Analysis of Cellular Heterogeneity for Enhancing the Protein Production in Industrial Scales

Christopher Sauer

Thesis submitted for the degree of
Doctor of Philosophy

Institute for Cell and Molecular Biosciences
Newcastle University

November 2016

Abstract

The ability of microorganisms such as *B. subtilis* to differentiate into subpopulations with distinct phenotypes has been an important feature for their persistence in the face of environmental fluctuations. This so called heterogeneity provides *B. subtilis* with a selective advantage in its natural habitat, despite such bet-hedging strategies being energetically unfavourable processes. Unfortunately, heterogeneous gene expression in an industrial setting will reduce process efficiency by diverting substrate and energy consumption away from product formation.

This thesis investigates the relationship between cellular heterogeneity and enzyme production by analysing different physiological and developmental pathways during the high level production of the native xylanase XynA and the heterologous amylase AmyM. The potential cellular heterogeneity of the quality control system was investigated using fluorescence microscopy and flow cytometry in different mutants.

For the investigation of potential strategies to reduce the heterogeneous expression and optimise protein production in biotechnologically relevant processes, plasmid and integration based systems were compared and analysed by using fluorescent reporter fusions reflecting the enzymes transcription level and induced secretion stress.

Furthermore, this thesis investigated the approach of generating a synthetic expression library to discover new expression units demonstrating high expression levels and reduced heterogeneous activities. The library was screened using flow cytometry and modules were assessed using the production of the industrially relevant enzyme XynA.

Finally, controlled gene expression levels are of importance for the *in silico* design of novel pathways (synthetic biology). However, the expression of integrated genes can depend on their chromosomal location. This thesis assessed the impact of genomic location and orientation on gene expression levels by using an unbiased transposon approach for the random integration of an expression cassette comprising a divergent fluorescence and enzyme reporter system. In addition, the gene dosage effect was investigated in an industrially relevant setting using XynA as model enzyme.

Declaration

I certify that this thesis contains my own work, except where acknowledged, and that no part of this material has previously been submitted for a qualification at this or any other university.

Christopher Sauer

Acknowledgements

First of all, I would like to thank my supervisors Prof. Leendert Hamoen, Prof. Colin Harwood and Dr. Tjeerd van Rij for giving me the chance to be part in such an amazing project and their enormous help during my Ph.D. It was a very interesting experience to tackle objectives from different angles. Leendert has been an interesting person with many qualities: his straightforwardness was sometimes disarming but this has strongly contributed to the success of the project. Without Colin I would have missed vital stories about scientific and daily life, not to forget the invitations to his home for great celebrations and barbeque evenings. I will miss his British humour and ability to raise issues in a way that I would not feel threatened, which is a rare personal quality nowadays. Tjeerd has been kind and always helpful and supportive for all problems that occurred.

A very big thank you goes to Rita Cruz who has been my closest friend and colleague during our joint Ph.D. programme. We survived the moving house every nine months and celebrated together on many occasions. Our work had an almost symbiotic character.

One of the biggest supporters though was my girlfriend Angie that kept listening and motivating myself and I don't know where I would be without her – probably still not able to speak an applicable English.

I would like to thank past and present members of the Hamoen lab, including Simon, Henrik, Tom, Declan and all the people in the CBCB that made my stay in the UK so positive, including the weekly football games, visits to the pub and the (in)famous CBCB Christmas parties. I am also very grateful for my kind colleagues at DSM, especially Marcel and Daphne which strongly supported me during my stay.

Contributions to this thesis by colleagues are indicated in the text, and particularly those of Rita Cruz (together we performed the 24 h enzyme activity assay, whereas she carried out experiments to measure secretion stress and enzyme activities and AmyM production in protease deficient strains; Chapter 4) and Simon Syvertsson (constructed pSS125, carried out transposon frequency assay, back-crossing of strains and arbitrary primed PCR; Chapter 6)

This Ph.D. was supported by the European Commission funded Marie Curie Initial Training Network ATRIEM (Project No. 317228).

Content

List of Figures

List of Tables

Abbreviations

ATP	Adenosine triphosphate
AU	Arbitrary Units (fluorescence intensity on a 16-bit scale)
bp	Base pairs, prefixes: kilo- (k) or mega- (M)
BRB	Bacillus strain lacking protease(s)
cat	Chloramphenicol acetyltransferase
CCD	Charge-coupled device (digital camera)
cDNA	Complementary DNA
chDNA	Chromosomal DNA
CTD	C-terminal domain
Da	Dalton - atomic mass unit, prefix: kilo (k)
DAPI	4’,6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleoside triphosphates
EDTA	Ethylenediaminetetraacetic acid
EPS	Exopolysaccharides
ery	Erythromycin
FACS	Fluorescence-activated cell sorting
g	Gram - metric SI prefixes can be applied
GFP	Green fluorescent protein
GOI	Gene of interest
h	Hour
IPTG	Isopropyl -d-1-thiogalactopyranoside
kan	Kanamycin
l	Litre - metric SI prefixes can be applied
ITR	Inverse terminal repeats
LB	Luria-Bertani (medium)
M	Molar - metric SI prefixes can be applied
m	Metre - metric SI prefixes can be applied
mCherry	Red fluorescent protein

MCS	Multiple cloning site
MilliQ	Uultrapure deionised water
min	Minute
mRNA	Messenger RNA
MTP	Microtitre plate
MU	Miller Units
N	In DNA sequences: can be substituted with any of the following nucleotides: A, T, G or C (RNA: U instead of T)
NAD	Nicotinamide adenine dinucleotide
OD_{600}	Optical density at 600 nm
ONPG	o-Nitrophenyl-β-D-galactopyranosid
PBS	Phosphate buffered saline
PC	Phase contrast
PCR	Polymerase chain reaction
qPCR	Quantitative PCR
RBS	Ribosome binding site
RFP	Red fluorescent protein
RLT	Lysis buffer for lysing cells and tissues prior to RNA isolation
RNA	Ribonucleic acid
RNAP	RNA Polymerase
RPE	Buffer for washing membrane-bound RNA
rRNA	Ribosomal RNA
rpm or ×*g*	revolutions per minute or gravity
s	Second - metric SI prefixes can be applied
SDS	Sodium dodecyl sulfate
sfGFP	Superfolder GFP
spec	Spectinomycin
TCA	Trichloroacetic acid
t_m	Annealing temperature
TRIS	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
UP element	Sequence that provides contact for the interaction with RNAP α-CTD
UTR	Untranslated region
X-gal	5-Brom-4-chlor-3-indoxyl-β-D-galactopyranosid

Chapter 1

Introduction

1.1 *Bacillus subtilis* - A soil-dwelling bacterium

The soil provides microorganisms with a habitat that includes an extraordinarily wide range of nutrient sources. However, this milieu can be perturbed by rapid changes in its chemical and physical parameters. These environmental fluctuations require a high degree of adaptability among their inhabitants to both detect these variations and adjust their physiology and development accordingly (Msadek, 1999). *Bacillus subtilis* is one of many microorganisms that has successfully adapted to the complex and heterogeneous environment of the soil. It is a non-pathogenic, rod-shaped Gram-positive bacterium with a low GC-content (*Firmicutes*). The genome of the best-studied representative of this species, *B. subtilis subsp. subtilis* strain 168, encodes more than 4100 genes with a total length of 4.2 Mbp. The genome was sequenced by an international consortium of more than 30 laboratories (Kunst *et al.*, 1997) and reannotated in 2009 (Barbe *et al.*, 2009). It was the first non-pathogenic bacterium to be shown to be transformable with exogenous DNA (Anagnostopoulos and Spizizen, 1961) and, as a result has been intensely studied for more than six decades. *B. subtilis* becomes competent naturally and homologous extracellular DNA is taken up by the competence machinery and recombined with high efficiency into the chromosome. As a result, this bacterium is a widely accepted model organism for academic and industrial research.

Competence development is one of many strategies *B. subtilis* has evolved to improve its survival potential under less favourable environmental conditions. Other strategies include the production of flagella to seek new nutrient sources (Guttenplan *et al.*, 2013, Mukherjee and Kearns, 2014) and macromolecular hydrolases for the utilisation of environmental proteins, starch and lignocellulosic compounds (Wu *et al.*, 1991). To outcompete competitive species under nutrient limiting conditions, *B. subtilis* synthesises antibiotics such as subtilin to improve survivability (Stein, 2005). Once the cellular stress levels exceed certain limits, the vegetative cycle is

terminated and sporulation is initiated (Driks, 2002, Piggot and Hilbert, 2004). The capability to form endospores allows *B. subtilis* to overcome periods with little or no influx of nutrients and to survive under several stress situations including extreme temperatures and pH values and radiation. The differentiation process into dormant endospores has been intensively studied to uncover the underlying mechanisms that allow a survival for years or even millions of years (Vreeland *et al.*, 2000, Errington, 2003, Higgins and Dworkin, 2012). Taken together, the ability of *B. subtilis* to differentiate into various physiological and morphological forms has been an important feature of its ability to persist in the face of the fluctuating conditions in its natural habitat.

1.2 Cellular heterogeneity

1.2.1 Gene regulation

In bacteria, gene regulation includes the modulation of gene expression at a specific time that reflects an adaptation to rapidly changing conditions in their habitat. Gene regulation is primarily achieved at the level of transcription which is initiated by the recruitment of the RNA polymerase (RNAP) (Figure 1.1). The core RNAP consists of several subunits (β, β', and α_2) including the N-terminal domain (α-NTD) responsible for the assembly of the β and β' subunits and the C-terminal domain (α-CTD) potentially binding the upstream promoter (UP) element on the DNA (Browning and Busby, 2004). Its purpose is the transcription of DNA into messenger RNA (mRNA). Transcription initiation is achieved by the RNAP holoenzyme, which contains the subunits of the core enzyme and is assembled by interacting with the σ subunit to facilitate the recognition of promoter elements. Additionally, this σ factor directs the RNAP holoenzyme to target promoter sequences and facilitates the DNA unwinding (open complex) close to the transcriptional start before it becomes discharged (Wosten, 1998). *B. subtilis* comprises multiple σ factors including vegetative factors (σ^A, σ^B, σ^C, σ^D, σ^H, σ^L) and sporulation-specific factors (σ^E, σ^F, σ^G, σ^K) for an adaptive gene regulation in fluctuating environments (Haldenwang, 1995). The housekeeping σ^A factor is involved in gene regulation during vegetative growth and early sporulation, whereas the alternative σ^D factor is required for motility development and, as observed later, an essential part for the bimodal formation of flagella. Other σ factors are involved in the general stress response (σ^B) or competence development (σ^H) (Haldenwang, 1995). Since the interaction of σ factors with RNAPs is a transient process and new σ factors are recruited for a subsequent transcription initiation once

it becomes discharged, σ factors negatively affect genes being regulated by competing σ factors which simultaneously attempt to form the RNAP holoenzyme (Osterberg *et al.*, 2011). Further gene regulation can occur by the activity of anti-σ factors which antagonise their associated σ factors (e.g. motility) and therefore negatively regulate their activity, whereas anti-anti-σ factors release σ factors by binding their anti-σ factors.

Once the RNAP holoenzyme is assembled it binds to the promoter elements which include the UP element, -35 and -10 regions (located 35 and 10 bp upstream from the transcriptional start) and the extended -10 element (Figure 1.1). However, their influence on transcription initiation varies from promoter to promoter emphasising again the diversity of different sigma factor recognition sites. Additionally, the intervening sequence between the -35 and -10 region has been shown to influence the RNAP binding efficiency and open complex formation (Liu *et al.*, 2004). Transcription factors can up- or downregulate the expression of genes by binding promoter regions and act solely as activators or repressors, whereas others operate as either depending on the specific promoter (e.g. Spo0A) (Perez-Rueda and Collado-Vides, 2000, Fujita *et al.*, 2005). By enhancing the promoters' affinity for the RNAP, activators simply improve transcription initiation. In contrast, repressors reduce transcription initiation by steric hindrance, looping and the modulation of activators (Browning and Busby, 2004). The activity of transcription factors can be regulated by antagonism (e.g. SinI binds SinR) or degradation processes (e.g. ComK).

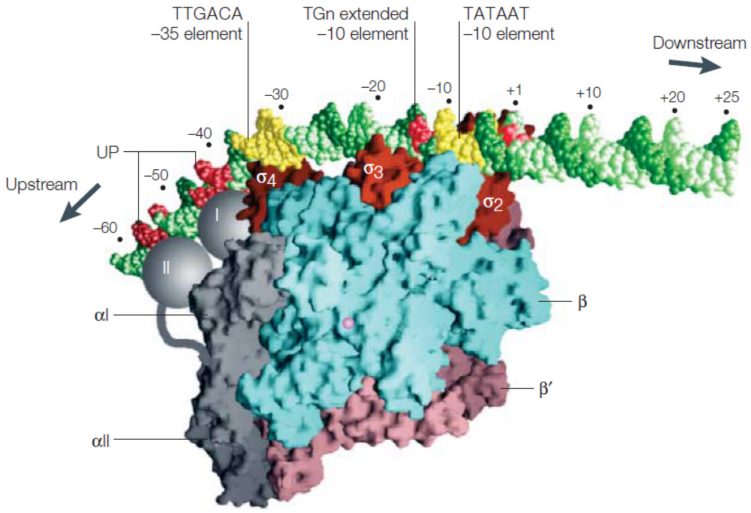


Figure 1.1: RNA polymerase interacting with specific promoter regions. Crystallographic model of RNAP holoenzyme binding the -35 (TTGACA) and -10 (TATAAT) regions (yellow) and the UP and TGn elements (red). RNAP containing the different β subunits (blue and pink), α-CTDs (grey) and σ domains (red). Domains I and II of α-CTDs bind the UP element. The active site of the enzyme is depicted by the Mg^{2+} ion (magenta). DNA strands are presented in green (from Browning and Busby (2004)).

1.2.2 The source of phenotypic heterogeneity

Microbial populations generally perceive environmental changes by recognising small molecules using complex sensing systems. These systems lead to the activation of distinct pathways and changes in gene expression that reflect an appropriate response to fluctuations in the environment. In the past, it has generally been assumed that isogenic cells growing under identical conditions respond equally to a homogeneously distributed signal. This view has changed in recent years with the advent of fluorescence microscopy and flow cytometry that has facilitated the analysis of gene expression at the single-cell level. This research revealed distinct cell types engaged in a variety of physiological and developmental pathways within a single population. The resulting heterogeneity reflects phenotypic cell-to-cell variation that occurs independently of environmental or genetic changes amongst individual cells in the population (Avery, 2006, Balazsi *et al.*, 2011, Levine *et al.*, 2013). However, varying environments and genetic differences are difficult to be excluded completely during a typical experiment since minimally fluctuating environments and mutations on the single-cell level are difficult to be verified. Recent studies therefore showed that phenotypic diversity is caused by specific mechanisms without the requirement for environmental inputs (Blake *et al.*, 2006, Suel *et al.*, 2007). Furthermore, the mutation rate that is necessary to explain phenotypic heterogeneity between single cells would be higher than any known mutation system indicating that the origin of phenotypic variations is not a reflection of genetic differences (Ackermann, 2015). Instead, the cause of the observed heterologous cell behaviour is the stochastic distribution of small molecules (Huh and Paulsson, 2011) and associated stochastic gene expression (Elowitz *et al.*, 2002), leading to the expression of different subsets of genes. Such stochastic effects on gene expression levels have been identified as the main source for phenotypic heterogeneity and this is supported by a recent study showing that population variability was reduced in *B. subtilis* exhibiting larger cell sizes (Suel *et al.*, 2007). Other mechanisms contributing to phenotypic cell-to-cell variability include cellular aging (Levy *et al.*, 2012), periodic oscillations in the cell cycle (Lenz and Sogaard-Andersen, 2011) and modulating effects of quorum sensing by diffusible molecules (Anetzberger *et al.*, 2012). Taken together, these molecular mechanisms that lead to phenotypic heterogeneity support the survival of cell populations in fluctuating environments.

1.2.3 Noise as a driver for phenotypic heterogeneity

Gene expression noise contributes to phenotypic heterogeneity between genetically identical cells. Depending on its intensity, noise can significantly impair biological rhythms such as circadian clocks, influencing the process of cell development (Barkai and Leibler, 2000). Low noise can be induced by systems with high abundances of regulator proteins leading to small phenotypic variations, whereas low amounts of regulator proteins result in high noise with strong fluctuations in the phenotype. The impact of different noise levels was demonstrated in an expression system that revealed high transcriptional rates leading to low noise as compared with low transcriptional rates showing strong phenotypic variations (Ozbudak *et al.*, 2002). Another example is the bidirectional DNA replication in bacteria that reflects a high processing velocity due to abundant deoxynucleotides (dNTPs), leading to low variations and little noise (Mirkin and Mirkin, 2005). Consequently, during stationary phase when nutrients become limited, the exceeding demand for dNTPs but finite availability leads to increased noise levels.

Cellular processes include two types of noise: intrinsic and extrinsic noise. Intrinsic noise reflects the stochasticity of gene expression within single cells that is caused by molecular scale events such as competition for low abundance transcription factors, leading to different gene regulation outcomes. In contrast, extrinsic noise represents the fluctuating amounts of extracellular components, leading to variable gene expression levels at the intercellular level. Both types of noise were studied by expressing two genes encoding different fluorescent proteins under the control of identical promoters (Elowitz *et al.*, 2002). This expression system revealed a correlation between both fluorescent proteins in the absence of intrinsic noise with the production of identical amounts of both fluorescent reporter proteins (Figure 1.2A). The introduction of extrinsic noise to these cells nevertheless led to heterogeneity. In the presence of intrinsic noise (Figure 1.2B), the expression of both genes becomes uncoordinated, also leading to phenotypic variations.

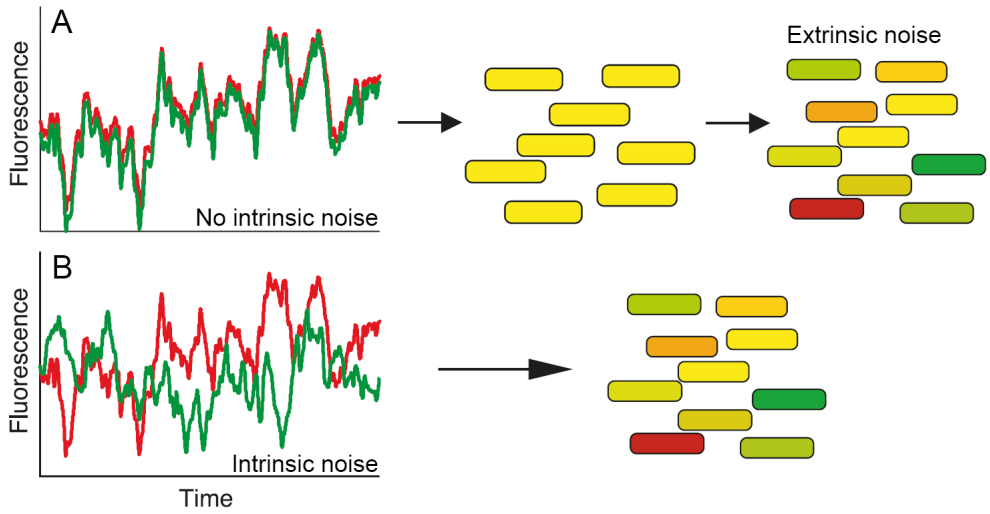


Figure 1.2: Influence of intrinsic and extrinsic noise on gene expression. (A) Both fluorescent proteins (red & green) correlate in their expression over time in the absence of intrinsic noise. When extrinsic noise was introduced, the amounts of the fluorescent proteins varied between the cells. (B) When intrinsic noise was introduced, gene expression becomes uncoordinated (modified from Elowitz *et al.* (2002)).

1.2.4 Heterogeneity and bistability in B. subtilis

Heterogeneity in *B. subtilis* was studied originally by analysing sporulation, a differentiation process induced in response to nutrient limitation but involving a minority of cells in the population (Freese, 1972, Chung *et al.*, 1994). Subsequent studies have involved a more detailed analysis of sporulation and other pathways (Piggot and Hilbert, 2004, Veening *et al.*, 2005) to investigate heterogeneous cell fates (Lopez *et al.*, 2009a) including the development of competence for the uptake of extracellular DNA (Smits *et al.*, 2005, Gamba *et al.*, 2015), the production of flagella to seek fresh nutrients (Kearns and Losick, 2005, Cozy and Kearns, 2010, Diethmaier *et al.*, 2011) and formation of an extracellular matrix for biofilm development (Chai *et al.*, 2008, Guttenplan and Kearns, 2013, Diethmaier *et al.*, 2011). These heterogeneous pathways reveal populations of cells that are either switched on or off for a specific developmental pathway (Figure 1.3). These distinct subpopulations are caused by stochastic fluctuations, a state determined as bistability (Figure 1.4B-C) (Veening *et al.*, 2008b). The importance of individual cells expressing different sets of genes is the increase in fitness associated with population heterogeneity; a so-called bet-hedging strategy under rapidly changing conditions (Veening *et al.*, 2008b, Veening *et al.*, 2008c). In contrast, homogenous populations are only adapted for a limited range of environmental conditions.

The phosphorylation states of three master regulators, Spo0A, ComA and DegU, were identified as the major triggers for the development of the various cell types observed within populations of *B. subtilis* (Veening *et al.*, 2008a, Lopez *et al.*, 2009a, Lopez and Kolter, 2010). Low levels of phosphorylated Spo0A induce so-called cannibalism and matrix production cascades (Gonzalez-Pastor *et al.*, 2003, Hamon and Lazazzera, 2001), whereas high levels of Spo0A~P lead to the initiation of sporulation and concomitant termination of the vegetative cell cycle in a subpopulation of cells (Fujita *et al.*, 2005). Phosphorylation of ComA induces surfactin production and the development of competence in a subpopulation of cells (Nakano *et al.*, 1991, Hamoen *et al.*, 1995). Phosphorylation of DegU drives the production of proteases, creating so-called 'miner' cells, which facilitate the degradation of extracellular proteins and peptides (Verhamme *et al.*, 2007, Veening *et al.*, 2008a). In contrast, the decision to differentiate into either sessile or motile cells is not based on the phosphorylation states of these master regulators, but on the alternative sigma factor, SigD, since cells only develop flagella once SigD reaches a threshold level (Cozy and Kearns, 2010). These bistable and heterogeneous processes are regulated by complex interconnected systems that include positive and double-negative feedback loops that determine the distinct and different cell fates.

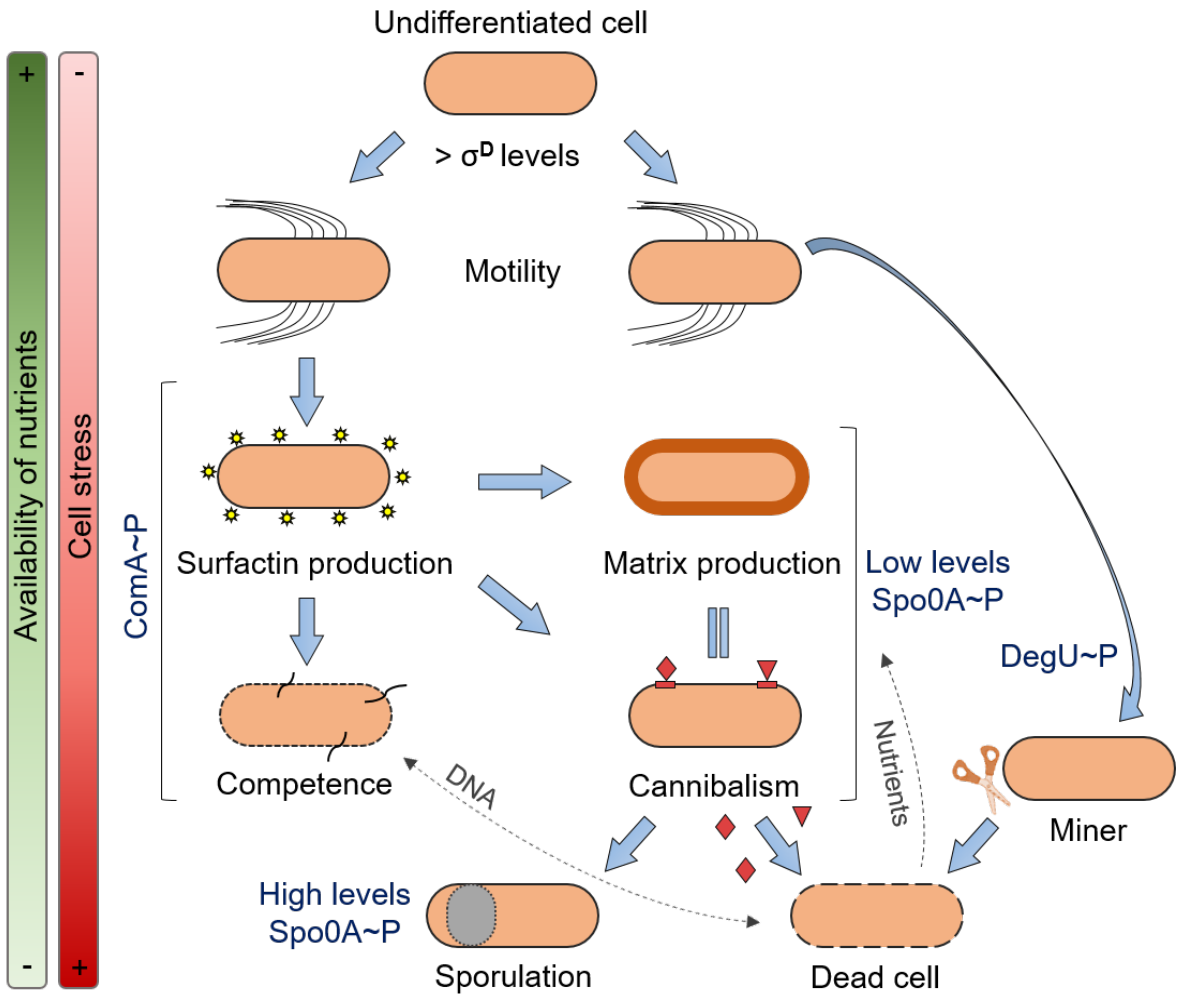


Figure 1.3: Overview of the differentiation process that determine the formation of distinct cell types of *B. subtilis*. Motility is induced by increased SigD levels in a fraction of cells. Subsequent cell fates are triggered by phosphorylation of the three master regulators ComA~P, DegU~P and low and high levels of Spo0A~P. Blue arrows indicate the process of differentiation. Equal sign reflects induction of both differentiations in the same subpopulation. Decrease of available nutrients and increased cell stress during the progress of differentiation are shown by gradients.

1.2.5 Regulation of motility

B. subtilis has the capability to develop different forms of motility that allow an active and passive movement across liquid and solid media to seek fresh nutrients. The active locomotion includes the feature of swimming by rotating the peritrichous flagella in liquid media (Nishihara and Freese, 1975), whereas the swarming ability reflects a social form of surface movement across solid media (Kearns and Losick, 2003). The passive locomotion occurs through sliding on surfaces with reduced tension due to surfactin production and is flagella-independent (Kinsinger *et al.*, 2003). Both swarming and sliding only occur in undomesticated strains since frameshift mutations disrupt the synthesis of surfactin and SwrA in laboratory strains such as 168 (Patrick and Kearns, 2009).

The approximately 26 flagella in *B. subtilis* (Guttenplan *et al.*, 2013) are generally structured into three main parts: basal body, hook and filament. The basal body is fixed in the membrane with a basal ring that results from the polymerisation of several FliF subunits. It contains the secretion machinery to transport essential units for flagellar synthesis at the extracellular side (Mukherjee and Kearns, 2014) and embeds the energy driving components required to drive rotation. The stator elements, MotA and MotB, direct the conversion of power, generated by proton and sodium motive forces across the cytoplasmic membrane (Zhou *et al.*, 1998, Ito *et al.*, 2004, Chan *et al.*, 2014), into the mechanical rotation of the flagella. The hook, consisting of a hook structure (FlgE) and its cap (FlgD), is attached to the basal body and alters the angle of the filament rotation. The filament reflects the propeller of the flagellum that is connected to the hook by junction proteins (FlgK, FlgL). Its long helical structure is composed of eleven flagellin (Hag) monomers, which polymerise into protofilaments by the interaction of their N-terminal and C-terminal domains, and its cap (FliD) (Mukherjee and Kearns, 2014). Prior to assembly, the flagellin monomers are secreted through the hollow tube of the flagellum in an unfolded state by a type III secretion mechanism (Evans *et al.*, 2013, Stern and Berg, 2013).

The direction of rotation of flagella is regulated by chemotaxis. This system facilitates the sensing of chemical gradients and modulating the direction by rotation to facilitate either straight swimming towards attractants or tumbling to induce a random walk that is biased away from repellents. The counter clockwise rotation propels the bacterium towards attractants while a clockwise rotation leading to an uncoordinated flagella rotation and cell tumbling. This system is based on a complex chemotaxis pathway that uses three adaptation systems: reversible methyl-accepting chemotaxis proteins (MCP) and associated histidine kinase CheA, the CheC–CheD–CheY~P adaptation system and the CheV adaptation system (Rao *et al.*, 2008). The probability of direct swimming towards attractants results from the activation of the CheA (CheA~P), via MCPs, and subsequent phosphotransferase to its cognate response regulator, CheY~P. CheA~P in turn directs the flagellum to rotate in a counter clockwise direction. This two-component system thus controls the active movement of *B. subtilis* by mediating the received signals from the chemoreceptor to the flagellar motor.

Most of the motility and chemotaxis genes are encoded within the large 27 kbp *fla*/*che* operon (Kearns and Losick, 2005, Mukherjee and Kearns, 2014). The operon is regulated by the σ^A-dependent housekeeping promoter, $P_{fla/che}$, that drives the

expression of the genes for basal body and hook formation. Once this development is completed, the activity of the alternative sigma factor σ^D is increased through antagonism of the anti-σ^D factor FlgM (Caramori *et al.*, 1996). This leads to the expression of genes that are linked to filament assembly and rotation and eventually to production of motile cells (Figure 1.3). Since motility shows a bimodal development (Chapter 3), revealing distinct subpopulations of motile and sessile cells, the latter tending to form chains due to impaired autolysin activity (Chen *et al.*, 2009), researchers investigated the role of σ^D in more detail (Cozy and Kearns, 2010). This showed that the proportion of motile cells could be increased by moving the *sigD* gene more proximal to the σ^A-dependent $P_{fla/che}$ promoter, due to a higher read-through of $P_{fla/che}$ and two other potential σ^D-dependent promoters (Cozy and Kearns, 2010).

1.2.6 Surfactin as trigger for the development of biofilms and cannibalism

The formation of biofilms leads to heterogeneous communities that contain several cell types, including motile cells, matrix producers and a subpopulation of sporulating cells (Vlamakis *et al.*, 2008). Such differentiation processes require specific communication on an intercellular level that is based on signalling molecules. For the induction of matrix producer cells, two molecules, the pheromone ComX and the lipopeptide surfactin, are essential prerequisites for biofilm development in *B. subtilis*. The production of surfactin allows cells to communicate via quorum-sensing molecules. This process is strongly regulated by another quorum-sensing pathway initiated by the secreted ComX that is expressed in most cells, however, only a small proportion of cells respond to this signal. The membrane kinase ComP senses ComX and becomes activated, leading to the phosphorylation of the master regulator ComA (ComA~P) (Figure 1.4A) (Magnuson *et al.*, 1994). The phosphorylated form of this response regulator induces the sequential transcription of the long *srfAA-AD* operon resulting in the production of surfactin (Nakano *et al.*, 1991). The mechanisms for the translocation of both ComX and surfactin into the extracellular environment remain unknown. Surfactin itself leads to pore formation in the membranes of other cells in the population (Sheppard *et al.*, 1991), triggering differentiation into matrix producers (Chai *et al.*, 2008). Interestingly, while cells producing and secreting surfactin cannot become biofilm initiating cells, matrix producing cells are immune to the ComX signal preventing them from producing surfactin; this classifies the quorum-sensing surfactin molecule as paracrine signal (Lopez *et al.*, 2009c).

An extracellular matrix, containing proteins, nucleic acids and particularly exopolysaccharides (EPS), is responsible for holding cells together by forming bundled chains (Hall-Stoodley *et al.*, 2004). For the production of the extracellular matrix, both the *epsA-O* operon and *tasA* are essential to coordinate the development process (Branda *et al.*, 2001, Branda *et al.*, 2006). These genes are negatively regulated by the SinR repressor and its antagonist, SinI, relieves repression by preventing SinR from binding to cognate binding sites associated with the biofilm-dependent genes (Kearns *et al.*, 2005). SinR is further modulated by the activity of SlrR, which forms a heteromeric complex. However, SinR also controls *slrR* at the transcriptional level by modulating its activity to repress genes involved in motility and autolysis (Chai *et al.*, 2010). This bilateral influence and formation of several feedback loops results in the generation of stochastic differences leading to a bimodal gene expression (Newman and Lewis, 2013). Since matrix production is initiated by surfactin, the activation of the membrane kinase KinC leads to the phosphorylation of Spo0A~P, triggering the induction of SinI. For the induction of the matrix production and development of cannibalism, low levels of Spo0A~P are required, resulting in both differentiation processes occurring in the same subpopulation (Lopez *et al.*, 2009b) (Figure 1.3).

Cannibalism is a differentiation process that is triggered in a proportion of the cell population under nutrient-limiting conditions, aimed at avoiding or delaying the energy intensive sporulation process. Cannibalism is induced in cells with low levels of Spo0A~P, leading to the production and secretion of the two killing factors SdpC, a bacteriocins induced by repressing AbrB, and the antibiotic-like peptide SkfA (sporulation killing factor A). The expression of these toxins is accompanied by the simultaneous expression of their respective immunity genes, SdpI and SkfEF (Claverys and Havarstein, 2007) (Figure 1.3). SdpI is an integral membrane protein protecting the cells by binding SdpC, leading to an increased production of SdpI. The ABC transporter SkfEF is responsible for the export of SkfA (Gonzalez-Pastor *et al.*, 2003). Since the killer factor-induced cells are immune to the cognate toxins that benefit from the nutrients released by their sensitive sibling cells. This led Losick and colleagues to refer to this process as a form of cannibalism (Gonzalez-Pastor *et al.*, 2003, Ellermeier *et al.*, 2006).

1.2.7 Competence development

The quorum-sensing molecule surfactin is not only involved in the induction of biofilm formation and differentiation into cannibalistic cells but its phosphorylated master regulator ComA~P also initiates natural competence development in a different subpopulation of cells other than the matrix producers and cannibals (Nakano *et al.*, 1991, Kunst *et al.*, 1994) (Figure 1.3). Interestingly, competent cells benefit from the actions of cannibals since the toxin-dependent cell lysis of cohabiting cells releases DNA that might be taken up and integrated into the chromosome to facilitate adaptation to unfavourable conditions. Additionally, this feature has contributed to the success of *B. subtilis* as a model and industrial organism since chromosome modifications or the introduction of plasmids can be undertaken with ease.

Competence can be induced in the same subpopulation as that producing surfactin, indicating that their regulatory cascades are similarly triggered (Nakano *et al.*, 1991). The developmental pathway of competence begins with the sensing of ComX by ComP, leading to the phosphorylation of the master regulator ComA (ComA~P) (Figure 1.4A) (Magnuson *et al.*, 1994). In addition to initiating the expression of surfactin, ComA~P leads to the production of ComS which, in turn, activates the transcription factor ComK (Nakano *et al.*, 1991). ComK is the master regulator of competence and its auto-stimulation is responsible for a positive-feedback loop that leads to the bistable competence system (Figure 1.4B) (Smits *et al.*, 2005). The auto-stimulation of ComK is controlled *via* transcriptional regulation at its promoter P_{comK}, the generation of noise generated by proteolytic degradation pathways (Hamoen *et al.*, 2003, Suel *et al.*, 2007) and the reduction of noise *via* Kre (*ykyB*) that was recently identified modulating the bimodal expression of ComK by affecting its mRNA stability (Gamba *et al.*, 2015).

On one hand ComK reduces the expression of Rok which represses the promoter of *comK* (Maamar and Dubnau, 2005) and therefore induces its own expression while, on the other hand, ComK overproduction leads to the down-regulation of *comS,* reflecting a putative negative-feedback loop since ComS is required for the activation of ComK (Figure 1.4A) (Suel *et al.*, 2006). Interestingly, Spo0A~P is involved in the positive regulation of the *comK* expression either directly by antagonizing the repressor Rok (Mirouze *et al.*, 2012) or indirectly by repressing both AbrB (low levels of Spo0A~P) and SinR (high levels of Spo0A~P) which repress the *rok* expression (Hahn *et al.*, 1995, Fujita *et al.*, 2005, Schultz *et al.*, 2009).

Another repressor for *comK* expression is the nutritional repressor CodY (Dubnau and Losick, 2006)

Once sufficient amounts of ComK are produced, the genes encoding the structural components, *comC*, *comGA* and *comP,* are expressed, leading to the formation of the DNA uptake machinery responsible for the import of single-stranded DNA molecules of up to 20 kb that, if containing homology with the host's chromosome can be integrated via homologous recombination (Chen and Dubnau, 2004) (Chapter 3).

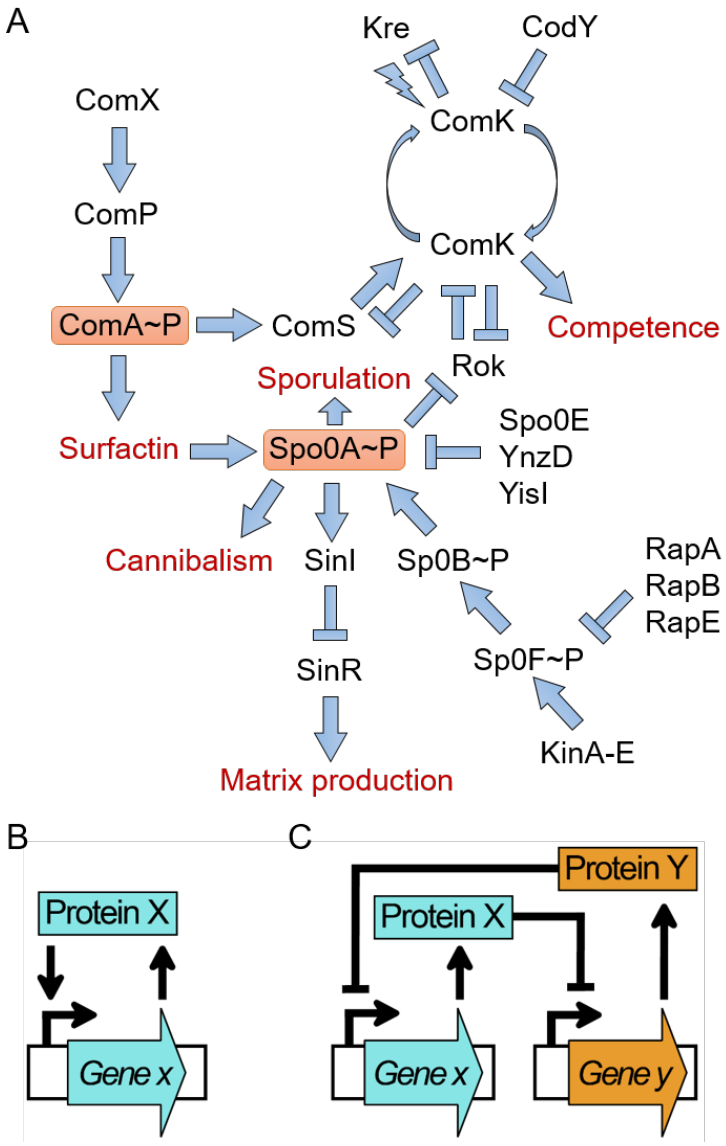


Figure 1.4: Overview of the regulatory processes that determine the formation of distinct cell types of *B. subtilis*. (A) Surfactin, matrix production, cannibalism, competence, sporulation and their master regulators ComA~P and Spo0A~P (box). Bistability requires feedback loops maintaining a specific state. (B) The positive transcriptional autoregulatory feedback loop is denoted by gene *x* that drives its own expression (e.g. *comK*). (C) In the double-negative feedback loop the expression of gene *x* is repressed by protein Y, whereas the transcription of gene *y* is negatively regulated by protein X, once it is sufficiently produced inducing high levels of protein X. Symbols indicate the activation (arrow), inhibition (T-bar) or modulation (flash) of processes.

1.2.8 Initiation of sporulation

The capability to initiate sporulation and develop resistant endospores allows *B. subtilis* and related bacteria to persist for long periods of time under extreme environmental conditions. The underlying differentiation process is controlled by the master regulator Spo0A, once it is activated by phosphorylation in response to specific environmental and intracellular signals including high cell density and nutrient limitation (Figure 1.3). Sporulation reflects another bistable regulatory system (Lopez *et al.*, 2009a, Veening *et al.*, 2008c). Whilst low levels of Spo0A~P trigger the differentiation of cells into matrix producers and cannibals, high levels of Spo0A~P are mandatory for sporulation initiation, reflecting the cessation of the vegetative cycle (Chapter 3). Spo0A~P induces the expression of its own transcript, directly influencing the expression of 121 genes and indirectly the expression of further ~400 genes (Molle *et al.*, 2003, Piggot and Hilbert, 2004). Interestingly, Spo0A~P is not responsible for establishing the bistable state during spore development, but rather the activity of the phosphorelay responsible for activating Spo0A (Figure 1.4A), since it was shown that cells with pre-activated Spo0A~P lack such bistable behaviour (Veening *et al.*, 2008c, Ireton *et al.*, 1993). The phosphorelay itself consists of Spo0A, Spo0B and Spo0F. Feeding signals into this phosphotransferase pathway are five histidine sensor kinases, namely KinA, KinB, KinC, KinD and KinE. The major kinases, KinA and KinB, sense different signals including the decreased cell respiration caused by limited nutrient availability (Eswaramoorthy *et al.*, 2010, Kolodkin-Gal *et al.*, 2013), whereas the other kinases (KinC-E) are not pivotal for sporulation. Once Spo0F becomes phosphorylated, the phosphate is transferred to Spo0B which then triggers the activation of the master regulator Spo0A~P. However, the Spo0A levels are regulated through a positive feedback loop reflected by the stimulation of its own expression. Furthermore, the activation of Spo0A is modulated by negative-feedback loops indirectly by the phosphatase activities of RapA, RapB and RapE on Spo0F~P and directly on Spo0A~P *via* the Spo0E, YnzD and YisI phosphatases that remove the phosphate group (Ohlsen *et al.*, 1994, Perego, 2001). These negative-feedback loops contribute to the generation of noise in the sporulation initiation pathway and therefore to the bistability of the system (Veening *et al.*, 2005).

1.2.9 Heterogeneity and its impact for the industry

The ability of *B. subtilis* to differentiate into subpopulations with distinct phenotypes is a bet-hedging strategy that, despite being an energetically unfavourable process, provides this bacterium with a selective advantage in its natural habitat, the soil. In this thesis we consider the impact of this differentiation process on industrial processes. Heterogeneous gene expression is of particular relevance in an industrial setting since it has the potential to reduce process efficiency by diverting substrate and energy consumption away from product formation (Figure 1.5) (Piersma *et al.*, 2013, Veening *et al.*, 2008b). Consequently, a major aim of the work in this thesis is to develop tools that quantify heterogeneity under industrially relevant conditions (Chapter 3). Once the degree of heterogeneity under industry-like conditions is understood, it is possible to consider the development of strategies that reduce the proportion of low-producer cells with the aim of increasing heterologous protein production (Chapter 4).

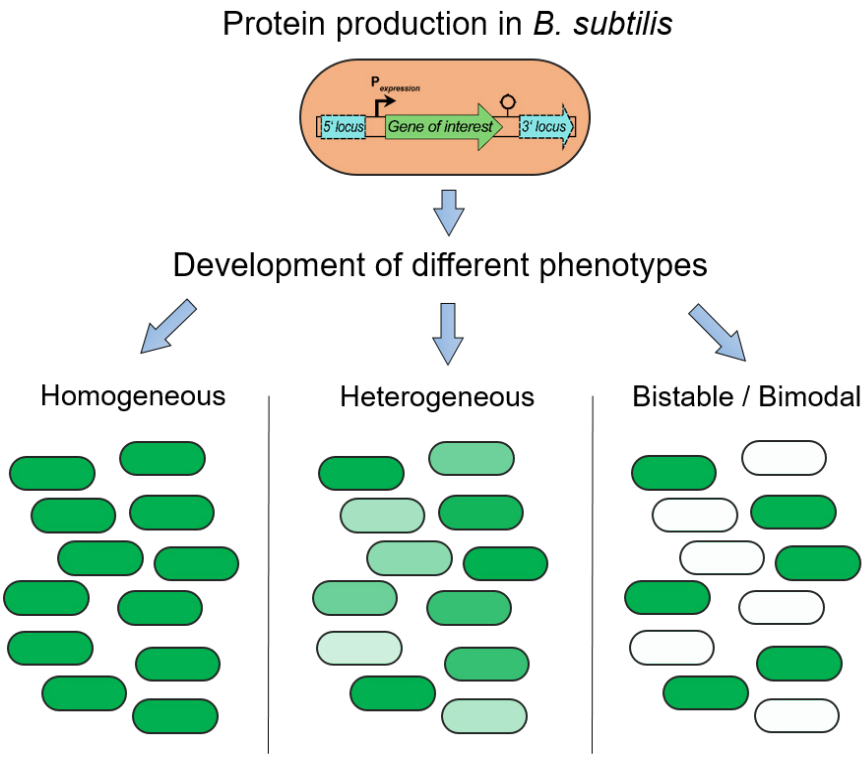


Figure 1.5: Overview of potential phenotypes during protein production in *B. subtilis*. Differentiation processes lead to a variety of cell types producing the protein of interest in a homogeneous, heterogeneous or bistable / bimodal fashion.

1.3 Protein production in *B. subtilis*

1.3.1 Protein synthesis

Once the DNA is transcribed into mRNA during transcription, the energy intensive translation process is initiated in bacteria provided that the mRNA is sufficiently stable. This process consists of several steps including the initiation, elongation and termination phase. Initiation is mostly the rate limiting step (de Smit and van Duin, 1994). The initiation phase starts by pairing the 3' end of the 16S rRNA with the Shine-Dalgarno (SD) sequence, that is located at the ribosome binding site (RBS) of the mRNA and the precise placement of the initiator tRNA (fMet-tRNA) that is responsible for the correct selection of the initiation codon (AUG) at the mRNA supported by various initiation factors and rRNA (Sprengart and Porter, 1997). In an industrial context, by altering the RBS consensus' sequence (AGGAGG), the production of proteins can be improved but specific interactions between the 5' untranslated region (UTR) and the gene of choice reduce translation initiation. Subsequently, the 70S complex is formed by the assembly of the 50S and 30S ribosome subunits and the elongation is initiated, which can be stalled by mRNA secondary structures and rare codons. Stalling during elongation can ultimately lead to mRNA degradation (Rocha *et al.*, 1999). Once the polypeptide chain is growing, starting with the fMet-tRNA at the P site, amino acids are transferred by tRNAs, whereby the entry of new tRNAs to the A site is facilitated by the elongation factor Tu. Elongation proceeds until the ribosome reaches a stop codon (UGA, UAA or UAG). These stop codons are recognised by release factors (RF1, RF2) resulting in the termination of the protein synthesis and subsequently secretory proteins are transported to the membrane where they get translocated. Importantly, signal sequences play a crucial role since they are recognised by the translocation machineries to transport unfolded proteins from the cytoplasm into the cell envelope where the mature protein folds into its native conformation after it is released from the translocase and the signal peptide cleaved off (Tjalsma *et al.*, 2000, Harwood and Cranenburgh, 2008).

1.3.2 Secretion stress during protein production

B. subtilis and close relatives *B. licheniformis* and *B. amyloliquefaciens* are widely used for the high-level protein production in biotechnologically relevant processes (Harwood and Cranenburgh, 2008). The ability of *Bacillus* species to secrete high amounts of proteins directly into the culture broth, by-passing the outer membrane

barrier of Gram-negative bacteria such as *E. coli,* reduces downstream processing costs significantly, an important factor in large scale applications such as industrial enzyme manufacturing. *E. coli* is still the preferred host for high-added-value, low to medium scale production when downstream processing costs are less significant. The high secretion capacity is one reason for *B. subtilis* being the preferred workhorse for industrial scale protein production (Harwood, 1992). However, despite a good understanding of the elements of its protein secretion pathway, generally, the translocation of heterologous proteins into the culture medium, as compared with native proteins, has proven to be more challenging than expected. Understanding the specific bottlenecks involved has been the subject of very active research over the last two decades and which has implicated protein folding, quality control mechanisms and transport across the cell wall as significant bottlenecks to high-level production (Bolhuis *et al.*, 1999, Schumann, 2007). As a result, the production and secretion of certain target proteins can induce high levels of secretion stress at the interface between cell membrane and cell wall, as detected by the membrane-bound CssS sensor kinase. Once activated, CssS transmits a signal to its cognate response regulator, CssR (Hyyrylainen *et al.*, 2001, Darmon *et al.*, 2002, Westers *et al.*, 2006, Noone *et al.*, 2012, Sarvas *et al.*, 2004). Activation of the CssR regulon leads to the induction of quality control proteases HtrA and HtrB, that degrade misfolded and aggregated secretory proteins at the membrane/wall interface (Noone *et al.*, 2000, Noone *et al.*, 2012). The expression of these quality control proteases in response to secretion stress has been used to discriminate between cells in the population actively producing heterologous proteins (producer cells) from non-producers (Westers *et al.*, 2004, Ploss *et al.*, 2016, Trip *et al.*, 2011). Consequently, the production of target proteins strongly depends on the regulation of these quality control proteases indicating the importance of these proteins for industrial productivity (Chapter 4). Currently, strategies to re-engineer *B. subtilis* to overcome this bottleneck and improve the protein production showed extremely variable results. In addition, the two additional proteases, WprA and HtrC, also belong to the quality control system but are induced in response to cell wall rather than secretion stress (Margot and Karamata, 1996, Krishnappa *et al.*, 2014, Bisicchia *et al.*, 2010), but to date there are no studies describing WprA and HtrC being responsible for significant degradation of industrial enzymes. Importantly, the expression of *htrC* has been only verified in a recent study showing the RNA of *htrC* in a mutant lacking all extracellular proteases and the quality control proteases *htrA*, *htrB* and *wprA* (Pohl *et al.*, 2013).

1.3.3 Influence of extracytoplasmic proteases

The passage of proteins from the cytoplasm, across the cytoplasmic membrane and thick cell wall and into the culture medium is a strongly regulated process. The medium itself is a complex environment in which proteins and peptides are degraded to serve as important sources of nutrient substrates (Msadek, 1999). In industrial fermentations, carbon and nitrogen sources are usually fed to the culture at rates that limit growth but maximise productivity. In commercial media, the amino acids and oligopeptides that are the preferred nitrogen sources for *B. subtilis* are derived from hydrolysed proteins, and this required the activity of proteases naturally secreted by this bacterium, namely AprE, Bpr, Epr, Vpr, Mpr, NprB, NprE and WprA. Whilst AprE and NprE are responsible for ~95% of the protease activity in the extracellular environment (Wu *et al.*, 1991), the remaining proteases may still represent limiting factors for the production of specific target proteins. While native *B. subtilis* secretory proteins are generally resistant to degradation by these proteinases due to adaptations during evolution, whereby recent studies indicated degradation events of cognate proteins (Krishnappa *et al.*, 2013, Krishnappa *et al.*, 2014), the same is not necessarily true for target heterologous proteins. This represents a conundrum, the naturally produced proteases are required for amino acid and oligopeptide production, but potentially also degrade the very proteins the industrial process is designed to produce indicating the importance to fine-tune medium composition and protease deletions (Chapter 4) (Pohl and Harwood, 2010, Pohl *et al.*, 2013, Krishnappa *et al.*, 2013, Krishnappa *et al.*, 2014).

1.3.4 B. subtilis in an industrial context

B. subtilis has been a model organism for more than six decades. Its genetic tractability and capacity to secrete proteins has contributed to its wide-spread use for the production of native industrial enzymes (e.g. proteases and amylases) and an increasing interest by the biotechnology industry is using this bacterium for the production of recombinant proteins to improve yields at lower costs (Harwood, 1992, van Dijl and Hecker, 2013). To facilitate the high-throughput construction of industrial strains producing commercially valuable heterologous proteins (Lam *et al.*, 1998, Jensen *et al.*, 2000, Pohl and Harwood, 2010, Zobel *et al.*, 2015), advanced genetic engineering strategies such as Golden Gate (Engler *et al.*, 2008, Engler *et al.*, 2009) and Gibson Assembly® (Gibson *et al.*, 2009) have been adapted to this organism. Additionally, the recently developed CRISPR/Cas9 system is being increasingly

targeted for genome editing in *B. subtilis,* potentially replacing more traditional genome editing technologies (Zhang *et al.*, 2016, Altenbuchner, 2016, Peters *et al.*, 2016, Westbrook *et al.*, 2016). These technologies are used to improve productivity by, for example, developing efficient expression cassettes designed to optimise the expression of specific genes of interest. However, identifying promoters that function predictably in all contexts is not trivial, as outlined in Chapter 5 (Harwood, 1992, Mijakovic *et al.*, 2005, Brockmeier *et al.*, 2006).

1.3.5 Industrially relevant model enzymes: XynA and AmyM

The demand for industrial enzymes with specific catalytic and physical properties (e.g. temperature stability, pH tolerance) is increasing steadily. As the range of potential industrial enzymes increases, so too does the need to optimise and adapt production hosts to improve process efficiency. For any particular protein, the process bottlenecks are likely to reflect a unique combination of host and product characteristics, such as interaction with pathway components, rate of protein folding, sensitivity to host proteases etc. XynA and AmyM are industrially relevant enzymes that are produced for applications in the baking process. XynA, a native *B. subtilis* xylanase (Bernier *et al.*, 1983, Wolf *et al.*, 1995, Pollet *et al.*, 2009), belonging to the endo-class of enzymes that is able to hydrolyse the β-1,4-xylose linkages of xylans, is a main component of the polysaccharide hemicellulose (Ruller *et al.*, 2006). The hydrolysis of hemicellulose causes a reduction in the polymerisation of the flour substrate and increases the dough's capacity to bind water. In addition, it degrades lignocellulose and this is relevant for several chemical and fuel applications (Kulkarni *et al.*, 1999). An important characteristic of XynA from *Bacillus* species is its absence of cellulase activity, an important requirement for its application in the paper industry (Torronen and Rouvinen, 1997). In contrast to XynA, AmyM is an α-amylase obtained originally from *Geobacillus stearothermophilus* (Dauter *et al.*, 1999). It is of similar industrial importance since the main cause of bread staling is the re-crystallization of starch, and enzymes such as AmyM slows this process down by hydrolysing the α-1,4 glycosidic bonds in amylopectin, generating maltose (Tanyildizi *et al.*, 2005). In addition to its anti-staling activity, AmyM also improves crumb softness, slice ability and elasticity (Diderichsen and Christiansen, 1988, Goesaert *et al.*, 2009). AmyM is a stable protein and can be produced at high concentrations in *B. subtilis* (Grupta *et al.*, 2003).

1.4 Synthetic engineering of *B. subtilis*

The interest in systematically re-designing and engineering biological elements has grown dramatically in recent years, driven by engineering principles. Referred to as synthetic biology, this developing technology provides technological solutions for assembling and integrating standardised parts into synthetic circuits that provide novel solutions for current challenges in the biotechnological, diagnostic, therapeutic and global health sectors (Rooke, 2013, Lienert *et al.*, 2014, Slomovic *et al.*, 2015). *B. subtilis* has been a popular chassis for the development of this technology, principally because of its well-established genetics and biochemistry, and industrial importance. In the context of this thesis, the technological advances associated with synthetic biology holds out the prospect of improved production systems for industrial enzymes whose commercial exploitation is currently limited by costly and time-consuming classical strain development strategies and lower than expected yields.

To synthetically engineer *B. subtilis* several strategies can be applied (Figure 1.6) including the synthesis of expression modules to create large libraries achieving strong and homogeneous gene expression levels (Chapter 5). The development of such modular systems requires modern and automated high-throughput devices. By systematically varying standardised parts including regions for transcription and translation initiation and their high-throughput synthesis, such approaches reflect an accelerated evolution by testing more than thousand different combinations for the expression of a gene of choice at once. Other approaches include the deployment of synthetic translocation units for an improved secretion efficiency to artificially manipulate the secretion process increasing its efficiency during all growth phases with available cellular resources. The engineering of the quality control system and extracellular protein stability requires a quantitative multi-omics analysis that reveals insights about secretion stress responses and protease targets in the growth medium supported by complex statistical and mathematical analyses (van Dijl and Hecker, 2013). Nevertheless, the systematic approach, deleting large areas of the *B. subtilis* genome that leads to improved production strains lacking unnecessary elements (Morimoto *et al.*, 2008, Manabe *et al.*, 2011), is a research direction with the potential of creating a *Bacillus* factory. While such engineering approaches are promising in respect of achieving higher production levels of specific proteins, exaggerated claims are being made about its potential emphasising the requirement of including the chassis into the initial design to better predict the response of hosts.

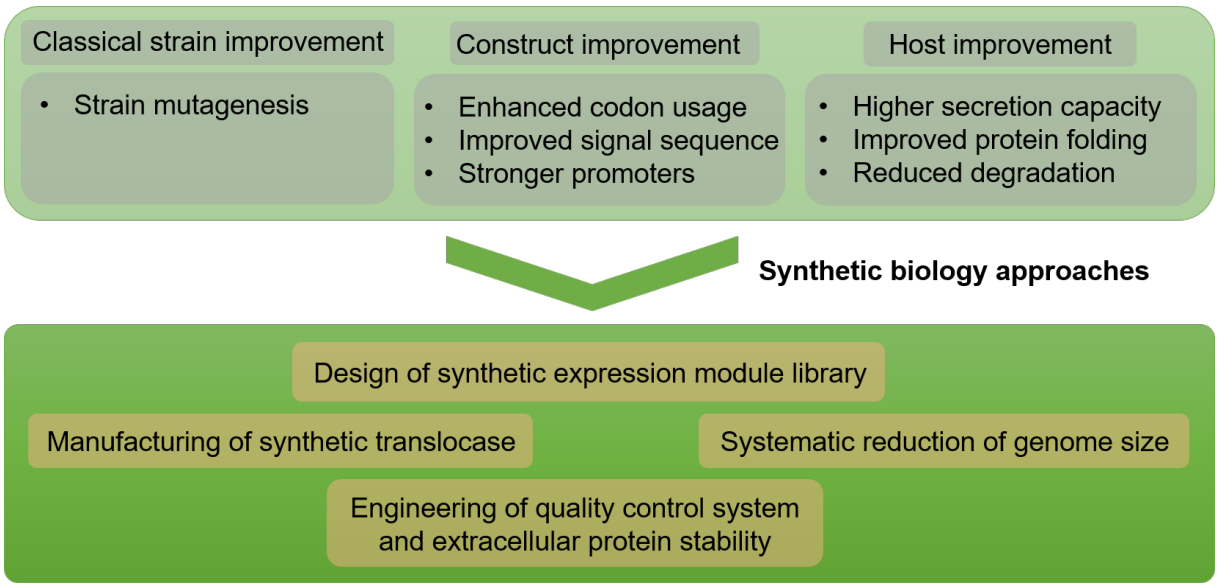


Figure 1.6: Classical strategies advanced by synthetic biology approaches for next-generation applications in *B. subtilis* to improve the protein production in industrially relevant settings.

1.5 Chromosome replication in *B. subtilis*

Bacterial chromosomes vary enormously in size, from around 0.5 Mbp to 15 Mbp, their size reflecting the environment they inhabit (Konstantinidis and Tiedje, 2004, Dini-Andreote *et al.*, 2012). Bacteria generally have one or a very small number of chromosomes that are usually covalently closed structures, although some species (e.g. *Streptomyces sp.*) have linear chromosomes (Casjens, 1998). The physical structures of the bacterial chromosome show much less structural complexity than those of eukaryotic organisms, which are invariably linear and highly structured on a backbone of chromatin. *B. subtilis* contains a single circular chromosome of 4.2 Mbp, and 75% of the coding sequences are located on the leading strand (Kunst *et al.*, 1997, Rocha, 2004, Barbe *et al.*, 2009). As is the case with all studied bacteria, chromosomal replication is initiated by DnaA at the origin of replication (*oriC*), and the two replication forks thus migrate bidirectionally to the replication terminus situated approximately 179 degrees from the origin. Because the time taken to complete one round of replication is at least 40 minutes at 37°C, with cell division taking place up to 20 minutes later, in rapidly-growing cells with mean generation times of 20 min, multiple rounds of replication are initiated on the same chromosome. As a result, genes close to *oriC* are present in higher copy numbers than genes close to the terminus. This has led us to address the question of the importance of gene location and orientation with respect to the expression of genes involved in targeted protein production (Chapter 6).

Each replication fork consists of a complex of two DNA polymerases III enzymes with proofreading and exonuclease activities and a helicase required to unwind the DNA. The duplex DNA at the fork includes a leading strand, whose synthesis is basically processive, and a lagging strand that is synthesised discontinuously using RNA primers and subsequently joined by a succession of repair processes. *B. subtilis* encodes two DNA polymerases α-subunits; PolC which has an exonuclease domain and is involved in leading strand synthesis and DnaE, associated with the lagging strand synthesis. RNA primase generates the 10 to 12 nucleotide primers required for initiating the synthesis of the discontinuous Okazaki fragments by DNA polymerase. The RNA primers are subsequently removed, replaced with deoxynucleotides and ligated by the combined activities of nucleases, polymerases and DNA ligase. Once the replication forks arrive at the terminus (*ter*) region, a terminator protein binds to the *ter* sites and therefore creates a protein-DNA complex leading to the arrest of both replication forks (Bussiere and Bastia, 1999). XerC and XerD subsequently catalyse a site-specific recombination event at so-called *dif* sites to resolve the chromosome dimer to monomers (Blakely *et al*., 1993).

1.6 Aims

The major aim of this thesis was to investigate the relationship of cellular heterogeneity and enzyme production and to analyse whether a heterogeneous or bistable regulation of developmental and physiological pathways in *B. subtilis* leads to a reduction in protein production of XynA or AmyM. To examine this hypothesis, and to identify potentially unproductive subpopulations, a range of technologies (e.g. microbioreactors, fluorescence microscopy and flow cytometry) were used to analyse cultures at both the population and single-cell level. This required the combinatorial usage of fluorescent reporters to reflect the physiological state and enzyme expression level of single cells.

An additional aim was to investigate the next-generation DNA synthesis approach for generating a synthetic expression module library that was combined with a fluorescent reporter. This facilitated a high-throughput screen for expression modules with both high expression levels and diminished heterogeneous activities for incorporation into modular parts for future academic and industrial applications.

Chapter 2

Materials and Methods

2.1 Enzymes and chemicals

All enzymes for DNA modification and restriction digestion were purchased from New England Biolabs (NEB®) and Thermo Fisher Scientific. Chemicals and components for media were purchased from BDH Laboratory Supplies, Melford, Roth and Sigma-Aldrich.

2.2 Strains and Media

The strains and plasmids used for the work described in this thesis are listed in Table 2.1 and were verified by restriction digestion and subsequent sequencing. To analyse and align sequences, Clone Manager (Sci-Ed Software) and SnapGene Viewer (SnapGene) were employed.

For *B. subtilis*, the wild type strain 168 was used unless otherwise mentioned. For the cultivation of bacterial strains Luria-Bertani medium (LB; 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl), 2×PY medium (1.28% (w/v) peptone, 0.8% (w/v) yeast extract, 0.4% (w/v) NaCl), an adjusted rich medium (2% (w/v) yeast extract, 2.5% (v/v) Glucose, 0.75% (v/v) KH_2PO_4, 0.5% (v/v) sodium glutamate, 0.15% (v/v) $MgCl_2$, 0.1% (v/v) $CaCl_2$ x $2H_2O$, 0.005% (v/v) $MnSO_4$ x H_2O, 0.0025% (v/v) $FeCl_3$ x $6H_2O$) from the patent of ROCHE® (Publication number: EP0405370 A1), the chemically-defined medium (1.4% (v/v) K_2HPO_4, 0.6% (v/v) KH_2PO_4, 0.2% (v/v) K-glutamate, 0.12% (v/v) $Na_3C_6H_5O_7$ × $2H_2O$, 0.02% (v/v) $MgSO_4$ [$7H_2O$], 2% (v/v) glucose, 0.1% (w/v) yeast extract, 0.02% (v/v) casamino acids, 0.00022% (v/v) $(NH_4)_5Fe(C_6H_4O_7)_2$)) and Spizizen-plus medium (1.4% (v/v) K_2HPO_4, 0.6% (v/v) KH_2PO_4, 0.2% (v/v) K-glutamate, 0.12% (v/v) $Na_3C_6H_5O_7$ × $2H_2O$, 0.02% (v/v) $MgSO_4$ × $7H_2O$, 0.5% (v/v) glucose, 0.002% (v/v) tryptophan, 0.02% (v/v) casamino acids, 0.00022% (v/v) $(NH_4)_5Fe(C_6H_4O_7)_2$)) were used. Where necessary, agar plates or liquid media were supplemented with a final concentration of the following antibiotics: spectinomycin (100 µg/ml), erythromycin (2 µg/ml), neomycin/kanamycin (12.5 µg/ml), chloramphenicol (10 µg/ml) and tetracycline (20 µg/ml) for *B. subtilis*. For *E.*

coli ampicillin (100 µg/ml) was added. Nutrient agar with a final concentration of 1% (w/v) (Oxoid), 2×PY agar (1.28% (w/v) peptone, 0.8% (w/v) yeast extract, 0.4% (w/v) NaCl, 1.5% (w/v) agar) or LB agar (1.5% (w/v)) were used for growth on solid media. Optical density measurements were performed using the Libra S22 spectrophotometer (Biochrom). For obtaining appropriate OD_{600} values ≤1.0 cell suspensions were diluted with MilliQ water containing 0.85% (w/v) NaCl.

Strain or plasmid	Genotype	Construction or Reference
B. subtilis		
168	Wild type, *trpC*	Kunst *et al.* (1997)
bSS353	W168 *trpC*; *lacA*::*tet; aprE*::*lacI, cat*	S. Syvertsson (unpub.)
BCS002	W168 (BSBI) *aprE*::P_{spoIIE}-*sfGFP*, *spec*	168 × pCS05 (StarGate®)
BCS003	W168 *aprE*::P_{hag}-*sfGFP*, *spec*	168 × pCS05 (StarGate®)
BCS005	W168 *aprE*::P_{gapB}-*sfGFP*, *spec*	168 × pCS05 (StarGate®)
BCS006	W168 *aprE*::P_{gapA}-*sfGFP*, *spec*	168 × pCS05 (StarGate®)
BCS009	W168 *aprE*::P_{srfAA}-*sfGFP*, *spec*	168 × pCS05 (StarGate®)
BCS028	W168 *aprE*::P_{comGA}-*sfGFP*, *spec*	168 × pCS05 (StarGate®)
BCS031	W168 *aprE*::P_{htrA}-*sfGFP*, *spec*	168 × pCS05 (StarGate®)
BCS024	W168 *nprE*::P_{15}-*xynA*, *ery*	168 × pCS31 (StarGate®)
BCS040	W168 *nprE*::pCS31, *ery*	168 × pCS31
BCS056	W168 *nprE*::P_{15}-*xynA*, *ery*; *aprE*::P_{spoIIE}-*sfGFP*, *spec*	BCS024 × BCS002
BCS057	W168 *nprE*::P_{15}-*xynA*, *ery*; *aprE*::P_{hag}-*sfGFP*, *spec*	BCS024 × BCS003
BCS058	W168 *nprE*::P_{15}-*xynA*, *ery*; *aprE*::P_{gapB}-*sfGFP*, *spec*	BCS024 × BCS005
BCS059	W168 *nprE*::P_{15}-*xynA*, *ery*; *aprE*::P_{gapA}-*sfGFP*, *spec*	BCS024 × BCS006
BCS060	W168 *nprE*::P_{15}-*xynA*, *ery*; *aprE*::P_{srfAA}-*sfGFP*, *spec*	BCS024 × BCS009
BCS061	W168 *nprE*::P_{15}-*xynA*, *ery*; *aprE*::P_{comGA}-*sfGFP*, *spec*	BCS024 × BCS028
BCS062	W168 *nprE*::P_{15}-*xynA*, *ery*; *aprE*::P_{htrA}-*sfGFP*, *spec*	BCS024 × BCS031
BCS095	W168 *aprE*::P_{ctc}-*sfGFP*, *spec*	168 × pCS05 (StarGate®)
BCS101	W168 *nprE*::P_{15}-*xynA*, *ery*; *aprE*::P_{ctc}-*sfGFP*, *spec*	BCS024 × BCS095
BCS109	W168 *nprE*::pCS31, *ery*; *aprE*::P_{spoIIE}-*sfGFP*, *spec*	BCS040 × BCS002
BCS110	W168 *nprE*::pCS31, *ery*; *aprE*::P_{hag}-*sfGFP*, *spec*	BCS040 × BCS003
BCS111	W168 *nprE*::pCS31, *ery*; *aprE*::P_{gapB}-*sfGFP*, *spec*	BCS040 × BCS005
BCS112	W168 *nprE*::pCS31, *ery*; *aprE*::P_{gapA}-*sfGFP*, *spec*	BCS040 × BCS006
BCS113	W168 *nprE*::pCS31, *ery*; *aprE*::P_{srfAA}-*sfGFP*, *spec*	BCS040 × BCS009
BCS114	W168 *nprE*::pCS31, *ery*; *aprE*::P_{comGA}-*sfGFP*, *spec*	BCS040 × BCS028
BCS115	W168 *nprE*::pCS31, *ery*; *aprE*::P_{htrA}-*sfGFP*, *spec*	BCS040 × BCS031
BCS116	W168 *nprE*::pCS31, *ery*; *aprE*::P_{ctc}-*sfGFP*, *spec*	BCS040 × BCS095
BCS117	W168 *nprE*::pCS31, *ery*; *aprE*::P_{spoIIE}-*sfGFP*, *spec*; pCS74 (P_{amyQ}), *neo*	BCS040 × BCS002 × pCS74
BCS118	W168 *nprE*::pCS31, *ery*; *aprE*::P_{hag}-*sfGFP*, *spec*; pCS74 (P_{amyQ}), *neo*	BCS040 × BCS003 × pCS74
BCS119	W168 *nprE*::pCS31, *ery*; *aprE*::P_{gapB}-*sfGFP*, *spec*; pCS74 (P_{amyQ}), *neo*	BCS040 × BCS005 × pCS74
BCS120	W168 *nprE*::pCS31, *ery*; *aprE*::P_{gapA}-*sfGFP*, *spec*; pCS74 (P_{amyQ}), *neo*	BCS040 × BCS006 × pCS74
BCS121	W168 *nprE*::pCS31, *ery*; *aprE*::P_{srfAA}-*sfGFP*, *spec*; pCS74 (P_{amyQ}), *neo*	BCS040 × BCS009 × pCS74
BCS122	W168 *nprE*::pCS31, *ery*; *aprE*::P_{comGA}-*sfGFP*, *spec*; pCS74 (P_{amyQ}), *neo*	BCS040 × BCS028 × pCS74
BCS123	W168 *nprE*::pCS31, *ery*; *aprE*::P_{htrA}-*sfGFP*, *spec*; pCS74 (P_{amyQ}), *neo*	BCS040 × BCS031 × pCS74
BCS124	W168 *nprE*::pCS31, *ery*; *aprE*::P_{ctc}-*sfGFP*, *spec*; pCS74 (P_{amyQ}), *neo*	BCS040 × BCS095 × pCS74
BCS125	W168 *nprE*::pCS31, *ery*; *aprE*::P_{spoIIE}-*sfGFP*, *spec*; pCS73 (P_{amyQ}-*amyM*), *neo*	BCS040 × BCS002 × pCS73
BCS126	W168 *nprE*::pCS31, *ery*; *aprE*::P_{hag}-*sfGFP*, *spec*; pCS73 (P_{amyQ}-*amyM*), *neo*	BCS040 × BCS003 × pCS73
BCS127	W168 *nprE*::pCS31, *ery*; *aprE*::P_{gapB}-*sfGFP*, *spec*; pCS73 (P_{amyQ}-*amyM*), *neo*	BCS040 × BCS005 × pCS73
BCS128	W168 *nprE*::pCS31, *ery*; *aprE*::P_{gapA}-*sfGFP*, *spec*; pCS73 (P_{amyQ}-*amyM*), *neo*	BCS040 × BCS006 × pCS73
BCS129	W168 *nprE*::pCS31, *ery*; *aprE*::P_{srfAA}-*sfGFP*, *spec*; pCS73 (P_{amyQ}-*amyM*), *neo*	BCS040 × BCS009 × pCS73
BCS130	W168 *nprE*::pCS31, *ery*; *aprE*::P_{comGA}-*sfGFP*, *spec*; pCS73 (P_{amyQ}-*amyM*), *neo*	BCS040 × BCS028 × pCS73
BCS131	W168 *nprE*::pCS31, *ery*; *aprE*::P_{htrA}-*sfGFP*, *spec*; pCS73 (P_{amyQ}-*amyM*), *neo*	BCS040 × BCS031 × pCS73
BCS132	W168 *nprE*::pCS31, *ery*; *aprE*::P_{ctc}-*sfGFP*, *spec*; pCS73 (P_{amyQ}-*amyM*), *neo*	BCS040 × BCS095 × pCS73
BCS133	W168 pCS73 (P_{amyQ}-*amyM*), *neo*	168 × pCS73
BCS134	W168 pCS74 (P_{amyQ}), *neo*	168 × pCS74
BCS195	W168 pCS58 (P_{amyQ}-*xynA*), *neo*	168 × pCS58
BCS205	W168 *rrnO-23S*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125
BCS206	W168 *ctc*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125
BCS207	W168 *mfd*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125
BCS208	W168 *yckB*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125
BCS209	W168 *ydgG*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125
BCS210	W168 *speA*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125
BCS211	W168 *yojF*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125
BCS212	W168 *bshB2*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125
BCS213	W168 *brxA*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125
BCS214	W168 *yqhS*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125
BCS215	W168 *yqeT*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125
BCS216	W168 *recJ*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125

Strain or plasmid	Genotype	Construction or Reference
BCS217	W168 *yvmB*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125
BCS218	W168 *spsC*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125
BCS219	W168 *nupG*::pSS125 (P_{spac}-*lacZ*; P_{veg}-*sfGFP*), *kan; lacA*::*tet; aprE*::*lacI, cat*	bSS353 × pSS125
BCS234	W168 *amyE*::P_{xyl}-*htrA*, *spec*	168 × pCS61
BCS237	W168 pCS73 (P_{amyQ}-*amyM*), *neo*; *amyE*::P_{xyl}-*htrA*, *spec*	BCS133 × BCS234
BCS262	BRB07 *lacA*::P_{htrA}-*sfGFP*, *spec*	(Pohl *et al.*), Rita Cruz (unpub.)
BCS263	BRB09 *lacA*::P_{htrA}-*sfGFP*, *spec*	(Pohl *et al.*), Rita Cruz (unpub.)
BCS264	BRB10 *lacA*::P_{htrA}-*sfGFP*, *spec*	(Pohl *et al.*), Rita Cruz (unpub.)
BCS265	BRB13 *lacA*::P_{htrA}-*sfGFP*, *spec*	(Pohl *et al.*), Rita Cruz (unpub.)
BCS274	W168 pCS67 (P_{amyQ}-*amyM*|*mCherry*), *neo*	168 × pCS67 (StarGate®)
BCS282	BRB07 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS67 (P_{amyQ}-*amyM*|*mCherry*), *neo*	BCS262 × pCS67
BCS283	BRB08 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS67 (P_{amyQ}-*amyM*|*mCherry*), *neo*	BCS306 × pCS67
BCS284	BRB09 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS67 (P_{amyQ}-*amyM*|*mCherry*), *neo*	BCS263 × pCS67
BCS285	BRB10 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS67 (P_{amyQ}-*amyM*|*mCherry*), *neo*	BCS264 × pCS67
BCS286	BRB11 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS67 (P_{amyQ}-*amyM*|*mCherry*), *neo*	BCS307 × pCS67
BCS287	BRB12 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS67 (P_{amyQ}-*amyM*|*mCherry*), *neo*	BCS308 × pCS67
BCS288	BRB13 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS67 (P_{amyQ}-*amyM*|*mCherry*), *neo*	BCS265 × pCS67
BCS290	W168 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS67 (P_{amyQ}-*amyM*|*mCherry*), *neo*	BCS311 × pCS67
BCS291	W168 *amyE*::P_{xyl}-*htrB*, *spec*	168 × pCS68
BCS292	W168 *amyE*::P_{xyl}-*htrC*, *spec*	168 × pCS69
BCS293	W168 *amyE*::P_{xyl}-*wprA*, *spec*	168 × pCS70
BCS294	W168 pCS58 (P_{amyQ}-*xynA*), *neo*; *amyE*::P_{xyl}-*htrA*, *spec*	BCS195 × BCS234
BCS295	W168 pCS58 (P_{amyQ}-*xynA*), *neo*; *amyE*::P_{xyl}-*htrB*, *spec*	BCS195 × BCS291
BCS296	W168 pCS58 (P_{amyQ}-*xynA*), *neo*; *amyE*::P_{xyl}-*htrC*, *spec*	BCS195 × BCS292
BCS297	W168 pCS58 (P_{amyQ}-*xynA*), *neo*; *amyE*::P_{xyl}-*wprA*, *spec*	BCS195 × BCS293
BCS298	W168 pCS73 (P_{amyQ}-*amyM*), *neo*; *amyE*::P_{xyl}-*htrB*, *spec*	BCS133 × BCS291
BCS299	W168 pCS73 (P_{amyQ}-*amyM*), *neo*; *amyE*::P_{xyl}-*htrC*, *spec*	BCS133 × BCS292
BCS300	W168 pCS73 (P_{amyQ}-*amyM*), *neo*; *amyE*::P_{xyl}-*wprA*, *spec*	BCS133 × BCS293
BCS306	BRB08 *lacA*::P_{htrA}-*sfGFP*, *spec*	(Pohl *et al.*), Rita Cruz (unpub.)
BCS307	BRB11 *lacA*::P_{htrA}-*sfGFP*, *spec*	(Pohl *et al.*), Rita Cruz (unpub.)
BCS308	BRB12 *lacA*::P_{htrA}-*sfGFP*, *spec*	(Pohl *et al.*), Rita Cruz (unpub.)
BCS310	W168 pCS76 (P_{amyQ}-*mCherry*), *neo*	168 × pCS76
BCS311	W168 *lacA*::P_{htrA}-*sfGFP*, *spec*	(Pohl *et al.*), Rita Cruz (unpub.)
BCS315	BRB07 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS76 (P_{amyQ}-*mCherry*), *neo*	BCS262 × pCS76
BCS316	BRB08 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS76 (P_{amyQ}-*mCherry*), *neo*	BCS306 × pCS76
BCS317	BRB09 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS76 (P_{amyQ}-*mCherry*), *neo*	BCS263 × pCS76
BCS318	BRB10 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS76 (P_{amyQ}-*mCherry*), *neo*	BCS264 × pCS76
BCS319	BRB11 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS76 (P_{amyQ}-*mCherry*), *neo*	BCS307 × pCS76
BCS320	BRB12 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS76 (P_{amyQ}-*mCherry*), *neo*	BCS308 × pCS76
BCS321	BRB13 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS76 (P_{amyQ}-*mCherry*), *neo*	BCS265 × pCS76
BCS323	W168 *lacA*::P_{htrA}-*sfGFP*, *spec*; pCS76 (P_{amyQ}-*mCherry*), *neo*	BCS311 × pCS76
BCS333	W168 *amyE*::P_{15}-*xynA*|*mCherry*, *kan*	168 × pCS38 (StarGate®)
BCS335	W168 *nprE*::P_{infC}-*mCherry*, *kan*	168 × pCS78 (StarGate®)
BCS340	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant1-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS341	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant2-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS342	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant3-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS343	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant4-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS344	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant5-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS346	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant7-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS352	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant12-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS355	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant15-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS357	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant17-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS358	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant18-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS360	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant20-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS361	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant21-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS363	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant23-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS364	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant24-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS365	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant25-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS366	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant26-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS367	W168 *amyE*::P_{15}-*amyM*|*mCherry*, *kan*	168 × pCS38 (StarGate®)
BCS369	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*empty-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS370	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*veg-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS371	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::P_{15}-*sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS372	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*design-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
BCS373	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*empty -xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS374	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*veg-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS375	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::P_{15}-*xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS376	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*design-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS377	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant1-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS378	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant2-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS379	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant3-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS380	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant4-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS381	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant5-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS382	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant7-xynA*, *spec*	BCS335 × pCS80 (StarGate®)

Strain or plasmid	Genotype	Construction or Reference
BCS384	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant12-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS386	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant15-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS388	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant17-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS389	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant18-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS390	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant20-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS391	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant21-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS393	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant23-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS394	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant24-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS395	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant25-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS396	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*variant26-xynA*, *spec*	BCS335 × pCS80 (StarGate®)
BCS397	W168 *lacA*::P_{htrA}-*sfGFP*, *spec*; *amyE*::P_{15}-*amyM*|*mCherry*, *kan*	BCS311 × BCS367
BCS398	W168 *lacA*::P_{htrA}-*sfGFP*, *spec*; *amyE*::P_{15}-*xynA*|*mCherry*, *kan*	BCS311 × BCS333
BCS400	W168 *nprE*::P_{infC}-*mCherry*, *kan*; *amyE*::*library-sfGFP*, *spec*	BCS335 × pCS66 (StarGate®)
E. coli		
TOP10	*mcrA*, Δ(*mrr-hsd*RMS-*mcrBC*), Φ80 Δ(*lacZ*)M15, Δ*lacX74*, *deoR*, *recA1*, *araD139*, Δ(*ara-leu*)7697, *galU*, *galK*, *rpsL(SmR)*, *endA1*, *nupG*	Casadaban and Cohen (1980)
NEB® 10-beta	Φ80 Δ(*lacZ*)M15, *recA1*, *endA1*, Δ(*mrr-hsdRMS-mcrBC*), *mcrA* Δ(*mrr-hsdRMS-mcrBC*), *fhuA*	NEB®
Plasmid		
pAPNC213	*aprE*::(MCS P_{spac}, *lacI*, *spec*); *bla*	Morimoto *et al.* (2002)
pSS125	*ery bla himar9* Tn*YLB*-1::*(kan* t_{ezrA}/t_{braB} P_{veg}-*sfGFP* P_{spac}-*lacZ* t_{ywoG}/t_{ywoF}*)*	Sauer *et al.* (2016)
pSS148	*amyE*::(*spec,* P_{xyl}-*mCherry*); *bla*	S. Syvertsson, unpub.
pPG49	*amyE*::(*spec*, *sfGFP*); *bla*	Pamela Gamba, unpub.
pCS01	*aprE*::(MCS, *spec*); *bla*	pAPNC213
pCS02	*aprE*::(MCS, *sfGFP*, *spec*); *bla*	pCS01
pCS03	*aprE*::(MCS, *sfGFP*, *spec*); *bla*	pCS02
pCS04	*aprE*::(*kan*$_{(E)}$, *sfGFP*, *spec*); *bla*	pCS03
pCS05	*aprE*::(*kan*$_{(E)}$, *sfGFP*, *spec*); *bla*	pCS04
pCS10	*nprE*::(*kan*$_{(E)}$, *sfGFP*, *spec*); *bla*	pCS05
pCS13	*aprE*::(*kan*$_{(E)}$, *mCherry*, *ery*); *bla*	pCS05
pCS31	*nprE*::(*kan*$_{(E)}$, *ery*); *bla*	pCS10
pCS38	*amyE*::(*cat*$_{(E)}$, *kan*); *bla*	pCS75
pCS41	*nprE*::(*kan*$_{(E)}$, *kan*); *bla*	pCS31
pCS47	*amyE*::(*spec,* P_{xyl}-*divIVA*); *bla*	pSS148
pCS58	P_{amyQ}-*xynA*, *reppUB*, *neo*, *bleo*	pCS73
pCS61	*amyE*::(*spec,* P_{xyl}-*htrA*); *bla*	pCS47
pCS66	*amyE*::(*sfGFP*, *spec*); *bla*	pCS75
pCS67	P_{amyQ}-*amyM*-*mCherry*, *reppUB*, *neo*, *bleo*	pCS73
pCS68	*amyE*::(*spec,* P_{xyl}-*htrB*); *bla*	pCS47
pCS69	*amyE*::(*spec,* P_{xyl}-*htrC*); *bla*	pCS47
pCS70	*amyE*::(*spec,* P_{xyl}-*wprA*); *bla*	pCS47
pCS73	P_{amyQ}-*amyM*, *reppUB*, *neo*, *bleo*	DSM
pCS74	P_{amyQ}, *bla*, *cat*$_{(d)}$, *reppUB*, *neo*, *bleo*	DSM
pCS75	*amyE*::(*cat*$_{(E)}$, *spec*); *bla*	DSM
pCS76	P_{amyQ}-*mCherry*, *reppUB*, *neo*, *bleo*	pCS67
pCS78	*nprE*::(P_{infC}-mCherry, *kan*); *bla*	pCS41
pCS80	*amyE*::(*cat*$_{(E)}$, *xynA*, *spec*); *bla*	pCS75

Table 2.1: Bacterial strains and plasmids. The list describes the origin of the strains and plasmids used in this study. Under construction the recipient is shown first and the donor second. Antibiotic resistances: ampicillin (*bla*), chloramphenicol (*cat*), erythromycin (*ery*), kanamycin (*kan*), neomycin (*neo*), spectinomycin (*spec*), tetracycline (*tet*). Abbreviations: *E. coli (E)*, defective *(d)*.

2.3 Transformation of bacteria

2.3.1 B. subtilis

For the transformation of *B. subtilis* the cells were inoculated into Spizizen-plus medium at 37°C, with shaking at 165 rpm at an Innova 2300 platform shaker (New Brunswick Scientific). For the period of starvation, competence medium (starvation medium) lacking ammonium ferric citrate and tryptophan was used.

The required strain was grown in LB medium with shaking at 37°C overnight. The following morning the OD_{600} of the culture was measured and diluted into 10 ml of competence medium to achieve a starting OD_{600} of 0.1. The cells were then grown in a 100 ml shake flask until an OD_{600} between 0.4 and 0.6 where the culture was diluted 1:1 with pre-warmed starvation medium and incubation continued for around 1.5 h with shaking at 37°C. Cells became maximal competent ~1 h later and were centrifuged at 3,000 ×*g* in a benchtop centrifuge for 10 min at room temperature. Ninety per cent of the exhausted growth medium was removed and the remaining medium was used to resuspend the pellet, concentrating the cells ~10-fold. The DNA (50-150 ng for chDNA, 100-2,000 ng for plasmid, 500-4,000 ng for PCR fragment) was mixed with at least 100 µl of competent cells in a 2 ml Eppendorf tube with shaking for 1 h at 37°C. As a negative control, 10 µl of MilliQ water was added to one tube. Following incubation, 100 µl of the transformation mixture was plated on to selective nutrient or 2×PY agar plates containing the appropriate antibiotic(s). After the liquid was completely absorbed, the plates were incubated at 37°C overnight.

Competent *B. subtilis* cells were frozen after the step of concentration, aliquoted in to Eppendorf tubes, flash frozen using liquid nitrogen and stored at -80°C. The frozen competent cells were useable for several weeks. For an instant transformation, DNA was added to a tube of frozen competent cells, and incubated with shaking at 37°C for 1 h, following the procedure described above.

To extract chDNA from a donor strain for transfer to a recipient strain, the donor strain was grown overnight in either LB or 2×PY and between 0.5-1.0 ml of culture was centrifuged at 16,000 ×*g* to pellet the cells. The resulting pellets were resuspended in 100 µl saline (0.85% (w/v) NaCl) containing 10 µl RNase A solution (R6148, Sigma-Aldrich) and 2 µl lysozyme (50 mg/ml). The suspensions were incubated at 37°C for 5 min and then 200 µl cell lysis solution (25 mM EDTA, 2 % (w/v) SDS) was added and mixed thoroughly. After the addition of 115 µl of protein precipitation solution (10 M NH_4Ac) the tubes were vortexed for 5 s with a subsequent centrifugation step at 16,000 ×*g* for 10 min at 4°C. To precipitate the DNA, 375 µl of supernatant was mixed with 375 µl iso-propanol before the tubes were inverted and centrifuged at 16,000 ×*g* for 10 min. The pellets were washed with 100 µl of 70% (v/v) ethanol and briefly centrifuged for 1 min. The liquid was carefully removed by pipetting and re- centrifuged for 1 min to remove the remaining trace of ethanol. The pellet was resuspended in 100 µl of MilliQ water and the quality of the DNA was determined with a NanoDrop spectrophotometer (Thermo Fisher Scientific).

For the verification of successful integrations into the chromosome the inserts were amplified using primers that annealed outside the integration locus. If of the expected size, the PCR products were sent for sequencing to validate the sequences.

2.3.2 E. coli

For the transformation of *E. coli*, competent cells were either purchased from NEB® or prepared to achieve chemically competent cells following the method of Dagert and Ehrlich (Dagert and Ehrlich, 1979). Four millilitres of an overnight culture of Top10 strain (Casadaban and Cohen, 1980) were inoculated in to 400 ml LB medium and incubated with shaking at 37°C until the culture reached an OD_{600} ~0.3. From this point onwards all steps were performed on ice or 4°C while using the centrifuge. The culture was split into 8 Falcon tubes and incubated for 10 min on ice before the cells were harvested by centrifuging in a benchtop centrifuge at 3,000 ×*g* for 8 min. The supernatant was discarded and 20 ml of 100 mM $CaCl_2$ added slowly to resuspend the pellet which was then incubated for 30 min on ice. The cell suspensions were combined in to four Falcon tubes and then centrifuged for 8 min at 3,000 ×*g*. The four cell pellets were resuspended with 2 ml of 100 mM $CaCl_2$ containing 15% (v/v) glycerol and incubated for 15 min on ice. The final step involved aliquoting into 100 µl lots in reaction tubes, snap-freezing in liquid N_2 and storing at -80°C. The frozen competent cells were useable for several months.

For both, prepared and purchased competent cells, a slightly adjusted protocol from NEB® was used to perform transformation reactions with high efficiency. The competent cells were thawed on ice until the ice crystals dissolved. Following the addition of approximately 100 ng of plasmid DNA, the tubes were carefully flicked to mix the DNA with the competent cells. After incubation on ice for 30 min, the tubes were incubated at 42°C for 30 s to heat shock the cells. The cells were chilled on ice for a couple of minutes before 400 µl of LB was added to each tube. The transformation mixture was then incubated at 37°C with shaking for 1 h and centrifuged at full speed on a bench centrifuge for 1 min. The pellet was resuspended in 80 µl of medium, plated on selective LB agar plates and incubated at 37°C overnight.

2.4 Molecular methods

2.4.1 Polymerase chain reaction (PCR)

For the amplification of genes, 0.02 U/µl of Q5 or Phusion polymerase (NEB®) were used for reactions that required proofreading enzymes *e.g.* Gibson assembly® or to prior sequencing. To detect transformants containing the correct insert DNA in plasmid constructs (colony PCR) the MyTaq™ red mix (Bioline) was used.

For each reaction (50 µl), oligonucleotide primers were added at a final concentration of 0.5 µM, dNTPs at 200 µM and 1× reaction buffer when using Q5 or Phusion polymerase. Template DNA was added at the following concentrations: chDNA between 1 ng and 1 µg, plasmid DNA between 1 pg and 10 ng. When using the MyTaq™ red mix for colony PCR the final concentration of the primers was 0.4 µM. When performing the PCR, the T3000 thermocycler (Thistle Scientific®) or T100™ thermal cycler (Bio-Rad) was used. Oligonucleotides were supplied by Integrated DNA Technologies (Leuven, Belgium) and were designed to have an annealing temperature (t_m) close to 58.4°C Howley *et al.* (1979). The t_m and extension time was adjusted based on the oligonucleotides and product size, respectively. The PCR protocol for the Q5 or Phusion polymerase as follows: Initial denaturation at 98°C for 30 s. Then the cycle was repeated 30 times followed with a denaturation at 98°C for 10 s, annealing of the primers for 20 s (55°C – 65°C) and elongation of the product at 72°C (~25 s/kb). After the 30 cycles, a final elongation was performed for 5 min at 72°C. For the MyTaq™ red mix, an initial denaturation at 95°C for 120 s was completed before 30 cycles of a denaturation at 95°C for 15 s, annealing for 15 s (55°C-65°C), extension at 72°C (~30 s/kb) and a final extension at 72°C for 5 min.

2.4.2 PCR product processing

Verification of amplified PCR products was done by agarose gel electrophoresis. Agarose gels (1% (w/v)) contained 0.015% (v/v) SafeView (cat. no. G108, Applied Biological Materials Inc.) or 0.0125% (v/v) Nancy-520 (01494, Sigma-Aldrich) and 1× TAE-buffer (40 mM Tris-acetate, 1 mM EDTA). The gels were electrophoresed at a constant 90 V (~90 mA) for 45 min. A 1kb DNA ladder from Promega® or the 1kb plus ladder from Invitrogen™ was used as molecular size markers. The DNA was visualised with ultraviolet light using a gel documentation system (GelDoc, Bio-Rad). Amplified PCR products were purified using the QiaQuick PCR purification kit (cat. no. 28106, Qiagen®), the NucleoSpin® Gel and PCR clean-up-kit from Macherey-Nagel (REF 740609.250) or, to extract DNA bands from gel, the GeneJET gel extraction kit (cat. no. K0691, Thermo Fisher Scientific) following the supplied protocols.

2.4.3 Traditional cloning

For traditional cloning both insert and vector DNA were digested with 10 U of each restriction enzyme (NEB®) for 3 hours following the manufacturer's guidelines. After the complete digest the samples were incubated at high temperatures (65°C or 80°C) to inactivate the enzymes and purified with the appropriate kit (Section 2.4.2). Subsequently, vector and insert DNA were added in the ratio of 1:3. For ligation, 7.5 U of T4 ligase (Thermo Fisher Scientific) and ligase buffer were added and incubated either for 1 h at room temperature or overnight at 5°C. For the transformation of *E. coli,* half of the ligation mixture was used (Section 2.3.2). To purify plasmid DNA from *E. coli,* the NucleoSpin® Plasmid kit (REF 740588.250, Machery-Nagel) or the QIAprep spin miniprep kit (cat. no. 27106, Qiagen®) was used following the manufacturer's protocol. The resulting plasmid DNA was analysed by restriction digestion, targeting both the vector and expected insert DNA. After digestion, an agarose gel electrophoresis was performed to visualise the DNA and positive clones showing an insert of the correct size were sent for sequencing to verify the DNA.

Name	Sequence (5’ to 3’)	Restriction site	Description
oCS001	GCGCCTCGAGAGGTCGACTCTAGAGGATCC	*Xho*I	pAPNC213 (fwd)
oCS002	AATTACTAGTGTGAGCGCAACGCAAGCTTC	*Spe*I	pAPNC213 (rev)
oCS003	AATTTCTAGAAAACCGGTAGATCTCACGTG	*Xba*I	*sfGFP* (fwd)
oCS004	AATTACTAGTATTTGTAGAGCTCATCCATG	*Spe*I	*sfGFP* (rev)
oCS005	GTAAGCGAATTTGCCACCATGGGTACCCTGCAGATGAG		*comGA* (rev)
oCS006	CTTTTCTATTGAATCCAATTTTCTAGAGGATCCCCGGG		*comGA* (fwd)
oCS007	AATTGAATTCCGGCGGAGACGCTTGCAGTGGGCTTACATGG	*EcoR*I, *BsmB*I	Introducing kanamycin (fwd)
oCS008	AATTGGATCCCCAAGGAGACGTCAGAAGAACTCGTCAAGAAG	*BamH*I, *BsmB*I	Introducing kanamycin (rev)
oCS009	GGAGACGTCAGAAGAACTCG		Plasmid backbone pCS05 (fwd)
oCS010	TTGGATTCAATAGAAAAGGTAAGCG		Plasmid backbone pCS05 (rev)
oCS014	AATTCGTCTCCCGGCTGTTGACGAAAGAGACAAAC	*BsmB*I	P_{spoIIE} - StarGate® (fwd)
oCS016	AATTCGTCTCCCCAACATTTCCTCTCCTCCTGTTATATTCGTTGCCTGTC	*BsmB*I	P_{spoIIE} - StarGate® (rev)
oCS017	AATTCGTCTCCCGGCATCGCGGAAAATAAACGAAG	*BsmB*I	P_{hag} - StarGate® (fwd)
oCS019	AATTCGTCTCCCCAACATTTCCTCTCCTCCGAATATGTTGTTAAGGCACG	*BsmB*I	P_{hag} - StarGate® (rev)
oCS020	AATTCGTCTCCCGGCTGATGATCGATTCTTTCGGG	*BsmB*I	P_{gapA} - StarGate® (fwd)
oCS022	AATTCGTCTCCCCAACATTTCCTCTCCTCCTTGCTGGACATTATATGTCC	*BsmB*I	P_{gapA} - StarGate® (rev)
oCS023	AATTTCTAGAACTCTTGAATGAACCACCAC	*Xba*I	5’ *nprE* flanking region (fwd)
oCS024	AATTGGATCCGCTCCGTTAGAAACAGCGTC	*BamH*I	3’ *nprE* flanking region (rev)
oCS025	AATTGCGGCCGCGACGGTTCATTCTTCTCTCC	*Not*I	Plasmid backbone pCS10 (fwd)
oCS026	AATTCCGCGGCGTGACCTGTAGCAGAATTC	*Sac*II	Plasmid backbone pCS10 (rev)
oCS027	AATTTCTAGAATAGGATCCGGAATTGACTCAAGCTTCAC	*Xba*I, *Bam*HI	StarGate® backbone (fwd)
oCS028	AATTCCGCGGATAGCGGCCGCGCTAACGCCCGAATTCCG	*Sac*II, *Not*I	StarGate® backbone (rev)
oCS029	AATTCCGCGGTATGATTCTTCTCGCTTCCG	*Sac*II	StarGate® backbone (fwd)
oCS030	AATTTCTAGATAGGTATATCATCTCTCGCC	*Xba*I	StarGate® backbone (rev)
oCS054	AATTCGTCTCCCGGCGTTTAGTGGAAATGATTGCGGC	*BsmB*I	P_{srfAA} - StarGate® (fwd)
oCS055	AATTCGTCTCCCCAACATTTCCTCTCCTCCCTTTATAAGCAGTGAACATGTGC	*BsmB*I	P_{srfAA} - StarGate® (rev)
oCS058	AATTCGTCTCCCGGCATCCACGCTGTGTAAAAATTTTAC	*BsmB*I	P_{15} - StarGate® (fwd)
oCS060	AATTCTCGAGAAACCGTGTGCTCTACGACC	*Xba*I	Introducing erythromycin (fwd)
oCS061	AATTTCTAGACTGCGCAAAAGACATAATCG	*Xho*I	Introducing erythromycin (rev)
oCS062	AATTCTCGAGCCGTTAGCGTTTAAGTACATCC	*Xba*I	Exclusion spectinomycin (fwd)
oCS063	AATTTCTAGACCTCTTGTGAAATTAGAGAACGC	*Xho*I	Exclusion spectinomycin (rev)
oCS064	AATTCGTCTCCCGGCAAGCGGACGGTGTCATTAAC	*BsmB*I	P_{gapB} - StarGate® (fwd)
oCS066	AATTCGTCTCCCCAACATTTCCTCTCCTCCGCCATCCTAATTTGTTACTATC	*BsmB*I	P_{gapB} - StarGate® (rev)
oCS068	AATTCGTCTCCCCAAATGTAAATCGCTCCTTTTTTAATTAATTTC	*BsmB*I	P_{15} - StarGate® (rev)
oCS072	AATTCGTCTCCCGGCCGGACTCTATAGGATGTTTC	*BsmB*I	P_{comGA} - StarGate® (fwd)
oCS073	AATTCGTCTCCCCAACATTTCCTCTCCTCCAACGCATATTGTAGAAAAAGAAG	*BsmB*I	P_{comGA} - StarGate® (rev)
oCS074	AATTCGTCTCCCCAACCACACTGTTACGTTAGAAC	*BsmB*I	*xynA* - StarGate® (rev)
oCS080	AATTCGTCTCAACATATGTAAATCGCTCCTTTTTTAATTAATTTC	*BsmB*I	P_{15} - StarGate® (rev)
oCS082	AATTCGTCTCTATGTTTAAGTTTAAAAAGAATTTCTTAG	*BsmB*I	*xynA* - StarGate® (fwd)
oCS093	AAGCACACGCAGGTCATTTG		Integration check into *aprE* (fwd)
oCS094	CCATCCGTCGATCATGGAAC		Integration check into *aprE* (rev)
oCS097	AATTTCTAGAGAGAAGTTCAAAAATATTATTGAC	*Xba*I	Synthetic promoter (fwd)
oCS098	AATTGGTACCGCTTACCTTTTCTATTGAATCC	*Kpn*I	Synthetic promoter (rev)
oCS102	AATTCGTCTCCCGGCCTCAATAAATAGCTCATTCTC	*BsmB*I	P_{htrA} - StarGate® (fwd)
oCS103	AATTCGTCTCCCCAACATGTTCACTCCGTTTCTC	*BsmB*I	P_{htrA} - StarGate® (rev)

Name	Sequence (5’ to 3’)	Restriction site	Description
oCS108	AATTCGTCTCCCGGCGAAGACTGCTATCCTCATCG	*BsmB*I	P_{ctc} - StarGate® (fwd)
oCS109	AATTCGTCTCCCCAAATTCAGCACCATCCTCTTGTC	*BsmB*I	P_{ctc} - StarGate® (rev)
oCS221	AGTAGTTCCTCCTTATGTAC		Gibson assembly® backbone pSS148 (fwd)
oCS222	TTTAATTCTGCGTGACATCC		Gibson assembly® backbone pSS148 (rev)
oCS262	TCATTCGGTTAGACAGCGG		Integration check into *nprE* (fwd)
oCS263	TTGCTCTTAACTGGACGCG		Integration check into *nprE* (rev)
oCS314	TACTGTTACGTGGCAAAACTAG		P_{amyQ} - Gibson assembly® (fwd)
oCS316	AAAAAAGGAGCGATTTACATATGTTTAAGTTTAAAAAGAATTTCTTAGTTG		*xynA* - Gibson assembly® (fwd)
oCS317	CTAGTTTTGCCACGTAACAGTACTACCACACTGTTACGTTAGAACTTC		*xynA* - Gibson assembly® (rev)
oCS318	ATGTAAATCGCTCCTTTTTTAATTAATTTC		P_{amyQ} - Gibson assembly® (rev)
oCS345	CTAGTACATAAGGAGGAACTACTATGGATAACTATCGTGATGAAAAC		*htrA* - Gibson assembly® pCS47 (fwd)
oCS346	CCTCAGTGGTTGGCTCCAATTACGAAGTTTTCTCTTCTTTTTGATC		*htrA* - Gibson assembly® pCS47 (rev)
oCS373	CAGTGATGACCACGTCACAACGTCGAAAATTGGATAAAGTGG		Chloramphenicol - Δ*htrA* (fwd)
oCS374	GAATCACCGAAGTTTGCCACACCGACTGTAAAAAGTACAGTC		Chloramphenicol - Δ*htrA* (rev)
oCS375	CAGTTTTGGAAATCAGCGGCCGTCGAAAATTGGATAAAGTGG		Chloramphenicol - Δ*htrB* (fwd)
oCS376	GTGAAGAGTCACCAAAGCTCACCGACTGTAAAAAGTACAGTC		Chloramphenicol - Δ*htrB* (rev)
oCS379	GAAAGGTAATGAGAATGAGGTC		5’ Δ*htrA* - Gibson assembly® (fwd)
oCS380	TTGTGACGTGGTCATCACTG		5’ Δ*htrA* - Gibson assembly® (rev)
oCS381	GTGGCAAACTTCGGTGATTC		3’ Δ*htrA* - Gibson assembly® (fwd)
oCS382	CGTTTCGGAGAATTTTCACTTC		3’ Δ*htrA* - Gibson assembly® (rev)
oCS383	TGGATTATCGACGTGATGGC		5’ Δ*htrB* - Gibson assembly® (fwd)
oCS384	GCCGCTGATTTCCAAAACTG		5’ Δ*htrB* - Gibson assembly® (rev)
oCS385	GAGCTTTGGTGACTCTTCAC		3’ Δ*htrB* - Gibson assembly® (fwd)
oCS386	TGCTTTCTGTCTGCTTGGTC		3’ Δ*htrB* - Gibson assembly® (rev)
oCS427	CTAGTACATAAGGAGGAACTACTATGGATTATCGACGTGATGG		*htrB* - Gibson assembly® pCS47 (fwd)
oCS428	CCTCAGTGGTTGGCTCCAACTTAGCTTGAACTGCTTTCTG		*htrB* - Gibson assembly® pCS47 (rev)
oCS429	CTAGTACATAAGGAGGAACTACTATGGTGGATTACGAACGTGA		*htrC* - Gibson assembly® pCS47 (fwd)
oCS430	CCTCAGTGGTTGGCTCCAATATTAACTGCCTAATTGGTCTG		*htrC* - Gibson assembly® pCS47 (rev)
oCS431	CTAGTACATAAGGAGGAACTACTATGAAACGCAGAAAATTCAGCTC		*wprA* - Gibson assembly® pCS47 (fwd)
oCS432	CCTCAGTGGTTGGCTCCAAGTTACTTTTCAACAACAACTTTTGC		*wprA* - Gibson assembly® pCS47 (rev)
oCS440	AATGTAAAAAGGAGGACAAGTCAATGAAAAAGAAAACGCTTTCTTTATTTG		pCS73 - Gibson assembly® (fwd)
oCS441	ATTGACTTGTCCTCCTTTTTCAGTCGCTAGTTTTGCCACG		pCS73 - Gibson assembly® (rev)
oCS442	AAAAAGGAGGACAAGTCAATGGTCAGCAAGGGAGAGG		*mCherry* - Gibson assembly® (fwd)
oCS443	CGGGCTGTATGACTGGAATAATACAAGCAAGTGCATATCCTG		*mCherry* - Gibson assembly® (rev)
oCS453	GGGAAGCGTTCACAGTTTCG		Integration check into *amyE* (fwd)
oCS454	CCCGCTCCGATTAAAGCTAC		Integration check into *amyE* (rev)
oCS462	TTAATTAATTTCCTCTCCCTCTC		Exclusion of *amyM* - pCS67 (rev)
oCS468	CTGTTACGTGGCAAAACTAG		Exclusion of *amyM* - pCS67 (fwd)
oCS492	GGTCATTACAGTGAGGTACG		Amplification of synthetic library (fwd)
oCS493	ACTGACAGACTGGTAACTCG		Amplification of synthetic library (rev)
oCS523	AATTCGTCTCCCGGCGGCCGCTTCTTTGTATTCTG	*BsmB*I	P_{infC}-*mCherry* - StarGate® (fwd)
oCS524	AATTCGTCTCCCCAATTATTACTCGAGTAAGGATCCTTTG	*BsmB*I	P_{infC}-*mCherry* - StarGate® (rev)
oCS525	AGCGCGTCTCCCCGCATCCACGCTGTGTAAAAATTTTAC	*BsmB*I	P_{15} - StarGate® (fwd)
oCS526	AATTCGTCTCAGTCGCTAGTTTTGCCACGTAACAG	*BsmB*I	*amyM* - StarGate® (rev)
oCS527	AATTCGTCTCAGTCGTTACCACACTGTTACGTTAGAAC	*BsmB*I	*xynA* - StarGate® (rev)
oCS528	AATTCGTCTCTCGACTGAAAAAAGGAGCGATTTACAATGGTCAGCAAGGGAGAGG	*BsmB*I	*mCherry* - StarGate® (fwd)
oCS529	AGCGCGTCTCCTTATCAAGCAAGTGCATATCCTGAA	*BsmB*I	*mCherry* - StarGate® (rev)
oCS543	AGTGGAAGTTCTAACGTAACAGTGTGGTAGTTTCGAAAAAAGGCCGCCCC		pCS75 - Gibson assembly® (fwd)

Name	Sequence (5' to 3')	Restriction site	Description
oCS544	CTAAGAAATTCTTTTTAAACTTAAACATGGAGACGCGCTTTACGCC		pCS75 - Gibson assembly® (rev)
oCS545	ATGTTTAAGTTTAAAAAGAATTTCTTAG		*xynA* - Gibson assembly® (fwd)
oCS546	CTACCACACTGTTACGTTAG		*xynA* - Gibson assembly® (rev)
oCS547	AGCGCGTCTCCCCGCCGCCAATACGGAGAAATCTG	*BsmB*I	Synthetic library - StarGate® (fwd)
oCS548	AGCGCGTCTCCACATGAATCTACCTCCTTTTCTAGAAC	*BsmB*I	Variant 1 - StarGate® (rev)
oCS549	AGCGCGTCTCCACATTTAGACATCCTCCTTAACAAGATTG	*BsmB*I	Variant 2 - StarGate® (rev)
oCS550	AGCGCGTCTCCACATGATATCTACCTCCTTTGTAAATCC	*BsmB*I	Variant 3 - StarGate® (rev)
oCS551	AGCGCGTCTCCACATGATATCTACCTCCTTTTCTAGAAC	*BsmB*I	Variant 4,26 - StarGate® (rev)
oCS552	AGCGCGTCTCCACATGATATCTACCTCCTTCATTACAT	*BsmB*I	Variant 5 - StarGate® (rev)
oCS553	AGCGCGTCTCCACATGATATCTACCTCCTTTGTAAAGTC	*BsmB*I	Variant 7,12,18 - StarGate® (rev)
oCS554	AGCGCGTCTCCACATTAGTAATCCCTCCTTTGTAAAGTC	*BsmB*I	Variant 15,24 - StarGate® (rev)
oCS555	AGCGCGTCTCCACATGATATCTACCTCCTTTGTAAGTC	*BsmB*I	Variant 17 - StarGate® (rev)
oCS556	AGCGCGTCTCCACATTTAGTACACCTCCTTTTCTAGAAC	*BsmB*I	Variant 20 - StarGate® (rev)
oCS557	AGCGCGTCTCCACATGATATCTACCTCCTTCACTACA	*BsmB*I	Variant 21 - StarGate® (rev)
oCS558	AGCGCGTCTCCACATGATATCTACCTCCTTTTCTAGAC	*BsmB*I	Variant 23 - StarGate® (rev)
oCS559	AGCGCGTCTCCACATTAGTAATCCCTCCTTCACTACA	*BsmB*I	Variant 25 - StarGate® (rev)
oCS560	AATTCGTCTCTGATAATCCACGCTGTGTAAAAATTTTAC	*BsmB*I	P_{15} - pCS66 - StarGate® (fwd)
oCS561	AGCGCGTCTCTCCATATGTAAATCGCTCCTTTTTTAATTAATTTC	*BsmB*I	P_{15} - pCS66 - StarGate® (rev)
oCS562	AGCGCGTCTCCCCGCATCCACGCTGTGTAAAAATTTTAC	*BsmB*I	P_{15} - pCS80 - StarGate® (fwd)
oCS563	AGCGCGTCTCCACATATGTAAATCGCTCCTTTTTTAATTAATTTC	*BsmB*I	P_{15} - pCS80 - StarGate® (rev)
oCS564	AGCGCGTCTCCCCCGCACGTTGATATAATTTAAATTTTATTTGAC	*BsmB*I	*veg* - pCS80 - StarGate® (fwd)
oCS565	AGCGCGTCTCCACATTGCATCCACCTCACTACATTTA	*BsmB*I	*veg* - pCS80 - StarGate® (rev)
oCS566	AGCGCGTCTCCCCGCGAGAAGTTCAAAAATATTATTGAC	*BsmB*I	*design* - pCS80 - StarGate® (fwd)
oCS567	AGCGCGTCTCCACATTGACTTGTCCTCCTTTTTAC	*BsmB*I	*design* - pCS80 - StarGate® (rev)
oCS568	GTTCCAATACGGAGAAATCG		Sequencing - synthetic library (fwd)
oCS569	TCTTCTCCTTTGCTCATCTG		Sequencing - synthetic library (rev)
oCS570	CGAAAGAGAACGATCAGAGC		qPCR - *htrA* (fwd)
oCS571	AAATGCCAAGAGCAAGACTG		qPCR - *htrA* (rev)
oCS572	GAAGCTGATTGGACATTCTG		qPCR - *htrB* (fwd)
oCS573	TTTTCACAGCCGTTTCTTGC		qPCR - *htrB* (rev)
oCS574	TTACGAACGTGAGGAAGAAC		qPCR - *htrC* (fwd)
oCS575	TGTATCCAGCCCTTCATTTG		qPCR - *htrC* (rev)
oCS576	GAAAATTCAGCTCGGTTGTG		qPCR - *wprA* (fwd)
oCS577	TTTGCCGTTTTCCAGAGCAG		qPCR - *wprA* (rev)
XynA (362)	CCTAAATGGCGAACCTGTAGTC		qPCR - *xynA* (fwd)
XynA (363)	AAACGCTGTCAATGGGTCTG		qPCR - *xynA* (rev)
AmyM (368)	ACATCAGCAACTGGGACGAC		qPCR - *amyM* (fwd)
AmyM (369)	GCTTCACCGCATCAATCC		qPCR - *amyM* (rev)

Table 2.2: Oligonucleotides used in this study. Acronyms as follows: forward (fwd), reverse (rev)

2.4.4 StarGate® - de novo cloning

StarGate® was originally designed by IBA GmbH and modified in collaboration with DSM to fit the host organism *B. subtilis* (Van Rij *et al.*, 2013). The StarGate® system allows the direct and efficient cloning of DNA fragments in one reaction resulting in the successful integration of the assembled DNA into the chromosome of *B. subtilis* via homologous recombination. The Golden Gate assembly is a similar procedure but under different reaction conditions (Engler *et al.*, 2008, Engler *et al.*, 2009). The StarGate® system is based on the digestion of target DNA with the Type II restriction endonuclease *Esp*3I that digests both vector and insert(s) that carry *Esp*3I recognition sites. The designed restrictions sites from vector and insert(s) are compatible to each other, leading to a specific ligation of fragments without remaining scars. For a simple StarGate® reaction, 3 µl of both vector (4 nM) and insert (4 nM – 8nM) were combined with 5 µl 5× T4 DNA ligase buffer (Thermo Fisher Scientific), 1 µl T4 DNA ligase (1 U/µl) (cat. no. 15224-017, Thermo Fisher Scientific), 0.5 µl Esp3I (10 U/µl) (cat. no. ER0451, Thermo Fisher Scientific), made up to 25 µl with 12.5 µl MilliQ water. This mixture was incubated for 3 h at 30°C, another hour at 37°C and then used for transformation of *B. subtilis* (Section 2.3.1). For the synthetic library, the reaction volume for the StarGate® reation was doubled to 50 µl and contained 12 nM of unmethylated pCS66, 24 nM of amplified expression module fragments and double the amount of enzymes. This mix was then used for transformation, however, 950 µl of competent cells were used in a 14 ml Falcon tube (cat. no. 352059, BD BIOSCIENCES) and shaken at 37°C for 1 h before plating on 2×PY selective agar plates (BioAssay dishs, low profile, cat. no. 240845, Thermo Fisher Scientific) and incubated overnight at 30°C wrapped in plastic foil.

2.4.5 Gibson assembly® technology

Gibson assembly uses overlapping DNA fragments to join DNA fragments in a single reaction using either the Gibson Assembly® Master Mix (E2611S, NEB®) or NEBuilder® HiFi DNA Assembly Master Mix (E2621L, NEB®). The difference between these two is the use of a proof-reading polymerase in the NEBuilder® HiFi mix leading to improved efficiency and accuracy. The reaction includes an exonuclease activity leading to single-stranded 3’ overhangs, a polymerase activity filling up gaps after the annealing of DNA fragments and a DNA ligase activity sealing nicks in the DNA (Gibson *et al.*, 2009). For a Gibson reaction involving a single insert, 2.5 µl of the provided master mix (5 µl for more than one insert) was used. For

maximum cloning efficiency, a ratio of purified insert to vector DNA (50-100 ng) of 2.5 to 1 was applied. In cases when inserts were smaller than 200 bp a ratio of 5:1 was used. The DNA mixture was incubated at 50°C for 30 min (for more than one insert 60 min) leading to the complete assembly of insert and vector DNA. Subsequently 2 µl of the Gibson mix was used to transform chemical competent *E. coli* cells, self-made or from NEB®, as described previously (Section 2.3.2). However, it is possible to transform competent *B. subtilis* directly with the Gibson mix: this reduces the overall cloning time since it avoids the need to use *E. coli* as an intermediate host, but increases the probability of false positive clones.

2.4.6 Deletion of genes via Gibson assembly®

The method to assemble DNA fragments seamlessly supports the design of deletion constructs that are otherwise performed by more difficult approaches. For this protocol, three fragments are needed as follows: 5' flanking region of a gene of interest (GOI) that will be deleted, an antibiotic resistance gene with overlapping sequence and a 3' flanking region of the chosen GOI. The 5' ends of the oligos, to amplify the antibiotic resistance gene, contained a 15- to 30-nucleotide overhang that is homologous to the 15 to 30 bases of the 5' end of the adjacent DNA fragment (GOI). After the separate amplification and purification of the three products (Section 2.4.2), the fragments were combined with the Gibson assembly® master mix (E2611S, NEB®) at a ratio of 1:1:1 and incubated for 30 min at 50°C. A PCR reaction using the primers from the outer flanking regions (5' ends) followed to amplify the combined product. If necessary, the correct band was extracted from an agarose gel and subsequently purified. Competent *B. subtilis* was transformed with at least 500 ng of the purified product (Section 2.3.1).

2.4.7 Construction of StarGate® vectors

The plasmid pAPNC213 was linearised by using oCS001/oCS002 for the amplification step to remove *lacI* and the inducible P_{spac} promoter. The PCR product was treated with *Dpn*I and cleaned up to remove any remaining template DNA. The linearised plasmid was kinated by using T4 polynucleotide kinase and subsequently ligated using T4 DNA Ligase, and in turn incubated for 30 min at room temperature before the enzymes were deactivated by heating up to 65 °C for 20 min. Competent *E. coli* cells were transformed with the mix resulting in the plasmid pCS01.

To create pCS02, sfGFP was amplified from pPG49 with the primers oCS003/004. Both pCS01 vector and *sfGFP* fragment were digested with *Xba*I/*Spe*I, treated with *Dpn*I, heat inactivated, purified and ligated using T4 DNA Ligase. Competent *E. coli* cells were transformed with the mix resulting in the plasmid pCS02.

The vector pCS02 was linearised by using oCS005/oCS006 for the amplification step to remove several bp and circularisised by using the method used for pCS01 stated above resulting in the plasmid pCS03.

The kanamycin resistance gene was amplified from pCS75 with primers oCS007/oCS008 to introduce *BsmB*I restriction sites. Both pCS03 vector and kanamycin fragment were digested with *EcoR*I/*Bam*HI, treated with *Dpn*I, heat inactivated, purified and ligated using T4 DNA Ligase. Competent *E. coli* cells were transformed with the mix resulting in the plasmid pCS04.

To create the final backbone for StarGate® reactions, pCS04 was linearised by using oCS009/oCS010 for the amplification step to remove several bp and circularisised by using the method used for pCS01 stated above resulting in the plasmid pCS05.

Another StarGate® backbone was created by amplifiying nprE flanking regions (5', oCS023/oCS024; 3', oCS025/oCS026) and essential regions for plasmid replication in *E. coli* (oCS027/oCS028) and sequential digesting with *Xba*I/*BamH*I and then *Not*I/*Sac*II to obtain pCS10. pCS13 was created by changing *sfGFP* to *mCherry* and spectinomycin to erythromycin using oCS062/oCS063 to linearise pCS05 and exclude spectinomycin and insert erythromycin using oCS060/oCS061 and circularisised by using the method used for pCS01 stated above.

pCS31 was created by linearising pCS10 to remove mCherry and circularisising by using the method used for pCS01 stated above. pCS41 was constructed by using pCS31 as template DNA to exchange erythromycin with kanamycin resistance.

2.4.8 Construction of vectors for overexpression of enzymes

The vectors for overexpression of XynA and AmyM were constructed by using the plasmid pCS73 (DSM). pCS58 was created by using Gibson Assembly® by amplifying pCS73 with the primers oCS314/oCS318 to remove *amyM* and the amplification of *xynA* from chDNA of 168 via oCS316/oCS317.

pCS67 was constructed by using Gibson Assembly® with the following fragments: linearised pCS73 via oCS440/oCS441 and *mCherry* via oCS442/oCS443.

The plasmid pCS76 was created by linearising pCS67 using oCS462/oCS468 to remove *amyM* and circularising by using the method used for pCS01 stated above.

2.4.9 Construction of vectors for overexpression of proteases

The backbone for the overexpression vectors was plasmid pSS148. To create pCS47 mCherry was excluded by linearising pSS148 using oCS221/oCS222. Subsequently, the protease genes were amplified: *htrA* (oCS345/oCS346), *htrB* (oCS427/oCS428), *htrC* (oCS429/oCS430) and *wprA* (oCS431/oCS432), and combined with the vector fragment using Gibson Assembly® resulting in pCS61, pCS68, pCS69 and pCS70.

2.4.10 Construction of vectors for expression module library

pCS38 was constructed by removing the spectinomycin resistance gene from pCS75 and inserting kanamycin. pCS66 was created by amplifying *sfGFP* from pCS05 and introducing it into pCS75 after removing cat (performed by Daphne Groothuis).

pCS78 was constructed by amplifying P_{infC}-*mCherry* via oCS523/oCS524 and using pCS41 for a StarGate® reaction.

pCS80 was created by using Gibson Assembly® with the primers oCS543/oCS544 to linearise pCS75 and remove *amyM* and oCS545/oCS546 to amplify *xynA* from chDNA of 168.

2.4.11 RNA extraction

RNA was extracted for the quantitative reverse transcription. Target strains were incubated in LB with appropriate antibiotics in 24 well plates (part no. PDW10ML24CLIDBCS, Corning Axygen) shaking at 250 rpm on a New Brunswick Scientific Escella E24 Shaker at 37°C overnight. In the morning, 2 ml pre-cultures in LB were prepared by diluting the overnight cultures to a starting OD_{600} of 0.05 without the addition of antibiotics and incubated with the same conditions as overnight. As the cultures reached exponential phase (OD_{600} between 0.2-0.8) the cultures were diluted, in quadruple, into 1.5 ml pre-warmed LB to an OD_{600} of 0.02. Growth was continued in FlowerPlates® in the BioLector® (m2p-labs) shaking with 800 rpm at 37°C and a humidity of ~95% to prevent an evaporation of the medium. Growth was monitored by measuring scattered light (λ_{620}) every 15 min. When the cells were in

exponential phase (4 h), 0.5 ml of culture from two wells was collected and vortexed with 1 ml of RNA protect (cat. no. 76506, Qiagen) in a 2 ml Eppendorf tube and incubated for 1 min at room temperature. The tubes were then centrifuged for 2 min at 15,781 ×*g* at room temperature and the supernatant discarded. The pellets were stored at -80°C. The remaining two cultures were processed in the same way when reaching stationary phase (20 h).

While extracting RNA, the samples were kept on ice whenever possible, gloves were changed as much as possible and RNaseZap® solution (cat. no. R2020, Sigma-Aldrich) used on pipettes and benches to destroy RNases. For the isolation of RNA, the RNeasy mini kit (cat. no. 74104, Qiagen) was used with an adjusted protocol. The pellets were thawed and resuspended in the supplied 1 ml RLT buffer (containing 10 µl of 2-Mercapto-ethanol per 1 ml of RLT). The suspension was transferred to tubes containing 0.1 mm beads (lysing matrix B, cat. no. 116911050, MP BIOMEDICALS) to mechanical lyse the cells with the Precellys device (cat. no. S6009, Zymo Research) for 3×10 s with a short break in between. Subsequently, the tubes were centrifuged for 10 min at 25,000 ×*g* at 4°C and then the clear lysate was pipetted into fresh 2-ml Eppendorf tubes. To the lysate, 600 µl ethanol (96-100% (v/v)) was added before the tubes were inverted and up to 700 µl of lysate containing ethanol was pipetted to the RNeasy mini column. After a centrifugation step for 15 s at 8,000 ×*g,* the flow-through was discarded and if there was more than 700 µl of lysate this step was repeated. With the addition of 700 µl Buffer RW1 to the column and centrifugation for 15 s at 8,000 ×*g* the flow-through was discarded. This step was repeated and the washed column was put in a new 2-ml collection tube and 500 µl of supplied RPE buffer were added to the column and centrifuged for 15 s at 8,000 ×*g,* again discarding the supernatant. This step was repeated but with a prolonged centrifugation time of 2 min. Subsequently, the column was placed on a new collection tube and further centrifuged for 2 min to dry the column thoroughly. After putting the column in a new 1.5 ml collection tube, 50 µl RNase free water was added and another centrifugation step was performed for 1 min at 8,000 ×*g* to elute the RNA. The elution was repeated a second time with the same elute.

To completely remove any remaining DNA, the eluted RNA samples were treated with the RNase-free DNase (cat. no. 79254, Qiagen). The supplied protocol was used and followed by a cleaning procedure to remove the residual DNase. RLT buffer (350 µl) was added and then 250 µl of ethanol (96-100% (v/v)), inverting the tubes in between. Then the sample was applied to the RNeasy mini column and

centrifuged for 15 s at 8,000 ×*g*. The column was transferred into a new 2 ml collection tube before 500 µl RPE was pipetted and a centrifugation step followed at 15 s at 8,000 ×*g*. This step was repeated and the washed column was put in a new collection tube and dried while centrifuging at full speed for 2 min. To elute the RNA, the column was placed in a new 1.5-ml collection tube and 50 µl of RNase free water was added, with second elution step using the same sample. The RNA concentration was measured by the NanoDrop (Thermo Fisher Scientific) with a preferred reading at A260/A280 >1.8 (ideally 2) and A260/A230 >1.8 (ideally 2.0 to 2.2). The purified RNA samples were stored at -80°C until they were used for the reverse transcription PCR.

2.4.12 Reverse transcription PCR

The reverse transcription (RT) creates complementary DNA (cDNA) from RNA samples (Section 2.4.11) as a first step in quantitative RT-PCR. RT uses a combination of dNTPs and the enzyme reverse transcriptase and is followed by quantitative PCR using fluorescent probes. The RT reaction was performed with the high capacity cDNA reverse transcription kit (cat. no. 4368814, Applied Biosystems) using no RNase inhibitor. As a control for residual DNA contamination, a reaction without reverse transcriptase was performed in parallel. For the reaction a 2×RT master mix was prepared on ice following the manufacturer's protocol but using double the amount of RT buffer, dNTP mix, RT random primers, reverse transcriptase and nuclease-free water in a total reaction volume of 20 µl. For each RT-PCR, 20 µl of master mix was combined with 20 µl of RNA sample in either a 96-well plate or reaction tube that was then sealed and briefly centrifuged to eliminate air bubbles. The plate or tubes were loaded into the T100^{TM} thermal cycler (Bio-Rad) to perform the RT-PCR with the following cycles: 10 min at 25°C, 120 min at 37°C and 5 min at 85°C. The finished samples were stored at -20°C for long-term storage.

2.4.13 Quantitative PCR

After the RT-PCR (Section 2.4.12), finished samples containing 40 µl of cDNA were diluted 5-fold with 160 µl RNase free MilliQ water or, if necessary, up to 10-fold. For quantitative PCR (qPCR), the primers were designed to ensure they were free of potential secondary structures and specific to the target sequence leading to an amplicon length between 100 and 150 bp. The resulting qPCR primers were verified using the Primer3Plus tool (Untergasser *et al.*, 2007). For validation, the primers

were probed for their primer efficiency using chDNA from *B. subtilis* strain 168 to achieve efficiencies between 96-100%. The cDNAs were analysed using the iQ™ SYBR® green supermix (cat. no. 1708882, Bio-Rad) in a C1000 thermal cycler (CFX96 real-time system, Bio-Rad). For each qPCR reaction, 12.5 µl SYBR green supermix, 0.75 µl of each primer (0.3 µM final concentration), 6 µl of MilliQ water and 5 µl template cDNA were added, giving a total reaction volume of 25 µl. The *B. subtilis fbaA* and *sdhA* genes were used as reference genes to calculate the transcription levels of the target genes. The qPCR was carried out according to manufacturer's description: initial denaturation for 300 s at 95°C followed by 40 cycles of denaturation for 10 s at 95°C, the annealing of oligos for 45 s at 58°C and the elongation for 45 s at 72°C. Afterwards, the melting curve was performed by denaturing the DNA for 10 s at 95°C and then the samples were heated up from 65°C to 95°C for 10 s with incremental temperatures of 0.5°C for each cycle. For the analysis, an Excel of the data was used to calculate the fold increase of RNA transcripts of the target genes compared to the transcripts of the reference genes. To perform the qPCRs with high accuracy and reproducible methodology, guidelines were followed to facilitate the interpretation and general comparison between data sets (Pfaffl, 2001, Bustin *et al.*, 2009).

2.4.14 Plasmid DNA extraction

For the extraction of plasmid DNA the GeneJET Plasmid Miniprep Kit was used (Thermo Scientific, #K0502, #K0503). *E. coli* was cultured overnight in LB medium containing ampicillin at 37°C. After harvesting the cells for 1 min at 16,000 ×*g* the supernatant was discarded and the pellet resuspended in 250 µl resuspension buffer (100 µg/ml RNase A). By the addition of 250 µl lysis solution and gently inverting the tubes the cells were lysed. After an incubation time of 1 min at room temperature, the lysate was neutralised by adding 350 µl neutralisation solution and inverting the tubes to precipitate genomic DNA, proteins and cell debris. After the centrifugation at 16,000 ×*g* for 5 min, the supernatant was carefully transferred to the spin column and subsequently spun down and the flow-through was discarded. To wash the column 500 µl of wash solution (64% (v/v) ethanol) was added to remove remaining proteins and RNA. After repeating the washing procedure the column was dried and to remove residual wash solution by centrifuging at 16,000 ×*g* for 1 min. To elute the plasmid DNA, MilliQ water was added to the middle of the column, incubated at room temperature for 1 min and spun down at 16,000 ×*g* for 2 min. The quality of the

eluted plasmid DNA was determined by using a NanoDrop spectrophotometer (Thermo Fisher Scientific).

2.4.15 DNA sequencing

The automated DNA sequencing approach follows the method of Sanger that is based on chain termination by dideoxynucleotides during *in vitro* DNA elongation reactions. During a single PCR reaction dideoxynucleoside triphosphates, labelled with fluorescent dyes, are incorporated into the newly synthesised DNA strand leading to the termination of the elongation process and generation of different sized fragments. Since the dyes emit light at different wavelengths, the fragments are detected using a laser beam and the information is collected by automated systems. This results in the generation of chromatograms that show coloured peaks reflecting the sequence of the template DNA (up to ~700-800 bp).

2.5 Fluorescence microscopy

2.5.1 Continuous culture microscopy

Bacterial cultures were grown in a variety of media until exponential phase, diluted in fresh medium and grown to a required stage of the growth cycle, before 100 µl of cells were pelleted and resuspended in an equal volume of PBS (or less if it was necessary to concentrate the cells). Microscope slides were mounted with a GeneFrame (Thermo Fisher Scientific) and 125 µl of 1.25% (w/v) of agarose in PBS were pipetted into the frame and, when set, 0.3 µl of the resuspended culture was spotted on the surface. For microscopy either the Nikon Eclipse Ti with a built-in perfect focus system from Nikon or the Zeiss Axiovert 200M was used. Both microscopes were mounted with a CoolSNAP HQ² CCD camera (Photometrics®) and controlled by MetaMorph software (Molecular Devices). The excitation/emission wavelengths (in nm) and exposure times of fluorophores with the Nikon microscope were the following: DAPI (340-380/435-485) for 500 ms and 1000 ms, GFP (460-500/510-560) for 500 ms and 1000 ms, mCherry, Nile red or FM95-5 (550-600/615-665) for 500 ms and 1000 ms, respectively. For the Zeiss microscope they were as follows: DAPI (350/460) for 500 ms and 1000 ms, GFP (470/525) for 500 ms and 1000 ms, mCherry, Nile red or FM95-5 (560/630) for 500 ms and 1000 ms, respectively. For phase-contrast, an exposure time of 100 ms was selected. To stain cell membranes 0.2 µg/ml Nile Red (9-diethylamino-5-benzo-[α]-phenoxazinone) or 3

µg/ml FM5-95 (N-(3-trimethyl-ammoniumpropyl)-4-(6-(4-(diethylamino)phenyl)hexatrienyl)pyridinium dibromide) were used. For the staining of the nucleoid, to allow the measurement of GFP or/and mCherry fluorescence in every cell via NucTracer, 2 µg/ml DAPI was used (Section 2.6).

2.5.2 Time-lapse microscopy

Cells for time-lapse microscopy were grown in minimal competence medium (Section 2.2) to exponential phase, diluted in fresh medium and, after a further 2 h, diluted ~100-fold in fresh medium to be able to spot single cells on the microscopic slide. To ensure the physical separation and aeration of the cells they were spotted in a gene frame (Thermo Fisher Scientific) on to thin stripes of 1.4% (w/v) 2-hydroxyethyl agarose (low gelling temperature agarose/type VII, cat. no. A4018, Sigma-Aldrich) containing PBS with 10% (v/v) of minimal competence medium. Time-lapse microscopy was performed either with an IX71 microscope (Olympus®) for at least 16 h at 30°C with a DeltaVision® system or with a Nikon Eclipse Ti mounted with a CoolSNAP HQ² CCD camera (Photometrics®) at 32°C. Both systems were equipped with a temperature-controlled incubation box. Neutral density filters were used to reduce fluorescence intensity by ~50%. Exposure times and excitation/emission wavelengths were as follows: GFP (470/525) for 1000 ms and mCherry (545/620) for 1000 ms. For phase contrast 50 ms exposure time was selected and pictures were taken every 15 min.

2.6 Microscopy analysis

Microscopy images were analysed using the NucTracer software that assigns the DAPI stained nucleoid as identifier for the region of interest to measure GFP or/and mCherry fluorescence in each cell. NucTracer (Syvertsson *et al.*, 2016) is based on the plugin ObjectJ that supports vector objects identifying images on a transparent layer in ImageJ (Vischer *et al.*, 1994, Schneider *et al.*, 2012, Sauer *et al.*, 2016).The analysis was performed with 16-bit tiff files blanking the measured values against the average background value of the image.

2.7 β-Galactosidase assay

For the measurement of β-galactosidase activities, cells were collected in exponential and stationary phase, alongside the samples used for microscopy (Section 2.5.1),

and subsequently flash frozen using liquid nitrogen and stored at -80°C. The activity assays were performed according to the protocol of Miller (Miller, 1972), whereby cells were first resuspended in working buffer (1.067% (v/v) $Na_2HPO_4 \times 2H_2O$, 0.551% (v/v) $NaH_2PO_4 \times H_2O$, 0.074% (v/v) KCl, 0.024% (v/v) $MgSO_4$, adjusted to pH 7.0 and 50 mM β-mercaptoethanol [freshly added every time]) and their OD_{600} determined. The cells were then lysed by the addition of 20 mg/ml lysozyme and incubation for 30 min at room temperature. For the start of the experiment, 2-nitrophenyl-β-D-galactopyranoside (ONPG, 20 mM), dissolved in working buffer without β-mercaptoethanol, was added at a 1:8 volume ratio and the time was measured until a yellow colour developed. Reactions were terminated by the addition of 1 M Na_2CO_3 at a ratio of 1:2.25 followed by the measurement of the OD_{550} and A_{420} values. The β-galactosidase activity in Miller Units (MU) was calculated using the following equation:

$$Miller\ Units = \frac{1000 * (A_{420} - (1.75 * OD_{550}))}{t * v * OD_{600}}$$

To resolve the equation, the values of *A* (absorbance at specified wavelength), *OD* (optical density at specified wavelength), *t* (reaction time) and *v* (reaction volume) were inserted.

2.8 Transposon screen of expression cassette in *B. subtilis*

The construction of the plasmid containing an expression cassette, encoding *lacZ* and *sfGFP,* and the *himar1* transposase for an unbiased integration into the chromosome of *B. subtilis* was based on work described by Lampe and Le Breton (Lampe *et al.*, 1996, Le Breton *et al.*, 2006). In brief, *B. subtilis* was transformed with the transposon plasmid pSS125 leading to colonies on erythromycin selective plates at 30°C. After some colonies were cultured in LB medium for 3-5 h, they were streaked on kanamycin selective plates overnight at 50°C to prevent further replication of the plasmid and to force the integration into the chromosome via the transposase. Subsequently, arbitrary PCRs based on Knobloch (Knobloch *et al.*, 2003) were performed of strains, encoding only kanamycin resistance, to identify the location of the transposed expression constructs. The resulting strains were tested for their β-galactosidase and GFP activities (Section 2.6 and 2.7) using 1 mM IPTG for the induction of the P_{spac} promoter driving the expression of β-galactosidase from the start of the experiment. Samples for the activity and fluorescence assays were harvested during exponential phase, in stationary phase and 2 h after the induction of

P_{spac} in stationary phase. For the analysis of the transposon strains, LB at 30°C and 37°C and minimal competence medium at 37°C were used.

2.9 A microbioreactor system

The BioLector® is a microbioreactor system that allows, more than the usual plate readers, high-throughput fermentations with the simultaneous measurement of biomass (λ_{620}), fluorophores, pH and dissolved oxygen by keeping a stable humidity around 95%. After growing pre-cultures in 24-well plates (part no. PDW10ML24CLIDBCS, Corning Axygen) at 37°C shaking at 250 rpm in an Excella E24 incubator shaker (New Brunswick Scientific) the cells were inoculated in FlowerPlates® B or BOH (cat. no. MTP-48-B or BOH, m2p-labs) with a starting OD_{600} between 0.02 and 0.05 in pre-warmed medium.
Before adding the cells, the appropriate amount of medium was removed to synchronize the wells to achieve an equal biomass. To calculate relative fluorescent activities (fluorescence per unit biomass), the relative fluorescent activity of the wild type strain lacking the *gfp* gene was subtracted from the test samples. As a result, the background signal of medium and bacteria was removed and subsequently these values were multiplied with the factor 1000.

2.10 Activity assays of industrial enzymes

2.10.1 Xylanase assay

Xylanase activities were assayed with the EnzChek® ultra xylanase assay kit (cat. no. E33650, Thermo Fisher Scientific) following the supplied protocol. The principle of the assay is based on the hydrolysis of the substrate *o*-nitro-phenyl-β-D-xylobioside that is bound to a fluorescent marker. After mixing the substrate with the diluted xylanase sample (1:100), the reaction mix was incubated in the dark for 15 or 30 min at room temperature. The increase in fluorescence signal resulting from the xylanase activity is measured using a plate reader (Infinite 200 Pro, Tecan® or FLUOstar OPTIMA, BMG LABTECH®) with the excitation/emission wavelength of 358/455 (Section 2.11)

2.10.2 Amylase assay

To determine α-amylase activity the Phadebas® amylase assay was used (cat. no. 1302, Phadebas®). The amylase activity is linked to the degradation of a cross-linked

starch polymer that leads to the solubilisation of the blue chromophore from the insoluble substrate. The absorbance at 620 nm reflected the α-amylase activity of the sample. The protocol was modified as follows: 1 Phadebas® tablet was added to 5 ml of 100 mM phosphate buffer (K_2HPO_4 17.42 g; KH_2PO_4 13.61 g, pH 5.5) and thoroughly vortexed. During the assay, the insoluble substrate was constantly mixed to provide an equal amount of substrate to each sample. After the addition of 180 µl substrate solution to a MTP, the plate pre-heated to 60°C for 5 min. The assay was started by pipetting 20 µl sample into the MTP and stopped after 20 min with the addition of 70 µl of a 1 M TRIS solution (Tris (hydroxymethyl)aminomethane, cat no. 648310, Molecular Biology Grade - CAS 77-86-1 - Calbiochem). The plate was centrifuged at 3,000 ×*g* for 10 min and 100 µl of supernatant was transferred into a fresh MTP. The A_{620} was measured in a plate reader.

2.11 Plate reader assay

To measure fluorescence, absorbance or optical density, the FLUOstar OPTIMA (BMG LABTECH®) and Infinite 200 Pro (Tecan®) plate readers were used. For the analysis of xylanase activity (Section 2.10.1), the fluorescence signal was measured in a Nunc® 96-well plate (flat, black) with a gain of 75. The signal reads were performed with 20 flashes with a settle time of 50 ms. To analyse amylase activity (Section 2.10.2), the A_{620} was monitored in a Nunc® 96-well plate (flat, transparent). For the assays, the fluorescence or absorbance values for samples containing just the substrate solution and water, was subtracted from the activities of the samples to remove the background signal.

2.12 Precipitation of proteins from culture supernatant

For the analysis of secreted proteins of engineered strains, the culture supernatant was treated with trichloroacetic acid (TCA) to precipitate proteins, followed by a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The protocol started with the inoculation of the strains in the appropriate medium from an overnight culture with an OD_{600} between 0.02 and 0.05, and incubation with shaking at 37°C. The cultures were grown for at least 4 hours and then the cells were harvested by centrifugation at 3,893 x*g* for 30 min at 4°C. The supernatant was carefully transferred into a glass bottle, a pre-cooled solution of TCA added to a final concentration of 12% (v/v) and the mixture incubated overnight on ice. The precipitated proteins were harvested by centrifuging the samples at 30,000 x*g* for 70

minutes at 4°C. Most of the supernatant was decanted and with the remaining supernatant (~1 ml) the pellet was resuspended and centrifuged for an additional 20,000 x*g* for 15 min at 4°C. The pellet was then washed 3 times with ethanol (96-100% (v/v)), each time, with a centrifugation step at 20,000 x*g* for 15 min at 4°C. After drying the pellet at 65°C, it was dissolved in a solution of urea (8 M)/thiourea (2 M). This mixture was centrifuged again at 20,000 x*g* for 30 min at room temperature, this time collecting the supernatant. The protein concentration was determined using the 2-D Quant Kit from GE Healthcare Life Sciences. Equal amounts of protein were boiled in sodium dodecyl sulfate (SDS) loading buffer for 10 min. The samples were then loaded in a 12% (w/v) SDS polyacrylamide gel and the electrophoresis was started in a SDS running buffer (70 volts for 20 min followed by 150 volts for one hour). The gel was stained for 10 min with coomassie brilliant blue, destained for 10 min with a destain solution, and incubated in MilliQ water at room temperature overnight in order to distinguish the protein bands of the secreted proteins.

2.13 Synthetic expression module library

2.13.1 Bioinformatics design

A synthetic expression module library was used to identify strong expression units in *B. subtilis* based on the σ^A-dependent P_{veg} promoter. For the bioinformatics design, sequence parts of P_{veg} were synthetically modified with randomised sequence. These modified parts were combined and this resulted in a library of 12,096 different variants. However, key elements consisting of optimised ribosome binding site (RBS) and sigma factor recognition sequences were conserved to ensure high gene expression. The features of the synthetic library included:

- the UP element upstream the -35 region of sigma factor recognition (TTGACA) was modified by altering the percentage of AT content (25%, 50%, 75% and 100%)
- the spacer between the -35 and -10 (TATAAT) regions was randomised with the restriction that it maintained a high AT content of 75%
- the spacer between position -10 and the RBS, was modified with the restriction that the nucleotide at position +1 maintained a balance between purines and pyrimidines. The distance was increased in steps of 10 from 10 bp to 30 bp with no restriction on the %AT content

- the 8 bp-spacer between RBS and start codon was modified while maintaining a %AT content of 75%

The associated combinatorial design, resulting in a synthetic expression module library, containing 12,096 oligonucleotides, was created using the programming language Python and sent to the BROAD Technology Lab (Cambridge, MA 02142) for synthesis. The resulting oligonucleotide pool was quantified using a Qubit® fluorometer with a measured concentration of 1.44 ng/µl.

2.13.2 Cloning procedure

Before the synthesised library was used for a StarGate® reaction, the oligonucleotide pool was enriched for full-length synthesis products. This step was done by performing a PCR using 25 µl NEBNext® high-fidelity 2× PCR master mix (cat. no. M0541S, NEB®), 19 µl MilliQ water, 2.5 µl of each primer (10 µM - oCS492/oCS493) and 1 µl of the 1:100 diluted oligonucleotide pool. The PCR was as follows: Initial denaturation at 98°C for 30 s, 25 cycles consisting of the denaturation at 98°C for 10 s, annealing at 58.4°C for 20 s and an elongation at 72°C for 15 s followed by the final extension at 72°C for 300 s.

After the purification of the PCR products (Section 2.4.2) StarGate® reactions with subsequent transformations were performed as described in Section 2.4.4 leading to a library of 152,000 transformants. The colonies were scrapped off the plates and combined in a flask and stored at -80°C.

2.13.3 Sequencing

To investigate the quality of the expression module library with respect to its conformity to the original bioinformatics design, 192 clones were picked with the QPix 450 colony picker (Molecular Devices LLC), cultured overnight, chromosome extracted and sent for sequencing.

To determine whether the activities of the strong expression units that were identified by the sequencing process were applicable for industrially relevant strains, high-expressing GFP transformants were handpicked and fused to *xynA* to investigate the correlation. For this approach both strains with the same expression module but producing either GFP or XynA were cultured in a BioLector® for the measurement of GFP and XynA activities (Section 2.9).

2.14 Flow cytometry

Flow cytometry facilitates the high-throughput counting, the detection of fluorescence signal and sorting of cells based on fluorescence intensity.

2.14.1 Fluorescence-activated cell sorting (FACS)

To investigate *B. subtilis* cells via flow cytometry, overnight cultures of specific strains were prepared in 5 ml of LB medium containing the appropriate antibiotics shaking at 165 rpm at an Innova 2300 platform shaker (New Brunswick Scientific) at 30°C or 37°C. For cultures growing at 37°C, cells were pre-cultured for 3 h before they were diluted and grown to the desired phase of growth.

For cells growing at 30°C, 0.33 ml of overnight culture was added into 100 ml of LB (without antibiotics) shaking for 3 h and 15 min at 165 rpm at an Innova 2300 platform shaker at 30°C (Chapter 5). After reaching an OD_{600} of ~0.25, the flask was put immediately on ice to prevent further growth and protein folding. Afterwards the cells were centrifuged at 3,000 ×*g* for 10 min at 4°C (including cells grown at 37°C). The supernatant was removed and the cells were resuspended in 15 ml of ice-cold PBS and cell chaining was reduced by ultrasonication using the lowest power for 10 s. This suspension was immediately run through the FACSAria™ Fusion flow cytometer system (BD Biosciences) at 4°C using the excitation/emission wavelength of 488/530-30-A for GFP and 561/610-20-A for mCherry and the forward and side scatter used to measure cell volume and morphology, respectively. Any doublets or dead cells were excluded using the forward and side scatter data. Twelve bins were generated with a 43 channel gap to create non-adjacent bins allowing smaller cross contaminations between the bins. For the synthetic library (30°C) the following number of cells were sorted: bin 1, 1 million cells; bin 2-8, 250,000 cells; bin 9-10, 100,000 cells and bin 11-12, 25,000 cells whereby bin 1 reflected the lowest and bin 12 the highest GFP intensity. After the sorting, the individual bins were cultured in LB medium at room temperature overnight and in the morning they were incubated for further 8 h at 30°C. The cells of the bins were then subjected to FACScan (Section 2.14.2).

2.14.2 Fluorescence-activated cell scan (FACScan)

FACScan is based on flow cytometry, however, cells are only screened for their fluorescence activity. To analyse cells via FACScan, appropriate strains were pre-cultured for 3 h at 37°C from an overnight culture before they were diluted again and cultured until the required phase of growth. Cells (including the sorted cells grown at 30°C; Section 2.14.1) were pelleted and subsequently resuspended in ice-cold PBS, diluting the cells 100-fold to remove the remaining autofluorescence of the medium. For the scanning of 50,000 events, the LSRFortessa™ X-20 (BD Biosciences) was applied using the excitation/emission wavelength of 488/530-30-A for GFP and 561/610-20-A for mCherry and the forward and side scatter to measure volume and morphology, respectively. Any doublets or dead cells were excluded using the forward and side scatter data. Cultures which were waiting for their measurement were kept on ice to prevent further protein folding. For the evaluation of the results the programmes FACSDiva™, FlowJo and MATLAB (MathWorks®) were used.

Chapter 3

Physiology and development of *B. subtilis* during the expression of industrial enzymes

3.1 Introduction

This chapter is focused on the analysis of industrial enzyme expression and its effects on physiological and developmental processes in *B. subtilis*. *B. subtilis* has been the subject of intense basic and applied research for more than six decades. For example, the investigation of key metabolic pathways associated with the central carbon metabolism, glycolysis and gluconeogenesis, has provided information about cell physiology that is of importance in both academic and industrial contexts. The energy, generated during glycolysis and subsequent pathways not only powers cell growth and motility but also a range of other processes that optimise competitiveness and survival in its natural environment. A key adaptation of *B. subtilis* is its bet-hedging strategy (Veening *et al.*, 2008c); since the organism is not able to predict the outcome of any stressful situation, cells in the population differentiate into distinct physiological cell types, each with different survival potentials in different environments. In its natural environment *B. subtilis* responds to stress in a variety of ways. For example, in response to nutrient stress, chemotaxis and motility genes of the SigD regulon are induced with the aim of seeking new sources of nutrients (Guttenplan *et al.*, 2013). If this and other strategies such as the uptake of DNA (Dubnau, 1991a, Dubnau, 1991b) or production of extracellular proteases (Veening *et al.*, 2008a) are unsuccessful, the vegetative cell cycle is terminated and the sporulation cycle initiated, leading to the formation of highly resistant endospores (Errington, 2003, Higgins and Dworkin, 2012, Piggot and Hilbert, 2004). Each particular physiological cell type exists in a bistable relationship with other cells in the population so that two subpopulations are formed. In the case of motility, motile cells are single cells while sessile cells form chains (Kearns and Losick, 2005). This observation, in combination with the expression of a specific extracellular enzyme,

raises the question as to whether the motile cells produce more or less of the enzyme than the non-motile cells. Similarly, the ability to swarm and colonise might be advantageous as it allows *B. subtilis* to migrate over surfaces and be more competitive than other microorganisms co-existing in the same environment (Julkowska *et al.*, 2005, Ron and Rosenberg, 2001). In recent years' tools have been developed to address just such a question. For example, secretion stress is likely to reflect the amount and type of protein secreted by an individual cell. In response to secretion stress, cells induce the expression of genes encoding quality control proteases such as HtrA and HtrB. These proteins degrade misfolded proteins within the cell envelope (Noone *et al.*, 2012, Trip *et al.*, 2011). The sigma factor B (SigB), regulated via an anti-sigma factor and signal transduction cascade, activates about 150 stress genes in response to general stress (Hecker *et al.*, 2007). By monitoring both secretion stress and general stress, the impact of industrial enzyme production can be determined in a population on a cell-by-cell basis. If stress responses are unsuccessful, competence genes and sporulation genes are induced, the former an attempt to increase metabolic versatility, the latter to generate quiescent resistant cells for long-term survival (Dubnau, 1991a, Dubnau, 1991b, Errington, 2003, Higgins and Dworkin, 2012, Piggot and Hilbert, 2004). In each case only a subpopulation of the cells respond, leading to a phenotypically heterogeneous population (Dubnau and Losick, 2006, Smits *et al.*, 2005, Gamba *et al.*, 2015, Veening *et al.*, 2005). For industry both competence and sporulation are processes that need to be avoided in industrial processes as they are energetically costly and, in case of sporulation, irreversible leading to reduced productivity in up-scaled fermentations.

In an industrial context, it is important to understand exactly how *B. subtilis* responds to high-level industrial enzyme production; specifically, to understand which processes are controlled in a bistable manner during the production phase. While traditionally such questions have been addressed at a whole population level, technologies now exist to address this at the level of individual cells in the population. Therefore, a number of physiological pathways were selected to investigate *B. subtilis* under production-like conditions with two approaches; firstly, by assaying xylanase and amylase activities in industrial-like media while monitoring the physiology on a population level, and secondly, by analysing *B. subtilis* on a single-cell level to determine the degree of heterogeneity of physiological processes.

3.2 Results and discussion

3.2.1 Production of industrial enzymes in a microbioreactor

In industrial fermentations, for any given strain, the yield of extracellular enzymes is dependent on the composition of the growth medium and conditions. Highly complex media are designed to achieve high cell densities, although the presence of other proteins in the culture medium lowers the purity of the target enzyme product and increases downstream costs. In contrast, chemically defined media enriched with yeast extract and various carbon sources support lower production costs. We therefore analysed enzyme production under industry-like conditions, in two types of media: a rich medium (ROCHE®, EP0405370 A1) and a chemically-defined medium enriched with glucose and yeast extract (Section 2.2).

To analyse the influence of culture medium on enzyme productivity in *B. subtilis* under production-like conditions two industrial enzymes were used: XynA, a native *B. subtilis* xylanase, and AmyM, a heterologous α–amylase from *Geobacillus stearothermophilus* (Figure 3.1). The gene encoding XynA (x*ynA*) was integrated into the chromosome at the *nprE* locus with its expression being driven by the strong and constitutive promoter P_{15}. Because several attempts to clone the *amyM* gene onto the chromosome were unsuccessful, this gene was expressed from the P_{amyQ} promoter via an autonomously replicating vector pCS73, based on the pUB110 replicon and with a copy number of around 48 ± 2 per cell (Leonhardt, 1990). In the case of XynA, the negative control was the empty plasmid pCS31 that was integrated at the *nprE* locus. For AmyM, the self-replicating plasmid pCS74 was used. The approximate sizes of the secreted proteins are ~20.5 kDa for XynA and ~75.4 kDa for AmyM.

A microbioreactor (BioLector®) was used to analyse growth and enzyme activities of strains encoding XynA and AmyM in rich and defined medium (Figure 3.1). The microbioreactor was set up with 48-well FlowerPlates® with 1-ml culture volumes and shaking at 900 rpm at 37°C with high humidity (92-95%) to reduce evaporation of the medium (Section 2.9). The gains for biomass were altered for the particular medium, because of instrument limitations, leading to gains of 10/60 (biomass/GFP) in rich medium and 20/100 in chemically-defined medium. The differences were calculated by growing cells in both media with the respective gains to equalise prior achieved values.

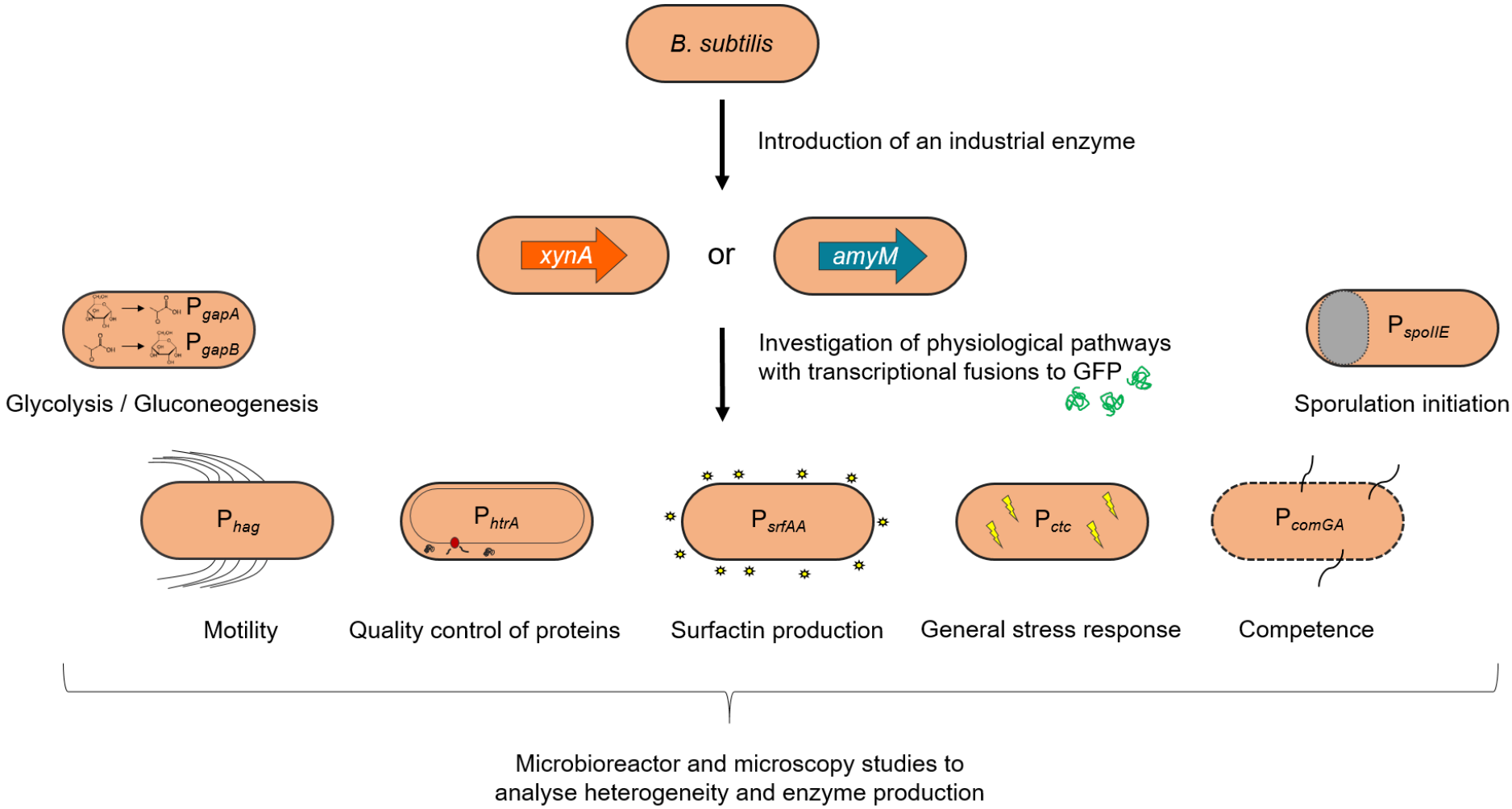


Figure 3.1: Overview of the procedure for the investigation of the physiology and development of *B. subtilis* during the production of an industrial enzyme. The genes encoding, XynA and AmyM were introduced into *B. subtilis* W168 under the control of constitutive promoters. In the resulting strains, selected promoter regions were transcriptionally fused to GFP to monitor physiological processes. To analyse the mechanisms on a population and single cell level under the influence of an expressed enzyme, studies in the microbioreactor and subsequent microscopy studies were performed.

In rich medium the cells displayed a biphasic exponential growth phase, with the second phase slower and shorter than the first (Figure 3.2A). This biphasic growth behaviour is likely to be due to the release and reuse of overflow metabolites; the glycolytic utilisation of glucose in the medium (measured with Combur^4 Test® N) leads to the production of pyruvate and acetyl CoA and by-products such as lactate, acetoin and acetate (Sonenshein, 2007). These compounds help to restore NAD^+ by the conversion of pyruvate to lactate and to the synthesis of ATP by forming acetate from acetyl CoA. Only after the full consumption of glucose these fermentation products are further metabolised *via* the citric acid cycle to generate ATP (Sonenshein, 2007). In defined medium no such biphasic growth behaviour was observed (Figure 3.2B).

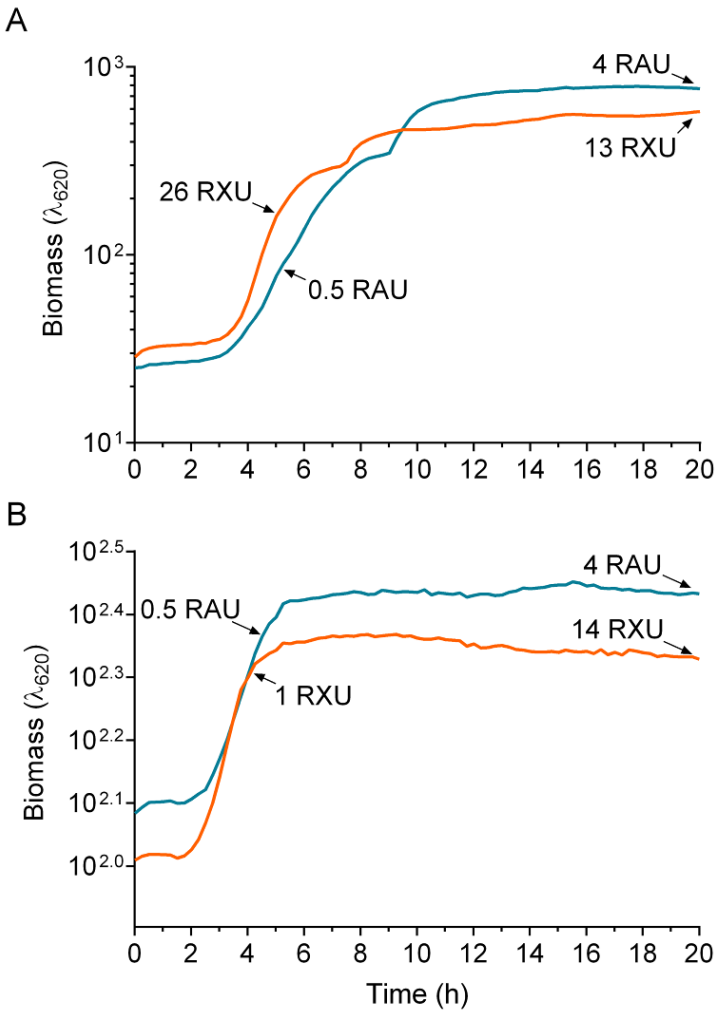


Figure 3.2: Growth and enzyme production of strains expressing XynA and AmyM in rich and defined medium. Two samples, in exponential and stationary phase, were taken in rich (A) and defined medium (B) to identify the activities of both enzymes. The activities were divided by the biomass to reflect the activity per unit biomass and shown on the graph: RXU indicates relative XynA activity (orange) and RAU relative AmyM activities (blue). Arrows indicate time points at which the assays of four biological replicates were performed.

Samples were taken during exponential phase (4-5 h) and stationary phase (20 h). Enzyme activities were assayed using the EnzChek® Ultra Xylanase Assay Kit for XynA and the Phadebas® Amylase test for AmyM as described earlier (Section 2.10). The enzyme activities of negative controls, with no expression construct, were subtracted from that of cells with the expression construct. The data reveal differences in the kinetics of XynA or AmyM synthesis (Figure 3.2). In rich medium, the highest relative activity of XynA (26 RXU) was observed in exponential phase, declining by ~50% at late stationary phase. In contrast, the activity of AmyM was ~8-fold higher in stationary phase than in exponential phase. In defined medium the strains encoding both enzymes showed the same growth and activity dynamics; in late exponential / transition phase the enzyme activity in the culture medium was low, but increased significantly in stationary phase.

To investigate the difference in the behaviour of strains expressing XynA and AmyM in rich medium, the enzyme activities were determined throughout a 24 h growth experiment (Figure 3.3). Cells carrying the empty construct were used to monitor native enzyme production. The data revealed an oscillating pattern of XynA activity. In the growth phase (1-7 h) the XynA activity reached 76 RXU and further increased to 110 RXU between 8 h and 14 h (Figure 3.3A). In contrast, the relative activity decreases after this peak to ~37 RXU. This indicates that despite being a native *B. subtilis* enzyme, it is susceptible to the minor proteases release into the culture medium and during stationary phase the rate of XynA degradation might be higher than its rate of synthesis. Whilst the most productive phase for XynA synthesis was during exponential phase, that of AmyM was during stationary phase, with a significant level of synthesis being observed from 14 h onwards (Figure 3.3B). The AmyM activity reached a maximum at ~21 h and despite being a heterologous enzyme, retains stable in the growth medium. The differences in the dynamics of XynA and AmyM synthesis lead to questions about why AmyM synthesis or secretion or both are delayed, despite both constructs using constitutive promoters that are generally active during both exponential and stationary phase these differences in expression were studied in chapter 4.

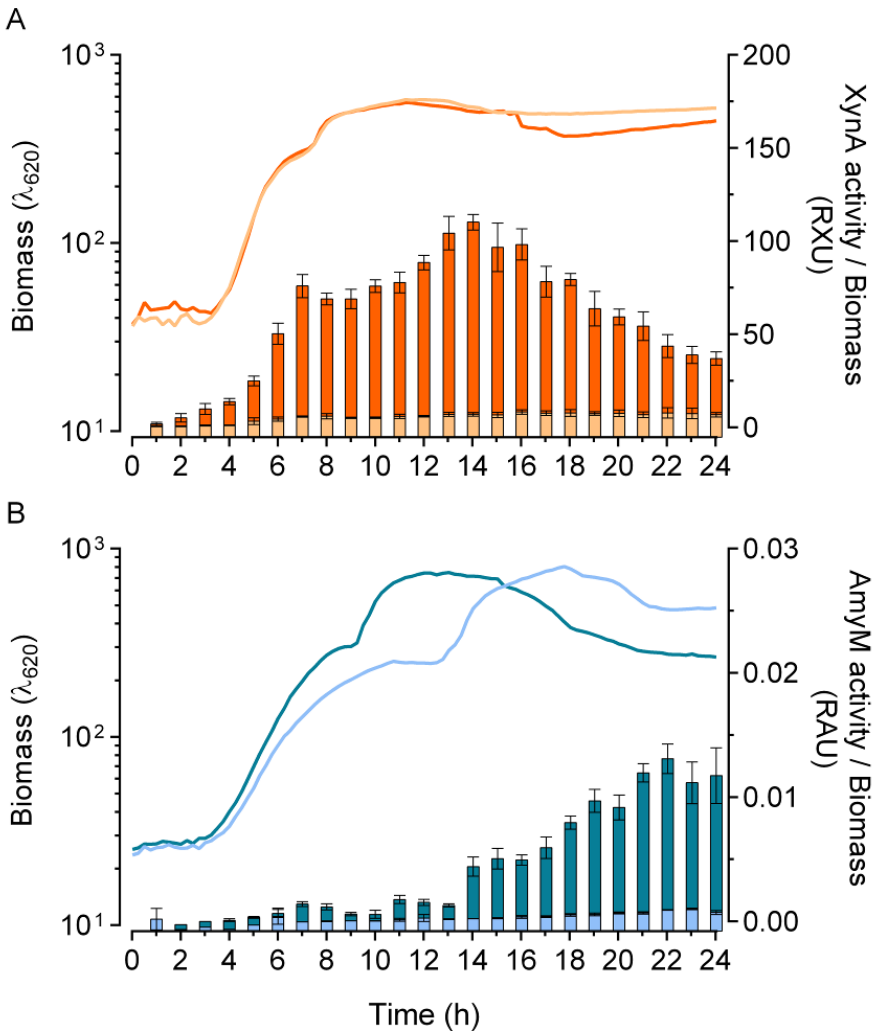


Figure 3.3: Dynamics of 24 h of growth and enzyme production by strains expressing XynA and AmyM in rich medium. The activities were divided by the biomass to achieve an activity per cell. (A) Growth curves and enzyme activities (RXU) of strains producing XynA (dark) and the negative control (light) are shown. (B) Growth curves and enzyme activities (RAU) of strains producing AmyM (dark) and the negative control (light) are displayed. Error bars indicate the standard deviation of three biological replicates.

3.2.2 Investigation of physiological pathways

The previous section showed that the kinetics of production and release of XynA and AmyM were different, despite their genes being transcribed from strong constitutive promoters. We therefore decided to investigate the effect of promoters associated with a range of different physiological and developmental pathways under production-like conditions. Eight promoters (P_{gapA}, P_{gapB}, P_{hag}, P_{htrA}, P_{srfAA}, P_{ctc}, P_{comGA}, P_{spoIIE}) were chosen and transcriptionally fused to a gene encoding a monomeric, superfolding green fluorescent protein, and integrated into the *aprE* locus (Figure 3.1) (Pedelacq *et al.*, 2006). To ensure the inclusion of all the regulatory elements, the promoter regions of the relevant transcriptional units were amplified on a fragment ~400 bp upstream from the start codon. In addition, the first

36 nucleotides of the *B. subtilis comGA* gene were added to improve translation initiation and the native ribosome binding site (RBS) was replaced with one known to be efficient at initiating translation (Veening *et al.*, 2004, Vellanoweth and Rabinowitz, 1992). The resulting constructs were sensitive enough to identify even low promoter activities. Finally, this strategy led to the disruption of both *nprE* (empty plasmid) and *aprE* genes to mimic industry-like *B. subtilis* strains that are engineered to prevent the synthesis of the neutral protease B and subtilisin E, which account for more than 90% of the protease activity in the culture medium (Trip *et al.*, 2011, Wu *et al.*, 1991).

For the analysis, the cells were cultured in the BioLector® monitoring GFP fluorescence through growth. The level of fluorescence reflects the level of activity of the respective physiological process in the population as a whole. Both rich and defined media were used to determine activity differences based on the medium. In the XynA expressing strains in rich medium, glycolysis (P_{gapA}) was turned on during exponential phase (between 4-7 h) in response to the presence of adequate concentrations of glucose in the medium for conversion into ATP and reducing power (Figure 3.4A). As soon as glucose was consumed, the genes for glycolysis (P_{gapA}) were switched off and those for gluconeogenesis (P_{gapB}) were switched on (8-10 h) to generate glucose from lactate and other by-products to provide cells with alternative carbon and energy sources (Figure 3.4B). P_{hag} (motility) was activated at the same time as P_{gapA} but at a lower level and switched off after ~8 hours (Figure 3.4C). The genes for surfactin expression (P_{srfAA}) were active in exponential phase (5-8 h). Although the surfactin promoter is induced, the presence of a frameshift mutation in strain 168 means that no surfactin is synthesised (Nakano *et al.*, 1992). The processes of quality control of proteins (P_{htrA}), competence (P_{comGA}) and general cell stress (P_{ctc}) were not activated in rich medium (Figures 3.4D, F & G). Interestingly, sporulation (P_{spoIIE}) was initiated in stationary phase (~12 h) skipping the development of competence and leading directly to the formation of endospores (Figure 3.4H). In the case of chemically-defined medium, with glucose as the main carbon source, the glycolysis genes (P_{gapA}) were not only activated in exponential phase but also remained active in late stationary phase compared with the off-state in rich medium (Figure 3.5A), while the gluconeogenesis gene (P_{gapB}) remained turned off. This indicated that glucose was in excess through growth and that *B. subtilis* converted by-products that were accumulated during the degradation of glucose via the overflow metabolism (Sonenshein, 2007).

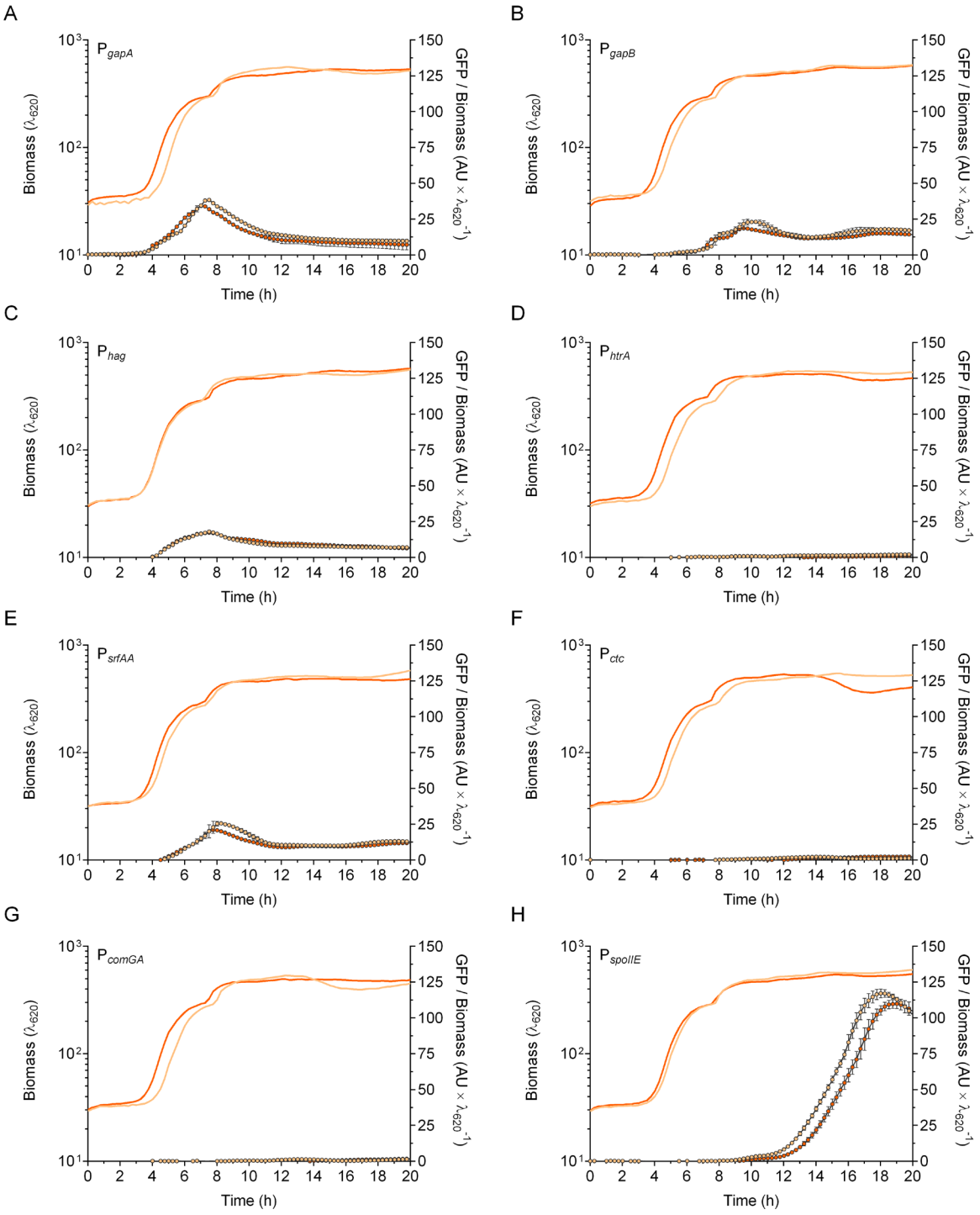


Figure 3.4: Physiological and developmental responses of cells in response to XynA production in rich medium. Eight promoters were selected to monitor the physiology and development of *B. subtilis via* the transcriptional fusion to GFP (A) P_{gapA} (glycolysis), (B) P_{gapB} (gluconeogenesis), (C) P_{hag} (motility), (D) P_{htrA} (quality control of proteins), (E) P_{srfAA} (surfactin expression), (F) P_{ctc} (general cell stress), (G) P_{comGA} (competence), (H) P_{spoIIE} (sporulation). Biomass (smooth lines) and relative promoter activity (GFP per unit biomass; circles) were determined in isogenic strains with (dark) and without (light) the gene encoding XynA. Error bars indicate the standard deviation of three biological replicates.

This energy generating process (glycolysis) is likely to prevent the initiation of sporulation (Figure 3.5H) compared to that in rich medium. Motility genes (P_{hag}) were activated during exponential phase with an increasing GFP activity towards late stationary phase (Figure 3.5C). As motility is an energy intensive process, the correlation between P_{gapA} and P_{hag} (Pearson correlation coefficient $R^2 = 0.96$) implies a requirement for energy provided by glycolysis. The reporter for general stress (P_{ctc}) showed an increasing activity during the transition phase as the cells prepare for unfavourable conditions (Banse *et al.*, 2008). Both P_{htrA} (quality control) and P_{comGA} (competence) were switched off during growth. In general, the fluorescence dynamics of XynA-expressing and control strains were comparable.

The data indicates that the production of XynA, a protein native to *B. subtilis*, does not markedly affect the expression of the eight promoters chosen to monitor cell physiology and development. In contrast, to the data of the XynA-expressing cells, the strains expressing AmyM showed significant differences in the GFP dynamics in both media compared with the non-producing strain (Figure 3.6 & 3.7). In rich medium the promoter used to monitor the transcription of the quality control of proteins (P_{htrA}) was highly induced in the AmyM producing cells but not in the non-producing control strain (Figure 3.6D). Similar results were observed in defined medium, although the extent of the induction was considerably lower. Taken together, the data indicate that the production of AmyM, a heterologous protein, exerts a greater burden on the cell, resulting in secretion stress and the induction of the CssR regulon that activates *htrA* and *htrB* (Trip *et al.*, 2011, Westers *et al.*, 2004, Westers *et al.*, 2006).

In rich medium, motility (P_{hag}), glycolysis (P_{gapA}), gluconeogenesis (P_{gapB}) and surfactin expression (P_{srfAA}) were all turned on in exponential phase and noticeable differences between non-producing and AmyM-producing cells were observed in surfactin expression (Figure 3.6). Sporulation was turned on in the non-producing cells in the late stationary phase (~17 h) with a low expression and without the induction in the AmyM producer cells (Figure 3.6H). General cell stress (P_{ctc}) and competence (P_{comGA}) showed low to no levels of expression.

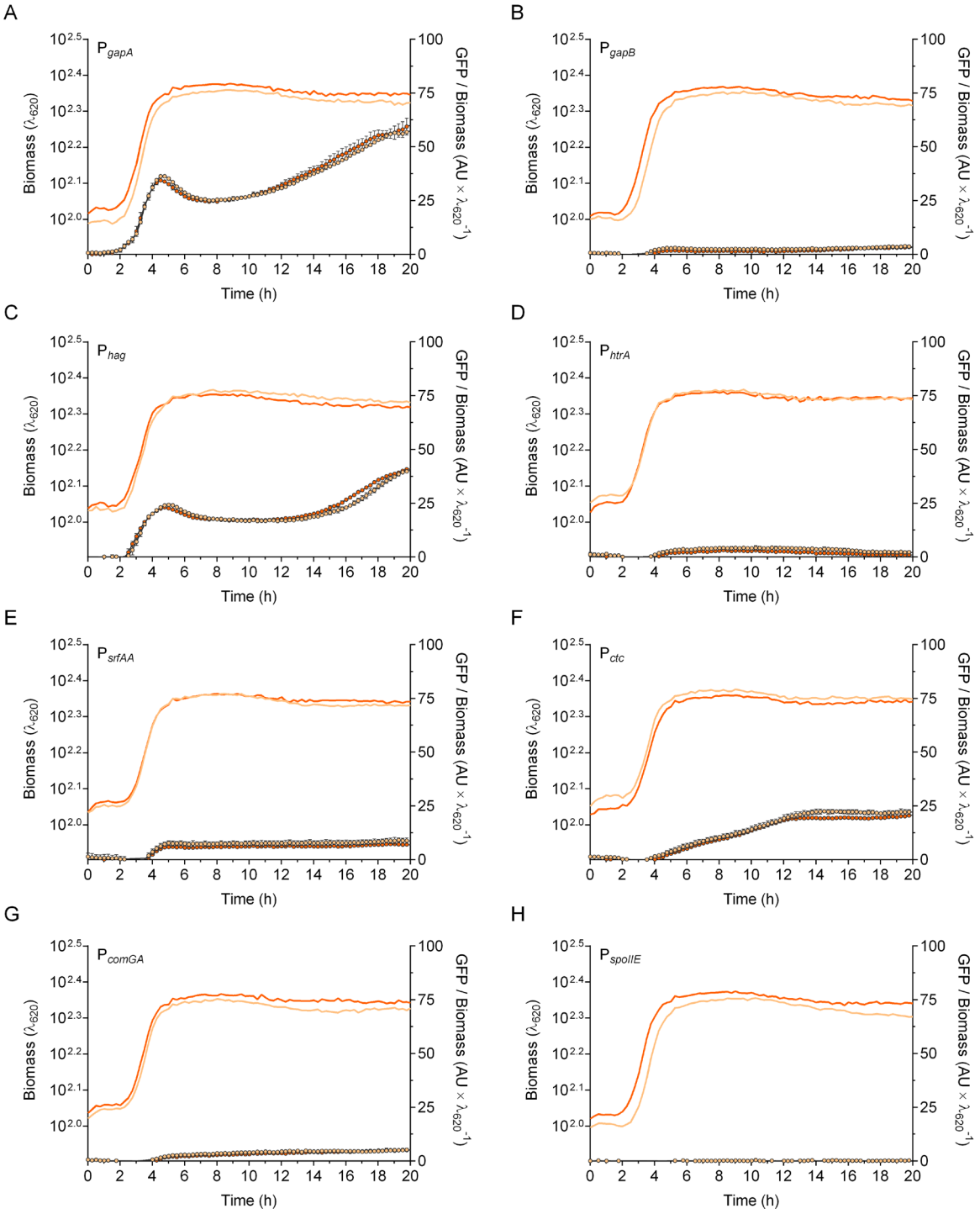


Figure 3.5: Physiological and developmental responses of cells in response to XynA production in defined medium. Eight promoters were selected to monitor the physiology and development of *B. subtilis via* the transcriptional fusion to GFP (A) P_{gapA} (glycolysis), (B) P_{gapB} (gluconeogenesis), (C) P_{hag} (motility), (D) P_{htrA} (quality control of proteins), (E) P_{srfAA} (surfactin expression), (F) P_{ctc} (general cell stress), (G) P_{comGA} (competence), (H) P_{spoIIE} (sporulation). Biomass (smooth lines) and relative promoter activity (GFP per unit biomass; circles) were determined in isogenic strains with (dark) and without (light) the gene encoding XynA. Error bars indicate the standard deviation of three biological replicates.

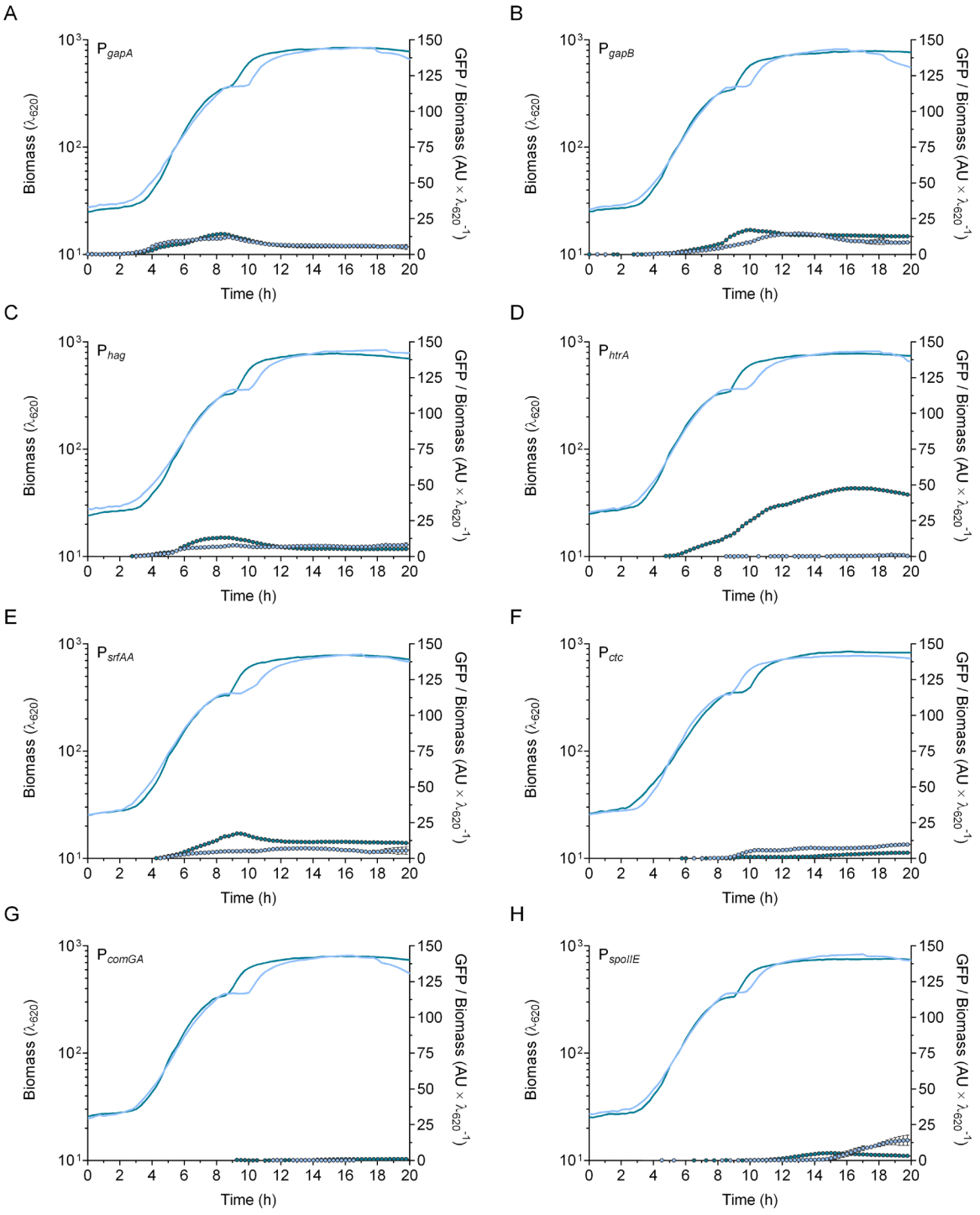


Figure 3.6: Physiological and developmental responses of cells in response to AmyM production in rich medium. Eight promoters were selected to monitor the physiology and development of *B. subtilis via* transcriptional fusion to GFP (A) P_{gapA} (glycolysis), (B) P_{gapB} (gluconeogenesis), (C) P_{hag} (motility), (D) P_{htrA} (quality control of proteins), (E) P_{srfAA} (surfactin expression), (F) P_{ctc} (general cell stress), (G) P_{comGA} (competence), (H) P_{spoIIE} (sporulation). Biomass (smooth lines) and relative promoter activity (GFP per unit biomass; circles) were determined in isogenic strains with (dark) and without (light) the gene encoding AmyM. Error bars indicate the standard deviation of three biological replicates.

In contrast to rich medium, general stress was upregulated in defined medium in AmyM producing cells in late growth from 7 hours onwards (Figure 3.7F). Additionally, cells producing AmyM were more active for glycolysis and motility during exponential and stationary phase than non-producing cells (Figures 3.7A & C) perhaps providing an explanation for the high stress in the producer strains as the cells are more glycolytic and moving around seeking new nutrients. Similar to the strains carrying the *xynA* gene, the GFP expression profiles relating to glycolysis and motility overlap (R^2 = 0.97) which further supports the hypothesis of the direct connection of motility and glycolytic processes, presumably reflecting changes in core carbon metabolism and the burden on the cells resulting from the production and secretion of AmyM. Competence was activated on a low level in stationary phase, potentially a problem in large-scale fermentations where such processes reduce the productivity of the protein of interest (Figure 3.7G). Surfactin expression was activated on a low level in control and producer strains (Figure 3.7E) while similar to the strains encoding XynA, gluconeogenesis and sporulation (Figures 3.7B & H) were not active during fermentation (Figures 3.5B & H).

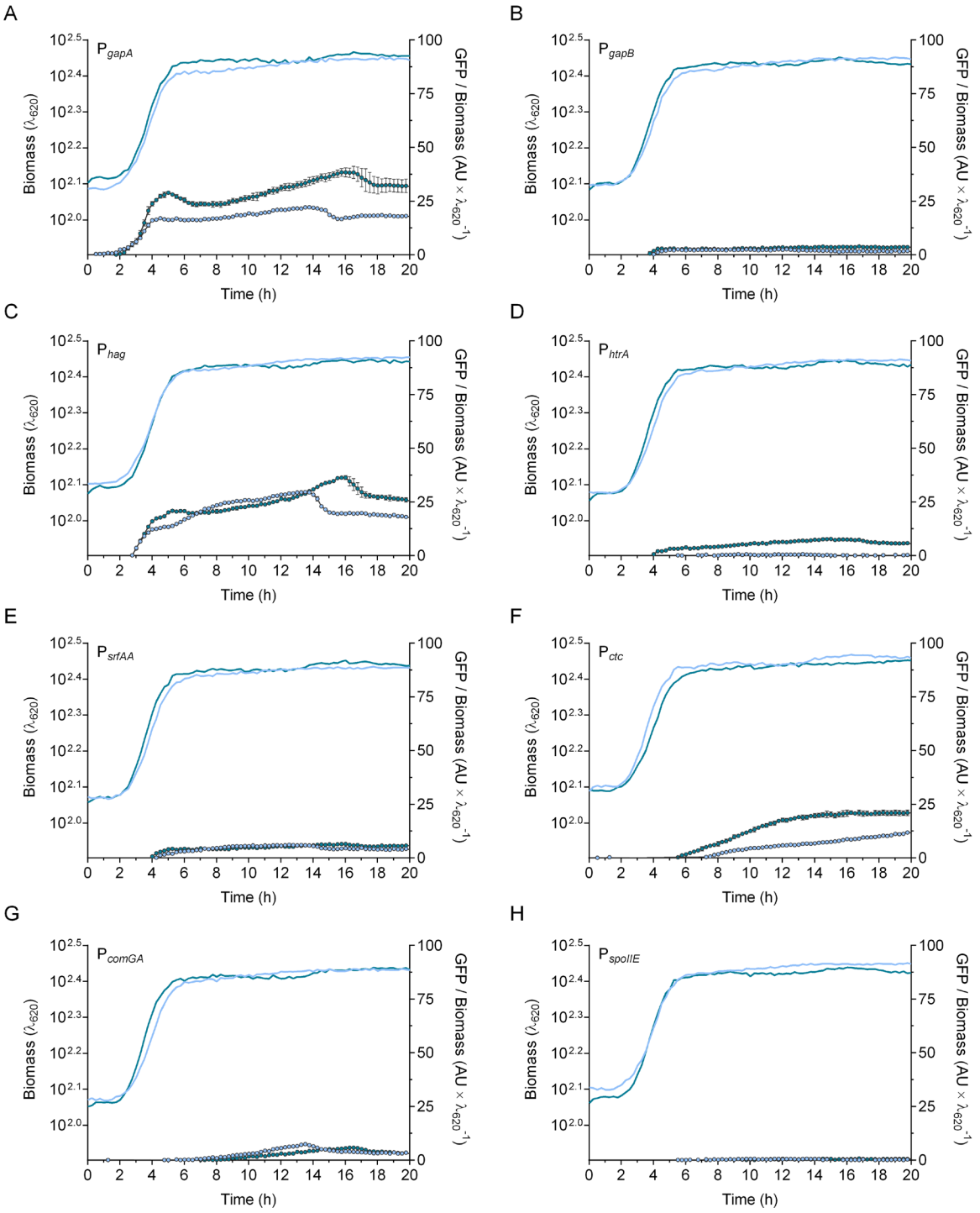


Figure 3.7: Physiological and developmental responses of cells in response to AmyM production in defined medium. Eight promoters were selected to monitor the physiology and development of *B. subtilis via* transcriptional fusion to GFP (A) P_{gapA} (glycolysis), (B) P_{gapB} (gluconeogenesis), (C) P_{hag} (motility), (D) P_{htrA} (quality control of proteins), (E) P_{srfAA} (surfactin expression), (F) P_{ctc} (general cell stress), (G) P_{comGA} (competence), (H) P_{spoIIE} (sporulation). Biomass (smooth lines) and relative promoter activity (GFP per unit biomass; circles) were determined in isogenic strains with (dark) and without (light) the gene encoding AmyM. Error bars indicate the standard deviation of three biological replicates.

3.2.3 Analysis of heterogeneity during enzyme expression

In Section 3.2.2, the activities of eight promoters associated with cell physiology were analysed at the population level. In this section the activities of seven of these promoters, namely P_{gapA}, P_{gapB}, $P_{hag,}$ P_{htrA}, P_{srfAA}, P_{ctc} and P_{comGA} were analysed at the single cell level using fluorescence microscopy. The aim was to determine promoter activity and cell heterogeneity in response to industrial enzyme production. Fluorescence microscopy additionally is able to provide information on cell morphology and viability. Cultures were grown in either rich or defined media in a BioLector® microbioreactor using FlowerPlates® with comparable growth kinetics to that observed in the Figures 3.4-3.7. Fluorescence microscopy was performed on cells during exponential and stationary phase. The fluorescent signal was determined from at least 140 cells and analysed using the ObjectJ plugin within ImageJ (Section 2.6) (Syvertsson *et al.*, 2016).

To investigate the impact of XynA production, the expression of four promoters P_{gapA}, P_{gapB}, P_{hag} and P_{srfAA} was analysed. As can be seen in the representative images in the Figures 3.8A & B, irrespective of the production of XynA, when grown on rich medium, all four promoters were active, albeit to very different extents. P_{gapA} (glycolysis) and P_{hag} (motility) gave the strongest GFP signals in both exponential and stationary phase (Figures 3.8 & 3.9). In contrast, P_{gapB} (gluconeogenesis) and P_{srfAA} (surfactin expression) were only weakly expressed. The data correlates well with the previous data in which *gapA* and *gapB* are differentially expressed in relation to the need for glycolysis and gluconeogenesis (Figure 3.4). The distribution of the signal strengths in the cells indicates a clear Poissonian distribution in the case of three of the four promoter analysed (Figures 3.8 & 3.9), namely P_{gapA}, P_{gapB}, and P_{srfAA}. The exception was $P_{hag,}$ in which there appears to be two subpopulations, presumably reflecting separate populations of motile and non-motile cells (Cozy and Kearns, 2010). It is striking that the production of XynA had little impact on either the levels of expression or the distribution of the signal strengths in both exponential and stationary phase.

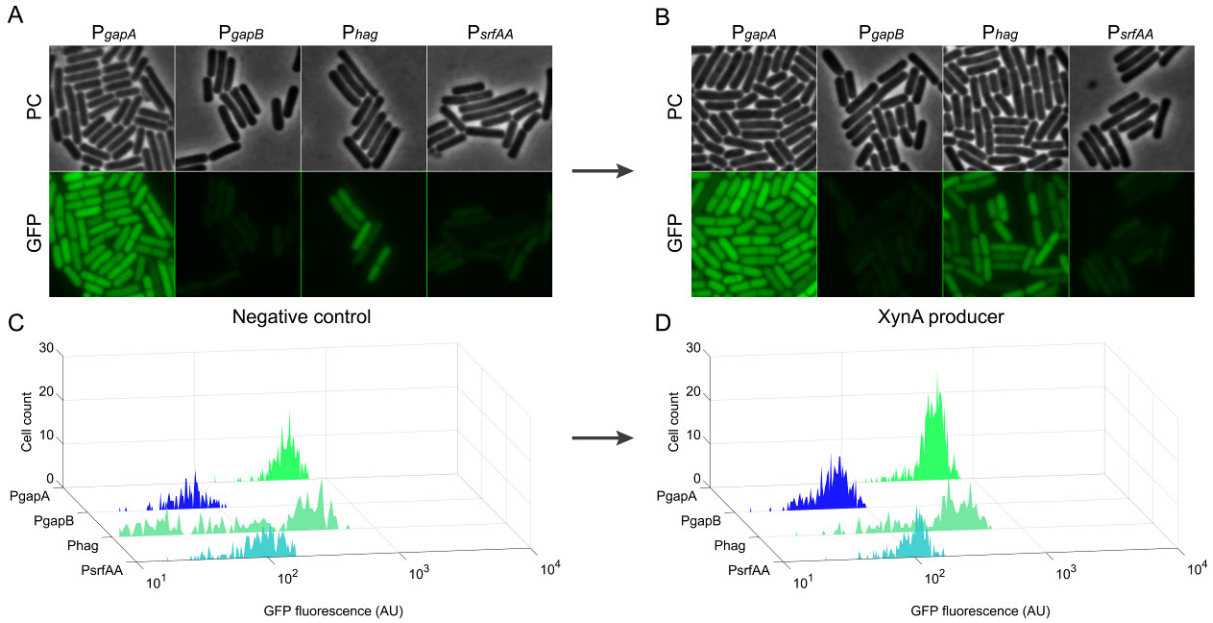


Figure 3.8: Fluorescence microscopy of XynA producing strains in rich medium during exponential phase. The strains were analysed after 5-6 h of growth in non-producing (A, C) and XynA producing strains (B, D). (A, B) Phase contrast and GFP images show the activities of the selected promoters P_{gapA}, P_{gapB}, P_{hag} and P_{srfAA}. (C, D) Fluorescence distributions of the promoter activities were determined with the plugin ObjectJ in ImageJ identifying the GFP signal at the single cell level.

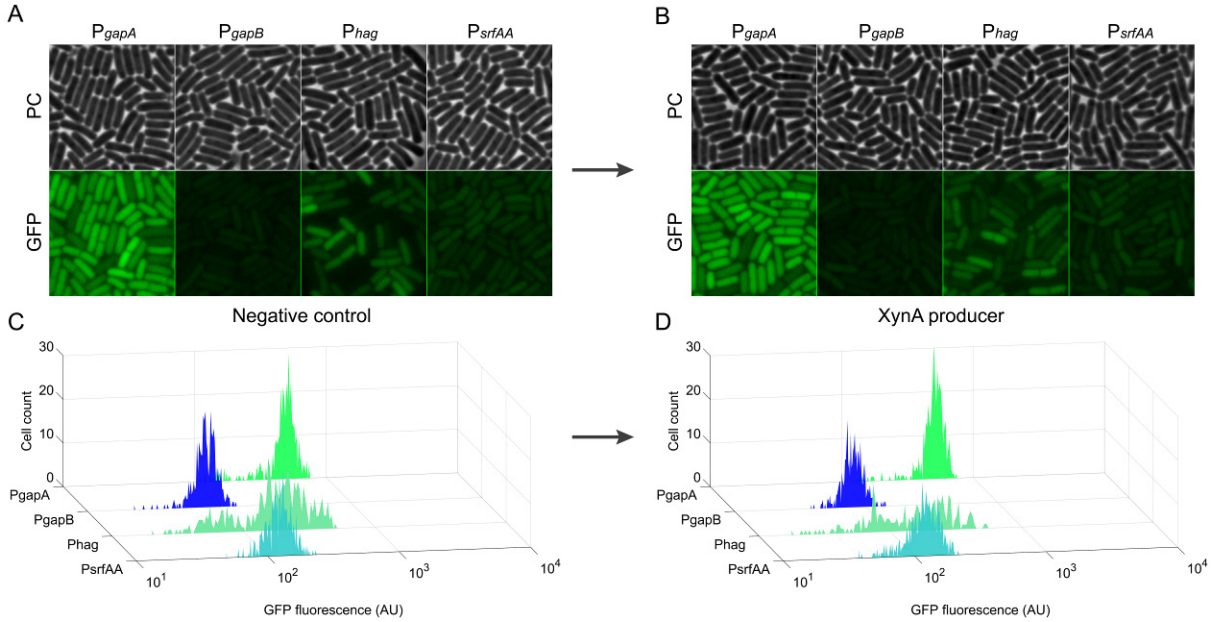


Figure 3.9: Fluorescence microscopy of XynA producing strains in rich medium during stationary phase. The strains were analysed after 16-17 h of growth in non-producing (A, C) and XynA producing strains (B, D). (A, B) Phase contrast and GFP images show the activities of the selected promoters P_{gapA}, P_{gapB}, P_{hag} and P_{srfAA}. (C, D) Fluorescence distributions of the promoter activities were determined with the plugin ObjectJ in ImageJ identifying the GFP signal at the single cell level.

The experiments were repeated in defined medium but with the addition of a strain with P_{ctc}. The pattern of promoter activities and signal distributions were broadly similar (Figures 3.10 & 3.11) to those observed in rich medium except that P_{hag} showed a more normal distribution with only a thin tail in exponential phase (Figure 3.10) and virtually no tail in stationary phase (Figure 3.11), indicating that the majority of the cells were likely to have flagella. In addition, the P_{srfAA} and P_{ctc} promoters were more active in stationary phase. Interestingly, single cells showed gluconeogenic processes (P_{gapB}) responding to low levels of glucose in the medium supporting the survival of those single cells. This is contrary to the results obtained with the microbioreactor showing activity of P_{gapB} in rich but not defined medium (Figures 3.4B & 3.5B). As with the cells grown in rich medium, no impact of XynA production on the expression of these promoters was observed.

Next, we studied the impact of synthesising the heterologous AmyM protein on the same set of promoters. In contrast, to the data obtained for XynA production, cells actively synthesising AmyM showed more differences at the single cell level. The reporter for quality control of proteins (P_{htrA}) showed the most significant differences between control and cells actively synthesising AmyM. In both growth media the expression of P_{htrA} was switched on and showed a high level of heterogeneity in the AmyM producer strains reflecting secretion stress triggered by the heterologous enzyme (Figures 3.12-3.15).

In rich medium, motility was activated in exponential phase with a heterogeneous pattern (Figure 3.12), however, this pathway was barely active during stationary phase (Figure 3.13) similar to the results from the microbioreactor experiment (Figure 3.6C). The GFP distributions of P_{srfAA} and P_{gapB} were heterogeneous in rich medium with no difference between control and AmyM producer strains but P_{gapB} was switched on in stationary phase, reflecting a response to starving conditions (Figure 3.13) (de Jong *et al.*, 2012). A more heterogeneous expression of P_{gapA} (glycolysis) was observed in the control strain in stationary phase with cells being inactive for glycolytic processes indicating a different physiology than in the producer strains.

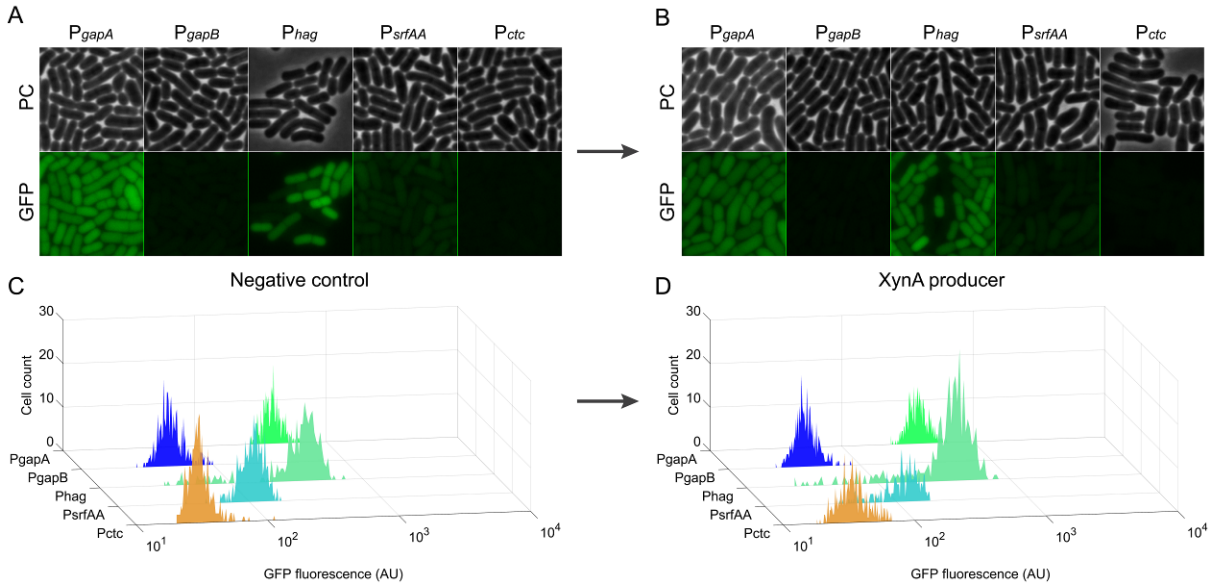


Figure 3.10: Fluorescence microscopy of XynA producing strains in defined medium during exponential phase. The strains were analysed after 5-6 h of growth in non-producing (A, C) and XynA producing strains (B, D). (A, B) Phase contrast and GFP images show the activities of the selected promoters P_{gapA}, P_{gapB}, P_{hag}, P_{srfAA} and P_{ctc}. (C, D) Fluorescence distributions of the promoter activities were determined with the plugin ObjectJ in ImageJ identifying the GFP signal at the single cell level.

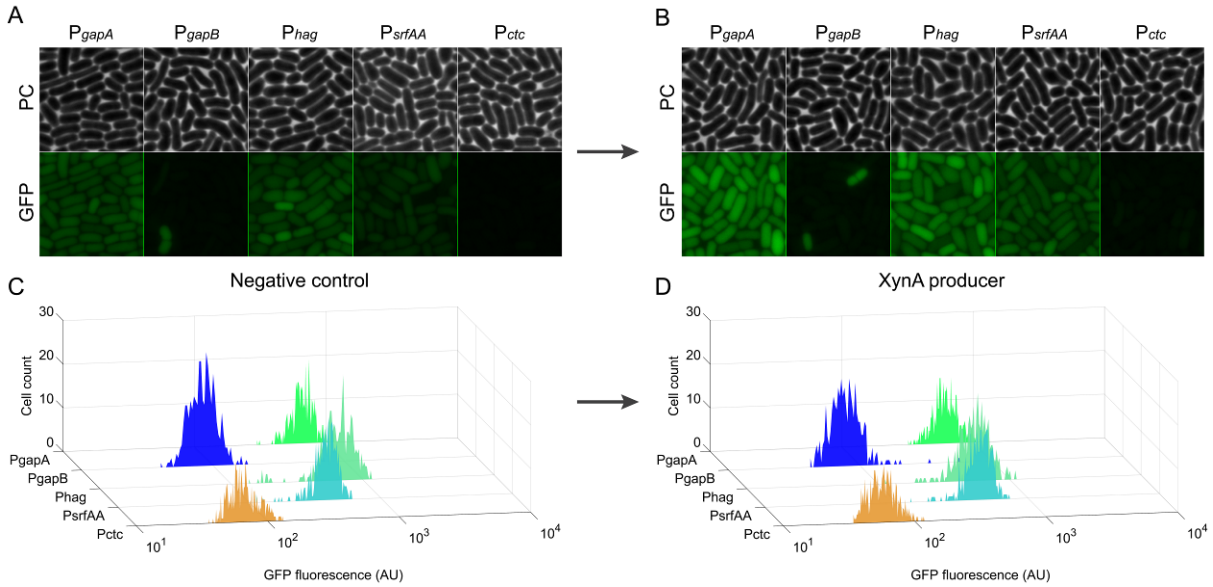


Figure 3.11: Fluorescence microscopy of XynA producing strains in defined medium during stationary phase. The strains were analysed after 16-17 h of growth in non-producing (A, C) and XynA producing strains (B, D). (A, B) Phase contrast and GFP images show the activities of the selected promoters P_{gapA}, P_{gapB}, P_{hag}, P_{srfAA} and P_{ctc}. (C, D) Fluorescence distributions of the promoter activities were determined with the plugin ObjectJ in ImageJ identifying the GFP signal at the single cell level.

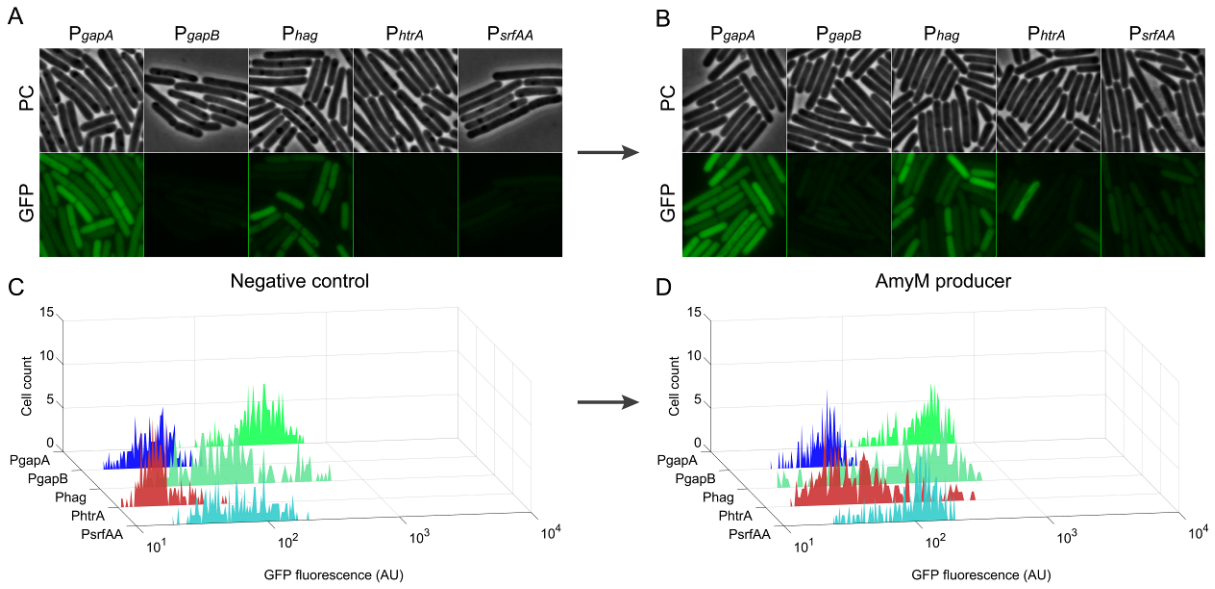


Figure 3.12: Fluorescence microscopy of AmyM producing strains in rich medium during exponential phase. The strains were analysed after 5-6 h of growth in non-producing (A, C) and AmyM producing strains (B, D). (A, B) Phase contrast and GFP images show the activities of the selected promoters P_{gapA}, P_{gapB}, P_{hag}, P_{htrA} and P_{srfAA}. (C, D) Fluorescence distributions of the promoter activities were determined with the plugin ObjectJ in ImageJ identifying the GFP signal at the single cell level.

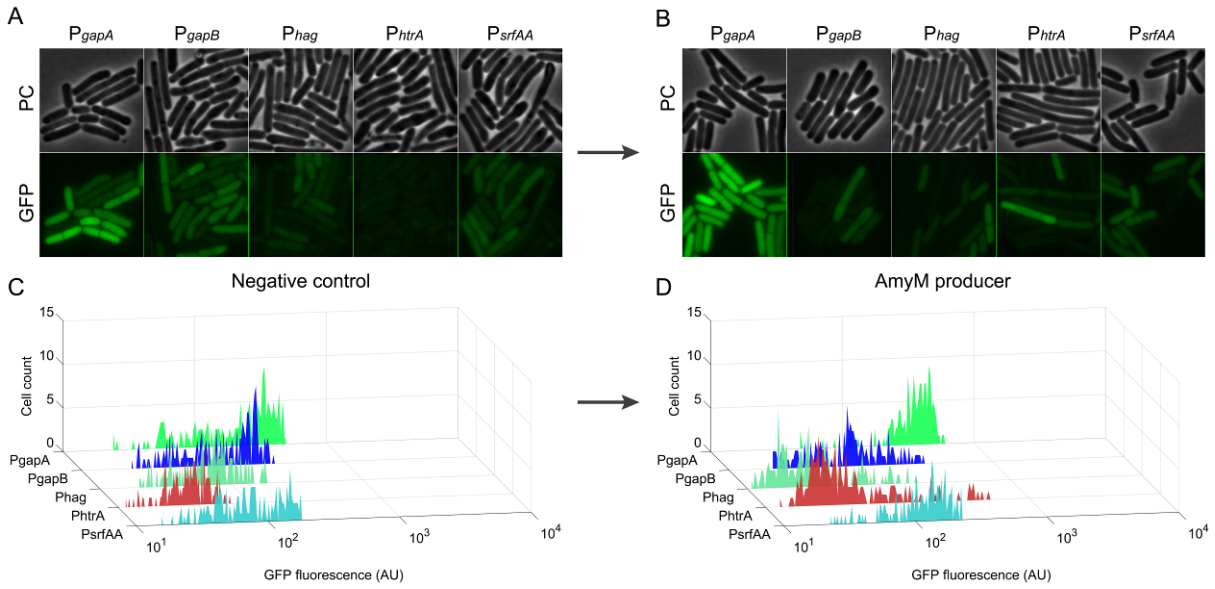


Figure 3.13: Fluorescence microscopy of AmyM producing strains in rich medium during stationary phase. The strains were analysed after 16-17 h of growth in non-producing (A, C) and AmyM producing strains (B, D). (A, B) Phase contrast and GFP images show the activities of the selected promoters P_{gapA}, P_{gapB}, P_{hag}, P_{htrA} and P_{srfAA}. (C, D) Fluorescence distributions of the promoter activities were determined with the plugin ObjectJ in ImageJ identifying the GFP signal at the single cell level.

In contrast to rich medium, repeating the experiments in defined medium with the addition of a strain with P_{comGA}, cells were still motile in stationary phase, as indicated by the expression of P_{hag} (Figure 3.14). This correlates with the expression data of P_{hag} that were observed in the *xynA* expressing strains in defined medium (Figure 3.11) and the BioLector® data (Figures 3.5C & 3.7C). While performing microscopy, the cells with the empty vector looked sick compared to the *amyM* expressing strain (Figure 3.15A & B). Similar to the BioLector® data (Figure 3.7A), a lower expression of P_{gapA} was observed in the control strain during stationary phase (Figure 3.15D). The surfactin expression was increased in the AmyM producing strain indicating a need to explore for nutrients. Interestingly, both expressing and control strains showed single cells that were expressing competence (P_{comGA}) to take up DNA from the environment, which is consistent with the findings in the microbioreactor (Figure 3.7G). However, there was a difference in the observed activity for P_{ctc} (general stress) by microscopy and the microbioreactor, the latter showing a high stress signal for the *amyM* expressing strains. This discrepancy could relate to the sampling time. The production of AmyM had impact on the levels of expression, mostly on the quality control of proteins showing a widely distributed GFP signal in exponential and stationary phase and in both rich and defined medium. A stronger induction was also observed for the surfactin expression and glycolytic processes in the AmyM producer strains in stationary phase.

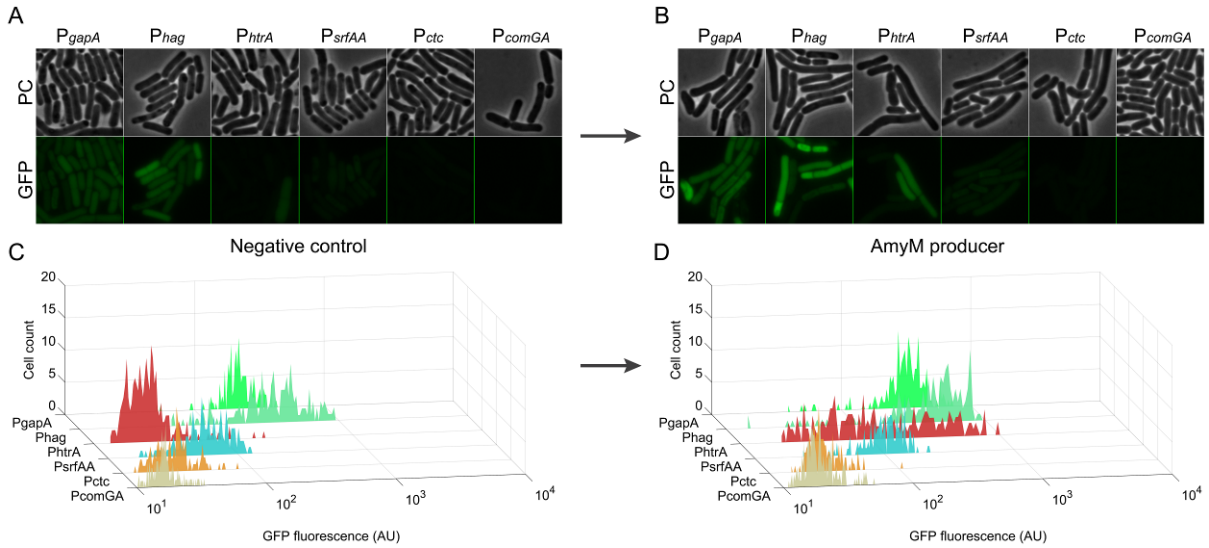


Figure 3.14: Fluorescence microscopy of AmyM producing strains in defined medium during exponential phase. The strains were analysed after 5-6 h of growth in non-producing (A, C) and AmyM producing strains (B, D). (A, B) Phase contrast and GFP images show the activities of the selected promoters P_{gapA}, P_{hag}, P_{htrA}, P_{srfAA}, P_{ctc} and P_{comGA}. (C, D) Fluorescence distributions of the promoter activities were determined with the plugin ObjectJ in ImageJ identifying the GFP signal at the single cell level.

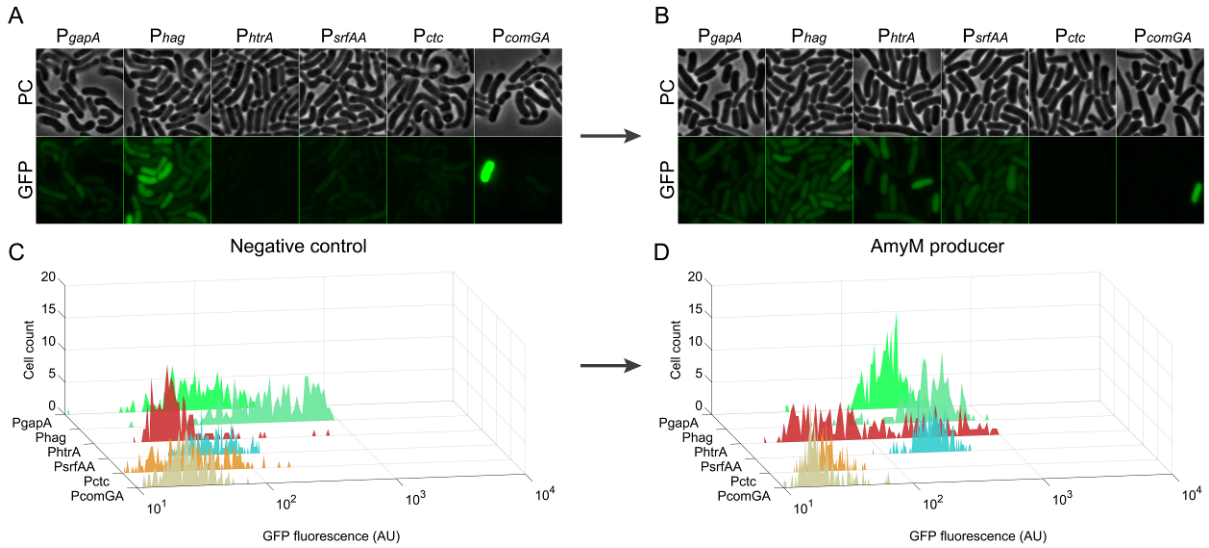


Figure 3.15: Fluorescence microscopy of AmyM producing strains in defined medium during stationary phase. The strains were analysed after 16-17 h of growth in non-producing (A, C) and AmyM producing strains (B, D). (A, B) Phase contrast and GFP images show the activities of the selected promoters P_{gapA}, P_{hag}, P_{htrA}, P_{srfAA}, P_{ctc} and P_{comGA}. (C, D) Fluorescence distributions of the promoter activities were determined with the plugin ObjectJ in ImageJ identifying the GFP signal at the single cell level.

3.3 Conclusion

The aim of this chapter was to analyse the impact of high-level production of the industrial enzymes XynA and AmyM on *B. subtilis* cell physiology and development.

The activities of both XynA and AmyM were determined in rich and defined medium showing a production of XynA during exponential phase and decline in stationary phase in rich medium, presumably because the rate of degradation was higher than the rate of synthesis. In defined medium, stationary phase degradation was less pronounced, emphasising the importance of growth medium, especially for industrial-like purposes.

In contrast to the XynA production profile, AmyM production increased during transition to, and in, stationary phase, indicating a different organisation of expression and secretion of AmyM. It is not clear what factors are responsible for this difference in kinetics, although several possibilities have been considered. These include differences in the genetic constructs between the integrated *xynA* and the *amyM* gene located on an autonomously replicating plasmid, with distinct promoters and different initiations of replication. Another reason might be the upregulation of quality control proteases at an early stage that leads to a degradation of the heterologous AmyM. It could be as well related to some missing elements at the early stage of growth e.g. chaperones that assist the folding of the protein and lead to the delay in the accumulation and secretion of AmyM into the growth medium. To further analyse the kinetics of XynA and AmyM, the transcription of genes encoding both enzymes was investigated in chapter 4 using an identical, plasmid-based system.

In case of the native enzyme XynA, no GFP expression differences were observed at the population level of the eight different physiological and developmental pathways between XynA producing and non-producing strains in either rich or defined media. Additionally, the investigation at the single-cell level, allowing us to monitor cellular heterogeneity under high-level enzyme production, confirmed this observation.

In contrast, AmyM expressing strains showed very clear differences with respect to the quality control protease reporter P_{htrA} in both media on the population and single-cell level. The upregulation of the quality control protease *via* the CssR regulon shows that translocation and folding of the heterologous AmyM generates secretion stress compared to the native XynA. Interestingly, the AmyM producing strains showed bistable induction of $P_{htrA,}$ reflecting the secretion stress induced by this enzyme in rich and defined medium. We predicted that the P_{htrA} reporter fusion might be able to similarly reveal cells that produce different amounts of AmyM and

we tested this hypothesis in chapter 4 by combining the P_{htrA} reporter with a transcriptional mCherry fusion reflecting the AmyM transcription.

In the case of both the XynA and AmyM producing strains there was a direct correlation between glycolytic activity and motility in defined medium. This indicates a connection between the energy utilising development and usage of flagella (P_{hag}) and energy generating mechanisms (P_{gapA}). Given the requirement for high-energy consumption, motility, which is of little value in a stirred fermenter, is likely to negatively influence process efficiency in the later stages of large-scale bioprocessing. In comparison to defined medium, motility seems to be repressed in rich medium, which could reflect the activity of the global transcriptional regulator CodY that down-regulates flagellar gene expression (Bergara *et al.*, 2003). In contrast, to the activation of glycolysis in the late stages of growth in defined medium, sporulation was turned on in rich medium during stationary phase when glycolytic processes ceased lacking to provide energy for the starving bacteria.

In defined medium the differences in glycolytic and motility expression, seen in the microbioreactor between AmyM producing and non-producing strains, were reflected at the single-cell level with a higher heterogeneity, including more sessile cells due to the inactivity of SigD (Kearns and Losick, 2005), in the non-producing strains in stationary phase. These processes were more uniform during AmyM production, showing an adaptation of *B. subtilis* to such conditions. This finding leads to the conclusion that the more AmyM a cell produces the more active the glycolytic pathway is and the need to search for new nutrient sources.

Furthermore, microscopy revealed that in defined medium the XynA producing and non-producing strains displayed an altered morphology with bulky cells. The rich medium on the other hand contains almost 10-times more magnesium than defined medium, and this difference in magnesium concentration could be responsible for maintaining a normal rod shape. In contrast, to this bulky cell shape, in the AmyM related experiments the non-producing cells looked very sick compared to the AmyM producing strains. Since the difference between those strains was the empty plasmid in the non-producing strains coding for two more genes, ampicillin and a defective chloramphenicol, this might be an additional burden for the cells and therefore possibly a reason for the impaired cell morphology.

Chapter 4

Impact of proteases on secretion stress and production of industrial enzymes

4.1 Introduction

In the previous chapter the production of both XynA and AmyM and their impact on the physiology of *B. subtilis* were analysed during fermentation in rich and chemically-defined medium on a population and single cell level. The results showed different kinetics in the production of both enzymes, particularly the expression of the heterologous AmyM effected cells leading to the upregulation of the protein quality control system (P_{htrA}). The production of XynA had no influence on the cells. In this chapter, the effect of quality proteases on both enzymes, and the secretion stress associated with AmyM production are discussed.

In industrial fermentation processes a limiting factor for recombinant enzyme production are proteases that are degrading proteins in the cell envelope or in the culture medium. The so called feeding proteases of *B. subtilis* (AprE, Bpr, Epr, Vpr, Mpr, NprB and NprE) are secreted and facilitate the survival of *B. subtilis* by breaking down extracellular proteins to recover amino acids and peptides (Pohl *et al.*, 2013, Krishnappa *et al.*, 2013, Krishnappa *et al.*, 2014). These proteases have been shown to target heterologous proteins leading to reduced yields (Wu *et al.*, 1993, Wu *et al.*, 2002, Westers *et al.*, 2008). In addition to the feeding proteases, there are four so called quality control proteases (HtrA, HtrB, HtrC, WprA) that are active at the membrane and cell wall interface. These proteases are responsible for the clearance of blocked translocases and degrade slowly folding or misfolded proteins (Stephenson and Harwood, 1998, Jensen *et al.*, 2000, Sarvas *et al.*, 2004, Pohl and Harwood, 2010, Pohl *et al.*, 2013). HtrA and HtrB are membrane bound and regulated by the two-component system CssRS that is induced by secretion and heat stress (Darmon *et al.*, 2002, Westers *et al.*, 2004, Westers *et al.*, 2006, Noone *et al.*, 2012). HtrC, a homologue of HtrA and HtrB, is induced by cell wall stress and

controlled by the WalRK two-component system (Fabret and Hoch, 1998, Bisicchia *et al.*, 2010, Pohl *et al.*, 2013, Krishnappa *et al.*, 2014). WprA is a cell wall-associated protease and is involved in the proteolytic regulation of HtrA and HtrB (Margot and Karamata, 1996, Krishnappa *et al.*, 2014). These quality control proteases are important for the successful secretion of a protein of interest into the growth medium. To investigate the relationship between quality control proteases and the production kinetics of XynA and AmyM, the impact of the artificial expression of quality control proteases on the production of both industrial enzymes was analysed. In particular, HtrA was shown to be a valuable secretion stress marker for the expression of homologous and heterologous proteins using a transcriptional fusion to its promoter (Trip *et al.*, 2011). However, these studies analysed secretion stress on the population level lacking information about heterogeneity and actual production of a protein of interest. Since industrial fermentations aim to reduce a heterogeneous enzyme expression, we investigated the heterogeneity and relation of secretion stress and enzyme expression at the single cell level to analyse protease deficient strains and compare the plasmid and integration based expression of AmyM.

4.2 Results and discussion

4.2.1 Enzyme kinetics and influence of quality control proteases

In the previous chapter, significant differences in the expression kinetics between XynA and AmyM were observed. We first wanted to know whether the differences in enzyme production kinetics are related to differences in cellular mRNA levels. To ensure that the expression conditions were the same, a plasmid based system was used for the expression of both XynA and AmyM. The plasmid uses P_{amyQ} from *B. amyloliquefaciens* as promoter, also used for industrial fermentations (Ploss *et al.*, 2016), and pUB110 as replicon which gives a copy number of ~48 ± 2 per cell (Leonhardt, 1990). The different strains were grown in LB medium in the BioLector® using a gain of 50 (biomass) to determine cell growth (Figure 4.1A & B). Enzyme activities were examined using the EnzChek® Ultra Xylanase Assay Kit for XynA and the Phadebas® Amylase assay for AmyM as mentioned earlier (Section 2.10) (carried out together with Rita Cruz). Despite the use of the same plasmid system, both enzymes clearly showed different production kinetics. The XynA expression rapidly increased during growth (~1-7 h) to ~17 RXU and slowly reached a plateau in stationary phase (~12.5 h) with a relative xylanase activity of ~23 RXU (Figure 4.1A).

XynA levels remained stable in LB medium. This is in contrast to growth in rich medium (Chapter 3, Figure 3.3A), where XynA is degraded in the stationary phase. In contrast to XynA production, the synthesis of AmyM is delayed until stationary phase and started approximately after ~7.5 h reaching a maximum at ~11.5 h with 0.0013 RAU. This late start of AmyM synthesis was comparable to previous results (Chapter 3, Figure 3.3B).

To determine whether *amyM* and *xynA* are transcribed at the same level, mRNA was extracted of both XynA and AmyM producing strains during the transition phase and stationary phase (Section 2.4.11), and quantified with qPCR (Section 2.4.12 & 2.4.13). In the XynA producing strain, the expression amounted to ~971-fold in the end of exponential phase and ~552-fold in stationary phase (Figure 4.1C). Interestingly, there is still a high level of *xynA* mRNA in stationary phase but no detectable increase in enzyme production. Possibly, XynA synthesis and degradation is balanced out during stationary phase (Figure 4.1A). The qPCR results for AmyM showed a substantially higher mRNA level in the transition phase than in stationary phase (Figure 4.1C), which is the opposite of what is seen when actual enzyme levels were measured. Why this is the case is unclear.

The qPCR data indicate that transcription is not the secretion dynamic determining factor for XynA and AmyM synthesis. To further pursue the investigation of factors that influence the production of both enzymes, the effect of the quality control proteases was studied, since previous results in chapter 3 have shown an upregulation of the secretion stress reporter P_{htrA}, and since the discrepancy between mRNA levels and enzyme production might suggest a proteolytic degradation. For the analysis of induced secretion and cell wall stress, the expression of the four quality control proteases HtrA, HtrB, HtrC and WprA was determined at the mRNA level in XynA and AmyM producing strains. Samples were taken during the transition phase in cells growing in LB (Figure 4.2). For XynA producing cells, the results showed a small induction of *htrA* (~2-fold) and *htrB* (~3.6-fold) whereas in the AmyM producing strains *htrA* (~26-fold) and *htrB* (~51-fold) expression was strongly induced. This strong induction in AmyM producing cells is in agreement with the upregulation of P_{htrA} described in chapter 3 and underlines the fact that heterologous AmyM generates substantially more stress than expression of the native, coevolved, XynA protein. The expression of *htrC* (~1.2-fold) and *wprA* (~0.9-fold) in XynA

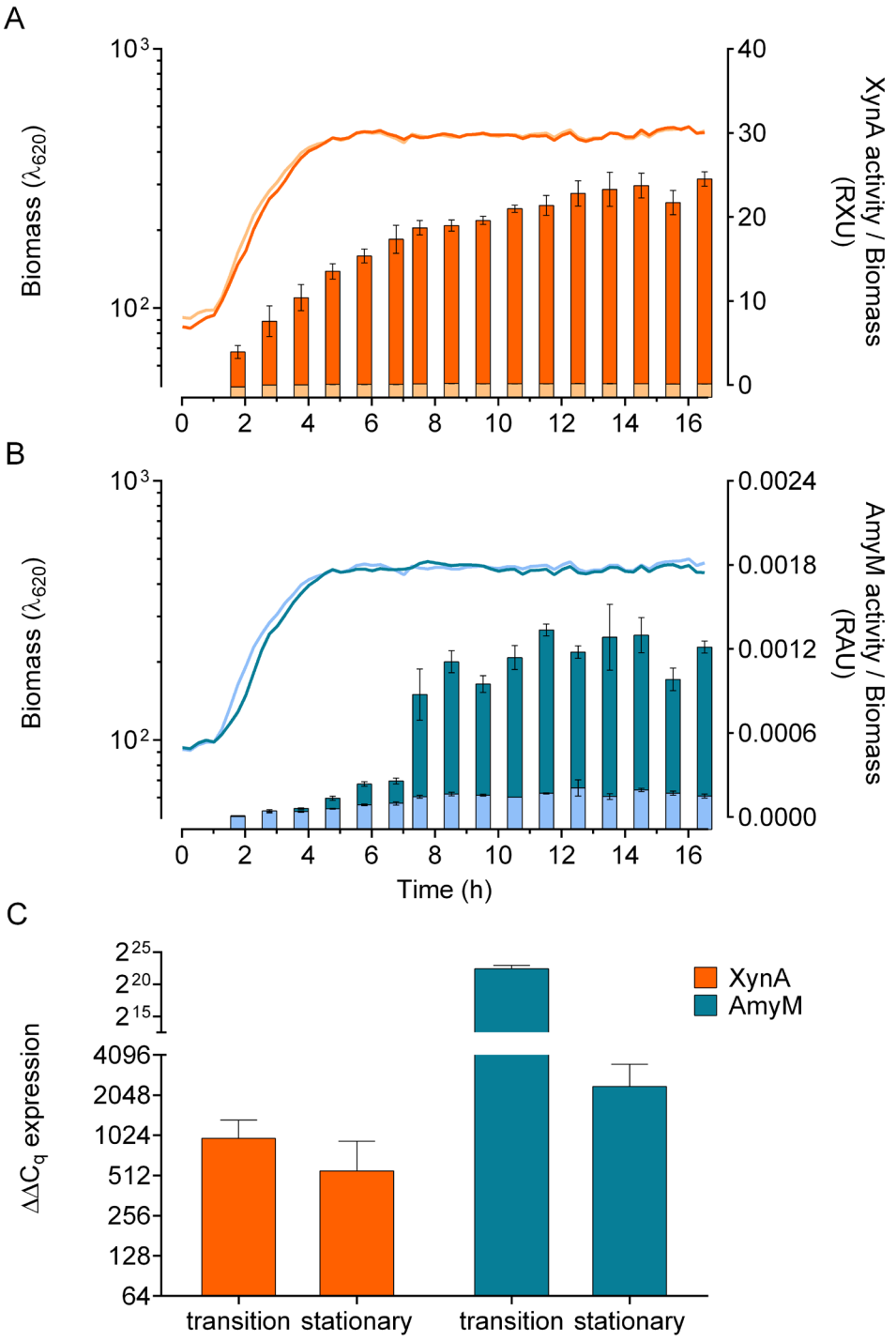


Figure 4.1: Growth and enzyme production by strains expressing XynA or AmyM in LB medium. The activities were divided by the biomass (smooth lines) to achieve an activity per cell (bars). (A) Growth curves and enzyme activities (RXU) of strains producing XynA (dark) and the negative control (light) are displayed. (B) Growth curves and enzyme activities (RAU) of strains producing AmyM (dark) and the negative control (light) are shown. (C) Relative quantification of *xynA* (orange) and *amyM* (blue) expression ($\Delta\Delta C_q$) in transition and stationary phase determined by qPCR. Error bars indicate the standard deviation of three biological replicates.

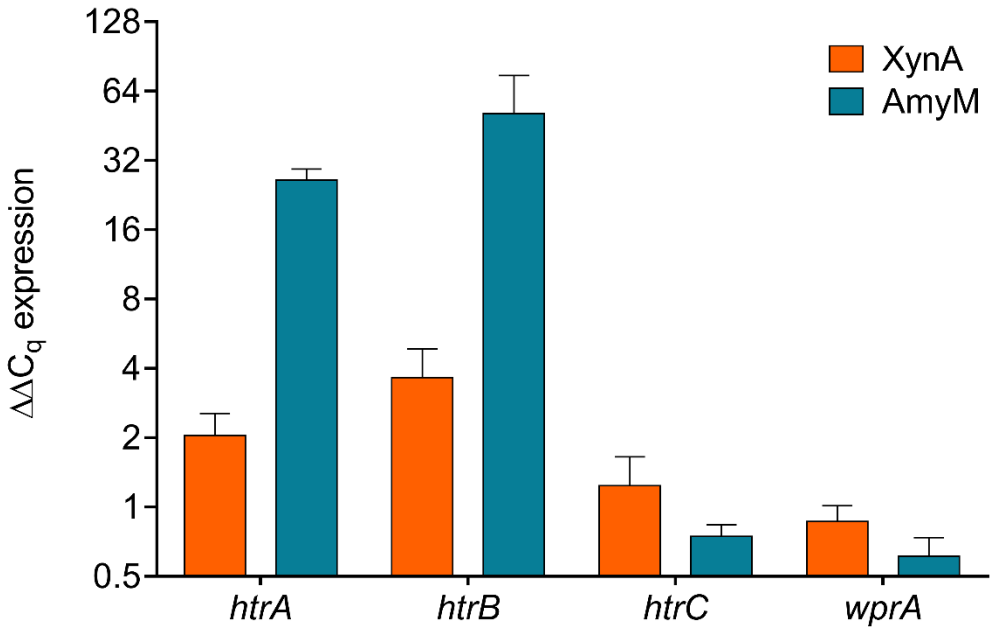


Figure 4.2: Transcriptional expression of quality control proteases in strains producing XynA or AmyM. Relative quantification of *htrA*, *htrB*, *htrC* and *wprA* expression ($\Delta\Delta C_q$) in XynA (orange) or AmyM (blue) producing strains compared to non-producing strains during transition phase in LB medium determined by qPCR. Error bars indicate the standard deviation of three biological replicates.

producing strains was similar to that of the non-producing strain. In the AmyM producing strain the expression of *htrC* (~0.7-fold) and *wprA* (~0.6-fold) was slightly lower compared to the non-producing strain, indicating that the cell wall associated secretion stress might be lower.

The upregulation of the quality proteases indicates secretion stress and the necessity of these proteins to relieve this stress. To investigate this further, strains were constructed that express an extra copy of HtrA, HtrB, HtrC or WprA under control of the strong xylose-inducible promoter P_{xyl} (Feucht and Lewis, 2001). To verify the overexpression, mRNA levels were determined with qPCRs (Section 2.4.11-2.4.13). The qPCR data reveal that all four proteases were overexpressed (Figure 4.3B) when cells were grown in the presence of xylose, with *htrA* ~28.8-fold, *htrB* ~22.5-fold, *htrC* ~18.7-fold and *wprA* approximately ~4.4-fold. After this successful validation, the xylose-inducible constructs were introduced into the XynA and AmyM production strains and the loss of XynA and AmyM production in the medium determined (Figure 4.3C & D) during transition phase and stationary phase (t_1, t_2; Figure 4.3A) when a quality control protease was overexpressed. The positive control, set to 100%, consists of a strain synthesising XynA or AmyM but lacking the xylose induction constructs. The activities of the overexpression strains were then calculated by comparing to the positive control in presence or absence of xylose and

subsequently the differences (Δ) of both activities (± xylose) were calculated to achieve the loss of activity when overexpressing the specific protease. A different version of this figure showing the positive controls and strains with and without xylose induction can be found in the appendix A (Figure A.6). Interestingly, XynA production is reduced when HtrA is overexpressed (~23%) suggesting that HtrA is involved in degradation processes of XynA (Figure 4.3C). Overexpression of HtrB and HtrC had a lower impact (~7%) and WprA had virtually no influence on the synthesis of XynA, both in transition and stationary phase. The AmyM production was clearly reduced in strains overexpressing HtrA (~19%) showing similar results compared to XynA production during stationary phase (Figure 4.3D). Overexpression of HtrC and WprA also led to a decreased AmyM activity (~12.5%) during stationary phase, whereas HtrB likely plays a minor role in degradation processes.

The results indicated that the membrane bound protease HtrA affects XynA and AmyM production. Since HtrA and HtrB are under cross-regulation mediated by the CssRS two component system leading to elevated levels of one protease when the other is deleted (Darmon *et al.*, 2002, Noone *et al.*, 2001, Krishnappa *et al.*, 2014), we examined how the deletion of *htrA* or *htrB* would affect production of XynA and AmyM. To monitor the effect on secretion stress, we included the reporter fusion P_{htrA}-*gfp* of chapter 3. Cultures were grown in the BioLector® to enable measurement of biomass and GFP signal in parallel (Figure 4.4A & B) (carried out by Rita Cruz). The XynA producer strain showed close to no expression of P_{htrA} (~3 AU × λ_{620}^{-1}) but with the added deletion of *htrA* (~62 AU × λ_{620}^{-1}) or *htrB* (~37 AU × λ_{620}^{-1}) the GFP signals raised significantly. This increase in secretion stress was ~1.2-fold (Δ*htrA*) and ~1.7-fold (Δ*htrB*), respectively, compared to the deletion strains with no enzyme being expressed (Figure 4.4A). Activity assays of XynA were then performed (Section 2.10) to analyse the actual enzyme production of the deletion strains of *htrA* or *htrB*. Samples were taken at three time points (t_1, t_2, t_3; Figure 4.4A & C) (carried out by Rita Cruz). The XynA production was comparable between all three time points (Figure 4.4C) indicating that the synthesis reached a plateau at t_1 similar to figure 4.1A. However, the production was highest when the strain contained both proteases (~13.3 RXU) but as soon as either *htrA* or *htrB* were deleted, the production decreased to ~8.1 RXU or ~5 RXU, respectively. This amounts to a reduced XynA activity of 64% (Δ*htrA*) and 166% (Δ*htrB*).

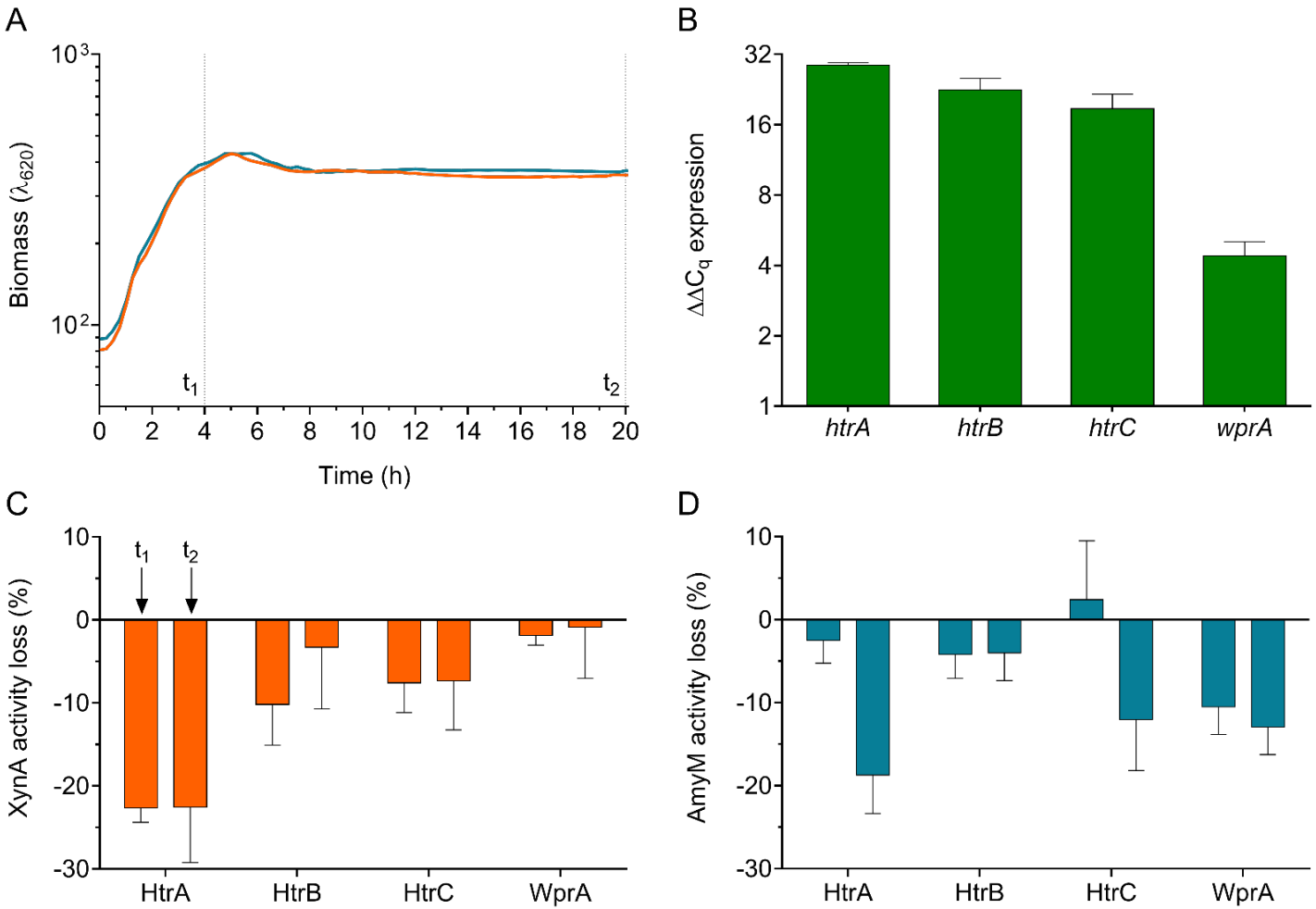


Figure 4.3: Overexpression of quality control proteases in strains producing XynA or AmyM in LB medium. Strains producing XynA (orange) or AmyM (blue) and carrying the xylose-inducible promoter P_{xyl} were induced with 0.2% of xylose to overexpress HtrA, HtrB, HtrC or WprA. (A) Averaged growth curves indicate strains in the presence (dark) and absence (light) of xylose. Samples for activity assays were taken during transition (t_1) and stationary phase (t_2). (B) Relative quantification of transcriptional expression ($\Delta\Delta C_q$) of *htrA*, *htrB*, *htrC* and *wprA* when induced with xylose compared to non-inducing conditions determined by qPCR. (C) Loss of XynA and (D) AmyM activities were calculated by subtracting the activities in absence from the activities in presence of xylose during transition (t_1, left bar) and stationary phase (t_2, right bar) when overexpressing a quality control protease. Enzyme activities were calculated by comparing quality control protease overexpression strains with positive controls (set to 100%) only expressing XynA or AmyM in presence or absence of xylose. Activities of 168 were subtracted from the samples. Error bars indicate the standard deviation of three biological replicates.

In contrast to the XynA producer cells, the secretion stress levels were highly elevated in the AmyM producing strain (~141 AU × λ_{620}^{-1}) during exponential phase. By deleting either *htrA* or *htrB* in the AmyM producer strain, the secretion stress increased to more than twice as much with an increasing intensity of ~300 AU × λ_{620}^{-1} (t_1, Figure 4.4B). The increase in stress response in AmyM producer cells amounted to ~5.4-fold (Δ*htrA*) and ~14-fold (Δ*htrB*). The enzyme production of AmyM increased from t_1 until t_3 (Figure 4.4D) similar to previous experiments (Figure 4.1B) (carried out by Rita Cruz). The absence of one quality control protease (HtrA or HtrB) does not significantly change the AmyM production but it does lead to a high induction in secretion stress. These results suggest that both proteases, which are normally upregulated when XynA and AmyM are synthesised (Figure 4.2), are playing a key role in the processing of XynA in the cell envelope. For the AmyM production either HtrA or HtrB is likely to be sufficient for the quality control, however, secretion stress is highly induced when one protease is deleted.

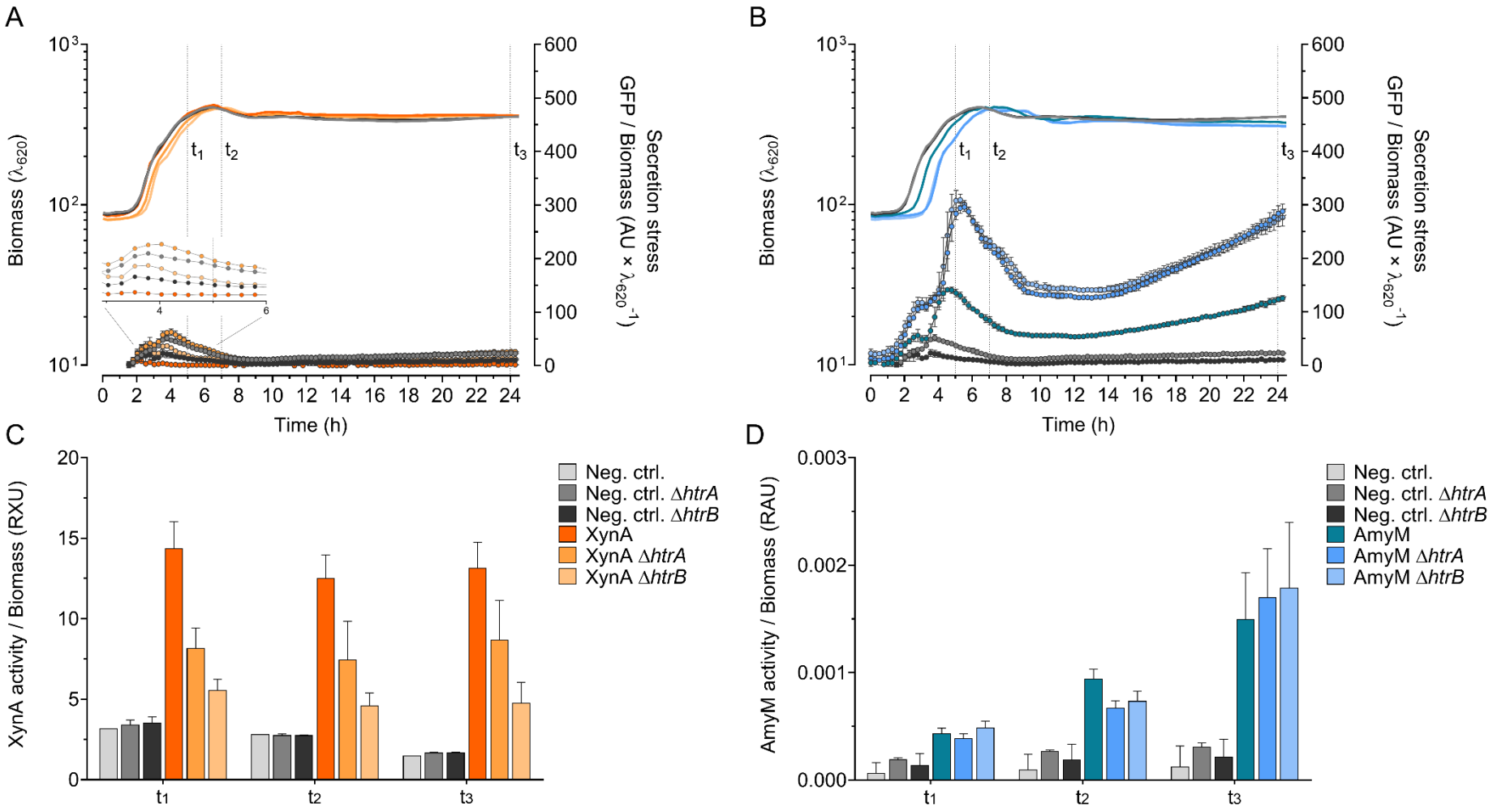


Figure 4.4: Secretion stress and enzyme activities of strains producing XynA or AmyM and lacking *htrA* or *htrB*. (A-B) Biomass (smooth lines) and secretion stress, reflected by the relative promoter activity of P_{htrA}-*gfp* (GFP per unit biomass; circles), were monitored for 24 h of strains producing (A) XynA (orange) or (B) AmyM (blue) including negative controls (grey to black) using the BioLector®. Samples for activity assays were taken at three time points (t_1-t_3). (C) Enzyme activities (RXU) of strains producing XynA (dark) lacking *htrA* (medium) or *htrB* (light) and the negative controls are displayed. (D) Enzyme activities (RAU) of strains producing AmyM (dark) lacking *htrA* (medium) or *htrB* (light) and the negative controls are shown. The activities were divided by the biomass to achieve an activity per cell. Error bars indicate the standard deviation of three biological replicates.

4.2.2 Analysis of secretion stress and AmyM production in protease deficient strains

The strongest induction of P_{htrA} is seen when the heterologous AmyM protein is expressed, especially when *htrA* or *htrB* is absent. To investigate the role of other proteases in the secretion stress caused by AmyM production, a number of protease deficient strains (Figure 4.5A) were made containing the P_{htrA} secretion stress reporter. The strains were transformed with a high copy plasmid containing P_{amyQ}-*amyM* immediately followed by the *mcherry* reporter gene, resulting in a bicistronic operon to monitor transcription of the *amyM* gene. So called 'BRB'-strains lacking feeding and quality control proteases (Figure 4.5A) were used as hosts (Pohl *et al.*, 2013). Strains BRB7 up to BRB13 were selected to have strains that lacked all feeding proteases (BRB7) and a few strains that on top of this lacked several quality control proteases (BRB8 to BRB13). First, the effect on AmyM production was analysed (Figure 4.5B & C) (carried out by Rita Cruz). Amylase activities (A_{620}) were measured at different points during growth and the results show that the production in all strains is comparable, except for BRB13 which shows a significantly lower activity (Figure 4.5B). The activities of BRB11-BRB13 in exponential phase (t_1, t_2; Figure 4.5B), indicated by very small bars, could not be quantified due to sensitivity limitations by the plate reader in the lower range. To validate the amount of AmyM produced in the BRB strains, samples were collected in stationary phase and analysed with SDS-PAGE (Figure 4.5C). All strains showed a similar amount of AmyM at ~75.4 kDa, except for BRB13 which showed no visible band, and BRB7 showing a weak protein band. The latter is not in line with the AmyM activity assays for reasons that are currently unknown.

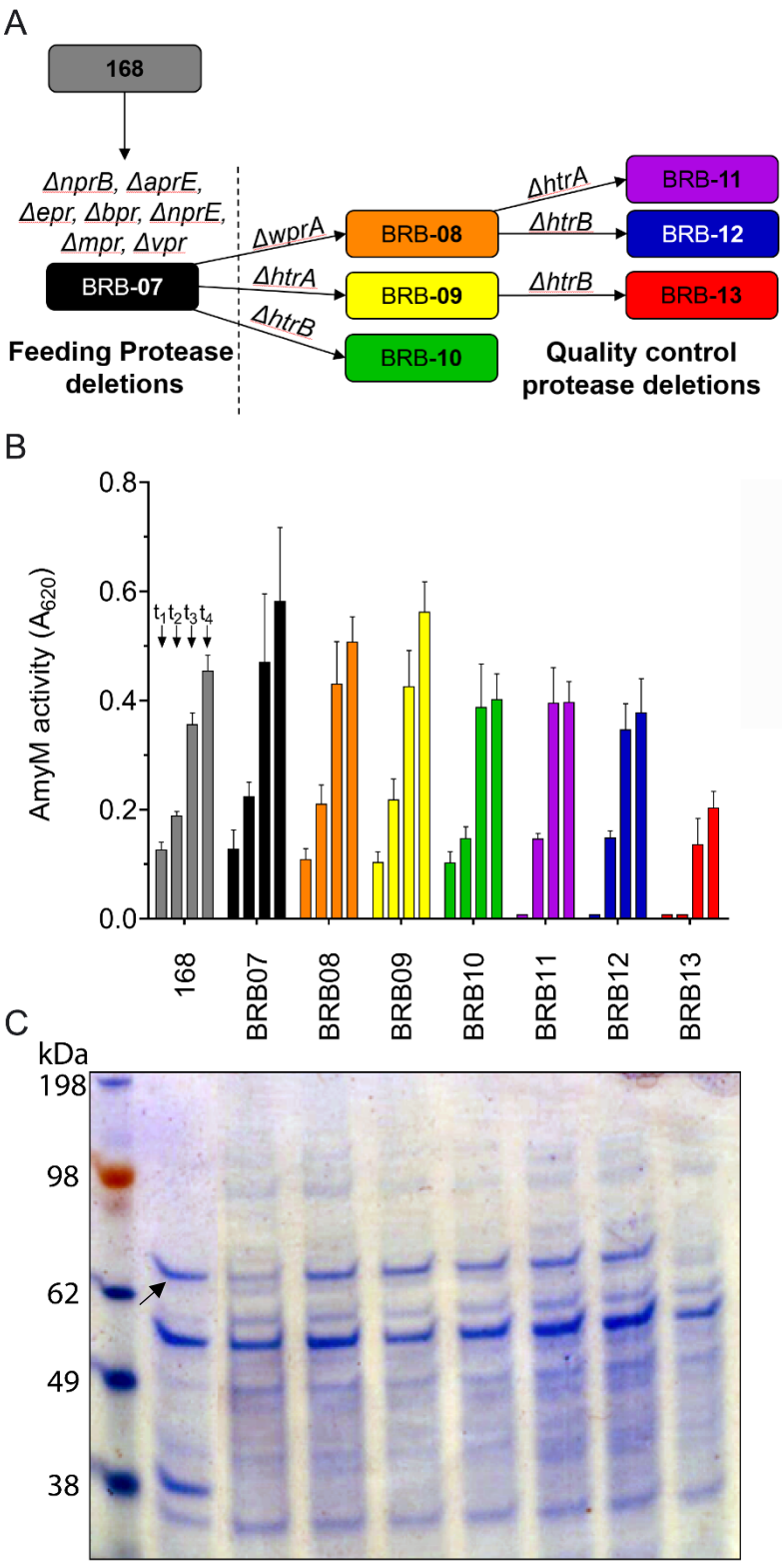


Figure 4.5: AmyM production in protease deficient strains. The strains 168 and BRB7 to BRB13 producing AmyM were subjected to enzyme activity assays after growing in LB medium. (A) Genetic background of BRB strains lacking feeding proteases (BRB7) and quality control proteases (BRB8 to BRB13). Figure was adapted from Pohl *et al.* (Pohl *et al.*, 2013). (B) AmyM production was monitored during growth (t_1, t_2: exponential phase; t_3, t_4: stationary phase). Error bars indicate the standard deviation of nine biological replicates. (C) SDS-PAGE shows proteins of protease deficient strains in stationary phase. The heterologous enzyme AmyM is shown by an intense band (arrow) at ~75.4 kDa. The SeeBlue® Plus2 Prestained Standard marker (Thermo Fisher Scientific) was used as ladder.

Since subsequent deletions of feeding and quality control proteases do not increase AmyM production, the question arises whether the absence of all these proteases has a clear effect on secretion stress, or whether only deletion of *htrA* or *htrB* increases secretion stress, as we have shown before (Figure 4.4A & B). To test this, the BioLector® was used to monitor growth and fluorescent levels, whereby secretion stress was reflected by GFP levels (P_{htrA}-*gfp*) and *amyM* expression by mCherry levels (Figure 4.6). As a negative control a plasmid was constructed (pCS76) expressing *mcherry* alone under the P_{amyQ} promoter. The growth rates of all strains were comparable, except for the BRB13 background (Figure 4.6A). In fact, strain BRB13 also displayed the highest secretion stress, up to ~2.5-fold higher than BRB9 and ~3.7-fold higher compared to 168 (Figure 4.6B). Interestingly, although the negative controls with plasmids only expressing mCherry revealed low secretion stress levels (168, BRB7, BRB8), the deletion of *htrA* or *htrB* led to increased stress (BRB9 to BRB12) similar to figure 4.4B and with erasing both proteases (BRB13) the GFP signal was in fact above the intensities of all AmyM producing strains, except for BRB13 (+AmyM). The expression of *amyM*, reflected by mCherry intensities, was also monitored during growth (Figure 4.6C). The most striking differences in mCherry levels showed strain BRB13 (+AmyM), displaying by far the lowest activity. Strains containing either a *htrA* or *htrB* deletion or the wild type strain 168 demonstrated ~3-4-fold higher mCherry intensities than BRB13 (+AmyM). These differences in mCherry fluorescence of the AmyM producing strains seem to be linked to secretion stress levels since strains with higher secretion stress (GFP) show lower enzyme expression (mCherry). Especially the deletion of both *htrA* and *htrB* in BRB13 (+AmyM) impacted the cells remarkably in growth, secretion stress and enzyme expression leading to the highest secretion stress and lowest enzyme expression levels. The results of BRB13 lacking *amyM* reveal that the loss of both HtrA and HtrB are disadvantageous for cells not even producing an industrial enzyme.

The investigation of secretion stress and enzyme production showed high secretion stress and low AmyM expression levels in strains lacking both HtrA and HtrB. However, the analysis was focused on the population level lacking important information about heterogeneity of secretion stress and enzyme expression in cells and how related these processes are with each other. To investigate the heterogeneity and their relationship in more detail an analysis at the single-cell level was performed. For this purpose, the BRB strains were subjected to FACScan

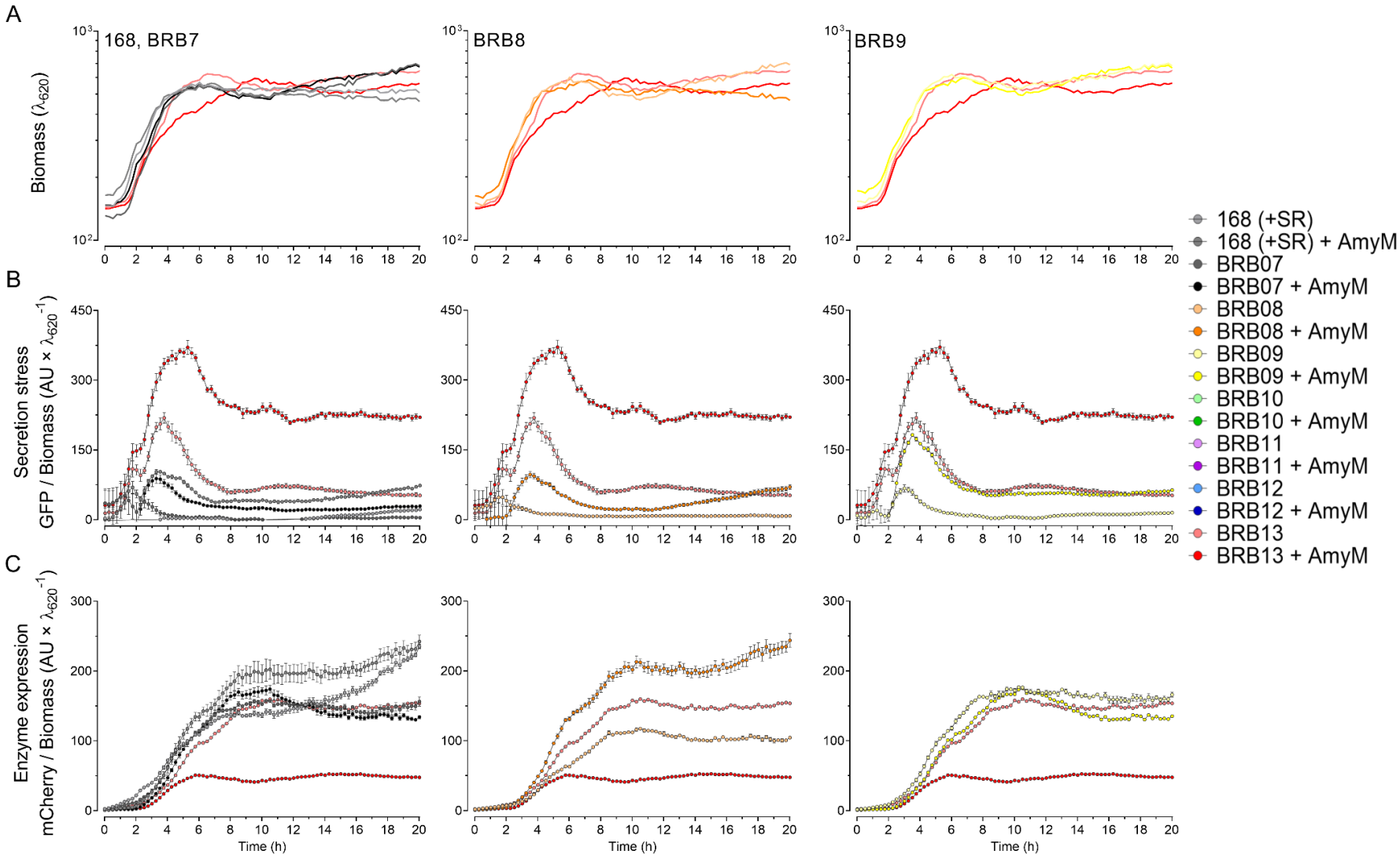


Figure 4.6: (See caption on following page)

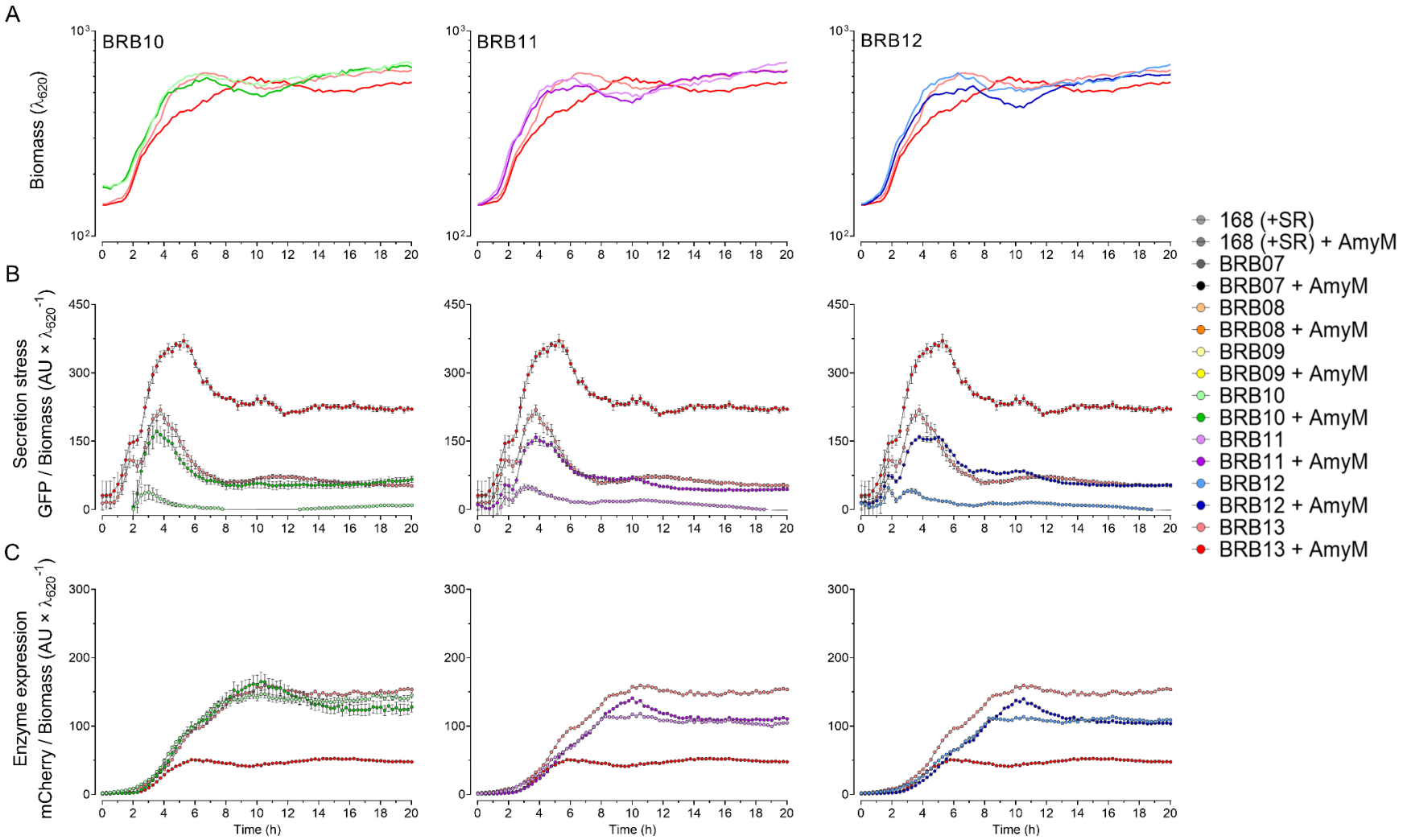


Figure 4.6: Secretion stress and enzyme expression of protease deficient strains producing AmyM. Strains carrying secretion stress reporter P_{htrA}-*gfp*, a plasmid expressing AmyM and mCherry simultaneously and negative controls (only mCherry) were subjected to BioLector® experiments for 20 h in LB medium. (A) Biomass (smooth lines), (B) secretion stress (GFP) and (C) enzyme expression (mCherry) (relative promoter activity - GFP or mCherry per unit biomass; circles) were monitored in 168 and BRB7 up to BRB13. Results of BRB13 were introduced as reference lines. Error bars indicate the standard deviation of three biological replicates.

analyses. Cells were harvested in the stationary phase and approximately 50,000 events were measured per sample for GFP (secretion stress) and mCherry (enzyme expression) levels. As shown in figure 4.7A, the cellular GFP levels showed a heterogenic bimodal distribution for strains 168, BRB7 and BRB8, revealing a population of cells with low secretion stress levels and a population of cells displaying higher stress levels. BRB9 to BRB13 showed a monomodal GFP distribution with a peak at high GFP levels indicating that all cells induced the secretion stress reporter fusion. Interestingly, the absence of either *htrA* or *htrB* converts the bimodal response into a classic monomodal response with increased secretion stress levels (Figure 4.7A). This strong upregulation of secretion stress fits to the findings of previous studies were the deletion of either *htrA* or *htrB* led to elevated levels of the other due to cross-regulation mediated by the CssRS two component system (Darmon *et al.*, 2002, Noone *et al.*, 2001, Krishnappa *et al.*, 2014). The strains BRB9 to BRB13 showed an increasingly higher GFP response, similar to the BioLector® data (Figure 4.6B). The RFP signal, indicative of *amyM* expression, showed a much more homogenous distribution in the different strains and no obvious bimodal pattern. However, as expected from the Biolector® results (Figure 4.6C), the average RFP signals differed substantially (Figure 4.7B). The FACScan analysis enabled us to correlate the GFP with mCherry fluorescence per cell. As shown in figure 4.7C, there is a clear positive correlation between *amyM* expression levels and the induction of P_{htrA}. The plots of the strains 168, BRB7 and BRB8 reveal two populations with one displaying a lower stress and enzyme expression against the other one with higher stress and enzyme expression. The more quality control proteases were deleted the more uniform the distribution (BRB9-12) with a higher correlation caused by the cross-regulation of the CssRS system (Figure 4.7C). With the deletion of both *htrA* and *htrB* in BRB13 two populations are formed with one expressing much less mCherry ($\sim 10^3$) compared to the other ($\sim 10^4$) while experiencing similar stress levels. The FACScan analysis confirmed the BioLector® data (Figure 4.6B), indicating that removal of either *htrA* or *htrB* or both increased secretion stress in BRB9 to BRB13 suggesting a compensating effect of cells demanding more HtrA due to a reduced proteolytic activity (Noone *et al.*, 2000). In contrast to the BRB strains, the correlation of secretion stress and enzyme expression in 168 was lower than expected due to the formation of two populations and cells lacking plasmids experiencing no secretion stress and enzyme expression (Figure 4.7C).

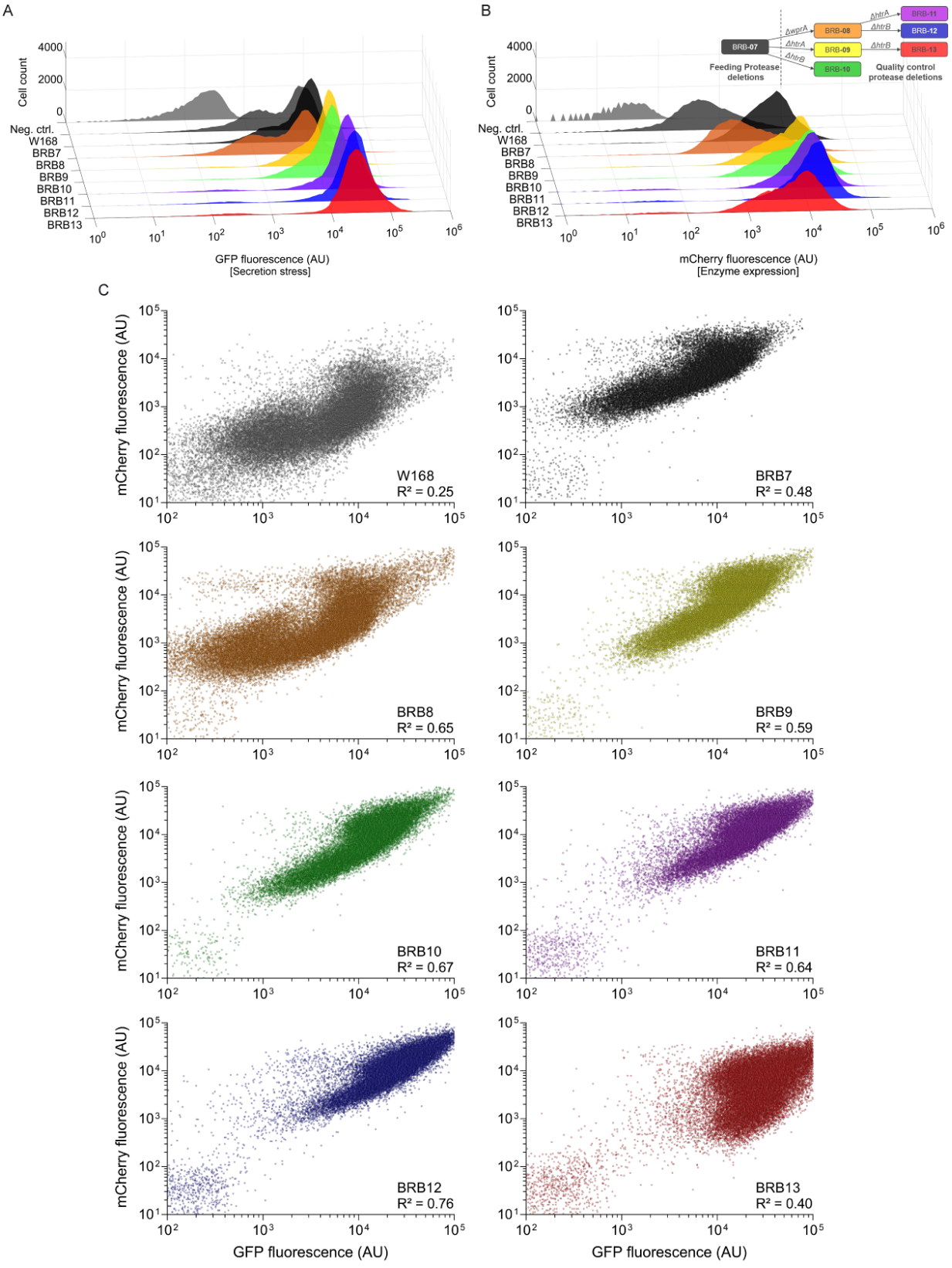


Figure 4.7: Secretion stress and enzyme expression of protease deficient strains producing AmyM during early stationary phase in *B. subtilis*. BRB strains carrying a plasmid expressing AmyM and mCherry simultaneously and secretion stress reporter P_{htrA}-*gfp* were subjected to FACScan to measure (A) secretion stress (GFP) and (B) enzyme expression (mCherry) by scanning 50,000 events per strain for both fluorescence signals using a LSRFortessa™ X-20 (BD Biosciences) after growth at 37°C in LB medium. The deletion scheme of the BRB strains is attached. (C) Correlation of secretion stress with enzyme expression.

4.2.3 Heterogeneity of secretions stress and AmyM production

In the previous section a relationship between enzyme expression and secretion stress was shown in protease deficient strains at the population and single-cell level. This correlation was stronger the more quality control proteases were removed resulting in more secretion stress but particularly in 168 this relationship was lower displaying a bimodal secretion stress response. To investigate whether this bimodality was caused by the promoter P_{amyQ} inducing *amyM*, the plasmid based expression system was compared with a genomic integrated *amyM* system. In addition, the integrated construct containing the constitutive P_{15} (Stewart *et al.*, 1998) promoter was driving the expression of AmyM and mCherry. To compare both expression systems, a FACScan analysis was performed. As shown in figure 4.8, this analysis revealed remarkable differences in heterogeneity and signal strength between plasmid and integration based systems. The secretion stress was observed as very bimodal for the plasmid compared to the relatively monomodal distribution of the integrated construct (Figure 4.8A). This heterogeneity is also reflected by the noise (variance/mean) (Ozbudak *et al.*, 2002) of the secretion stress induced by the plasmid with ~8200 AU against ~3100 AU of the integration. Considering the bimodal distribution of the plasmid, the average GFP value would be expected to be lower than the integration but the plasmid based system showed an average GFP signal of ~6500 AU, whereas the integrated *amyM* expression system showed an average GFP fluorescence of ~5800 AU. This higher secretion stress of the plasmid is reflected by a lower enzyme expression (mCherry) in figure 4.8B, compared to the integrated construct, supporting previous results of the BioLector® (Figure 4.6). The mCherry levels of the plasmid also revealed a wider signal distribution in comparison to the integrated version. The latter showed an overall higher *amyM* expression (mCherry signal) and a bimodal distribution signal with overlapping peaks between 10^3 AU and 10^4 AU. The correlation of secretion stress and enzyme expression of both expression constructs revealed two populations in the plasmid based system compared to the integrated *amyM* system with one population following a clear trend with the higher the stress the higher the enzyme expression (Figure 4.8C & D).

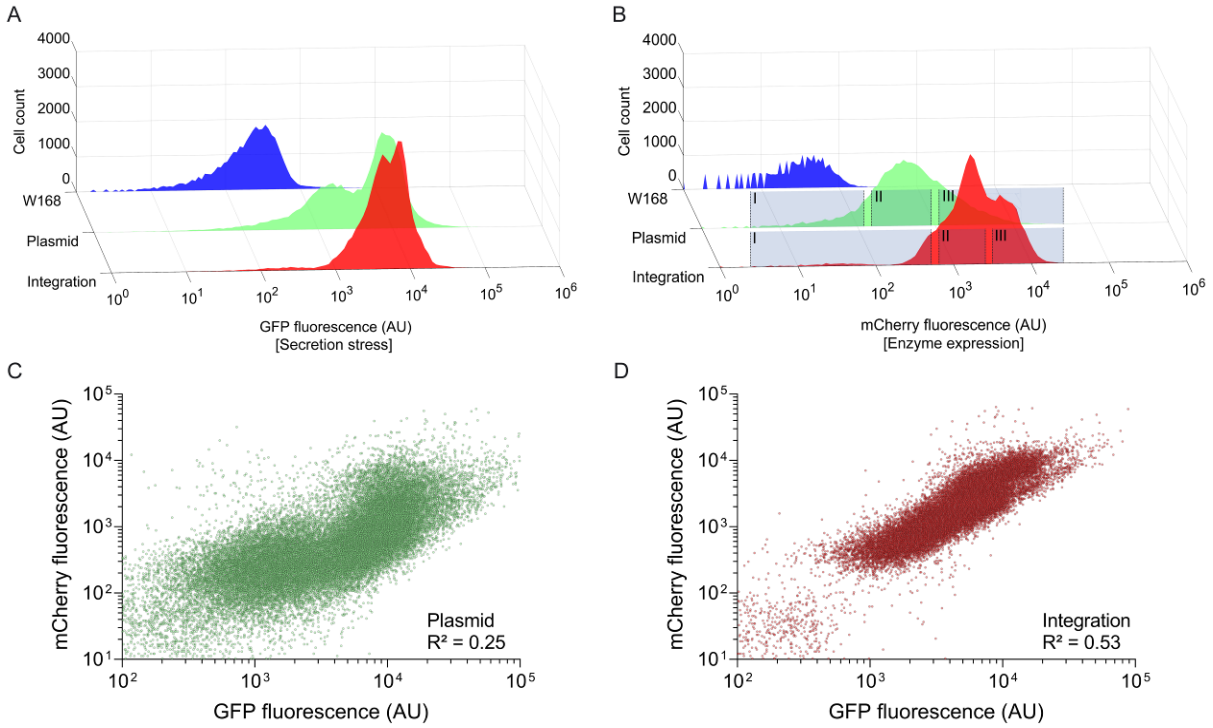


Figure 4.8: Secretion stress and enzyme expression of plasmid and integration based system producing AmyM during stationary phase in *B. subtilis*. Strains carrying a plasmid or an integrated construct expressing AmyM and mCherry simultaneously and secretion stress reporter P_{htrA}-*gfp* were subjected to FACScan to measure (A) secretion stress (GFP) and (B) enzyme expression (mCherry) by scanning 50,000 events per strain for both fluorescence signals using a LSRFortessa™ X-20 (BD Biosciences). After the scanning process, cells were sorted into three bins (I-III) with varying fluorescence intensities by a FACSAria™ Fusion flow cytometer system (BD Biosciences). Correlation of secretion stress with enzyme expression of (C) plasmid and (D) integrated based construct.

To investigate whether the bimodality of secretion stress caused by the plasmid based *amyM* expression was a result of the P_{amyQ} promoter, time-lapse microscopy of both AmyM expression constructs was performed (Figure 4.9). During microscopy the same trend was observed, showing a better correlation between stress and *amyM* expression when using the integrated construct (Figure 4.9C & D). However, the bimodality of secretion stress and enzyme expression was not observed this time suggesting that this bistability might be a cell chaining issue that was seen in cells growing in LB medium during the FACScan experiments (Figure 4.8A & C). The cells carrying the plasmid were observed to be longer than cells containing the integrated version (Figure 4.9A & B) which was presumably caused by cell division defects that occurred due to the plasmid based *amyM* expression. During the FACS experiments these defects in cell division possibly led to more chained cells that highly expressed *amyM*, reflecting one population, and another population with short and single cells showing a lower enzyme expression.

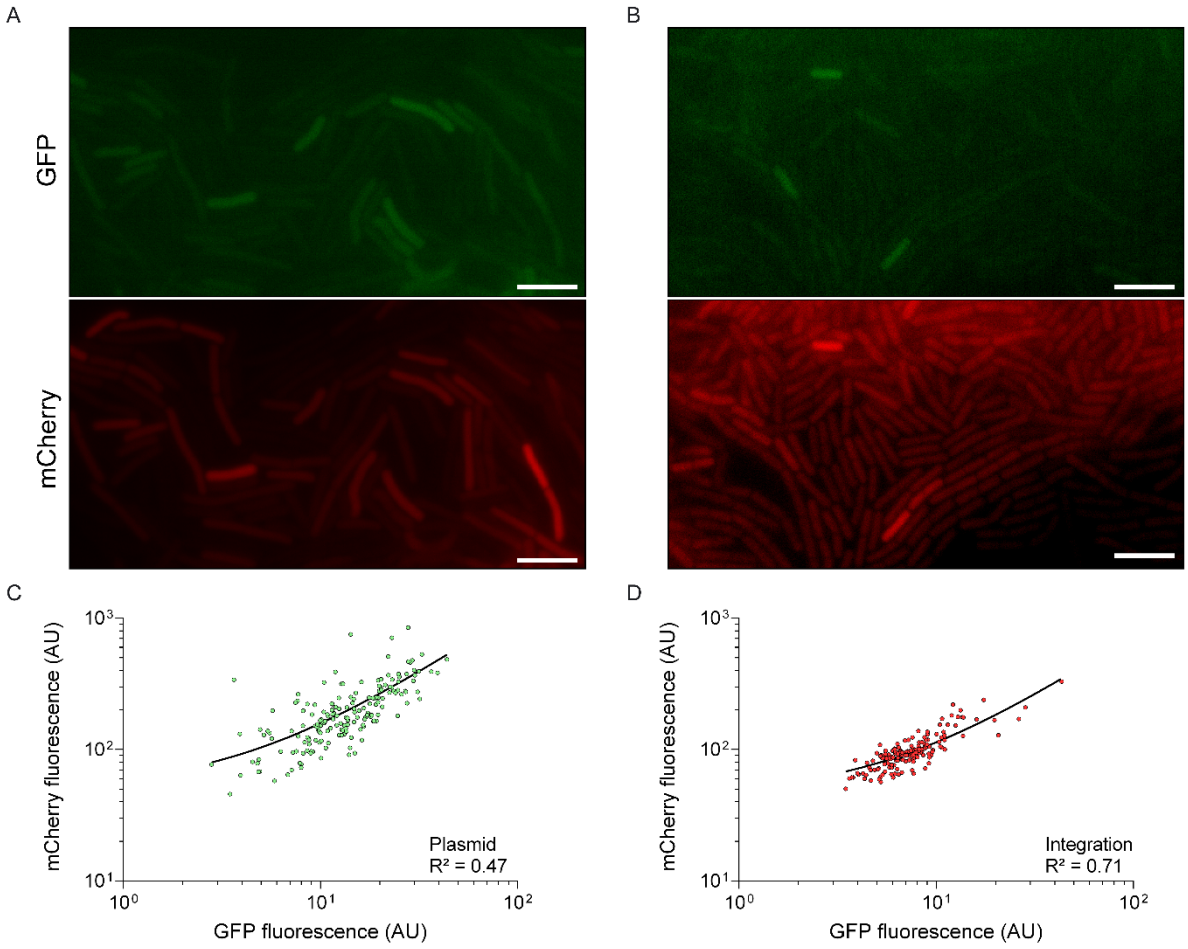


Figure 4.9: Secretion stress and enzyme expression of plasmid and integration based system producing AmyM in a microcolony of *B. subtilis*. Strains carrying secretion stress reporter P_{htrA}-*gfp* and a plasmid or an integrated construct expressing AmyM and mCherry simultaneously were subjected to time-lapse microscopy. The panels are split into two parts, each showing the following rows of frames: secretion stress (GFP; top) and enzyme expression (RFP; bottom) (A) Montage showing a microcolony of a plasmid based expression of AmyM. (B) Montage showing a microcolony of an integrated based expression of AmyM. Scale bars are 5 µm.

From an industrial point of view, it is of interest to reduce heterogeneity in enzyme expression and construct strains where all cells produce similar high-levels of the protein of interest. To obtain mutants that highly express *amyM* and to analyse whether cells maintain their expression levels or show again a similar heterogeneity, cells of the plasmid and integrated based expression construct were sorted into three bins (I, II, III; Figure 4.8B) with varying mCherry intensities (bin I ≙ low signal, bin II ≙ medium signal, bin III ≙ high signal). After that, cells were regrown and subjected to FACScan to determine the mCherry signals of each bin to investigate whether the cells retained the mCherry activities. The sorted cells expressing *amyM* from the plasmid (Figure 4.8B) showed an identical behaviour between the three sorted bins (Figure 4.10A & B).

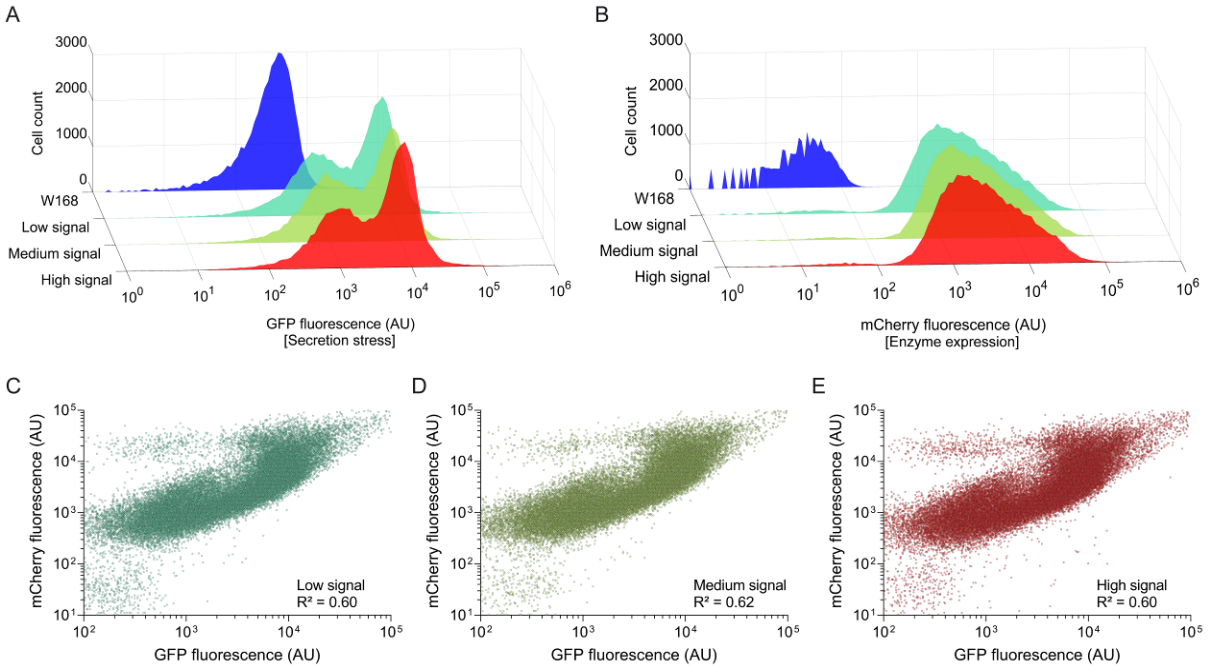


Figure 4.10: Secretion stress and enzyme expression of plasmid based system producing AmyM after cell sorting during stationary phase in *B. subtilis*. Strains carrying a plasmid expressing AmyM and mCherry simultaneously and secretion stress reporter P_{htrA}-*gfp* were subjected to FACS and cells were sorted into three bins with varying mCherry intensities (low, medium and high signal; bins I-III, Figure 4.8) by a FACSAria™ Fusion flow cytometer system (BD Biosciences) and subsequent growth at 37°C in LB medium. Cells were measured for (A) secretion stress (GFP)and (B) enzyme expression (mCherry) by FACScan with the scan of 50,000 events per strain for both fluorescence signals using a LSRFortessa™ X-20 (BD Biosciences). Correlation of secretion stress with enzyme expression of sorted cells with (C) low, (D) medium and (D) high signal.

The unaffected bimodality (GFP) indicates that the strains were not mutated and the bimodality was likely induced by cell chaining. In contrast to the plasmid, cells carrying the integrated construct revealed a different behaviour between the three bins showing that cells obtained from bin I with the lowest mCherry signal displayed again the lowest mCherry signal with an average intensity of ~4800 AU (Figure 4.11B). Bin II showed an average mCherry signal of ~8000 AU and bin III displayed a slight bimodal pattern with the same peak of bin II and an additional peak with increased intensity that was observed earlier in figure 4.8B. The overall fluorescence of both peaks averaged at ~10900 AU revealing that a high enzyme expression (mCherry signal) leads to high secretion stress (GFP signal) levels (Figure 4.11A & B).

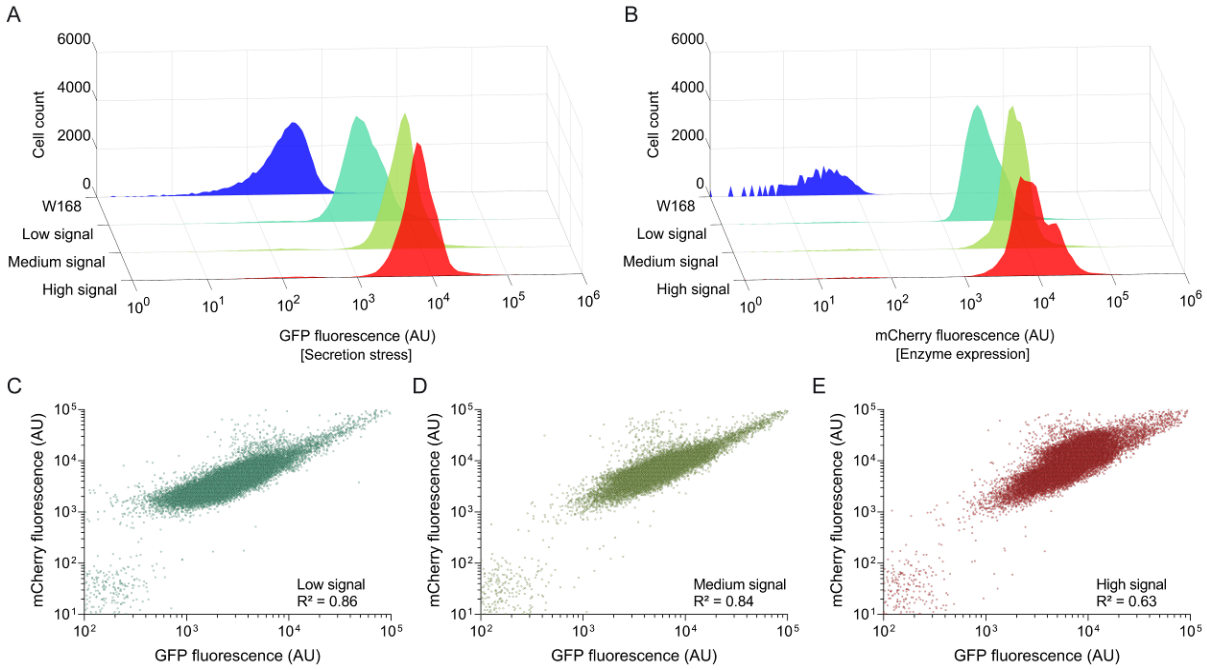


Figure 4.11: Secretion stress and enzyme expression of integration based system producing AmyM after cell sorting during stationary phase in *B. subtilis*. Strains carrying an integrated construct expressing AmyM and mCherry simultaneously and secretion stress reporter P_{htrA}-*gfp* were subjected to FACS and cells were sorted into three bins with varying mCherry intensities (low, medium and high signal; bins I-III, Figure 4.8) by a FACSAria™ Fusion flow cytometer system (BD Biosciences) and subsequent growth at 37°C in LB medium. Cells were measured for (A) secretion stress (GFP) and (B) enzyme expression (mCherry) by FACScan with the scan of 50,000 events per strain for both fluorescence signals using a LSRFortessa™ X-20 (BD Biosciences). Correlation of secretion stress with enzyme expression of sorted cells with (C) low, (D) medium and (D) high signal.

To test whether the observed mCherry levels of the different binned populations (Figure 4.10B & 4.11B) match with the AmyM production, cells from the three bins were regrown after the sorting process and the production of AmyM measured in stationary phase (Figure 4.12). *amyM* expressed from a plasmid showed no significant differences in AmyM activities (~40%) between the three bins (Figure 4.10B). However, for the integrated *amyM* construct, a trend of increased production was observed (~72% - 100%) in the bins from low to high mCherry signals (Figure 4.11B). The actual enzyme expression was thus ~2-fold higher with the integrated construct than with the plasmid. The AmyM production reflects the mCherry intensities in the FACScan data indicating that the integrated a*myM* construct became mutated, presumably as a consequence of high secretion stress induced by the strong P_{15} promoter. Possibly, the secretion stress of strains carrying the plasmid was not high enough to induce a bimodal response in their enzyme expression due to the heterogeneous P_{amyQ} activity. Nevertheless, the plasmid put sufficient pressure on the cells so that removal of antibiotics led to cells that lost the plasmid.

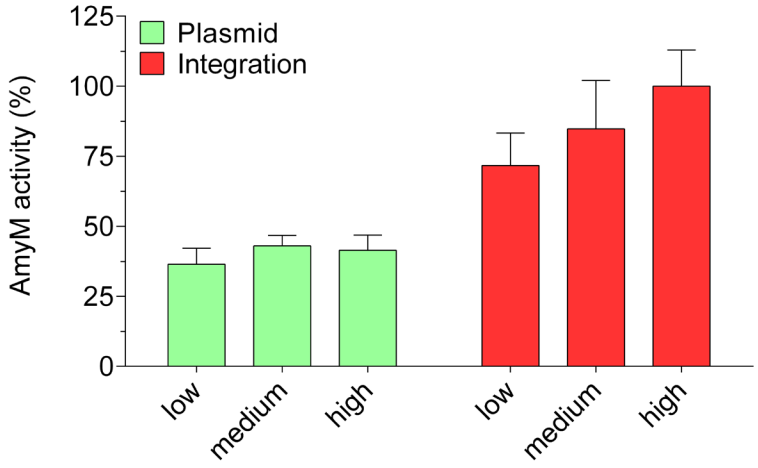


Figure 4.12: AmyM production of plasmid and integrated based system during stationary phase in *B. subtilis*. Cells carrying a plasmid (green) or integrated construct (red) expressing AmyM and mCherry simultaneously and secretion stress reporter P_{htrA}-*gfp* were subjected to AmyM activity assays after sorting cells into three bins with varying mCherry intensities (low, medium and high signal; bins I-III, Figure 4.8) by a FACSAria™ Fusion flow cytometer system (BD Biosciences) and subsequent growth at 37°C in LB medium. Enzyme activities of integrated based construct with high signal (red) was set to 100%. Error bars indicate the standard deviation of three biological replicates.

4.3 Conclusion

An important conclusion from this chapter is that the difference between *xynA* and *amyM* expression was not determined by mRNA levels. This would imply that transcription is not the limiting factor for the different production dynamics of both enzymes but possibly subsequent secretion and quality control processes. This led us to investigate the role of the four quality control proteases HtrA, HtrB, HtrC and WprA revealing a lower induction of *htrA* and *htrB* in cells producing XynA compared to a strong induction in the AmyM producing cells. Interestingly, the absence of these control proteases did not improve production, rather showing a reduced XynA production, indicating that degradation by these proteases does not account for the difference in XynA and AmyM production.

Our analyses showed a strong upregulation of secretion stress with the AmyM expression. By using an extensive combination of protease deficient strains we show that removal of either *htrA* or *htrB* results in strong increase of stress, but with no effect on production. However, removal of both control proteases resulted in very high secretion stress, affected growth and lower AmyM production. Apparently, cells require at least one of the two proteases to facilitate the quality control and production of AmyM.

Subsequently the AmyM production showed a bimodal distribution in secretion stress, possibly induced by cell chaining, but with the deletion of *htrA* or *htrB* the

observed bistability disappeared and only a single high peak was detected. This is a consequence of the compensating effect of cells requiring more HtrA (P_{htrA}) due to lower proteolytic activity (Noone *et al.*, 2000) and the upregulation of either HtrA or HtrB with the removal of the other facilitated by the CssRS two-component system (Noone *et al.*, 2001, Darmon *et al.*, 2002, Krishnappa *et al.*, 2014). The integrated construct lacked this bimodal distribution. The high *amyM* expression of the integrated construct resulted in mutated subpopulations since cells that were sorted for varying mCherry intensities and regrown, produced cell populations with different secretion stress levels that were related to the original mCherry levels. These populations also produced different AmyM levels. In contrast, cells carrying the plasmid showed no difference compared to before the sorting, leading to the conclusion that the strains were not mutated.

Chapter 5

Enhancement of *veg* expression module with synthetic biology approach

5.1 Introduction

In the previous chapter, the effect of quality proteases on industrial enzymes, and the secretion stress induced by the plasmid and integration based expression of *amyM* was discussed. The results showed the importance of quality control proteases being involved in the regulation of industrial enzymes. For the plasmid based expression a strong bimodal secretion stress response was observed compared to a more monomodal response of the integrated construct. Since bimodal promoter activities reduce the enzyme production in an industrial context, this chapter discusses a synthetic approach to enhance expression modules towards a high and monomodal gene expression.

Gene expression, transcription of DNA to RNA, and translation from RNA to protein is a multistep highly regulated process that has been optimised in organisms for maximal fitness. Industrial biotechnology aims to achieve maximal productivity of an enzyme, however, this endeavour requires a different set of biological parts than provided by nature. Synthetic biology explores the space between the natural diversity with the aim to better understand the individual parts and their interaction to allow predictable designs that is crucial for engineering microbes with tuned pathways for maximal productivity. In both academic and industrial environments, bacterial promoters are studied for the effect of mutations on regulation and transcription to understand the function of different parts. This approach has significantly contributed to our current understanding of promoters in respect of the importance of the -35 and -10 regions, transcription start and regulation by the different sigma factors. Nevertheless, this research has not yet resulted in algorithms that leads to the design of promoters with desired strength for a given gene.

In industrial biotechnology, promoter behaviour can be the key for success in strain construction and process development. Especially mutation studies have

shown the importance of single nucleotides for the achievement of stronger promoters (Vellanoweth and Rabinowitz, 1992, Jensen and Hammer, 1998, Liu *et al.*, 2004, Davis *et al.*, 2011, Liebeton *et al.*, 2014). Since industrial researchers seek to increase the expression and secretion of heterologous proteins in *B. subtilis*, the necessity to find promoters that are in line with the expected expression levels is highly demanding. However, the approach of combining promoter elements and the insertion, deletion or mutation of nucleotides is complex and time-consuming. Furthermore, promoters that show heterogeneous activities are bottlenecks for the production of enzymes since this behaviour likely leads to reduced protein yields (Ploss *et al.*, 2016).

Well-characterised promoters are one objective of current studies, however, the additional prediction of elements upstream and downstream of promoter sequences is more challenging since larger numbers of combinations are necessary to effectively evaluate these expression units. Such expression modules that lead to predictable gene expression levels are crucial in microbial engineering. To overexpress industrial enzymes in *B. subtilis,* strong expression modules are required for the high production of desired products. A strong expression module removes bottlenecks of transcription and translation during enzyme production and allows an easy, fast and high-throughput strain engineering, whereas less efficient expression modules require the integration of multiple copies, which is laborious and causes strain instability. The improved efficiency in strain engineering further opens the possibility to construct and assay larger numbers of strains with upfront designed and varying elements that allow faster strain optimisation and feasible biotechnological processes.

To achieve strong expression modules, next generation approaches are required that are directly linked to the manufacturing of synthetic modules (Mijakovic *et al.*, 2005). However, synthetic constructs based on systematic and randomised strategies are competing with its host endogenous transcriptional and translational machinery leading to additional burdens that cells have to compensate (Gorochowski *et al.*, 2016). For the achievement of enhanced expression modules in *B. subtilis*, it is thus of importance to use a native expression module as genetic basis to introduce synthetic modifications. Such modifications can include the combination of transcriptional and translational components but simultaneously leaving essential key elements unchanged e.g. for the sigma factor recognition (-35 and -10 regions) and ribosome binding to achieve a substantial gene expression (Kosuri *et al.*, 2013,

Goodman *et al.*, 2013). Since synthetic expression constructs can contain non-conserved sequence it is of interest to investigate their influence on gene expression levels. This expression varies due to gene specific interations with their upstream sequence (Salis *et al.*, 2009) indicating that large synthetic libraries are required to pinpoint expression modules showing an improved gene expression.

Therefore, an expression library containing ~12,000 sequences was created using a combinatorial design to detect expression units with high and monomodal gene expression levels in *B. subtilis*. The expression levels of the expression modules were analysed by using flow cytometry and microbioreactors with GFP fluorescence as read-out for gene expression. The expression modules were fragmented into three domains:

- UP element upstream the -35 region
- Promoter including the -35 and -10 regions, the spacer between and the sequence until the transcriptional start (+1)
- 5' UTR containing the spacer between the +1 region and RBS, the RBS itself and the part between RBS and translational start ATG

All synthetic parts were combined, generated as oligo pool and subsequently characterised by measuring the *gfp* expression. We investigated this expression library to find out which synthetic modules can be used to achieve a particular expression level in synthetic circuits and to obtain a dynamic range of expression levels. Additionally, we analysed whether the combination of synthetically engineered parts of the module library improve the strength of the native *veg* expression module. The industrial relevance of the novel expression modules was evaluated by using the strongest GFP expression units for the expression of the industrially relevant enzyme xylanase and by comparing its production with the associated GFP fluorescence.

5.2 Results and discussion

5.2.1 Design and characterisation of synthetic expression module library

The housekeeping σ^A-dependent P_{veg} promoter is one of the strongest native promoters in the *Bacillus* genome (Lam *et al.*, 1998, Radeck *et al.*, 2013). Therefore, we selected the P_{veg} promoter as a starting point for synthetic engineering by modifying its UP-element, promoter and 5' UTR region. By synthetically modifying the *veg* expression module (Figure 5.1A) and creating a large expression library, we aimed to achieve synthetic modules with distinct expression levels exhibiting a range from low to high for synthetic circuits and production strains. The design of the expression library was based on combining genetic elements that are required for an ideal transcriptional and translational initiation. The regions known to be important for a reliable expression (-35 and -10 regions, RBS) were included in the design and kept constant. The sequences in-between these conserved regions were changed to investigate their effect on expression.

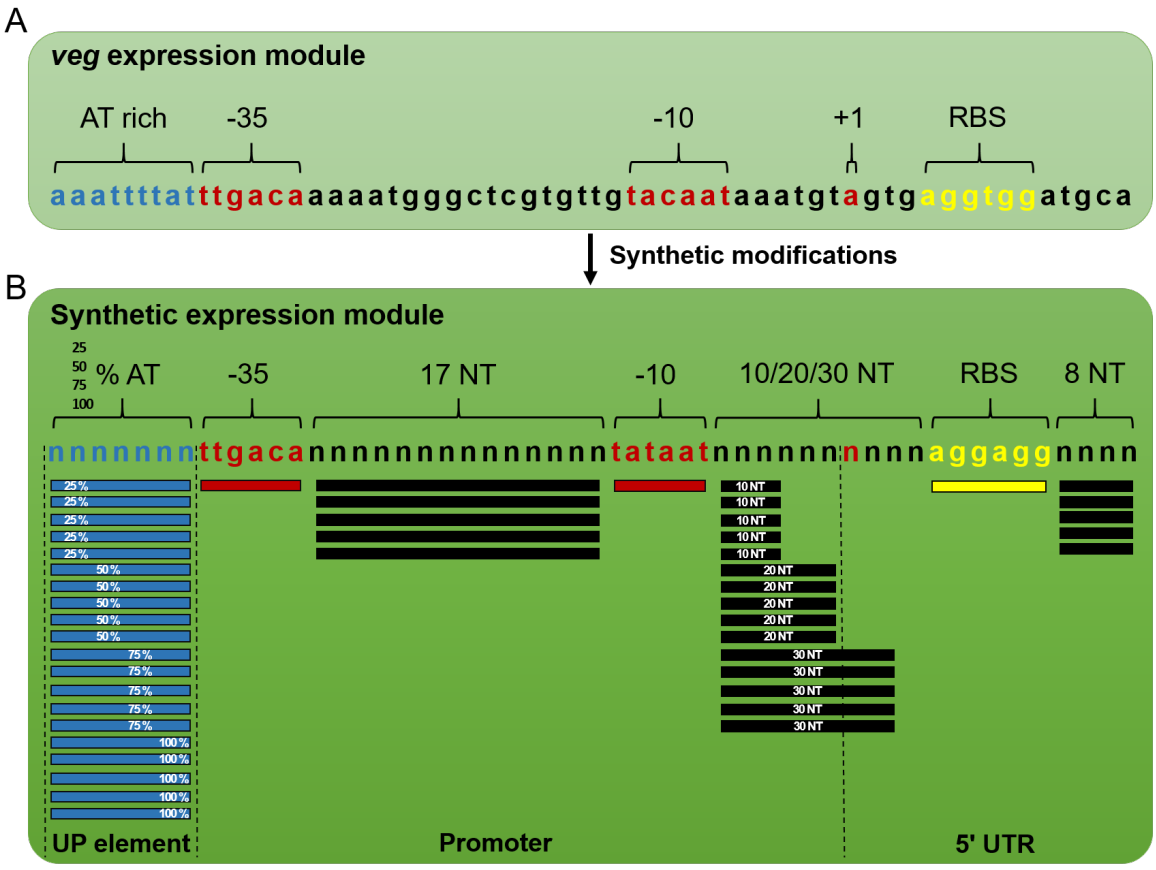


Figure 5.1: Design of synthetic expression module. (A) Sequence of the *veg* expression module. (B) Synthetic expression module with synthetic modifications subdivided in three domains: UP-element, promoter and 5' UTR. The domains were modified in their AT content (blue) of 25%, 50%, 75%, 100% (5 variants each), spacer (5 variants) between the -35 and -10 regions (red), -10 region, spacer (15 variants) between -10 region and RBS, +1 region (red), RBS (yellow), and the 8 NT spacer (5 variants) between RBS and start codon. Nucleobases (n) are reflected by either adenine, thymine, guanine or cytosine.

The UP-element was investigated by analysing the AT vs. GC content upstream of the -35 region by exchanging the sequence with 20 different DNA sequences containing a variable AT content. These sequences were randomised with restriction to their AT content. Five fragments for each of the following four different parts were generated containing either 25% AT, 50% AT, 75% AT or 100% AT content (Figure 5.1B), whereby sequences containing a higher AT content increase the interaction of the UP element with the α-subunit of the RNA polymerase (Aiyar *et al.*, 1998, Caramori and Galizzi, 1998, Meijer and Salas, 2004, Phan *et al.*, 2012). Five variants (75% AT content) for the 17 nucleotide spacer between the promoter sequences TTGACA (-35 region) and TATAAT (-10 region) to allow an ideal sigma factor A recognition and improve RNA polymerase binding (Moran *et al.*, 1982, Helmann, 1995, Jarmer *et al.*, 2001). 15 variants of the spacer between the -10 region and RBS (five for each: 10, 20 and 30 bp) including the +1 region (5' UTR) for the start of transcription and the RBS itself to allow optimal ribosome binding. Since the consensus of the Shine-Dalgarno sequence is 'AGGAGG' being complementary to the 3' end of the 16S rRNA (Osada *et al.*, 1999), this sequence was selected as RBS in the expression module design. Five variants (75% AT content) of the 8 NT spacer between RBS and start codon to improve translation initiation efficiency (Vellanoweth and Rabinowitz, 1992). These variants of four different parts were combined, with the addition of the native *veg* expression module sequences (Figure 5.1A), to achieve a library of 12,096 synthetic expression units (21x6x16x6).

To quantify growth and development of the strain library reflected by transcriptional and translational activities of individual cells, mCherry levels were used as internal reference. The mCherry expression module containing the constitutive P_{infC} promoter (Nicolas *et al.*, 2012) was integrated in the *nprE* locus of 168 (Figure 5.2A). The obtained strain was used as host for the expression library and by applying StarGate® reactions the synthetic library was cloned upstream of *gfp* and integrated as a single copy in the *amyE* locus (Figure 5.2B). The production of GFP and level of fluorescence was taken as a measure for gene expression of the synthetic library and these GFP intensities were normalised by mCherry levels.

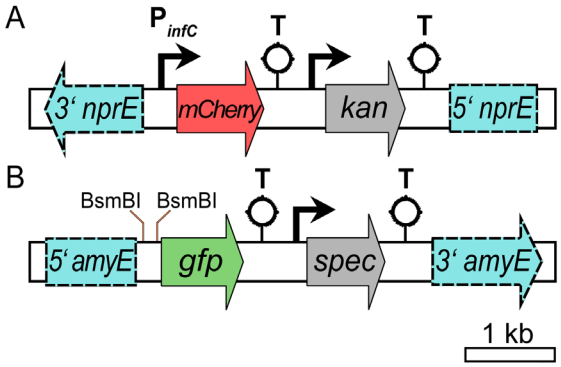


Figure 5.2: Schematic overview of the backbones for the internal reference and expression library. (A) P_{infC} promoter is fused to mCherry as internal reference and colonies are selected by kanamycin resistance (*kan*). (B) BsmBI restriction sites indicate the insert location of the expression modules upstream of *gfp*. The colonies containing the expression modules are selected by spectinomycin resistance (*spec*). Terminators (T) prevent read-through into adjacent genes. Disrupted integration loci (*amyE*; *nprE*) are indicated with dashed lines.

The experimental design was validated by generating a limited number of transformants containing the synthetic expression modules. A total of 192 colonies were randomly picked and sequenced to verify the synthetic library. This revealed that only ~34% of the expression module sequences matched the bioinformatics design of ~12,000 variants. The other sequences carried point mutations or deletions leading to a more diverse expression library than anticipated. After removing duplicates and non-active modules due to deletions, 135 colonies were grown overnight and subjected to end point measurements of GFP and mCherry to verify the dynamic range of the expression modules (Appendix A, Table A.1). The strength of expression modules was measured by the ratio of GFP activities, reflected by synthetic expression modules, and mCherry levels, reflected by the constitutively active mCherry reporter (Figure 5.3). The heterogenic distribution showed GFP/mCherry ratios from ~10 reflecting strong expression modules down to ~0.02 revealing weak expression units with an average of ~1.8. This ~500-fold difference of expression strength indicates an enormous dynamic expression range and large numbers of variations in the synthetic module sequences.

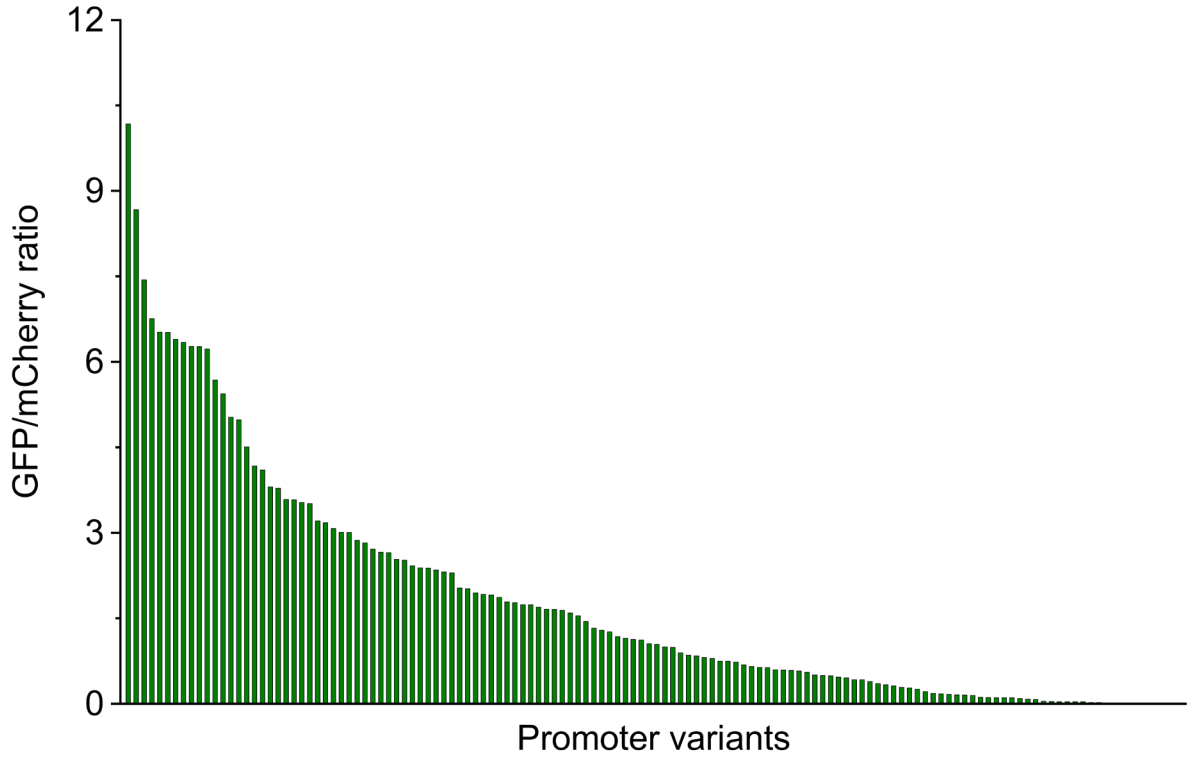


Figure 5.3: GFP/mCherry ratio of library variants after overnight growth. 135 expression modules were randomly picked from plate and grown overnight. Subsequently end point measurements of GFP and mCherry were performed in a plate reader to determine expression strength reflected by the GFP/mCherry ratio.

The proof of principle demonstrated that all parts were in place for generating the strain library, sorting the library by FACS into bins with a range of GFP intensities and analysing the different bins by DNA sequencing. A library of ~152,000 clones was generated reflecting a ~13-fold clonal coverage of the ~12,000 variants in the synthetic library. The strain library was grown until early exponential phase to subsequently measure the cells fluorescence by FACScan and determine their expression module strength (GFP) and growth and development (mCherry) (Figure 5.4). The GFP fluorescence intensity reflects gene expression levels with an intensity range between $>10^0$ AU and $>10^5$ AU. The GFP fluorescence plot shows a bimodal distribution with one peak of cells showing lower GFP intensities ($\sim10^2$-10^3 AU) and the second peak with highly elevated GFP levels ($\sim10^4$ AU) (Figure 5.4A).

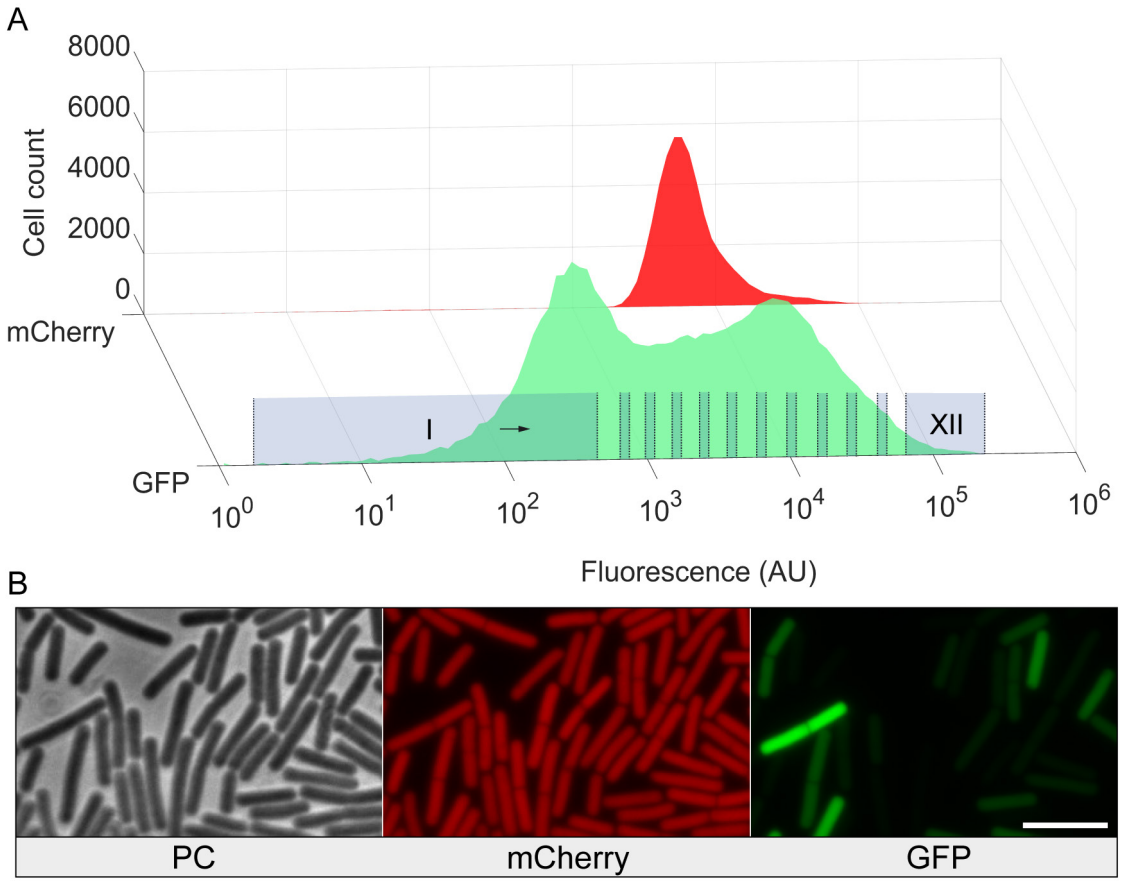


Figure 5.4: Gene expression of the library during exponential phase. Expression library of 152,000 clones was grown until early exponential phase. (A) Cells were subjected to FACScan to detect mCherry and GFP fluorescence signals using a LSRFortessa™ X-20 (BD Biosciences). After the scanning process, cells were sorted into 12 nonadjacent log-spaced bins (I to XII) with increasing GFP intensities by a FACSAria™ Fusion flow cytometer system (BD Biosciences). (B) Microscopic analysis of heterogenic expression module library. Scale bar is 5 µm.

Overall, there is a high cell count of GFP fluorescence between $>10^2$ AU and 10^4 AU indicating a large diversity in the activities of the expression modules. The measurement of mCherry fluorescence showed a single sharp peak ($<10^4$ AU) suggesting that growth and development was similar in most cells (Figure 5.4A). The large range of gene expression was confirmed by inspecting a sample of the unsorted strain library and performing microscopic analyses of phase contrast, red florescence (mCherry) and green florescence (GFP) (Figure 5.4B). This showed a strong heterogeneity in GFP fluorescence reflecting large variations in the activities of the expression modules and a homogeneous distribution of the red florescence indicating a similar growth and development of the cells (mCherry).

The library was then subjected to FACS to sort cells into 12 nonadjacent log-spaced bins with increasing GFP intensities (Figure 5.4A, blue bins from I to XII). The sorting was validated by analysing three populations with low, medium and high GFP activities (I, VI, XII) and fluorescence microscopy to determine the success of sorting indicated by increasing GFP levels on the single-cell level (Figure 5.5). The results confirm that bin one has the lowest GFP signal, bin VI a moderate GFP signal and bin XII the strongest GFP signal. The growth status (mCherry) of the cells is similar. Clearly, the FACS sorting resulted in populations of cells with different *gfp* expression levels.

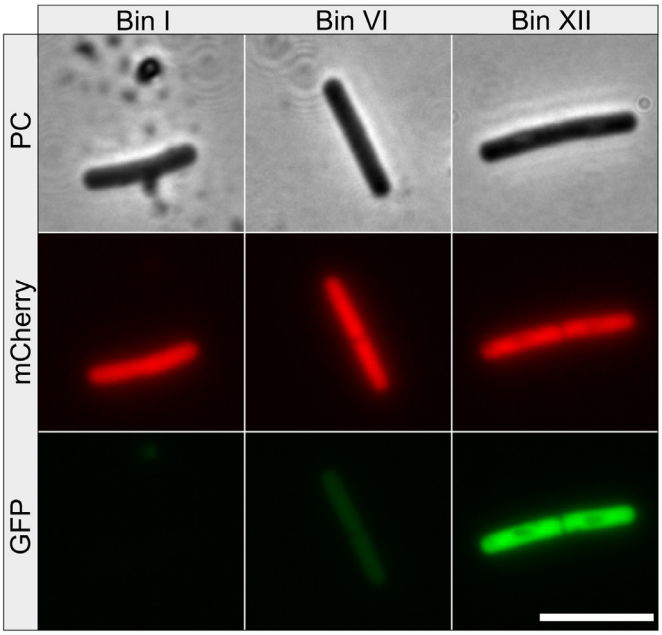


Figure 5.5: Microscopic images of representative cells in bins I, VI and XII. Immediately after sorting, cells were subjected to microscopic analyses. Phase contract (PC), red florescence (mCherry) and green fluorescence (GFP) images of individual cells are shown. Scale bar is 5 µm.

The 12 subpopulations of each bin were further characterised by a FACScan analysis (Figure 5.6). The FACScan of each individual bin confirmed the preliminary microscopic observation that subpopulations with different GFP intensities were sorted by FACS. The scanned bins showed the expected GFP signal range with a low fluorescence in bin I and a steadily increasing GFP intensity up to bin XII (Figure 5.6A). Each bin shows a peak in its GFP activity, however, all bins and especially bin X and XI show a long tail with a substantial amount of cells with lower GFP intensities (>10^1 AU). These intensities either reflect expression modules with lower strength,

despite the log-spaced sorting of the cells, or heterogeneously active units. The analysis of the growth status revealed a comparable distribution of mCherry signal (<10^4 AU) (Figure 5.6B) compared with the cells before sorting (Figure 5.4A). Similar to the tails of the GFP intensities, the cells in the bins, particularly bin X and XI, display a tail with cells showing low RFP values. Since these tails of GFP and mCherry seem to form separate populations, both fluorescent proteins were plotted against each other to investigate the population characteristic (Figure 5.6C). This analysis shows four populations in bin X and XI with one being virtually off for both GFP (10^2 AU to 10^3 AU) and mCherry (10^2 AU), the second with increased GFP (10^4 AU) and mCherry (10^2 AU to 10^3 AU) signals, the third with the highest GFP (10^4 AU to 10^5 AU) and mCherry (10^3 AU to 10^4 AU) levels and the fourth with lower GFP (10^3 AU) and similar mCherry (10^3 AU to 10^4 AU) fluorescence (Figure 5.6C). The first population presumably reflects cells carrying no expression construct or reporter for growth and development. The second population contains cells that are showing an appropriate expression module activity but vary from the growth status (mCherry) likely being in an earlier growth phase compared to cells of the third population that apparently reflects the culture of interest with the most cell counts. The fourth population likely shows cells that were carried over from the previous bins I to IX due to lower GFP signals. The other bins contain different amounts of populations, mainly population one and three and in the higher bins (V to XII) also population four indicating that the integrity of cell sorting was insufficient. To test whether the sorted cells would maintain their expression module activities and show the increasing GFP intensities of bin I to XII on the single-cell level, cells of each subpopulation were streaked on agar plates and after an overnight incubation a colony of each bin was randomly picked and analysed via fluorescence microscopy (Figure 5.7).

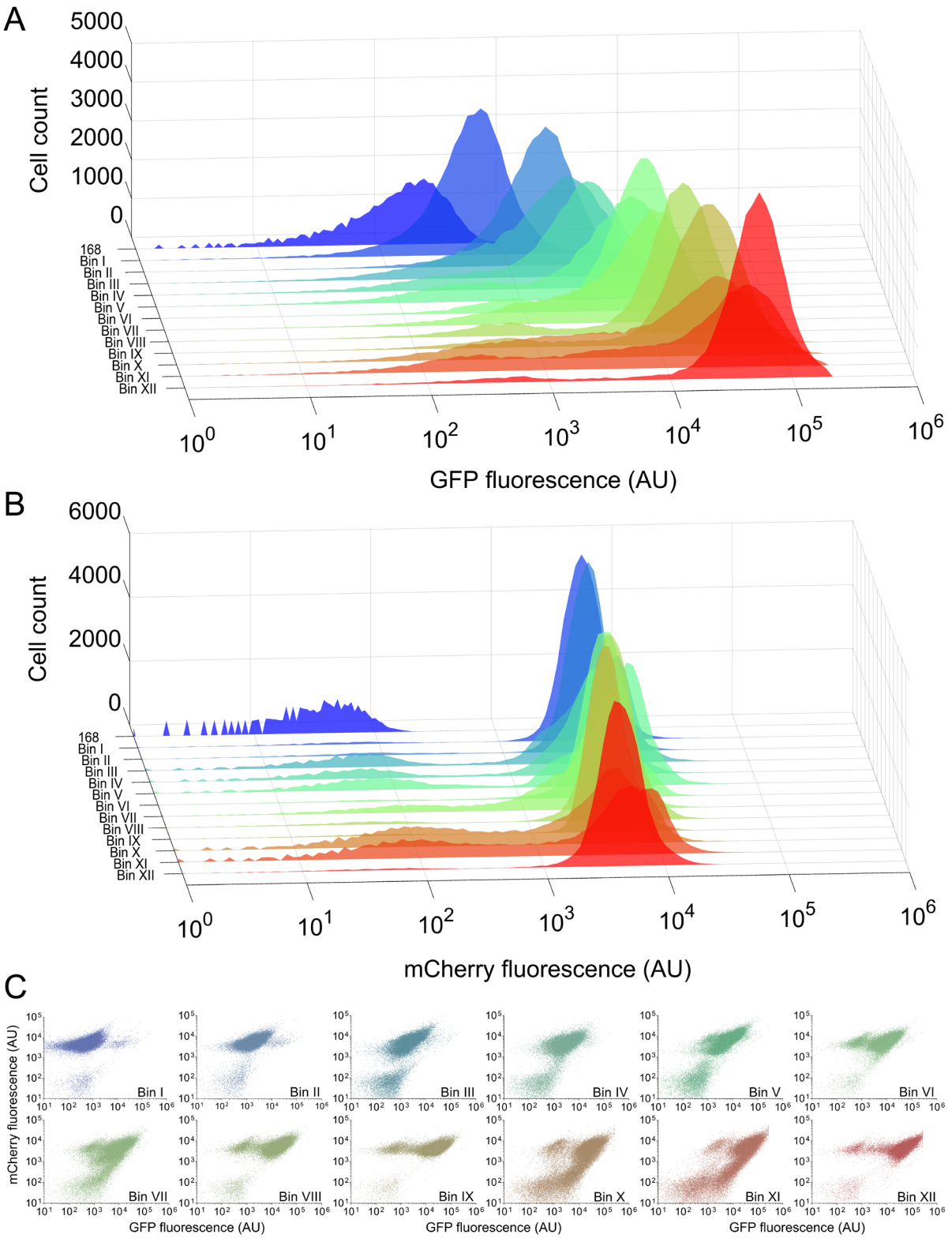


Figure 5.6: GFP and mCherry expression of subpopulations in bin I to XII. The subpopulations of the 12 bins and the parent strain 168 were grown overnight and subjected to FACScan by scanning 50,000 events. (A) GFP fluorescence (AU) and (B) mCherry fluorescence are shown. (C) GFP against mCherry fluorescence of individual cells.

The microscopic pictures reveal an increasing GFP signal of the colonies of bin I to XII, however, cells of bin X show a lower than expected GFP level suggesting that the picked colony belonged to the previously discussed tail of bin X including cells with lower fluorescence (Figure 5.6A).

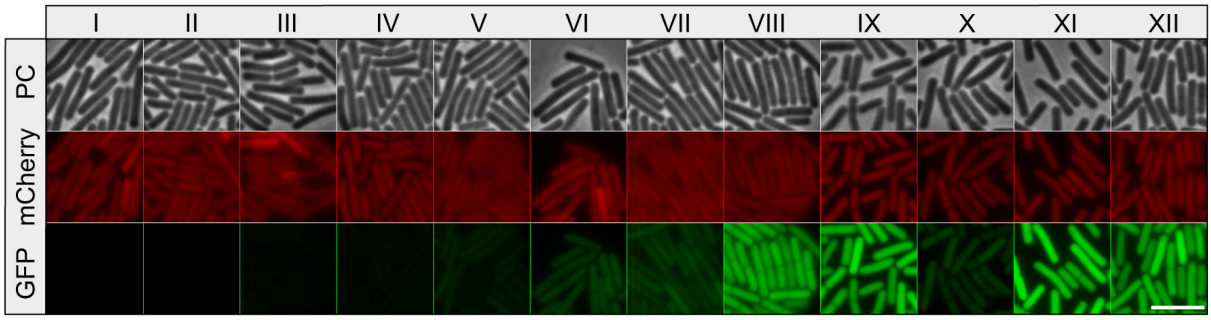


Figure 5.7: Microscopic images of cell populations in bin I to XII. Cells of all 12 bins were streaked on separate agar plates and incubated overnight. One colony of each bin was randomly picked and subjected to microscopy. Phase contract (PC), red fluorescence (mCherry) and green fluorescence (GFP) images are shown. Scale bar is 5 µm.

5.2.2 Industrial application for synthetic expression modules

The expression module library showed that synthetic modifications of the native *veg* expression unit results in low up to highly elevated module activities. The GFP levels were used as read-out for the expression strength but the expression will be gene dependent due to differences in translation efficiency and interactions of the 5' UTR and the gene of choice (Goodman *et al.*, 2013). To investigate this gene dependency in expression strength we screened for synthetic modules that show high and monomodal GFP intensities and applied these units for the expression of the industrially relevant enzyme xylanase to analyse if the native *veg* expression module can be improved by combining synthetically engineered parts of the expression library. XynA was used as a model case to identify expression modules that result in high enzyme production titres. 16 expression modules with high GFP levels were selected (Appendix A, Table A.2), verified for their GFP response and transcriptionally fused to *xynA* to correlate GFP activity with XynA production to examine if expression modules identified by the FACS analysis can be applied for an industrial setting. To verify the high GFP intensities of the selected expression units, firstly one variant (26) was compared to strains carrying an empty control lacking an expression unit upstream of *gfp* (empty), the native *veg* expression module (*veg*) and the originally designed expression unit (design) being the prototype of the synthetic library (Appendix A, Table A.2). The design unit was constructed similarly to the synthetic library containing all the key elements but aiming for a sequence with high

AT content (~75%) and adjusted nucleotides based on the base conservation of Helmann, 1995. These strains were grown on a plate overnight and with the use of blue light, snapshots of the grown colonies were taken to visually compare the GFP differences (Figure 5.8). By measuring the GFP values with ImageJ, Variant 26 shows by far the strongest GFP signal (~78 AU) followed by the design module displaying half the signal strength (~38 AU), the *veg* expression module with 3.5-fold lower intensity (~22 AU) and the empty control with 14-fold lower GFP activity (~5.5 AU). These GFP differences indicate the potential of the synthetic library.

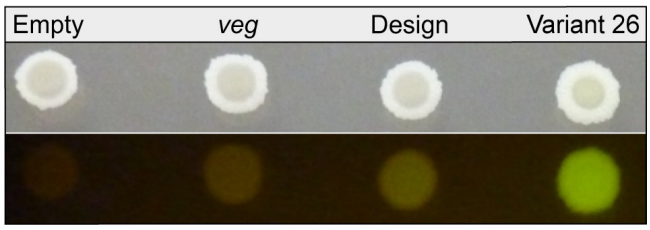


Figure 5.8: White and blue light images of *B. subtilis* colonies on agar plate. White light image (top) and blue light image (bottom) are shown. First colony with no expression module upstream of *gfp* (empty), vegetative expression module (*veg*), design expression module (design) and expression variant 26.

Since the first test was successful, all 16 expression variants were subjected to BioLector® experiments while measuring the cells biomass, expression module activity (GFP) and metabolic activity (mCherry) simultaneously to compare the strengths of the modules on the population level (Figure 5.9). This revealed enormous differences on the GFP level during exponential phase (~3 h) with a ~13-fold higher GFP signal of variant 26 compared to the *veg* expression unit suggesting a striking impact of the synthetic modifications on the native *veg* module (Figure 5.9B). The other selected variants showed GFP signals that were similar or lower than variant 26 but higher compared to the controls including the empty module, *veg* unit, the industrial phage P_{15} module (Chapter 3 & 4) and the synthetic rational design unit. The module activities decreased from transition phase (~4 h) until early stationary phase (~6 h) indicating a degradation of GFP, however, GFP is a very stable protein and should maintain during the assay but perhaps the sensors of the device were disturbed by the entry into stationary phase with a large biomass increase. The measurement of biomass and mCherry reporter (metabolic activity), respectively, revealed a similar behaviour between the strains during the assay (Figure 5.9A & C). Only the strains carrying the empty module showed a marginally higher mCherry signal and variant 26 a lower mCherry intensity compared to the other strains indicating a slightly different growth behaviour.

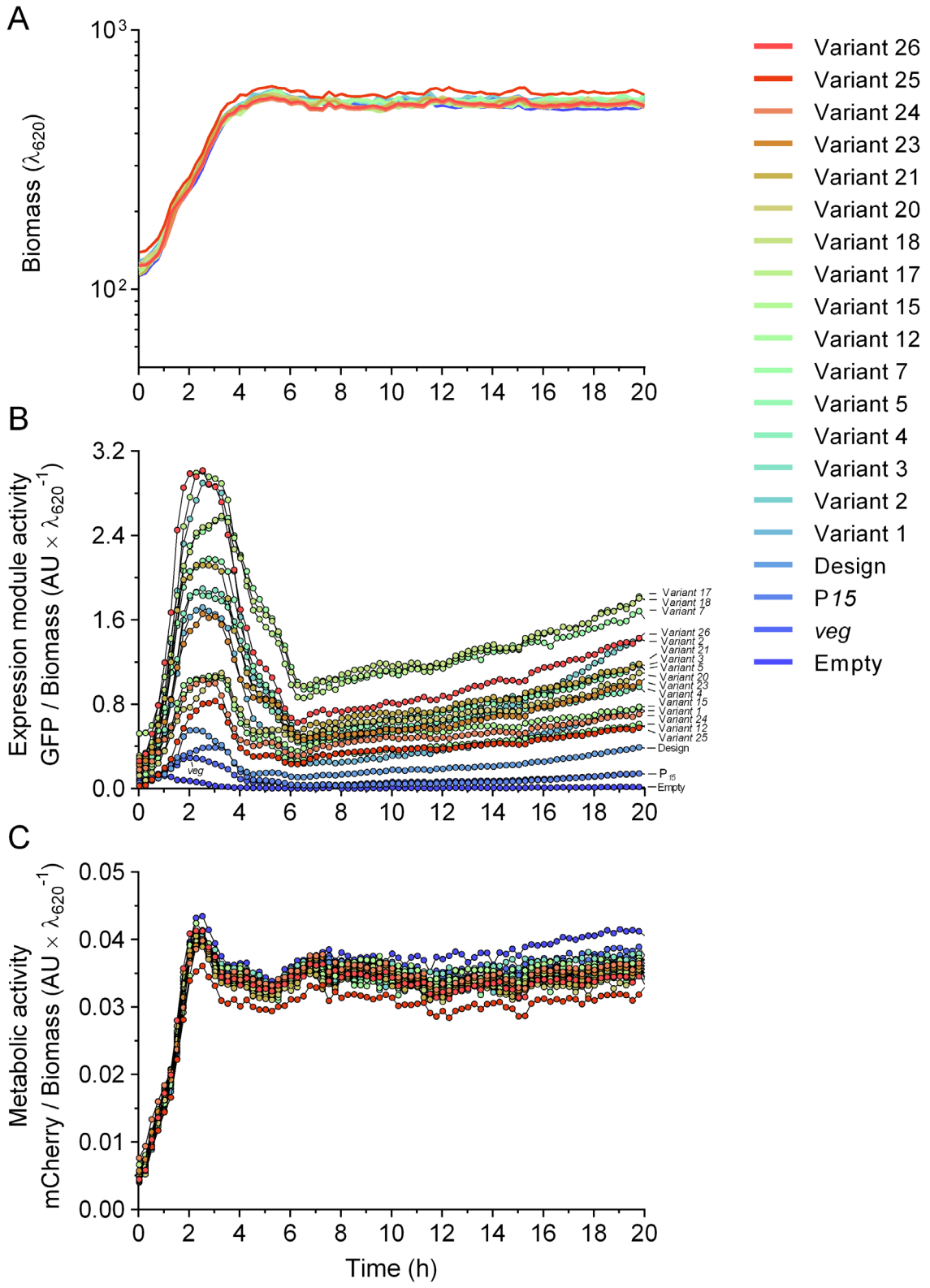


Figure 5.9: Growth, expression module comparison by GFP levels and metabolic activities by mCherry levels. Strains containing sixteen selected synthetic expression modules (1, 2, 3, 4, 5, 7, 12, 15, 17, 18, 20, 21, 23, 24, 25, 26) and control strains carrying no expression unit (empty), vegetative expression module (*veg*), constitutive phage module (P_{15}) or synthetic rationally designed expression module (design) (blue). (A) Biomass, (B) GFP intensities of the different expression modules and (C) mCherry fluorescence reflecting the metabolic activity of each population are shown (relative activities - GFP or mCherry per unit biomass). Strains were grown in quadruplicate in a BioLector® for 20 h in LB medium and the average values are shown.

Since the BioLector® studies of the expression variants displayed large differences in their activities on the population level, cells were measured during exponential phase for their GFP and mCherry fluorescence on the single-cell level by FACScan to analyse the heterogeneity of synthetic units that likely leads to a reduced protein production in an industrial context (Figure 5.10). The FACScan revealed, similar to the BioLector® experiments, substantial differences on the GFP level with an up to ~16-fold higher GFP signal of variant 26 compared to the *veg* module demonstrating a remarkable influence of the synthetic modifications on its native unit (Figure 5.10A). The other expression variants revealed comparable GFP levels with higher intensities than the control strains containing the *veg* expression module, P_{15} unit and the design module. The BioLector® and FACScan results of the expression variants differ in some cases which is likely a matter of timing of the FACScan screening. Apart from the increased GFP levels, variant three shows three peaks likely indicating a mixed population due to picking more than one colony and the variants one, four and seven show a bimodal distribution with one minor peak of cells displaying close to no fluorescence and one peak with highly elevated GFP levels. The mCherry signals were almost identical confirming a similar growth and development of cells with the exception of variant twelve displaying a tail of cells with lower fluorescence (Figure 5.10B).

The analysis of the expression module variants by BioLector® and FACScan experiments showed remarkably the enhancement of the *veg* module by synthetic modifications. This improvement on the GFP level revealed that up to ~16-fold differences can be achieved, however, it is difficult to predict XynA production levels with the identical expression units since there are likely to be gene specific interactions between the 5' UTR and 5' end of the particular gene. To determine whether the XynA production correlates with GFP levels and to identify strongly *xynA* expressing modules, the previously analysed 16 expression modules variants were fused to *xynA* and subsequently analysed in the BioLector® for their XynA production in exponential, transition and stationary phase, and mCherry activities (Figure 5.11).

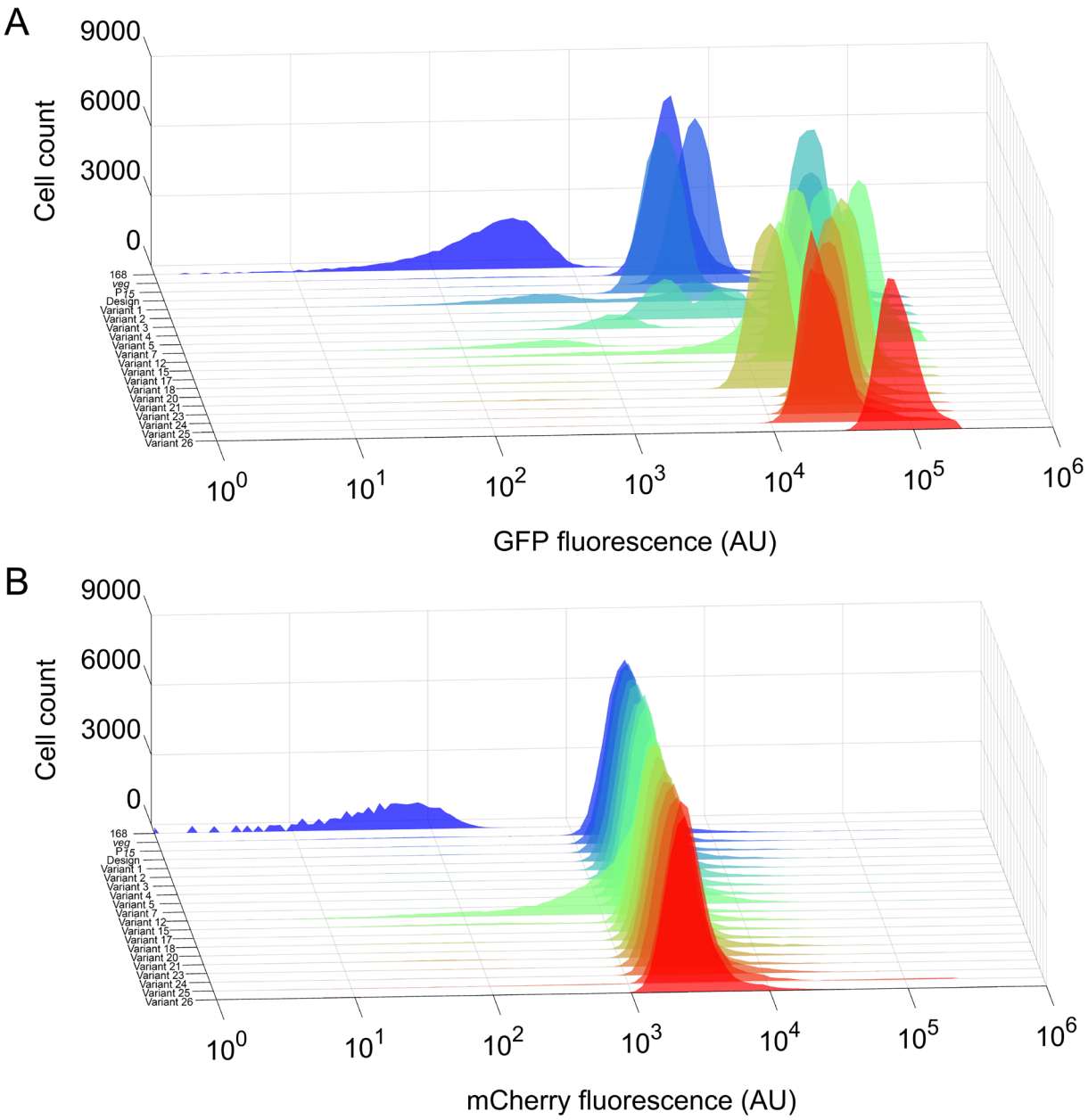


Figure 5.10: Characterisation of expression modules. Sixteen expression module variants (1, 2, 3, 4, 5, 7, 12, 15, 17, 18, 20, 21, 23, 24, 25, 26) and control strains carrying the vegetative expression module (*veg*), constitutive phage expression module (P_{15}) and rational designed expression module (design). (A) GFP fluorescence intensities and (B) mCherry fluorescence signals of expression units are shown. Cultures were grown in LB medium until exponential phase and subjected to FACScan by scanning 50,000 events per strain for both fluorescence signals using a LSRFortessa™ X-20 (BD Biosciences).

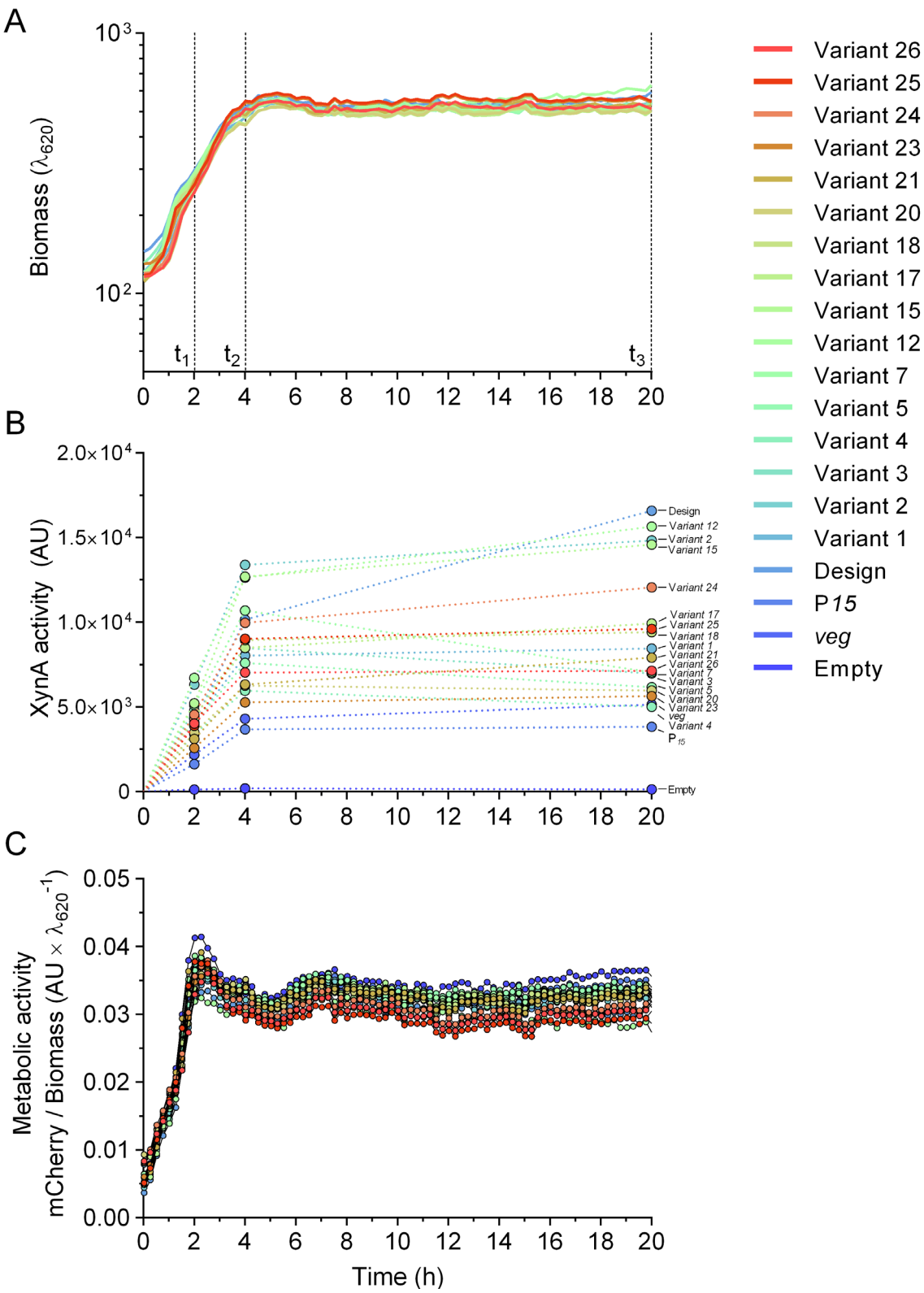


Figure 5.11: Growth, XynA production and metabolic activity of strains with strong expression module. Strains containing expression modules (1, 2, 3, 4, 5, 7, 12, 15, 17, 18, 20, 21, 23, 24, 25, 26) and control strains carrying no expression module (empty), the vegetative expression module (*veg*), constitutive phage expression module (P_{15}) and rational designed expression unit (design) (blue). (A) Biomass, (B) XynA production in exponential, transition and stationary phase (t_1, t_2, t_3) and (C) mCherry expression reflecting metabolic activity of each population are shown (relative activities - GFP or mCherry per unit biomass). Strains were grown in quadruplicate in a BioLector® for 20 h in LB medium and the average values are shown.

The experiments revealed that cells doubled their XynA production from exponential to transition phase in the culture medium (Figure 5.11B; t_1 to t_2) but afterwards the XynA production reached a plateau in stationary phase, except for the design expression module that revealed a remarkable ~3.2-fold higher XynA production compared to the *veg* unit. A few strains showed a marginally increased (variants 12, 15, 17, 21, 24) or decreased XynA production (variants 3, 4, 5, 7) compared to transition phase, and, most selected variants showed XynA levels above that of the control strains (empty, *veg* and P_{15} modules). The dynamics of the biomass and mCherry reporter for growth and development, respectively, were similar between the strains during the BioLector® assay (Figure 5.11A & C). For the correlation of both GFP activities and XynA production levels, the ratios of GFP/mCherry and XynA/mCherry of three time points (Figure 5.11 & 5.12; t_1, t_2, t_3) were calculated to achieve the expression module strength associated with the growth status of the cells represented by the mCherry reporter. These ratios were then sorted for an increasing GFP/mCherry ratio (Figure 5.12A &B) suggesting that identical variants with a distinct gene of interest show different GFP activity and XynA production levels. Interestingly, variant 2 showed a high GFP intensity and XynA production reflecting an expression module that might be of interest for both academia and industry. The correlation of both ratios revealed a poor relationship between GFP activity and XynA production (Figure 5.12C) confirming that expression modules identified by the FACS analysis can only be partially translated to other genes due to gene specificity. This fact is supported by the originally design expression module revealing low GFP activities but substantial XynA production levels. Hence, strains carrying synthetic expression units showed elevated XynA production levels with up to ~3.2-fold more than the native *veg* expression unit but compared to the up to ~16-fold higher GFP fluorescence levels in the earlier assays, this reduced improvement underlines the expression differences when using an identical expression unit with a different gene of interest. Furthermore, the reduced fold-differences in XynA production between highest and lowest expressing strains might indicate the difficult process proteins undergo once they are being secreted (quality control) and translocated into the culture medium (degradation).

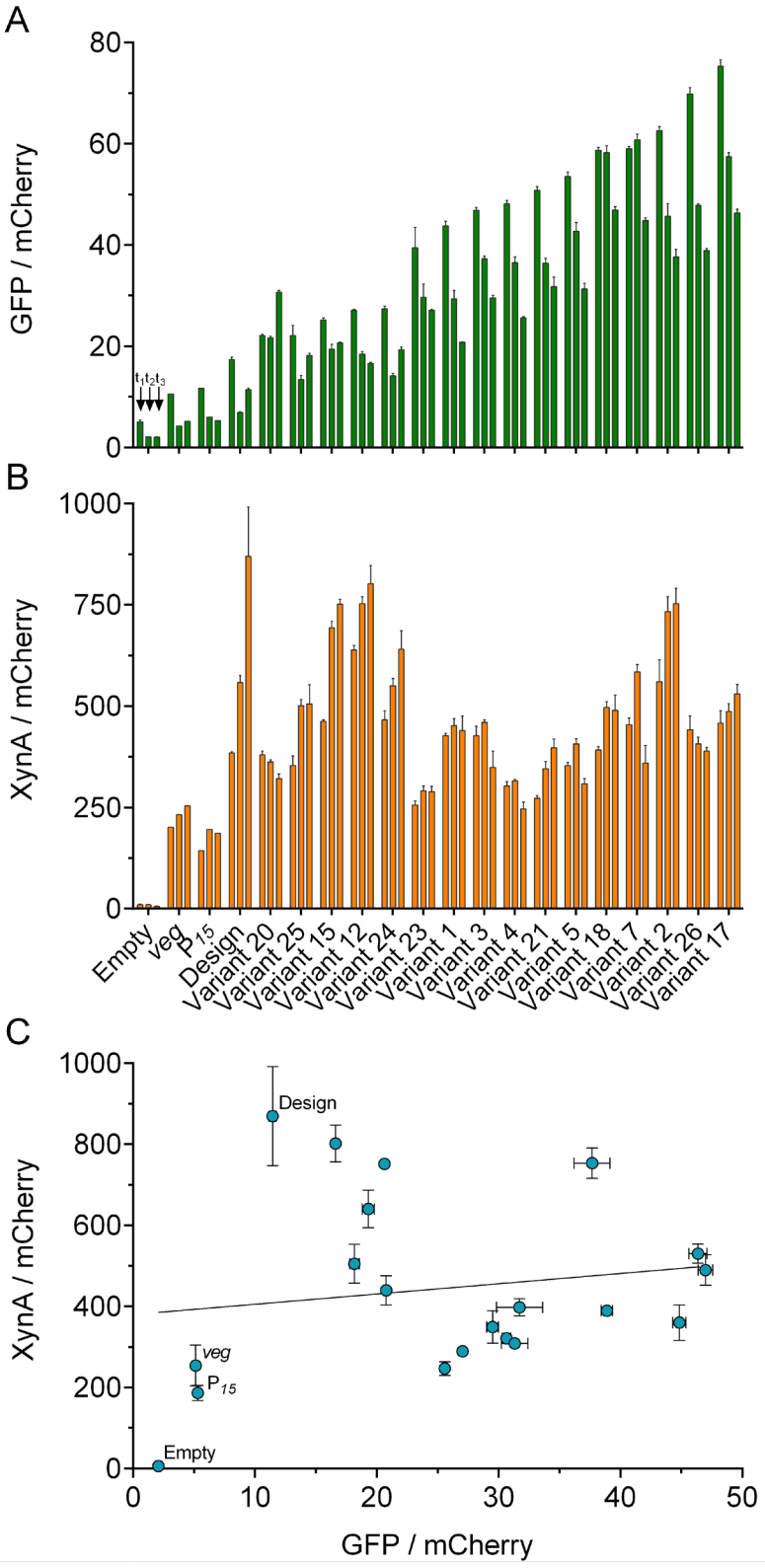


Figure 5.12: Correlation between GFP expression and XynA production of selected expression modules. Strains containing expression modules (1, 2, 3, 4, 5, 7, 12, 15, 17, 18, 20, 21, 23, 24, 25, 26) and control strains carrying no expression module (empty), the vegetative expression module (*veg*), constitutively phage expression module (P_{15}) and rational designed expression module (design.) Ratios of (A) GFP/mCherry and (B) XynA/mCherry are shown for each strain in exponential, transition and stationary phase (t_1, t_2, t_3). (C) Correlation between GFP expression and XynA production in stationary phase is shown by the regression line (R^2 = 0.02). Strains were grown in quadruplicate in a BioLector® for 20 h in LB medium and the average values are shown. Error bars indicate the standard deviation of four biological replicates.

5.3 Conclusion

The native *veg* expression module was enhanced by applying a rational synthetic design approach to modify its sequence that led to strong expression modules for the high expression of *gfp* and *xynA*. The synthetic library containing ~12,000 combinations led to improved expression units demonstrated by a dynamic range of GFP levels between $>10^0$ AU and $>10^5$ AU. The cell sorting into 12 nonadjacent log-spaced bins to separate the GFP intensities from low to high confirmed this dynamic range and the averaged GFP intensities of the strongest expression modules in bin twelve showed a two log higher GFP signal compared to the weakest expression modules of bin one. This broad range in *gfp* expression allows targeting expression modules with a particular expression that is required for the development of synthetic circuits in both academic and industrially relevant settings.

A selection of sixteen strong GFP expression modules was tested in combination with the industrial enzyme product XynA. Although selected for their strong GFP intensity on plate, in liquid culture these expression modules showed a remarkable range in their module activities with variant 26 exhibiting an up to ~13-fold higher *gfp* expression compared to the *veg* expression module. For industrial production purposes a heterogeneous gene expression is undesired as it results in fluctuation outputs between production batches. Therefore, the level of heterogeneity in *gfp* expression was determined revealing that three out of sixteen expression modules showed a bimodal distribution in GFP fluorescence. Consequently, such expression modules would be discarded for the industrial production of bioproducts. The strong expression modules were then tested for the expression of xylanase. The different xylanase expressing strains showed a large range in their xylanase production levels. Nevertheless, the selected modules showed an up to ~3.2-fold higher XynA production than the native *veg* expression module. This demonstrated that expression modules with high *gfp* expression levels also result in high xylanase production levels and that the selection of strong *gfp* expression modules can be used to increase the production of XynA. The improvement in XynA production is remarkable in respect of an industrial context where such a production increase is difficult to achieve. However, *gfp* expression and XynA production levels poorly correlate indicating that modules optimised for *gfp* expression are not necessarily matching the expected XynA production levels emphasizing the gene specific interaction of the 5' UTR and the 5' end of the gene of choice. Further optimisation of

the expression module screening system to directly select the best expression modules for the enzyme of interest would be a much desired tool.

Chapter 6

Influence of chromosomal location on gene expression

6.1 Introduction

The focus of previous chapters was primarily on the approaches to optimise the production of the industrial enzymes XynA and AmyM. This chapter focuses on the influences of chromosomal location and transcriptional orientation on productivity in general. To this end we have developed a transposon system for the randomised integration of a novel insertion cassette comprising divergent reporter genes.

The organization of a bacterial chromosome is far less complex than that of eukaryotic genomes (Kaplan *et al.*, 2009, Le *et al.*, 2013). While eukaryotic genomes are compartmentalized in a membrane bound nucleus, and their DNA spatial organized into heterochromatin that directly affects gene expression (Akhtar *et al.*, 2013, Wilson *et al.*, 1990), prokaryotic organisms generally have a singular genome that can be either linear or more often circular in structure. It is also distributed throughout the cytoplasm which it shares with other cellular components such as the ribosome and structural proteins such as MreB and FtsZ. In bacteria with covalently closed circular (ccc) chromosomes, replication proceeds bidirectionally from the origin of replication (*oriC*) until the two replication forks meet within the termination region (*ter*) approximately 178.7° from *oriC* (Bussiere and Bastia, 1999, Murray and Koh, 2014, Kunst *et al.*, 1997, Barbe *et al.*, 2009). For bacteria such as *B. subtilis* and *E. coli*, it takes about 40 minutes to replicate the entire chromosome at 37°C. When cells are growing rapidly (*e.g.* mean generation time <40 min), new rounds of replication are initiated at *oriC* before the previous round reaches the terminus. As a result, the copy number of genes close to *oriC* can increase exponentially (*e.g.* 2, 4, 8, …) in comparison with genes close to *ter* (Sousa *et al.*, 1997, Couturier and Rocha, 2006). This is likely to influence the supply of gene products in rapidly growing cells and indeed it is usual to find multiple copies of ribosomal operons in the regions either side of *oriC*.

In addition to gene location and its potential influence on gene dosage, the orientation of a gene might influence the activity of genes. Interestingly genomic sequencing studies revealed (Kunst *et al.*, 1997, Barbe *et al.*, 2009) that most genes (~75%) in *B. subtilis* are encoded on the leading strand being co-directional with the replication of DNA (Zeigler and Dean, 1990, McLean *et al.*, 1998, Rocha, 2004). In other organisms, this co-directionality leads to reduced fork collapses and collisions between replication apparatus and RNA polymerase (Washburn and Gottesman, 2011, Mirkin and Mirkin, 2005). In the case of ribosomal genes, 94% are encoded on the leading strand, which indicates a strong bias for the localisation of strongly expressed genes on this strand (McLean *et al.*, 1998). What is not clear is whether this bias is important for replication, transcription or both.

In previous studies it was shown that gene dosage depends on the proximity towards the origin of replication in rapidly growing *E. coli* cells by altering the genomic position of a reporter gene (Block *et al.*, 2012). A different study presented results that concluded that chromosome organisation rather than gene dosage impacts the gene expression in *E. coli* (Bryant *et al.*, 2015). A limitation of previous studies is that integration sites have not been selected randomly, and this may have unconscientiously biased the outcome. To overcome any potential bias, a random integration system was designed to explore the impact of location and orientation on gene expression in *B. subtilis*. To this end a transposon-based integration cassette, containing divergently expressed heterologous reporter genes, was constructed to facilitate random integrations into the bacterial chromosome with the subsequent measurement of reporter genes expression. This research was published and is attached in the appendix B.

6.2 Results and discussion

6.2.1 Transposon-mediated integration

For an unbiased integration approach an existing Mariner-derived transposon system was applied (Lampe *et al.*, 1996, Le Breton *et al.*, 2006) that was previously used to modify chromosomes in a number of organisms (Rubin *et al.*, 1999, Choi and Kim, 2009, Picardeau, 2010). The constructed transposon plasmid, pSS125 (Appendix A, Figure A.7) (constructed by Simon Syvertsson), originated from pMarB (Le Breton *et al.*, 2006), and consisted of an expression cassette encoding *lacZ* and *sfGFP* driven by the inducible P_{spac} promoter and constitutively active P_{veg} promoter, respectively (Figure 6.1). This cassette was flanked by transcription terminators, selected from the

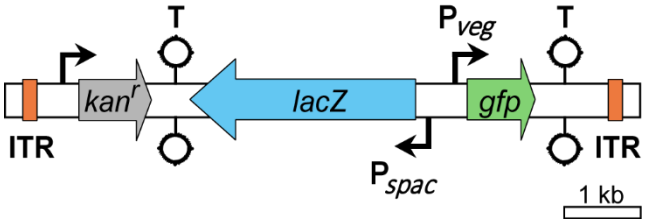


Figure 6.1: Scheme of the Mariner transposon enclosing the bidirectional expression cassette. The inducible P_{spac} promoter and the σ^A-dependent P_{veg} promoter were fused to *lacZ* and *gfp*, respectively. For the repression of the inducible P_{spac} promoter the *lacI* gene was integrated at the genomic *aprE* locus (not shown). The cassette is flanked by strong transcriptional terminators (T) to prevent read-through from and into adjacent genes. To select for the transposed expression cassette a kanamycin resistance served as antibiotic marker (*kan*^{r}). The inverse terminal repeats (ITR) were used for the recognition and the precise excision of the expression cassette. The 1 kilobase (kb) block indicates the size of the elements.

WebGeSTer database (Mitra *et al.*, 2011), which prevented the read-through from and into adjacent genes once integrated. The *himar1* transposase was responsible for the excision at the inverse terminal repeats (ITR) that flanked the expression cassette (Figure 6.1) and TA dinucleotides randomly located throughout the chromosome.

The transposition frequency of the transposon on pSS125 was compared with that of pMarB by comparing the ratio of colonies that were Ery^r/Kan^r with those that were Ery^s/Kan^r, since the Kan^r phenotype is associated with the transposon and Ery^r with the donor plasmid (Section 2.8) (carried out by Simon Syvertsson). This ratio was, on average, around 1.2:1 (n = 4) of pSS125 compared to 29.1:1 (n = 6) of pMarB. This discrepancy could be due to the presence of the expression cassette of pSS125 resulting in a reduced transposition efficiency. Following transposition, *B. subtilis* cells were grown on nutrient agar plates supplemented with X-Gal, a colorimetric β-galactosidase substrate that generates blue colonies, and the inducer of the P_{spac} promoter IPTG (1 mM IPTG). Blue colonies were chosen randomly and to

ensure that only a single copy of the transposon was present, the selected strains were back-crossed into *B. subtilis* strain W168 (*lacA*::*tet, aprE*::*lacI*, *cat*) (carried out by Simon Syvertsson). The isolated clones were cultured in a microtitre plate to screen for significant growth defects (Figure 6.2) before 14 transposon-containing clones were selected and their insertion site on the chromosome identified (Figure 6.3 & 6.4) via arbitrary primed PCR (carried out by Simon Syvertsson), following a modification of the protocol from Knobloch (Knobloch *et al.*, 2003). As a result, we were able to achieve a set of strains with transposed expression cassettes distributed throughout the *B. subtilis* chromosome (Figure 6.4 & Table 6.1) to elucidate the role of genome location and orientation on heterologous gene expression.

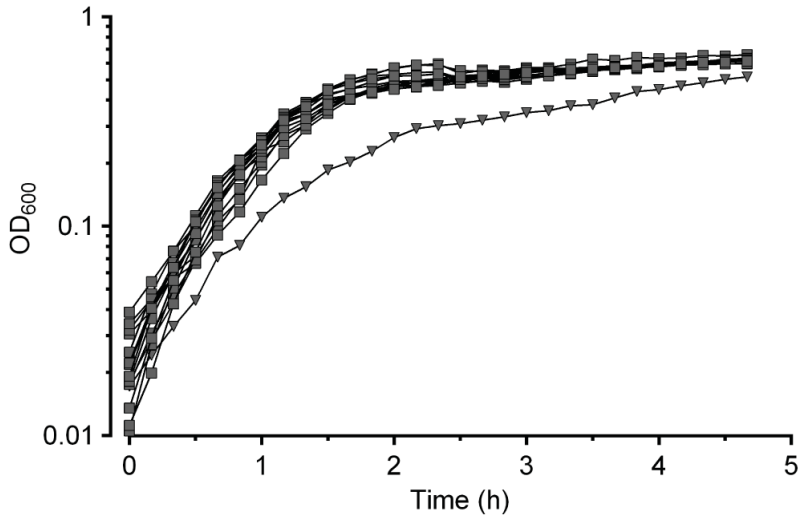


Figure 6.2: Microplate reader screen for the growth analysis of 15 transposon mutants. One strain deviated from the norm (triangles) and was excluded from further measurements. The other 14 strains were selected for further experiments (squares) and are listed in table 6.1. The mean values of three technical replicates are shown.

Locus	Genetic map	Functional assignment
rrnO-23S	1°	ribosomal RNA-23S, translation
ctc	5°	general stress protein
yckB	30.7°	similar to amino acid ABC transporter (binding protein)
ydgG	52°	similar to transcriptional regulator (MarR family)
speA	131.1	arginine decarboxylase, polyamine biosynthesis
bshB2	181.2	biosynthesis of bacillithiol
yojF	181.3	unknown
brxA	196.4	de-bacillithiolation of S-bacillithiolated OhrR and MetE
yqhS	216.9	similar to 3-dehydroquinate dehydratase
yqeT	224.1	similar to ribosomal protein L11 methyltransferase
recJ	241.3	DNA repair
yvmB	307.8	unknown
spsC	332.0	spore coat polysaccharide synthesis
nupG	342.1	purine uptake

Table 6.1: List of the integration sites of the transposed expression cassettes. The loci, genetic locations and the (predicted) functions of the disrupted genes are shown.

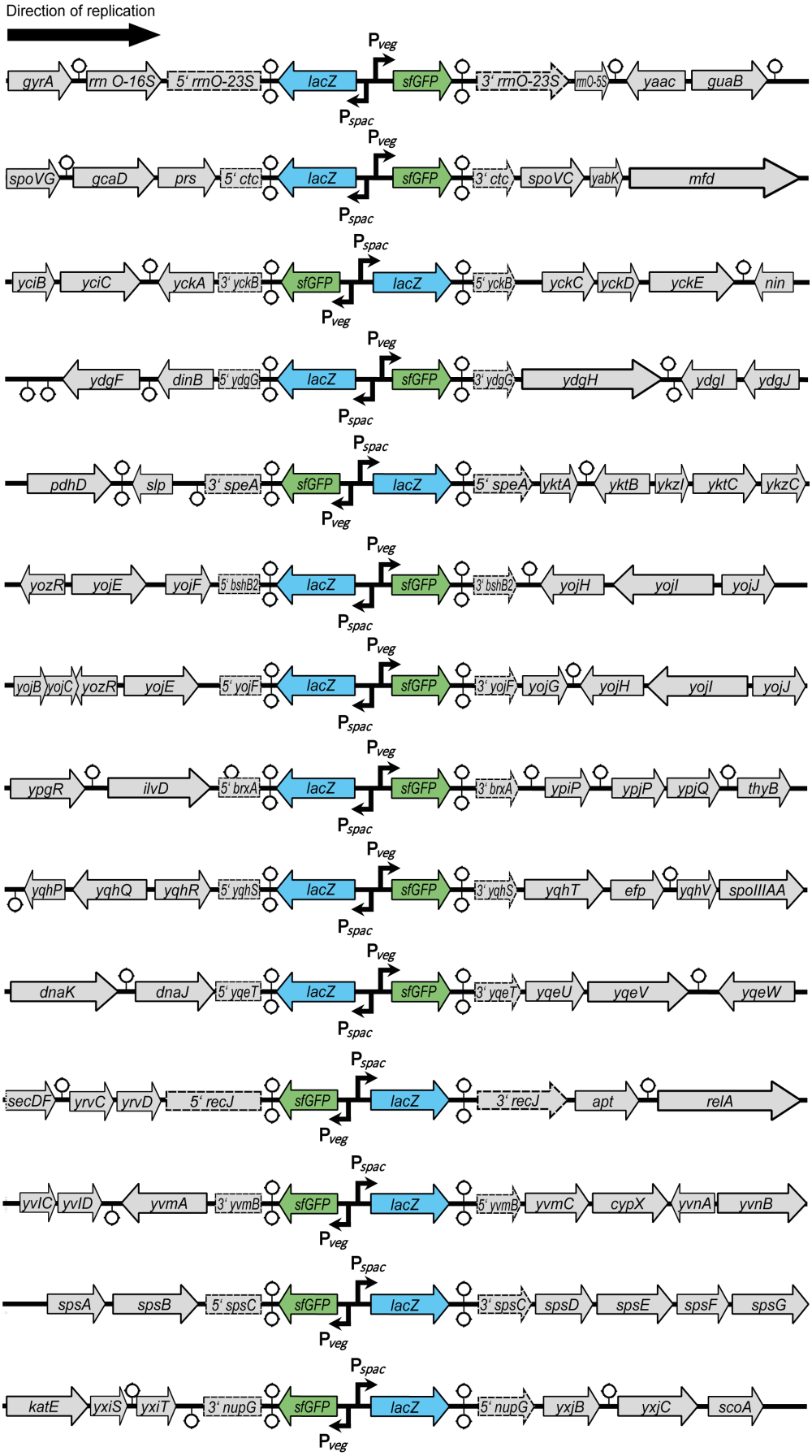


Figure 6.3: Schematic overview of the transposon integration loci in their chromosomal contexts. Genes were disrupted (dotted lines) by the integration of the transposed expression cassettes. The black arrow indicates the directionality of DNA replication.

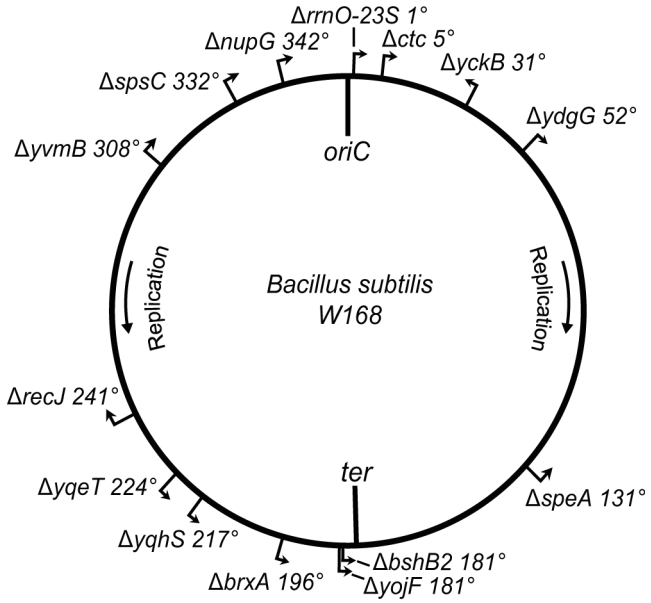


Figure 6.4: Chromosomal map of the transposon insertions. The locations of the disrupted genes are indicated in degrees (1° ≙ 11,708 base pairs) and the arrows display the orientation of the *gfp* reporter gene. The terminus region (*ter*) at approximately 178.7° and the origin of replication (*oriC*) are indicated including the direction of the DNA replication machinery.

6.2.2 Effect of genomic location

To analyse the influence of the genomic location of the transposed expression cassette on the heterologous gene expression, the activities of the β-galactosidase and GFP reporter genes were analysed by enzymatic assays and microscopy following growth in LB at 37°C (Sections 2.6 & 2.7). To validate the transcriptional isolation of the constructs *via* the bidirectional transcription terminators, the β-galactosidase activity of eight of the transposon strains were analysed in the absence of the P_{spac} inducer, IPTG. This analysis revealed very low β-galactosidase activities in each of the strains with maximal values of 6 Miller Units (MU). This confirmed that the terminations were functioning well by blocking transcriptional read-through from adjacent promoters (Figure 6.5).

In contrast, when the 14 strains were induced in exponential phase (OD_{600} ~0.5) with 1 mM IPTG, and their β-galactosidase activities and GFP fluorescence intensity measured, both markers showed a very similar result reflecting a genomic location dependent level of activity (Figure 6.6A). Another replicate of the experiment is shown in the appendix A (Figure A.8A). The reporter activities were highest (~1340 MU for β-galactosidase and ~43 AU for GFP) when the expression cassette was integrated at the site closest to the origin of replication. The difference in β-galactosidase activity between the most proximal and distal *oriC* locations was ~5.1-fold.

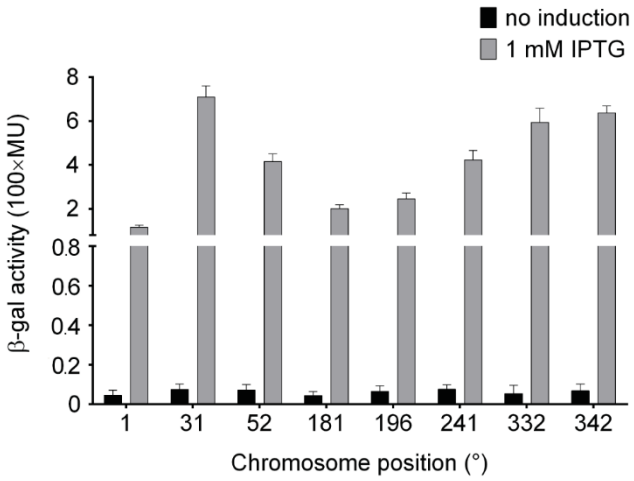


Figure 6.5: Analysis of the transcriptional read-through into the *lacZ* gene from upstream chromosomal genes via β-galactosidase activity assays. The β-galactosidase activities of eight locations were measured in stationary phase after 3.5 h of growth in LB medium at 37°C in the presence of 1 mM of IPTG (grey bars) and in the absence of the inducer (black bars). Mean values with standard deviations of three biological replicates are shown. The chromosomal locations of the expression cassettes are indicated in degrees.

In the case of GFP fluorescence intensity, a 3.6-fold difference was observed between the most proximal and distal *oriC* locations (Figure 6.6B). The β-galactosidase data fits well with earlier studies as the copy number of genes close to the origin of replication is roughly 5-fold higher in LB medium in exponential phase compared to the terminus region (Murray and Koh, 2014). These data reflect the crucial interdependency between the frequency of replication initiation and gene dosage. Following the trend lines, the closer the constructs were positioned to the terminus the lower were the measured β-galactosidase activities (~261 MU) and GFP intensities (~12 AU). To confirm that the location-dependent expression levels are due to different gene dosages, DNA copy numbers of the integration loci were determined by qPCR (carried out by Laura Bohorquez), which revealed a ~4.7-fold difference between the most proximal and distal *oriC* locations supporting previous results and showing that gene dosage clearly influences gene expression in rapidly growing cells (Sauer *et al.*, 2016).

Despite the clear trend lines reflecting location and reporter gene expression (Figure 6.6A), some very clear outliers were observed. The most significant outlier is located at ~1° (1° ≙ 11,708 base pairs) on the chromosome that represents an insertion into the *rrnO-23S* gene close to *oriC*. Despite the highest GFP signal that was measured during the assays, the transposed expression cassette displayed the lowest β-galactosidase activity. After sequencing this particular strain, no mutation or modifications were identified. This points to a situation where the expression of *lacZ*

gene on the lagging strand appears to be compromised by the highly transcribed *rrnO-16S* operon on the leading strand. Despite the presence of the bifunctional transcriptional terminators, this could be due to read-through into the *lacZ* gene, resulting in RNA polymerase clashes and/or the generation of antisense *lacZ* RNA and interference with translation (Figure 6.5). Any read-through might actually enhance GFP expression (Figure 6.6A). The integration into *spsC* gene (332°) resulted in lower than expected GFP expression, but without obvious explanation for its strong deviation from the trend line.

To further study the influence of genomic location on expression and to understand its relevance to enzyme production, the two industrial enzymes *xynA* and *amyM* were cloned separately in five of the same loci, three being close to *oriC* and two near *ter* (Figure 6.6D & Table 6.1). The measured activities of the native enzyme XynA (Section 2.10) during exponential phase showed an up to 1.6-fold higher activity when expressed from a location closer to *oriC* than *ter* (~178.7°), which reflects the trend observed with the transposon–encoded cassette expressing β-galactosidase and GFP. The lower activity ratio between *oriC* and *ter* might be a consequence of the additional processing of this enzyme, namely its translocation across the membrane and secretion into the growth medium. However, the measured activities of the heterologous enzyme AmyM, integrated in various chromosomal locations, showed no significant difference between the selected loci (Figure 6.6D). This could be due to the production of AmyM starting towards the end of exponential phase and during stationary phase, as previously discussed (Chapter 3 & 4), when gene dosage effects are likely to be minimal, while the current measurements were performed in exponential phase when AmyM production is low.

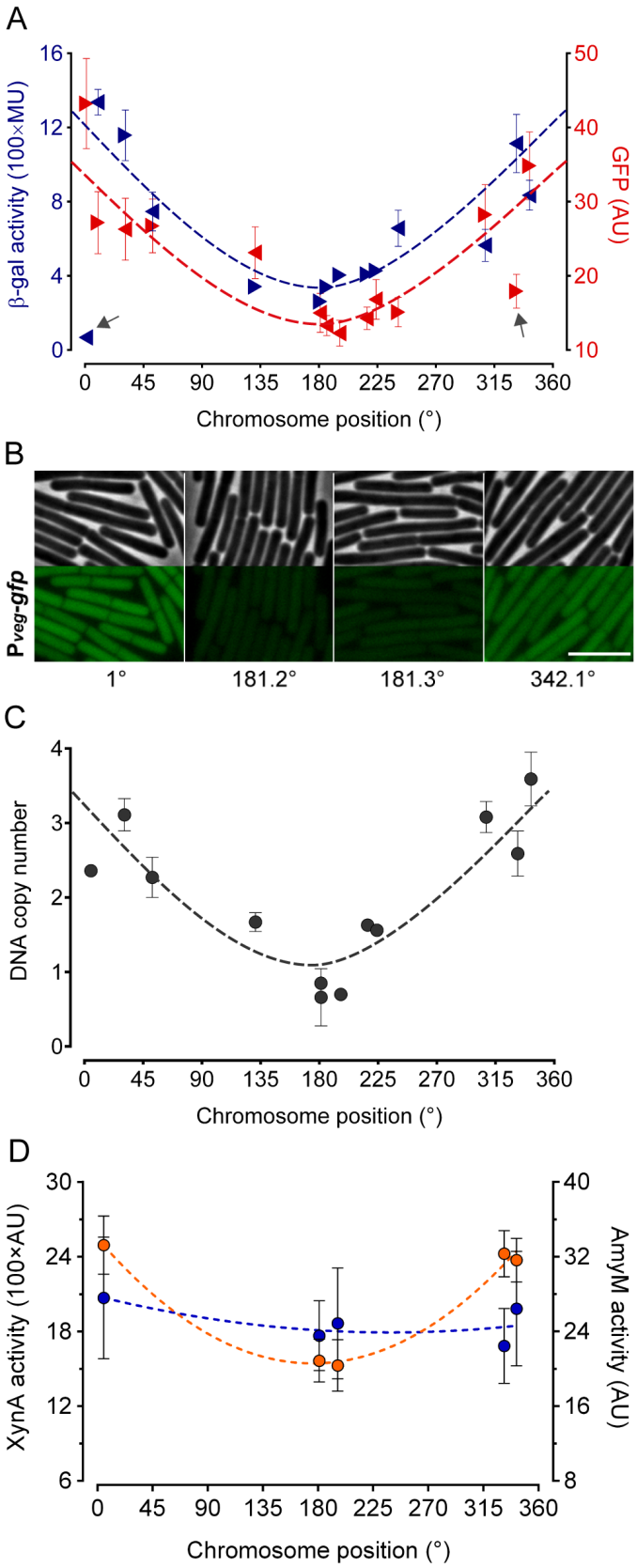


Figure 6.6: Effect of genomic location on gene expression levels. (A) β-galactosidase activities (blue) and GFP intensities (red) were examined in exponential phase after culturing 14 transposon mutants in LB medium at 37°C until the cells reached an OD_{600} ~0.5. The chromosomal locations of the expression cassettes are indicated in degrees. The transcriptional orientation of the reporter genes is illustrated by triangles. The goodness of fit for β-galactosidase amounts to R^2 = 0.86 (excluding locus at 1°) and GFP to R^2 = 0.70. Grey arrows indicate two outliers. Average values are displayed with standard deviations of two technical replicates for β-galactosidase and at least 100 cells for GFP. (B) Micrographs

of two expression cassettes close to *oriC* (1°, 342.1°) and two near *ter* (181.2°, 181.3°) demonstrate location dependent GFP expression levels based on P_{veg}-*gfp*. Scale bar is 5 µm. (C) DNA copy numbers of transposon locations measured by qPCR. Error bars indicate standard deviations of three biological replicates. Goodness of fit is R^2 = 0.78. (D) Enzyme activities of xylanase (orange) and amylase (blue) were measured in exponential phase. The five chromosomal locations of the enzymes are indicated in degrees. The goodness of fit for XynA amounts to R^2 = 0.99 and AmyM to R^2 = 0.51. Average values are shown with standard deviations of three biological replicates.

6.2.3 Influence of transcriptional direction

In *B. subtilis* there is a clear gene orientation bias, with ~75% of the genes on the leading strand, and therefore coinciding with the net direction of DNA replication (Kunst *et al.*, 1997, Barbe *et al.*, 2009). In contrast, in *E. coli* only 55% of the genes are located on the leading strand (Blattner *et al.*, 1997). To investigate the significance, if any, of this bias on *B. subtilis* gene expression, the transposed strains were analysed to determine the activities of their reporter genes. For this approach the monitored values of the expression cassettes were analysed and sorted according to their distances from *oriC* following the removal of the outliers at 1° and 332° (Figure 6.7). As the terminus is located at approximately 178.7° (Kunst *et al.*, 1997, Barbe *et al.*, 2009), genes in the region of *oriC* (358° to 360°) are physically the most distant from *ter.* To understand the significance of orientation, the β-galactosidase and GFP production values for genes either facing *oriC* or *ter* were plotted, and the resulting expression curves were compared using GraphPad Prism® 6. No significant differences could be observed for any of the matched slopes. This comparison was done with 12 transposon strains, representing only a limited number of insertion sites, nevertheless, the effect of location on the expression of both reporter genes is clearly illustrated (Figure 6.7). Another replicate is shown in the appendix A (Figure A.8C & D). Despite the bias that around 75% of genes are located on the leading strand (Kunst *et al.*, 1997, Barbe *et al.*, 2009), gene orientation appears to have no effect on expression levels. This provides evidence that the single collision events between individual genes and the replisome, some of which are avoided due to the activity of rho in removing the elongation complexes ahead of replisome progresses, has little or no impact on transcription (Washburn and Gottesman, 2011). In contrast, multiple clashes between the replisome and RNA polymerase could reduce the efficiency of replication, indicating that leading strand bias is for the benefit of replication rather than transcription (Sauer *et al.*, 2016).

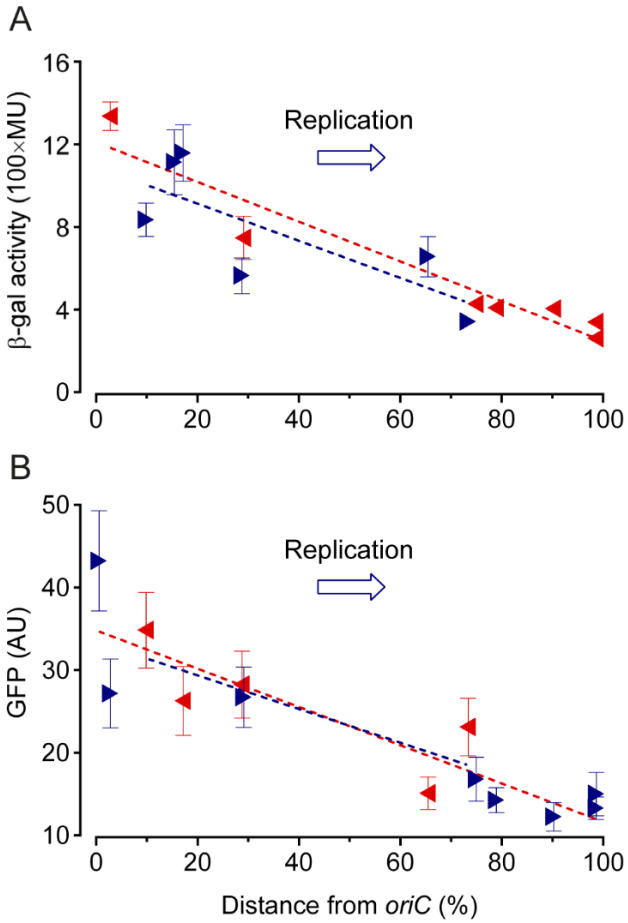


Figure 6.7: Influence of transcriptional orientation on gene expression. β-galactosidase activities (A) and GFP intensities (B) from 13 transposon mutants (from figure 6.6A) were sorted with respect to their distance from *oriC*. The outliers at 1° (β-galactosidase) and 332° (GFP) were removed. The goodness of fit for β-galactosidase amounts to R^2 = 0.60 (excluding locus at 1°) and GFP to R^2 = 0.51. Average values are shown with standard deviations of three biological replicates.

6.2.4 Impact of growth rate reduction

B. subtilis can be cultured in a variety of growth media including minimal or rich media (Section 2.2). However, the growth rate is noticeably reduced when growing in the former rather than the latter. To investigate the influence of growth rate and insertion site on gene expression, β-galactosidase and GFP expression levels were measured for two transposon cassettes inserted close to *oriC* (181° and 191°) and two close to *ter* (5° and 342°). The rationale behind these experiments is that the *ori*:*ter* ratio is likely to be reduced at slower growth rates and the corresponding reduction in the *ori*:*ter* ratio should be reflected in a reduced expression ratio. This hypothesis was confirmed in a previous study that showed an *ori*:*ter* ratio of 3:1 in minimal medium compared with 5:1 in rich media (Murray and Koh, 2014). However, surprisingly, the same study indicated that the *ori*:*ter* ratio is not effected when *B. subtilis* was grown in LB medium at different growth temperatures (Murray and Koh, 2014).

We therefore investigated the influence on gene expression of the above mention transposon cassettes in LB at 30°C and 37°C, when the respective mean generation times (mgts) were 50 min and 27 min, and in minimal competence medium at 37°C when the mgt was 45 min. Compared with the mgt in LB at 37°C, the other two conditions reduced the mgt to approximately half. Following growth of the strains with transposon cassettes at locations close to either *oriC* or *ter* in rich medium at two temperatures and minimal medium at 37°C, the β-galactosidase and GFP expression levels were determined. The averaged expression levels of the two transposon cassettes close to *oriC* were divided by the averaged expression levels of the two transposon cassettes close to *ter* (Figure 6.8). In the case of *lacZ* expression, a significant reduction in the *ori:ter* ratio (3.5-2.5) was observed between growth at 37°C in LB medium and minimal competence medium (Figure 6.8). In contrast, no clearly significant differences were observed in the GFP expression levels under the various conditions for unknown reasons. Additionally, previous work indicated that the *ori*:*ter* ratio was similar for cells growing in LB medium at 37°C and 30°C (Murray and Koh, 2014), which was confirmed by our results showing no significant differences in the expression levels of β-galactosidase and GFP (Figure 6.8).

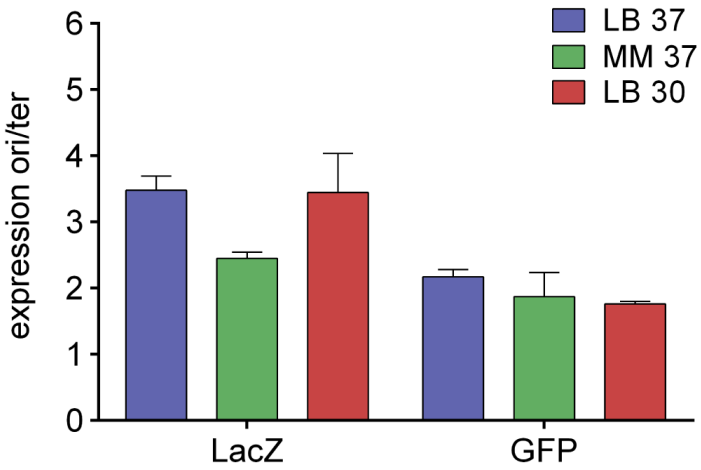


Figure 6.8: Impact of growth conditions on gene expression. The *ori*/*ter* ratio of β-galactosidase activities and GFP intensities were determined by dividing the average expression levels of two integration sites near *oriC* by that of two loci close to *ter* after cells were cultured under different conditions: LB medium (blue) and minimal medium (green) at 37°C, and in LB medium at 30°C (red). Average values are shown with standard deviations of three biological replicates.

6.2.5 Change in gene expression during stationary phase

The stationary phase is coupled with dramatic physiological and developmental changes that result in the reduction of DNA replication as a consequence of the lower growth rate due to the limitation of nutrients. Thus, the previously demonstrated influence of gene location on gene expression would be expected to hold during stationary phase. Unexpectedly, during stationary phase β-galactosidase and GFP showed expression profiles to that found during the exponential growth (Figure 6.9A & B). These contradicting results indicate that the expression levels of the reporter genes were still influenced by their genomic location despite the lower replication initiation frequency. However, this gene dosage effect could be explained by the relative stabilities of the β-galactosidase and GFP proteins, reflecting their actual synthesis during exponential phase (Figure 6.6). In an attempt to clarify this issue, β-galactosidase expression was induced with IPTG (1 mM) two hours after the transition from exponential to stationary phase (Figure 6.9A & C). In this case no significant differences were observed between β-galactosidase activity levels for transposon cassettes located regions on the chromosome and, as expected, the activity level was lower compared to that observed in exponential phase (Figure 6.9C). To confirm the observed results of β-galactosidase and to examine whether this is a result of reduced gene dosage, we measured the DNA copy numbers of the integration loci by qPCR (carried out by Laura Bohorquez), which revealed a similar trend supporting the fact of decreased gene dosage during stationary phase (Figure 6.9D). The data confirm that gene expression is dependent on the relationship between location, the frequency of replication initiation and the resulting impact on gene dosage.

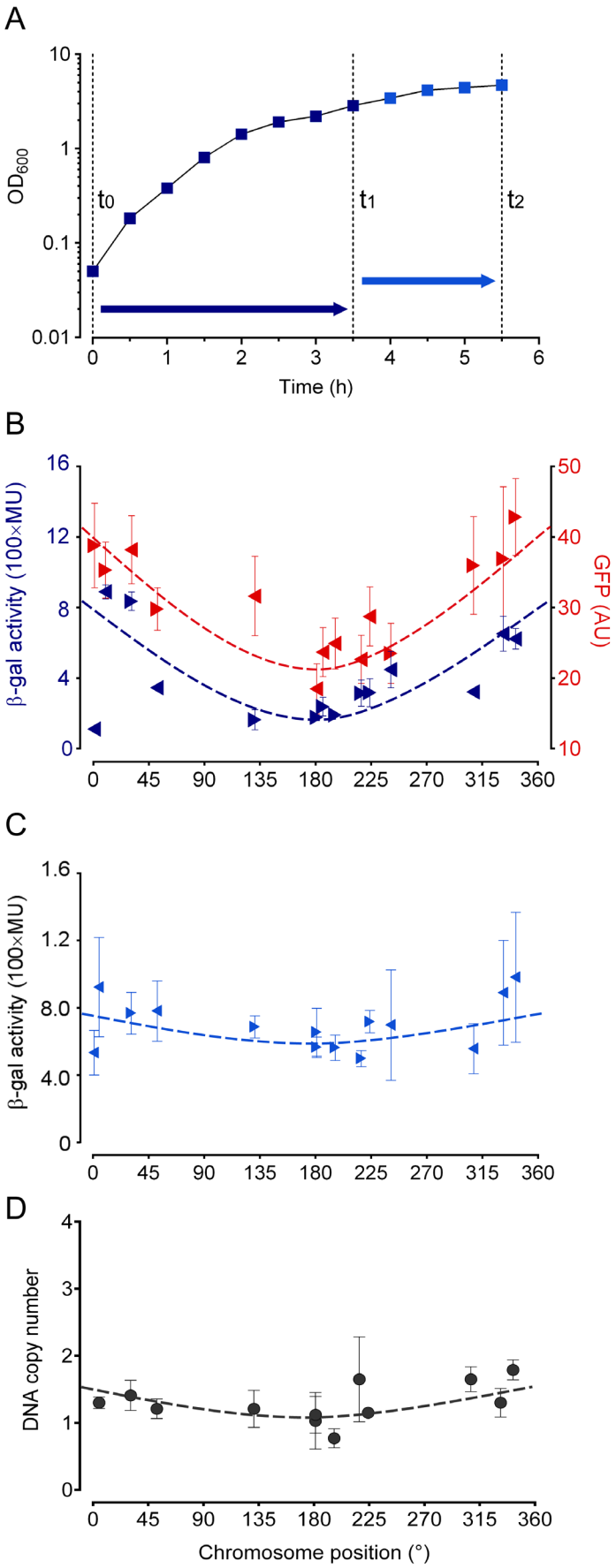


Figure 6.9: Effect of stationary growth on gene dosage. (A) Transposon mutants were cultured in LB medium at 37°C with the induction of *lacZ* with 1 mM IPTG from the start of the experiment (t_0) with the collection of cells after 3.5 hours or induced in stationary phase (t_1) and harvested after 2 hours (t_2). (B) β-galactosidase activities (blue) and GFP fluorescence intensities (red) were measured in stationary

phase after growing 14 transposon strains in LB medium at 37°C until the cultures reached an OD_{600} ~3.0. The chromosomal locations of the reporter cassettes are shown in degrees. The reporter gene orientation is indicated by triangles. The goodness of fit for β-galactosidase amounts to R^2 = 0.77 (excluding locus at 1°) and GFP to R^2 = 0.80. Mean values are displayed with standard deviations of two technical replicates for β-galactosidase and at least 100 cells for GFP. (C) Activities of β-galactosidase were assayed after the IPTG induction in stationary phase for 2 hours (t_1-t_2). Average values with standard deviations of three biological replicates are presented. The transcriptional orientation of *lacZ* is displayed by triangles. The goodness of fit is R^2 = 0.34. (D) DNA copy numbers of transposon locations measured by qPCR. Error bars indicate standard deviations of three biological replicates. Goodness of fit is R^2 = 0.42.

6.3 Conclusion

To determine the influence of genomic location and orientation on the heterologous gene expression in *B. subtilis*, an unbiased chromosome integration approach was adopted involving a transposon-based expression cassette that integrates randomly. The expression cassette encoded two divergently expressed reporters, namely β-galactosidase and GFP that were used to monitor expression at different locations and orientations. The data clearly demonstrate that gene dosage influences gene expression levels under conditions when the replication initiation frequency is high. The expression of the reporter genes strongly depends on their genomic location and this was reflected in their proximity to the origin of replication: the activities of the reporter genes were up to 5.1-fold (β-galactosidase) and 3.6-fold (GFP) higher when located close to *oriC* than *ter*. This outcome reflects similar findings to that of a previous study in *E. coli* (Block *et al.*, 2012), but differs from the results of another study in *E. coli* identifying chromosome organisation as the influencing factor for gene expression levels (Bryant *et al.*, 2015). A second important observation was that the orientation of the β-galactosidase and GFP genes with respect to lagging and leading strand did not have a significant impact on expression. While initially unexpected given the strong orientation bias observed in *B. subtilis,* the data indicate that transcription is not effected by the passage of replisome machinery, and therefore the orientation bias may be for the benefit of replication rather than transcription.

In view of the increasing adaption to other bacterial species for commercial exploitation, the strategy of utilising the Mariner transposon to identify genome integration sites that maximise gene expression is of great potential industrial importance. The transposon could even be used to deliver heterologous genes to the genome of organisms lacking suitable recombination systems. The economic importance is also emphasised in the case of the native *xynA* gene, encoding a commercially important xylanase. Xylanase activity expressed during logarithmic

growth was up to 1.6-fold higher when *xynA* was located close to *oriC* compared to that of *ter*. Taken together, the experiments provide important knowledge on an industrial bacterium for which even small increases in productivity are of economical importance.

Chapter 7

Final Discussion

The main aim of this thesis was to analyse whether heterogeneous or bistable regulatory processes in *B. subtilis* decrease the overall protein production of the industrial enzyme XynA or AmyM.

We began by studying the production of both enzymes during growth, revealing very different production profiles. In case of XynA, the production was initiated during exponential phase reaching a plateau in chemically defined and LB medium during stationary phase. In rich medium the production subsequently decreased possibly due to increased rate of proteolytic degradation. In contrast, AmyM production was initiated when exponential growth ceased and cells entered the stationary phase, suggesting different regulatory mechanisms that prevent its production during exponential growth. To determine the cause for these differences, transcriptional expression levels of both enzymes were measured to examine whether different mRNA levels are responsible for this disparity. This appeared not to be the case, and the results suggested that subsequent secretion and quality control processes are involved in the delayed AmyM production.

After confirming the presence of XynA or AmyM, different physiological and developmental pathways were investigated at the population and single-cell level by creating transcriptional reporter fusions with GFP to determine their degree of heterogeneity during growth. This revealed no differences between XynA producing and non-producing strains at both population and single-cell levels. In contrast, strains producing the heterologous AmyM showed a heterogeneous induction of secretion stress during exponential phase. The results suggest that the strong artificial expression of the native XynA only resulted in negligible amounts of secretion stress, presumably due to its rapid folding and secretion capacity. Since AmyM is originated from *G. stearothermophilus*, comprising a signal sequence that is foreign to the secretion system of *B. subtilis* and being folded in an unfamiliar

environment, the upregulated secretion stress reflects the difficulties for the successful secretion of AmyM.

To determine the relationship of secretion stress and *amyM* expression at the single-cell level, a transcriptional mCherry fusion reflecting the AmyM transcription was constructed and analysed using flow cytometry. This analysis revealed the formation of two populations with one subpopulation showing low secretion stress and *amyM* expression levels, and a population showing high production and high secretion stress. Further experiments confirmed a bistable induction of secretion stress but the removal of either *htrA* or *htrB* abolished the bistable distribution. Presumably, this was caused by compensating effects of cells due to lower proteolytic activity by which either protease was consequently upregulated to facilitate the quality control of proteins including AmyM. By deleting both proteases simultaneously, cells showed highly elevated secretion stress and lower AmyM production levels, and impaired growth, indicating the importance of the presence of at least one quality control protease.

Since *amyM* was expressed from a plasmid based system, displaying a bimodal secretion stress response, an integrated version was constructed. This abolished the bistable secretion stress, suggesting that rather cell division defects in cells carrying the plasmid, indicated by microscopy, are accountable for the observed bistability. However, integration of *amyM* into the chromosome appeared to accumulate mutations that influenced the overall AmyM production, possibly due to extreme cell stress induced by high expression levels of *amyM* and *mCherry*. These results indicate that cell division defects during the plasmid based expression lead to the reduction of AmyM production, thus resulting in a decreased process efficiency in large-scale applications. Consequently, to reduce industrial production losses, genome integrated constructs are preferred over plasmid based expression systems, which also require the continuous addition of applicable antibiotics.

To obtain synthetic modules with high and homogeneous expression levels we modified the sequence of the native *veg* module by varying its UP element, promoter region and 5’ UTR. Using next-generation DNA synthesis, we generated ~12,000 synthetic modules for the expression of GFP. The obtained expression library was screened using flow cytometry and microbioreactor experiments resulting in the improvement of the native *veg* expression unit by ~13-fold and simultaneous exclusion of modules with heterogeneous activities. The library demonstrated a dynamic range of GFP expression levels between $>10^0$ AU and $>10^5$ AU, providing a

useful range for synthetic modules with desirable expression levels. However, the activities appeared to be associated with the expression of the GFP reporter, since exchanging GFP with XynA led to no correlation between GFP and XynA levels. This shows that gene specific effects of the 5' UTR and the 5' end of the gene should be taken into account. For future industrial applications this can be circumvented by directly selecting the optimal expression unit for the specific enzyme, by constructing translational fusions of a gene of choice to a fluorescent reporter.

Finally, gene dosage effects are rarely addressed in *B. subtilis*, but for engineering novel pathways precise gene expression levels are required to fulfil computational predictions. In this thesis, genomic location and gene orientation were investigated by using an unbiased approach for the integration of an expression cassette. This revealed a strong gene dosage effect in rapidly growing cells reflected by activity differences when genes were located close to *oriC* compared to *ter* of ~5.1-fold (β-galactosidase), ~3.6-fold (GFP) and 1.6-fold (XynA). The different results with XynA compared to β-galactosidase and GFP may be explained by the nature of the protein itself being a secretory protein that undergoes quality control processes. In contrast, gene orientation with respect to DNA leading and lagging strand synthesis does not affect gene expression levels, indicating that the replisome does not influence transcription during replication.

Appendix A

This section contains additional and alternative versions of figures and sequences. The published article of chapter 6 is attached in the end.

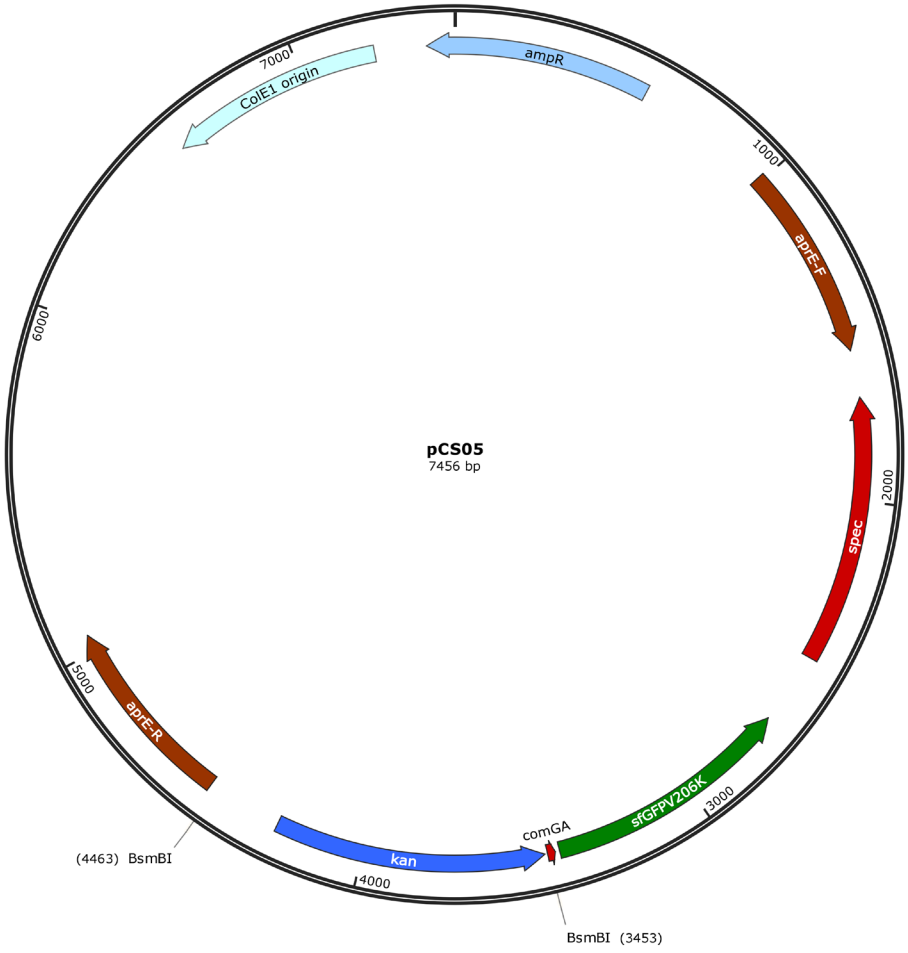


Figure A.1: pCS05 – StarGate® backbone to create transcriptional GFP fusions (Section 2.4.7).

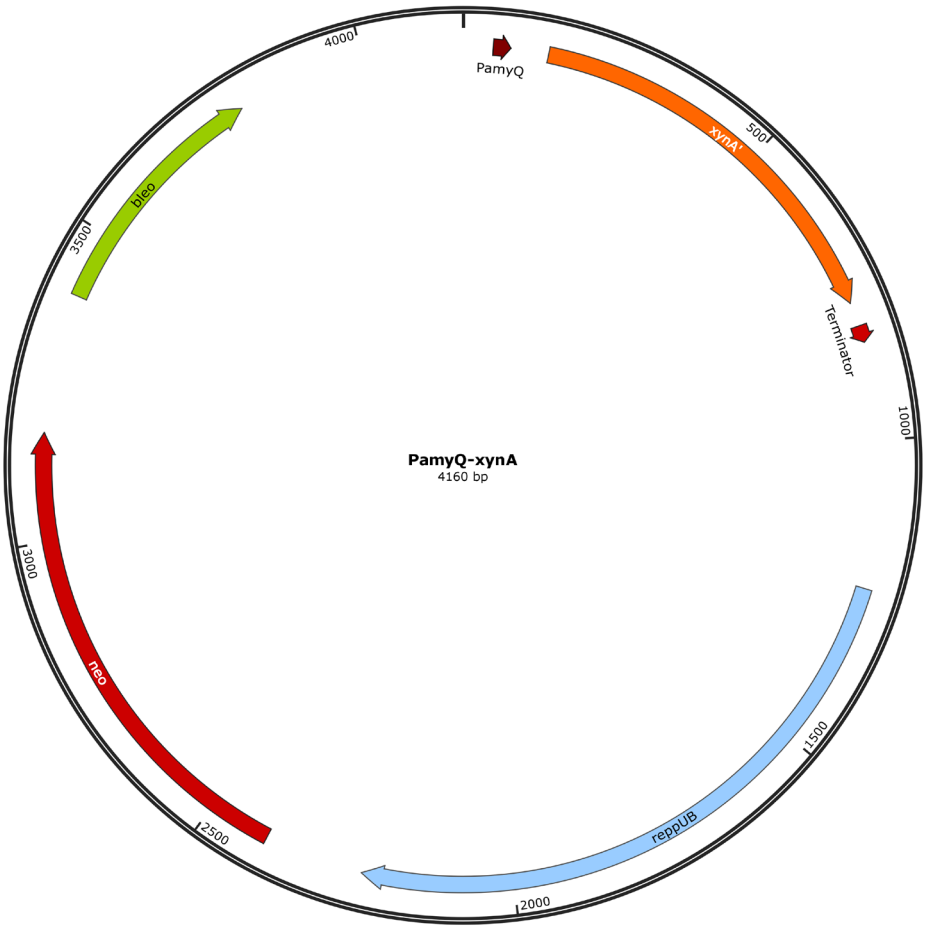


Figure A.2: pCS58 – *xynA* expression vector (Section 2.4.8).

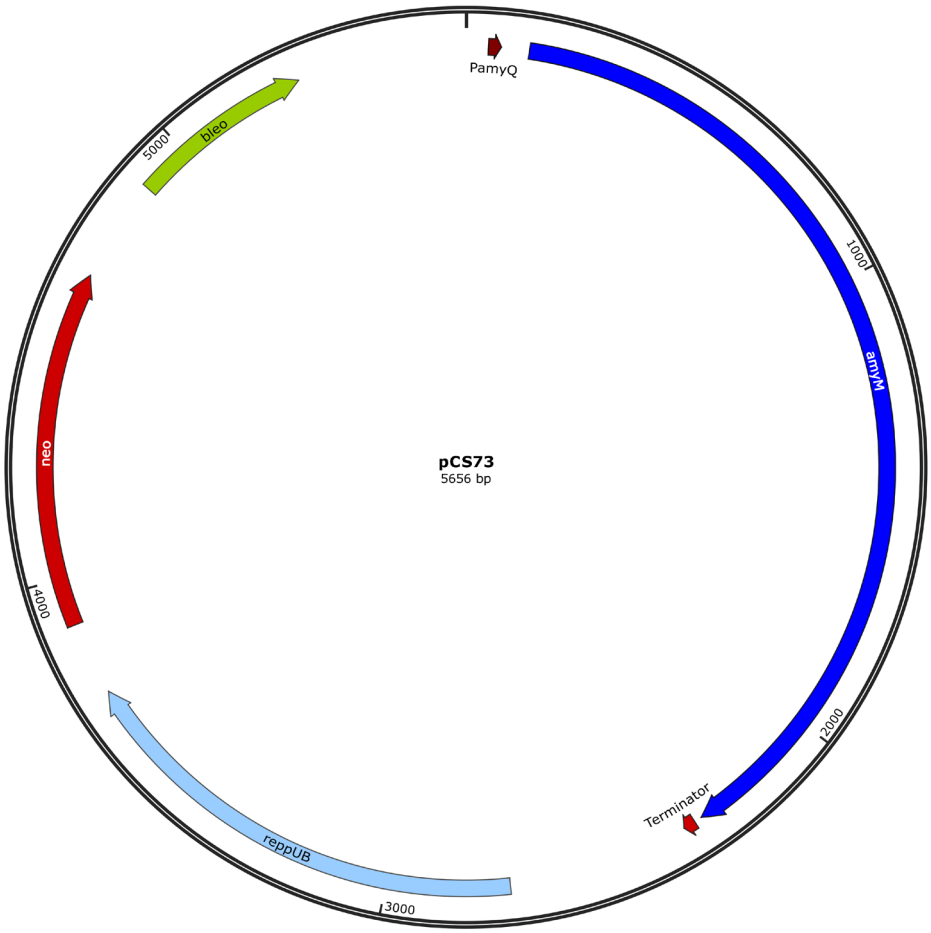


Figure A.3: pCS73 – *amyM* expression vector (Section 2.4.8).

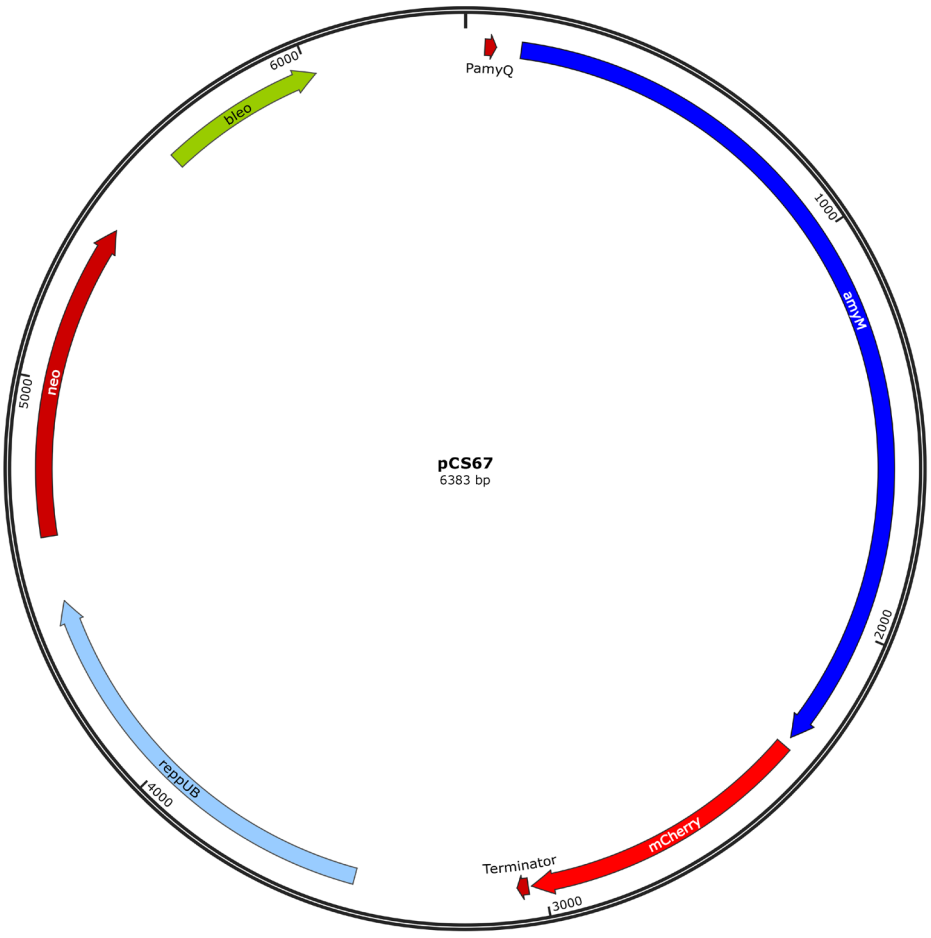


Figure A.4: pCS67 – *amyM-mCherry* expression vector (Section 2.4.8).

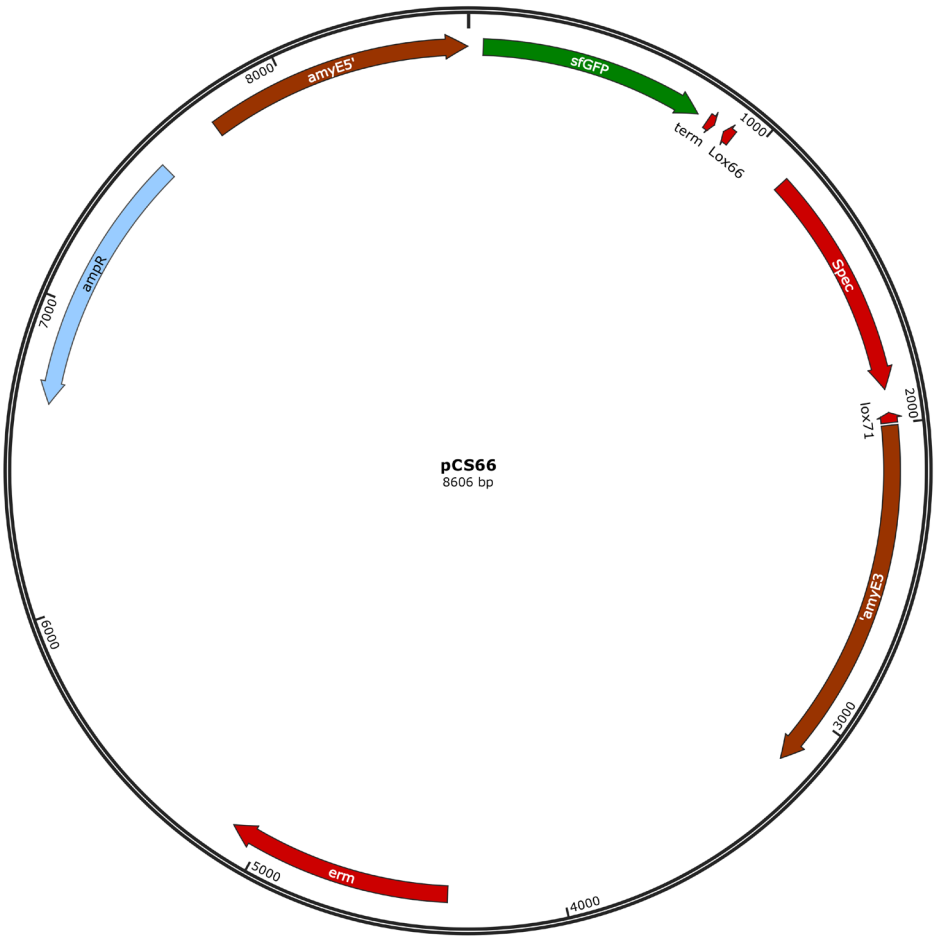


Figure A.5: pCS66 – StarGate® backbone for synthetic expression module library (Section 2.4.8).

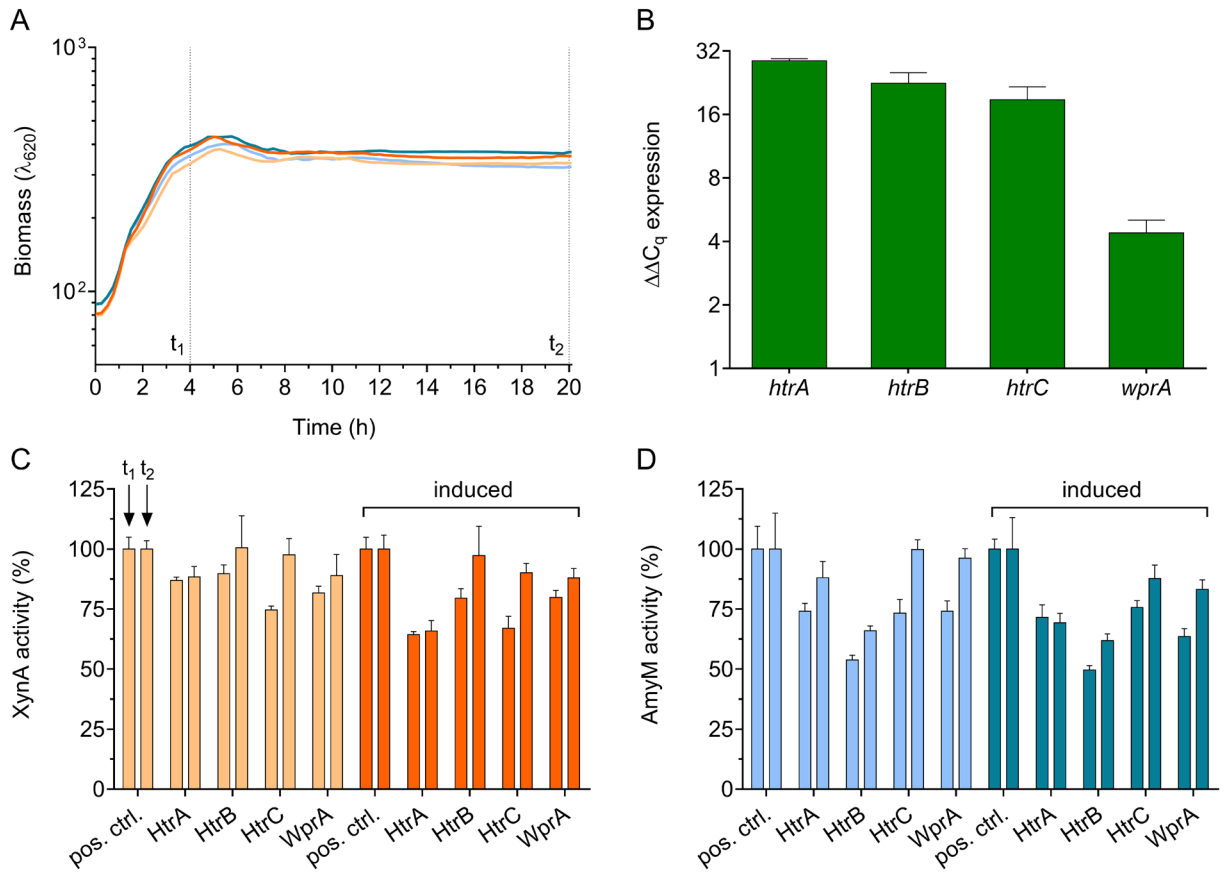


Figure A.6: An alternative version of figure 4.3 is shown. The activities of the overexpression strains of the positive controls and strains with and without xylose induction are shown (light – without xylose; dark – with xylose).

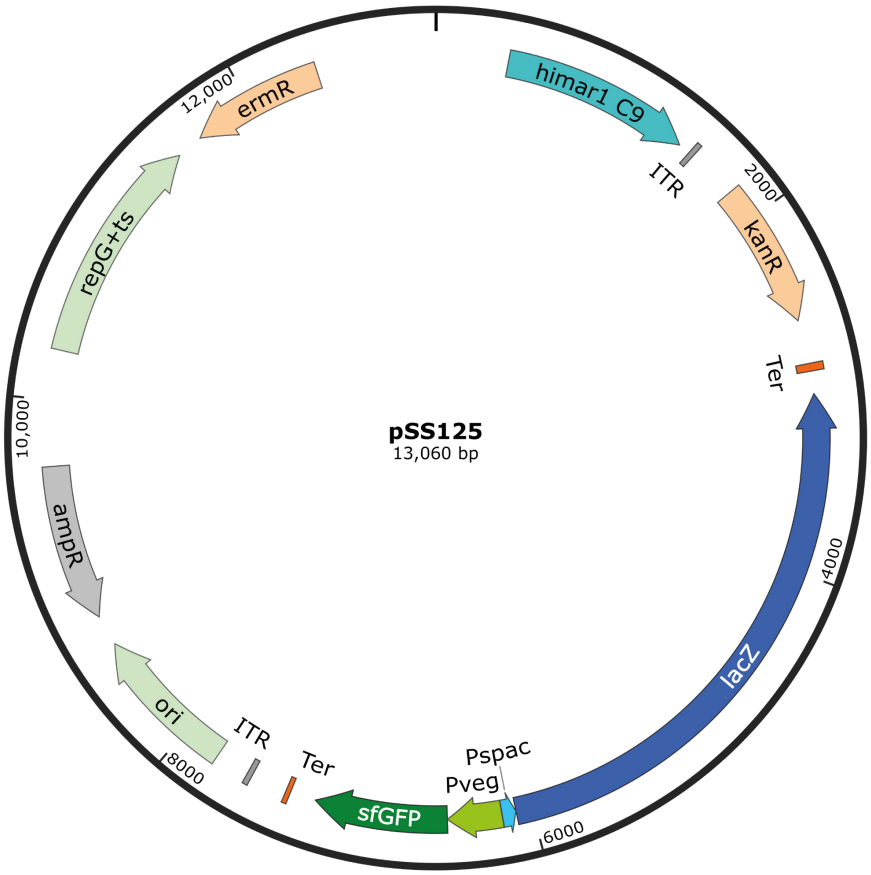


Figure A.7: pSS125 vector for random integration of expression cassette (Section 6.2.1). Vector map of plasmid pSS125 containing the *himar*1 transposase gene (cyan), kanamycin resistance gene (light orange), ITRs (grey), transcriptional terminators (orange), P_{veg} driving the expression of *sfGFP* (green) and P_{spac} driving the expression of *lacZ* (blue) (Sauer *et al.*, 2016).

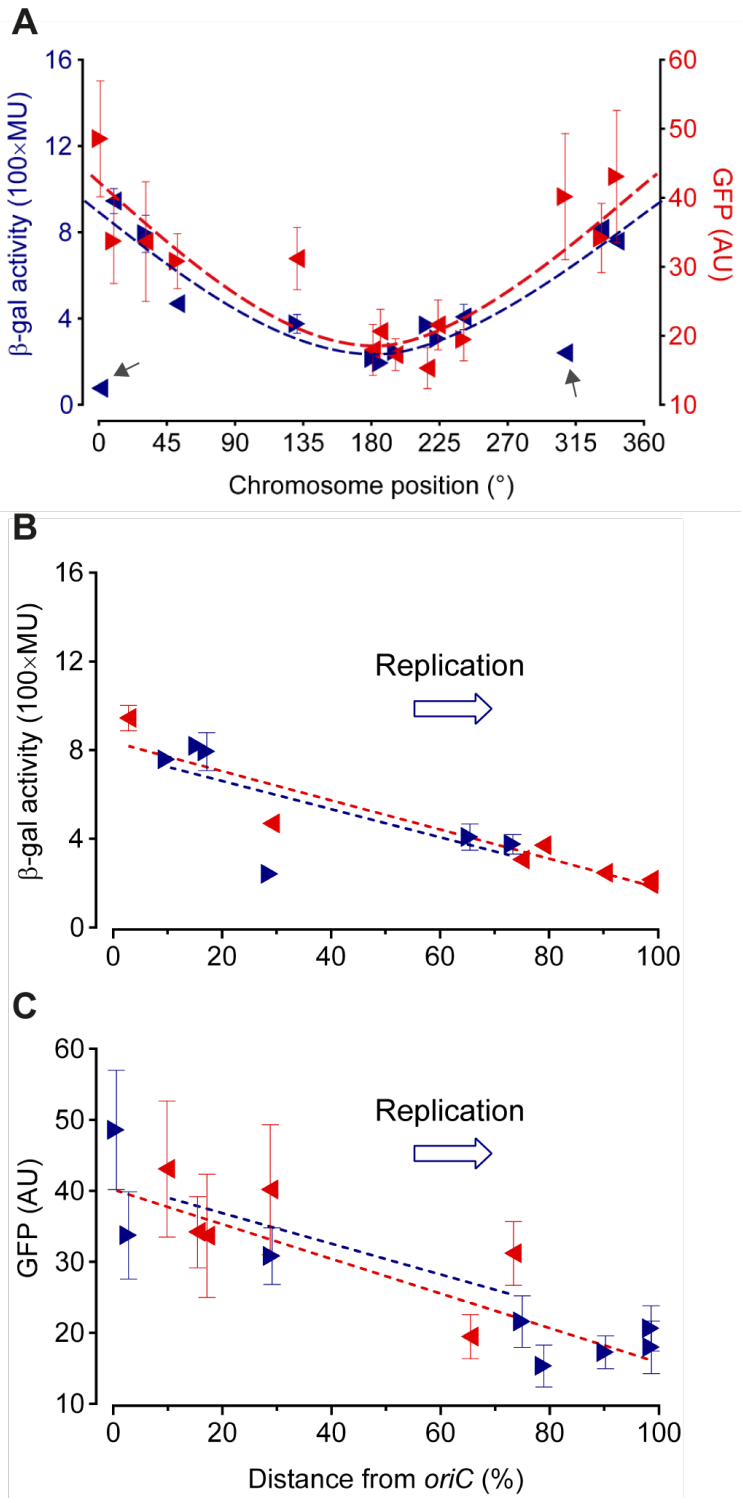


Figure A.8: Effect of chromosomal location and transcription direction on gene expression (Section 6.2.2 & 6.2.3). Replicate that shows effect of chromosomal location and transcription direction on gene expression. (A) β-galactosidase activities (blue) and GFP fluorescence (red) are plotted against genomic location (degrees). Goodness of fit for LacZ with R^2 = 0.79 (excluding locus at 1°) and GFP with R^2 = 0.79. β-galactosidase activities (B) and GFP fluorescence (C) are plotted against the distance of the integration loci from *oriC*. Reporter genes that are transcribed in the direction of DNA replication are shown in blue and genes that are transcribed against replication direction are shown in red. Error bars indicate the standard deviation of two technical replicates in case of β-galactosidase and at least 100 cells for GFP. Goodness of fit for LacZ with R^2 = 0.48 (blue), R^2 = 0.87 (red) and for GFP with R^2 = 0.81 (blue), R^2 = 0.52 (red) (Sauer *et al.*, 2016).

Table A.1 – Sequences and intensities of 135 clones

ID	**Sequence** (5' to 3')	Length	Mean Intensity	GFP/mCherry ratio
1	TCTAGGGAATTTGCCCGCGCATATTGACATTAAAATAGACTTTGCATATAATCCCTGTAGAAGCGAGTAATACAATCTTGTTAAGGAGGGATTACTA	97	8105	1.267339
2	GGCCACGCGGGTAGTGCCGCCTCATTGACAACTGAAGATATATTCTATATAATCTTCTAGAAAAGGAGGATGTCTAA	77	8193	1.057705
3	TTGGTATAGTCCGTACAACCACGGTTGACATTAAAATAGACTTTGCATATAATAAAGTCGTCTAAGGAGGTAGATATC	78	7570	0.6432548
4	ATAGCTATTATGTTCTTGAAAAATTGACAACTGAAGATATATTCTATATAATAAATGTAGTGAAGGAGGATGTCTAA	77	10241	3.010352
5	TAAATTATAATATTTATTAAATTATGACATTAAAATAGCTGCATATAATTGAGAGTCCGATTATCAAAAGGAGCTAATGTA	81	7609	0.09453823
6	ACGTTGATATAATTTAATTTTATTTGACATAATAGACAACTTATGTTATAATATTGATCAGCAAGGAGGGATTACTA	77	8131	0.5098191
7	TATTTTAATTTATATATAAAAATTGACATAATAGACAACTTATGTTATAATAAGACTTCGAAGGAGGTAGATATC	75	11280	1.658975
8	TCATTTGATGTACATAGCATAATATTGACATAATAGACAACTTATGTTATAATGGACTTACAAAGGAGGGATTACTA	77	12073	2.663871
9	TAAATTATAATATTTATTAAATTATTGACATAATAGACAACTTATGTTATAA**C**GTTCTAGAAAAGGAGGGATTACTA	77	8328	0.5010794
10	TCATTTGATGTCATAGCATAATATTGACAAAAATGGGCTCGTGTTGTATAATGGACTTTACAAAGGAGGATGTCTAA	77	8789	2.034797
11	TAAATTATAATATTTATTAAATTATTGACAAAAATGGGCTCGTGTTGTATAATGGACTTTACAAAGGAGGTAGATATC	78	9119	6.270516
12	ACGTTGATATAATTTAAATTTTATTTGACACACTTGAGAAATATTTATATAATCCCTGTAGAAGCGAGTATACAATCTTGTTAAGGAGGTGTACTAA	97	8740	1.131306
13	TTATATATAAATATAATTATATTTGACAAAAATGGGCTCGTGTTGTATAATAAGATTTCGAAGGAGGTGTACTAA	75	9549	4.984845
14	GTCAGGGCGGCTACGCGCTGCCCATTGACAACACTAGTAATTTTAGTATAATATTGATCAGCAAGGAGGTAGATATC	77	8533	0.8531193
15	ATAGCCTATTATGTTCTTGAAAAATTGACATAATAGACAACTTATGTTATAATAAATGTAGTGAAGGAGGGATTACTA	78	13819	3.180354
16	ACTGTTGGCGACCTTCGATGAAACTTGACATAATAGACAACTTATGTTATAATCGTTCTAAAAAAGGAGGATGCA	75	8040	0.7339268
17	TTAATATATATTAAATTAATATATTTGACAACTGAAGATATATTCTATATAATAAGATCTTCGAAGGAGGTGTACTAA	78	8380	1.046157
18	GCTGCGCGCACCGAGGCTTCAGGTTGACAAAAATGGGCTCGTGTTGTATAATCCTAACTAGTGTATGTAACGAAGGAGGTAATGTA	86	11363	3.581688
19	GAGTAATAAGTATCTTCATATCTATTGACATAATAGACAACTTATGTTATAATAAATGTAGTGAAGGAGGTGTACTA	77	7794	3.008833
20	ACTGTTGGCGACCTTCGATGAAACTTGACAACTGAAGATATATTCTATATAATAAATGTAGTGAAGGAGGCTAATGTA	78	7878	0.577912
21	TTAATATATATTAAATTAATATATTTGACATAATAGACAACTTATGTTATAATCGTTCTAGAAAAGGAGGATTCTAA	77	7517	0.1124195
22	ACGTTGATATATTTAAATTTTATTTGACACACTTGAGAAATATTTATATAATCCCTGTAGAAGCGAGTAATACAATCTTTTAAGGAGGATGCA	93	8172	2.023115

23	GTCAGGGCGGCTACGCGCTGCCCATTGACATTAAAATAGACTTTGCATATAATATAGTGAGCGTATTCATCCTATCAGAAGC TAAGGAGGATTACTA	97	8484	0.7484356
24	TCATTTGATGTACATAGCATAATATTGACAACAACTAGTAATTTTAGTATAATTGAGTTAGGTCCTCTCAAAAAAGGAGGTAGA TATC	88	8297	1.294145
25	ACGTTGATATAATTTAAATTTTATTGACATAATAGACAACTTATGTTATAATCCTGTAGAAGCGAGTAATACAATCTGGTTAAG GAGGATGTCTAA	96	12996	6.342831
26	ATAGCCTATTATGTTCTTGAAAAATTGACATAATAGACAACTTATGTTATAATACTTGACGTCTGAGCAATAAAGGAGGCTAAT GTA	87	9424	0.58835
27	TCATTTGATGTACATAGCTAATATTGACATAATAGACAACTTATGTTATAATACTGACGTCTGAGCAATAAAGGAGGCTAATGT A	85	8105	1.870365
28	GTCAGGGCGGCTACGCGCTGCCCATTGACAACTGAAGATATATTCTATATAATAAATGTAGTGAGGAGGGATTACTA	77	7756	2.42423
29	GCTGCGCGCACCGAGGCCTTCAGGTTGACAACAACTAGTAATTTTAGTATAATTACAATCGTTTCAGTTCTTCGGGCGAAAA AAGGGGGTTACTA	95	7808	-0.05
30	TTGGTATAGTCCGTACAACCACGGTTGACACACTTGAGAAATATTTATATAATCCTAACTAGTGTATGTAACGAAGGAGGGAT TACTA	88	8531	0.2616885
31	GGGGAGGCTGCATGCCCCTGACCCTTGACACACTTGAGAAATATTTATATAATATTGATCAGCAAGGAGGGATTACTA	78	7553	0.03831269
32	GGCCACGCGGGTAGTGCCGCCTCATTGACAACAACTAGTAATTTTAGTATAATAAGTCTTCGAAGGAGGGTTACTA	76	7714	0.04369643
33	TTATATATTAAAATATAATTATATTTGACATAAAATAGACTTTGCATATAATGGACTTACAAAGGAGGACTAA	73	7892	0.02632753
34	TTAATATATATTAAATTAATATATTTGACATAATAGACAACTTATGTTATAATAAATGTAGTGAAGGAGGGATTACTA	78	8653	7.441073
35	AATATTTTATTAATTTATAAAATATTGACAACAACTAGTAATTTTAGTATAATCGTTCTAGAAAAGGAGGGATTACA	77	13171	5.441929
36	TCTAGGGAATTTGCCCGCAGCATATTGACAACTGAAGATATATTCTATATAATGGACTTTACAAAGGAGGATGTCTAA	78	8383	4.104962
37	TTGGTATAGTCCGTACAACCACGGTTGACAACTGAAGATATATTCTATATAATCCTAACTAGTGTATGTAACGAAGGAGGATG CA	85	8181	0.2807372
38	TTATATATTAAAATATAATTATATTTGACACACTTGAGAAATATTTATATAATAAGATCTTCGAAGGAGGTAGATATC	78	8636	2.870158
39	GACTCATTAATTAACAGATTTGTATTGACACACTTGAGAAATATTATATAATGGACTTTACAAAGGAGGATGCA	74	7908	0.04491133
40	TTAATATTAAAATATAATTATATTTGACACACTTGAGAAATATTTATATAATAAATGTAGTGAAGGAGGGATTACTA	77	12392	2.535275
41	AATATTTTATTAATTTATAAAATATTGACATAATAGACAACTTATGTTATAATGGACTTTACAAAGGAGGATGTCTAA	78	9214	1.740436
42	TAATTATAATATTTATTAAATTTGACAAAAATGGGCTCGTGTTGTATAATAATGTAGTGAAGGAGGTGTACTA	73	11111	2.301767
43	TTTCACCGTCTGATAACGGCGGAATTGACACACTTGAGAAATATTTATATAATTTGAGAGCTCCGATTATCAAAAGGAGGTGT ACTAA	88	8433	0.03757876

44	GACTCATTAATTAACAGATTTGTATTGACATAATAGACAACTTATGTTATAATGGACTTTACAAAGGAGGATGTCTAA	78	8034	0.8425905
45	TCTAGGGAATTTGCCCGCAGCATATTGACATTAAAATAGACTTTGCATATAATAAGATCTTCGAAGGAGGGATTACTA	78	9091	0.09524722
46	CTTGGGTAGACTCAGGCAACTCATTTGACACACTTGAGAAATATTTATATAATCGTTCTAGAAAAGGAGGTAGATATC	78	11374	1.546669
47	GGCCACGCGGGTAGTGCCGCCTCATTGACAACAACTAGTAATTTTAGTATAATATTAATCAGCAAGGAGGTGTACTAA	78	8218	0.423913
48	AATATTTTATTAATTTATAAAATATTGACACACTTGAGAAATATTTATATAATAAATGTAGTGAAGGAGGATGTCTA	77	9610	0.5996491
49	TTTCACCGTCTGATAACGGCGGAATTGACAAAAATGGGCTCGTGTTGTATAATTCAGAAATAGCTAGTTCTGCAAGGAGGGA TTACTA	88	7873	0.2157018
50	GTCAGGGCGGCTACGCGCTGCCCATTGACAACTGAAGATATATTCTATATAATAAAGTCGTCTAAGGAGGGATTACTA	78	9311	1.77631
51	CTTGGGTAGACTCAGGCAACTCATTTGACATAATAGACAACTTATGTTATAATATTGATCAGCAAGGAGGTAGATATC	78	8921	0.4707828
52	ACGTTGATATAATTTAAATTTTATTTGACAACTGAAGATATATTCTATATAATATTGATCAGCAAGGAGGATGTCTAA	78	8274	-0.0264253
53	TTTCACCGTCTGATAACGGCGGAATTGACAACAACTAGTAATTTTAGTATAATATTGATCAGCAAGGAGGTGTACTAA	78	9406	0.1096263
54	GGGGAGCTGCATGCCCCTGACCCTTGACAACAACTAGTAATTTTAGTATAATCCCTGTAGAAGCGAGTAATACAATCTTGTTA AGGAGGCTAATGTA	97	7982	2.655921
55	GGGGAGGCTGCATGCCCCTGACCCTTGACAACTGAAGATATATTCTATATAATCGTTCTAGAAAAGGAGGATGTCTAA	78	7922	2.382841
56	GGCCACCGGGTACCTCATTGACACACTTGAGAAATATTTATATAATATTGATCAGCAAGGAGGGATTACTA	71	7939	0.1872822
57	TCTAGGGAATTTGCCCGCAGCATATTGACAAAAATGGGCTCGTGTTGTATAATTACAATACGTTCAGTTCTTCGGGCGAAAAA AGGAGGGATTACTA	97	8289	1.121212
58	TCATTTGATGTACATAGCATAATATTGACACACTTGAGAAATATTTATATAATATTGATCAGCAAGGAGGCTAATGTA	78	7910	0.1493776
59	GTCAGGGCGGCTACGCGCTGCCCATTGACATAATAGACAACTTATGTTATAATCGTTCTAGAAAAGGAGGGATTACTA	78	11082	3.782702
60	TTAATATATATTAAATTAATATATTTGACAACAACTAGTAATTTTAGTATAATTTGAGAGCTCCGATTATCAAAAGGAGTGTACT AA	87	7614	0.04699289
61	AATATTTTATTAATTTATAAAATATTGACAAAAATGGGCTCGTGTTGTATAATCCTAACTAGTGTATGTACGAAGGAGGATGTC TAA	87	7903	0.796909
62	GAGTAATAAGTATCTTCATATCTATTGACAACAACTAGTAATTTTAGTATAATCCCTGTAGAACGAGTAATACAATCTTGTTAA GGAGGGATTACTA	97	16982	8.672205
63	CTTGGGTAGACTCAGGCAACTCATTTGACACACTTGAGAAATATTTATATAATAAATGTAGTGAAGGAGGCTAATGTA	78	7864	0.7492545
64	TCATTTGATGTACATAGCATAATATTGACAACAACTAGTAATTTTAGTATAATATAGTGAGCGTATTCATCCTATCAGAACTAA GGAGGTAGATATC	97	10306	1.446706
65	GAGTAATAAGTATCTTCATATCTATTGACAAAAATGGGCTCGTGTTGTATAATGGACTTTACAAAGGAGGTGTCTAA	77	8971	6.396535
66	CTTGGGTAGACTCAGGCAACTCATTTGACATTAAAATAGACTTTGCATATAATAAAGTCGTCTAAGGAGGTGTACTAA	78	8727	0.1759966

67	ATAGCCTATTATGTTCTTGAAAAATTGACACACTTGAGAAATATTTATATAATAAGATCTTCGAAGGAGGCTAATGTA	78	8316	0.169162
68	GAGTAATAAGTATCTTCATATCTATTGACATTAAAATAGACTTTGCATATAATGGACTTTACAAAGGAGGTGTACTAA	78	8968	3.590414
69	GGGGAGGCTGCATGCCCCTGACCCTTGACAACAACTAGTAATTTTAGTATAATAAATGTAGTGAAGGAGGGATTACTA	78	11232	1.595969
70	GGCCACGCGGGTAGTGCCGCCTCATTGACATTAAAATAGACTTTGCATATAATATAGTGAGCGTATTCATCCTATCAGAAGC TAAGGAGGTAGATATC	98	9891	0.5552799
71	TAAATTATAATTTTATTAAATTATTGACAACTGAAGATATATTCTATATAATTGAGTTAGGTCCCTCAAAAAAGGAGGTGTACTA A	86	7843	6.761287
72	GAGTAATAAGTATCTTCATATCTATTGACAACTGAAGATATATTCTATATAATCCCTGTAGAAGCGAGTAATACAATCTTGTTA AGGAGGGATTACTA	98	15209	6.516927
73	ACGTTGATATAATTTAAATTTTATTTGACACACTTGAGAAATATTTATATAATAAATGTAGTGAAGGAGGTGTACTAA	78	9711	6.228245
74	GCTGCGCGCACCGAGCCTTCAGGTTGACAACAACTAGTAATTTTAGTATAATCTAACTAGTGTATGTAACGAAGGAGGATGC A	83	7661	0.293262
75	GGGGAGGCTGCATGCCCCTGACCCTTGACAACAACTAGTAATTTTAGTATAATAAATGTAGTGAAGGAGGCTAATGTA	78	8155	1.741395
76	TCATTTGATGTACATAGCATAATATTGACATTAAAATAGACTTTGCATATAATGGACTTTACAAAGGAGGTAGATATC	78	11880	3.208977
77	CTTGGGTAGACTCAGGCAATCATTTGACACACTTGAGAAATATTTATATAATAAATGTAGTGAAGGAGGATGTCTAA	77	9490	0.4230582
78	TAAATTATAATATTTATTAAATTATTGACAAAAATGGGCTCGTGTTGTATAATCCCTGTAGAAGCGAGTAATACAATCTTGTTAA GGAGGATGCA	95	8505	1.699786
79	CTTGGGTAGCTCAGGCAACTCATTTGACAACTGAAGATATATTCTATATAATAAGATCTTCGAAGGAGGTGTACTAA	77	9968	0.8981788
80	TTATATATTAAAATATAATTATATTTGACAAAAATGGGCTCGTGTTGTATAATCGTTCTAGAAAAGGAGGCTAATGTA	78	10280	6.272795
81	GGCCACGCGGGTAGTGCCGCCTCTTGACAATGAAGATATATTCTATATAATAAGATCTTCGAAGGAGGGATTACTA	76	8159	0.1098921
82	TCTAGGGAATTGCCCCAGCATATTGACAACAACTAGTAATTTTAGTATAATAAATGTAGTGAAGGAGGTGTACTAA	76	8363	2.828392
83	TTTCACCGTCTGATAACGGCGGAATTGACAACTGAAGATATATTCTATATAATTACAATACGTTTCAATTCTTCGGGCGAAAAA AGGAGGGATTACTA	98	7634	1.001409
84	GGTATAGTCCGTACAACCACGGTTGACATAATAGACAACTTATGTTATAATGGACTTTACAAAGGAGGCTAATGTA	76	7544	1.178853
85	AATATTTTATTAATTTATAAAATATTGACATTAAAATAGACTTTGCATATAATAAATGTAGTGAAGGAGGTAGATATC	78	11841	4.175819
86	GACTCATTAATTAACAGATTTGTATTGACAACAACTAGTAATTTTAGTATAATGGACTTACAAAGGAGGGATTACTA	77	13778	6.523531
87	TTATATATTAAAATATAATTATATTTGACATAATAGACAACTTATGTTATAATAAGATCTTCGAAAGAGGTAGTATC	77	6849	0.04182769
88	GCTGCGCGCACCGAGGCCTTCAGGTTGACATAATAGACAACTTATGTTATATAAATGTAGTGAAGGAGGATGTCTAA	77	7546	0.3157062
89	TCATTTGATGTACATAGCATAATATTGACATTAAAATAGACTTTGCATATAATTCAGAAATAGCTAGTTCTGCAAGGAGGCTAA TGTA	88	7127	0.02405248

90	GCTGCGCCACGAGGCCTTCAGGTTGACATAATAGACAACTTATGTTATAATCGTTCTAGAAAAGGAGGCTAATGTA	76	9413	3.536561
91	TTAATATATATTAAATTAATATATTTGACATAATAGACAACTTAGTTATAATACTTGACGTTCTGAGCAATAAAGGAGGATGCA	84	7344	0.08615246
92	GAGTTTGATAATCATATACCATTTTGACATAATAGACAACTTATGTTATAATTTGAGAGCTCGATTATCAAAAGGAGGGATTAC TA	86	9571	0.8130589
93	AATATTTTATTAATTTATAAAATATTGACAACAACTAGTAATTTTAGTATAATGGACTTTACAAAGGAGGGATTACTA	78	12852	3.807342
94	GTCAGGGCGGCTCGCGCTGCCCATTGACATTAAAATAGCTTTGCATATAATCGTTCTAGAAAAGGAGGCTAATGTA	76	7746	0.1106121
95	TAAATTATAATATTTATTAAATTATTGACATAATAGACAACTTATGTTTAATGGACTTTACAAAGGAGGTGTCTAA	76	7925	0.01753959
96	ATAGCCTATTATGTTCTTGAAAAATTGACAAAAATGGGCTCGTGTTGTATAATATTGATCAGCAAGGAGGCTAATGTA	78	7689	0.1591009
97	GGCCACGCGGGTAGTGCCGCCTTTGACATAATAGACAACTTATGTTATAATGATCCTTAGAGATGCTAACTGCCAAGATAAG GAGGATGTCTAA	94	10403	1.909599
98	ACTGTTGGCGACCTTCGATGAAACTTGACATTAAAATAGACTTTGCATATAATAAATGTAGTGAAGGAGGTAGATATC	78	18764	2.317011
99	GCTGCGCGCACCGAGGCCTTCAGGTTGACAACTGAAGATATATTCTATATAATTACAATACGTTTCAGTTCTTCGGGCGAAA AAAGGAGGTGTACTAA	98	8238	0.337011
100	GAGTAATAAGTATCTTCATATCTATTGACACACTTGAGAAATATTTATATAATGGACTTTACAAAGGAGGATGTCTAA	78	7549	1.947256
101	ACGTTGATATAATTTAAATTTTATTTGACACACTTGAGAAATATTTATATAATTGAGTTAGGTCCTCTCAAAAAAGGAGGGATTA CTA	88	7644	0.9922604
102	TCGAGGCCTACGCGACCTCGGCGGTTGACAACTGAAGATATATCTATATAATAAATGTAGTGAAGGAGGGATTACTA	77	8126	0.4944326
103	TTTCACCGTCTGATAACGGCGGATTGACAAAAATGGGCTCATGTTGTATAATGGACTTACAAAGGAGGTGTACTAA	76	11741	4.511729
104	TAAATTATAATATTTATTAAATTATTGACAACAACTAGTAATTTTAGTATAATTACAATACGTTTCAGTTCTTCGGGCGAAAAAA GGAGGATGCA	95	7304	0.3952517
105	GAGTAATAAGTATCTTCATATCTATTGACAACTGAAGATATATTCTATATAATCGTTCTAGAAAAGGAGGTGTACTAA	78	11270	5.031819
106	ACGTTGATATAATTTAAATTTTATTTGACAACTGAAGATATATTCTATATAATAAATGTAGTGAAGGAGGTAGATATC	78	21601	10.17472
107	TTATATATTAAAATATAATTATATTTGACAACAACTAGTAATTTTAGTATAATTTGAGACTCCGATATCAAAGGAGCGATTACT	84	7853	0.03387059
108	ACTGTTGGCGACCTTCGATGAAACTTGACACACTTGAGAAATATTTATATAATTGAGTTAGGTCCTCTCAAAAAAAGAGGTGT CTAA	87	6848	0.0790394
109	TTGGTATAGTCCGTACAACCCGGTTGACAATGAAGATATATTTATATAACGTTCTAGAAAAGGAGGATGCA	71	7498	0.02669633
110	GTCAGGCGGCTCGCGCTGCCCATTGACATTAAAATAGACTTTGCATATAATAAATGTAGTGAAGGAGGTAGATATC	76	10748	2.351243
111	TTGGTATAGTCCGTCAACCACGGTTGCATTAAAATAGACTTTGCATATATAAGATCTTCGAAGGAGTAGATATC	74	7190	0.00096432
112	TCTAGGGAATTTGCCCGCAGCATATTGACATTAAAATAGACTTTGCATATAATTACAATACGTTTCAGTTCTTCGGGCGAAAA AAGGAGGTGTACTAA	98	8250	0.3560951

113	TCGAGGCCTACGCGACCTCGGCGGTTGACACACTTGAGAAATATTTATATAATAAATGTAGTGAAGGAGGATGCA	75	7045	0.1566203
114	TCGAGGCCTACGCGACTCGGCGGTTGACAACTGAAGATATATTCTATATAATACTTGCGTCTGAGCATAAAGGAGGTAGATA TC	84	8154	0.594943
115	TCTAGGGAATTTGCCCGCAGCATATTGACATAATAGACAACTTATGTTATAATATAGTGAGCGTATTCATCCTATCAGAAGCT AAGGAGGATGTCTAA	98	6761	0.658475
116	AATATTTTATTAATTTATAAAATATTGACAACAACTAGTAATTTTAGTATAATATGATCAGCAAGGAGGCTAATTTA	77	8417	0.1202741
117	TTTCACCGTCTGATAACGGCGGAATTGACAAAAATGGGCTCGTGTTGTATAATCGTCTAGAAAAGGAGGGATTACTA	77	11591	1.662642
118	GGCCACGCGGGTAGTGCCGCCTCATTGACATTAAAATAGACTTTGCATATAATATTGATCAGCAAGGAGGCTAATGTA	78	6877	0.0387688
119	GACTCATTAATTAACAGATTTGTATTGACATTAAAATAGACTTTGCATATAATAAATGTAGTGAAGGAGGATGTCTAA	78	11007	1.926078
120	GTCAGGGCGGCTACGCGCTGCCCATTGACACACTTGAGAAATATTTATATAATAAAAAGGAGGATGTCTAA	71	6875	0.03935468
121	ACTGTTGGCGACTTCGATGAAATTGACAACACTAGTAATTTAGTATAATATTGATCAGCAAGGAGGTAGATATC	74	9206	1.331146
122	ACTGTTGGCGACCTTCGATGAAACTTGACACACTTGAGAAATATTTATATAATCCTAACTAGTGTATTAACGAAGGAGGTGTA CTAA	87	7725	0.4572104
123	TCTAGGGAATTTGCCGCAGCATATTGACAACTGAAGATATATTCTATATAATCGTTCTAGAAAAGGAGGCTAATGTA	77	8416	2.716569
124	ACGTTGATATAATTTAAATTTTATTTGACAACTGAAGATATATTCTATATAATCGTTCTAGAAAAGGAGGATGTCTAA	78	9709	3.077818
125	TTAATTATATATTTATTAAATTATTGACATAATAGACACTTTGTTATATTTGAGAGCTCCGATATCAAAAGGAGGGATTACTA	83	6690	-0.0141098
126	TCGAGCCTACGCGACCTCGGCGGTTGACACACTTGAGAAATATTTATATAATGGTTCTAGAAAAGGAGGTGTACTA	76	9208	3.51595
127	TTGGTATAGTCCGTACAACCACGGTTGACAACAACTAGTAATTTTAGTATAATCGTTCTAGAAAAGGAGGTGTACTAA	78	8416	2.520707
128	TAAATTATAATATTATTAAATTATTGACACACTTGAGAAATATTTATATAATAAATGTAGTGAAGGAGGTAGATATC	77	16372	5.683824
129	GGGGAGGCTGCATGCCCCTGACCCTTGACATAATAGACACTTATGTTATAATCTAACTAGTGTATGTAACGAAGGAGGATGT CTAA	86	9041	1.152009
130	GAGTAATAAGTATCTTCATATCTATTGACAACAACTAGTAATTTTAGTATAATTCAGAAATAGCTAGTTGCAAGGAGGTGTACT AA	86	8059	0.6862287
131	ACTTGATATAATTTAAATTTTATTTGACAACTGAAGATATATTCTATATAATAAATGTAGTGAAGGAGGATGTCTAA	77	9014	1.789065
132	AATATTTTATTAATTTATAAAATATTGACAACAACAGTAATTTTAGTATAATTTGAGAGCTCGATTATCAAAAGGAGGATGCA	83	7506	0.01311249
133	CTTGGGTAGACTCAGGCAACTCATTTGACAACTGAAGATATATTCTATATAATCCCTTAGAAGCGAGTAATAAAATCTTGTTAA GGAGGGATTAT	95	13697	2.387006
134	TAAATTATAATATTTATTAAATTATTGACAACTGAAGATATATTCTATATAATAAAGTCGTCTAAGGAGGTAGATATC	78	9145	1.641379
135	GAGTTTAGATAATCATATACCATTGACAACAACTAGTAATTTTAGTATAATATTGATCAGCAAGGAGGTAGATATC	76	6967	0.6376973

Table A.1: Sequences and intensities (mean, GFP/mCherry) of 135 clones for the verification of the expression module library: ~34% of the sequences matched the bioinformatics design of ~12,000 promoter variants (Section 5.2.1).

Table A.2 – Sequences of selected expression modules

ID	**Sequence** (5' to 3')
veg	ACGTTGATATAATTTAAATTTTATTTGACAAAAATGGGCTCGTGTTGTACAATAAATGTAGTGAGGTGGATGCA
P_{15}	ATCCACGCTGTGTAAAAATTTTACAAAAAGGTATTGACTTTCCCTACAGGGTGTGTAATAATTTAATTAAAGAAAATGAGAGGGAGAGGAAATTAATTAAAAA AGGAGCGATTTACAT
design	GAGAAGTTCAAAAATATTATTGACATTTATATGTAAATATGTTATAATGCAGATATAGGCGAAAGGCCAATGTAAAAAGGAGGACAAGTCA
Variant 1	TAGGGAATTTGCCCGCAGCATATTGACAACTGAAGATATATTCTATATAATCGTTCTAGAAAAGGAGGTAGATTC
Variant 2	GAGTAATAAGTATCTTCATATCTATTGACAACAACTAGTAATTTTAGTATAATCCCTTTAGAAGCGAGTAATACAATCTTGTTAAGGAGGATGTCTAA
Variant 3	GAGTAATAAGTATCTTCATATCTATTGACAAAAATGGGCTCGTGTTGTATAATGGATTTACAAAGGAGGTAGATATC
Variant 4	CGAGGCCTACGCGACCTCGGCGGTTGACATAATAGACAACTTATGTTATAATCGTTCTAGAAAAGGAGGTAGATATC
Variant 5	GACTCATTAATTAACAGATTTGTATTGACAAAAATGGGCTCGTGTTGTATAATAAATGTAATGAAGGAGGTAGATATC
Variant 7	GACTCATTAATTAACAGATTTGTATTGACAACTGAAGATATATTCTATATAATGGACTTTACAAAGGAGGTAGATATC
Variant 12	ATAGCCTATTATGTTCTTGAAAAATTGACAACAACTAGTAATTTTAGTATAATGGACTTTACAAAGGAGGTAGATATC
Variant 15	ACGTTGATATAATTTAAATTTTATTTGACAACAACTAGTAATTTTAGTATAATGGACTTTACAAAGGAGGGATTACTA
Variant 17	TCTAGGGAATTTGCCCGCAGCATATTGACATTAAAATAGACTTTGCATATAATGGACTTACAAAGGAGGTAGATATC
Variant 18	GACTCATTAATTAACAGATTTGTATTGACATTAAAATAGACTTTGCATATAATGGACTTTACAAAGGAGGTAGATATC
Variant 20	GAGTAATAAGTATCTTCATATCTATTGACATAATAGACAACTTATGTTATAATCGTTCTAGAAAAGGAGGTGTACTAA
Variant 21	GACTCATTAATTAACAGATTTGTATTGACATAATAGACAACTTATGTTATAATAAATGTAGTGAAGGAGGTAGATATC
Variant 23	TATTTTTAATTTATATATAAAAAATTGACAACTGAAGATATATTCTATATAATCGTCTAGAAAAGGAGGTAGATATC
Variant 24	GAGTAATAAGTATCTTCATATCTATTGACAAAAATGGGCTCGTGTTGTATAATGGACTTTACAAAGGAGGGATTACTA
Variant 25	AATATTTTATTAATTTATAAAATATTGACAACTGAAGATATATTCTATATAATAAATGTAGTGAAGGAGGGATTACTA
Variant 26	GAGTAATAAGTATCTTCATATCTATTGACAACTGAAGATATATTCTATATAATCGTTCTAGAAAAGGAGGTAGATATC

Table A.2: Overview of three control expression modules (*veg*, P_{15}, design) and the 16 variants used for the expression of both *gfp* and *xynA* (Section 5.2.2).

Appendix B

ACS SyntheticBiology

Letter

pubs.acs.org/synthbio

Effect of Genome Position on Heterologous Gene Expression in *Bacillus subtilis*: An Unbiased Analysis

Christopher Sauer,[†,‡,∥] Simon Syvertsson,[†,∥] Laura C. Bohorquez,[§] Rita Cruz,[†,‡] Colin R. Harwood,[†] Tjeerd van Rij,[‡] and Leendert W. Hamoen*[,†,§]

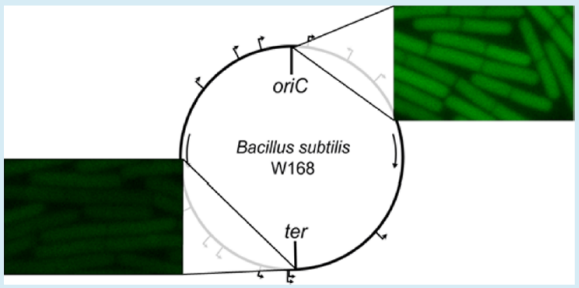
[†]Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Richardson Road, NE2 4AX Newcastle, United Kingdom

[‡]DSM Biotechnology Center, P.O. Box 1, 2600 MA Delft, The Netherlands

[§]Bacterial Cell Biology, Swammerdam Institute for Life Sciences, University of Amsterdam, De Boelelaan, 1081 HZ Amsterdam, The Netherlands

Supporting Information

ABSTRACT: A fixed gene copy number is important for the *in silico* construction of engineered synthetic networks. However, the copy number of integrated genes depends on their genomic location. This gene dosage effect is rarely addressed in synthetic biology. Two studies in *Escherichia coli* presented conflicting data on the impact of gene dosage. Here, we investigate how genome location and gene orientation influences expression in *Bacillus subtilis*. An important difference with the *E. coli* studies is that we used an unbiased genome integration approach mediated by random transposon insertion. We found that there is a strong gene dosage effect in fast growing *B. subtilis* cells, which can amount to a 5-fold difference in gene expression. In contrast, gene orientation with respect to DNA replication direction does not influence gene expression. Our study shows that gene dosage should be taken into account when designing synthetic circuits in *B. subtilis* and presumably other bacteria.



KEYWORDS: *Bacillus subtilis, gene expression, genome location, transposon, GFP, LacZ*

Bacillus subtilis is widely used as bacterial model system and for the production of enzymes and vitamins. It is genetically tractable, easily culturable and regarded as safe. This has resulted in the development of a wide array of genetic tools for use in both academic and industrial environments, and has made the organism a popular cloning chassis in synthetic biology.[1] The natural genetic transformation system of *B. subtilis* facilitates the integration of DNA into the genome by homologous recombination.[2,3] In contrast to the use of plasmids as genetic carriers, genome integration assures that heterologous genetic material is stably maintained at a copy number that reflects that of the chromosome. The latter is important when genetic modules (*e.g.* biobricks) are used for the construction of engineered synthetic metabolic pathways and regulatory networks.[1,4] However, since DNA replication originates at a fixed position on the genome (origin of replication), and rapidly growing bacteria initiate new rounds of replication before the previous rounds have been completed, the copy number of integrated genes will fluctuate. As a result, gene dosage rises with increasing proximity to the origin of replication.[5,6] Another aspect that might affect the activity of integrated genes is the transcriptional orientation. The majority of native genes are codirectional with DNA replication[7−9] in order to reduce collisions between RNA polymerase and the DNA replication machinery.[10,11] It is therefore possible that the orientation of integrated genes will influence expression. In this study, we set out to determine to which extent genome location and gene orientation influences heterologous gene expression in *B. subtilis*.

Several groups have studied the effects of genome location on gene expression using the model organism *Escherichia coli*. By varying the position of a reporter gene on the genome, Block and co-workers have shown that expression is only influenced by gene dosage and not by the orientation of the gene relative to the direction of DNA replication.[12] They found that in rapidly growing cultures of *E. coli* the maximum expression levels were approximately 9-fold higher for gene locations close to the origin of replication compared to locations close to the replication terminus.[12] More recently, Bryant and co-workers reported that gene expression can vary up to 300-fold for a reporter fusion that was integrated at different locations on the *E. coli* genome.[13] Moreover, they concluded that gene dosage had only a minor influence on expression, and that the observed differences are

Received: February 16, 2016

related to specific features involved in chromosome organization. The discrepancy between these reports may therefore reflect the particular locations chosen for the integration of the reporter fusions.[12,13] To prevent this complication, we developed an unbiased reporter integration approach for *B. subtilis*.

Instead of choosing the insertion sites of reporter genes, we used a transposon-mediated random insertion approach and selected clones exhibiting a wide range of expression activities. Following this approach, we found that expression of a *lacZ* reporter gene can differ up to 5-fold based on its chromosomal location. This difference in expression correlated strongly with the location of the reporter cassette relative to the origin of replication, and was not influenced by gene orientation with respect to the DNA replication direction. Thus, gene dose is an important factor influencing heterologous gene expression in *B. subtilis*, and this should be taken into account when designing *in silico* synthetic regulatory circuits, and also when expressing commercially valuable heterologous proteins.

To be able to make an unbiased assessment of the influence of genomic location on gene expression in *B. subtilis*, we made use of the Mariner transposon.[14,15] This transposon only requires a TA dinucleotide sequence for transposition and consequently it is known to insert randomly into the genome.[16−18] To determine the influence of gene orientation on expression, a bidirectional reporter cassette was constructed comprising the constitutively active P_{veg} promoter and the inducible P_{spac} promoter, fused, respectively to *gfp* and *lacZ* reporter genes (Figure 1). To control the P_{spac} promoter, the *lacI* repressor

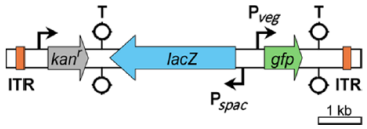


Figure 1. Schematic overview of the Mariner transposon containing the bidirectional reporter cassette. The housekeeping SigA-dependent P_{veg} promoter and the inducible P_{spac} promoter are fused to *gfp* and *lacZ*, respectively, and located between strong intrinsic transcription terminators (T) to prevent read-through from and into adjacent genes. The transposons are selected by kanamycin resistance (kan^R). The inverse terminal repeats (ITR) used by the transposase are indicated. The *lacI* gene coding for the repressor of the P_{spac} promoter is integrated at the genomic *aprE* locus (not indicated).

gene was integrated into the genome at the *aprE* locus. Transcriptional read-through from and into adjacent genes was avoided by the inclusion of strong bidirectional terminators at both ends of the cassette (Figure 1). These bidirectional *B. subtilis* terminators were identified using the WebGeSTer database.[19] The reporter cassette was subsequently cloned into the Mariner transposon. Despite the size of the cassette (~6 kb), the transposon was still active, albeit with a transposition frequency that was an order of magnitude lower than the wild type transposon (Supporting Information Table S1).

Following transposition, cells were grown on nutrient agar plates containing X-gal and IPTG, and colonies showing a range of color intensities, from dark blue to almost white, were selected at random for detailed analysis. To ensure that the transposon insertions did not influence growth, the growth rates of the selected clones were measured in microtiter plates. Fourteen clones that showed normal growth rates (Supporting Information Figure S1) were selected for further analyses. The locations of the transposon insertions on the *B. subtilis* genome are shown in Figure 2 and detailed genetic information about these loci is provided in the Supporting Information (Figure S2).

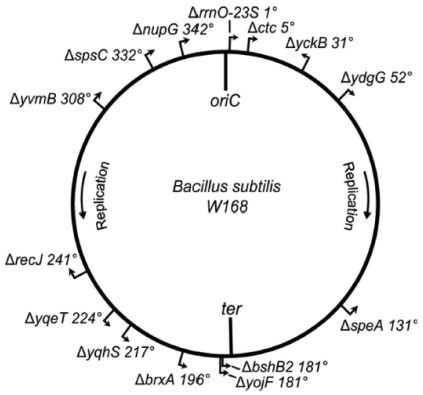


Figure 2. Chromosomal location of selected transposon mutants. Arrows indicate the orientation of the *gfp* marker and the location is given in degrees (1° ≙ 11,708 base pairs). The origin of replication (*oriC*) and replication terminus (*ter*) are indicated. In *B. subtilis* the *ter* region is located approximately 178.7° on the *B. subtilis* genome. DNA replication directions are indicated.

The 14 transposon insertion strains were grown in LB medium at 37 °C in the presence of 1 mM IPTG, and samples were taken during exponential growth (OD_{600} ~ 0.5) to determine β-galactosidase activity and fluorescence intensity (GFP) measurements. Figure 3A shows the β-galactosidase activity and GFP fluorescence intensity with respect to the locations of the transposons on the genome. The orientation of the reporter genes is indicated by arrowheads. The trend lines show a clear correlation between chromosome location and gene expression for both reporters. The maximum difference in expression between a gene located close to the replication origin (*oriC*) or to the terminus (*ter*) is approximately 5-fold. The differences in GFP expression levels are clearly visible by eye (Figure 3B). The experiment was carried out in triplicate and a representative set of data is shown in the Supporting Information (Figure S3A).

To determine whether the observed activities were influenced by read-through from adjacent genes, β-galactosidase activities were also measured in the absence of IPTG. In the eight transposon clones tested, the absence of IPTG resulted in very low β-galactosidase activities, maximally of 6 Miller Units (Supporting Information Figure S4), indicating that the reporter cassette is transcriptionally isolated by the flanking bidirectional terminators, and therefore not influenced by transcriptional read-through from adjacent genes.

It has been reported that in *B. subtilis* the copy number of a gene close to *oriC* can be ~5 times higher than that of a gene close to the *ter* (*ori*:*ter* ratio) when growing exponentially (~2.5 doublings/h) in LB medium.[20] To examine the relation between expression and gene dosage, DNA copy numbers of the different loci were determined using qPCR (Figure 3C). We found a comparable *ori*:*ter* ratio as described before,[20] and a distribution of DNA copy numbers of the different loci that reflects the trends of *lacZ* and *gfp* expression, indicating that gene dosage is a major factor influencing gene expression in fast growing cells.

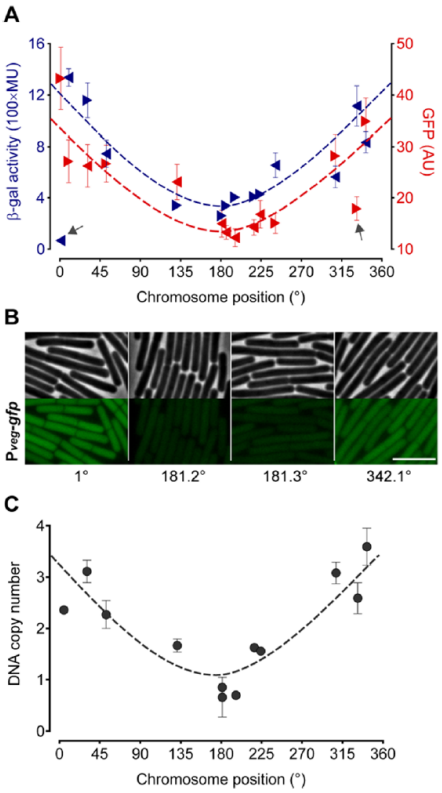


Figure 3. Effect of chromosomal location on gene expression. Cells were grown in LB medium at 37 °C and samples were taken during exponential growth ($OD_{600} \sim 0.5$). (A) β-galactosidase activities (blue) and GFP fluorescence (red) are plotted relative to the chromosomal locations (degrees) of the transposons. Error bars indicate standard deviations of two technical replicates in case of β-galactosidase and at least 100 cells for GFP. The triangles indicate transcriptional direction of *lacZ* or *gfp*. Two outliers are marked by gray arrows. Goodness of fit for β-galactosidase with $R^2 = 0.86$ (excluding locus at 1°) and GFP with $R^2 = 0.70$. (B) Visual differences in GFP expression illustrated by fluorescence micrographs of four different transposons. Scale bar 5 μm. (C) DNA copy numbers of transposon locations determined by qPCR. Error bars indicate standard deviations of three biological replicates. Goodness of fit with $R^2 = 0.78$.

Despite the clear relationship between expression levels and the genomic location of the reporters, a few transposon insertions deviated from the trend lines (see arrows Figure 3A). For example, the transposon integrated close to *oriC* (~1°) showed a much lower β-galactosidase activity than expected, even though the sequences of the *lacZ* gene and P_{spac} promoter did not reveal any mutations. This transposon was inserted into the nonessential ribosomal gene *rrnO-23S* that is part of the *rrnO-16S* operon (Supporting Information Figure S2). Possibly, the terminator of the transposon cassette is not fully able to prevent read-through from the strongly transcribed *rrnO-16S* operon. The transposon insertion at 332° (*spsC*) showed a significantly lower than expected GFP activity for reasons that are currently unknown (Figure 3A).

Like most other bacteria, the orientation of genes on the *B. subtilis* genome is strongly correlated with the direction of DNA replication.[7,9] To determine whether gene orientation affected expression, β-galactosidase activities and GFP fluorescence intensities were analyzed with respect to their orientation, as indicated in Figure 4. The expression levels

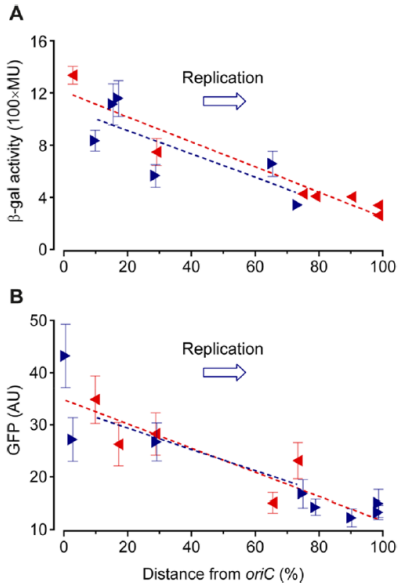


Figure 4. Effect of gene orientation on gene expression. β-Galactosidase activities (A) and GFP fluorescence intensities (B) from Figure 3 are plotted against the distance of the transposon insertion sites from the origin of replication (*oriC*). Triangles indicate the transcriptional direction of the reporter genes. Reporter genes that are transcribed in the same direction as DNA replication are shown in blue and genes that are transcribed against the DNA replication direction are shown in red. Error bars indicate the standard deviation of two technical replicates in case of β-galactosidase and at least 100 cells for GFP. Goodness of fit for β-galactosidase with $R^2 = 0.60$ (blue), $R^2 = 0.91$ (red) and for GFP with $R^2 = 0.82$ (blue), $R^2 = 0.66$ (red).

were plotted against the distance of the transposon insertion sites from the origin of replication. The outliers at 1° and 332° were discarded in this analysis. The resulting trend lines are comparable (Figure 4), indicating that gene orientation did not have a strong influence on expression levels. One replicate is shown in Supporting Information Figure S3B–C. At first sight this might seem surprising considering the strongly biased orientation of the genes on the genome, however, this bias is likely related to the negative effect of RNA polymerases colliding with the DNA replication machinery, resulting in less efficient DNA replication rather than effecting transcription efficiency.

B. subtilis grows considerably slower in minimal medium containing glucose and amino acids than in nutritionally rich LB medium.[21,22] This is also reflected in a lower *ori*:*ter* ratio which in minimal medium approaches 3:1.[20] If expression differences are indeed due to gene dosage effects, this should be reflected in a reduced expression ratio when grown in minimal medium. To test this, four transposon strains were selected; two with the reporter cassette close to the terminus (181° and 191°) and two with the reporter cassette close to the origin of replication (5° and 342°). The strains were grown in LB medium and minimal medium with mean generation times at 37 °C of 27 and 45 min, respectively. Indeed, growth in minimal medium reduced the difference in *lacZ* expression from around 3.5-fold to 2.5-fold compared to LB (Figure 5). However, the *gfp* marker, for unknown reasons, did not show a clear reduction in expression.

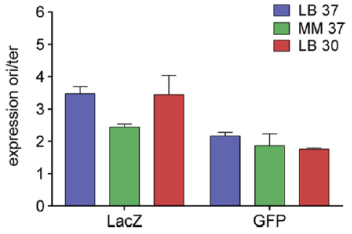


Figure 5. Influence of growth conditions on gene expression. The fold expression differences of *β*-galactosidase and GFP under different growth conditions were calculated by dividing the average expression levels of two loci close to the origin by that of two loci close to the terminus of replication. Cells were grown in LB medium (blue) and minimal medium (green) at 37 °C, and in LB medium at 30 °C (red). Error bars indicate the standard deviation of three biological replicates.

Murray and Koh have shown that the *ori*:*ter* ratio is similar when *B. subtilis* is grown in LB medium at 30 and 37 °C,[20] despite the mean generation time at 30 °C (50 min) being virtually double that at 37 °C (27 min). This appears to be reflected in the unchanged expression activities of both marker genes (Figure 5).

Since DNA replication is markedly reduced when exponential growth ceases, the differential expression of different transposon strains should decrease as well. However, when we measured *β*-galactosidase activities and GFP intensities during stationary phase, the differential expression profiles with respect to location were still observed (Figure 6A,B. Replicate shown in Figure S5). Both *β*-galactosidase and GFP are stable proteins and a possible explanation might be that the observed trends are still visible in the stationary growth phase as a consequence of the difference in reporter protein expression during logarithmic growth. To test this, *lacZ* was induced only in the stationary phase (Figure 6A) and, as expected, the observed *β*-galactosidase levels were much smaller compared to the induction of *lacZ* during logarithmic growth (compare *Y*-axis scales). More importantly, the difference in *lacZ* expression between the transposon strains was almost completely abolished (Figure 6C). To confirm that this effect is a consequence of a reduction in gene dosage, the DNA copy numbers of the different loci were determined using qPCR (Figure 6D). Again the distribution of DNA copy numbers followed closely the trend observed for *lacZ* expression.

Our unbiased selection approach shows that gene dosage as a consequence of multiple replication forks plays a considerable role in gene expression in *B. subtilis*, which is in full agreement with the study of Block and co-workers,[12] but does not support the conclusions from the study by Bryant and co-workers.[13] Similar to Block and co-workers, we also found that the orientation of reporter genes has no influence on expression levels. The importance of gene dosage in *B. subtilis* is not restricted to heterologous proteins as was shown in two recent studies that revealed the impact of chromosomal location of genes involved in the regulation of motility, biofilm development and sporulation.[23,24]

Clearly, gene dosage effects have to be taken into account when locating synthetic circuits in bacteria. Moreover, a potential 5-fold increase in enzyme production represents a commercially significant improvement. In fact, our transposon reporter system might be adapted to improve the production of enzymes that are expressed from the genome, since we show that even large DNA constructs (at least 6 kb) can be randomly

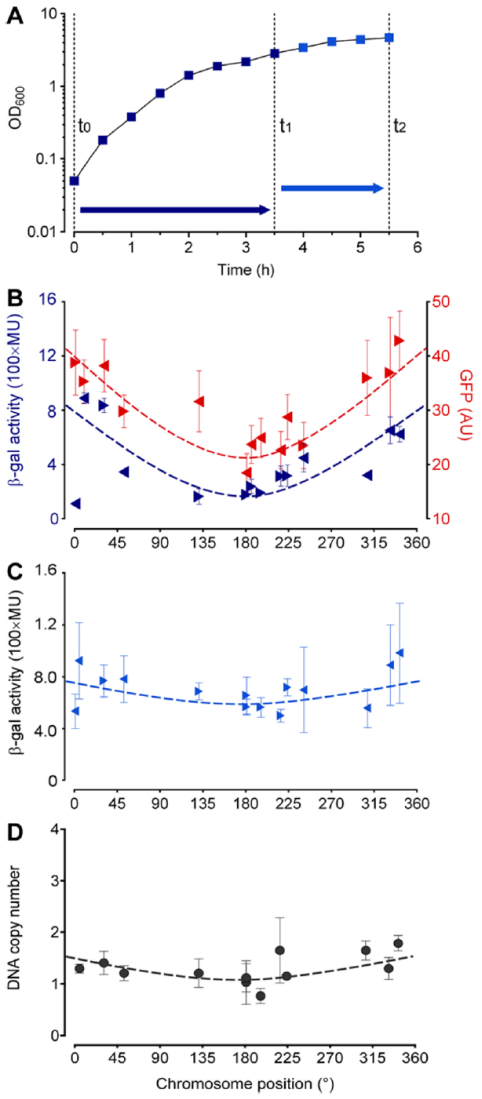


Figure 6. Reporter gene expression in stationary growth. (A) Transposon strains were grown in LB at 37 °C and *lacZ* expression was induced with 1 mM IPTG at the start of growth (t_0) after which cells were harvested 3.5 h later (t_1), or at the beginning of the stationary phase (t_1) after which cells were harvested 2 h later (t_2). (B) *β*-galactosidase activities (blue) and GFP fluorescence (red) are plotted against chromosomal location (degrees) of the transposons when 1 mM IPTG was added at the start (t_0) of growth. Error bars indicate the standard deviation of two technical replicates in case of *β*-galactosidase and at least 100 cells for GFP. Goodness of fit for *β*-galactosidase with $R^2 = 0.77$ (excluding locus at 1°) and GFP with $R^2 = 0.80$. (C) *β*-galactosidase activities when 1 mM IPTG was added in the beginning of the stationary growth phase (t_1). Average values with standard deviations of three independent biological replicates are shown. Direction of the triangles indicates transcriptional direction of *lacZ* or *gfp*. Goodness of fit for *β*-galactosidase with $R^2 = 0.34$. (D) DNA copy numbers of transposon locations determined by qPCR. Error bars indicate standard deviations of three biological replicates. Goodness of fit with $R^2 = 0.42$.

inserted into the *B. subtilis* genome by the Mariner transposon. In light of the broad host-range of the Mariner transposon, such

strategy is especially useful for production organisms for which no convenient genome recombination system is available.

METHODS

Bacterial Strains and Growth Conditions. For cloning purposes *E. coli* DH5α was used. The GFP and β-galactosidase experiments were performed in *B. subtilis* strain W168 (*lacA*::*tet aprE*::*lacI*, *cat*). Cells were grown in Luria−Bertani (LB) medium with the addition of 10 μg/mL kanamycin for overnight cultures. No antibiotics were added when cells were assayed. Minimal medium was based on Spizizen's minimal medium and consists of 2 g/L $(NH4)_2SO_4$, 14 g/L K_2HPO_4, 6 g/L KH_2PO_4, 1 g/L sodium citrate, 2 g/L $MgSO_4$, 5 g/L glucose, 2 g/L tryptophan, 0.2 g/L casamino acids, and 2.2 g/L ammonium ferric citrate.[25] Precultures were performed by growing cells from the overnight culture in fresh media until they reached exponential phase. Then all cultures were diluted to OD_{600} ~ 0.05 and grown until the samples reached an OD_{600} ~ 0.5 and 3.0.

Plasmid and Strain Construction. A unique cloning site was created in pMarB[15] by linearizing the plasmid with the PCR primer pair oSS330/oSS331 (Supporting Information Table S1). The PCR product was digested with *Mun*I and ligated, resulting in plasmid pSS121. The reporter cassette was assembled using splicing by overlap extension PCR.[26] *gfp* was amplified from pSG1729-*gfp* using oligo primer pair oSS346/oSS258, P_{veg} was amplified from chromosomal DNA of *B. subtilis* W168 with primer pair oSS259/oSS260, and *lacZ* was amplified from pMUTIN4 with primer pair oSS261/oSS262. The P_{spac} promoter was introduced by overhangs in primers oSS260 and oSS261. The three fragments were joined stepwise using oSS346 as the terminal *gfp* primer, and oSS347 as the terminal *lacZ* primer, which also introduced *Bam*HI sites at both ends of the final construct. The components were first assembled into pUC19. The bidirectional transcriptional terminator *ezrA/braB* was PCR amplified from the chromosomal DNA of *B. subtilis* W168 using primer pair oSS344/oSS345 containing *Sac*I restriction sites, and subsequently cloned into the *Sac*I site of pUC19 resulting in pSS116. The *Bam*HI site of pSS116 was used to introduce the P_{veg}-*gfp*/P_{spac}-*lacZ* fragment (pSS117). A bidirectional transcriptional terminator was incorporated into the *Sph*I site of pSS117 by amplifying the *ywoG/ywoF* terminator using primer pair oSS348/oSS349 (pSS118). Both terminators were identified using the WebGeSTer database.[19] *Eco*RI and *Hind*III sites were used to excise the reporter cassette from pSS118 and for cloning into *Mun*I digested pSS121, after the inset and vector were blunt-end following a treatment with the Klenow fragment of DNA polymerase I. The resulting transposon vector was called pSS125 (Supporting Information Figure S6).

Transposon Selection. The transposon procedure was based on that described by Lampe and Le Breton.[14,15] *B. subtilis* W168 was transformed with pSS125 and transformants were selected on erythromycin containing plates at 30 °C. A few colonies were taken and cultured in LB medium for 3−5 h at 30 °C and spread on nutrient agar plates with kanamycin at several dilutions from 10^{-3} to 10^{-5}. The plates were incubated overnight at 50 °C, the nonpermissive temperature for the plasmid, favoring the integration of the P_{veg}-*gfp*/P_{spac}-*lacZ* fragment and the kanamycin-resistance cassette by the plasmid-encoded transposase. Undesired strains that still contained the plasmid were identified *via* the screening of plates containing erythromycin positive clones. Strains which were only resistant to kanamycin were used for further experiments. Transposon insertion efficiencies were determined by comparing the number of colonies with the desired kanamycin resistance, revealing a successful integration of the expression cassette, with the number of colonies having both kanamycin and erythromycin resistance, reflecting the presence of the plasmid (Supporting Information Table S2). Colonies were plated on nutrient agar plates containing 40 μg/mL of X-gal and 10 μM IPTG to select clones by color.

To determine the transposon integration sites, arbitrary PCR was used as described by Knobloch.[27] In short, PCR reactions were performed using GoTaq polymerase (Promega) and primer pairs Arb1/MarB1 or Arb1/MarB2. The thermocycler program began with an initial denaturation at 95 °C for 5 min, 6 cycles at 95 °C for 1 min, 30 °C for 1 min, 72 °C for 1 min, 30 cycles at 95 °C for 30 s, 50 °C for 30 s, 72 °C for 1 min and a final extension at 72 °C for 7 min. In the second PCR step primer pairs Arb2/MarB1 or Arb2/MarB2 used 2 μL of the first reaction as a template. For the second PCR the following program was used: 95 °C for 5 min, 6 cycles at 95 °C for 30 s, 50 °C for 30 s, 72 °C for 30 s, 30 cycles at 95 °C for 30 s, 58 °C for 30 s, 72 °C for 30 s, and a final extension at 72 °C for 7 min. PCR products were sequenced using primer MarB1N or MarB2N.

β-Galactosidase and GFP Assays. To measure β-galactosidase activities, cells were harvested and flash frozen using liquid nitrogen and stored at −80 °C. β-Galactosidase activity assays were performed according to the standard protocol of Miller,[28] whereby cells were first lysed by adding 20 mg/mL lysozyme at room temperature for 30 min.

To determine GFP levels, cells were washed in PBS after sampling. A sample of 0.3 μL was applied to a GeneFrame (AbGene, Surrey, UK) containing 1.25% agarose in PBS, supplemented with 2 μg/mL of DAPI. For microscopic measurements, a Zeiss 200 M microscope was used with the following excitation filters and exposure times: GFP (470/525) for 500 ms; DAPI (350/460) for 1000 ms. Microscopy images were analyzed using in-house software (NucTracer) that uses the DAPI stained nucleoid as identifier for the region of interest to measure GFP fluorescence in the cell. NucTracer is based on the ImageJ plugin ObjectJ that supports graphical vector objects identifying images on a transparent layer.[29,30]

DNA Copy Numbers Analysis. DNA copy number analysis was performed as described by Murray and Koh.[20] In short, chromosomal DNA was isolated during exponential (OD_{600} ~ 0.5) and stationary phase (OD_{600} ~ 3.0). As a control for single copy number, DNA from *B. subtilis* spores was isolated. Regions of origin (*oriC*) and terminus (*ter*) of replication and the transposon locations were amplified using the primers listed in Supporting Information Table S2. qPCR was performed using iQ SYBR green supermix (Bio-Rad) in a C1000 thermal cycler (CFX96 real-time system, Bio-Rad). A relative quantification analysis ($\Delta\Delta C_q$) was performed by determining quantification cycles (C_q) to calculate the ratios of transposon location compared to *ter* and by normalizing to the *ori*:*ter* ratio of chromosomal DNA of spores that contain one chromosome reflecting an equivalent ratio of 1. Two loci were excluded from the analysis (1° and 241°) due to primer inefficiencies during qPCR.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.6b00065.

Table S1: Transposon insertion efficiency of the expression cassette comparing pSS125 to pMarB. Table S2: Primers used in this study. Figure S1: Growth rate measurement of transposons. Figure S2: Overview of the chromosomal transposon insertion sites. Figure S3: Effect of chromosomal location and transcription direction on gene expression. Figure S4: Transcriptional read-through measured by β-galactosidase assays in the absence of IPTG. Figure S5: Expression differences in stationary growth. Figure S6: Plasmid map of pSS125. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: l.w.hamoen@uva.nl.

Author Contributions

‖C.S. and S.S. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

C.S. and R.C. were supported by the European Commission funded Marie Curie Initial Training Network ATRIEM (Project No. 317228). S.S. was supported by a BBSRC DTG PhD studentship. L.C.B was supported by the European Commission funded Marie Curie Innovative Training Network AMBER (Project No. 317338), and L.W.H was supported by an NWO STW-Vici (12128) grant.

■ REFERENCES

(1) Radeck, J., Kraft, K., Bartels, J., Cikovic, T., Durr, F., Emenegger, J., Kelterborn, S., Sauer, C., Fritz, G., Gebhard, S., et al. (2013) The Bacillus BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with Bacillus subtilis. *J. Biol. Eng. 7*, 29.

(2) Niaudet, B., Janniere, L., and Ehrlich, S. D. (1985) Integration of linear, heterologous DNA molecules into the Bacillus subtilis chromosome: mechanism and use in induction of predictable rearrangements. *J. Bacteriol. 163*, 111−120.

(3) Young, M. (1983) The mechanism of insertion of a segment of heterologous DNA into the chromosome of Bacillus subtilis. *Microbiology 129*, 1497−1512.

(4) van Dijl, J. M., and Hecker, M. (2013) Bacillus subtilis: from soil bacterium to super-secreting cell factory. *Microb. Cell Fact. 12*, 3.

(5) Sousa, C., de Lorenzo, V., and Cebolla, A. (1997) Modulation of gene expression through chromosomal positioning in Escherichia coli. *Microbiology 143* (Pt 6), 2071−2078.

(6) Couturier, E., and Rocha, E. P. (2006) Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol. Microbiol. 59*, 1506−1518.

(7) Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., et al. (1997) The complete genome sequence of the gram-positive bacterium Bacillus subtilis. *Nature 390*, 249−256.

(8) McLean, M. J., Wolfe, K. H., and Devine, K. M. (1998) Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. *J. Mol. Evol. 47*, 691−696.

(9) Zeigler, D. R., and Dean, D. H. (1990) Orientation of genes in the Bacillus subtilis chromosome. *Genetics 125*, 703−708.

(10) Mirkin, E. V., and Mirkin, S. M. (2005) Mechanisms of transcription-replication collisions in bacteria. *Molecular and cellular biology 25*, 888−895.

(11) Washburn, R. S., and Gottesman, M. E. (2011) Transcription termination maintains chromosome integrity. *Proc. Natl. Acad. Sci. U. S. A. 108*, 792−797.

(12) Block, D. H., Hussein, R., Liang, L. W., and Lim, H. N. (2012) Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Res. 40*, 8979−8992.

(13) Bryant, J. A., Sellars, L. E., Busby, S. J., and Lee, D. J. (2015) Chromosome position effects on gene expression in Escherichia coli K-12. *Nucleic Acids Res. 42*, 11383−11392.

(14) Lampe, D. J., Churchill, M. E., and Robertson, H. M. (1996) A purified mariner transposase is sufficient to mediate transposition in vitro. *EMBO J. 15*, 5470−5479.

(15) Le Breton, Y., Mohapatra, N. P., and Haldenwang, W. G. (2006) In vivo random mutagenesis of Bacillus subtilis by use of TnYLB-1, a mariner-based transposon. *Applied and environmental microbiology 72*, 327−333.

(16) Choi, K. H., and Kim, K. J. (2009) Applications of transposon-based gene delivery system in bacteria. *J. Microbiol. Biotechnol. 19*, 217−228.

(17) Picardeau, M. (2010) Transposition of fly mariner elements into bacteria as a genetic tool for mutagenesis. *Genetica 138*, 551−558.

(18) Rubin, E. J., Akerley, B. J., Novik, V. N., Lampe, D. J., Husson, R. N., and Mekalanos, J. J. (1999) In vivo transposition of mariner-based elements in enteric bacteria and mycobacteria. *Proc. Natl. Acad. Sci. U. S. A. 96*, 1645−1650.

(19) Mitra, A., Kesarwani, A. K., Pal, D., and Nagaraja, V. (2011) WebGeSTer DB–a transcription terminator database. *Nucleic Acids Res. 39*, D129−135.

(20) Murray, H., and Koh, A. (2014) Multiple Regulatory Systems Coordinate DNA Replication with Cell Growth in Bacillus subtilis. *PLoS Genet. 10*, e1004731.

(21) Morimoto, T., Loh, P. C., Hirai, T., Asai, K., Kobayashi, K., Moriya, S., and Ogasawara, N. (2002) Six GTP-binding proteins of the Era/Obg family are essential for cell growth in Bacillus subtilis. *Microbiology 148*, 3539−3552.

(22) Real, G., and Henriques, A. O. (2006) Localization of the Bacillus subtilis murB gene within the dcw cluster is important for growth and sporulation. *Journal of bacteriology 188*, 1721−1732.

(23) Chai, Y., Norman, T., Kolter, R., and Losick, R. (2011) Evidence that metabolism and chromosome copy number control mutually exclusive cell fates in Bacillus subtilis. *EMBO J. 30*, 1402−1413.

(24) Narula, J., Kuchina, A., Lee, D. Y., Fujita, M., Suel, G. M., and Igoshin, O. A. (2015) Chromosomal Arrangement of Phosphorelay Genes Couples Sporulation and DNA Replication. *Cell 162*, 328−337.

(25) Spizizen, J. (1958) Transformation of Biochemically Deficient Strains of Bacillus Subtilis by Deoxyribonucleate. *Proc. Natl. Acad. Sci. U. S. A. 44*, 1072−1078.

(26) Horton, R. M. (1995) PCR-mediated recombination and mutagenesis. SOEing together tailor-made genes. *Mol. Biotechnol. 3*, 93−99.

(27) Knobloch, J. K., Nedelmann, M., Kiel, K., Bartscht, K., Horstkotte, M. A., Dobinsky, S., Rohde, H., and Mack, D. (2003) Establishment of an arbitrary PCR for rapid identification of Tn917 insertion sites in Staphylococcus epidermidis: characterization of biofilm-negative and nonmucoid mutants. *Applied and environmental microbiology 69*, 5812−5818.

(28) Miller, J. H. (1972) Assay of β-galactosidase, in *Experiments in Molecular Genetics*, 3rd ed., pp 352−355, Cold Spring Harbor Laboratory Press, New York.

(29) Vischer, N. O. E., Huls, P. G., and Woldringh, C. L. (1994) Object-Image: An Interactive Image Analysis Program Using Structured Point Collection. *Binary 6*, 160−166.

(30) Syvertsson, S., Vischer, N. O., Gao, Y., and Hamoen, L. W. (2016) When Phase Contrast Fails: ChainTracer and NucTracer, Two ImageJ Methods for Semi-Automated Single Cell Analysis Using Membrane or DNA Staining. *PLoS One 11*, e0151267.

Bibliography

Ackermann, M. 2015. A functional perspective on phenotypic heterogeneity in microorganisms. *Nat Rev Microbiol,* 13**,** 497-508.

Aiyar, S. E., Gourse, R. L. & Ross, W. 1998. Upstream A-tracts increase bacterial promoter activity through interactions with the RNA polymerase alpha subunit. *Proc Natl Acad Sci U S A,* 95**,** 14652-7.

Akhtar, W., de Jong, J., Pindyurin, A. V., Pagie, L., Meuleman, W., de Ridder, J., Berns, A., Wessels, L. F., van Lohuizen, M. & van Steensel, B. 2013. Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell,* 154**,** 914-27.

Altenbuchner, J. 2016. Editing of the *Bacillus subtilis* Genome by the CRISPR-Cas9 System. *Appl Environ Microbiol,* 82**,** 5421-7.

Anagnostopoulos, C. & Spizizen, J. 1961. Requirements for Transformation in *Bacillus subtilis*. *J Bacteriol,* 81**,** 741-6.

Anetzberger, C., Schell, U. & Jung, K. 2012. Single cell analysis of *Vibrio harveyi* uncovers functional heterogeneity in response to quorum sensing signals. *BMC Microbiol,* 12**,** 209.

Avery, S. V. 2006. Microbial cell individuality and the underlying sources of heterogeneity. *Nat Rev Microbiol,* 4**,** 577-87.

Balazsi, G., van Oudenaarden, A. & Collins, J. J. 2011. Cellular decision making and biological noise: from microbes to mammals. *Cell,* 144**,** 910-25.

Banse, A. V., Chastanet, A., Rahn-Lee, L., Hobbs, E. C. & Losick, R. 2008. Parallel pathways of repression and antirepression governing the transition to stationary phase in *Bacillus subtilis*. *Proc Natl Acad Sci U S A,* 105**,** 15547-52.

Barbe, V., Cruveiller, S., Kunst, F., Lenoble, P., Meurice, G., Sekowska, A., Vallenet, D., Wang, T., Moszer, I., Medigue, C. & Danchin, A. 2009. From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology,* 155**,** 1758-75.

Barkai, N. & Leibler, S. 2000. Circadian clocks limited by noise. *Nature,* 403**,** 267-8.

Bergara, F., Ibarra, C., Iwamasa, J., Patarroyo, J. C., Aguilera, R. & Marquez-Magana, L. M. 2003. CodY is a nutritional repressor of flagellar gene expression in *Bacillus subtilis*. *J Bacteriol,* 185**,** 3118-26.

Bernier, R., Desrochers, M., Jurasek, L. & Paice, M. G. 1983. Isolation and Characterization of a Xylanase from *Bacillus subtilis*. *Appl Environ Microbiol,* 46**,** 511-4.

Bisicchia, P., Lioliou, E., Noone, D., Salzberg, L. I., Botella, E., Hubner, S. & Devine, K. M. 2010. Peptidoglycan metabolism is controlled by the WalRK (YycFG) and PhoPR two-component systems in phosphate-limited *Bacillus subtilis* cells. *Mol Microbiol,* 75**,** 972-89.

Blake, W. J., Balazsi, G., Kohanski, M. A., Isaacs, F. J., Murphy, K. F., Kuang, Y., Cantor, C. R., Walt, D. R. & Collins, J. J. 2006. Phenotypic consequences of promoter-mediated transcriptional noise. *Mol Cell,* 24**,** 853-65.

Blakely, G., May, G., McCulloch, R., Arciszewska, L. K., Burke, M., Lovett, S. T. & Sherratt, D. J. 1993. Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. *Cell,* 75**,** 351-61.

Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. & Shao, Y. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science,* 277**,** 1453-62.

Block, D. H., Hussein, R., Liang, L. W. & Lim, H. N. 2012. Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Res,* 40**,** 8979-92.

Bolhuis, A., Tjalsma, H., Smith, H. E., de Jong, A., Meima, R., Venema, G., Bron, S. & van Dijl, J. M. 1999. Evaluation of bottlenecks in the late stages of protein secretion in *Bacillus subtilis*. *Appl Environ Microbiol,* 65**,** 2934-41.

Branda, S. S., Chu, F., Kearns, D. B., Losick, R. & Kolter, R. 2006. A major protein component of the *Bacillus subtilis* biofilm matrix. *Mol Microbiol,* 59**,** 1229-38.

Branda, S. S., Gonzalez-Pastor, J. E., Ben-Yehuda, S., Losick, R. & Kolter, R. 2001. Fruiting body formation by *Bacillus subtilis*. *Proc Natl Acad Sci U S A,* 98**,** 11621-6.

Brockmeier, U., Caspers, M., Freudl, R., Jockwer, A., Noll, T. & Eggert, T. 2006. Systematic screening of all signal peptides from *Bacillus subtilis*: a powerful strategy in optimizing heterologous protein secretion in Gram-positive bacteria. *J Mol Biol,* 362**,** 393-402.

Browning, D. F. & Busby, S. J. 2004. The regulation of bacterial transcription initiation. *Nat Rev Microbiol,* 2**,** 57-65.

Bryant, J. A., Sellars, L. E., Busby, S. J. & Lee, D. J. 2015. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res,* 42**,** 11383-92.

Bussiere, D. E. & Bastia, D. 1999. Termination of DNA replication of bacterial and plasmid chromosomes. *Mol Microbiol,* 31**,** 1611-8.

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J. & Wittwer, C. T. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem,* 55**,** 611-22.

Caramori, T., Barilla, D., Nessi, C., Sacchi, L. & Galizzi, A. 1996. Role of FlgM in sigma D-dependent gene expression in *Bacillus subtilis*. *J Bacteriol,* 178**,** 3113-8.

Caramori, T. & Galizzi, A. 1998. The UP element of the promoter for the flagellin gene, *hag*, stimulates transcription from both SigD- and SigA-dependent promoters in *Bacillus subtilis*. *Mol Gen Genet,* 258**,** 385-8.

Casadaban, M. J. & Cohen, S. N. 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J Mol Biol,* 138**,** 179-207.

Casjens, S. 1998. The diverse and dynamic structure of bacterial genomes. *Annu Rev Genet,* 32**,** 339-77.

Chai, Y., Chu, F., Kolter, R. & Losick, R. 2008. Bistability and biofilm formation in *Bacillus subtilis*. *Mol Microbiol,* 67**,** 254-63.

Chai, Y., Norman, T., Kolter, R. & Losick, R. 2010. An epigenetic switch governing daughter cell separation in *Bacillus subtilis*. *Genes Dev,* 24**,** 754-65.

Chan, J. M., Guttenplan, S. B. & Kearns, D. B. 2014. Defects in the flagellar motor increase synthesis of poly-gamma-glutamate in *Bacillus subtilis*. *J Bacteriol,* 196**,** 740-53.

Chen, I. & Dubnau, D. 2004. DNA uptake during bacterial transformation. *Nat Rev Microbiol,* 2**,** 241-9.

Chen, R., Guttenplan, S. B., Blair, K. M. & Kearns, D. B. 2009. Role of the sigmaD-dependent autolysins in *Bacillus subtilis* population heterogeneity. *J Bacteriol,* 191**,** 5775-84.

Choi, K. H. & Kim, K. J. 2009. Applications of transposon-based gene delivery system in bacteria. *J Microbiol Biotechnol,* 19**,** 217-28.

Chung, J. D., Stephanopoulos, G., Ireton, K. & Grossman, A. D. 1994. Gene expression in single cells of *Bacillus subtilis*: evidence that a threshold mechanism controls the initiation of sporulation. *J Bacteriol,* 176**,** 1977-84.

Claverys, J. P. & Havarstein, L. S. 2007. Cannibalism and fratricide: mechanisms and raisons d'etre. *Nat Rev Microbiol,* 5**,** 219-29.

Couturier, E. & Rocha, E. P. 2006. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol Microbiol,* 59**,** 1506-18.

Cozy, L. M. & Kearns, D. B. 2010. Gene position in a long operon governs motility development in *Bacillus subtilis*. *Mol Microbiol,* 76**,** 273-85.

Dagert, M. & Ehrlich, S. D. 1979. Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. *Gene,* 6**,** 23-8.

Darmon, E., Noone, D., Masson, A., Bron, S., Kuipers, O. P., Devine, K. M. & van Dijl, J. M. 2002. A novel class of heat and secretion stress-responsive genes is controlled by the autoregulated CssRS two-component system of *Bacillus subtilis*. *J Bacteriol,* 184**,** 5661-71.

Dauter, Z., Dauter, M., Brzozowski, A. M., Christensen, S., Borchert, T. V., Beier, L., Wilson, K. S. & Davies, G. J. 1999. X-ray structure of Novamyl, the five-domain "maltogenic" alpha-amylase from *Bacillus stearothermophilus*: maltose and acarbose complexes at 1.7A resolution. *Biochemistry,* 38**,** 8385-92.

Davis, J. H., Rubin, A. J. & Sauer, R. T. 2011. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res,* 39**,** 1131-41.

de Jong, I. G., Veening, J. W. & Kuipers, O. P. 2012. Single cell analysis of gene expression patterns during carbon starvation in *Bacillus subtilis* reveals large phenotypic variation. *Environ Microbiol,* 14**,** 3110-21.

de Smit, M. H. & van Duin, J. 1994. Translational initiation on structured messengers. Another role for the Shine-Dalgarno interaction. *J Mol Biol,* 235**,** 173-84.

Diderichsen, B. & Christiansen, L. 1988. Cloning of a maltogenic alpha-amylase from *Bacillus stearothermophilus*. *FEMS Microbiology Letters,* 56**,** 53-60.

Diethmaier, C., Pietack, N., Gunka, K., Wrede, C., Lehnik-Habrink, M., Herzberg, C., Hubner, S. & Stulke, J. 2011. A novel factor controlling bistability in *Bacillus subtilis*: the YmdB protein affects flagellin expression and biofilm formation. *J Bacteriol,* 193**,** 5997-6007.

Dini-Andreote, F., Andreote, F. D., Araujo, W. L., Trevors, J. T. & van Elsas, J. D. 2012. Bacterial genomes: habitat specificity and uncharted organisms. *Microb Ecol,* 64**,** 1-7.

Driks, A. 2002. Overview: Development in bacteria: spore formation in *Bacillus subtilis*. *Cell Mol Life Sci,* 59**,** 389-91.

Dubnau, D. 1991a. Genetic competence in *Bacillus subtilis*. *Microbiol Rev,* 55**,** 395-424.

Dubnau, D. 1991b. The regulation of genetic competence in *Bacillus subtilis*. *Mol Microbiol,* 5**,** 11-8.

Dubnau, D. & Losick, R. 2006. Bistability in bacteria. *Mol Microbiol,* 61**,** 564-72.

Ellermeier, C. D., Hobbs, E. C., Gonzalez-Pastor, J. E. & Losick, R. 2006. A three-protein signaling pathway governing immunity to a bacterial cannibalism toxin. *Cell,* 124**,** 549-59.

Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. 2002. Stochastic gene expression in a single cell. *Science,* 297**,** 1183-6.

Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. 2009. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One,* 4**,** e5553.

Engler, C., Kandzia, R. & Marillonnet, S. 2008. A one pot, one step, precision cloning method with high throughput capability. *PLoS One,* 3**,** e3647.

Errington, J. 2003. Regulation of endospore formation in *Bacillus subtilis*. *Nat Rev Microbiol,* 1**,** 117-26.

Eswaramoorthy, P., Duan, D., Dinh, J., Dravis, A., Devi, S. N. & Fujita, M. 2010. The threshold level of the sensor histidine kinase KinA governs entry into sporulation in *Bacillus subtilis*. *J Bacteriol,* 192**,** 3870-82.

Evans, L. D., Poulter, S., Terentjev, E. M., Hughes, C. & Fraser, G. M. 2013. A chain mechanism for flagellum growth. *Nature,* 504**,** 287-90.

Fabret, C. & Hoch, J. A. 1998. A two-component signal transduction system essential for growth of *Bacillus subtilis*: implications for anti-infective therapy. *J Bacteriol,* 180**,** 6375-83.

Feucht, A. & Lewis, P. J. 2001. Improved plasmid vectors for the production of multiple fluorescent protein fusions in *Bacillus subtilis*. *Gene,* 264**,** 289-97.

Freese, E. 1972. Sporulation of bacilli, a model of cellular differentiation. *Curr Top Dev Biol,* 7**,** 85-124.

Fujita, M., Gonzalez-Pastor, J. E. & Losick, R. 2005. High- and low-threshold genes in the Spo0A regulon of *Bacillus subtilis*. *J Bacteriol,* 187**,** 1357-68.

Gamba, P., Jonker, M. J. & Hamoen, L. W. 2015. A Novel Feedback Loop That Controls Bimodal Expression of Genetic Competence. *PLoS Genet,* 11**,** e1005047.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd & Smith, H. O. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods,* 6**,** 343-5.

Goesaert, H., Leman, P., Bijttebier, A. & Delcour, J. A. 2009. Antifirming effects of starch degrading enzymes in bread crumb. *J Agric Food Chem,* 57**,** 2346-55.

Gonzalez-Pastor, J. E., Hobbs, E. C. & Losick, R. 2003. Cannibalism by sporulating bacteria. *Science,* 301**,** 510-3.

Goodman, D. B., Church, G. M. & Kosuri, S. 2013. Causes and effects of N-terminal codon bias in bacterial genes. *Science,* 342**,** 475-9.

Gorochowski, T. E., Avcilar-Kucukgoze, I., Bovenberg, R. A., Roubos, J. A. & Ignatova, Z. 2016. A Minimal Model of Ribosome Allocation Dynamics Captures Trade-offs in Expression between Endogenous and Synthetic Genes. *ACS Synth Biol,* 5**,** 710-20.

Grupta, R., Gigras, P., Mohapatra, H., Goswami, V. K. & Chauhan, B. 2003. Microbial α-amylase: a biotechnological perspective. *Process. Biochem.,* 38**,** 1599-1616.

Guttenplan, S. B. & Kearns, D. B. 2013. Regulation of flagellar motility during biofilm formation. *FEMS Microbiol Rev,* 37**,** 849-71.

Guttenplan, S. B., Shaw, S. & Kearns, D. B. 2013. The cell biology of peritrichous flagella in *Bacillus subtilis*. *Mol Microbiol,* 87**,** 211-29.

Hahn, J., Roggiani, M. & Dubnau, D. 1995. The major role of Spo0A in genetic competence is to downregulate *abrB*, an essential competence gene. *J Bacteriol,* 177**,** 3601-5.

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

Hall-Stoodley, L., Costerton, J. W. & Stoodley, P. 2004. Bacterial biofilms: from the natural environment to infectious diseases. *Nat Rev Microbiol,* 2**,** 95-108.

Hamoen, L. W., Eshuis, H., Jongbloed, J., Venema, G. & van Sinderen, D. 1995. A small gene, designated comS, located within the coding region of the fourth amino acid-activation domain of *srfA*, is required for competence development in *Bacillus subtilis*. *Mol Microbiol,* 15**,** 55-63.

Hamoen, L. W., Venema, G. & Kuipers, O. P. 2003. Controlling competence in *Bacillus subtilis*: shared use of regulators. *Microbiology,* 149**,** 9-17.

Hamon, M. A. & Lazazzera, B. A. 2001. The sporulation transcription factor Spo0A is required for biofilm development in *Bacillus subtilis*. *Mol Microbiol,* 42**,** 1199-209.

Harwood, C. R. 1992. *Bacillus subtilis* and its relatives: molecular biological and industrial workhorses. *Trends Biotechnol,* 10**,** 247-56.

Harwood, C. R. & Cranenburgh, R. 2008. Bacillus protein secretion: an unfolding story. *Trends Microbiol,* 16**,** 73-9.

Hecker, M., Pane-Farre, J. & Volker, U. 2007. SigB-dependent general stress response in *Bacillus subtilis* and related gram-positive bacteria. *Annu Rev Microbiol,* 61**,** 215-36.

Helmann, J. D. 1995. Compilation and analysis of *Bacillus subtilis* sigma A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. *Nucleic Acids Res,* 23**,** 2351-60.

Higgins, D. & Dworkin, J. 2012. Recent progress in *Bacillus subtilis* sporulation. *FEMS Microbiol Rev,* 36**,** 131-48.

Howley, P. M., Israel, M. A., Law, M. F. & Martin, M. A. 1979. A rapid method for detecting and mapping homology between heterologous DNAs. Evaluation of polyomavirus genomes. *J Biol Chem,* 254**,** 4876-83.

Huh, D. & Paulsson, J. 2011. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat Genet,* 43**,** 95-100.

Hyyrylainen, H. L., Bolhuis, A., Darmon, E., Muukkonen, L., Koski, P., Vitikainen, M., Sarvas, M., Pragai, Z., Bron, S., van Dijl, J. M. & Kontinen, V. P. 2001. A novel two-component regulatory system in *Bacillus subtilis* for the survival of severe secretion stress. *Mol Microbiol,* 41**,** 1159-72.

Ireton, K., Rudner, D. Z., Siranosian, K. J. & Grossman, A. D. 1993. Integration of multiple developmental signals in *Bacillus subtilis* through the Spo0A transcription factor. *Genes Dev,* 7**,** 283-94.

Ito, M., Hicks, D. B., Henkin, T. M., Guffanti, A. A., Powers, B. D., Zvi, L., Uematsu, K. & Krulwich, T. A. 2004. MotPS is the stator-force generator for motility of alkaliphilic *Bacillus*, and its homologue is a second functional Mot in *Bacillus subtilis*. *Mol Microbiol,* 53**,** 1035-49.

Jarmer, H., Larsen, T. S., Krogh, A., Saxild, H. H., Brunak, S. & Knudsen, S. 2001. Sigma A recognition sites in the *Bacillus subtilis* genome. *Microbiology,* 147**,** 2417-24.

Jensen, C. L., Stephenson, K., Jorgensen, S. T. & Harwood, C. 2000. Cell-associated degradation affects the yield of secreted engineered and heterologous proteins in the *Bacillus subtilis* expression system. *Microbiology,* 146 (Pt 10)**,** 2583-94.

Jensen, P. R. & Hammer, K. 1998. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Appl Environ Microbiol,* 64**,** 82-7.

Julkowska, D., Obuchowski, M., Holland, I. B. & Seror, S. J. 2005. Comparative analysis of the development of swarming communities of *Bacillus subtilis* 168 and a natural wild type: critical effects of surfactin and the composition of the medium. *J Bacteriol,* 187**,** 65-76.

Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., LeProust, E. M., Hughes, T. R., Lieb, J. D., Widom, J. & Segal, E. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature,* 458**,** 362-6.

Kearns, D. B., Chu, F., Branda, S. S., Kolter, R. & Losick, R. 2005. A master regulator for biofilm formation by *Bacillus subtilis*. *Mol Microbiol,* 55**,** 739-49.

Kearns, D. B. & Losick, R. 2003. Swarming motility in undomesticated *Bacillus subtilis*. *Mol Microbiol,* 49**,** 581-90.

Kearns, D. B. & Losick, R. 2005. Cell population heterogeneity during growth of *Bacillus subtilis*. *Genes Dev,* 19**,** 3083-94.

Kinsinger, R. F., Shirk, M. C. & Fall, R. 2003. Rapid surface motility in *Bacillus subtilis* is dependent on extracellular surfactin and potassium ion. *J Bacteriol,* 185**,** 5627-31.

Knobloch, J. K., Nedelmann, M., Kiel, K., Bartscht, K., Horstkotte, M. A., Dobinsky, S., Rohde, H. & Mack, D. 2003. Establishment of an arbitrary PCR for rapid identification of Tn917 insertion sites in Staphylococcus epidermidis: characterization of biofilm-negative and nonmucoid mutants. *Appl Environ Microbiol,* 69**,** 5812-8.

Kolodkin-Gal, I., Elsholz, A. K., Muth, C., Girguis, P. R., Kolter, R. & Losick, R. 2013. Respiration control of multicellularity in *Bacillus subtilis* by a complex of the

cytochrome chain with a membrane-embedded histidine kinase. *Genes Dev,* 27**,** 887-99.

Konstantinidis, K. T. & Tiedje, J. M. 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc Natl Acad Sci U S A,* 101**,** 3160-5.

Kosuri, S., Goodman, D. B., Cambray, G., Mutalik, V. K., Gao, Y., Arkin, A. P., Endy, D. & Church, G. M. 2013. Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc Natl Acad Sci U S A,* 110**,** 14024-9.

Krishnappa, L., Dreisbach, A., Otto, A., Goosens, V. J., Cranenburgh, R. M., Harwood, C. R., Becher, D. & van Dijl, J. M. 2013. Extracytoplasmic proteases determining the cleavage and release of secreted proteins, lipoproteins, and membrane proteins in *Bacillus subtilis*. *J Proteome Res,* 12**,** 4101-10.

Krishnappa, L., Monteferrante, C. G., Neef, J., Dreisbach, A. & van Dijl, J. M. 2014. Degradation of extracytoplasmic catalysts for protein folding in *Bacillus subtilis*. *Appl Environ Microbiol,* 80**,** 1463-8.

Kulkarni, N., Shendye, A. & Rao, M. 1999. Molecular and biotechnological aspects of xylanases. *FEMS Microbiol Rev,* 23**,** 411-56.

Kunst, F., Msadek, T., Bignon, J. & Rapoport, G. 1994. The DegS/DegU and ComP/ComA two-component systems are part of a network controlling degradative enzyme synthesis and competence in *Bacillus subtilis*. *Res Microbiol,* 145**,** 393-402.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Cordani, J. J., Connerton, I. F., Cummings, N. J., Daniel, R. A., Denziot, F., Devine, K. M., Dusterhoft, A., Ehrlich, S. D., Emmerson, P. T., Entian, K. D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S. Y., Glaser, P., Goffeau, A., Golightly, E. J., Grandi, G., Guiseppi, G., Guy, B. J., Haga, K., Haiech, J., Harwood, C. R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M. F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S. M., Levine, A.,

Liu, H., Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R. P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S. H., Parro, V., Pohl, T. M., Portelle, D., Porwollik, S., Prescott, A. M., Presecan, E., Pujic, P., Purnelle, B., et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature,* 390**,** 249-56.

Lam, K. H., Chow, K. C. & Wong, W. K. 1998. Construction of an efficient *Bacillus subtilis* system for extracellular production of heterologous proteins. *J Biotechnol,* 63**,** 167-77.

Lampe, D. J., Churchill, M. E. & Robertson, H. M. 1996. A purified mariner transposase is sufficient to mediate transposition *in vitro*. *EMBO J,* 15**,** 5470-9.

Le Breton, Y., Mohapatra, N. P. & Haldenwang, W. G. 2006. *In vivo* random mutagenesis of *Bacillus subtilis* by use of TnYLB-1, a mariner-based transposon. *Appl Environ Microbiol,* 72**,** 327-33.

Le, T. B., Imakaev, M. V., Mirny, L. A. & Laub, M. T. 2013. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science,* 342**,** 731-4.

Lenz, P. & Sogaard-Andersen, L. 2011. Temporal and spatial oscillations in bacteria. *Nat Rev Microbiol,* 9**,** 565-77.

Leonhardt, H. 1990. Identification of a low-copy-number mutation within the pUB110 replicon and its effect on plasmid stability in *Bacillus subtilis*. *Gene,* 94**,** 121-4.

Levine, J. H., Lin, Y. & Elowitz, M. B. 2013. Functional roles of pulsing in genetic circuits. *Science,* 342**,** 1193-200.

Levy, S. F., Ziv, N. & Siegal, M. L. 2012. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. *PLoS Biol,* 10**,** e1001325.

Liebeton, K., Lengefeld, J. & Eck, J. 2014. The nucleotide composition of the spacer sequence influences the expression yield of heterologously expressed genes in *Bacillus subtilis*. *J Biotechnol,* 191**,** 214-20.

Lienert, F., Lohmueller, J. J., Garg, A. & Silver, P. A. 2014. Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nat Rev Mol Cell Biol,* 15**,** 95-107.

Liu, M., Tolstorukov, M., Zhurkin, V., Garges, S. & Adhya, S. 2004. A mutant spacer sequence between -35 and -10 elements makes the P_{lac} promoter hyperactive and cAMP receptor protein-independent. *Proc Natl Acad Sci U S A,* 101**,** 6911-6.

Lopez, D. & Kolter, R. 2010. Extracellular signals that define distinct and coexisting cell fates in *Bacillus subtilis*. *FEMS Microbiol Rev,* 34**,** 134-49.

Lopez, D., Vlamakis, H. & Kolter, R. 2009a. Generation of multiple cell types in *Bacillus subtilis*. *FEMS Microbiol Rev,* 33**,** 152-63.

Lopez, D., Vlamakis, H., Losick, R. & Kolter, R. 2009b. Cannibalism enhances biofilm development in *Bacillus subtilis*. *Mol Microbiol,* 74**,** 609-18.

Lopez, D., Vlamakis, H., Losick, R. & Kolter, R. 2009c. Paracrine signaling in a bacterium. *Genes Dev,* 23**,** 1631-8.

Maamar, H. & Dubnau, D. 2005. Bistability in the *Bacillus subtilis* K-state (competence) system requires a positive feedback loop. *Mol Microbiol,* 56**,** 615-24.

Magnuson, R., Solomon, J. & Grossman, A. D. 1994. Biochemical and genetic characterization of a competence pheromone from *B. subtilis*. *Cell,* 77**,** 207-16.

Manabe, K., Kageyama, Y., Morimoto, T., Ozawa, T., Sawada, K., Endo, K., Tohata, M., Ara, K., Ozaki, K. & Ogasawara, N. 2011. Combined effect of improved cell yield and increased specific productivity enhances recombinant enzyme production in genome-reduced *Bacillus subtilis* strain MGB874. *Appl Environ Microbiol,* 77**,** 8370-81.

Margot, P. & Karamata, D. 1996. The *wprA* gene of *Bacillus subtilis* 168, expressed during exponential growth, encodes a cell-wall-associated protease. *Microbiology,* 142 (Pt 12)**,** 3437-44.

McLean, M. J., Wolfe, K. H. & Devine, K. M. 1998. Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. *J Mol Evol,* 47**,** 691-6.

Meijer, W. J. & Salas, M. 2004. Relevance of UP elements for three strong *Bacillus subtilis* phage phi29 promoters. *Nucleic Acids Res,* 32**,** 1166-76.

Mijakovic, I., Petranovic, D. & Jensen, P. R. 2005. Tunable promoters in systems biology. *Curr Opin Biotechnol,* 16**,** 329-35.

Miller, J. H. 1972. Assay of β-galactosidase. *Experiments in molecular genetics.* 3rd ed. New York: Cold Spring Harbor Laboratory Press.

Mirkin, E. V. & Mirkin, S. M. 2005. Mechanisms of transcription-replication collisions in bacteria. *Mol Cell Biol,* 25**,** 888-95.

Mirouze, N., Desai, Y., Raj, A. & Dubnau, D. 2012. Spo0A~P imposes a temporal gate for the bimodal expression of competence in *Bacillus subtilis*. *PLoS Genet,* 8**,** e1002586.

Mitra, A., Kesarwani, A. K., Pal, D. & Nagaraja, V. 2011. WebGeSTer DB--a transcription terminator database. *Nucleic Acids Res,* 39**,** D129-35.

Molle, V., Fujita, M., Jensen, S. T., Eichenberger, P., Gonzalez-Pastor, J. E., Liu, J. S. & Losick, R. 2003. The Spo0A regulon of *Bacillus subtilis*. *Mol Microbiol,* 50**,** 1683-701.

Moran, C. P., Jr., Lang, N., LeGrice, S. F., Lee, G., Stephens, M., Sonenshein, A. L., Pero, J. & Losick, R. 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *Mol Gen Genet,* 186**,** 339-46.

Morimoto, T., Kadoya, R., Endo, K., Tohata, M., Sawada, K., Liu, S., Ozawa, T., Kodama, T., Kakeshita, H., Kageyama, Y., Manabe, K., Kanaya, S., Ara, K., Ozaki, K. & Ogasawara, N. 2008. Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. *DNA Res,* 15**,** 73-81.

Morimoto, T., Loh, P. C., Hirai, T., Asai, K., Kobayashi, K., Moriya, S. & Ogasawara, N. 2002. Six GTP-binding proteins of the Era/Obg family are essential for cell growth in *Bacillus subtilis*. *Microbiology,* 148**,** 3539-52.

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

Mukherjee, S. & Kearns, D. B. 2014. The structure and regulation of flagella in *Bacillus subtilis*. *Annu Rev Genet,* 48**,** 319-40.

Murray, H. & Koh, A. 2014. Multiple regulatory systems coordinate DNA replication with cell growth in *Bacillus subtilis*. *PLoS Genet,* 10**,** e1004731.

Nakano, M. M., Corbell, N., Besson, J. & Zuber, P. 1992. Isolation and characterization of *sfp*: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in *Bacillus subtilis*. *Mol Gen Genet,* 232**,** 313-21.

Nakano, M. M., Xia, L. A. & Zuber, P. 1991. Transcription initiation region of the *srfA* operon, which is controlled by the *comP-comA* signal transduction system in *Bacillus subtilis*. *J Bacteriol,* 173**,** 5487-93.

Newman, J. A. & Lewis, R. J. 2013. Exploring the role of SlrR and SlrA in the SinR epigenetic switch. *Commun Integr Biol,* 6**,** e25658.

Nicolas, P., Mader, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M., Aymerich, S., Becher, D., Bisicchia, P.,

Botella, E., Delumeau, O., Doherty, G., Denham, E. L., Fogg, M. J., Fromion, V., Goelzer, A., Hansen, A., Hartig, E., Harwood, C. R., Homuth, G., Jarmer, H., Jules, M., Klipp, E., Le Chat, L., Lecointe, F., Lewis, P., Liebermeister, W., March, A., Mars, R. A., Nannapaneni, P., Noone, D., Pohl, S., Rinn, B., Rugheimer, F., Sappa, P. K., Samson, F., Schaffer, M., Schwikowski, B., Steil, L., Stulke, J., Wiegert, T., Devine, K. M., Wilkinson, A. J., van Dijl, J. M., Hecker, M., Volker, U., Bessieres, P. & Noirot, P. 2012. Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science,* 335**,** 1103-6.

Nishihara, T. & Freese, E. 1975. Motility of *Bacillus subtilis* during growth and sporulation. *J Bacteriol,* 123**,** 366-71.

Noone, D., Botella, E., Butler, C., Hansen, A., Jende, I. & Devine, K. M. 2012. Signal perception by the secretion stress-responsive CssRS two-component system in *Bacillus subtilis*. *J Bacteriol,* 194**,** 1800-14.

Noone, D., Howell, A., Collery, R. & Devine, K. M. 2001. YkdA and YvtA, HtrA-like serine proteases in *Bacillus subtilis*, engage in negative autoregulation and reciprocal cross-regulation of *ykdA* and *yvtA* gene expression. *J Bacteriol,* 183**,** 654-63.

Noone, D., Howell, A. & Devine, K. M. 2000. Expression of *ykdA*, encoding a *Bacillus subtilis* homologue of HtrA, is heat shock inducible and negatively autoregulated. *J Bacteriol,* 182**,** 1592-9.

Ohlsen, K. L., Grimsley, J. K. & Hoch, J. A. 1994. Deactivation of the sporulation transcription factor Spo0A by the Spo0E protein phosphatase. *Proc Natl Acad Sci U S A,* 91**,** 1756-60.

Osada, Y., Saito, R. & Tomita, M. 1999. Analysis of base-pairing potentials between 16S rRNA and 5' UTR for translation initiation in various prokaryotes. *Bioinformatics,* 15**,** 578-81.

Osterberg, S., del Peso-Santos, T. & Shingler, V. 2011. Regulation of alternative sigma factor use. *Annu Rev Microbiol,* 65**,** 37-55.

Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. 2002. Regulation of noise in the expression of a single gene. *Nat Genet,* 31**,** 69-73.

Patrick, J. E. & Kearns, D. B. 2009. Laboratory strains of *Bacillus subtilis* do not exhibit swarming motility. *J Bacteriol,* 191**,** 7129-33.

Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. 2006. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol,* 24**,** 79-88.

Perego, M. 2001. A new family of aspartyl phosphate phosphatases targeting the sporulation transcription factor Spo0A of *Bacillus subtilis*. *Mol Microbiol,* 42**,** 133-43.

Perez-Rueda, E. & Collado-Vides, J. 2000. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res,* 28**,** 1838-47.

Peters, J. M., Colavin, A., Shi, H., Czarny, T. L., Larson, M. H., Wong, S., Hawkins, J. S., Lu, C. H., Koo, B. M., Marta, E., Shiver, A. L., Whitehead, E. H., Weissman, J. S., Brown, E. D., Qi, L. S., Huang, K. C. & Gross, C. A. 2016. A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell,* 165**,** 1493-506.

Pfaffl, M. W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res,* 29**,** e45.

Phan, T. T., Nguyen, H. D. & Schumann, W. 2012. Development of a strong intracellular expression system for *Bacillus subtilis* by optimizing promoter elements. *J Biotechnol,* 157**,** 167-72.

Picardeau, M. 2010. Transposition of fly mariner elements into bacteria as a genetic tool for mutagenesis. *Genetica,* 138**,** 551-8.

Piersma, S., Denham, E. L., Drulhe, S., Tonk, R. H., Schwikowski, B. & van Dijl, J. M. 2013. TLM-Quant: an open-source pipeline for visualization and quantification of gene expression heterogeneity in growing microbial cells. *PLoS One,* 8**,** e68696.

Piggot, P. J. & Hilbert, D. W. 2004. Sporulation of *Bacillus subtilis*. *Curr Opin Microbiol,* 7**,** 579-86.

Ploss, T. N., Reilman, E., Monteferrante, C. G., Denham, E. L., Piersma, S., Lingner, A., Vehmaanpera, J., Lorenz, P. & van Dijl, J. M. 2016. Homogeneity and heterogeneity in amylase production by *Bacillus subtilis* under different growth conditions. *Microb Cell Fact,* 15**,** 57.

Pohl, S., Bhavsar, G., Hulme, J., Bloor, A. E., Misirli, G., Leckenby, M. W., Radford, D. S., Smith, W., Wipat, A., Williamson, E. D., Harwood, C. R. & Cranenburgh, R. M. 2013. Proteomic analysis of *Bacillus subtilis* strains engineered for improved production of heterologous proteins. *Proteomics,* 13**,** 3298-308.

Pohl, S. & Harwood, C. R. 2010. Heterologous protein secretion by *bacillus* species from the cradle to the grave. *Adv Appl Microbiol,* 73**,** 1-25.

Pollet, A., Vandermarliere, E., Lammertyn, J., Strelkov, S. V., Delcour, J. A. & Courtin, C. M. 2009. Crystallographic and activity-based evidence for thumb flexibility and its relevance in glycoside hydrolase family 11 xylanases. *Proteins,* 77**,** 395-403.

Radeck, J., Kraft, K., Bartels, J., Cikovic, T., Durr, F., Emenegger, J., Kelterborn, S., Sauer, C., Fritz, G., Gebhard, S. & Mascher, T. 2013. The *Bacillus* BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with *Bacillus subtilis*. *J Biol Eng,* 7**,** 29.

Rao, C. V., Glekas, G. D. & Ordal, G. W. 2008. The three adaptation systems of *Bacillus subtilis* chemotaxis. *Trends Microbiol,* 16**,** 480-7.

Rocha, E. P. 2004. The replication-related organization of bacterial genomes. *Microbiology,* 150**,** 1609-27.

Rocha, E. P., Danchin, A. & Viari, A. 1999. Translation in *Bacillus subtilis*: roles and trends of initiation and termination, insights from a genome analysis. *Nucleic Acids Res,* 27**,** 3567-76.

Ron, E. Z. & Rosenberg, E. 2001. Natural roles of biosurfactants. *Environ Microbiol,* 3**,** 229-36.

Rooke, J. 2013. Synthetic biology as a source of global health innovation. *Syst Synth Biol,* 7**,** 67-72.

Rubin, E. J., Akerley, B. J., Novik, V. N., Lampe, D. J., Husson, R. N. & Mekalanos, J. J. 1999. *In vivo* transposition of mariner-based elements in enteric bacteria and mycobacteria. *Proc Natl Acad Sci U S A,* 96**,** 1645-50.

Ruller, R., Rosa, J. C., Faca, V. M., Greene, L. J. & Ward, R. J. 2006. Efficient constitutive expression of *Bacillus subtilis* xylanase A in *Escherichia coli* DH5alpha under the control of the *Bacillus* BsXA promoter. *Biotechnol Appl Biochem,* 43**,** 9-15.

Salis, H. M., Mirsky, E. A. & Voigt, C. A. 2009. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol,* 27**,** 946-50.

Sarvas, M., Harwood, C. R., Bron, S. & van Dijl, J. M. 2004. Post-translocational folding of secretory proteins in Gram-positive bacteria. *Biochim Biophys Acta,* 1694**,** 311-27.

Sauer, C., Syvertsson, S., Bohorquez, L. C., Cruz, R., Harwood, C. R., van Rij, T. & Hamoen, L. W. 2016. Effect of Genome Position on Heterologous Gene Expression in *Bacillus subtilis*: An Unbiased Analysis. *ACS Synth Biol,* 5**,** 942-7.

Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods,* 9**,** 671-5.

Schultz, D., Wolynes, P. G., Ben Jacob, E. & Onuchic, J. N. 2009. Deciding fate in adverse times: sporulation and competence in *Bacillus subtilis*. *Proc Natl Acad Sci U S A,* 106**,** 21027-34.

Schumann, W. 2007. Production of recombinant proteins in *Bacillus subtilis*. *Adv Appl Microbiol,* 62**,** 137-89.

Sheppard, J. D., Jumarie, C., Cooper, D. G. & Laprade, R. 1991. Ionic channels induced by surfactin in planar lipid bilayer membranes. *Biochim Biophys Acta,* 1064**,** 13-23.

Slomovic, S., Pardee, K. & Collins, J. J. 2015. Synthetic biology devices for *in vitro* and *in vivo* diagnostics. *Proc Natl Acad Sci U S A,* 112**,** 14429-35.

Smits, W. K., Eschevins, C. C., Susanna, K. A., Bron, S., Kuipers, O. P. & Hamoen, L. W. 2005. Stripping *Bacillus*: ComK auto-stimulation is responsible for the bistable response in competence development. *Mol Microbiol,* 56**,** 604-14.

Sonenshein, A. L. 2007. Control of key metabolic intersections in *Bacillus subtilis*. *Nat Rev Microbiol,* 5**,** 917-27.

Sousa, C., de Lorenzo, V. & Cebolla, A. 1997. Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiology,* 143 (Pt 6)**,** 2071-8.

Sprengart, M. L. & Porter, A. G. 1997. Functional importance of RNA interactions in selection of translation initiation codons. *Mol Microbiol,* 24**,** 19-28.

Stein, T. 2005. *Bacillus subtilis* antibiotics: structures, syntheses and specific functions. *Mol Microbiol,* 56**,** 845-57.

Stephenson, K. & Harwood, C. R. 1998. Influence of a cell-wall-associated protease on production of alpha-amylase by *Bacillus subtilis*. *Appl Environ Microbiol,* 64**,** 2875-81.

Stern, A. S. & Berg, H. C. 2013. Single-file diffusion of flagellin in flagellar filaments. *Biophys J,* 105**,** 182-4.

Stewart, C. R., Gaslightwala, I., Hinata, K., Krolikowski, K. A., Needleman, D. S., Peng, A. S., Peterman, M. A., Tobias, A. & Wei, P. 1998. Genes and

regulatory sites of the "host-takeover module" in the terminal redundancy of *Bacillus subtilis* bacteriophage SPO1. *Virology,* 246**,** 329-40.

Suel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. 2006. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature,* 440**,** 545-50.

Suel, G. M., Kulkarni, R. P., Dworkin, J., Garcia-Ojalvo, J. & Elowitz, M. B. 2007. Tunability and noise dependence in differentiation dynamics. *Science,* 315**,** 1716-9.

Syvertsson, S., Vischer, N. O., Gao, Y. & Hamoen, L. W. 2016. When Phase Contrast Fails: ChainTracer and NucTracer, Two ImageJ Methods for Semi-Automated Single Cell Analysis Using Membrane or DNA Staining. *PLoS One,* 11**,** e0151267.

Tanyildizi, M. S., Özer, D. & Elibol, M. 2005. Optimization of α-amylase production by *Bacillus* sp. using response surface methodology. 40**,** 2291-2296.

Tjalsma, H., Bolhuis, A., Jongbloed, J. D., Bron, S. & van Dijl, J. M. 2000. Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol Mol Biol Rev,* 64**,** 515-47.

Torronen, A. & Rouvinen, J. 1997. Structural and functional properties of low molecular weight endo-1,4-beta-xylanases. *J Biotechnol,* 57**,** 137-49.

Trip, H., van der Veek, P. J., Renniers, T. C., Meima, R., Sagt, C. M., Mohrmann, L. & Kuipers, O. P. 2011. A novel screening system for secretion of heterologous proteins in *Bacillus subtilis*. *Microb Biotechnol,* 4**,** 673-82.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. & Leunissen, J. A. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res,* 35**,** W71-4.

van Dijl, J. M. & Hecker, M. 2013. *Bacillus subtilis*: from soil bacterium to super-secreting cell factory. *Microb Cell Fact,* 12**,** 3.

Van Rij, E. T., Renniers, A. C. H. M., Pragai, Z. & Van Peij, N. N. M. E. 2013. *Screening method*. PCT/EP2012/063802.

Veening, J. W., Hamoen, L. W. & Kuipers, O. P. 2005. Phosphatases modulate the bistable sporulation gene expression pattern in *Bacillus subtilis*. *Mol Microbiol,* 56**,** 1481-94.

Veening, J. W., Igoshin, O. A., Eijlander, R. T., Nijland, R., Hamoen, L. W. & Kuipers, O. P. 2008a. Transient heterogeneity in extracellular protease production by *Bacillus subtilis*. *Mol Syst Biol,* 4**,** 184.

Veening, J. W., Smits, W. K., Hamoen, L. W., Jongbloed, J. D. & Kuipers, O. P. 2004. Visualization of differential gene expression by improved cyan fluorescent protein and yellow fluorescent protein production in *Bacillus subtilis*. *Appl Environ Microbiol,* 70**,** 6809-15.

Veening, J. W., Smits, W. K. & Kuipers, O. P. 2008b. Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol,* 62**,** 193-210.

Veening, J. W., Stewart, E. J., Berngruber, T. W., Taddei, F., Kuipers, O. P. & Hamoen, L. W. 2008c. Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc Natl Acad Sci U S A,* 105**,** 4393-8.

Vellanoweth, R. L. & Rabinowitz, J. C. 1992. The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* in vivo. *Mol Microbiol,* 6**,** 1105-14.

Verhamme, D. T., Kiley, T. B. & Stanley-Wall, N. R. 2007. DegU co-ordinates multicellular behaviour exhibited by *Bacillus subtilis*. *Mol Microbiol,* 65**,** 554-68.

Vischer, N. O. E., Huls, P. G. & Woldringh, C. L. 1994. Object-Image: An Interactive Image Analysis Program Using Structured Point Collection. *Binary,* 6**,** 160-166.

Vlamakis, H., Aguilar, C., Losick, R. & Kolter, R. 2008. Control of cell fate by the formation of an architecturally complex bacterial community. *Genes Dev,* 22**,** 945-53.

Vreeland, R. H., Rosenzweig, W. D. & Powers, D. W. 2000. Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. *Nature,* 407**,** 897-900.

Washburn, R. S. & Gottesman, M. E. 2011. Transcription termination maintains chromosome integrity. *Proc Natl Acad Sci U S A,* 108**,** 792-7.

Westbrook, A. W., Moo-Young, M. & Chou, C. P. 2016. Development of a CRISPR-Cas9 Tool Kit for Comprehensive Engineering of *Bacillus subtilis*. *Appl Environ Microbiol,* 82**,** 4876-95.

Westers, H., Darmon, E., Zanen, G., Veening, J. W., Kuipers, O. P., Bron, S., Quax, W. J. & van Dijl, J. M. 2004. The *Bacillus* secretion stress response is an indicator for alpha-amylase production levels. *Lett Appl Microbiol,* 39**,** 65-73.

Westers, H., Westers, L., Darmon, E., van Dijl, J. M., Quax, W. J. & Zanen, G. 2006. The CssRS two-component regulatory system controls a general secretion stress response in *Bacillus subtilis*. *FEBS J,* 273**,** 3816-27.

Westers, L., Westers, H., Zanen, G., Antelmann, H., Hecker, M., Noone, D., Devine, K. M., van Dijl, J. M. & Quax, W. J. 2008. Genetic or chemical protease inhibition causes significant changes in the *Bacillus subtilis* exoproteome. *Proteomics,* 8**,** 2704-13.

Wilson, C., Bellen, H. J. & Gehring, W. J. 1990. Position effects on eukaryotic gene expression. *Annu Rev Cell Biol,* 6**,** 679-714.

Wolf, M., Geczi, A., Simon, O. & Borriss, R. 1995. Genes encoding xylan and beta-glucan hydrolysing enzymes in *Bacillus subtilis*: characterization, mapping and construction of strains deficient in lichenase, cellulase and xylanase. *Microbiology,* 141 (Pt 2)**,** 281-90.

Wosten, M. M. 1998. Eubacterial sigma-factors. *FEMS Microbiol Rev,* 22**,** 127-50.

Wu, S. C., Yeung, J. C., Duan, Y., Ye, R., Szarka, S. J., Habibi, H. R. & Wong, S. L. 2002. Functional production and characterization of a fibrin-specific single-chain antibody fragment from *Bacillus subtilis*: effects of molecular chaperones and a wall-bound protease on antibody fragment production. *Appl Environ Microbiol,* 68**,** 3261-9.

Wu, X. C., Lee, W., Tran, L. & Wong, S. L. 1991. Engineering a *Bacillus subtilis* expression-secretion system with a strain deficient in six extracellular proteases. *J Bacteriol,* 173**,** 4952-8.

Wu, X. C., Ng, S. C., Near, R. I. & Wong, S. L. 1993. Efficient production of a functional single-chain antidigoxin antibody via an engineered *Bacillus subtilis* expression-secretion system. *Biotechnology (N Y),* 11**,** 71-6.

Zeigler, D. R. & Dean, D. H. 1990. Orientation of genes in the *Bacillus subtilis* chromosome. *Genetics,* 125**,** 703-8.

Zhang, K., Duan, X. & Wu, J. 2016. Multigene disruption in undomesticated *Bacillus subtilis* ATCC 6051a using the CRISPR/Cas9 system. *Sci Rep,* 6**,** 27943.

Zhou, J., Lloyd, S. A. & Blair, D. F. 1998. Electrostatic interactions between rotor and stator in the bacterial flagellar motor. *Proc Natl Acad Sci U S A,* 95**,** 6436-41.

Zobel, S., Kumpfmuller, J., Sussmuth, R. D. & Schweder, T. 2015. *Bacillus subtilis* as heterologous host for the secretory production of the non-ribosomal cyclodepsipeptide enniatin. *Appl Microbiol Biotechnol,* 99**,** 681-91.