTGACAACG • Pr-MZ36 AGCGCGTCTCCTATGGAAGTTACTGACGTAAG	• VS-44 GCGATTAACTAATAAGGAGGACAAACAGCGCGTCTCCCATG • Pr-MZ38 AGCGCGTCTCCTTATTTTTGACACCAGACCAACTG
StarGate cloning for Bs-MZ67	• Amplification of P_{veg} • Amplification of $lacZ^{ALGG}$	• VS-43 AGCGCGTCTCCCCGCAATTTTGTCAAAATAATTTTATTGACAACG • Pr-MZ36 AGCGCGTCTCCTATGGAAGTTACTGACGTAAG	• VS-44 GCGATTAACTAATAAGGAGGACAAACAGCGCGTCTCCCATG • Pr-MZ37 AGCGCGTCTCCTTATTAGCCGCCAAGAGCATAATTTTCGCTATAATTTTCATCATTTGCAGCTTTTTGACACCAGACCAACTG
pMZ14	• Amplification on pCS72 • Amplification on gDNA of Bs-MZ52 of phs-sspB • Amplification on gDNA of BSB1 of *clpX* • Amplification on gDNA of BSB1 of *clpP*	• Pr-MZ100 GACTCACACAGAAGACAAAAAGTAAGCGCGTCTCCATAAGTTTAAAC • Pr-MZ98 GGCGAAAATGAGACGTTGATCGGCAAAATGTGAGCACTCACAATTCA • Pr-MZ71 CTGAATTTAAAGGAGGTTCATGAGAATGTTTAAATTTAACGAGG • Pr-MZ73 ACAGAATGAATTGCTGGAGGGATACAACATGAATTTAATACCTACAGTCATTGAAC	• Pr-MZ97 AAATGAATTGTGAGTGCTCACATTTTGCCGATCAACGTCTCATTTTC • Pr-MZ70 TCTCATGAACCTCCTTTAAATTCAGTTACTTCACAACGCGTAATG • Pr-MZ72 GTTGTATCCCTCCAGCAATTCATTCTGTTTATGCAGATGTTTTATCTTGGC • Pr-MZ99 TTTGTTTAAACTTATGGAGACGCGCTTACTTTTTGTCTTCTGTGTGAG
pMZ15	• Amplification on pCS72 • Amplification on gDNA of Bs-MZ52 of phs-sspB	• Pr-MZ102 TTCGATTCTGACCCATCGTAATTGAGCGCGTCTCCATAAGTTTAAAC	• Pr-MZ97 AAATGAATTGTGAGTGCTCACATTTTGCCGATCAACGTCTCATTTTC

	• Amplification on gDNA of BSB1 of *clpX* • Amplification on gDNA of BSB1 of *clpP*	• Pr-MZ98 GGCGAAAATGAGACGTTGATC GGCAAAATGTGAGCACTCACA ATTCA • Pr-MZ75 CTGAATTTAAAGGAGGTTCATG AGAATGACAGATAAACGCAAA GATGG • Pr-MZ77 ACAGAATGAATTGCTGGAGGG ATACAACATGTCATACAGCGG CGAAC	• Pr-MZ70 TCTCATGAACCTCCTTTAAATT CAGTTACTTCACAACGCGTAAT G • Pr-MZ76 GTTGTATCCCTCCAGCAATTCA TTCTGTTTATTCACCAGATGCC TGTTG • Pr-MZ101 TTTGTTTAAACTTATGGAGACG CGCTCAATTACGATGGGTCAGA ATC
pMZ16	• Amplification on pCS72 • Amplification on gDNA of Bs-MZ52 of phs-sspB • Amplification on gDNA of BSB1 of *clpX* • Amplification on gDNA of BSB1 of *clpP*	• Pr-MZ100 GACTCACACAGAAGACAAAAA GTAAGCGCGTCTCCATAAGTTT AAAC • Pr-MZ98 GGCGAAAATGAGACGTTGATC GGCAAAATGTGAGCACTCACA ATTCA • Pr-MZ80 CATTACGCGTTGTGAAGTAATG TTTAAATTTAACGAGGAAAAA GG • Pr-MZ82 AAGATAAAACATCTGCATGAT GAATTTAATACCTACAGTCATT G	• Pr-MZ97 AAATGAATTGTGAGTGCTCACA TTTTGCCGATCAACGTCTCATT TTC • Pr-MZ79 TTCCTCGTTAAATTTAAACATT ACTTCACAACGCGTAAT • Pr-MZ81 CTGTAGGTATTAAATTCATCAT GCAGATGTTTTATCTTGGC • Pr-MZ99 TTTGTTTAAACTTATGGAGACG CGCTTACTTTTTGTCTTCTGTGT GAG
pMZ17	• Amplification on pCS72 • Amplification on gDNA of Bs-MZ52 of phs-sspB	• Pr-MZ102 TTCGATTCTGACCCATCGTAAT TGAGCGCGTCTCCATAAGTTTA AAC • Pr-MZ98	• Pr-MZ97 AAATGAATTGTGAGTGCTCACA TTTTGCCGATCAACGTCTCATT TTC • Pr-MZ83

	• Amplification on gDNA of BSB1 of *clpX* • Amplification on gDNA of BSB1 of *clpP*	GGCGAAAATGAGACGTTGATC GGCAAAATGTGAGCACTCACA ATTCA • Pr-MZ84 ATTACGCGTTGTGAAGTAATGA CAGATAAACGCAAAGATG • Pr-MZ86 AACAGGCATCTGGTGAATGAT GTCATACAGCGGCGAAC	TCTTTGCGTTTATCTGTCATTAC TTCACAACGCGTAAT • Pr-MZ85 CGTTCGCCGCTGTATGACATCA TTCACCAGATGCCTGTTGC • Pr-MZ101 TTTGTTTAAACTTATGGAGACG CGCTCAATTACGATGGGTCAGA ATC
pMZ18	• Amplification of pCS72 • Amplification of *clpX clpP*from pMZ14	• Pr-MZ100 GACTCACACAGAAGACAAAAA GTAAGCGCGTCTCCATAAGTTT AAAC • Pr-MZ138 GGCGAAAATGAGACGTTGATC GGCAAATTTTGTCAAAATAATT TTATTGACAACGTCTTATTAAC GTTGATACCGGTTAAATTTTAT TTGACAAAAATGGGCTCGTGTT GTACAATAAATG	• Pr-MZ139 CTTTTTCCTCGTTAAATTTAAA CATGTTTGTCCTCCTTATTAGTT AATCGCTAGCACATTTATTGTA CAACACGAGCCCATTTTTGTCA AATAAAATTTAACCGGTATCAA CGTTAATAAGAC • Pr-MZ99 TTTGTTTAAACTTATGGAGACG CGCTTACTTTTTGTCTTCTGTGT GAG
Bs-MZ150	• Amplification of P_{SG22}*sfgfp*^{ALGG}	• Pr-MZ156 AGCGCGTCTCCCCGCAATTTTG TCAAAATAATTTTATTG	• Pr-MZ55 AGCGCGTCTCCTTATTAGCCGC CAAGAGCATAATTTTCGCTATA ATTTTCATCATTTGCAGCTTTAT AAAGTTCGTCCATACCGTGAG
Bs-MZ151	• Amplification of P_{SG29} *sfgfp*^{ALGG}	• VS-43 AGCGCGTCTCCCCGCAATTTTG TCAAAATAATTTTATTGACAAC G	• Pr-MZ55 AGCGCGTCTCCTTATTAGCCGC CAAGAGCATAATTTTCGCTATA ATTTTCATCATTTGCAGCTTTAT AAAGTTCGTCCATACCGTGAG
Bs-MZ152	• Amplification of P_{SG34} *sfgfp*^{ALGG}	• Pr-MZ153 AGCGCGTCTCCCCGCCCTTTCT TCTTGACTTGATTTCAC	• Pr-MZ55 AGCGCGTCTCCTTATTAGCCGC CAAGAGCATAATTTTCGCTATA

			ATTTTCATCATTTGCAGCTTTAT AAAGTTCGTCCATACCGTGAG
Bs-MZ153	• Amplification of P_{SG36} *sfgfp*^{ALGG}	• Pr-MZ151 AGCGCGTCTCCCCGCACTGCGT CAATACACGTTGAC	• Pr-MZ55 AGCGCGTCTCCTTATTAGCCGC CAAGAGCATAATTTTCGCTATA ATTTTCATCATTTGCAGCTTTAT AAAGTTCGTCCATACCGTGAG
Bs-MZ154	• Amplification of P_{SG37} *sfgfp*^{ALGG}	• Pr-MZ150 AGCGCGTCTCCCCGCGTTAAGA TGGCAAGCTTGAC	• Pr-MZ55 AGCGCGTCTCCTTATTAGCCGC CAAGAGCATAATTTTCGCTATA ATTTTCATCATTTGCAGCTTTAT AAAGTTCGTCCATACCGTGAG
Bs-MZ155	• Amplification of P_{SG38} *sfgfp*^{ALGG}	• Pr-MZ149 AGCGCGTCTCCCCGCAAAAGG GCTTAAATGTTTGC	• Pr-MZ55 AGCGCGTCTCCTTATTAGCCGC CAAGAGCATAATTTTCGCTATA ATTTTCATCATTTGCAGCTTTAT AAAGTTCGTCCATACCGTGAG
Bs-MZ113	• Amplification of P_{veg} • Amplification of*mkate*	• VS-43 AGCGCGTCTCCCCGCAATTTTG TCAAAATAATTTTATTGACAAC G • Pr-MZ157 AGCGCGTCTCCTATGTCAGAAC TAATCAAAGAG	• Pr-MZ87 AGCGCGTCTCCCATAGTTTGTC CTCCTTATTAGTTAATCGC • Pr-MZ158 AGCGCGTCTCCTTATTAGCGAT GTCCCAGTTTAG
Bs-MZ159	• Amplification of *mkate*^{ALGG}	• Pr-MZ157 AGCGCGTCTCCTATGTCAGAAC TAATCAAAGAG	• Pr-MZ159 AGCGCGTCTCCTTATTAGCCGC CAAGAGCATAATTTTCGCTATA ATTTTCATCATTTGCAGCGCGA TGTCCCAGTTTAGACG

Note: For the remaining strains of P_{SG22}, P_{SG29}, P_{SG34}, P_{SG37}, P_{SG38}, the forward primer on each of the promoters was the same one used as in the strains with *sfgfp*. While the reverse one was Pr-MZ87.

CHAPTIRE 9

Résumé Détaillé en Français

9 Résumé Détaillé en Français

9.1 Introduction

La croissance du marché des enzymes industrielles a encouragé les scientifiques à mettre au point des outils biologiques pour la production et la sécrétion de protéines. Des micro-organismes tels que *Bacillus subtilis, Escherichia coli, Saccharomyces cerevisiae* sont très utilisés dans l'industrie pour la production de protéines. Les tentatives d'amélioration de la production et de la sécrétion de protéines dans les micro-organismes sont un sujet important de la recherche actuelle. Dans mon projet de doctorat, nous nous sommes concentrés sur *B. subtilis* comme organisme modèle pour étudier la production de protéines recombinantes.

Bacillus subtilis est une bactérie Gram positif en forme de bacille. Les bactéries Gram-positives (G^{+ve}) ont une seule membrane sur la face interne d'un peptidoglycane épais, contrairement aux bactéries Gram-négatives (G^{-ve}), qui ont une membrane interne et externe, et un espace périplasmique en forme de gel entre elles. De plus, *B. subtilis* peut former des endospores dormantes résistantes dans des conditions de stress environnemental ou en cas de carence nutritionnelle (Earl et al. 2010). *B. subtilis* n'est pas pathogène et est donc considéré comme GRAS (Genetically Recognized as Safe) par la Food and Drug Administration américaine. *B. subtilis* est utilisé commercialement depuis longtemps en raison de ses enzymes sécrétées comme les amylases, les lipases et les protéases, de ses propriétés antipathogènes et de son utilisation dans la production alimentaire du Natto. *B. subtilis* est génétiquement modifiable, facile à manipuler et naturellement compétente (Solomon & Grossman 1996). En raison de ces propriétés bénéfiques, cet organisme modèle a fait l'objet de nombreuses études.

Les espèces *Bacillus subtilis* et *Escherichia coli* sont les procaryotes les plus couramment utilisés pour la production industrielle de protéines recombinantes (Westers et al. 2004). Les *Bacillus* sp. contribuent à environ 60 % des enzymes commerciales disponibles, tandis que *E. coli* est principalement utilisé pour la production industrielle de protéines pharmaceutiques (Westers et al. 2004). Cependant, la production de protéines recombinantes dans les micro-organismes s'accompagne de quelques difficultés. La surproduction d'une protéine hétérologue entraîne une diminution du taux de croissance des microorganismes (Dong et al. 1995 ; Kafri, Metzl-Raz, et al. 2016). De plus, la production de protéines hétérologues s'accompagne de rendements sécrétés inférieurs aux quantités prévues (20-25 g/L) chez les espèces de *Bacillus* (Ploss et al. 2016). Chez *B. subtilis*, les goulots

d'étranglement ont été identifiés au niveau de la sécrétion, ce qu'on appelle le stress de sécrétion. Les tentatives de compréhension des goulots d'étranglement se concentrent sur l'étude des mécanismes de sécrétion, des complexes et interactions des protéines membranaires et des voies de signalisation cellulaire pour obtenir les souches super sécrétrices. Mais jusqu'à aujourd'hui, les conséquences d'un niveau de production élevé sur la physiologie cellulaire et le taux de croissance de *B. subtilis* n'ont pas encore été documentées. Lorsque la cellule est submergée de circuits synthétiques, on s'attend à ce que, jusqu'à un certain point, ce système perturbe la cellule en consommant toutes sortes de ressources (machines, énergie, métabolites, etc.). Par conséquent, dans mon projet de doctorat, nous voulions (1) comprendre la physiologie cellulaire, étudier les effets de la surproduction de protéines sur la physiologie cellulaire et sur différents processus cellulaires, (2) identifier les goulots d'étranglement et le type de ressource limitante, (3) et surmonter cette limitation pour améliorer la production de protéines.

9.2 Résultats

La première partie de mon projet de doctorat est consacrée à l'analyse des conséquences de la production élevéede protéines sur *B. subtilis*. Mes recherches se sont concentrées sur le niveau d'expression des gènes et sur les conséquences surla physiologie cellulaire. L'hypothèse la plus probable actuellement est que la surexpression d'un gène hétérologue est coûteuse pour la cellule en termes de ressources cellulaires et d'occupation de l'espace cytosolique. Par conséquent, on s'attend à une diminution du taux de croissance cellulaire. Des études récentes sur *B. subtilis* ont montré que ces micro-organismes peuvent tolérer de façon inattendue la production de très grandes quantités de protéines sans avoir d'effet sur le taux de croissance des cellules cultivées sur un milieu LB riche (Sauer et al. 2018). Dans le présent travail, nous visons à revisiter ces résultats et à conclure fermement sur l'effet de la surproduction de protéines sur la physiologie cellulaire en utilisant divers milieux de croissance de composition définie (du milieu pauvre au milieu riche). Pour cela, nous avons construit des mutants de souches de *B. subtilis* exprimant des gènes hétérologues, et la croissance et l'expression des gènes mutants ont été suivies dans différents milieux de croissance.

Le niveau de surexpression des gènes a été contrôlé par trois moyens différents :

1. Le système d'expression (c'est-à-dire la " force " des séquences génétiques qui contrôlent la transcription et la traduction),

2. Le gène d'intérêt (c'est-à-dire des séquences codantes courtes ou longues),
3. L'emplacement du gène d'intérêt sur le chromosome circulaire (c'est-à-dire proche ou éloigné de l'origine de la réplication).

9.2.1 Le choix des promoteurs

Tout d'abord, nous avons construit un ensemble de promoteurs inductibles basés sur l'architecture du promoteur inductible à l'IPTG P_{hs}. L'IPTG induit l'expression en se liant au répresseur LacI de *E. coli*, empêchant ainsi LacI de se lier aux séquences d'ADN de l'opérateur *lacO* (présent dans le P_{hs}), près de la région de liaison de l'ARN polymérase. Par conséquent, l'utilisation d'un promoteur inductible par IPTG nécessite d'abord l'insertion du gène *lacI* dans le génome de *B. subtilis*. P_{hs} a été associé en tandem avec des promoteurs constitutifs (avec le P_{hs} en aval d'un promoteur constitutif tel que $P_{const}P_{hs}$) afin d'augmenter le niveau d'expression par rapport au P_{hs} seul en présence d'IPTG. L'utilisation de différents promoteurs constitutifs permettra d'obtenir une gamme de force. En outre, les différents promoteurs, seuls ou en tandem, ont été couplés à des régions d'initiation dela traduction (TIR) variables. Un résumé des promoteurs et des TIR est présenté au Tableau 9-1.

Les promoteurs synthétiques inductibles ont été fusionnés à un gène *gfp*. Les constructions génétiques ont ensuite été intégrées dans le chromosome par un seul croisement à un locus proche de l'origine de réplication pour maintenir un dosage élevé du gène. La croissance des souches et la fluorescence de la GFP ont été suivies par "Live Cell Array (LCA)". Il s'agit d'une méthodologie à haut débit qui permet de suivre la croissance des cellules de *B. subtilis* et l'expression d'un gène d'intérêt dans un lecteur de plaque 96 puits (Botella et al. 2010 ; Buescher et al. 2012). Pour évaluer la force des différentes constructions génétiques synthétiques, des mesures ont été effectuées en milieu riche.

Tableau 9-1 Promoteurs constitutifs et éléments TIR constituant chacun des promoteurs synthétiques inductibles

Nom	Promoteur	TIR
sP_1	P_{ylxM}	$^{gtlX}TIR_{gtlX}$
sP_2	P_{fbaA}	$^{fbaA}TIR_{short}$
sP_3	P_{ylxM}	$^{hs}TIR_{fbaA}$
sP_4	P_{rrnJP2}	$^{gtlX}TIR_{gtlX}$
sP_5	-	$^{gtlX}TIR_{gtlX}$

Les résultats présentés en Figure 9.1 montrent une gamme de production de GFP. De plus, sans induction à l'IPTG, il n'y a pas de production de GFP, ce qui signifie que les promoteurs sont bien contrôlés. Nous nous attendions à ce que l'action de deux promoteurs en tandem entraîne des niveaux d'expression plus élevés que le P_{hs} natif. Cependant, les résultats ont montré que P_{hs} et sP_4 ($P_{rrnJP2}^{gtlX}TIR_{gtlX}$*gfp*) sont les promoteurs les plus puissants parmi les promoteurs synthétiques.

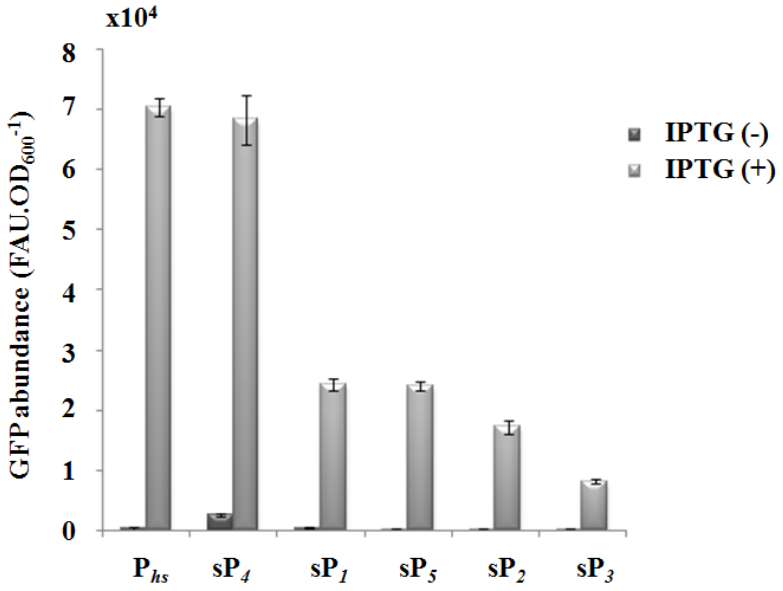

Figure 9.1 Variation de l'abondance de GFP avec les promoteurs synthétiques (sP)

Les mesures en LCA en milieu riche CHG avec induction IPTG (1mM) montrent l'abondance de la GFP lorsqu'elle est exprimée sous le contrôle des promoteurs synthétiques (sP) et du P_{hs}.

9.2.2 Le choix de protéine

Ensuite, nous avons décidé de construire une protéine de fusion entre GFP et β-Galactosidase de *E. coli*. Cette fusion devrait produire une protéine coûteuse en acides aminés. De plus, son expression devrait être facilement surveillée par fluorescence. La protéine de fusion a été construite en ajoutant un peptide de liaison de 12 acides aminés (ASGGGGGSGGGGS) pour faciliter le repliement des deux protéines. La protéine chimérique résultante était de 140 kDa. La fusion *gfp-lacZ* a été réalisée en aval des promoteurs inductibles susmentionnés. La croissance et l'expression des gènes ont été surveillées par LCA. Les mesures ont été effectuées dans des milieux définis pour les souches portant la GFP seule ou la fusion GFP-LacZ. Les résultats ont montré une diminution de la fluorescence des souches porteuses de GFP-LacZ par rapport aux souches porteuses de GFP [Figure 9.2A]. Cependant, le taux de croissance des souches produisant la GFP ou la GFP-LacZ était similaire à celui de la souche témoin cultivée dans le même milieu. Ensuite, nous

avons effectué un western blot pour vérifier la quantité de GFP produite par les souches produisant GFP et GFP-LacZ sous P_{hs} [Figure 9.2B]. Le western blot indiquait que la quantité de GFP dans la protéine de fusion produite était beaucoup plus faible que celle des souches exprimant la GFP native, ce qui est conforme aux résultats obtenus en LCA. En raison de la très faible teneur en protéines détectée dans les cellules, la protéine de fusion n'était évidemment pas un bon candidat pour étudier l'impact de la surproduction de protéines sur la physiologie cellulaire. En effet, il n'y a pas eu d'effet négatif sur la cellule au niveau de la croissance. Nous avons donc décidé de ne pas surproduire de chimère pour étudier plus avant les conséquences de la surproduction de protéines sur la physiologie cellulaire.

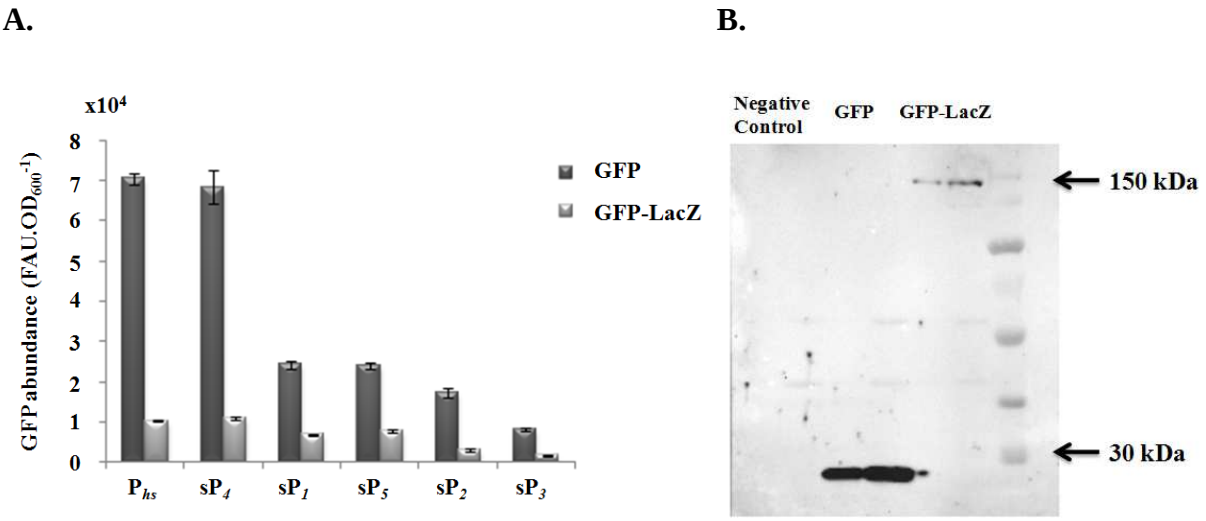

Figure 9.2 La protéine GFP-LacZ est moins abondante dans la cellule que la GFP native

(A) Mesures en LCA de l'abondance des GFP (Fluorescence $AU.OD_{600}^{-1}$) dans les souches exprimant la GFP native ou une fusion GFP-LacZ sous le contrôle des promoteurs inductibles par IPTG 1mM. Il y a une fluorescence nettement faible dans les souches porteuses de GFP-LacZ. (B) Western blot contre GFP des souches exprimant GFP ou GFP-LacZ contrôlées par *Phs*. L'extrait cellulaire a été chargé deux fois en quantités différentes pour chacune des souches. Les intensités des bandes dans les souches exprimant la fusion GFP-LacZ sont très légères par rapport à GFP.

La bêta-galactosidase est une protéine de grande taille (112 kDa) et il a été démontré qu'elle provoque un retard de croissance chez *E. coli* lorsqu'elle est produite massivement (Dong et al. 1995). Nous avons donc construit de nouvelles souches en intégrant une ou deux copies de *lacZ* à 2 *loci* dans le chromosome de *B. subtilis*. Le premier *locus* est *amyE*, un gène non essentiel, proche de l'origine de réplication et connu comme un locus d'intégration habituel chez *B. subtilis*. Le deuxième locus est *nprE*, un gène non essentiel également, situé près du *terminus* de réplication. La croissance cellulaire a été surveillée par LCA dans un milieu riche

et défini. L'effet sur le taux de croissance en milieu riche n'a pas été significatif même dans des conditions d'expression totalement induite (1 mM IPTG), et ce quel que soit le moment d'induction au cours de la croissance (IPTG ajouté au début de l'expérience ou injecté pendant la croissance exponentielle). Dans un milieu pauvre, le défaut de croissance attendu n'était ni plus fort ni significatif (M9 Pyruvate qui donne normalement un taux de croissance d'environ 0,3 h^{-1} pour une souche sauvage) [Figure 9.3]. Nous avons conclu que le niveau d'expression de *lacZ*, lorsqu'il est présent en deux exemplaires et exprimé sous le contrôle de P_{hs} en présence de 1 mM IPTG, n'était pas assez élevé pour provoquer une diminution du taux de croissance chez *B. subtilis.*

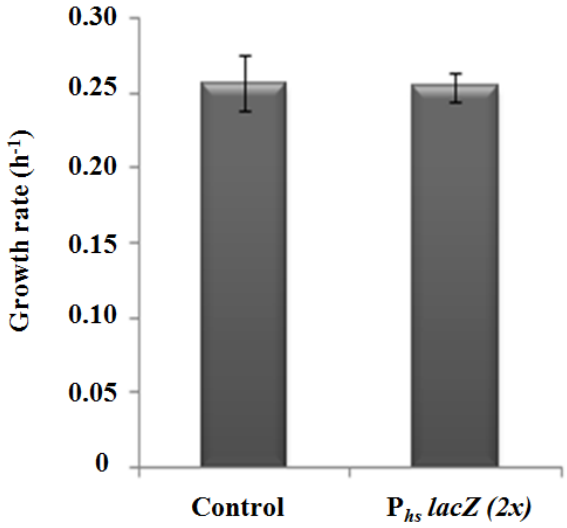

Figure 9.3 Taux de croissance d'une souche produisant LacZ et d'une souche témoin (Control)

La double intégration de lacZ a montré un taux de croissance similaire à celui de la souche témoin dans un milieu mal défini (M9P).

9.2.3 Surproduction de protéines et conséquences sur la physiologie cellulaire de *B. subtilis*

9.2.3.1 Effets sur le taux de croissance

Le promoteur utilisé dans nos stratégies précédentes était l'un des plus forts promoteurs inductibles, mais nous n'avons observé aucun défaut de croissance. Nous avons donc décidé d'utiliser le P_{veg} constitutif, le promoteur constitutif natif le plus fort de *B. subtilis* connu à ce jour (Guiziou et al. 2016). Une souche P_{veg}*lacZ* dans le locus *amyE* a été construite et des mesures en LCA ont été effectuées pour surveiller la croissance cellulaire. Tout d'abord, les mesures ont été effectuées dans le milieu S (un milieu minimal avec du glucose comme source de carbone, ce qui permet à la souche sauvage de *B. subtilis* d'atteindre un taux

de croissance d'environ 0,7 h^{-1}). Le taux de croissance de la souche P_{veg}*lacZ* a diminué d'environ 15 % [Figure 9.4].

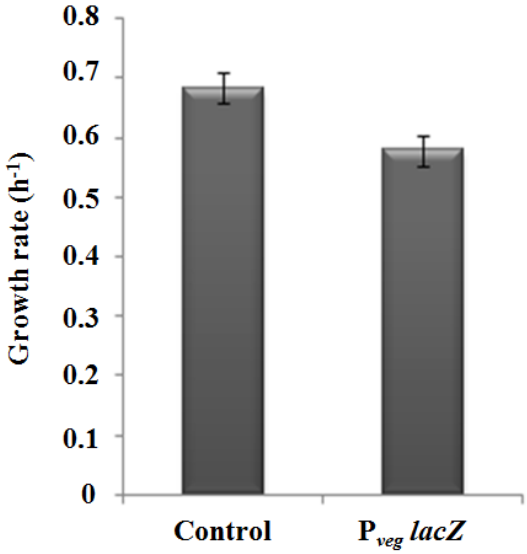

Figure 9.4 Réduction du taux de croissance sur *B. subtilis* lors de l'expression du gène *lacZ* sous le contrôle du promoteur P_{veg}

Dans un milieu synthétique (S), la souche P_{veg}*lacZ* présentait un taux de croissance de 15 % inférieur à celui d'une souche témoin (Control).

Afin de confirmer ce résultat, les gènes *gfp, mkate2* et *lacZ* ont été placés sous le contrôle de P_{veg} et intégrés au chromosome *B. subtilis* en une ou deux copies. La croissance a été surveillée par LCA dans des cellules cultivées dans des milieux pauvres (S), intermédiaires (SX c'est-à-dire S + asparagine, malate et glutamate), et riches (CHG). Ces milieux donnent lieu à des taux de croissance de la souche de type sauvage d'environ 0,7 h^{-1}, 1,0 h^{-1} et 2,0 h^{-1}, respectivement. La diminution du taux de croissance précédemment observée dans la souche P_{veg}*lacZ* a également été observée dans des cellules surproduisant l'une des trois protéines gratuites différentes à partir de copies de gènes simples et doubles. Les résultats sont présentés sur Figure 9.5, qui montre les taux de croissance des souches portant les gènes hétérologues (*gfp, mkate,* et *lacZ*) dans différents milieux. Le défaut de croissance était fort dans les milieux S et SX et augmentait avec les intégrations en double copie des gènes (indiquées par 2x). D'autre part, lorsque les souches étaient cultivées en milieu riche (CHG), la diminution du taux de croissance n'était pas aussi importante que dans les milieux minimaux définis, S et SX. C'était plus significatif avec les souches surproduisant la protéine gratuite à partir de deux copies de gènes.

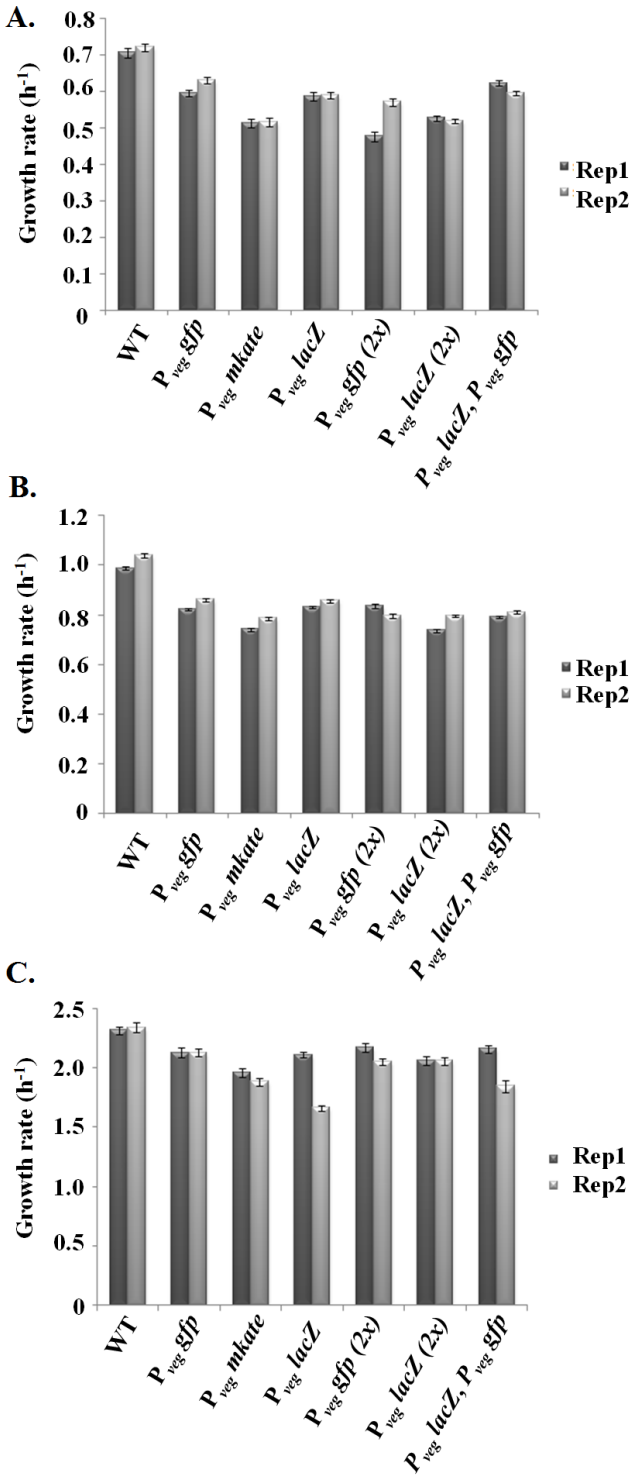

Figure 9.5 Ralentissement de la croissance de toutes les souches porteuses de gènes gratuits dans différents milieux définis.

Les graphiques montrent les taux de croissance des souches portant des copies simples ou doubles de gfp, mkate et lacZ cultivées dans (A) un milieu synthétique (avec glucose) S, (B) un milieu SX (S + malate, glutamate, asparagine), (C) un milieu riche CHG.

9.2.3.2 Effects sur la taille de la cellule

La surproduction de protéines chez *B. subtilis* ralentit non seulement la croissance, mais elle entraîne égalementune augmentation de la taille des cellules. Les souches de *B. subtilis* surproduisant les protéines gratuites ont été cultivées dans le milieu S et récoltées pendant la croissance exponentielle (OD_{600}=0.2). Premièrement, l'analyse microscopique a révélé une légère augmentation de la taille des cellules des souches surproduisant GFP et LacZ par rapport aux cellules de type sauvage. Pour une caractérisationplus précise, nous avons réalisé une analyse en cytométrie en flux. La cytométrie en flux permet d'analyser une population cellulaire au niveau unicellulaire et d'obtenir des analyses statistiques sur la taille et la fluorescence des populations cellulaires. À partir d'un diagramme de dispersion, 50 000 cellules ont été sélectionnées dans une ellipse identique utilisée pour sélectionner les données. Des histogrammes montrant la distribution de tailles de toutes les souchessont présentés ci-dessous [Figure 9.6A]. Ces résultats indiquent que la taille des cellules des souches surproductrices LacZ et GFP est supérieure à celle du type sauvage. Les valeurs Fsc des mutants ont été normalisées par rapport à celles du type sauvage [Figure 9.6B].

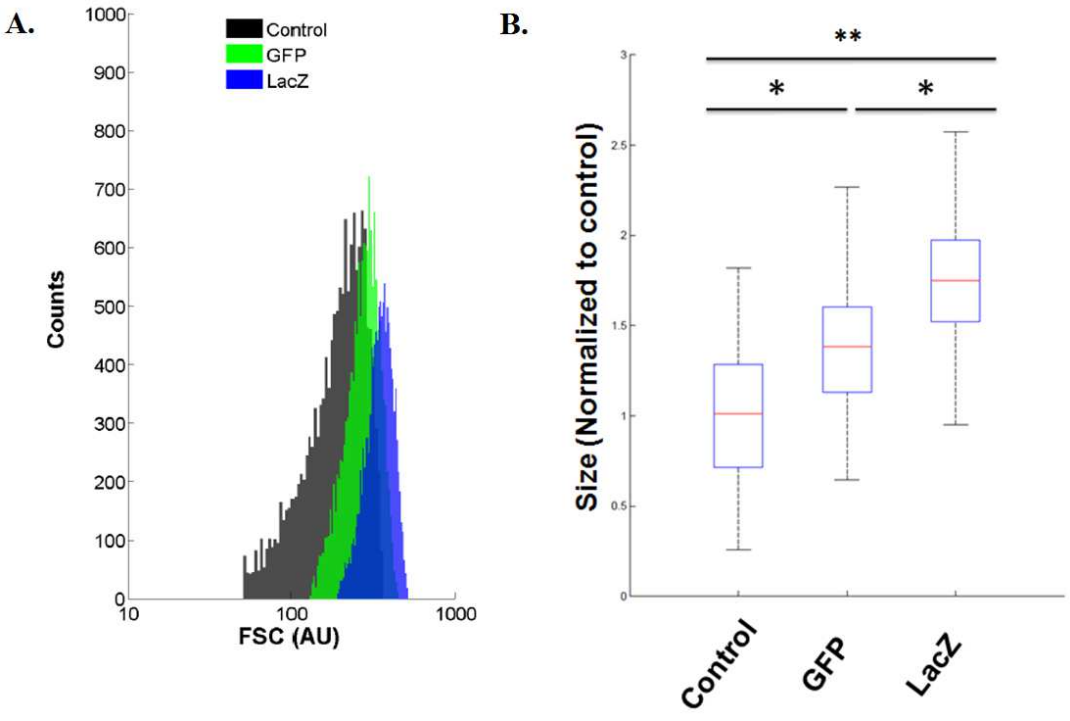

Figure 9.6 Données de cytométrie en flux

(A) Distribution de tailles des différentes souches analysées (souches témoins, GFP et LacZ). (B) Valeurs Fsc des souches mutantes présentées enboxplot normalisées par rapport à la valeur Fsc de la souche témoin.

9.2.4 Le taux de croissance est modifié lorsque la production de GFP atteint un certain seuil

Après avoir montré une charge sur le taux de croissance, nous nous sommes intéressés à trouver la quantité de protéines qui a contribué à cette charge. Pour ce faire, nous avons d'abord réalisé une analyse par gel SDS-PAGE de manière àvérifier l'abondance des protéines hétérologues par rapport aux protéines cellulaires solubles totales. Les pourcentages calculés pour les protéines hétérologues sont les suivants en respectant l'ordre sur le gel SDS-PAGE [Figure 9.7A] : (2) GFP_{Bsu} 9,07 %, (3) GFP^{ALGG}_{Bsu} 6,01 %, (4) $LacZ_{Bsu}$ 6,92 %, (5) GFP_{Bsu} 6,94 % et $LacZ_{Bsu}$ 5,49 %, (7) GFP_{Eco} 7,56 %. Pour une meilleure validation de la quantité de protéines réellement produites, nous avons effectué un western blot contre les GFP, et un résultat représentatif est illustré à la Figure 9.7C. La GFP représentait ~9,77% du total des protéines extraites, ce qui est proche des valeurs calculées sur la base de l'intensité du signal en SDS Page. De plus, en Figure 9.7D, les cellules cultivées en géloses LB portant le P_{veg}*mkate* (rose) et le P_{veg}*gfp* (vert) montrent une fluorescence des protéines visible à l'œil nu.

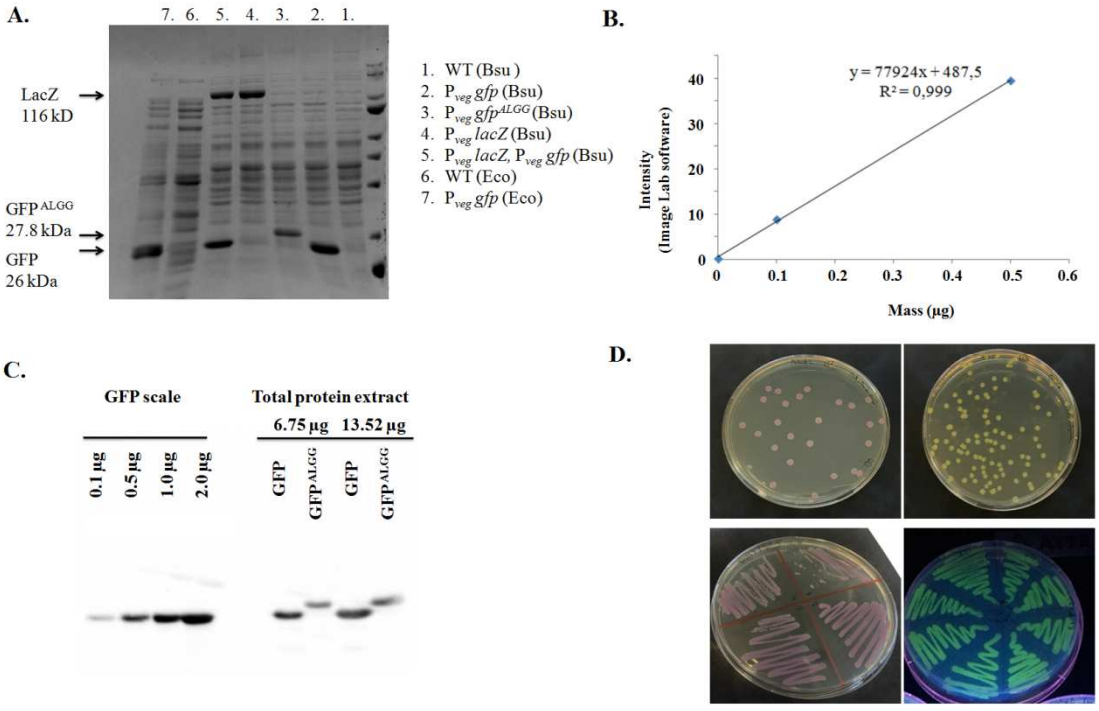

Figure 9.7 Quantification hétérologue des protéines par rapport à l'extrait total de protéines solubles

(A) Gel SDS-PAGE coloré au bleu de coomassie correspondant auxextraits totaux de protéines solubles pour le type sauvage et chacune des souches produisant GFP, LacZ et LacZ -GFP. (B) Graphique représentant l'intensité des bandes mesurées par le logiciel ImageLab par rapport à la masse protéique, qui a été utilisée pour quantifier la GFP dans le western blot (C) Western blot contre la GFP montrant une échelle de GFP de quantités connues, et la GFP produite dans un extrait protéique soluble total de 6,75 µg et 13,52µg. (D) Fluorescence des protéines des cellules produisant mKate2 et GFP indiquant l'intensité de la production.

9.2.5 La surproduction de *B. subtilis* par les GFP entraîne une réduction de la production d'autres protéines

La question suivante était de savoir pourquoi, comparativement aux connaissances actuelles sur *E. coli*, le taux de croissance défectueux n'est observé qu'après la surproduction d'une telle quantité d'une protéine gratuite? Afin de résoudre ce problème, nous avons effectué une analyse protéomique des souches surproductrices de GFP. Des analyses protéomiques ont été effectuées sur trois souches exprimant soit une copie de la cassette d'expression "P_{veg}*gfp*" (souche MZ139), soit deux copies de cette cassette d'expression (souche MZ140), et une souche sauvage (souche témoin MZ72). Les cellules ont été cultivées en milieu S jusqu'à ce qu'elles atteignent la phase exponentielle. En ce qui concerne la teneur totale en protéines solubles, il est intéressant de noter que le principal effet de la surproduction d'une protéine gratuite dans *B. subtilis* est une diminution de la quantité de nombreuses protéines dans les souches surproductrices par rapport à leur quantité respective dans la souche témoin. Dans la heat map ci-dessous [Figure 9.8], les 300 protéines les plus affectées sur les 800 protéines détectées sont présentées. Les valeurs tracées ont été générées en calculant le rapport $\log_{10}$ de chaque échantillon divisé par la moyenne des échantillons. Ensuite, les protéines ont été triées selon leur ratio de GFP2x.

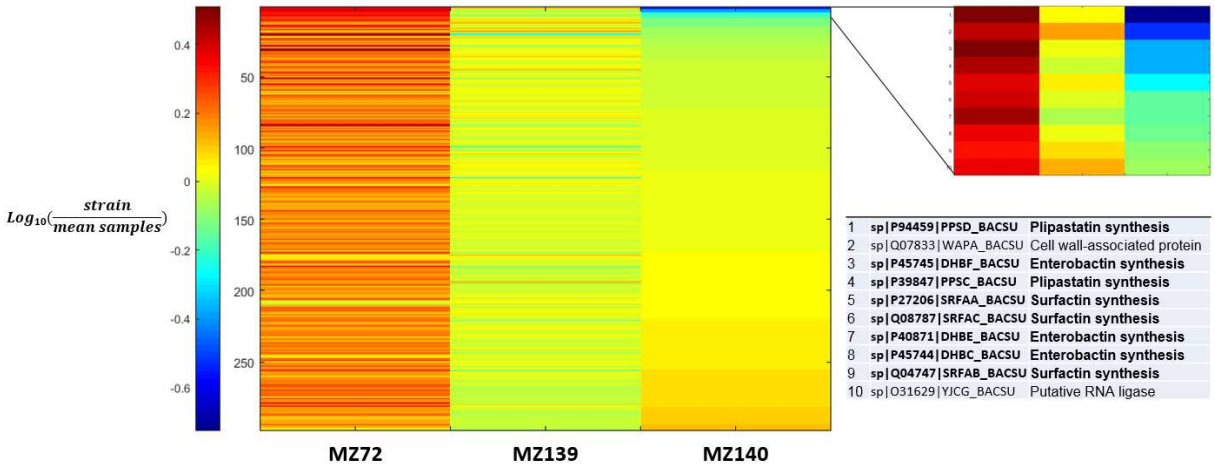

Figure 9.8 Représentation sur heat map des 298 protéines les plus affectées dans le type sauvage et les souches productrices de GFP. A droite, zoom sur les 10 protéines les plus affectées et leurs fonctions métaboliques respectives chez *B. subtilis*

Les numérations spectrales des protéines des souches analysées ont été comparées à la moyenne des numérations spectrales dans toutes les souches et les $\log_{10}$-ratios ont été utilisés pour la visualisation. MZ72 Control strain, MZ139 P_{veg}*gfp* une copie, et MZ140 P_{veg}*gfp* deux copies.

D'une manière générale, en regardant la heat map, nous observons que de nombreuses protéines sont moins produites dans les souches productrices GFP par rapport à la souche de type sauvage. De plus, on peut observer la différence dans la distribution des protéines entre les souches produisant la GFP. La partie supérieure de la heatmap révèle les protéines les plus affectées lors de la surexpression d'une GFP. Nous avons fait un zoom sur les 10 protéines les plus 'réprimées' dans la souche possédant deux copies de la cassette d'expression de la GFP (MZ140) [Figure 9.8- côté droit]. Parmi cette liste de dix protéines, huit appartiennent à des enzymes impliquées dans la synthèse de peptides non ribosomiques tels que la synthèse de la plipastatine, la synthèse de surfactine et la synthèse d'entérobactine (aussi appelée bacillibactine).

9.2.6 La dégradation de la protéine gratuite surproduite rétablit le taux de croissance

Chez *E. coli* et *B. subtilis*, lorsqu'un ribosome atteint l'extrémité 3' d'un ARNm sans atteindre un codon stop, l'ARNt codé par le gène *ssrA* ajoute un tag appelé tag ssrA. Dans *B. subtilis*, la protéase ClpXP reconnaît la protéine marquée ssrA et procède à sa dégradation. Nous avons décidé d'utiliser cette propriété et de générer un système de dégradation à base de ClpXP pour dégrader spécifiquement la protéine gratuite surproduite et libérer des acides aminés dans la cellule sans alléger la charge en ribosomes. Sur la base des résultats déjà publiés (Griffith & Grossman 2008 ; Guiziou et al. 2016), nous avons construit des souches produisant des GFP étiquetées à leur extrémité C-terminale avec cinq tags ssrA candidats (ALGG, DDAS, ADAN, ADCS, AASV) associés àla séquence AANDENYSENY (ou AGKTNSFNQNQNV) reconnue par la protéase ClpXP. Afin de dégrader spécifiquement la protéine gratuite surproduite et de libérer des acides aminés dans la cellule sans soulager la charge ribosomique, nous avons testé le système de dégradation basé sur ClpXP sur les protéines GFP et mKate2. Le taux de croissance a été surveillé dans différents milieux définis (S, SX et CHG) [Figure 9.9A, B, C respectivement]. En ciblant spécifiquement GFP^{ALGG}, nous avons observé une restauration du taux de croissance principalement dans les milieux S et SX. La restauration du taux de croissance a été d'environ 50%. Lors de la dégradation de la protéine $mKate2^{ALGG}$, une restauration du taux de croissance a été observée de manière significative dans les milieux S, SX et CHG. Nous avons conclu que la dégradation des protéines gratuites nouvellement synthétisées permettait le recyclage des acides aminés réutilisés par la cellule et la libération de l'espace cytosolique, ce qui permettrait une augmentation du taux de croissance. Par conséquent, la restauration du taux de croissance qui

en a résulté a indiqué que la charge causée sur la cellule par la protéine gratuite était principalement due à la limitation des acides aminés et/ou de l'espace cytosolique.

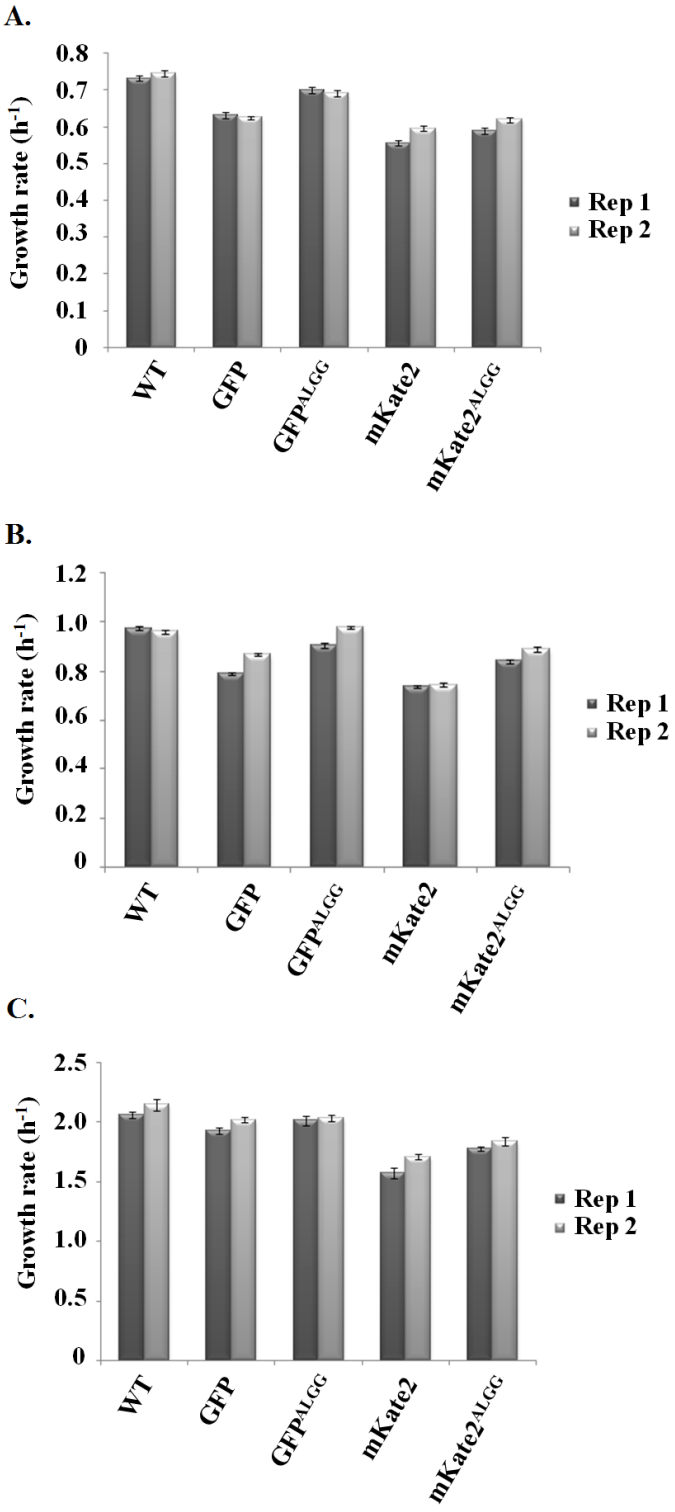

Figure 9.9 Restauration du taux de croissance lors de la dégradation des protéines

Les graphiques montrent des duplicatas de mesures en LCA où le taux de croissance a été suivi dans différents milieux pour différentes protéines gratuites afin de savoir si la dégradation des protéines va modifier le taux de croissance. (A) milieu S, (B) milieu SX, (C) milieu CHG.

9.2.7 Développement d'un outil de biologie synthétique pour affiner l'expression des gènes

Dans la deuxième partie de la thèse, nous avons cherché à développer un outil synthétique pour affiner l'expression des gènes en améliorant le système de dégradation ciblé. Nous avons d'abord caractérisé le travail effectué par Griffith et ses collaborateurs (2008) et nous avons constaté que leur système de dégradation ciblée est en effet pleinement efficace à des niveaux modérés de production de protéines marquées. Par conséquent, nous avons cherché à améliorer le système de dégradation afin d'augmenter le niveau de dégradation, car être capable d'ajuster finement les niveaux de protéines pourrait servir plusieurs applications biotechnologiques. Afin d'augmenter le niveau de dégradation d'une protéine d'intérêt, nous avons décidé de construire une souche dans laquelle les productions de la protéase ClpXP et de la SspB (un adaptateur de ClpX chex *E. coli*) sont augmentés. Nous avons conçu diverses structures opéroniques composées de *sspB*, *clpX* et *clpP* en aval de P_{hs} et de RBS synthétiques solides. La conception des opérons s'inspire des structures opéroniques naturelles de *B. subtilis*. Malheureusement, la dégradation de la protéine ciblée (GFP^{ALGG}) n'a pas été améliorée avec la surproduction de ClpXP et de SspB.

De plus, comme nous voulions mettre en place un outil de recyclage des acides aminés synthétiques basé sur l'utilisation de ClpXP, nous nous sommes demandé si ClpXP pouvait avoir des objectifs hors cible, surtout en cas de surproduction (en collaboration avec sspB). Par conséquent, pour étudier les conséquences d'une surproduction de ClpXP sur la physiologie cellulaire, nous avons construit un opéron P_{veg}*clpXclpP* en utilisant les gènes *clpX clpP* de *B. subtilis* et le plus fort promoteur connu à ce jour sur *B. subtilis*, P_{veg}. En examinant généralement le nombre de protéines détectées, on constate qu'il y avait plus de protéines dans la souche de type sauvage que dans la souche P_{veg}*clpXclpP*. Nous avons observé 316 protéines à régulation négative et 181 protéines à régulation positive dans la souche P_{veg}*clpXclpP* par rapport à la souche témoin. En Figure 9.10 figurent les catégories de protéines pour lesquelles l'abondance a changé entre les deux souches. En plus de ClpX et de ClpP, la plupart des protéines qui ont été régulées à la hausse sont liées aux protéines biosynthétiques d'acides aminés, à la synthèse de nucléotides ou aux synthétases d'ARNt. Les protéines qui étaient régulées à la baisse sont principalement impliquées dans la motilité et la chimiotaxie (synthèse duflagelle), ou contrôlées par le facteur sigma σ_D, (i.e. le facteur sigma qui active la synthèse du flagelle). Les autres catégories de protéines varient légèrement. Certaines protéines impliquées dans la réponse générale au stress et au stress oxydant, ou d'autres

protéases varient légèrement, mais aucune autre tendance claire n'est apparue. Par conséquent, les résultats indiquent qu'une surproduction de ClpXP a entraîné une réorganisation majeure du protéome.

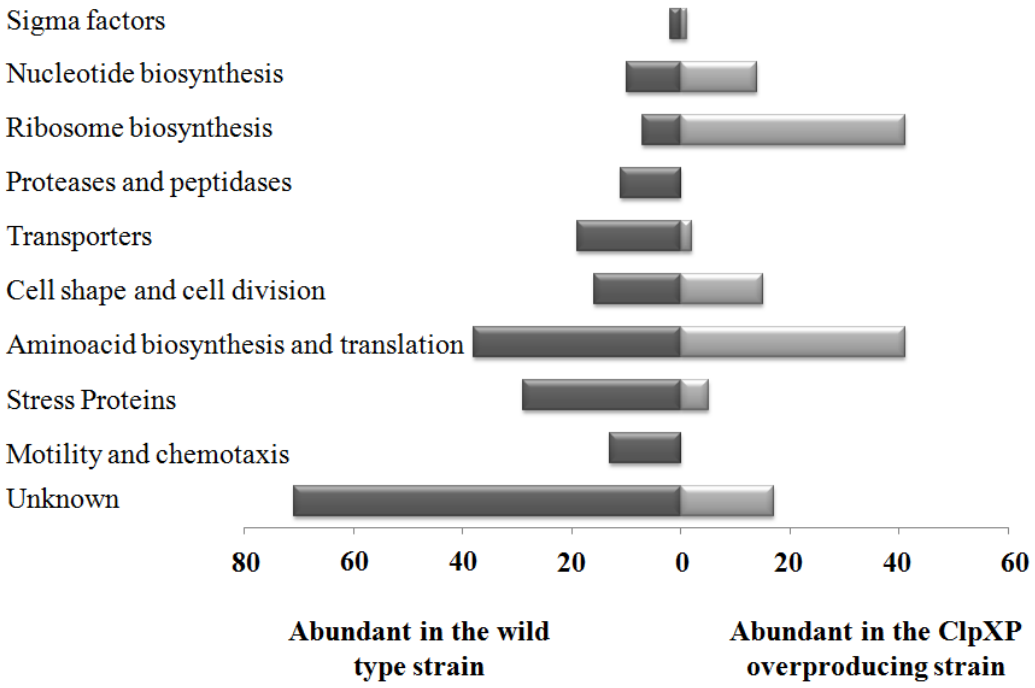

Figure 9.10 Nombre de protéines (regroupées en catégories) exprimées différemment entre le type sauvage et le mutant ClpXP

9.3 Conclusion

Mon projet de doctorat a apporté une nouvelle perspective pour comprendre les goulots d'étranglement de l'obtention d'usines de cellules microbiennes pour la production de protéines. Nous avons obtenu de nouveaux résultats et de nouvelles conclusions sur le comportement de *B. subtilis* lors de la surproduction d'une protéine hétérologue. *B. subtilis* a fait l'objet d'une étude de cas pour étudier le stress de sécrétion auquel elle fait face lorsqu'elle est soumise à une surproduction de protéines. Cependant, nous ne savions pas quelles étaient les conséquences au niveau de la production avant d'être sécrétées.

Nos résultats ont montré que *B. subtilis* ralentit son taux de croissance lorsqu'il est exposé à une surproduction gratuite de protéines. Nous avons étudié la réduction du taux de croissance dans différents milieux chimiquement définis et en utilisant différentes protéines rapportrices

(GFP, LacZ et mKate2). De plus, comme *E. coli* et *S. cerevisiae*, *B. subtilis* augmente sa taille en réponse à une production élevée de protéines.

Nous avons donné une analyse détaillée des données protéomiques pour la quantification des protéines relatives où nous avons montré une diminution générale de l'abondance des protéines endogènes dans les souches surproductrices de GFP. Une diminution générale du protéome est prévue en présence d'une protéine hétérologue inutile (Scott et al. 2014 ; Scott et al. 2010). En entrant plus en détail sur le type de protéines qui change le plus, nous avons découvert que les protéines non essentielles de grande taille liées aux synthétases peptidiques non ribosomales sont devenues moins abondantes avec l'augmentation de la production de GFP. La taille de ces protéines en fait des protéines coûteuses en termes de ressources pour la cellule, plutôt que d'être coûteuses en raison de leur nombre. Ces résultats nous ont permis de déduire une relation entre l'utilisation des acides aminés pour produire les protéines essentielles et non essentielles, y compris la protéine gratuite, l'occupation cytosolique des protéines et la réduction du taux de croissance. Cette hypothèse a été prouvée en appliquant un outil synthétique de dégradation ciblée des protéines contre les protéines gratuites (GFP, mKate2, et LacZ). Le résultat a montré une restauration du taux de croissance lors de la dégradation de GFP et mKate2, ce qui indique que les acides aminés libérés ont été réutilisés par la cellule. De plus, selon la taille de la protéine et son compactage, le type de ressource limitante peut être soit la composition en acides aminés, soit l'occupation cytosolique, soit les deux.

Enfin, nos résultats ouvrent la voie à d'autres études sur la surproduction de protéines sécrétées. La combinaison des informations sur le stress de sécrétion et la surproduction de la machine de sécrétion pour favoriser la production et la sécrétion de protéines hétéologues ouvrira de nouvelles perspectives pour améliorer les usines de cellules microbiennes.

References

Adrio, J.L. & Demain, A.L., 2014. Microbial enzymes: tools for biotechnological processes. *Biomolecules*, 4(1), pp.117–139.

Aïchaoui L, Jules M, Le Chat L, Aymerich S, Fromion V, Goelzer A, 2012. BasyLiCA: A tool for automatic processing of a Bacterial Live Cell Array. *Bioinformatics*, 28(20), pp.2705–2706.

Anderson, D.E., Gueiros-filho, F.J. & Erickson, H.P., 2004. Assembly Dynamics of FtsZ Rings in *Bacillus subtilis* and *Escherichia coli* and Effects of FtsZ-Regulating Proteins. *Journal of Bacteriology of bacteriology*, 186(17), pp.5775–5781.

Asaka, O. & Shoda, M., 1996. Biocontrol of Rizhoctonia solani Damping-off of Tomato with *Bacillus sutlis* RB14. *Applied and environmental microbiology*, 62(11), pp.4081–4085.

Atkinson, G.C., Tenson, T. & Hauryliuk, V., 2011. The RelA / SpoT Homolog (RSH) Superfamily : Distribution and Functional Evolution of ppGpp Synthetases and Hydrolases across the Tree of Life. *PLoS ONE*, 6(8), pp.1-21.

Baker, T.A. & Sauer, R.T., 2006. ATP-dependent proteases of bacteria: recognition logic and operating principles. *Trends Biochemical Sciences*, 31(12), pp.647–653.

Barbieri G,Voigt B, Albrecht D, Hecker M, Albertini A.M, Sonenshein L.A, Ferrari E,&Belitsky R.B., 2015. CodY Regulates Expression of the *Bacillus subtilis* Extracellular Proteases Vpr and Mpr. *Journal of Bacteriology*, 197(8), pp.1423–1432.

Basan M, Zhu M, Dai X, Warren M, Sévin D, Wang YP, &Hwa T., 2015. Inflating bacterial cells by increased protein synthesis. *Molecular Systems Biology*, 11, pp.1–7.

Battesti, A. & Gottesman S., 2017. Roles of adaptor proteins in regulation of bacterial proteolysis. *Current Opinion in Microbiology*, 16(2), pp.140–147

Beg Q. K., Vazquez A., Ernst J., Menezes M. A de, Bar-Joseph Z., Barabási A.-L., & Oltvai Z. N., 2007. Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proceedings of the National Academy of Sciences*, 104(31), pp.12663–12668..

Bertaux F, Kügelgen J von, Marguerat S, &Shahrezaei V., 2016. A unified coarse-grained theory of bacterial physiology explains the relationship between cell size , growth rate and proteome composition under various growth limitations. *bioRxiv*, pp.1–13.

Bertrand, Kevin P, Lenski,& Richard E. et al., 1989. Effects of Carriage and Expression of the

TnlO Tetracycline-Resistance Operon on the Fitness of Escherichia coli K12. ,*Molecular biology and evolution* 6(3) pp. 213–225.

Bi, E. & Lutkenhaus, J., 1991. FtsZ ring structure associated with division in *Escherichia coli*. *Nature*, 354, pp.161–164.

Borkowski O., Ceroni F., Stan GB.,&Ellis T., 2016. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Current Opinion in Microbiology*, 33, pp.123–130..

Borkowski O, Goelzer A, Schaffer M, Calabre M, Mäder U, Aymerich S, Jules M,&Fromion V, 2016. Translation elicits a growth rate-dependent , genome-wide , differential protein production in *Bacillus subtilis.* , *Molecular systems Biology*, 12(5), pp.1–14.

Botella E, Fogg M, Jules M, Piersma S, Doherty G, Hansen A, Denham EL, Le Chat L, Veiga P, Bailey K, Lewis PJ, van Dijl JM, Aymerich S, Wilkinson AJ, &Devine KM., 2010.

pBaSysBioII: An integrative plasmid generating *gfp* transcriptional fusions for high-throughput analysis of gene expression in *Bacillus subtilis*. *Microbiology*, 156(6), pp.1600–1608.

Bremer, H. & Dennis, P.P., 1996. Modulation of Chemical Composition and Other Parameters. *in Escherichia coli and Salmonella. Molecular Biology Neidhardt ed ASM Press*, 2(93), pp.1553–1569.

Brien, E.J.O., Utrilla, J. & Palsson, B.O., 2016. Quantification and Classification of *E . coli* Proteome Utilization and Unused Protein Costs across Environments. 12(6), pp.1–22.

Bron, S., Metier, W. & Haima, P., 1991. Plasmid instability and molecular cloning in *Bacillus subtilis*. ,*Research in Microbiology*, (142) pp.875–883.

Bruckner, R. & Titgemeyer, F., 2002. Carbon catabolite repression in bacteria : choice of the carbon source and autoregulatory limitation of sugar utilization. ,*FEMS Microbiology Letters*(209) pp141-148.

Bryant JA, Sellars LE, Busby SJ,&Lee DJ, 2014. Chromosome position effects on gene expression in *Escherichia coli* K12. , Nucleic Acids Research, 42(18), pp.11383–11392.

Buescher, J. et al., 2012. Global Network Reorganization During Dynamic Adaptations of Bacillus subtilis Metabolism. *Science*, (335) pp.1099-1102.

Cameron, D.E. & Collins, J.J., 2014. Tunable protein degradation in bacteria. *Nature Biotechnology*, 32(12), pp.1276–1281.

Carrio, M. & Villaverde, A., 2002. Construction and deconstruction of bacterial inclusion bodies. *Journal of Biotechnology*, 96, pp.3–12.

Caulier S, Nannan C, Gillis A, Licciardi F, Bragard C,&Mahillon J. et al., 2019. Overview of the Antimicrobial Compounds Produced by Members of the *Bacillus subtilis* Group. *Frontiers in Microbiology*, 10, pp.1–19.

Chai Y., Kolter R.,& Losick R., 2010. Reversal of an epigenetic switch cell chaining in *Bacillus subtilis* by protein instability. *Molecular Microbiology* , 78(1), pp.218–229.

Chan, C.M., Hahn, E. & Zuber, P., 2014. Adaptor bypass mutations of Bacillus subtilisspx suggest a mechanism for YjbH-enhanced proteolysis of the regulator Spx by ClpXP. *Molecular Microbiology*, 93(3), pp.426–438.

Chubukov V, Uhr M, Le Chat L, Kleijn RJ, Jules M, Link H, Aymerich S, Stelling J,&Sauer U. et al., 2013. Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis*. *Molecular Systems Biology*, 9(709), pp.1–13.

Churchward, G., Bremer, H. & Young, R., 1982. Macromolecular composition of bacteria. *Journal of Theoretical Biology*, 94(3), pp.651–670.

Cookson NA, Mather WH, Danino T, Mondragón-Palomino O, Williams RJ, Tsimring LS, &Hasty J., 2011. Queueing up for enzymatic processing: Correlated signaling through coupled degradation. *Molecular Systems Biology*, 7(1) pp.1-9.

Cooper S,&Helmstetter CE., 1968. Chromosome Replication and the Division of *Escherichia coli* B / r.*Molecular Biology* , 31, pp.519-540.

Cormack, B.P., Valdivia, R.H. & Falkow, S., 1996. FACS-optimized mutants of the green fluorescent protein (GFP). *Elsevier Science*, 173, pp.33–38.

Couturier, E. & Rocha, E.P.C., 2006. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Molecular Microbiology*, 59(5), pp.1506–1518.

Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, &Mann M, 2014. Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Molecular & Cellular Proteomics*, 13(9), pp.2513–2526.

Cox, J. & Mann, M., 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26(12), pp.1367–1372.

References

Dalebroux, Z.D. & Swanson, M.S., 2012. ppGpp : magic beyond RNA polymerase. *Nature Publishing Group*, 10(3), pp.203–212.

Daraba, A. & Alma, K., 2018. Hsp70-associated chaperones have a critical role in buffering protein production costs.*eLife* , pp.1–23.

Das, K. & Mukherjee, A.K., 2006. Assessment of mosquito larvicidal potency of cyclic lipopeptides produced by *Bacillus subtilis* strains. *Acta Tropica*, 97(2), pp.168–173.

Demain, A.L. & Vaishnav, P., 2011. Production of Recombinant Proteins by Microbes and Higher Organisms. *Comprehensive Biotechnology, Second Edition*, 3(3), pp.333–345.

Dennis, P.P. & Bremer, H., 2008. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 3(1).

Derré I, Rapoport G, Devine K, Rose M, &Msadek T., 1999. ClpE , a novel type of HSP100 ATPase , is part of the CtsR heat shock regulon of *Bacillus subtilis*. ,*Molecular Microbiology*, 32, pp.581–593.

Derre, I., Rapoport, G. & Msadek, T., 2000. The CtsR regulator of stress response is active as a dimer and specifically degraded in *vivo* at 37 8 C. ,*Molecular Microbiology*, 38, pp. 335-347.

Deuerling E, Mogk A, Richter C, Purucker M,&Schumann W., 1997. The ftsH gene of *Bacillus subtilis* is involved in major cellular processes such as sporulation , stress adaptation and secretion. , *Journal of Bacteriology*23, pp.921–933.

Deutscher J, Reizer J, Fischer C, Galinier A, Saier MH Jr, &Steinmetz M, 1994. Loss of protein kinase-catalyzed phosphorylation of hpr , a phosphocarrier protein of the phosphotransferase system , by mutation of the ptsh gene confers catabolite repression resistance to several catabolic genes of *Bacillus subtilis*. ,*Journal of Bacteriology*, 176(11), pp.3336–3344.

Dijl, J.Mvan. & Hecker, M., 2013. *Bacillus subtilis* : From soil bacterium to super- secreting cell factory. *Microbial Cell Factories*, 12 (3), pp 1-6.

Donachie, 1968.Relationship between cell size and time of initiation of DNA replication 1968,*Nature Publishing Group*, 219, pp. 1077-1079.

Dong, H., Nilsson, L. & Kurland, C.G., 1995. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *Journal of Bacteriology*, 177(6), pp.1497–1504.

Dougan DA, Mogk A, Zeth K, Turgay K, &Bukau B., 2002. AAA + proteins and substrate recognition, it all depends on their partner in crime.*FEBS Journal*, 529, pp.6–10.

Dougan, D.A., Truscott, K.N. & Zeth, K., 2010. The bacterial N-end rule pathway: expect the unexpected., *Molecular Microbiology*, 76, pp.545–558.

Earl, A.M., Losick, R. & Kolter, R., 2010. Ecology and genomics of *Bacillus subtilis*, *Trends in Microbiology*. , 16(6), pp.1–11.

Ehrlich, D., 1978. DNA cloning in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences*, 75(3), pp.1433–1436.

Ehrlich, S.D. et al., 1991. Plasmid replication and structural stability in *Bacillus subtilis*. *Microbiol*, pp.869–873.

Elowitz, M.B. & Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. *Nature*, pp.335–338.

Elsholz W., Hempel K., Michalik S., Gronau K., Becher D., Hecker M., &Gerth U., 2011. Activity Control of the ClpC Adaptor McsB in *Bacillus subtilis*. , *Journal Of Bacteriology*, 193(15), pp.3887–3893.

Elsholz AK, Michalik S, Zühlke D, Hecker M, &Gerth U., 2010. CtsR, the Gram-positive master regulator of protein quality control, feels the heat. ,*The EMBO Journal*, 29(21), pp.3621–3629.

Elsholz AK, Hempel K, Pöther DC, Becher D, Hecker M, &Gerth U, 2011. CtsR inactivation during thiol-specific stress in low GC, Gram+ bacteria. *Molecular Microbiology*, 79(3), pp.772–785.

Elsholz AKW, Birk MS, Charpentier E, &Turgay K, 2017. Functional Diversity of AAA+ Protease Complexes in *Bacillus subtilis*. *Frontiers in Molecular Biosciences*, 4, pp.1–15.

Emmert J., E. & Handelsman, B., 1999. Biocontrol of plant disease: a gram-positvie perspective. *FEMS Microbiology Letters*, 171, pp.1–9.

Engman, J. & Wachenfeldt, C. Von, 2015. Regulated protein aggregation : a mechanism to control the activity of the ClpXP adaptor protein YjbH. *Molecular Microbiology*, 95, pp.51–63.

References

Eymann C, Dreisbach A, Albrecht D, Bernhardt J, Becher D, Gentner S, Tam le T, Büttner K, Buurman G, Scharf C, Venz S, Völker U, &Hecker M., 2004. A comprehensive proteome map of growing *Bacillus subtilis* cells. *Proteomics*, 4(10), pp.2849–2876.

Fan H, Zhang Z, Li Y, Zhang X, Duan Y, &Wang Q., 2017. Biocontrol of Bacterial Fruit Blotch by *Bacillus subtilis* 9407 via Surfactin-Mediated Antibacterial Activity and Colonization. *Frontiers in Microbiology*, 8, pp.1–15.

Farrell, C.M., Grossman, A.D. & Sauer, R.T., 2005. Cytoplasmic degradation of ssrA-tagged proteins. *Molecular Microbiology*, 57(6), pp.1750–1761.

Flynn JM, Levchenko I, Seidel M, Wickner SH, Sauer RT, &Baker TA., 2001. Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis.*PNAS*, (19) pp. 10584-10589

Frees D, Savijoki K, Varmanen P, &Ingmer H, 2007. Clp ATPases and ClpP proteolytic complexes regulate vital biological processes in low GC, Gram-positive bacteria. *Molecular Microbiology*, 63(5), pp.1285–1295.

Fuhrmann J, Subramanian V, Kojetin DJ, &Thompson PR, 2017. Activity-based profiling reveals a regulatory link between oxidative stress and protein arginine phosphorylation. *Cell Chem Biol*, 23(8), pp.967–977.

Fuhrmann J, Schmidt A, Spiess S, Lehner A, Turgay K, Mechtler K, Charpentier E, &Clausen T. McsB is a protein arginine kinase that phosphorylates and inhibits the heat-shock regulator CtsR. *Science*, 324 (6), pp.1323–1328.

Fujita, Y., 2009. Carbon Catabolite Control of the Metabolic Network in *Bacillus subtilis*. *Bioscience, Biotechnology, Biochemistry*.73 (2), 245–259

Galinier A, Deutscher J, &Martin-Verstraete I., 1999. Phosphorylation of Either Crh or HPr Mediates Binding of CcpA to the *Bacillus subtilis* xyn *cre* and Catabolite Repression of the xyn Operon. *Journal of Molecular Biology*, pp.307–314.

Gao L, Han J, Liu H, Qu X, Lu Z, &Bie X., 2017. Plipastatin and surfactin coproduction by Bacillus subtilis pB2-L and their effects on microorganisms. *Antonie van Leeuwenhoek*.110(8), pp. 1007-1018

Geiger, T. & Wolz, C., 2014. Intersection of the stringent response and the CodY regulon in low GC Gram-positive bacteria. *International Journal of Medical Microbiology*, 304(2), pp.150–155.

References

Georgiou, G. & Valax, P., 1996. Expression of correctly folded proteins in *Escherichia coli*. *Current Opinion in Biotechnology*, 7 (2), pp.190–197.

Gerosa L, Kochanowski K, Heinemann M, &Sauer U., 2013. Dissecting specific and global transcriptional regulation of bacterial gene expression. *Molecular Systems Biology*, 9(658), pp.1–11.

Gerth U, Kock H, Kusters I, Michalik S,SwitzerRL, & Hecker M, 2008. Clp-Dependent Proteolysis Down-Regulates Central Metabolic Pathways in Glucose-Starved *Bacillus subtilis*. , *Journal Of Bacteriology*,190(1), pp.321–331.

GibsonDG, YoungL, ChuangRY, VenterJC, Hutchison IIICA & SmithHO, 2009. Enzymatic assembly of DNA molecules up to several hundrend kilobases. *Nature Methods*, 44(2), pp.149–159.

Goelzer A, Muntel J, Chubukov V, Jules M, Prestel E, Nölker R, Mariadassou M, Aymerich S, Hecker M, Noirot P, Becher D, &Fromion V, 2015. Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic Engineering*, 32, pp.232–243.

Goelzer A, Bekkal Brikci F, Martin-Verstraete I, Noirot P, Bessières P, Aymerich S, &Fromion V., 2008. Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis.BMC Systems Biology* , 2(20), pp.1–18.

Goelzer, A. & Fromion, V., 2011. Bacterial growth rate reflects a bottleneck in resource allocation. *Biochimica et Biophysica Acta - General Subjects*, 1810(10), pp.978–988.

Goelzer, A., Fromion, V.,& Scorletti G., 2009. Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6) pp.1210-1218.

Gottesman, S., 1996. Proteases and their targets in *Escherichia coli*. *Annual Review of Genetics*, 30(1), pp.465–506.

Gottesman, S., 2003. Proteolysis in Bacterial Regulatory Circuits. *Annual Review of Cell and Developmental Biology*, 19(1), pp.565–587.

Gourse, R.L. & Krasny, L., 2004. An alternative strategy for bacterial ribosome synthesis : Bacillus subtilis rRNA transcription. *The EMBO Journal*, 23(22), pp.4473–4483.

Green, R. & Rogers, E.J., 2013. Chemical Transformation of *E. coli*. *Methods Enzymol*., pp.3–6.

References

Griffith, K.L. & Grossman, A.D., 2008. Inducible protein degradation in *Bacillus subtilis* using heterologous peptide tags and adaptor proteins to target substrates to the protease ClpXP. , 70(4), pp.1012–1025.

Guiziou S., Sauveplane V., Chang HJ., Clerté C., Declerck N., Jules M., &BonnetJ., 2016. A part toolbox to tune genetic expression in *Bacillus subtilis*. *Nucleic Acids Research*, 44(15), pp. 7495-7508

Gyorgy A., . Jiménez IJ., Yazbek J., HuangHH., Chung H., Weiss R.,&Vecchio DD., 2015. Isocost Lines Describe the Cellular Economy of Genetic Circuits. *Biophysical Journal*, 109(3), pp.639–646.

Han L.,Suo F.,Jiang C., Gu J., Li N., Zhang N.,Cui W., &Zhou Z., 2017. Fabrication and characterization of a robust and strong bacterial promoter from a semi-rationally engineered promoter library in *Bacillus subtilis*. *Process Biochemistry*, 61(10), pp.56–62.

Handke, L.D., Shivers, R.P. & Sonenshein, A.L., 2008. Interaction of *Bacillus subtilis* CodY with GTP. , 190(3), pp.798–806.

Hartley, D.L. & Kane, J.F., 1986. Recovery and Reactivation of Recombinant Proteins. *Biochemical Society Transactions*, 16, pp.101–102.

Gerth U, Wipat A, Harwood CR, Carter N, Emmerson PT, &Hecker M., 1996. Sequence and transcriptional analysis of clpX, a class-III heat-shock gene of *Bacillus subtilis*. Science 181(1-2) pp. 77-83

Hong, H.A., Duc, L.H. & Cutting, S.M., 2005. The use of bacterial spore formers as probiotics. *FEMS Microbiology Reviews*, 29(4), pp.813–835.

Huang, K., Durand-heredia, J.,& Janakiraman A., 2013. FtsZ ring stability: of bundles, tubules, crosslinks, and curves. *Journal of Bacteriology*, 195(9), pp.1859–1868.

Irene S. Tan,, Weiss CA, Popham DL, &Ramamurthi KS., 2016. A quality control mechanism removes unfit cells from a population of sporulating bacteria. *Dev Cell*, 34(6), pp.682–693.

Janga, S.C.,Martínez-Antonio A., Salgado, H. & Collado-vides, J., 2006. Internal-sensing machinery directs the activity of the regulatory network in *Escherichia coli*. , *Trends in Microbiology*, 14(1), pp. 22-7.

Jürgen B., Hanschke R., Sarvas M., Hecker M., &Schweder T., 2001. Proteome and transcriptome based analysis of *Bacillus subtilis* cells overproducing an insoluble heterologous protein. *Applied Microbiology and Biotechnology*, 55(3), pp.326–332.

Kafri M, Metzl-Raz E, Jona G, Barkai N., 2016. The Cost of Protein Production. *CellReports*, 14(1), pp.22–31..

Karzai, A.W., Roche, E.D. & Sauer, R.T., 2000. The SsrA – SmpB system for protein tagging , directed degradation and ribosome rescue. , 7(6), pp.449–455.

Kenniston JA, Baker TA, Fernandez JM, &Sauer RT., 2003. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell*, 114(4), pp.511–520.

Kim, L., Mogk, A. & Schumann, W., 1996. A xylose-inducible *Bacillus subtilis* integration vector and its application.*Gene*, 181(1–2), pp.71–76.

Kirstein J, Molière N, Dougan DA, &Turgay K, 2009. Adapting the machine: Adaptor proteins for Hsp100/Clp and AAA+ proteases. *Nature Reviews Microbiology*, 7(8), pp.589–599.

Kirstein J., Strahl H., Molière N., Hamoen LW, & Turgay K.,, 2008. Localization of general and regulatory proteolysis in *Bacillus subtilis* cells. *Molecular Microbiology*, 70(3), pp.682–694.

Kirstein J., Dougan DA., Gerth U., Hecker M., &Turgay K, 2007. The tyrosine kinase McsB is a regulated adaptor protein for ClpCP. *The EMBO Journal*, 26(8), pp.2061–2070.

Kirstein J, Zühlke D, Gerth U, Turgay K, &Hecker M., 2005. A tyrosine kinase and its activator control the activity of the CtsR heat shock repressor.*The EMBO Journal* , 24(19), pp.3435–3445.

Klumpp, S. & Hwa, T., 2014. Bacterial growth: Global effects on gene expression, growth feedback and proteome partition. *Current Opinion in Biotechnology*, 28, pp.96–102.

Klumpp, S., Zhang, Z. & Hwa, T., 2009. Growth-rate dependent global effects on gene expression in bacteria. *Cell*, 139(7), pp.1366–1375.

Kock H., Gerth, U. & Hecker, M., 2004. The ClpP peptidase is the major determinant of bulk protein turnover in *Bacillus subtilis*. *Journal of Bacteriology*, 186(17), pp.5856–5864.

Kommineni S., Garg SK., Chan CM., &Zuber P.,2011. YjbH-enhanced proteolysis of Spx by ClpXP in *Bacillus subtilis* is inhibited by the small protein YirB (YuzO). *Journal of Bacteriology*, 193(9), pp.2133–2140.

References

Kriel A, Bittner AN, Kim SH, Liu K, Tehranchi AK, Zou WY, Rendon S, Chen R, Tu BP, &Wang JD. 2012. Direct regulation of gtp homeostasis by (p)ppgpp: a critical component of viability and stress resistance.*Molecular Cell* , 48(2), pp.231–241.

Kriel A, Brinsmade SR, Tse JL, Tehranchi AK, Bittner AN, Sonenshein AL, &Wang JD, 2014. GTP dysregulation in *Bacillus subtilis* cells lacking (p) ppgpp results in phenotypic amino acid auxotrophy and failure to adapt to nutrient downshift and regulate biosynthesis genes.*Journal of Bacteriology* , 196(1), pp.189–201.

Krüger E, Zühlke D, Witt E, Ludwig H, &Hecker M., 2001. Clp-mediated proteolysis in Gram-positive bacteria is autoregulated by the stability of a repressor.*The EMBO Journal* , 20(4):852-63.

Krüger E, Msadek T, Ohlmeier S, &Hecker M., 1997. The *Bacillus subtilisclpC* operon encodes DNA repair and competence proteins. *Microbiology Society*,143 (4), pp.1309-16.

Krüger E, Witt E, Ohlmeier S, Hanschke R, &Hecker M., 2000. The Clp proteases of Bacillus subtilis are directly involved in degradation of misfolded proteins. *Journal of Bacteriology*, 182(11), pp.3259–3265.

Kruger, E., Msadek, T. & Hecker, M., 1996. Alternate promoters direct stress-induced transcription of the *Bacillus subtilisclpC* operon.*Molecular Microbiology* , 20(4), pp. 713-23.

Kruger, E., Volker, U.W.E. & Hecker, M., 1994. Stress induction of *clpC* in *Bacillus subtilis* and its involvement in stress tolerance.*Journal of Bacteriology* , 176(11), pp.3360–3367.

Kubitschek HE., Baldwin WW., Schroeter SJ., & Graetzer R., 1984. Independence of buoyant cell density and growth rate in *Escherichia coli*. *Journal of Bacteriology*, 158(1), pp.296–299.

Kubitschek HE., Baldwin WW. & Graetzer, R., 1983. Buoyant density constancy during the cell cycle of *Escherichia coli*.*Journal of Bacteriology*, 155(3), pp.1027–1032.

Kumar, P., Dubey, R.C. & Maheshwari, D.K., 2012. *Bacillus* strains isolated from rhizosphere showed plant growth promoting and antagonistic activity against phytopathogens. *Microbiological Research*, 167(8), pp.493–499.

Kumar, V., Sangwan, P. & Singh, D., 2014. Industrial enzymes trends, scope and relevance. , *Nova Science Publisher*Book · July 2014.

Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessières P, Bolotin A, Borchert S, Borriss R, Boursier L, Brans A, &Braun M, 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*.*Nature*, 390(6657), pp.249–256.

References

Laloux, G. & Jacobs-Wagner, C., 2014. How do bacteria localize proteins to the cell pole? *Journal of Cell Science*, 127(1), pp.11–19.

Langella O, Valot B, Balliau T, Blein-Nicolas M, Bonhomme L, &Zivy M, 2017. X!TandemPipeline: A Tool to Manage Sequence Redundancy for Protein Inference and Phosphosite Identification. *Journal of Proteome Research*, 16, pp.494–503.

Leonhardt, H. & Alonso, J.C., 1991. Parameters affecting plasmid stability in *Bacillus subtilis*.*ScienceDirect Gene*, 103 (1), pp.107–111.

Levin PA, Kurtser IG, &Grossman AD., 1999. Identification and characterization of a negative regulator of FtsZ ring formation in *Bacillus subtilis*. , 96(17), pp.9642–9647.

Lin JT., Connelly MB., Amolo C, Otani S., & Yaver DS., 2005. Global Transcriptional response of *Bacillus subtilis* to treatment with subinhibitory concentrations of antibiotics that inhibit protein synthesis. *American Society for Microbiology*, 49(5), pp.1915–1926.

Liu, K., Bittner, A.N. & Wang, J.D., 2015. Diversity in (p)ppGpp metabolism and effectors. *Current Opinion in Microbiology*, (24), pp.72–79.

Ludwig H, Homuth G, Schmalisch M, Dyka FM, Hecker M, &Stülke J., 2001. Transcription of glycolytic genes and operons in *Bacillus subtilis*: evidence for the presence of multiple levels of control of the *gapA* operon. *Molecular Microbiology*, 41, pp.409–422.

Malakar, P. & Venkatesh, K. V., 2012. Effect of substrate and IPTG concentrations on the burden to growth of *Escherichia coli* on glycerol due to the expression of Lac proteins. *Applied Microbiology and Biotechnology*, 93(6), pp.2543–2549.

Marr, A.G., 1991. Growth Rate of *Escherichia coli*.*American Society for Microbiology*, 55(2), pp.316–333.

McGinness, K.E., Baker, T.A. & Sauer, R.T., 2006. Engineering Controllable Protein Degradation. *Molecular Cell*, 22(5), pp.701–707.

McKenney PT, Driks A, Eskandarian HA, Grabowski P, Guberman J, Wang KH, Gitai Z, &Eichenberger P, 2010. A distance-weighted interaction map reveals a previously uncharacterized layer of the *B. subtilis* spore coat. *Current Biology*, 20(10), pp.934–938.

Miethke, M., Hecker, M. & Gerth, U., 2006. Involvement of *Bacillus subtilis* ClpE in CtsR Degradation and Protein Quality Control.*Journal of Bacteriology* , 188(13), pp.4610–4619.

Mogk A., Homuth G., Scholz C., Kim L., Schmid FX., &Schumann W., 1997. The GroE chaperonin machine is a major modulator of the CIRCE heat shock regulon of *Bacillus subtilis*. *EMBO Journal*, 16(15), pp.4579–4590.

Molenaar D, van Berlo R, de Ridder D, &Teusink B, 2009. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular Systems Biology*, 5(323), pp.1–10.

Molière N, Hoßmann J, Schäfer H, &Turgay K., 2016. Role of Hsp100/Clp protease complexes in controlling the regulation of motility in *Bacillus subtilis*. *Frontiers in Microbiology*, (7), pp.1–16.

Molière, N. & Turgay, K., 2009. Chaperone-protease systems in regulation and protein quality control in *Bacillus subtilis.Research in Microbiology*, 160(9), pp.637–644.

Molle V, Nakaura Y, Shivers RP, Yamaguchi H, Losick R, Fujita Y, &Sonenshein AL., 2003. Additional Targets of the *Bacillus subtilis* Global Regulator CodY Identified by Chromatin Immunoprecipitation and Genome-Wide Transcript Analysis. , 185(6), pp.1911–1922.

Monod, J., 1949. The Growth of Bacterial Cultures.*Annual Review of Microbiology,* (3) pp. 371-394

Mori M, Hwa T, Martin OC, De Martino A, &Marinari E**.**, 2016. Constrained Allocation Flux Balance Analysis. *PLOS Computational Biology*, 12(16), pp.1–24.

Mukherjee S, Bree AC, Liu J, Patrick JE, Chien P, &Kearns DB**.**, 2015. Adaptor-mediated Lon proteolysis restricts *Bacillus subtilis* hyperflagellation. *Proceedings of the National Academy of Sciences*, 112(1), pp.250–255.

Nanamiya H, Kasai K, Nozawa A, Yun CS, Narisawa T, Murakami K, Natori Y, Kawamura F, &Tozawa Y, 2008. Identification and functional analysis of novel (p) ppGpp synthetase genes in *Bacillus subtilis*. *Molecular Microbiology*, (67), pp.291–304.

Nicolas, P. et al., 2012. Condition-dependent transcriptome architecture in *Bacillus subtilis*.*Science* , 335(6072), pp. 1103-1106.

O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, &Palsson BØ, 2013. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, 9(693).

Orth, J.D., Thiele, I. & Palsson, B.O., 2010. What is flux balance analysis? *Nature Biotechnology*, 28(3), pp.245–248.

Overton, T.W., 2014. Recombinant protein production in bacterial hosts. *Drug Discovery Today*, 19(5), pp.590–601..

Pan, Q., Garsin, D.A. & Losick, R., 2001. Self-reinforcing activation of a Cell-Specific transcription factor by proteolysis of an anti-sigmafactor in *B . subtilis*. *Molecular Cell*, 8(4), pp.873–883.

Partridge, S.R. & Errington, J., 1993. The importance of morphological events and

intercellular interactions in the regulation of prespore-specific gene expression during sporulation in *Bacillus subtilis.Molecular Microbiology*, 8(5), pp.945–955.

Pédelacq JD, Cabantous S, Tran T, Terwilliger TC, &Waldo GS., 2006. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology*, 24(1), pp.79–89.

Persuh, M., Mandic-mulec, I. & Dubnau, D., 2002. A MecA paralog , YpbH, binds ClpC, affecting both competence and sporulation.*Journal of Bacteriology*, 184(8), pp.2310–2313.

Ploss TN., Reilman E., MonteferranteCG., DenhamEL. , Piersma S., Lingner A., Vehmaanperä J., Lorenz P., &vanDijl JM., 2016. Homogeneity and heterogeneity in amylase production by *Bacillus subtilis* under different growth conditions. *Microbial Cell Factories*, 15(57) pp.1–16.

Pohl S, Bhavsar G, Hulme J, Bloor AE, Misirli G, Leckenby MW, Radford DS, Smith W, Wipat A, Williamson ED, Harwood CR, &Cranenburgh RM., 2013. Proteomic analysis of *Bacillus subtilis* strains engineered for improved production of heterologous proteins. *Proteomics*, 13(22), pp.3298–3308.

Potrykus, K. & Cashel, M., 2008. (p) ppGpp : Still Magical ?*Annual Reviews Microbiology* (62), pp. 35-51

Price CW, Fawcett P, Cérémonie H, Su N, Murphy CK, &Youngman P., 2001. Genome-wide analysis of the general stress response in *Bacillus subtilis*. *Molecular Microbiology*, 41, pp.757–774.

Palva I., 1982. Molecular cloning of a-amylase gene from *Bacillus amyloliquefaciens* and its expression in *B. subtilis*. *Gene*, 19(1), pp.81–87.

Pulschen AA, Sastre DE, Machinandiarena F, Crotta Asis A, Albanesi D, de Mendoza D, &Gueiros-Filho FJ., 2017. The stringent response plays a key role in *Bacillus subtilis* survival of fatty acid starvation. *Molecular Microbiology*, 103(4), pp.698–712.

Rodgers, P.B., 1993. Potential of biopesticides in agriculture. *Pesticide Science*, 39(2), pp.117–129.

Sadaie, Y. & Kada, T., 1983. Formation of competent Bacillus subtilis cells. *Journal of Bacteriology*, 153(2), pp.813–821.

Samarrai W., Liu DX., White AM., Studamire B., Edelstein J, Srivastava A, Widom RL., &Rudner R., 2011. Differential responses of *Bacillus subtilis* rRNA promoters to nutritional stress. *Journal of Bacteriology*, 193(3), pp.723–733.

Sauer C., Syvertsson S., Bohorquez LC., Cruz R., Harwood CR., van Rij T., &Hamoen LW, 2016. Effect of genome position on heterologous gene expression in *Bacillus subtilis*: an unbiased analysis. *ACS Synthetic Biology*, 5(9), pp.942–947.

Sauer C , Ver Loren van Themaat E, Boender LGM, Groothuis D, Cruz R, Hamoen LW, Harwood CR, &van Rij T., 2018. Exploring the nonconserved sequence space of synthetic expression modules in *Bacillus subtilis*. *ACS Synthetic Biology*, 7(7), pp.1773–1784.

Sauer RT, Bolon DN, Burton BM, Burton RE, Flynn JM, Grant RA, Hersch GL, Joshi SA, Kenniston JA, Levchenko I, Neher SB, Oakes ES, Siddiqui SM, Wah DA, &Baker TA., 2004. Sculpting the proteome with AAA+ proteases and disassembly machines. *Cell*, 119(1), pp.9–18.

Sauer, R.T. & Baker, T.A., 2011. AAA+ Proteases : ATP-fueled machines of protein destruction.*Annual Reviews of Biochemistry*, (80), pp. 587-612

Schaechter, B.Y.M. & Kjeldgaard, N., 1958. Dependency on medium and temperature of cellsize and chemical composition during balanced growth of *Salmonella typhimurium*.*Microbiology*, 19(3), , pp. 592-606

Schäfer H, Heinz A, Sudzinová P, Voß M, Hantke I, Krásný L, &Turgay K., 2019. Spx , the central regulator of the heat and oxidative stress response in *B . subtilis*, can repress transcription of translation-related genes. *Molecular Microbiology*, 111(111), pp.514–533.

Schallmey, M., Singh, A. & Ward, O.P., 2004. Developments in the use of *Bacillus* species for industrial production.*Canadian Reviews of Microbiology* , 17(1), pp.1–17.

Schleif, R., 1967. Control of Production of Ribosomal Protein. *Journal of Molecular Biology*, 27(1), pp.41–55.

Schlothauer T., Mogk A., DouganDA., Bukau B., & Turgay K., 2003. MecA, an adaptor protein necessary for chaperone activity. *PNAS*, 100(5), pp. 2306-2311.

Schulz, A. & Schumann, W., 1996. *hrcA*, the first gene of the *Bacillus subtilisdnaK* operon encodes a negative regulator of class I heat shock genes. *Journal of Bacteriology*, 178(4), pp.1088–1093.

Schumann, W., 2003. The Bacillus subtilis heat shock stimulon.*Cell Stress and Chaperones*, 8(3), pp.207–217.

Scott M.,Klumpp S, Mateescu EM, &Hwa T., 2014. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Molecular Systems Biology*, 10(8), pp.747–747.

Scott M, Gunderson CW, Mateescu EM, Zhang Z, &Hwa T., 2010. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330(6007), pp.1099–1102.

Sharpe ME, Hauser PM, Sharpe RG, &Errington J., 1998. *Bacillus subtilis* cell cycle as studied by fluorescence microscopy: constancy of cell length at initiation of DNA replication and evidence for active nucleoid partitioning. *Journal of Bacteriology*, 180(3), pp.547–555.

Si F, Treut G.,Sauls JT., Vadia S., Levin PA., & Jun S., 2019. Mechanistic origin of cell-size control and homeostasis in bacteria.*Current Biology* , pp.1–22.

Sharma RR., Singh D., &Singh R., 2013. Biological control of postharvest diseases of fruits and vegetables by microbial antagonists: *Applied and Environmental Microbiology*, 8(2), pp.274–302.

Siala, A., Hill, I.R. & Gray, T.R.G., 1974. Populations of spore-forming bacteria in an acid forest soil, with special reference to *Bacillus subtilis*. *Journal of General Microbiology*, 81(1), pp.183–190.

Siegal, M.L., 2015. Shifting Sugars and Shifting Paradigms.*PLOS Biology* , 13(2), pp.1–7.

Slack, F.J., Mueller, J.P. & Sonenshein, A.L., 1993. Mutations that relieve nutritional repression of the bacillus subtilis dipeptide permease operon.*Journal of Bacteriology*, 175(15), pp.4605–4614.

Sleight, S.C. & Sauro, H.M., 2013. Visualization of evolutionary stability dynamics and competitive fitness of *Escherichia coli* engineered with randomized multigene circuits. *ACS Synthetic Biology*, 2(9), pp.519–528.

Soga T, Ohashi Y, Ueno Y, Naraoka H, Tomita M, &Nishioka T., 2003. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *Journal of Proteome Research*, 2(5), pp.488–494.

Solomon, J.M. & Grossman, A.D., 1996. Who's competent and when: regulation of natural genetic competence in bacteria. *Elsevier Science*, 12(4), pp.10536–10540.

Sousa, C., De Lorenzo, V. & Cebolla, A., 1997. Modulation of gene expression through chromosomal positioning in *Escherichia coli*.*Microbiology*, 143(6), pp.2071–2078.

References

Slack FJ, Serror P, Joyce E, &Sonenshein AL., 1995. A gene required for nutritional repression of the *Bacillus subtilis* dipeptide permease operon.*Molecular Microbiology* , 15, pp.689–702.

Steinmetz M., Le Coq D., Aymerich S., Gonzy-Tréboul G., Gay P., 1985. The DNA sequence of the gene for the secreted *Bacillus subtilis* enzyme levansucrase and its genetic control sites. *MGG Molecular & General Genetics*, 200(2), pp.220–228.

Stoebel, D.M., Dean, A.M. & Dykhuizen, D.E., 2008. The cost of expression of *Escherichia coli lac* operon proteins is in the process, not in the products. *Genetics*, 178(3), pp.1653–1660.

Stülke J, Arnaud M, Rapoport G, &Martin-Verstraete I., 1998. PRD – a protein domain involved in PTS-dependent induction and carbon catabolite repression of catabolic operons in bacteria. *Molecular Microbiology*, 28, pp.865–874.

Stülke, J. & Hillen, W., 1999. Carbon catabolite repression in bacteria. *Current Opinion in Microbiology*, 2(2), pp.195–201.

Sun G., Sharkova E., Chesnut R., Birkey S., Duggan MF., Sorokin A., Pujic P., Ehrlich SD., & Hulett FM., 1996. Regulators of aerobic and anaerobic respiration in *Bacillus subtilis*.*Journal of Bacteriology* , 178(5), pp.1374–1385.

Turgay K., Hahn J., Burghoorn J., & Dubnau D. et al., 1998. Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. *The EMBO Journal*, 17(22), pp.6730–6738.

Thakore, Y, 2006.The biopesticide market for global agricultural use. *Biocontrol*. 2(3).

Tomoyasu, T., Bukau, B. & Arsene, F., 2000. The heat shock response of *Escherichia coli*. *International Journal of food Microbiology*, 55(1-3), pp.3–9.

Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, &Cox J., 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, 13(9), pp. 731-740.

Vadia, S. & Levin, P.A., 2016. Growth rate and cell size : A re-examination of the growth law. *ScienceDirect Current Opinion in Microbiology*, (24), pp.96–103.

Vadia S, Tse JL, Lucena R, Yang Z, Kellogg DR, Wang JD, &Levin PA, 2018. Fatty acids availability sets cell envelope capacity and dictates microbial cell size. *Curr Biol*, 27(12), pp.1757–1767.

Varma, A. & Palsson, B., 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and Environmental Microbiology*, 60(10), pp.3724–3731.

Vazquez A, Beg QK, Demenezes MA, Ernst J, Bar-Joseph Z, Barabási AL, Boros LG, &Oltvai ZN., 2008. Impact of the solvent capacity constraint on *E . coli* metabolism. BMC Systems biology , 2(7), pp.1–10.

Völker U, Engelmann S, Maul B, Riethdorf S, Völker A, Schmid R, Mach H,&Hecker M., 1994. Analysis of the induction of general stress proteins of *Bacillus subtilis*. *Microbiology Society*, 140(4), pp.741–752.

Verplaetse E, Slamti L, Gohar M, &Lereclus D., 2015. Cell differentiation in a *Bacillus thuringiensis* population during planktonic growth, biofilm formation, and host infection. *mBio*, 6(3), pp.1–10.

Villaverde, A. & Carrió, M.M., 2003. Protein aggregation in recombinant bacteria: Biological role of inclusion bodies. *Biotechnology Letters*, 25(17), pp.1385–1395.

Warner, J.B. & Lolkema, J.S., 2003. CcpA-dependent carbon catabolite repression in bacteria. *Micobiology and molecular biology reviews*, 67(4), pp.475–490.

Weart RB, Lee AH, Chien AC, Haeusser DP, Hill NS, &Levin PA, 2007. A metabolic sensor governing cell size in bacteria. *Cell*, 130 (2) pp.335–347.

Weart, R.B. & Levin, P.A., 2003. Growth rate-dependent regulation of medial FtsZ ring formation. *Journal of Bacteriology*, 185(9), pp.2826–2834.

Weiße AY, Oyarzún DA, Danos V, &Swain PS, 2015. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, 112(9), pp.1038–1047.

Wellington, S.R. & Spiegelmane, G.B., 1993. The kinetics of formation of complexes between *Escherichia coli* RNA polymerase and the *rrnB* P1 and P2 promoters of *Bacillus subtilis*. *Journal of Biological Chemistry*, (10), pp.7205–7214.

Wendrich TM, Blaha G, Wilson DN, Marahiel MA, &Nierhaus KH., 2002. Dissection of the mechanism for the stringent factor RelA.*Molecular Cell* , 10(4), pp.779–788.

Wendrich, T.M. & Marahiel, M.A., 1997. Cloning and characterization of a *relA/spoT* homologue from *Bacillus subtilis*. *Molecular Microbiology*, 26(1), pp.65–79.

Westers, L., Westers, H. & Quax, W.J., 2004. *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1694(1–3), pp.299–310.

References

Wiegert, T. & Schumann, W., 2001. SsrA-mediated tagging in *Bacillus subtilis*. *Journal of bacteriology*, 183(13), pp.3885–3889.

Winkler J, Seybert A, König L, Pruggnaller S, Haselmann U, Sourjik V, Weiss M, Frangakis AS, Mogk A,&Bukau B., 2010. Quantitative and spatio-temporal features of protein aggregation in Escherichia coli and consequences on protein quality control and cellular ageing.*The EMBO Journal* , 29(5), pp.910–923.

Wong, S., 1995. Advances in the use of *Bacillus subtilis* for the expression and secretion of heterologous proteins. *Current Opinion in Biotechnology*, 6(5) pp.517–522.

Wu SC, Ye R, Wu XC, Ng SC, &Wong SL., 1998. Enhanced secretory production of a single-chain antibody fragment from *Bacillus subtilis* by coproduction of molecular chaperones. *Journal of Bacteriology*, 180(11), pp.2830–2835.

Yan X, Hu S, Guan YX, &Yao SJ, 2012. Coexpression of chaperonin GroEL/GroES markedly enhanced soluble and functional expression of recombinant human interferon-gamma in *Escherichia coli*. *Applied Microbiology and Biotechnology*, 93(3), pp.1065–1074.

Yu Y, Yan F , He Y, Qin Y, Chen Y, Chai Y, &Guo JH, 2018. The ClpY-ClpQ protease regulates multicellular development in *Bacillus subtilis.Microbiology Society*, 164(5) pp.848–862.

Zhang K, Su L, Duan X, Liu L, &Wu J., 2017. High-level extracellular protein production in *Bacillus subtilis* using an optimized dual-promoter expression system. *Microbial Cell Factories*, 16(1), pp.1–15.

Zuber, U. & Schumann, W., 1994. CIRCE, a novel heat-shock element involved in regulation of heat-shock operon dnaK of *Bacillus subtilis*. *Journal of Bacteriology*, 176(5), pp.1359–1363.

université PARIS-SACLAY | ÉCOLE DOCTORALE Structure et dynamique des systèmes vivants (SDSV)

Titre: Ré-allocation des ressources cellulaires pour la production de protéines hétérologues chez *Bacillus subtilis*

Mots clés: Protéine gratuite, ressources cellulaires, surproduction de protéines, *Bacillus subtilis.*

Résumé: La synthèse de protéines recombinantes chez les microorganismes est d'un intérêt majeur pour la production de produits biopharmaceutiques, thérapeutiques et enzymatiques industriels. Cependant, la surproduction de protéines à un effet néfaste sur la physiologie cellulaire. Les ressources cellulaires (métabolites, énergie, machinerie moléculaire, espace cytosolique, *etc*.) sont en effet partagées entre les protéines de l'hôte et la protéine "gratuite". Cette surcharge non naturelle entraîne une croissance plus lente et des rendements en protéines plus faibles, un phénomène connu sous le nom de "burden". Dans mon projet de doctorat, il s'agissait (1) de déchiffrer les conséquences de la surproduction de protéines gratuites sur la physiologie cellulaire, (2) d'identifier le type de ressources limitantes, et (3) de surmonter cette limitation pour améliorer la production de protéines. Afin de déchiffrer les conséquences de la surproduction de protéines (1), nous avons analysé le taux de croissance, la production de protéines d'intérêt et le protéome de souches de *Bacillus subtilis* surproduisant divers niveaux de protéines rapportrices. Les protéines rapportrices ont été choisies de manière à être facilement quantifiables par fluorescence et par des tests d'activité (*i.e.* GFP, mKate2, LacZ, *etc*.). Pour obtenir les différents niveaux d'expression, nous avons construit des séquences synthétiques par assemblage de promoteurs constitutifs et inductibles et de régions d'initiation de traduction (TIR, RBS) variés. Nous avons ainsi montré que plus la quantité (et la taille) de la protéine produite était élevée, plus les taux de croissance étaient faibles et plus la taille des cellules était élevée. Par exemple, le taux de croissance a diminué de plus de 20 % lorsque la GFP était surproduite à plus de 5 % de la quantité totale de protéines solubles, selon des quantifications biochimiques et de fluorescence. Pour identifier le type de ressources limitantes (2), nous avons effectué une quantification relative des protéines sur les souches surproductrices de GFP et montré que certaines protéines non essentielles étaient moins abondantes dans ces souches. Nous avons ensuite dégradé spécifiquement les protéines rapportrices à l'aide d'un outil de biologie de synthèse précédemment mis au point pour *B. subtilis*, afin que les acides aminés puissent être recyclés dans le pool de ressources cellulaires. Avec une dégradation de 50-60% de GFP et mKate2, nous avons observé une restauration de 50% du taux de croissance. Ces résultats suggèrent que la quantité d'acides aminés (et par conséquent leur utilisation dans la synthèse des protéines) est le principal type de ressources limitantes. Pour améliorer la production de protéines (3), nous avons cherché à développer un système synthétique de recyclage des acides aminés basé sur le système de dégradation mentionné ci-dessus en surproduisant les protéases d'*E. coli* et *B. subtilis* (ClpXP) avec une protéine adaptatrice (SspB) d'*E. coli*. Cet outil pourrait permettre de dégrader spécifiquement des protéines non essentielles pour économiser des ressources cellulaires. Nous avons montré que la surproduction de ClpXP ou de SspB/ClpXP était suffisante pour permettre une dégradation complète des protéines produites à des niveaux bas et intermédiaires, et jusqu'à 50% des protéines fortement produites. Comme ClpXP est une protéase impliquée dans la réponse au stress, nous avons cherché à savoir si la surproduction de ClpXP pouvait avoir des conséquences négatives sur la physiologie cellulaire. Une quantification relative des protéines sur une souche surproductrice de ClpXP a montré que la surproduction de ClpXP provoque une réorganisation globale du protéome sans toutefois affecter le taux de croissance de la cellule.

université PARIS-SACLAY
ÉCOLE DOCTORALE
Structure et dynamique des systèmes vivants (SDSV)

Title: Re-allocation of cellular resources for the production of heterologous proteins in *Bacillus subtilis*

Key words: Gratuitous protein, cell resources, recombinant protein production, resource allocation, *Bacillus subtilis.*

Abstract: Recombinant protein production in microorganisms is of great interest for the production of biopharmaceuticals, therapeutics and industrial enzymes. However, recombinant protein production has always shown a harmful effect on the microorganism cell physiology when excessively produced. Cell resources (*i.e.* metabolites, energy, molecular machinery, cytosolic space, *etc*.) are used to produce the host's proteins and the overproduced gratuitous protein. As a result, this unnatural extra load typically leads to slower growth and lower protein yields, a phenomenon known as 'burden'. This burden comes from the fact that the recombinant protein has no benefit for the microorganism, and that it only uses cell resources at the expense of the production of the endogenous essential proteins. In my PhD project, the issues were (1) to decipher the consequences of gratuitous protein overproduction on the cell physiology, (2) to identify the limiting type of resources, and (3) to overcome this limitation to improve protein production. To address the first issue (1), we analyzed growth rates, production of several proteins of interest, and genome-wide proteomes of *Bacillus subtilis* strains overproducing various levels of reporter proteins. The reporter proteins were chosen so that they were easily quantifiable by fluorescence and β-galactosidase activity assays (*i.e.* GFP, mKate2, LacZ, *etc*.). To obtain the various levels of expression, we built synthetic sequences made of the assembly of various constitutive and inducible promoters and translation initiation regions (TIR, RBS). Hence, we showed that higher was the amount (and size) of the protein produced, lower were the rates of growth and higher were the cell sizes. For instance, the growth rate decreased down by over 20% when GFP was overproduced above 5% of the total soluble protein amount according to both biochemical and fluorescence assays. To further identify the limiting type of resources (2), we performed a relative protein quantification on the strains overproducing GFP at different levels. Hence, we showed that some non-essential proteins were less abundant in the strains overproducing GFP. We next targeted the reporter proteins for degradation using a synthetic tool previously engineered in *B. subtilis*, so that amino acids can be recycled back to the pool of cell resources. Degrading the reporter gratuitous protein should also relieve the constraint on the cytosolic density by liberating intracellular space. With a degradation of 50-60% of GFP and mKate2, we observed a 50% restoration of the growth rate. This result together with the proteome analysis suggested that the amount of amino acids (and consequently their utilization in protein synthesis) was the main limiting type of resources. To overcome this limitation and improve protein production (3), we aimed at exploring a synthetic, amino acid recycling system based on the above mentioned degradation system. We decided to improve the targeted degradation system by overproducing the *E. coli* and *B. subtilis* ClpXP proteases together with an *E. coli* adaptor protein SspB. This tool may allow to target proteins for degradation in order to save resources and improve the production of a protein of interest. We showed that the overproduction of either ClpXP or SspB/ ClpXP were sufficient to allow a complete degradation of the proteins produced low and intermediate levels, and up to 50% of degradation of the proteins highly produced. As ClpXP is a protease involved in stress responses, we aimed to know whether the overproduction of ClpXP may have negative consequences on the cell physiology. We therefore performed relative protein quantification on a strain overproducing ClpXP. The results showed that ClpXP overproduction causes a global reorganization on the proteome without affecting the growth rate of the cell.