UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL

Studies on the relationship between assembly of the spore coat and the biofilm

Bruno Alexandre Amaral Gonçalves

Mestrado em Microbiologia Aplicada

Dissertação orientada por:
Prof. Dra. Mónica Serrano
Prof. Dr. Rogério Tenreiro

2020

Acknowledgements

I would like to thank my supervisor Prof. Dr. Mónica Serrano for guiding me throughout this work, for all the advice, availability, comprehension, for the discussions and for continuously teaching me techniques and all the basics of working in a microbiology lab. I also thank Prof. Dr. Mónica for reviewing my writings and helping me make a better thesis with suggestions and tips, which were always very helpful and insightful. Thank you very much.

I would like to thank my co-supervisor Dr. Tiago Cordeiro for all the help and the discussions to find solutions for our scientific problems and Prof. Dr. Rogério Tenreiro for being my internal supervisor.

I would also like to thank Prof. Dr. Adriano Henriques, who accompanied my work as well and was always ready to provide helpful input and discussions to better my work and the project. I thank him for reviewing my thesis and helping in the making of this work, with an amazing curiosity and with suggestions to move forward and learn new things.

I would like to extend my special thanks to Diogo Martins, who was always there from the beginning, who taught me a lot and with whom this work benefitted for his availability, comprehension, and solidarity. For checking on me and making sure I wasn't lost in this new environment. Also, for the great environment and all the laughs I thank you very much.

To my colleagues in the laboratory, I would like to thank first, Sara Ramalhete and Aristides Mendes who were closest and to whom I frequently resorted for help and who provided a great working environment for me to work in. A big thanks to my other colleagues, also very available and ready to help, Beatriz Martins, Khira Amara, Leonor Duarte, Mariana Ferreira, Mónica Louro and Dr. Zoé Silva, and to the former colleagues whom I had the pleasure to meet Eleonora Marini and Mariana Barreira.

I would also like to thank Instituto de Tecnologia Química e Biológica António Xavier de Universidade Nova de Lisboa for receiving me as a master student and providing with the scientific environment and all the tools and conditions to execute this work.

This work was supported by the FCT ("Fundação para a Ciência e a Tecnologia") through the award PTDC/BIA-MIC/29293/2017. I also thank the FCT for the support in this project.

Lastly, I would like to thank my family for all the support and love throughout this year, for always being there for me and allowing me this amazing opportunity to study and work on my master's degree. To my friend I thank for the support, for checking on my progress and accompanying me side by side while they did their own master's thesis.

Resumo

O gene *yabG* é conservado em bactérias esporulantes e faz parte de uma assinatura genómica para a esporulação. YabG é uma protease tiólica que em *Bacillus subtilis*, é necessária para o processamento de um grupo de proteínas da superfície do esporo antes do seu *cross-linking* pela transglutaminase Tgl. O gene *veg* está localizado a jusante de *yabG* e este arranjo é conservado em organismos esporulantes. Veg foi implicado no controlo do desenvolvimento de biofilmes, o que sugere que os dois genes, *yabG* e *veg*, possam estabelecer uma ligação funcional entre o desenvolvimento de biofilmes e a montagem da superfície do esporo. A estrutura do biofilme é mantida por uma matriz extracelular. A formação de esporos pode ocorrer durante a formação do biofilme e após a lise da célula-mãe, estes são mantidos em associação com a estrutura. Suspeitamos que a composição da superfície do esporo possua características que permitem aos esporos a associação com a matriz do biofilme e a sua manutenção nesta estrutura. Aqui, iniciámos trabalho com vista a testar a possibilidade de que Tgl, YabG e Veg partilham uma ligação funcional. Caracterizámos a atividade de Tgl e, significativamente, demonstrámos que Veg tem um papel na montagem da superfície do esporo.

Os esporos de *B. subtilis* são estruturas celulares dormentes, altamente resistentes a uma grande variedade de fatores físicos e químicos e capazes de permanecer viáveis no ambiente durante períodos muito longos. O plano estrutural do esporo é conservado entre espécies de organismos esporulantes e consiste de um compartimento central, o núcleo, que contém o genoma, rodeado por uma camada de peptidoglicano designada por córtex, essencial para dormência e resistência ao calor, e finalmente por um manto proteico. O manto é composto por mais de 80 proteínas, que se organizam numa subestrutura interna lamelar, noutra externa electrodensa, e numa crosta glicosilada que forma a superfície do esporo. O manto protege o esporo contra agentes químicos diversos, incluindo enzimas capazes de degradar o córtex, anticorpos ou digestão por predadores. O manto influencia ainda a interação do esporo com moléculas capazes de induzir a germinação e a sua adesão a superfícies bióticas e abióticas. As proteínas do manto são sintetizadas na célula-mãe e recrutadas para a superfície do esporo em desenvolvimento mediante interações proteína-proteína específicas. Uma classe de proteínas morfogenéticas controla a localização inicial das proteínas do manto à superfície do esporo, enquanto que uma segunda classe dirige a sua deposição em redor da superfície do esporo. A montagem das diferentes subestruturas do manto resulta ainda da expressão sequencial, na célula mãe, de uma cascata de proteínas regulatórias que controla a transcrição dos genes codificantes para as proteínas do manto.

SafA é uma proteína morfogenética responsável pelo recrutamento de proteínas do manto interno. Esta proteína é recrutada para a superfície do esporo mediante uma interação com SpoIVA, uma ATPase que polimeriza à superfície do esporo e que forma uma camada basal no topo da qual o manto é formado. A deposição de SafA em redor do esporo é controlada por outra proteína morfogenética, SpoVID, e envolve uma interação direta entre as duas. SafA existe em três formas distintas na célula, $SafA^{FL}$, que corresponde à proteína completa, $SafA^{C30}$, correspondente à parte C-terminal da proteína completa, e $SafA^{N21}$, que corresponde à parte N-terminal. $SafA^{C30}$ e $SafA^{N21}$ são formadas por iniciação de tradução interna do mRNA *safA*, com início no codão 164. A região C30 é responsável pelo recrutamento das proteínas do manto interno enquanto que a região N21 possui sinais de localização sub-celular. Estes sinais consistem num domínio LysM de interação com o peptidoglicano do córtex, e uma região de 12 aminoácidos que permite a interação com SpoVID. O domínio LysM permite a interação de SafA com o córtex e a localização da proteína na interface córtex/manto interno. As diferentes formas de SafA são produzidas cedo na linha de expressão genética da célula-mãe, sob o comando de σ^E. Entre as proteínas recrutadas por SafA para o manto interno está uma transglutaminase designada Tgl que é produzida sob o controlo do factor σ^K, nas fases mais tardias da esporulação. Tgl é recrutada para o esporo em desenvolvimento num processo denominado localização guiada pelo substrato. A montagem de SafA é auto-regulatória, dado que o recrutamento de Tgl, resulta no *cross-linking* de SafA mediante ligações covalentes ε-(γ-glutamyl)lysil.

SafA forma complexos de elevado peso molecular, que são estabilizados por ligações ε-(γ-glutamyl)lysil catalisadas pela Tgl. *In vitro*, $SafA^{C30}$ forma um hexâmero oblongo, $(SafA^{C30})_6$, Tgl introduz *cross-links* nesta espécie. Recentemente foi descoberto que a atividade da Tgl é influenciada pela concentração local de $SafA^{C30}$. Com o aumento da concentração de $SafA^{C30}$, a atividade da Tgl diminui, sugerindo que a Tgl se liga covalentemente a $SafA^{C30}$ durante a reação. Este fenómeno pode representar um mecanismo para o controlo da atividade enzimática *in vivo*. A imobilização da Tgl

durante a reação impõe uma auto-limitação da extensão da reação que dependeria da concentração do substrato e da enzima.

Este estudo debruçou-se sobre a caracterização da reação de Tgl com a proteína completa, $SafA^{FL}$, e ainda com $SafA^{N21}$. O *cross-linking* das diferentes formas de SafA, sobre-produzidas e purificadas, foi testado através de ensaios enzimáticos. Nos ensaios realizados com $SafA^{FL}$, foi possível verificar uma atividade de *cross-linking* comparável à observada com $SafA^{C30}$ e a atividade da enzima é inversamente proporcional à concentração de substrato. A reação com $SafA^{N21}$ ocorre com uma eficiência muito reduzida, apesar de este substrato conter vários resíduos de glutamina e de lisina, sugerindo que esta região não está envolvida em *cross-linking* a si mesma ou a Tgl. Pode, no entanto, estar envolvida em *cross-linking* a outra proteína *in vivo* ou ao peptidoglicano do córtex. $SafA^{C30}$ apenas possui dois resíduos de lisina, nas posições 177 e 318 relativamente ao primeiro aminoácido da proteína completa. A substituição das duas lisinas presentes em $SafA^{C30}$ por alaninas levou a resultados interessantes acerca do papel destes aminoácidos e da Tgl na formação de complexos macromoleculares de SafA. As substituições levaram a uma redução na representação de uma espécie de alto peso molecular, presumivelmente o hexâmero $(SafA^{C30})_6$. A substituição K177A teve um efeito menos severo que a substituição K318A. Assim, o resíduo de lisina na posição 318 faz uma contribuição mais importante para o *cross-linking* pela Tgl. Este resíduo encontra-se perto de dois dos quatro resíduos de cisteína (C323 e C325) da região C30, e ainda de vários resíduos de glutamina. Quando as reações de *cross-linking de* $SafA^{C30}$ foram analisadas por SDS-PAGE em condições não redutoras, o efeito das substituições K117A e/ou K318A não é muito evidente, mas a redução na acumulação de $(SafA^{C30})_6$ é notória quando as reações são analisadas sob condições redutoras. Estes resultados sugerem que um modelo no qual $(SafA^{C30})_6$ é formado mediante interações não covalentes, e estabilizado por pontes dissulfureto. Tgl exerce uma atividade de *spotwelding* sobre este complexo, introduzido *cross-links* em posições específicas, mas não entre todos os monómeros que compõem o hexâmero. Estes resultados reforçam ainda a noção de que a Tgl estabiliza complexos pré-formados por oposição à polimerização *de novo*.

Outra proteína que foi alvo de estudo foi Veg. Veg é uma proteína de 86 aminoácidos cuja função e estrutura estão pouco estudadas. O gene *veg* localiza-se a jusante de *yabG*, um gene integrado numa assinatura genómica para a esporulação. YabG é uma protéase de cisteína necessária, em *B. subtilis*, para o processamento *in vivo* dos substratos da Tgl. YabG possui também um papel importante na montagem das camadas superficiais do esporo no patogénio humano *Clostridioides difficile*. Dois estudos distintos, em *B. subtilis*, atribuíram funções gerais a Veg, na formação de biofilmes e na germinação de esporos. Veg, por um lado, promove a formação de biofilme, através da inativação, directa ou indirecta, do repressor SinR. Por outro lado, Veg reduz a extensão e a eficiência da germinação. Aqui, expandimos o conhecimento destas funções estudando o efeito da sobre-expressão de *veg* em estágios tardios da esporulação. *veg* foi sobre-expresso nas fases tardias da esporulação, sob o controlo do promotor do gene σ^K-dependente *gerE* em cópia simples, inserido no locus não essencial *amyE*, e em cópia múltipla, num plasmídeo replicativo. Os dois alelos foram introduzidos numa estirpe selvagem e num mutante *veg* congénico, as estirpes inoculadas em meio de esporulação e os esporos resultantes purificados em gradiente de densidade. As proteínas do manto foram extraídas e analisadas por SDS-PAGE, espectrometria de massa e *immunobloting*. A expressão de P_{gerE}-*veg* resulta na redução da representação ou extractibilidade de dois componentes abundantes do manto externo, as proteínas CotG (36 kDa) e CotB (66 kDa), a primeira das quais é um determinante da organização estrutural desta subestrutura. Um conjunto adicional de bandas encontravam-se alteradas nos extratos do manto preparados a partir das estirpes P_{gerE}-*veg*. Através de espectrometria de massa foi possível identificar YdhD, uma hidrólase de peptidoglicano, implicada na germinação do esporo, correspondendo a uma banda de cerca de 46 kDa, YjdH, uma proteína de função desconhecida correspondendo a uma banda de cerca de 12 kDa e YtmK, uma proteína de ligação à cisteína de transportador ABC, correspondendo a uma banda de cerca de 37 kDa. A identificação de YdhD, que promove a capacidade de germinação em esporos, revela e reforça o papel de Veg na germinação em linha com estudos anteriores. Estudos de modelação molecular indicam que Veg contém um motivo *Sm-like* característico de proteínas que regulam a estabilidade e/ou a tradução de mRNAs, presentes em todos os domínios da vida. O modelo sugere que Veg, tal como as proteínas Sm e *Sm-like*, forma heptâmeros com um canal central e uma superfície carregada positivamente, com o potencial de ligação a ácidos nucleicos de cadeia simples. Estes resultados introduzem um nível novo de regulação no processo de montagem do manto e do

desenvolvimento de biofilmes. Veg pode atuar ao nível da estabilização de mRNA e da tradução para controlar a montagem do manto e, possivelmente influenciar do mesmo modo o desenvolvimento dos biofilmes.

O gene *veg* de *C. difficile* foi clonado sob o comando de um promotor T7*lac* em *Escherichia coli* para a sobre-produção e a purificação de Veg-His_6. A análise por SDS-PAGE revelou a presença de Veg-His_6, com um peso estimado de 11 kDa, nas diferentes frações. Foi estabelecido um protocolo geral para a super-produção e a purificação de Veg-His_6. Este sistema poderá ser utilizado para a caracterização bioquímica e estrutural de Veg. Por outro lado, a caracterização funcional do gene *veg* revelou a função deste gene na formação da superfície do esporo, uma descoberta que acentua a ideia de que existe uma relação forte entre a formação do biofilme e o processo de esporulação, e uma coordenação entre os dois processos.

Palavras-chave: manto, montagem, SafA, Tgl, Veg

Abstract

In the model organism *Bacillus subtilis*, the surface of the spore consists of a protein coat that affords protection against noxious chemicals and lytic enzymes and modulates spore germination and binding to cells and abiotic surfaces. The coat, formed by over 80 proteins, is divided into an inner layer, an outer layer and a crust. The coat proteins are synthesized in the mother cell and are recruited to the surface of the developing spores through specific protein-protein interactions and also according to successive waves of gene expression. SafA is a morphogenetic protein responsible for the assembly of the inner coat layer. SafA exists in three forms, the full-length protein, $SafA^{FL}$, $SafA^{C30}$ which corresponds to the C-terminal moiety of the protein, and $SafA^{N21}$, corresponding to the N-terminal moiety of the protein. The C30 region is responsible for recruiting the inner coat proteins, whereas the N-terminal region carries the signals for localization at the spore surface. Among the proteins that SafA recruits to the inner coat is a transglutaminase, called Tgl, in a process termed substrate-driven localization. Assembly of SafA is auto-regulatory in that once assembled, Tgl introduces ε-(γ-glutamyl)lysil bonds into SafA. $SafA^{C30}$ forms an oblong hexamer, $(SafA^{C30})_6$ and this species is cross-linked *in vitro* by Tgl. Strikingly, the activity of Tgl decreased with increasing $SafA^{C30}$ concentration, suggesting both that Tgl cross-linked itself to $SafA^{C30}$ and a mechanism for controlling the activity of the enzyme *in vivo*. Here we overproduced and purified $SafA^{FL}$ and $SafA^{N21}$ and tested for cross-linking by purified Tgl. Although Tgl proved to be much less efficient in cross-linking $SafA^{N21}$ than $SafA^{FL}$ or $SafA^{C30}$, its activity was inversely proportional to the concentration of all three substrates. Thus, *in vitro*, all three forms of SafA have the potential of controlling the activity of Tgl in a concentration-dependent manner. That $SafA^{N21}$ is a poor substrate for Tgl, in spite of the fact that it contains several glutamine and lysine residues, suggests that this region of SafA is not involved in cross-linking to itself or to Tgl in vivo and raises the possibility that it may be involved in cross-linking to another protein or directly to the cortex peptidoglycan. Moreover, we found that formation of $(SafA^{C30})_6$ involves disulfide bonds and that Tgl exerts a "spotwelding" activity on this species, *i.e.*, not all the monomers in $(SafA^{C30})_6$ are cross-linked by Tgl. Two lysine residues in $SafA^{C30}$, K177 and K318 are important for this activity but K318 makes a more important contribution. Since K318 is close to two of the four cysteine residues in $SafA^{C30}$, C323 and C325, and also in the close vicinity of several glutamine residues, a model is proposed in which disulfide bond formation at this position nucleates further cross-linking of $(SafA^{C30})_6$ by Tgl.

Veg, an 86 amino acid protein, was known to influence biofilm formation and spore germination. The *veg* gene is downstream of *yabG*, a gene that is part of a genomic signature for sporulation. *yabG* codes for a cysteine protease required for processing of the Tgl substrates prior to their cross-linking by Tgl. YabG also functions in the assembly of the spore surface layers in the human pathogen *Clostridioides difficile*. Here, we have shown that the overexpression of *veg* during the late stages of sporulation in *B. subtilis* impairs assembly of the abundant CotB and CotG, the latter a critical determinant of the structural organization of the outer coat. The assembly of other proteins, such as the peptidoglycan hydrolase YdhD and of a putative L-cysteine binding protein was also affected. Veg bears a Sm-like motif characteristic of proteins that regulate mRNA stability and translation in all domains of life. Molecular modelling studies suggest that, like the Sm proteins, Veg forms an heptamer with a central channel and a positively charged surface that may mediate interactions with single-stranded nuclei acids. Veg may thus define a novel level of control in spore coat assembly, influencing mRNA stability and/or translation and we propose that it functions in the same way in biofilm development. We overproduced and purified the *C. difficile* Veg protein, an 11 kDa protein with significant similarity to *B. subtilis* Veg. The structural and biochemical characterization of Veg will provide important insights into its function and its suspected functional relationship to YabG.

Keywords: spore coat, assembly, SafA, Tgl, Veg

Table of Contents

List of figures

Symbols and abbreviations

%	percentage
°C	degrees Celsius
::	fusion
Δ	deletion
σ	sigma factor
µl	microliter
µM	micromolar
A	adenine
Amp	ampicillin
Amp^R	ampicillin resistance
B&W	Bott and Wilson
BslA	biofilm surface layer protein A
C	cytosine
Cm	chloramphenicol
Cm^R	chloramphenicol resistance
ddH_2O	double distilled H_2O
DNA	deoxynucleic acid
DSM	Difco sporulation medium
DTT	dithiothreitol
g	acceleration of gravity
G	guanine
h	hours
H_2O	water
His_6	hexa-histidine tag
Km	kanamycin
Km^R	kanamycin resistance
kDa	kilodaltons
LA	Luria agar
LB	Luria broth
lb/in^2	pounds per square inch
M	molar
mg	milligram
min	minutes
mL	mililiter
mM	milimolar
mV	milivolt
nm	nanometer
Neo	neomycin
Neo^R	neomycin resistance
OD	optical density
OD_{600}	OD at 600 nm
OFM	outer forespore membrane
PBS	phosphate buffer saline
PCR	polymerase chain reaction
P_{gerE}	*gerE* promotor region
SafA	SpoVID-associated factor A
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate – polyacrilamide gel electrophoresis
SOE	splicing by overlap extension
T	thymine
TapA	TasA anchoring and assembly protein A

1. Introduction

1.1. *Bacillus subtilis*

The genus *Bacillus* is well known for its role in human and animal infection (*e.g.*, *Bacillus anthracis* and *Bacillus cereus*) but it is also important as a producer of antibiotics, insecticides, proteases, and other useful products (*e.g.*, *Bacillus subtilis*). Members of this genus are Gram-positive, aerobic endospore-forming bacteria commonly found in the soil (Cooper *et al.*, 2018).

The ability of *B. subtilis* to sporulate and to form biofilms in which spores are produced and the ease with which it can be genetically manipulated made this species one of the most studied bacteria and an experimental model organism for Gram-positive bacteria (Cohn, 1876; Sonenshein, Hoch and Losick, 2001, Dubnau and Lovett, 2001). The release of the complete *B. subtilis* genome in 1997 (Kunst *et al.*, 1997), has enabled global expression analysis at both the gene and protein levels (Earl *et al.*, 2008). Previous comparative genome studies have led to the identification of a genomic signature for sporulation, defined as those genes which are found in 90% of the endospore forming bacteria, and present in no more than 10% of the non-endospore-formers. Strikingly, after decades of genetics, transcriptomics and proteomics studies on sporulation in *B. subtilis*, the function of some of the signature genes on the process had gone unnoticed, yet some of the these genes were shown to have a role in biofilm development (Abecasis *et al.*, 2013). We take this as an indication of a close evolutionary and functional link between spore differentiation and biofilm formation. A link between spore differentiation and the formation of complex multicellular communities has also been observed in bacteria such as *Streptomyces coelicolor* or the gram-negative *Myxococcus xanthus*, but the spores and the biofilms formed by these organism are structurally and functionally different from those of *B. subtilis* and closely related organisms (Chater, 2001; Konovalova *et al.*, 2010).

Below, we outline key aspects of biofilm formation and spore development with emphasis on the assembly of the spore surface layers. We also give details on the function of a group of proteins whose study may help us understand the connection between biofilm formation and the assembly of the spore surface layers. By spore surface layers, we mean the spore coat and crust. One possibility is that the formation of the coat and crust layers, which requires the ordered assembly of a large number of proteins, and the biofilm matrix share common organizational principles and also some components, as at least two major components of the biofilm matrix are also coat-associated (see also below). Since spores are formed at a late stage in biofilm formation, the possible connection between coat and matrix assembly may provide the mature spores, once released from the mother cell, to interact with and become an integral and stable part of the biofilm matrix.

1.2. Biofilm formation in *B. subtilis*

Biofilms are communities of surface associated microorganisms encased in an extracellular matrix. They are a nearly universal bacterial trait and *B. subtilis* is no exception to the rule. It can form biofilms in a variety of surfaces, from liquid-air to solid-air to liquid-solid surfaces (Hall-Stoodley *et al.*, 2004). Biofilms are a fascinating phenomenon and one with great importance, as they can become very problematic in industrial, clinical, and other man-made environments where, by increasing bacterial resistance to antimicrobials and to removal, their formation is undesirable (Mah and O'Toole, 2001; Davies, 2003; Vlamakis *et al.*, 2013).

The development of the *B. subtilis* biofilm starts with formation of long-chains of non-motile cells, that secrete an extracellular matrix, originating from free motile cells that receive some sort of external signal to produce the biofilm (Branda *et al.*, 2006). Inside the biofilm, genetically identical cells express different genes, which makes a set of functionally distinct subpopulations. In addition to matrix-producing cells, there are motile cells and spores within the biofilm, with a specific spatial distribution (Branda *et al.*, 2001; López *et al.*, 2010). This differentiation and distribution are dynamic as cells of one type can alter their gene expression to become a different type of cell in the biofilm (or germinate, in the case of spores) (Figure 1.1) (López *et al.*, 2010).

The extracellular matrix of the *B. subtilis* biofilm is primarily composed of exopolysaccharide (EPS) and proteins. The EPS is synthesized by the products of the *epsA-O* operon (hereafter called *eps* operon), a 15 gene operon, which encodes the biosynthetic machinery that utilises sugar substrates to produce the exopolysaccharide (Nagórska *et al.*, 2010). The protein component of the *B. subtilis* biofilm matrix is composed of several uncharacterized proteins and a few that have been described. TasA

(translocation-dependent antimicrobial spore component), BslA (biofilm surface layer protein) and TapA (TasA anchoring and assembly protein) are three examples of such studied proteins (Branda *et al.*, 2006). TasA forms long amyloid-like fibers that are attached to the cell wall by TapA, which also has a role in fiber assembly (Romero *et al.*, 2011). These two proteins are codified in the *tapA-sipW-tasA* operon, which also expresses the signal peptidase SipW whose function is to process TapA and TasA during secretion to the external milieu (Terra *et al.*, 2012). BslA is another secreted protein important for surface hydrophobicity, complex colony morphology and pellicle formation (Branda *et al.*, 2001; Kobayashi and Iwano, 2012). Importantly, TasA and BslA are also spore-associated proteins (Stöver and Driks, 1999; Serrano *et al.*, 1999; Portinha, 2015).

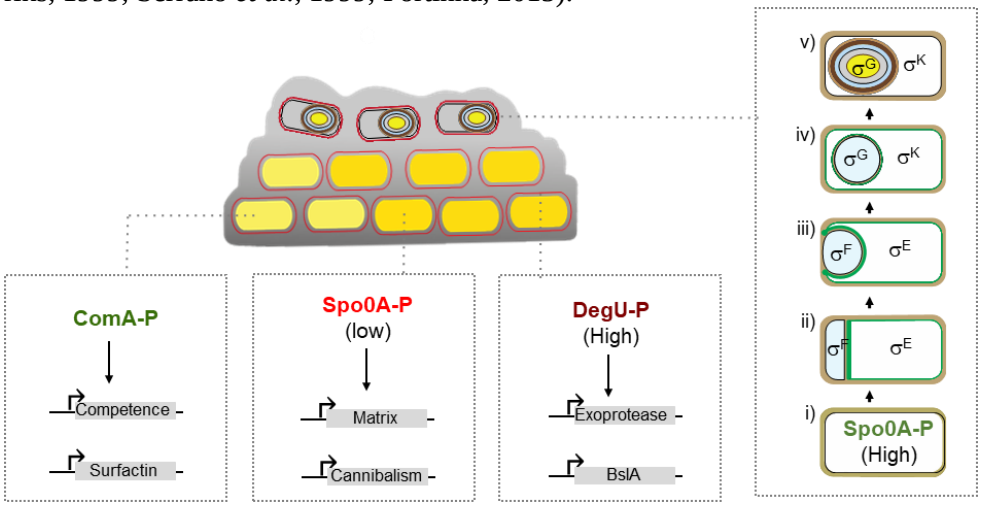

Figure 1.1. Regulation of gene expression during biofilm formation by *B. subtilis*. In the biofilm there are distinct subpopulations of cell types that exhibit different spatiotemporal distribution (*e.g.*, motile; competent, surfactin producers, matrix producers, exoprotease producers and sporulating cells). The master regulators that are responsible for the activation of specific gene expression cascades that lead to cell differentiation in the biofilm are indicated (ComA, Spo0A and DegU; adapted from Mielich-Süss and Lopez, 2014). The right panel represents the main stages of sporulation: *i*) a pre-divisional cell; *ii*) a cell that has completed asymmetric division forming a forespore and a mother cell; *iii*) a cell during engulfment; *iv*) a cell after the completion of engulfment of the forespore by the mother cell; *v*) assembly of the spore protective layers, the cortex peptidoglycan layer and the protein coat. At the end of the process, the mature spore is released upon lysis of the mother cell. Stages of sporulation are controlled by the timely expression and action of sporulation-specific sigma factors (σ^F, σ^E, σ^G, and σ^K).

1.3. Biofilm regulation

Biofilm formation is regulated by a complex transcriptional network that responds to environmental changes. Spo0A, DegU and ComA are the three essential master regulators that tightly control the transition from a motile, planktonic state to a sessile biofilm mode of growth (Figure 1.1) (Mielich-Süss and Lopez, 2015).

Spo0A is a major transcriptional regulator that controls the expression of more than 100 genes, including those needed for biofilm formation and sporulation. *spo0A* is part of the sporulation genomic signature (see above; Abecasis *et al.*, 2013) This regulator is activated through phosphorylation of an aspartate residue. Both phosphorylated and unphosphorylated Spo0A (Spo0A~P and Spo0A, respectively) are always present in the cell. Depending on the relative level of these two forms, gene expression is controlled differentially. Intermediate levels of Spo0A~P promote matrix gene expression, while high levels of this form induce sporulation gene expression. Spo0A~P concentration is controlled by four sensor kinases (KinA, KinB, KinC and KinD), either directly (in the case of KinC) or through a phosphorelay (Jiang *et al.*, 2000; Molle *et al.*, 2003; Fujita *et al.*, 2005).

Under the control of Spo0A~P is SinI, a SinR antirepressor protein (Shafikhani *et al.*, 2002). SinR is a DNA-binding protein that represses the expression of the *eps* and *tapA* operons (Kearns *et al.*, 2005; Chu *et al.*, 2008), and SinI acts by blocking SinR- mediated repression through the formation of a SinI-SinR heterodimeric complex, rendering SinR incapable of binding to DNA (Figure 1.2) (Bai *et al.*, 1993; Newman *et al.*, 2013). Spo0A~P represses yet the expression of a gene coding for another matrix gene repressor, *abrB.* AbrB represses expression of *bslA* in addition to the *tapA-sipW-tasA* and

eps operons (Hamon *et al.*, 2004). To add to this regulation a gene under SinR and AbrB dual control codes for a protein that can form a complex with SinR, acting as another antirepressor. Its product is a protein called SlrR, and it as two distinct modes of regulating biofilm formation. Firstly, the antirepression of the matrix genes by binding to SinR and forming a SinR-SlrR complex, which in a second instance acts as a repressor for genes involved in motility (*hag* gene, flagellin) and in cell separation (*lytABC* and *lytF*, autolysins) (Chu *et al.*, 2008). Furthermore, RemA, another DNA-binding protein, is also required for *slrR* and matrix gene expression and acts independently of any previously described pathways (Winkelman *et al.*, 2013). RemA (formerly known as YlzA), an 89 amino acid DNA-binding protein, not only promotes transcription of the *slrR* gene but also directly activates gene expression from the *eps* and *tapA* operons by binding to multiple sites upstream of both promoters. *remA* is also part of the sporulation genomic signature (Abecasis *et al.*, 2013; Winkelman *et al.*, 2013). Another protein that induces biofilm formation by inhibiting SinR, is Veg (Figure 1.2). Veg is an 86 amino acid, highly conserved, protein encoded by the *veg* gene, which is transcribed at very high levels during both exponential growth and sporulation (Ollington and Losick, 1981; Le Grice *et al.*, 1986; Lei *et al.*, 2013). Previous studies also reported delayed spore germination in *veg* mutants (Fukushima *et al.*, 2003).

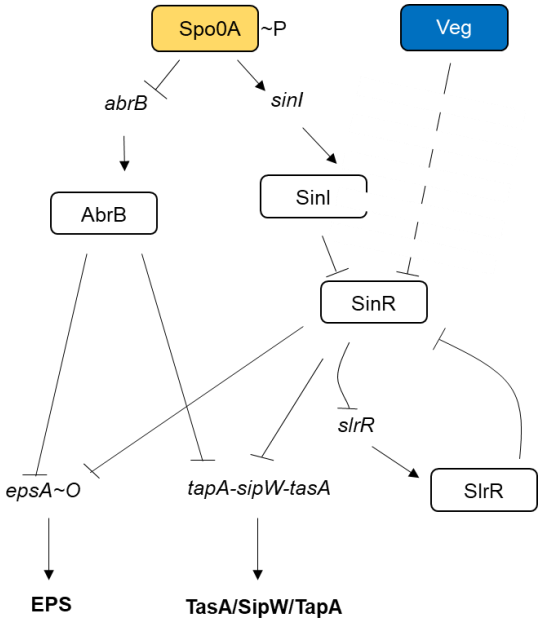

Figure 1.2. Spo0A, SinR and Veg regulation of matrix production operons. Arrows represent positive regulation or activation and blunt arrows represent negative regulation or repression. Dashed lines represent unknown pathways, where the level of regulation is not yet known.

An additional pathway is the DegS-DegU two component system. In this system DegS is the sensor kinase and DegU is the response regulator. In *B. subtilis*, DegU is a major regulator that controls several cellular processes, such as motility and competence. This two-component system regulates the expression of *blsA* and the *pgs* operon (poly-γ-glutamate synthesis) (Figure 1.1). A *degU* mutant is deficient in colony biofilm formation due to the loss of the surface hydrophobicity protein BlsA (Kobayashi, 2007; Verhamme *et al.*, 2009).

Lastly, ComA regulates the differentiation of naturally competent and surfactin- producing cells, activated by phosphorylation by a sensor membrane kinase, ComP. This kinase is previously activated by a pheromone termed ComX, that functions as a quorum-sensing signal, meaning that ComP is only activated a certain cell density is reached (Figure 1.1) (Magnuson *et al.*, 1994).

The phosphorylation of this three master regulators can be further fine-tuned through the Rap phosphatases which respond to specific cognate Phr peptides, representing a cell-cell communication signalling programme (Veening *et al.*, 2005).

1.4. Sporulation in *B. subtilis*

Cells that reach a high level of SpoOA~P activate the transcription of various key sporulation-specific genes, specifically *spoIIA*, *spoIIE* and *spoIIG* genes (Molle *et al.*, 2003). During sporulation, they differentiate into highly resistant, metabolically dormant cells, capable of resisting extreme heat, desiccation, and ionizing radiations. Estimates of spore longevity range from thousands to millions of years, although the lower end of the range is more likely (Figure 1.1) (Nicholson *et al.*, 2000).

Spores are made up of several concentric layers, as seen by electron microscopy: the core, which harbours a copy of the genome and is partially dehydrated, the cortex, lying between the inner and outer spore membranes and composed of a cell-wall like peptidoglycan, the coat, with two visibly separate layers, the inner and outer coat, and the outmost layer of the spore, the crust (Figure 1.3A) (Warth *et al.*, 1963; Aronson and Fitz-James, 1976; Mckenney *et al.*, 2013). Some bacteria, such as *B. anthracis* possess a different outmost layer called exosporium (Aronson and Fitz-James, 1976). This spore architecture protects the bacterial genome from a variety of stress conditions, conferring resistance to heat, UV radiation, oxidative stress, ingestion by bacterivores, and chemical or enzymatic degradation. Heat resistance is conferred by the partial dehydration of the core with most of the water being replaced by Ca^{2+}-dipicolinic acid. The cortex is essential to sustain this dehydrated state of the spore core and for spore mineralization (Setlow, 2007). The coat is necessary for the increased resistance to UV radiation and chemicals, such as peroxides, for protection of the spore cortex against the hydrolytic activity of enzymes, and it also provides physical protection from predators (Henriques and Moran, 2007).

Sporulation initiates with an asymmetric division, producing a smaller cell, the forespore, and a larger cell, the mother cell. The forespore is completely encased by the mother cell, in a stage called engulfment. The mother cell then proceeds to synthesize the cortex and assemble the proteinaceous coat, around the forespore (Figure 1.1). The process is finished when the mother cell lyses and releases the mature spore into the environment. The cycle completes when the spore germinates, under more favourable conditions, into a vegetative cell so the process can repeat itself (Higgins and Dworkin, 2012; Setlow *et al.*, 2017).

The morphological differentiation undergone by the sporulating cell during sporulation results from different patterns of gene expression. Gene expression is compartmentalized between the forespore and the mother cell and governed by sporulation specific RNA polymerase sigma factors (σ). σ^F and σ^G are active in the forespore and σ^E and σ^K are active in the mother cell, at different times; σ^F and σ^E are active before engulfment is complete, while σ^G and σ^K are active after engulfment completion (Figure 1.1) (Hilbert and Piggot, 2004).

σ^F is the first of the sporulation σ factors to become active, and it has a role in the activation of the subsequent σ factors. Its main function is to couple gene expression between the forespore and the mother cell as well as direct synthesis of the late forespore specific σ factor σ^G. σ^G also ensures coupling between cell specific gene expression in the two cells, while controlling expression of several sporulation genes, with functions in protection from hazardous conditions (e.g. *ssp* genes; encoding SASPs), and germination genes (e.g. *gerA* operon, involved in germination in response to alanine) (Hilbert and Piggot, 2004; De Hoon *et al.*, 2010). σ^E controls gene expression in the mother cell at the early stages of sporulation, after asymmetric division and before engulfment; σ^E regulates expression of many genes involved in spore morphogenesis (e.g. *spoIVA*, *cotE*, *spoVID* and *safA*, which encode scaffold proteins for coat assembly), in regulating activation of σ^G, and the *sigK* gene (the composite gene for the late mother cell specific factor, σ^K). σ^K controls gene expression, also in the mother cell, at later stages of sporulation, after engulfment of the spore; the genes under the σ^K regulon are involved in spore coat formation, spore maturation and regulation of σ^K-dependent genes. (Hilbert and Piggot, 2004; De Hoon *et al.*, 2010).

The patterns of sporulation gene expression and the stages of spore morphogenesis are intimately linked, helping us to understand the mechanisms that make this developmental process of differentiation possible.

1.5. The spore coat

The spore coat is one of the main components of the spore. It can be distinguished from the cortex, by transmission electron microscopy, because of its higher electron density. It is composed of more than 80 different proteins assembled in a multi-layered structure around the spore. Its assembly is coordinated with the end of cortex assembly (Driks and Eichenberger, 2016). As mention above, the

coat is required for the resistance to UV radiation and chemicals, such as peroxides, for protection of the spore cortex against the hydrolytic activity of enzymes, such as the lysozyme, and it also provides physical protection from predators (Setlow, 2006; Henriques and Moran, 2007). Being one of the outmost layers of the spore, the coat is one of the most variable layers of the spore, between sporulating species. The surface of the spore comes in contact with the environment, and different organisms find themselves in different niches, which is reflected in the composition of their spore surfaces layers (De Hoon *et al.*, 2010; Driks and Eichenberger, 2016). Interestingly, two of the main protein components of the biofilm matrix, TasA and BslA, are also associated with the spore coat (Serrano *et al.*, 1999; Stöver and Driks, 1999; Portinha, 2015: see also above).

Coat assembly is mainly a function of the mother cell (Henriques and Moran, 2007). The timing and location of the expression of the coat genes is essential for proper assembly of the coat layers and is under the control of σ^E and σ^K (Hilbert and Piggot, 2004; Costa *et al.*, 2007). Assembly begins in the outer forespore membrane (OFM), near the cortex, with the attachment of SpoIVA to the membrane, to form the basement layer of the coat. SpoIVA is recruited with SpoVM to the OFM. Both proteins are dependent on each other for their localization (Driks *et al.*, 1994; Price and Losick, 1999; Ramamurthi *et al.*, 2006). SpoIVA is also required for the localization of a morphogenetic protein, SpoVID (Beall *et al.*, 1993; Ozin *et al.*, 2000), and plays a role in cortex synthesis, since *spoIVA* mutants show not only a misassembled coat in the mother cell cytoplasm but also an absent cortex layer (Roels *et al.*, 1992; Stevens *et al.*, 1992; Catalano *et al.*, 2001). SpoVID, like SpoIVA, is necessary for proper coat assembly, although not for cortex synthesis, with *spoVID* mutants producing swirls of coat material that accumulate in the mother cell cytoplasm (Beall *et al.*, 1993). This information emphasizes the ability of coat proteins for self-assembly (Aronson *et al.*, 1992). SpoVID is a highly acidic protein with 575 amino acid residues. It possesses a LysM domain. In case of SpoVID this domain functions only as a protein-protein interaction module (Pereira *et al.*, 2019). The localization of SpoVID is achieved through interaction of the LysM domain of SpoVID with SpoIVA (Wang *et al.*, 2009). SpoVID interacts with another morphogenetic protein, SafA. SafA is suggested to anchor the coat to the cortex (Ozin *et al.*, 2000; Pereira *et al.*, 2019). A more detailed description of SafA is provided below. As Henriques and Moran (2007) summarized, the first stages in coat assembly involve a cascade of interactions, first between SpoVM and SpoIVA, then between SpoIVA and SpoVID, and lastly between SpoVID and SafA, positioning them near the forespore outer membrane (Figure 1.3B and C). The genes encoding for these proteins are all transcribed in the mother cell, under σ^E control (Hilbert and Piggot, 2004; Henriques and Moran, 2007). Additionally, SpoVM and SpoVID are essential for encasement of the coat around the spore (Nunes *et al.*, 2018).

Another morphogenetic protein, CotE, is recruited by SpoIVA and encased through interaction with SpoVID (de Francesco *et al.*, 2012). The *cotE* gene is expressed under σ^E and σ^K control (Zheng *et al.*, 1988; Costa *et al.*, 2007). *cotE* mutant spores lack the outer coat layer while retaining the lamellar structure of the inner coat, suggesting that inner and outer coat assembly are independent (Zheng *et al.*, 1988). CotE's function is thus attributed to the assembly of the outer coat (Figure 1.3), with some contribution to inner coat assembly, where some inner coat proteins are recruited to the inner coat in a CotE-independent manner but are not retained there in its absence (e.g. OxdD, CotS, CwjJ) (Zheng *et al.*, 1988; Chada *et al.*, 2003; Costa *et al.*, 2007). Outer coat assembly is regulated by CotE, but also by other morphogenetic proteins, with CotH and CotO being the most relevant, and both having functions in germination and lysozyme resistance in the spore (Naclerio *et al.*, 1996; Zilhão *et al.*, 1999; McPherson *et al.*, 2005). CotH is a kinase that regulates the assembly of two of the most abundant proteins of the outer coat, CotB and CotG, by phosphorylation (Nguyen *et al.*, 2016; Freitas *et al.*, 2020).

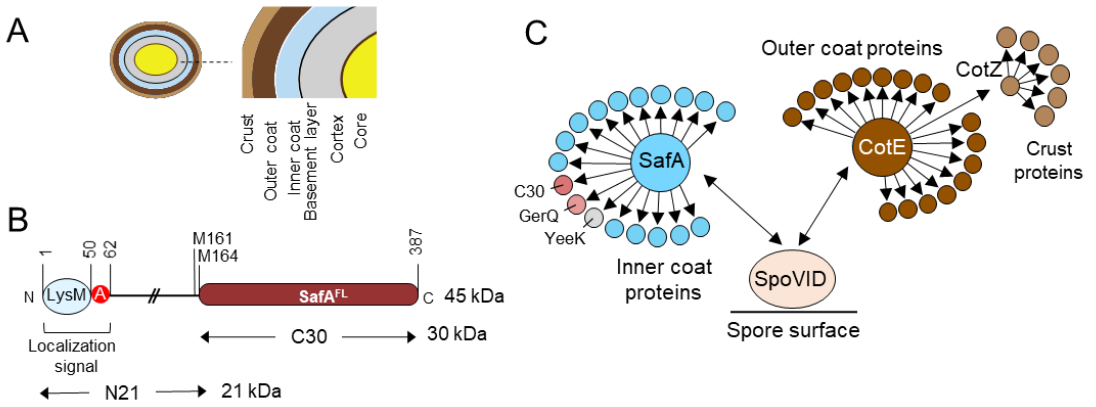

Figure 1.3 Spore coat assembly is governed by morphogenetic proteins that recruit other morphogenetic or structural proteins. SafA governs the assembly of the inner coat layer. A: Layers of the spore, with the core at the center and the crust at the surface; the spore coat is assembled underneath the crust and is the second outermost layer, composed of a basement layer, an inner coat and an outer coat. B: Structural organization of SafA. Three forms are present in cells; the full-length protein, SafA^{FL} (45 kDa), carries a localization signal in the form of a LysM domain and of region A, which binds to the encasement protein SpoVID. Through internal translation starting at Met codons 161 or 164, the SafA^{C30} form (30 kDa) is produced, and is recruited by SafA^{FL} to the coat. N21 (21 kDa) is the N-terminal form that includes the localization signals (LysM and region A) and a linker at its C-terminal end. C: SpoVID binds the morphogenetic proteins responsible for inner and outer coat assembly. These are SafA and CotE, respectively, and their localization to the coat is essential for the recruitment of several structural proteins and enzymes. CotE is also responsible for the recruitment of CotZ, the main crust morphogenetic protein. SpoVID is essential for spore encasement by all layers of the coat. Adapted from Fernandes *et al.*, 2019.

The structure of the outermost layer of the spore, the crust, is most influenced by CotZ, which recruits most crust proteins and CotO, responsible for encasement of the crust (Shuster *et al.*, 2019). Besides morphogenetic and structural proteins, enzymes also assemble in the coat, with many functions in coat maturation (*e.g.*, Tgl, YabG), spore germination (*e.g.*, CwlJ) (Henriques and Moran, 2007) or spore resistance (e.g. CotA). Tgl is a transglutaminase that controls the extractability of several coat proteins, such as GerQ and SafA (Ragkousi and Setlow, 2004; Fernandes *et al.*, 2019), while YabG is a protease thought to generate the substrates of Tgl from precursor proteins (Takamatsu *et al.*, 2000; Kuwana *et al.*, 2006). *yabG* is highly conserved among spore formers, and in pathogenic bacteria, such as *C. difficile*, was shown to be required for the proper assembly of the spore coat and exosporium, but also have an impact in late mother-cell gene expression (Abecasis *et al.*, 2013; Marini, 2020). CwlJ is a cortex lytic enzyme activated during spore germination. CotA is a copper-dependent laccase that contributes to resistance to UV radiation and reactive oxygen species by generating a pigment structurally similar to melanin (Hullo *et al.*, 2001; Martins et al., 2002; Mckenney, Driks and Eichenberger, 2013).

1.6. SafA

SafA is the morphogenetic protein responsible for the correct assembly of the spore inner coat layer. SafA can be found in three monomeric forms, a 45 kDa full-length form, SafA^{FL}, a 30 kDa C-terminal moiety, SafA^{C30}, and a 21 kDa N-terminal moiety, SafA^{N21}, produced through alternative translation initiation at codon 161 and translation arrest before codon 161, respectively (Figure 1.3B) (Ozin *et al.*, 2001a). SafA is mainly responsible for inner coat assembly, localizing to the coat-cortex interface (Takamatsu *et al.*, 1999; Fernandes *et al.*, 2018; Pereira *et al.*, 2019). It is recruited to the spore through an interaction with SpoIVA, and interacts with SpoVID at the beginning of engulfment (Costa *et al.*, 2006; Nunes *et al.*, 2018). This interaction with SpoVID is necessary for the localization of SafA^{FL} and SafA^{C30}, and, consequently, for connecting the cortex and the inner coat (Pereira *et al.*, 2019).

SafA has also a LysM domain localized in the N-terminal end of the protein, shown to be important for protein-peptidoglycan and protein-protein interactions (Figure 1.3B) (Buist *et al.*, 2008; Pereira et al., 2019), and a role in the adherence of the coat to the cortex and of the inner coat to the outer coat (Pereira *et al.*, 2019). Through its C30 portion, which interacts with SafA^{FL} as well as with itself, SafA forms a cap in the mother cell proximal (MCP) forespore pole, and a layer around the spore, after encasement (Nunes *et al.*, 2018). With the localization signal absent in SafA^{C30}, this form cannot localize to the inner coat, becoming dependent on SafA^{FL}, which contains the localization signals, the

LysM domain and region A (Nunes *et al.*, 2018; Pereira *et al.*, 2019). SafA is the main recruiter of Tgl, of which it is a substrate (Fernandes *et al.*, 2018; Fernandes *et al.*, 2019). Tgl localizes to the cortex-coat interface with SafA where it cross-links SafA into higher molecular weight forms ($SafA^{HMW}$) (Fernandes *et al.*, 2018).

$SafA^{C30}$, the C-terminal moiety of $SafA^{FL}$, forms an oblong hexamer in solution, as shown by size exclusion chromatography (SEC) and small-angle X-ray scattering (SAXS). SDS-PAGE analysis of $SafA^{C30}$, in the absence of Tgl, also showed a dimer of $SafA^{C30}$, suggesting that this species is highly stable. It has been proposed that the hexamer of $SafA^{C30}$ and higher molecular weight forms recruit Tgl to the developing spore surface. Tgl then must cross-link these complexes, in a self-limiting reaction, where Tgl becomes, itself, cross-linked into the forming structure (Fernandes *et al.*, 2019).

1.7. Tgl

Tgl is a 28 kDa transglutaminase (TGase), the smallest known TGase, responsible for catalysing amine incorporation and protein cross-linking, and thought to be specific of sporulation in *B. subtilis* (Fernandes *et al.*, 2015). TGases are capable of a myriad of reactions including deamidation, protein cross-linking, amine incorporation among other reactions. The best studied TGases are mammalian TGases, as is the example of type 2 tissue TGase or Factor XIII. These enzymes are produced in an inactive form and their activity is carefully regulated by Ca^{2+}, GTP and proteolysis. They also share structural homology and are members of the papain-like superfamily of cysteine proteases. Their catalytic sites are composed of catalytic triads of Cys-His-Asp or Cys-His-Asn, with restrictions to their access to modulate their activity (Griffin *et al.*, 2002; Lorand and Graham, 2003).

Tgl contains a narrow hydrophobic tunnel (15 Å long) where the catalytic site, composed of a partially redundant non-reciprocal catalytic dyad, is buried (Fernandes *et al.*, 2015). It is composed of Cys116 and Glu187 or Glu115, where Glu115 can replace Glu187, but not the opposite. It is structurally similar to cell wall endopeptidases possessing the same NplC/P60 catalytic core. The catalytic cysteine first forms an acyl enzyme intermediate with a glutamine (Q) donor substrate. Then, it engages a lysine (K) acceptor substrate to form the cross-linked product (Figure 1.4A). Since the catalytic cysteine of Tgl is hidden within the tunnel, recognition of the substrates by the enzyme must be conducted by other residues, namely the aromatic residues at both ends of the tunnel, which are also responsible for keeping water out of the tunnel. These residues play a major role in recognition of the substrates and in the localization of Tgl during spore development. The Q-side (front-side) of the tunnel plays the most important role in localization of Tgl since the Q-substrate is the first to interact with the enzyme, forming an acyl-enzyme intermediate. The K-side (back-side) also plays a role in localization, but to a lesser extent (Fernandes, 2014; Fernandes *et al.*, 2015; Fernandes *et al.*, 2019).

Tgl is produced in its active form, unlike other known TGases, which makes the temporal and spatial regulation of its activity the main form of control, in the absence of other regulatory factors (Fernandes *et al.*, 2019). Localization is in many cases important for protein function (Bauer *et al.*, 2015). Besides $SafA^{FL}$ and $SafA^{C30}$, Tgl has two more known substrates, GerQ and YeeK (Figure 1.4) (Ragkousi and Setlow, 2004; Takamatsu *et al.*, 2009; Fernandes *et al.*, 2019). In a process called substrate-driven localization, Tgl is recruited to the spore surface by its substrates. $SafA^{FL}$ and $SafA^{C30}$ are the main contributors for the recruitment of Tgl to the inner regions of the coat, while GerQ and YeeK appear mainly to delay encasement, by competing with Tgl for a binding interface in SafA. In the absence of $SafA^{FL}$ and $SafA^{C30}$, Tgl is not recruited to the spore surface, being produced, all the same, but remaining in the mother cell cytoplasm. Recruitment is controlled by the time of synthesis since Tgl is produced under the control of σ^K, after $SafA^{FL}$, $SafA^{C30}$ and GerQ at least. Although $SafA^{C30}$ can recruit Tgl to the spore surface in the presence of $SafA^{FL}$ it lacks the capacity to localize to the inner regions of the coat on its own. Without $SafA^{FL}$, $SafA^{C30}$ traps Tgl in the mother cell proximal forespore pole. Recruitment of Tgl by its pre-assembled substrates brings the possibility that its activity is controlled by the local concentration of its substrates (Ragkousi and Setlow, 2004; Zilhão *et al.*, 2005; Takamatsu *et al.*, 2009; Fernandes *et al.*, 2018; Pereira *et al.*, 2019).

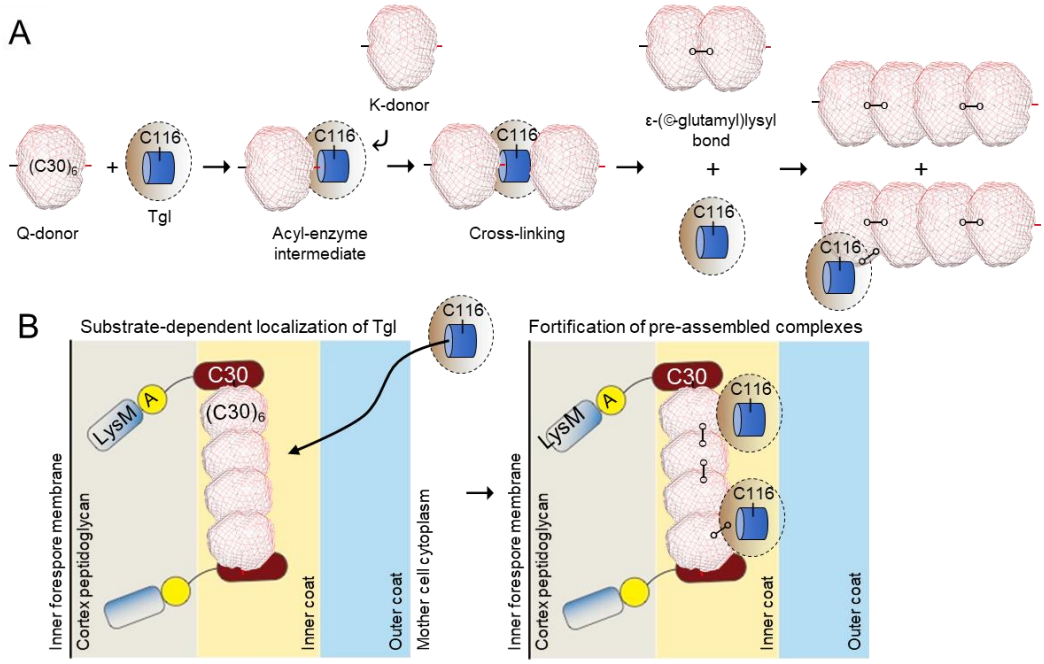

Figure 1.4. Control of Tgl localization and activity at the spore surface. A: reactions of $(SafA^{C30})_6$ with Tgl. We propose that the $SafA^{C30}$ hexamer $(SafA^{C30})_6$ acts both as a Q and K donor and becomes cross-linked by Tgl to form a dimer of hexamers. With time, this species becomes cross-linked into larger cross-linked oligomers, in what may represent *de novo* protein polymerization. Tgl itself is cross-linked into the forming structure. B: SafA is recruited early to the surface of the developing spore, prior to synthesis of Tgl. At the spore surface, SafA is found in at least two forms: $SafA^{FL}$ (full-length) and $SafA^{C30}$. $SafA^{FL}$ has a localization signal formed by a LysM domain and region A. $SafA^{C30}$ interacts with $SafA^{FL}$ via the corresponding region in the latter. As $SafA^{C30}$ lacks localization signals, this interaction is required for the localization of $SafA^{C30}$ to the coat. $SafA^{FL}$ and $SafA^{C30}$ are key determinants in the recruitment of Tgl to the coat (left panel), placing the enzyme in close proximity to its substrates. In a second step, Tgl cross-links $SafA^{C30}$ to fortify pre-assembled complexes of $SafA^{FL}$ and $SafA^{C30}$. During the process, Tgl itself also becomes cross-linked, which eventually limits its activity at the spore surface. The proteins are not drawn to scale. Adapted from Fernandes *et al.*, 2019.

1.8. Aims of this study

The presence of genes with roles in biofilm formation in the genomic signature for sporulation suggests an evolutionary and functional connection between the two processes. The finding that two major biofilm matrix components, TasA and BslA, are also found in association with the spore coat suggests that molecular features of the assembly of the two structures may be shared. One line of possible interactions involves the YabG and Veg proteins, Tgl and SafA. YabG is encoded by a gene found just upstream of *veg*, which has a role in biofilm formation, and this arrangement is conserved among sporeformers (Figure 1.5A). YabG, in turn is required for processing of the substrates that are cross-linked by the Tgl transglutaminase. In this work, we wanted to gather details on the function of these proteins, as a first step towards testing the idea of an evolutionary connection between assembly of the biofilm matrix and the spore coat.

Our specific goals were as follows:

1) To describe the molecular details of the cross-linking of the inner coat morphogenetic protein SafA, a well-characterized substrate of the Tgl transglutaminase;
2) To produce a model of the structural organization of the products of the Tgl-mediated cross-linking of SafA;
3) To begin the structural characterization of Veg;
4) Finally, to test whether Veg has a role in the assembly of the spore coat.

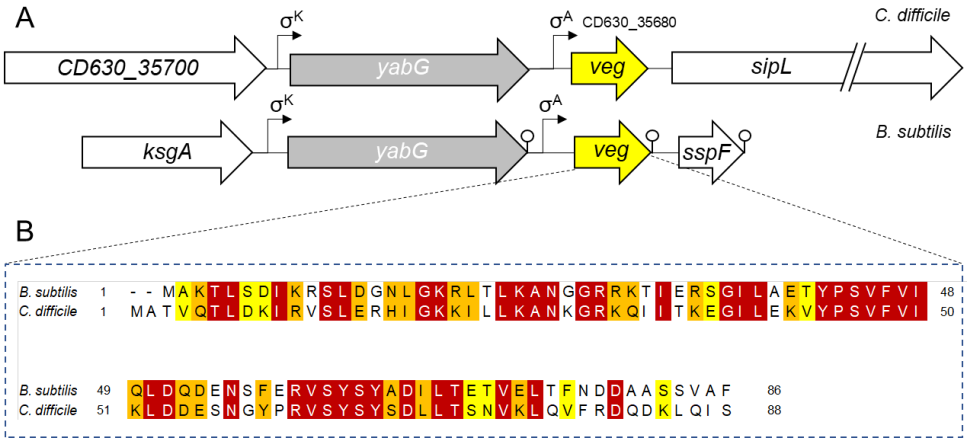

Figure 1.5. Veg is conserved among sporeformers. A: Genomic map of the region surrounding the *veg* gene in the *Clostridioides difficile* genome (upper panel) and in *B. subtilis* (lower panel). The position of *veg* can be seen relative to its neighboring genes; the positioning of *veg* downstream of *yabG* is conserved across sporeformers (not shown). B: Clustal W (https://www.genome.jp/tools-bin/clustalw) alignment of the Veg proteins from *C. difficile* and *B. subtilis*. The red background highlight identical residues. Residues are shadowed in orange or yellow according to the degree of conservation.

2. Materials and Methods

2.1. Microbiological techniques

2.1.1. Bacterial strains and growth conditions

The bacterial strains used in this study are listed in Table 1. *B. subtilis* strains used in this study are congenic derivatives of the wild-type strain MB24. *Escherichia coli* strain BL21(DE3) and its derivatives were used for overproduction of recombinant proteins and *E. coli* strain DH5α was used in cloning experiments and plasmid constructions made throughout this work. The strains were stored at -80°C in 15% (v/v) glycerol. Luria broth (LB) and Luria agar (LA) media were used for routine growth of *B. subtilis* and *E. coli* strains (Table 2). Antibiotics were added for selective growth (Table 3). All the growth media and solutions used in this work are listed Tables 2 and 6, respectively.

Overproduction of proteins in *E. coli*

E. coli strain BL21 (DE3) bearing the plasmids pLOM4, pCF70, pCF68, pBG2 and pBG1 (Table 4), were used for the overproduction of Tgl, $SafA^{FL}$, $SafA^{C30}$, $SafA^{N21}$, and Veg respectively. The strains were plated in LA medium with the corresponding antibiotic (Table 3), from frozen stocks, and grown overnight at 37°C. A colony was then pre-inoculated into 5 ml of LB with antibiotic for 6 h with agitation (150 rpm), and afterwards 1 ml of the pre-inoculum was transferred to 100 ml of auto-induction medium (Table 2; Studier, 2005) and incubated for 18 h at 37°C with agitation being harvested by centrifugation at 3500 x *g* for 6 min and the pellets stored at -80°C, if not used immediately.

2.1.2. Spore production and purification

Sporulation was induced using Difco sporulation medium (DSM; Sigma) (Harwood and Cutting, 1990). Strains were plated from frozen stocks, overnight, at 37°C in LA supplemented with the appropriate antibiotics. They were inoculated in LB medium for 6 h at 37°C with shaking (150 rpm). Afterwards, 1 ml of culture was inoculated in 100 ml of DMS medium, supplemented with 1 mM $Ca(NO_3)_2$, 0.01 mM $MnCl_2$ and 1 µM $FeSO_4$, for 24 h in the same conditions. After 24 h, the culture was centrifuged at 7519 x *g* for 10 min and resuspended in 100 ml of ddH_2O for 48 h. The spore suspension was centrifuged at 7519 x *g* for 10 min and resuspended in 1 ml of 20% (v/v) Gastrografin® (Bayer), and then transferred to a tube with 10 ml of 50% (v/v) Gastrografin®. This suspension was centrifuged at 4 ºC, 7519 x *g* for 30 min. The supernatant was carefully removed, and the pellet washed 5 times in 1 ml of ddH_2O (3 min, 7519 x *g* each wash). Finally, the pellet was resuspended in 500 µL of ddH_2O and spores were quantified by spectrophotometry at 580 nm. The same quantity of spores was resuspended in 20 µL of ddH_2O and 20 µL of 2 X SDS-PAGE loading buffer were added. The samples were boiled for 8 min and 20 µL of each sample loaded into a 15% (v/v) SDS-PAGE gel (Table 6). The resulting gel was incubated for 1h in Coomassie solution (Table 6) and afterwards in destaining solution until the background was clear. Both staining and destaining were done at room temperature and with agitation.

2.2. Molecular biology techniques

2.2.1. Plasmid construction

Inserts for cloning were obtained by PCR. PCR was performed using Phusion® High-Fidelity DNA Polymerase (Thermo Fischer Scientific), and reaction conditions were set based on primer melting temperature, amplified product size and the manufacturer's indications for the DNA polymerase (see primers in Table 5). PCR products were purified using DNA Clean and Concentrator™-5 (Zymo Research). Restriction and modification enzymes were used according to manufacturer's guidelines (Thermo Fischer Scientific). The amplified products were inserted, following purification and cleavage with appropriate restriction enzymes, into the plasmids. The plasmid constructions were sequenced to verify any mutations that could occur during PCR or cloning.

2.2.2. Competence development in *E. coli* and transformation

Competent cells were obtained with the $CaCl_2$ method (Sambrook *et al.*, 2012). A colony of *E. coli* was inoculated in 50 ml of LB medium and incubated at 37°C, with agitation, to an OD_{600} of 0.6.

The culture was centrifuged for 2 min at 5204 x *g* at 4°C. The pellet was resuspended in 25 ml of a cold solution of 0.1M $CaCl_2$. Supernatant was discarded. The suspension was incubated on ice for 1 h and centrifuged again. The supernatant was discarded, and the cell pellet was resuspended in 3.3 ml $CaCl_2$ with 15% glycerol.

Competent cells of the DH5α strain of *E. coli* were transformed as follows: 1 µL of plasmid or 10 µL of ligation mixture were added to 150 µL of competent cells and incubated on ice for 40 min. Transformation was done through heat-shock at 42°C for 90 seconds followed by 2 min on ice. The transformation mixture was incubated in 1 ml of LB medium at 37°C with agitation (150 rpm) for 1-2 h. The culture was centrifuged at 3500 x *g* for 2 min and 1 ml of supernatant removed. Cells in the pellet were resuspended and plated in LA supplemented with the appropriate antibiotic (Table 3). Plates were incubated overnight at 37°C and checked the next day for the appearance of transformant colonies.

For overproduction of proteins, plasmids were transformed into competent cells of strain BL21 (DE3) of *E. coli*, using the same procedure.

2.2.3. Extraction of plasmid DNA from *E. coli*

Single colonies were inoculated in 5 ml of LB with antibiotic and incubated at 37°C with agitation (150 rpm) overnight. 2 ml of cell cultures were pelleted (at 12000 x *g* for 5 min), resuspended in 360 µL of STET buffer, 24 µL of lysozyme (10 mg/ml) and 10 µL of RNaseA (10 mg/ml) and incubated at 37°C for 1 h, for cell lysis. Samples were then boiled (100°C) for 1 min, to denature chromosomal DNA, centrifuged at 12000 x *g* for 10 min and the sediment discarded. To precipitate the DNA, 0.7 volumes of isopropanol were added to each tube and tubes were centrifuged at 12000 x *g* for 45 min at 4°C. The supernatant was then discarded, and the tubes were left to dry at room temperature. The DNA was resuspended in 20 µL of ddH_2O and stored at -20°C. After analysis through agarose gel electrophoresis, the plasmids of interest were extracted and purified using the NZY miniprep kit (NZYtech), used according to manufacturer's guidelines, to obtain highly pure DNA.

2.2.4. Agarose gel electrophoresis

DNA samples were run in 1% (w/v) agarose gels in TAE 1 X buffer at 120 mV. Ethidium bromide 0.014% (v/v) was added to the gels. The Orange G Loading buffer (Table 6) was added to each sample before application into the gels. DNA was visualized under UV light and photographed. The 1kb plus DNA ladder (Invitrogen) was used as molecular weight marker.

2.2.5. Competence development in *B. subtilis* and transformation

B. subtilis was inoculated from a frozen stock (-80°C) to LA supplemented with the appropriate antibiotic and incubated overnight at 37°C. A colony was inoculated into 5 ml of GM1 (Table 2) and incubated for 5 h at 37°C with agitation (150 rpm). 1 ml of the growth was transferred to 4 ml of GM2 (Table 2) and incubated for 2 h at 37°C with agitation (150 rpm) for competence induction. For transformation, 500 µl of competent cells were mixed with 5 µL of genomic DNA or 20 µL of plasmid DNA for 1 h at 37°C with agitation (150 rpm). The transformation mixture was plated in LA supplemented with the appropriate antibiotics. To verify insertion of genes into the *amyE* locus, transformants were plated in LA with 1% starch and tested for amylase presence with Iodine solution (Table 6).

2.2.6. Genomic DNA extraction from *B. subtilis*

B. subtilis was inoculated from a frozen stock (-80°C) to LA plates supplemented with the appropriate antibiotic and incubated overnight at 37°C. The cells were then collected and suspended in 500 µL of SET buffer with 25 µL of lysozyme (10 mg/ml), 12.5 µL of Sarkosyl 10% (v/v), 12.5 µL of Proteinase K (1mg/ml) and 5 µL of RNAseA (10mg/ml), and incubated for 2 h at 37°C. 500 µL of phenol:chloroform:isoamylic alcohol were then added and gently mixed by inversion. The solution was centrifuged at 12000 x *g* for 10 min and 400 µL of the top layer were transferred to a new tube. The bottom layer was discarded. After adding 400 µL of chloroform and gently mixing by inversion, the DNA containing solution was centrifuged at 12000 x *g* for 10 min. 250 µL of the top layer were transferred to a new tube and the bottom layer was discarded. Finally, the DNA solution was dialyzed against ddH_2O using a Milipore membrane filter (Sigma) for 30 minutes and stored at 4 ºC.

2.3. Biochemical techniques

2.3.1. Protein purification by affinity chromatography

For protein purification, the pellets of auto-induced cells were resuspended in 10 ml of the appropriate buffer, depending on the protein [Buffer A for Tgl (Plácido *et al.*, 2008), Lysis buffer for SafA proteins (for $SafA^{FL}$ and $SafA^{C30}$, 200 µL of DTT 1 M were added to the suspension), and French Press buffer for Veg; Table 6] and 0.5µL of Benzonase® were added for every protein, for lysis in the French Press. The cells were then lysed in the French Press at 19000 lb/in^2 into 2 ml tubes containing 20 µL of PMSF 100 mM. The lysate was centrifuged at 12000 g for 30 min at 4°C, the pellet was discarded, and the supernatant kept and applied into a Ni^{2+} Sepharose® High Performance (GE Healthcare) His-trap affinity column. The columns were pre-equilibrated with the appropriate buffers and, after the supernatants were applied, washed with buffer with 10% (v/v) glycerol. Trapped proteins were eluted with buffer containing rising concentrations of imidazole and stored in 1.5 ml microtubes for quantification by the Bradford method. Samples are dialyzed in dialysis buffer (Table 6) overnight at 4°C and then stored at -20°C.

2.3.2. Protein quantification

To quantify the purified proteins, the Bradford method was used. Briefly, 5 µL samples were mixed in 1 ml of diluted (1:5) Bio-Rad Protein Assay Dye Reagent Concentrate (Bio-Rad) inside plastic cuvettes (Bradford, 1976). The samples were quantified by spectrophotometry at 595 nm, where the absorbance was directly correlated to the quantity of protein in each sample, through the following equation, where y is the absorbance and x is the amount of protein in the 5 µL sample:

$$\text{Equation 2.1: } y = 0{,}0656x$$

2.3.3. Cross-linking assays

For cross-linking reactions, Tgl was incubated with each form of SafA, in 0.1M Tris-HCl, pH 8.0, for 2h at 37°C. In a first series of assays the concentration of Tgl was kept constant, at 4 µM, while the concentration of $SafA^{FL}$, $SafA^{C30}$ and $SafA^{N21}$ varied. In another series of assays the opposite was done where the concentration of Tgl varied while the concentration of $SafA^{FL}$, $SafA^{C30}$ and $SafA^{N21}$ was constant, at 12 µM. The assays were performed in a total volume of 100 µL where 5 µL samples were collected at 11 different time points. 2µL of SDS-PAGE reducing loading Buffer 10 X (Table 6) and 13µL of 0.1M Tris-HCl pH 8.0 were added to reaction samples, in order to stop the reaction, for a total volume of 20 µL. For assays with reducing and non-reducing SDS-PAGE loading buffer (Figure 3.7) samples were also boiled for 1 min to stop the reaction.

2.3.4. SDS-PAGE

Purified proteins, products of the cross-linking reactions, and spore coat extracts were analyzed by SDS-PAGE. Samples were loaded on 12.5%, for purified proteins, 10%, for cross-linking reactions, or, for coat extracts, 15% SDS-PAGE gels. The Precision Plus Protein™ Standard Unstained (Bio-Rad) ladder was used. 20 µL of each sample mixture were applied to the gels after incubation for 5 min at 100°C. Following electrophoretic resolution, the gels were incubated for 1h in Coomassie solution (Table 6) and then in destaining solution until the background was clear. Both staining and destaining were done at room temperature with agitation.

2.3.5. Western blot

Proteins were electrophoretically transferred from SDS-PAGE gels to nitrocellulose membranes (Supported Nitrocellulose, 0.45 µm; BioRad) at 100 V for 60 min using transfer buffer (Table 6). The membrane was incubated in 20 ml of blocking solution (5% milk in PBS-T) for 1 h with agitation. The blocking solution was then removed, and the antibody (anti-CotB and anti-CotG, diluted 1:300 and 1:10000, respectively) was added in 10 ml of PBS-T with 0.5% milk. The membrane was incubated overnight with the antibody solution at 4°C without agitation. The antibody solution was then discarded, and the membrane washed 3 times with PBS-T for 10 min, each wash. The anti-rabbit secondary

antibody was added, at a 1:10000 dilution in 10 ml PBS-T with 0.5% milk. The membrane was incubated for 30 min at room temperature with agitation. Finally, the membrane was washed 3 more times in PBS-T for 10 min each wash. Proteins were detected using the detection solution ("SuperSignal West Pico Chemiluminescent"; Thermo Scientific), exposed to a chemiluminescence film (Amersham Hyperfilm™ ECL, GE Healthcare) in the dark, and then developed and fixed (Kodak).

3. Results

Cross-linking of SafA and its forms by Tgl

Tgl is a bacterial transglutaminase that catalyzes the formation of ε-(γ-glutamyl)lysil bonds thereby covalently linking proteins to one another (Kobayashi *et al.*, 1996; Zilhão *et al.*, 2005). Recently, the recruitment and regulation of Tgl activity were characterized and a better understanding of its role in the stabilization of protein complexes within the spore coat was possible (Fernandes *et al.*, 2015; Fernandes *et al.*, 2019). In particular, it was shown that Tgl is recruited to the surface of the developing spore by its pre-assembled substrates, with $SafA^{FL}$ and $SafA^{C30}$ making the most important contibutions (Fernandes *et al.*, 2015; Fernandes *et al.*, 2019). These studies have shed light on how an ancient bacterial transglutaminase functions and interacts with its substrates during the assembly of a macromolecular structure, the spore coat (Lorand and Graham, 2003; Fernandes *et al.*, 2015; Fernandes *et al.*, 2019). The model of recruitment of Tgl by its substrates raised the possibility that the activity of the transglutaminase could itself be dependent on the local substrate concentration. Fernandes and co-authors showed that for $SafA^{C30}$ an increase in substrate concentration led to a decrease in Tgl activity *in vitro* (Fernandes *et al.*, 2019). Since *in vivo* at least three forms of SafA are present (Ozin *et al.*, 2001a), $SafA^{FL}$, $SafA^{N21}$ and $SafA^{C30}$, we wanted to extend these studies and characterize the activity of Tgl *in vitro* with all forms of SafA. We wanted to investigate whether all forms of SafA could be cross-linked by Tgl *in vitro* and if so, whether the inverse dependency on substrate concentration for enzyme activity was a property of all forms.

3.1. Overproduction and purification of Tgl and SafA substrates

Plasmids for the overproduction of both Tgl-His_6 (pLOM4) and $SafA^{FL}$-His_6 (pCF70) have been described and both proteins have been overproduced using an auto-induction method and Ni^{2+}-affinity chromatography (Plácido *et al.*, 2008; Fernandes, 2014).

Tgl-His_6 production was auto-induced in a 100 ml LB culture of *E. coli* BL21(DE3) (pLOM4) and cells were mechanically lysed in a French Press. The enzyme was then purified and eluted in 5 fractions from the His-trap® column equilibrated with Buffer A (Table 6) and its purity verified by SDS-PAGE (Figure 3.1A). Tgl-His_6 was eluted and was nearly pure at 150 mM imidazole (Figure 3.1A). The protein was eluted in five fractions of 1 ml, and was nearly pure and most concentrated in fraction 4, for which a concentration of 40 μM was obtained as calculated using the Bradford method (Bradford, 1976). After purification and quantification Tgl was dialyzed in Dialysis buffer and either used directly in enzymatic essays or frozen, for later use, at -20°C.

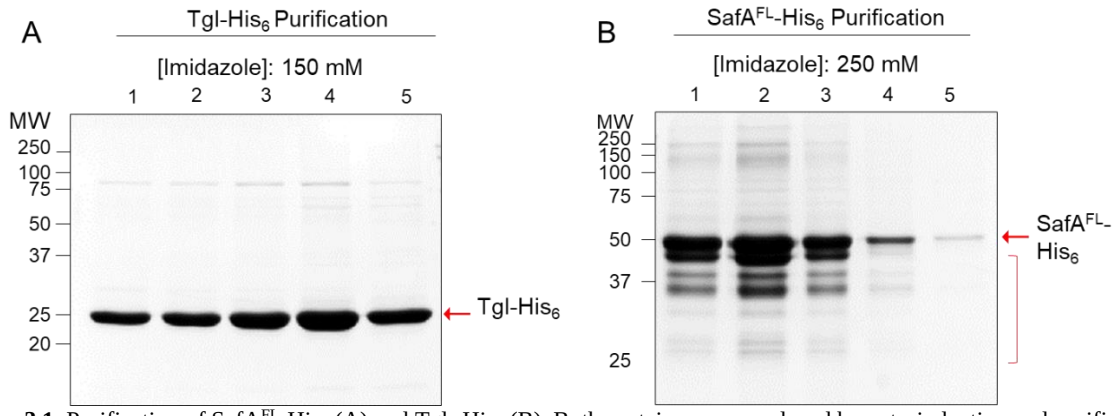

Figure 3.1. Purification of $SafA^{FL}$-His_6 (A) and Tgl- His_6 (B). Both proteins were produced by auto-induction and purified by Ni^{2+}-affinity chromatography on a His-trap column equilibrated with the appropriate buffer (Table 6). $SafA^{FL}$ was eluted at 250 mM of imidazole, and Tgl was eluted at 150 mM of imidazole. Red arrows mark the position of $SafA^{FL}$-His_6 or Tgl-His_6.The parentheses in B indicates possible degradation forms of $SafA^{FL}$-His_6.

In the case of SafA^{FL}-His$_6$, 200 ml of LB cultures of *E. coli* BL21(DE3) (pCF6) were auto-induced and lysed in a French Press. SafA^{FL}-His$_6$ purification was done as described before for its short form SafA^{C30} (Fernandes *et al.*, 2019). Proteins were eluted with Lysis Buffer with 250 mM imidazole and dialyzed in Dialysis buffer with 2 mM DTT to prevent disulfide bond formation. Five fractions of 1 ml were collected from the column and samples analyzed by SDS-PAGE (Figure 3.1B). SafA^{FL}-His$_6$ appeared essentially pure and in fraction 2, its concentration was of 14.1µM, as estimated using the Bradford method (Bradford, 1976).

Finally, for the purification of SafA^{N21}-His$_6$ we first constructed a plasmid with the *safA*^{N21} coding region fused to a *his-tag* (pBG2; Table 4). The 5' end of the *safA* open reading frame, ending in codon 164, corresponding to the N21 moiety of SafA (see Figure 1.3) was PCR-amplified with primers safA490Fw and safA980Rv (Table 5), the PCR products digested with XhoI and NcoI, and inserted between the same sites of pET16b to yield pBG2. In pBG2, expression of a *safA*^{N21}*-his$_6$* fusion is under the control of a PT7$_{lac}$ promoter. The plasmid was introduced into *E. coli* BL21(DE3). SafA^{N21}-His$_6$ was over-produced by auto-induction in LB cultures of 100 ml. Purification was performed by Ni^{2+}-affinity chromatography according to the methods described by Fernandes *et al.* (2019), for the SafA^{C30} protein. SafA^{N21}-His$_6$ accumulated mostly in the soluble fraction of a whole cell extract (Figure 3.2). The protein eluted nearly pure at 100 mM imidazole, as assessed by SDS-PAGE (Figure 3.2). Five fractions of 1 ml were collected for each of the imidazole concentrations. SafA^{N21}-His$_6$ showed the highest concentration, 50 µM, as estimated by the Bradford method (Bradford, 1976), in the second elution ration of 100 mM imidazole.

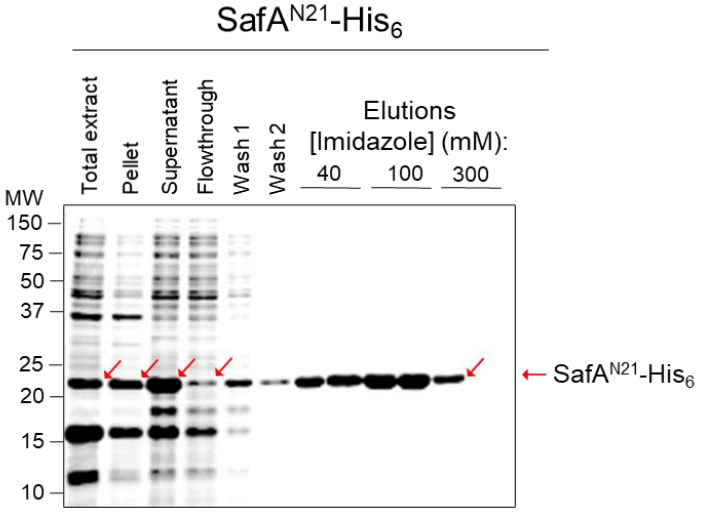

Figure 3.2. Purification of SafA^{N21}-His$_6$. The SafA^{N21}-His$_6$ protein was overproduced using an auto-induction regime and accumulated mostly in the soluble fraction ("supernatant" in the figure) of a whole cell extract. The supernatant was applied to a His-trap® column for purification of SafA^{N21}-His$_6$ by Ni^{2+}-affinity chromatography. The column was equilibrated with the appropriate buffer (Table 6), washed and SafA^{N21}-His$_6$ eluted with the indicated imidazole concentrations. The SafA^{N21}-His$_6$ eluted nearly pure at 100 mM of imidazole. Red arrows mark the position of SafA^{N21}- His$_6$.

The fractions showing the highest concentrations of Tgl-His$_6$, SafA^{FL}-His$_6$ and SafA^{N21}-His$_6$ were kept for further studies. From now on, the His$_6$ tag will be omitted from the designation of the proteins or simplicity.

3.2. Tgl activity decreases with increasing concentration of the SafA^{FL} substrate

As mentioned above, previous studies showed that, for SafA^{C30}, an increase in substrate concentration led to a decrease in Tgl activity *in vitro* (Fernandes *et al.*, 2019). We now extend these studies starting with the characterization of Tgl activity *in vitro* with the substrate SafA^{FL}. The full-length form of SafA was incubated with Tgl and enzymatic assays were performed, similarly to the assays with SafA^{C30}, as described in the Materials and Methods section. Samples were taken at different time points and the reaction stopped for those samples with the addition of SDS-PAGE reducing loading buffer 10 X; to monitor the progression of the cross-linking reaction, samples were analyzed by SDS-PAGE (Figure 3.3A). The disappearance of SafA^{FL} bands was quantified and plotted to better visualize

Tgl activity along time. We conducted this analysis for four different concentrations of SafA ^{FL} (12 µM, 6 µM, 3µM and 1.5 µM). We found that lower concentrations of SafA^{FL} led to a faster formation of higher molecular weight forms [(SafA^{FL})$_n$] and therefore, to a faster disappearance of its monomeric form (Figure 3.3B). For higher concentrations of the substrate, the opposite was true. Thus, Tgl behaved with SafA^{FL} as it did with SafA^{C30}, *i.e.*, the activity of Tgl decreased with increasing concentration of the SafA^{FL} substrate (Figure 8B).

Fernandes *et al.* (2019) also established that, during the cross-linking reaction, Tgl becomes part of the cross-linked products. With this in mind, we took an additional measurement and plotted the disappearance of the Tgl bans over time (Figure 3.3C). We found that also with the SafA^{FL} substrate, Tgl disappears over time, suggesting that it becomes part of the cross-linked products.

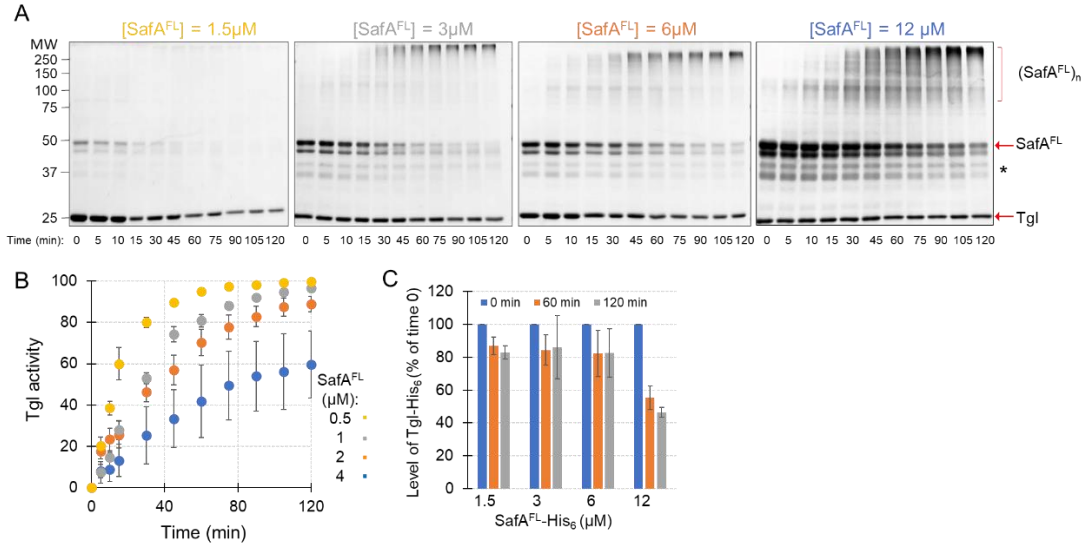

Figure 3.3. Reaction of Tgl with SafA^{FL} at different concentrations of SafA^{FL}. SafA^{FL} was incubated, at four different concentrations with Tgl at a fixed concentration (4 µM), in order to test the activity of Tgl at different concentrations of substrate. A: Samples were resolved by SDS-PAGE (10%) and stained with Coomassie solution. B: Tgl activity was monitored by SDS-PAGE and normalized as a function of the loss of SafA^{FL} over time, as measured using Image J. Red arrows mark the position of SafA^{FL} or Tgl. The asterisk denotes likely degradation products of SafA^{FL}. The parenthesis shows the position of cross-linked products of SafA^{FL} [(SafA^{FL})n]. C: decrease in the level of Tgl, at a fixed concentration of 4 µM, at the indicated times after the onset off the reaction (time 0), as a function of the concentration of SafA^{FL}.

3.3. SafA^{N21} is inefficiently cross-linked by Tgl

We went on to examine the activity of Tgl towards SafA^{N21}. This form of SafA, of 21 kDa, carries a LysM domain in its N-terminal end and a linker that interacts with SpoVID, through its region A (Figure 1.3B) (Costa *et al.*, 2006; Pereira *et al.*, 2019). The LysM domain was shown to act as both a protein-protein and a protein-peptidoglycan interaction module (Pereira *et al.*, 2019). To study the activity of Tgl towards SafA^{N21}, we built an *E. coli* overexpression system for this fragment of SafA and purified the protein by Ni^{2+}-affinity chromatography (Figure 3.2). Purified SafA^{N21} was incubated with Tgl for 2 h, and samples were taken at different time points for analysis, as described above for SafA^{FL} and SafA^{C30}. The samples were mixed with 10x SDS-PAGE reducing loading buffer to stop the reaction. Analysis of the reaction products by SDS-PAGE showed that SafA^{N21} disappeared over time, but at a much slower rate than SafA^{FL} or SafA^{C30} (Figure 3.3A and 3.4A). This was also seen by quantification of SafA^{N21} bands in Coomassie stained gels, where we measured the disappearance of the substrate over time at four different concentrations of the substrate protein SafA^{N21} (12 µM, 6 µM, 3 µM, 1.5 µM) (Figure 3.4B).

The activity of Tgl seems to decrease with increasing concentration of the SafA^{N21} substrate, as for SafA^{FL} or SafA^{C30} (Figure 3.4A and B). However, one difference that separates SafA^{N21} from SafA^{FL} or SafA^{C30} is the efficiency with which Tgl cross-linked SafA^{N21}. In the case of SafA^{N21}, the band corresponding to its monomeric form (in the 21 kDa region of the gel) did not disappear nearly as much and to the same extent as SafA^{FL} or SafA^{C30}. An additional band just below the SafA^{N21} band is visible

in the SDS-PAGE gels of the assays, possibly a cross-linked form of SafA^{N21} folded onto itself. However, this remains to be verified. Strikingly, SafA^{N21}, which encompasses the LysM domain and a linker between this domain and the SafA^{C30} part of the protein (Figure 3.4C), contains both lysine and glutamine residues. Possibly, the structure of SafA^{N21} is such that no glutamine and lysine residues are available to reach the active site of Tgl, which lies inside a narrow tunnel (Fernandes et al., 2015), making SafA^{N21} a poor substrate for Tgl. It is also possible that SafA^{N21} is cross-linked to another protein *in vivo*, or to the cortex peptidoglycan (Fernandes et al., 2018; Fernandes et al., 2019). In any case, the inefficient cross-linking of SafA^{N21} by Tgl, as compared to the full-length or SafA^{C30} forms of the protein, has important implications for the structural organization of SafA complexes within the coat (see the Discussion).

In all, our studies show that the activity of Tgl depends on and decreases with increasing SafA^{N21} concentration and that the cross-linking reaction is significantly less efficient than for SafA^{FL} or SafA^{C30}.

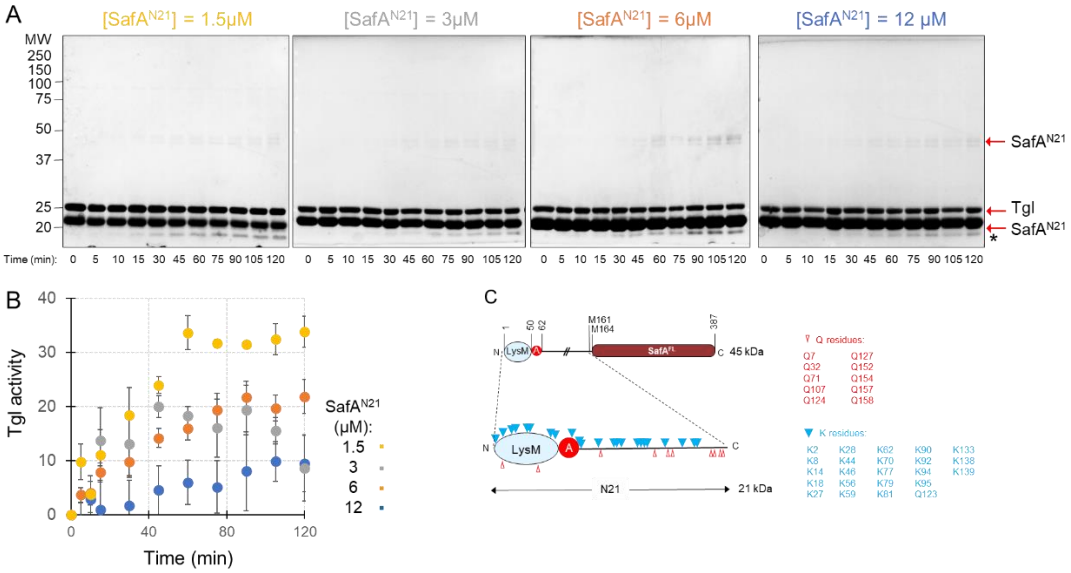

Figure 3.4. Reaction of Tgl with SafA^{N21} at different concentrations of SafA^{N21}. SafA^{N21} was incubated, at four different concentrations with Tgl at a fixed concentration (4 µM), in order to test the activity of Tgl at different concentrations of substrate. A: Samples were resolved by SDS-PAGE (10%) and stained with Coomassie solution. B: Tgl activity was monitored by SDS-PAGE and normalized as a function of the loss of SafA^{N21} over time, as measured using Image J. Red arrows mark the position of SafA^{N21} or Tgl. The asterisk denotes likely degradation products of SafA^{N21}. C: Structural organization of SafA^{N21}, the N-terminal form of SafA that includes the localization signals (LysM and region A) and a linker at its C-terminal end. The glutamine (10, in red) and the lysine (23, in blue) residues that could work as Q (donor) or K (acceptor) in Tgl-catalyzed cross-linking are indicated.

3.4. Tgl activity towards SafA^{FL} increases with increasing enzyme concentration

As we have shown above, the activity of Tgl depends inversely on the substrate concentration, for all forms of SafA that were studied, although this was less evident for SafA^{N21} which is a poor substrate for the enzyme (see above). This property, presumably resulting from the cross-linking of Tgl itself, could have important consequences for the control of Tgl activity *in vivo*. It seems possible, for example, that a high local concentration of Tgl could overcome the substrate-dependent inhibition of Tgl activity. To test this possibility and to obtain a picture of the dependency of enzyme activity on the concentration of Tgl, an additional series of assays was performed for the cross-linking reaction between Tgl and its SafA substrates SafA^{FL} and SafA^{N21} where the concentration of Tgl varied (4 µM, 2 µM, 1 µM, 0.5 µM) while the substrate concentration was fixed at 12 µM.

Similarly, to the previous assays, Tgl was incubated with each of its substrates for 2 h. Samples were taken at the same time point, the reaction stopped, and the products of the reaction were analysed by SDS-PAGE (Figure 3.5A and 3.6A). The disappearance of SafA^{FL} (Figure 3.5B) and SafA^{N21} (Figure 3.6B) was quantified and plotted as a function of time.

The results show that Tgl activity towards SafA^{FL} increases with increasing concentration of the enzyme (Figure 3.6B). In the case of SafA^{N21} an increase of Tgl activity with enzyme concentration was not evident (Figure 3.6B) most likely because this form of the SafA is not a good substrate for Tgl.

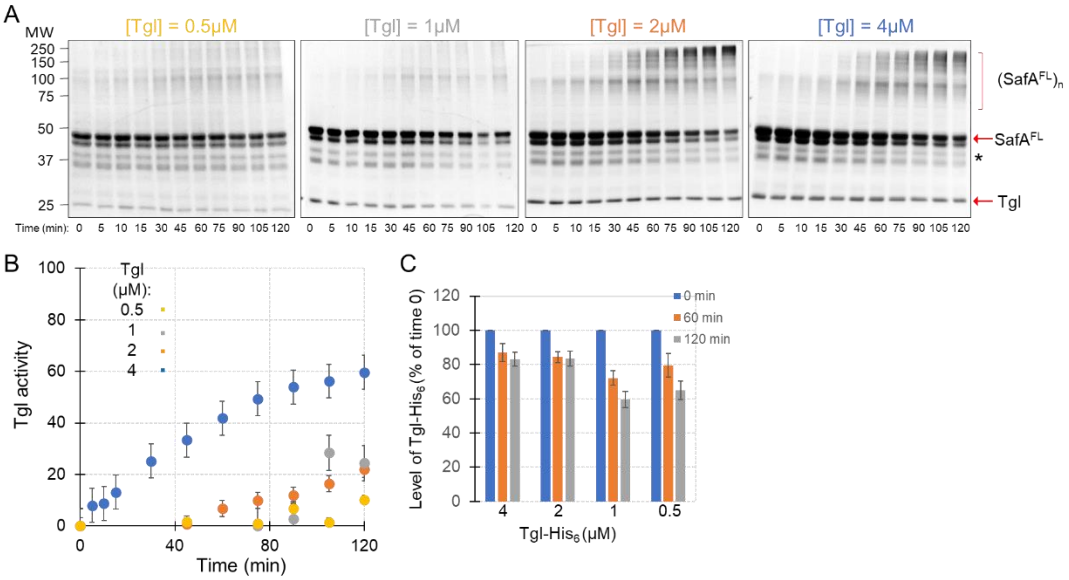

Figure 3.5. Reaction of Tgl with SafA^{FL} at different concentrations of Tgl. SafA^{FL} was incubated at a fixed concentration (12 μM) with Tgl at four different concentrations, in order to test the activity of Tgl at different concentrations of enzyme, but also the depletion of enzyme with time. A: Samples were resolved by SDS-PAGE (10%) and stained with Coomassie solution. B: Tgl activity was measured and normalized as a function of the loss of SafA^{FL} with time with time, through SDS-PAGE analysis; images were analyzed using Image J. Red arrows mark the position of SafA^{FL}or Tgl. The asterisk denotes likely degradation products of SafA^{FL}. The parenthesis shows the position of cross-linked products of SafA^{FL} [(SafA^{FL})$_n$]. C: decrease in the level of Tgl at the indicated times after time 0, as a function of the concentration of Tgl, for a fixed concentration of SafA^{FL} (12 μM).

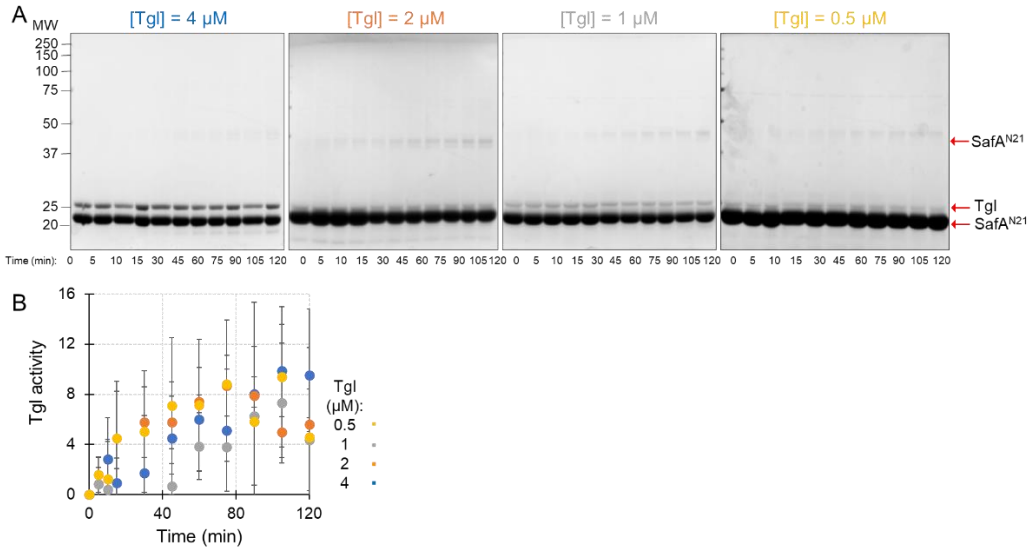

Figure 3.6. Reaction of Tgl with SafA^{N21} at different concentrations of Tgl. SafA^{N21} was incubated at a fixed concentration (12 μM) with Tgl at four different concentrations, in order to test the activity of Tgl at different concentrations of enzyme, but also the depletion of enzyme with time. A: Samples were resolved by SDS-PAGE (10%) and stained with Coomassie solution. B: Tgl activity was measured and normalized as a function of the loss of SafA^{N21} with time with time, through SDS-PAGE analysis; images were analysed using Image J. Red arrows mark the position of SafA^{N21} or Tgl.

3.5. SafA^{C30} forms disulfide cross-linked oligomers that are cross-linked by Tgl

As mentioned above, the cross-linking reaction catalyzed by Tgl needs the recognition of a glutamine residue (the donor) and a lysine residue (the acceptor) (Fernandes, 2014; Fernandes *et al.*, 2019). While SafA has 25 lysine residues, only two of them are present in the SafA^{C30} moiety of this protein, at positions 177 and 318, with the numbering starting from methionine 1 of the full-length protein (Figure 3.7B). We asked whether any of these lysines, or both, would have a role in the Tgl-mediated cross-link of SafA^{C30}. To examine the role and relative importance of the K177 and K318 residues in the cross-linking of SafA^{C30}, each of the lysine residues alone and the two in combination were substituted by alanine residues. The SafA^{C30}-His$_6$ proteins with the alanine substitutions were overproduced and purified from *E. coli* as described previously (Fernandes et al., 2019).

The different forms of SafA^{C30} were incubated with Tgl for 2 h at 37°C and equal volumes of the samples taken along time were mixed with either a reducing or a non-reducing SDS-PAGE loading buffer. Then, the samples were analyzed by SDS-PAGE and the gels stained with Coomassie (Figure 3.7A). SafA^{C30} incubated under the same conditions but in the absence of Tgl served as a control to establish the behavior of SafA^{C30} in the absence of Tgl and in the presence or absence of reducing agents in the loading dye.

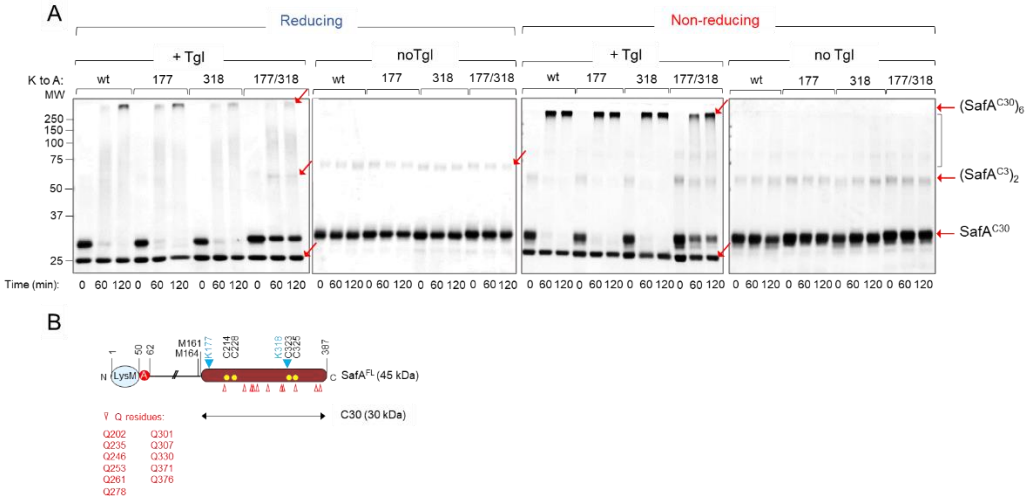

Figure 3.7. Reactions of Tgl with SafA$^{C30(K177A)}$, SafA$^{C30(K318A)}$ and SafA$^{C30(K177/318A)}$. The reactions were stopped at each time point under reducing and non-reducing conditions. The concentrations of all SafA^{C30} and Tgl were fixed at 12 µM for SafA^{C30} and 4 µM for Tgl. A: Samples were resolved by SDS-PAGE (10%) and stained with Coomassie solution. Red arrows mark the position of Tgl and the different forms of SafA^{C30}. The parenthesis shows the position of cross-linked forms of SafA^{C30}, above the dimer. B: Structural organization of SafA^{C30}, the C-terminal form of SafA. The glutamine (11, in red) and the lysine (2, in blue) residues that could work as Q donor or K acceptor substrates in Tgl-mediated cross-linking are indicated. The four cysteine residues present in SafA^{C30} are indicated as yellow dots.

When analyzed under non-reducing conditions, Tgl cross-linked most of SafA$^{C30(wt)}$ in a period of 120 min, resulting in a marked reduction in the SafA^{C30} monomeric form concomitant with the formation of a species that migrates above the 250 kDa marker, presumably an hexamer, (SafA^{C30})$_6$ (Figure 3.7A) (Fernandes et al., 2018). Most of SafA$^{C30(K177A)}$ and SafA$^{C30(K318A)}$ were also cross-linked by Tgl in 120 min; in contrast the cross-linking of SafA$^{C30(K117/318A)}$ was slightly reduced (Figure 3.7A). These observations suggest that the (SafA^{C30})$_6$ species is formed through or stabilized Tgl-mediated cross-linking and that K177 and K318 contribute to the cross-linking. In the absence of Tgl, formation of the(SafA^{C30})$_6$ species was not observed but a possible SafA^{C30}dimer, (SafA^{C30})$_2$, accumulated for the wt and mutant forms of SafA^{C30} under both reducing and non-reducing conditions (Figure 3.7A). This dimer is stable enough to resist the denaturing conditions of SDS-PAGE, consistent with the strong self-interaction of SafA^{C30} previously reported (Ozin *et al.*, 2001b).

When the same samples were analyzed under reducing conditions, however, the accumulation of (SafA^{C30})$_6$ is markedly reduced for both the WT and K177A forms of SafA^{C30} and even more for the K318A and K177A/K318A forms (Figure 3.7A). This suggests that formation of (SafA^{C30})$_6$ involves mainly disulfide bonds and that Tgl exerts a "spotwelding" activity in cross-linking and stabilizing (SafA^{C30})$_6$. One has to assume that the doubly cross-linked (SafA^{C30})$_6$, via disulfide and ε-(γ-

glutamyl)lysil bonds survives the reducing conditions of SDS-PAGE. This stabilization effect masks the differential effect of the K177A and K318A substitutions, but the analysis under reducing conditions suggests that residue K138 makes a more important contribution to the Tgl-mediated cross-linking of $(SafA^{C30})_6$.

Role of Veg during Sporulation

Veg, the product of the *veg* gene, is highly conserved among spore-formers; *veg* is known to influence both biofilm formation and spore assembly and germination, and its genomic position downstream of the *yabG* gene is also conserved among sporeformers (see Figure 1.5) (Fukushima *et al.*, 2003; Lei *et al.*, 2013). Since YabG is a spore coat-associated protease required, in *B. subtilis*, for the processing of the substrates that are subsequently cross-linked by Tgl (Kuwana *et al.*, 2006), we considered the possibility that Veg could somehow link the two processes, biofilm formation and spore coat assembly.

3.6. Overexpression of *veg* from the *gerE* promoter impairs assembly of the spore coat

A phenotypical analysis of *veg* mutants could provide further insights into the function of Veg during these two developmental processes of *B. subtilis*. Previous studies showed that overproduction of Veg stimulated biofilm formation via SinR repression (see above; Lei *et al.*, 2013). In conditions that induce sporulation, a Veg-GFP fusion protein was detected in vegetative cells and in sporulating cells, but the fluorescence of mother cells disappeared completely late in the process. The effect of overproduction of Veg at a final stage of sporulation has not yet been tested. Here, several mutants were constructed to study the role of Veg during sporulation.

We first placed the *veg* gene under the control of the promoter for the *gerE* gene (P_{gerE}), so that *veg* expression would take place in the mother cell, during the late stages of sporulation (Cutting et al., 1989; Steil et al., 2005; Eichenberger et al., 2004). To do this, we used genomic DNA from strain MB24 of *B. subtilis* to PCR-amplify the *veg* open reading frame using the oligonucleotide primers vegbs500fwd and vegbs768fwd and P_{gerE} using primers ger1D and ger480revvegbs. The two resulting DNA fragments were fused by SOE-PCR (splicing by overlap extension PCR) using ger1D and vegbs768rev as the forward and reverse primers, respectively. The amplified fused product, named P_{gerE}-*veg* was digested with BamHI and HindIII and inserted between the same sites of pDG364 to yield pBG6 and of pMK3 to produce pBG7. pDG364 is an integration vector that allows transfer of DNA fragments in single copy, by means of a marker replacement event (double cross-over) to the *amyE* locus of *B. subtilis*. pMK3, in turn, is a multicopy plasmid (Sullivan *et al.*, 1984).

Several strains were then constructed (see Table 1). Strains MB24 (wt) and MSB180 (a congenic derivative of MB24 carrying a Δ*veg*::*erm* mutation, hereinafter termed Δ*veg* for simplicity) were transformed with pBG6 to yield strains MSB183 (wt *amyE*::P_{gerE}-*veg*) and MSB182 (Δ*veg*, *amyE*::P_{gerE}-*veg*), respectively. Transformation of MB24 with the empty pMK3 vector yielded strain MSB185 (wt pMK3) while introduction of pMK3 into MSB180 produced strain MSB187 (Δ*veg* pMK3). Finally, pBG7 was introduced into MB24 to give MSB184 and into MSB180 to produce MSB186 (Δ*veg* pBG7).

Sporulation of the various strains was induced by growth and exhaustion in DSM and spores were density gradient-purified, as described in the Material and Methods section, 24h after the onset of sporulation. The coat proteins were extracted from the purified spores and electrophoretically resolved on 15% SDS-PAGE gels. The gels were then stained with Coomassie and the pattern of extractable coat proteins analyzed. The spore coat extracts prepared from the wt or the wt/*amyE*::P_{gerE}-*veg* did not differ significantly (Figure 3.8A). Likewise the pattern of extractable coat proteins from spores of the Δ*veg* and Δ*veg*/*amyE*::P_{gerE}-*veg* strains did not differ, nor did the pattern between the two groups of strains (Figure 3.8A). We conclude that disruption of the *veg* gene does not significantly impact the composition of the spore coat. Moreover, expression of *veg* from the P_{gerE} promoter in single copy, and therefore in the mother cell during the late stages of sporulation, also does not seem to cause any gross perturbations on the coat assembly process.

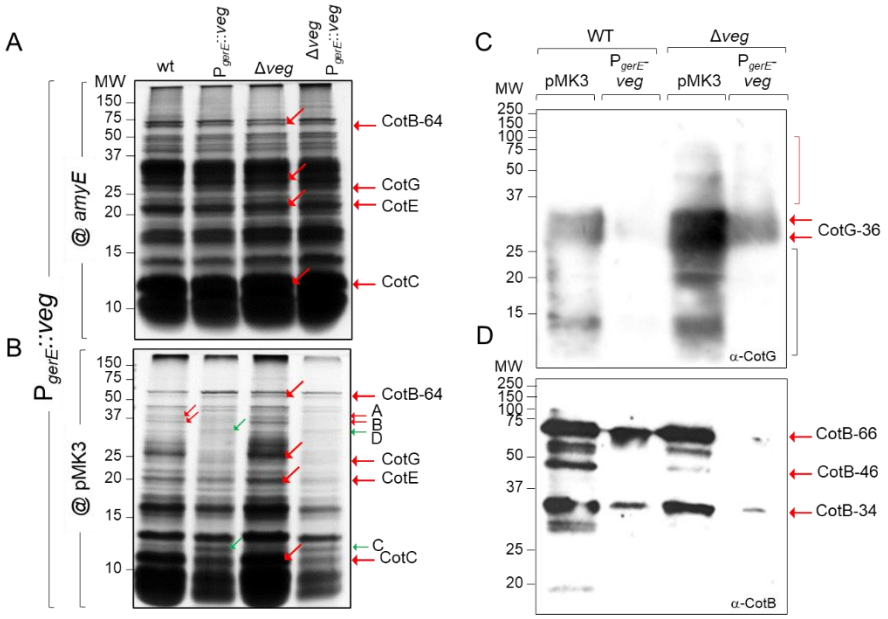

Figure 3.8. *B. subtilis* Veg impacts spore coat assembly. Spore coat protein pattern under Veg expression (A) and overexpression (B) in late stages of sporulation. A: Expression of *veg* under control of P_{gerE}, inserted in the *amyE* locus in the wild-type (lane 2) and in the *veg* mutant (lane 4) and its controls (lanes 1 and 3, respectively). B: Overexpression of *veg* under control of P_{gerE}, in pMK3 vector, in wild-type (lane 2) and in the *veg* mutant (lane 4) and its controls, with an empty pMK3 (lanes 1 and 3, respectively). C and D: Western blot detecting the presence of CotG (C) and CotB (D) in spore coat extracts of wild-type (MB24), *veg* (MSB180) and in the P_{gerE} -*veg* strains (MSB184 and MSB186). Overexpression of *veg* was obtained with the P_{gerE}::*veg* allele inserted into pMK3, a multicopy vector, transformed into *B. subtilis* cells. Red and green arrows show the position of the indicated species. The red and black parenthesis in panel D indicate mutimeric or possible processed forms of CotG, respectively.

We then examined the effects of the presence of the P_{gerE}-*veg* in multicopy on the composition of the coat. Coat protein extracts were prepared from spores of the wt/pMK3 and wt/pBG7 strains and in parallel, from spores of the Δ*veg*/pMK3 and Δ*veg*/pBG7 strains. Strikingly, we found that several species were significantly reduced or absent in spores of the wt/pBG7 and Δ*veg*/pBG7, *i.e.*, bearing the multicopy P_{gerE}-*veg* allele in comparison to their counterparts bearing the empty plasmid (Figure 3.8B). In spores of the wt/pBG7 strain, bands of about 66, 36, 46 (labeled as A) and 40 kDa (marked as B) were greatly reduced compared to the empty vector strains. In contrast, bands of about 37 (labeled as D) and 12 kDa (labeled C) appeared more represented in the extracts prepared from the strains (wt or Δ*veg*) bearing the P_{gerE}-*veg* multicopy allele (Figure 3.8B). Bands A to D were excised and subject to mass spectrometry analysis. While band B could not be identified, only one species was found for each of the bands A, B and D, as follows: A corresponds to YdhD, a peptidoglycan hydrolase with a predicted mass of 46.8 kDa; band C corresponds to YjdH, a protein of unknown function with a predicted mass of 15.2 kDa; D corresponds to a putative L-cysteine binding protein of an ABC-type transporter, YtmK, with a predicted mass of 32.3 kDa (Kodama *et al.*, 2000; Burguière *et al.*, 2005).

The 66 and 36 kDa bands could correspond to the abundant CotB-66 and CotG-36 species, as suggested by earlier work (Freitas et al., 2020). To verify this possibility, we used immunoblot analysis with anti-CotB and anti-CotG polyclonal antibodies of established specificity (Freitas *et al.*, 2020). CotG is usually detected in Coomassie-stained gels as a major species of about 36 kDa (Zilhão *et al.*, 1999; Freitas *et al.*, 2020). Immunoblot analysis, however, shows that CotG is also present as multimeric species above the 36 kDa species, and also as bands below the 36 kDa band (Zilhão *et al.*, 1999; Freitas *et al.*, 2020). In our analysis, CotG is detected in spores of the wt/pMK3 strain as two main species around 36 kDa and as several other bands below this position in the gel (Figure 3.8C). In spores of the Δ*veg*/ pMK3 strain, these same species are also detected, but appear more abundant or more extractable, and at least two species above the 36 kDa marker are also detected (Figure 3.8C). In contrast, the level and/or extractability of CotG in spores of the wt or Δ*veg* expressing P_{gerE}-*veg* from pBG7 is lower than in the counterpart strains carrying pMK3 (Figure 3.8D). CotG is required for the conversion of CotB from a form with 46 kDa (CotB-46) to a form of 66 kDa (CotB-66) (Zilhão *et al.*, 1999). CotB-66 is a

hyperphosphorylated form of CotB-46, that is formed through the action of the CotH kinase in the presence of CotG (Freitas *et al.*, 2020). The level and/or extractability of CotB-66 in spores of the wt or Δ*veg* expressing P_{gerE}-*veg* from pBG7 is also lower than in the counterpart strains carrying pMK3 (Figure 3.8D).

In all, our study provides the first clear phenotype associated with *veg* during sporulation, and more specifically, during assembly of the spore coat. It remains to be tested at what level Veg exerts its effects on the accumulation and/or assembly of the coat proteins it seems to control.

3.7. Overproduction and purification of *C. difficile* Veg

The way Veg functions in biofilm formation or sporulation is not known and to begin characterizing the protein, we PCR-amplified the *veg* gene of *C. difficile* and introduced it into a plasmid for the overproduction of a Veg-His_6 protein (*C. difficile* Veg is highly similar to its *B. subtilis* homologue and currently the subject of intense research in the group; Figure 1.5B). The PCR product was digested with XhoI and NcoI and inserted between the same sites of pET16b. The resulting plasmid was introduced into *E. coli* BL21(DE3) for overexpression of the *veg* gene and overproduction of Veg-His_6. Auto-induction for 18 h was used for the overproduction of Veg-His_6 accumulated to significant levels in the soluble fraction as judged by analyzing the supernatant and pellet of a whole cell lysate following breakage of the cells in a French pressure cell. The protein was then purified to nearly homogeneity by Ni^{2+}-affinity chromatography in an His-trap® column equilibrated with French Press buffer. Veg-His_6 was eluted with Start buffer at increasing concentrations of imidazole. The purification process was verified by SDS-PAGE; Veg-His_6 was recovered as a single species of about 11 kDa, close to the expected mass of the protein (10.5 kDa), at an imidazole concentration of 100 mM (Figure 3.9). The successful overproduction and purification of Veg-His_6 paves the way for further studies of the protein.

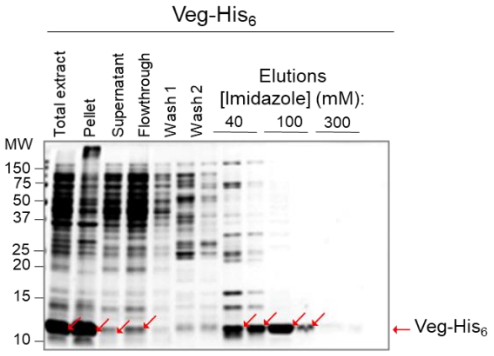

Figure 3.9. Purification of *C. difficile* Veg. Veg-His_6 was overproduced using an auto-induction regime. Veg-His_6 accumulated mostly in the soluble ration of a whole cell extract as indicated (lane labelled "supernatant"). The protein was applied to a His-trap® column equilibrated with French Press buffer (Table 6) for Ni^{2+}-affinity chromatography. The column was washed and Veg-His_6 eluted with French Press buffer containing the indicated imidazole concentrations. Veg-His_6 eluted nearly pure at 100 mM imidazole. Red arrows marks the position of Veg-His_6 in the different fractions throughout the purification process.

4. Discussion

The connection between biofilm formation and sporulation has been ever more supported in light of several studies. The influence of several proteins in both processes and the developmental regulation pathways that underlie them connect matrix production, competence, and sporulation in a continuous spectrum of gene expression and consequent phenotypical changes. Furthermore, the assembly of macromolecular complexes is a pivotal phenomenon through which both biofilms and spores build their structures. The spore coat is a macromolecular assembly of more than 80 different proteins, much of which interact with each other to assemble themselves in an orderly and cohesive manner. This reflects in the subsequent properties that spores possess that make them uniquely resistant, dormant cells.

4.1. Tgl activity decreases with increasing substrate concentration

Tgl is a spore-associated transglutaminase, produced late in the mother cell line of gene expression, under the control of σ^K; Tgl depends on the prior assembly of its substrates to localize at the surface of the developing spore (Kobayashi *et al.*, 1998; Fernandes, 2014; Fernandes *et al.*, 2019). The substrates that make the most important contribution to the localization of Tgl are SafA and $SafA^{C30}$. It was suggested, then, that the activity of Tgl could be, at least in part, dependent on the local concentration of substrates (Fernandes *et al.*, 2019). In vitro cross-linking assays with Tgl and $SafA^{C30}$ showed that in fact the activity of Tgl was inversely related to the concentration of the substrate (Fernandes *et al.*, 2019). The likely explanation is that Tgl cross-links itself to the nascent structure, and does so in a manner that increases with the substrate concentration. Since cross-linked Tgl is no longer able to be active in catalysis, this phenomenon could provide a way to limit, *in vivo*, the activity of the enzyme, which would eventually consume itself in the reaction. The reaction of Tgl with $SafA^{C30}$ *in vitro* is relevant, because $SafA^{C30}$ is produced *in vivo* through internal translation of the *safA* mRNA (Ozin et al., 2001).

In this work, we extend the characterization of the activity of Tgl *in vitro* to the other two forms of SafA present *in vivo*, $SafA^{FL}$ and $SafA^{N21}$. $SafA^{FL}$ was assayed with Tgl and the results showed that Tgl activity decreased as $SafA^{FL}$ concentration increased, as previously reported for $SafA^{C30}$. $SafA^{N21}$ was also assayed with Tgl and the activity of the enzyme again seemed to decrease with increasing concentration of the $SafA^{N21}$ substrate. We note, however, that the cross-linking reaction was much less effective, with almost no $SafA^{N21}$ bands disappearing from SDS-PAGE gels. The reason for this is unclear, since $SafA^{N21}$ consists of the LysM domain and the linker that connects this domain to the C30 domain (Figure 3.4C). $SafA^{N21}$ has several glutamine and lysine residues (Figure 3.4C). Although the structure of the $SafA^{N21}$ protein is not presently known, it is possible that the existing glutamine residues are not exposed and cannot participate in cross-linking to $SafA^{N21}$ itself or to Tgl. Thus, may not be a physiological substrate for Tgl. We cannot, however, exclude the possibility that *in vivo* $SafA^{N21}$ works as a Q- or K-donor in reactions that involve other proteins or that this part of the protein is cross-linked to the cortex peptidoglycan, as suggested (Fernanndes et al., 2018; Fernandes et al., 2019).

Tgl activity is thus limited by increasing concentrations of substrate. Tgl has other known substrates, GerQ and YeeK (Ragkousi and Setlow, 2004; Takamatsu *et al.*, 2009), and it seems likely that Tgl behaves in the same manner with any substrate. If so, then the self-limiting activity of Tgl through cross-linking to its natural substrates may be a way to limit unwanted reactions at the spore surface.

4.2. $SafA^{C30}$ forms disulfide cross-linked oligomers that depend on Tgl

Dimers of $SafA^{C30}$ are seen in SDS-PAGE gels, suggesting that they are highly stable species, able to resist the denaturing conditions of SDS-PAGE. The hexameric form of $SafA^{C30}$, $(SafA^{C30})_6$ is not detected by SDS-PAGE in the absence of Tgl (Figure 3.7). Thus, this species is not able to resist to the conditions of SDS-PAGE unless cross-linked by Tgl.

$(SafA^{C30})_6$ is detected under both reducing and non-reducing conditions. However, $(SafA^{C30})_6$ is more abundant when formation of this species is analyzed by SDS-PAGE under non-reducing conditions as compared to reducing conditions (Figure 3.7A). One likely explanation for this observation is that $(SafA^{C30})_6$ is formed by disulfide bonds and that the resulting complex is then subject to a "spotwelding"

activity through which Tgl introduces cross-links between some but not all of the monomers. Although we have not yet investigated this, it seems possible that $(SafA^{C30})_6$ is formed by self-assembly of $SafA^{C30}$ or $(SafA^{C30})_2$ [in which case $(SafA^{C30})_6$ could be a trimer of dimers] and that at some point in the assembly pathway disulfide bonds are formed between adjacent monomers. Tgl would then introduce cross-links at critical positions to stabilize the complex, in line with the idea that Tgl, as most other TGases that have been characterized, stabilizes pre-assembled complexes.

We do not presently know whether $SafA^{C30}$ self-associates even in the presence of reducing agents, which would show that, as we predict, the protein is able to self-assemble via non-covalent interactions, and at what point disulfide bonds are formed. A way to test this is to substitute the two cysteine residues of $SafA^{C30}$ by serines, to create two single and a double mutant. The study of this form of $SafA^{C30}$ would allow us to test the hypothesis that the assembly pathway leading to $(SafA^{C30})_6$ involves formation of non-covalent hexamer, then a disulfide cross-linked hexamer which is then stabilized by Tgl. It would also allow us to test whether the prior formation of disulfide bonds is a pre-requisite for cross-linking by Tgl. This is a possibility because the disulfide cross-linked hexamer could bring the appropriate glutamine and lysine residues within the right distance and orientation for the Tgl-mediated cross-linking.

Our observations with the alanine substitutions of $SafA^{C30}$ residues K177 and K318 are consistent with these ideas. Under non-reducing conditions, $SafA^{C30(K177A)}$, $SafA^{C30(K318A)}$, and $SafA^{C30(K177/318A)}$ formed oligomers in the presence of Tgl only, which are presumably only cross-linked through disulfide bonds. But under reducing conditions very little $(SafA^{C30})_6$ is detected and importantly, the reduction seems more important for the K318A than for the K177A substitution. This suggests that lysine 318 is more important for the stabilization of $(SafA^{C30})_6$ and interestingly, this residue is found in the middle of $SafA^{C30}$ close to cysteine residues at positions 323 and 325 (Figure 3.7B). It is tempting to suggest that disulfide bond formation involving these cysteine residues in adjacent $SafA^{C30}$ monomers would promote cross-linking of $(SafA^{C30})_6$ by Tgl. We also note that lysine 318 is in close proximity to several glutamine residues, which would be available presumably for cross-linking. In any event, it is also not clear whether the disulfide and Tgl-cross-linked $(SafA^{C30})_6$ is further cross-linked and if so, whether disulfide bonds are formed between $(SafA^{C30})_6$ units. However, our study suggests that at least $SafA^{FL}$ can be cross-linked into species with an apparent mass much higher than 250 kDa (Figure 3.3A).

Although disulfide bonds are rare in intracellular proteins, as the cytoplasm is a reducing environment, they have been observed in several spore surface proteins (Kailas *et al.*, 2011; Jiang *et al.*, 2015). For example, *B. subtilis* CotY is a cysteine-rich protein required for assembly of the outermost crust layer. CotY self-assembles into ordered one-dimensional fibres, two-dimensional sheets and three-dimensional stacks. The high degree of order within these structures appears to favor cooperative intracellular disulfide bond formation (Jiang *et al.*, 2015). Cooperative disulfide cross-linking during the formation of large assemblies may thus be a mechanism to overcome the reducing environment of the cytoplasm. Two alternatives, but not mutually exclusive possibilities are: *i*) late in sporulation, prior to mother cell lysis, the mother cell cytoplasm becomes less reducing; *ii*) disulfide cross-linking takes part, at least in part, following spore release from the mother cell.

4.3. Insights onto the structure of coat-associated SafA

Previous work has suggested a model in which $SafA^{FL}$ resides at the cortex/inner coat interface, with the LysM domain important in establishing the interaction with the cortex and the C30 domain in the coat. $SafA^{C30}$, made through internal translation would then be able to form complexes bridging $SafA^{FL}$ molecules (Figure 4.1A). Thus, $SafA^{FL}$ would provide positional cues for the assembly of $(SafA^{C30})_6$ around the spore. Tgl would then cross-link the $(SafA^{C30})_6$ units, fortifying the assembly and the cortex/inner coat interface (Figure 4.1B). Our work introduces an additional important detail in the model, by showing that: *i*) assembly of $(SafA^{C30})_6$ involves formation of disulfide bonds, *ii*) that Tgl cross-links this species, possibly in addition to introducing cross-links between adjacent $(SafA^{C30})_6$ complexes (Figure 4.1).

In addition, the observation that $SafA^{N21}$ contains several glutamine and lysine residues but is not a good substrate in cross-linking reactions to itself or to Tgl, raises the possibility that this region of the full-length protein is cross-linked to other proteins present at the cortex/inner coat interface or

directly to the cortex peptidoglycan (Fernandes *et al.*, 2018; Fernandes *et al.*, 2019) (Figures 3.4C and 4.1B).

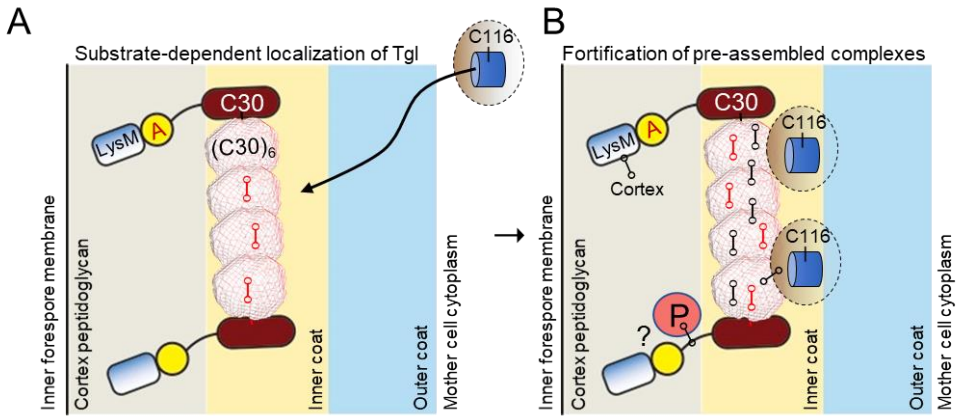

Figure 4.1. Cross-linking of SafA forms by Tgl. A: SafA is recruited early to the surface of the developing spore, prior to synthesis of Tgl. At the spore surface, $SafA^{FL}$ and $SafA^{C30}$ are key determinants in the recruitment of Tgl to the coat (left panel), placing the enzyme in close proximity to its substrates. B: In a second step, Tgl cross-links $SafA^{C30}$ and $SafA^{FL}$ to fortify pre-assembled complexes. Formation of $(SafA^{C30})_6$ may involve mainly disulfide bonds (red linkers) and Tgl exerts a "spotwelding" activity in cross-linking and stabilizing $(SafA^{C30})_6$ (black linkers). Tgl may also cross-link $SafA^{N21}$ to other coat proteins (P) or to the cortex. During the process, Tgl itself also becomes cross-linked, which eventually limits its activity at the spore surface. The proteins are not drawn to scale. Adapted from Fernandes *et al.*, 2019.

4.4. *B. subtilis* Veg impairs assembly of the outer coat

The *veg* gene is expressed under the control of the σ^A-containing form of RNA polymerase, (Moran *et al.*, 1982). A role has been attributed to the Veg protein in biofilm formation and also in spore germination (Fukushima *et al.*, 2003; Lei *et al.*, 2013). Veg promotes biofilm formation by repressing, directly or indirectly, the SinR repressor (Lei *et al.*, 2013). The specific function of Veg in both processes has remained unclear as no further studies were conducted after the initial observations were published. Here we built several strains to study the role of Veg in spore coat assembly (Figure 3.8). One important note is that the coat layer is responsible for spore germination, and YabG which is encoded by a gene found just upstream of *veg*, is also important for germination and is required for processing of the substrates that are cross-linked by the Tgl transglutaminase (Kuwana *et al.*, 2006; see also below). The results of coat extracts resolved by SDS-PAGE showed little difference between the wild-type strain and the *veg* mutant. Expression of *veg* in late stages of sporulation, under P_{gerE} in single copy yielded a comparable phenotype to that of the wild-type (Figure 3.8A). To increase expression of the fusion, P_{gerE}-*veg* was propagated into a multicopy plasmid (pMK3). SDS-PAGE analysis of the extractable coat proteins, from each strain, showed that a few bands were altered in the P_{gerE}-*veg* strains (Figure 3.8B). Bands corresponding to outer coat proteins, such as CotG and CotB-66 (Figure 3.8C and D), were reduced in the P_{gerE}-*veg* strains as confirmed by Western blot. From these results alone, we could infer that Veg, the product of the *veg* gene, somehow interferes with outer coat assembly. CotB and CotG are abundant components of the outer coat, and CotG is a key structural determinant of the striated structure of the outer coat (Freitas et al., 2020). Both CotB and CotG are phosphorylated by CotH, a serine kinase (Zilhão et al., 2005; Nguyen *et al.*, 2016; Freitas et al., 2020). CotB is produced as a 46 kDa protein, but the species of 66 kDa that accumulates in the coat is formed through extensive phosphorylation, a reaction that requires CotG (Nguyen *et al.*, 2016; Freitas et al., 2020). Thus, the reduction of CotG and CotB-66 is consistent (see also below).

Also missing were bands corresponding to a peptidoglycan hydrolase, YdhD, YjdH, a protein of unknown function, and YtmK, a putative L-cysteine binding protein of an ABC transporter (Kodama *et al.*, 2000; Bunai *et al.*, 2004). These peptides were identified by mass spectrometry. YdhD contains a cell-wall binding motif and is involved in germination; overproduction of this enzyme leads to reduced response to the germinant L-alanine and in some instances to germination without any germinants (Kodama *et al.*, 2000). Veg overproduction leads to a downregulation of YdhD. This could improve on L-alanine response and reduce the frequency of auto-germination.

The Veg protein shows structural homology to proteins bearing the nucleic acid binding Sm-like domain (Figure 4.2A). Sm and Sm-like proteins are members of a family of small proteins that are widespread throughout all kingdoms of life and have various roles in mRNA processing and translation (Bandyra and Luisi, 2013; Mura *et al.*, 2013). This domain is found as the core structure in Lsm (like-Sm) proteins and bacterial Lsm-related Hfq proteins and consists of an open beta-barrel with an SH3-like topology (Bandyra and Luisi, 2013; Mura *et al.*, 2013; Sauer, 2013). Sm-like proteins are ring-shaped oligomers of 6-7 members, which bind single stranded nucleic acids.

We produced a homology model of the Veg protein to assess whether its overall fold and the positions of conserved amino acids were maintained and to identify candidate residues that could be tested, in the future, through loss of side-chain mutagenesis for a role in nucleic acid binding. We used the structure of the protein with unknown function from *Streptococcus pneumoniae TIGR4* (protein data bank accession code: 3FB9). The homology model of Veg shows the characteristic heptameric fold of the Sm-like domain (Bandyra and Luisi, 2013; Mura *et al.*, 2013) (Figure 4.2). The model shows the presence of several surface-exposed positively charged lysine residues (Figure 4.2B). A surface electrostatic potential map representation of Veg shows that some parts of the surface of the oligomer are positively charged as is the inside of the ring (Figure 4.2C). These are features that could facilitate the interaction with the negative charged single stranded nucleic acids (Bandyra and Luisi, 2013; Mura *et al.*, 2013; Sauer, 2013).

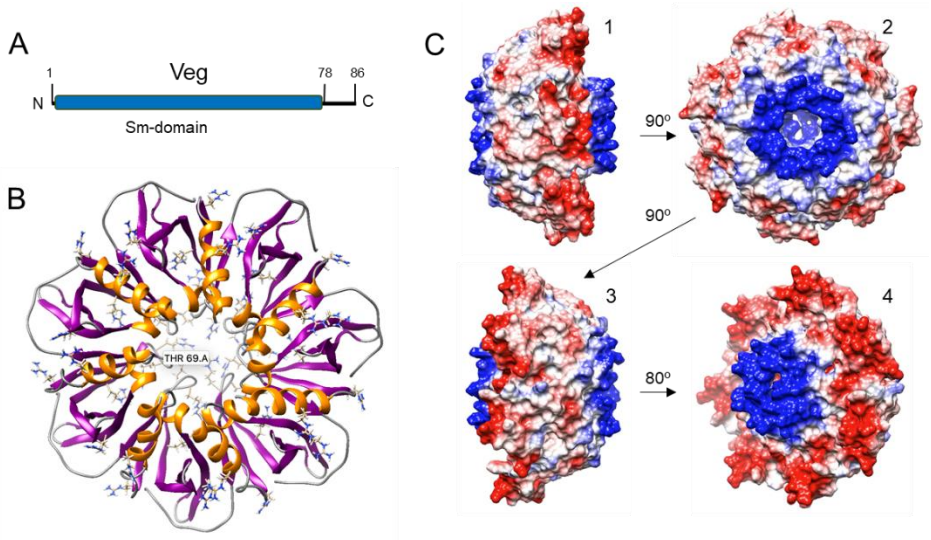

Figure 4.2. The structure of *C. difficile* Veg. A: Structural organization of Veg. Veg contains an Sm domain, an ancient RNA-binding motif with oligo(U) specificity. B: Homology model generated for Veg using as template a protein with unknown fuction from *S. penumoniae* TIGR4 (PDB code: 3FB9). This structural model show that the Veg protein may form an heptamer with a central channel, an arrangement that is characteristic of Sm-like proteins. C: Different views of the electrostatic surface map of the ring formed by the Veg heptamer. Red denotes positive charge and blue negative charge.

It seems plausible that Veg may affect the stability and/or translation of the mRNAs coding for some of the coat proteins, such as CotG and CotB. Interestingly, *cotG* and *cotB* are part of a gene cluster with the gene for the CotH kinase. Expression of *cotH* depends on a promoter located 812 nucleotides upstream of the coding region. The long sequence at 5′ end of *cotH* is most likely not translated and completely overlaps the divergent *cotG* gene. The two genes are adjacent on the chromosome and divergently transcribed by σ^K (Zilhão *et al.*, 2004). Since the overexpression of Veg leads to a decrease in the representation of CotB and CotG in spore coat extracts, a possibility is that Veg could exert its role at the level of the *cotH* mRNA.

Remarkably, CotB itself and the product of the downstream gene, *ywrJ*, also have a Sm-like fold. This strongly suggests that an important and previously unrecognized level of control of coat assembly is at the level of mRNA stability/translation. Veg and these other proteins may fine tune production and assembly of some key spore coat components, according to specific environmental conditions (Abhyankar *et al.*, 2016; Isticato *et al.*, 2020). Whether Veg, YwrJ and CotB have redundant functions remains to be tested.

4.5. Overproduction and purification of *C. difficile* Veg

For the first time, *C. difficile* Veg was cloned, overproduced and purified in order to characterize this small protein of unknown function that, in *B. subtilis*, has shown to participate in biofilm formation, spore coat assembly, and spore germination (Fukushima *et al.*, 2003; Lei *et al.*, 2013). Veg-His_6 was seen in the soluble fraction of the extract, the supernatant (Figure 3.9), as well has in the insoluble fraction, the pellet, and eluted in its most purified fraction at 100 mM imidazole. French press buffer was used during mechanical lysis and purification. The conditions for Veg-His_6 purification are standard conditions for protein purification and the protocol for Veg-His_6 purification may require optimization. The structural characterization of Veg will provide important cues on the function of a protein that seems to play an important role in developmental processes in sporeformers, from an evolutionary standpoint.

Our SDS-PAGE analysis allows only the visualization of its monomeric form. However, the presence of the Sm-like motif and our homology model (above) suggest that Veg self-interacts to form oligomers. SEC (size exclusion chromatography) could provide this information. Structural studies could also help further characterize Veg, such as SAXS (small-angle X-ray scattering) or NMR (nuclear magnetic resonance). The characterization of Veg can bring closer together two processes that are crucial in pathogenicity in humans and animals and spoilage in many industrial settings and opens the door for the development of new tools to overcome these many challenges.

Concluding remarks

The *yabG* gene is conserved among sporeformers and part of a genomic signature for endosporulation (Abecasis *et al.*, 2013). In *B. subtilis*, YabG is a protease required for the processing of a sub-set of proteins that are then cross-linked by Tgl. An important substrate of Tgl is SafA, the morphogenetic protein required for the assembly of the spore inner coat. Work in the laboratory has shown that YabG cleaves $SafA^{FL}$ (and $SafA^{C30}$) at an arginine residue at position 381 of the 387-residues long protein. Although $SafA^{FL}$ (and $SafA^{C30}$) is cross-linked efficiently by Tgl *in vitro*, as shown here, it is not yet known whether prior cleavage of the protein by YabG at arginine 381 improves the efficiency and/or the kinetics of cross-linking. The test of this possibility is an important aspect of the characterization of YabG, and a goal for future work.

The association of *yabG* and *veg* and the involvement of the latter in biofilm formation and spore coat assembly (this work) suggests a link between the two processes. If Veg proves to be an RNA-binding protein, it may control the stability and/or translation of mRNAs coding for proteins important for both biofilm formation and spore coat assembly. Biofilms are held together by an extracellular matrix. Spores are formed within biofilms and remain attached to the structure following lysis of the mother cell. We presume that the composition of the spore surface layers has features that allow spores to become and to remain embedded in the biofilm matrix and we suspect that this is the reason why the two processes share common regulatory proteins.

We have characterized the activity of Tgl in cross-linking a physiological substrate and in the future, we will further explore the functional connection between Tgl and YabG, and whether Tgl and YabG have a role in biofilm formation. We have initiated the characterization of Veg, and we have shown, for the first time, a role for this protein in spore coat assembly. In the future, we will analyse in detail the role of Veg in both coat assembly and biofilm formation, and the functional connection of Veg to YabG, suggested by the synteny of the two genes across genomes.

References

Abecasis, A.B., Serrano, M., Alves, R., Quintais, L., Pereira-Leal, J.B., and Henriques, A.O. (2013) A genomic signature and the identification of new sporulation genes. *J Bacteriol* **195**: 2101–2115 http://dx.doi.org/10.1128/JB.02110-12.http://jb.asm.org/. Accessed November 17, 2020.

Abhyankar, W.R., Kamphorst, K., Swarge, B.N., Veen, H. van, Wel, N.N. van der, Brul, S., *et al.* (2016) The Influence of Sporulation Conditions on the Spore Coat Protein Composition of *Bacillus subtilis* Spores. *Front Microbiol* **7**: 1636 http://journal.frontiersin.org/article/10.3389/fmicb.2016.01636/full. Accessed November 22, 2020.

Aronson, A.I., Ekanayake, L., and Fitz-James, P.C. (1992) Protein filaments may initiate the assembly of the *Bacillus subtilis* spore coat. *Biochimie* **74**: 661–667.

Aronson, A.I., and Fitz-James, P. (1976) Structure and Morphogenesis of the Bacterial Spore Coat. .

Bai, U., Mandic-Mulec, I., and Smith, I. (1993) SinI modulates the activity of SinR, a developmental switch protein of *Bacillus subtilis*, by protein-protein interaction. *Genes Dev* **7**: 139–148 https://pubmed.ncbi.nlm.nih.gov/8422983/. Accessed November 5, 2020.

Bandyra, K.J., and Luisi, B.F. (2013) Licensing and due process in the turnover of bacterial RNA. *RNA Biol* **10**: 627–635 https://www.tandfonline.com/doi/abs/10.4161/rna.24393. Accessed December 1, 2020.

Bauer, N.C., Doetsch, P.W., and Corbett, A.H. (2015) Mechanisms Regulating Protein Localization. *Traffic* **16**: 1039–1061.

Beall, B., Driks, A., Losick, R., and Moran, C.P. (1993) Cloning and characterization of a gene required for assembly of the *Bacillus subtilis* spore coat. *J Bacteriol* **175**: 1705–1716 https://pubmed.ncbi.nlm.nih.gov/8449878/. Accessed September 25, 2020.

Bradford, M.M. (1976) A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. .

Branda, S.S., Chu, F., Kearns, D.B., Losick, R., and Kolter, R. (2006) A major protein component of the *Bacillus subtilis* biofilm matrix. *Mol Microbiol* **59**: 1229–1238 https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2958.2005.05020.x. Accessed September 16, 2020.

Branda, S.S., González-Pastor, J.E., Ben-Yehuda, S., Losick, R., and Kolter, R. (2001) Fruiting body formation by *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **98**: 11621–11626 www.pnas.orgcgidoi10.1073pnas.191384198. Accessed September 16, 2020.

Buist, G., Steen, A., Kok, J., and Kuipers, O.P. (2008) LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol Microbiol* **68**: 838–847 https://pubmed.ncbi.nlm.nih.gov/18430080/. Accessed October 5, 2020.

Bunai, K., Ariga, M., Inoue, T., Nozaki, M., Ogane, S., Kakeshita, H., *et al.* (2004) Profiling and comprehensive expression analysis of ABC transporter solute-binding proteins of *Bacillus subtilis* membrane based on a proteomic approach. *Electrophoresis* **25**: 141–155 https://pubmed.ncbi.nlm.nih.gov/14730579/. Accessed November 22, 2020.

Burguière, P., Fert, J., Guillouard, I., Auger, S., Danchin, A., and Martin-Verstraete, I. (2005) Regulation of the *Bacillus subtilis* ytmI operon, involved in sulfur metabolism. *J Bacteriol* **187**: 6019–6030 http://jb.asm.org/. Accessed November 22, 2020.

Catalano, F.A., Meador-Parton, J., Popham, D.L., and Driks, A. (2001) Amino acids in the *Bacillus subtilis* morphogenetic protein SpoIVA with roles in spore coat and cortex formation. *J Bacteriol* **183**: 1645–1654 /pmc/articles/PMC95049/?report=abstract. Accessed September 26, 2020.

Chada, V.G.R., Sanstad, E.A., Wang, R., and Driks, A. (2003) Morphogenesis of *Bacillus* Spore Surfaces. *J Bacteriol* **185**: 6255–6261 http://jb.asm.org/. Accessed September 26, 2020.

Chater, K.F. (2001) Regulation of sporulation in Streptomyces coelicolor A3(2): A checkpoint multiplex? *Curr Opin Microbiol* **4**: 667–673 https://pubmed.ncbi.nlm.nih.gov/11731318/. Accessed November 17, 2020.

Chu, F., Kearns, D.B., McLoon, A., Chai, Y., Kolter, R., and Losick, R. (2008) A novel regulatory protein governing biofilm formation in *Bacillus subtilis*. *Mol Microbiol* **68**: 1117–1127 https://pubmed.ncbi.nlm.nih.gov/18430133/. Accessed September 17, 2020.

Cohn, F. (1876) Untersuchungen ueber Bakterien: Beitraege zur Biologie der Bacillen. In *Untersuchungen ueber Bakterien*. pp. 249–276.
Cooper, I.A., Russel, P., and Thwaite, J.E. (2018) *Bacillus*. In *Medical Microbiology*. Michael, B., and Irving, W. (eds). Elsevier, pp. 169–177 https://www.ncbi.nlm.nih.gov/books/NBK7699/. Accessed November 5, 2020.
Costa, T., Isidro, A.L., Moran, C.P., and Henriques, A.O. (2006) Interaction between coat morphogenetic proteins SafA and SpoVID. *J Bacteriol* **188**: 7731–7741 http://jb.asm.org/. Accessed October 6, 2020.
Costa, T., Serrano, M., Steil, L., Völker, U., Moran, C.P., and Henriques, A.O. (2007) The timing of cotE expression affects *Bacillus subtilis* spore coat morphology but not lysozyme resistance. *J Bacteriol* **189**: 2401–2410.
Davies, D. (2003) Understanding biofilm resistance to antibacterial agents. *Nat Rev Drug Discov* **2**: 114–122 https://www.nature.com/articles/nrd1008. Accessed September 16, 2020.
Driks, A., and Eichenberger, P. (2016) The Spore Coat. In *The Bacterial Spore: from Molecules to Systems*. American Society of Microbiology, pp. 179–200.
Driks, A., Roels, S., Beall, B., Moran, C.P., and Losick, R. (1994) Subcellular localization of proteins involved in the assembly of the spore coat of *Bacillus subtilis*. *Genes Dev* **8**: 234–244 https://pubmed.ncbi.nlm.nih.gov/8299942/. Accessed September 25, 2020.
Earl, A.M., Losick, R., and Kolter, R. (2008) Ecology and genomics of *Bacillus subtilis*. *Trends Microbiol* **16**: 269–275.
Fernandes, C.G. (2014) Assembly and function of a protein cross-linking enzyme during bacterial spore morphogenesis. .
Fernandes, C.G., Martins, D., Hernandez, G., Sousa, A.L., Freitas, C., Tranfield, E.M., *et al.* (2019) Temporal and spatial regulation of protein cross-linking by the pre-assembled substrates of a *Bacillus subtilis* spore coat transglutaminase. *PLOS Genet* **15**: e1007912 http://dx.plos.org/10.1371/journal.pgen.1007912. Accessed December 5, 2019.
Fernandes, C.G., Moran, C.P., and Henriques, A.O. (2018) Autoregulation of SafA assembly through recruitment of a protein cross-linking enzyme. *J Bacteriol* **200**.
Fernandes, C.G., Plácido, D., Lousa, D., Brito, J.A., Isidro, A., Soares, C.M., *et al.* (2015) Structural and Functional Characterization of an Ancient Bacterial Transglutaminase Sheds Light on the Minimal Requirements for Protein Cross-Linking. *Biochemistry* **54**: 5723–5734 https://pubs.acs.org/doi/10.1021/acs.biochem.5b00661. Accessed December 6, 2019.
Francesco, M.D. de, Jacobs, J.Z., Nunes, F., Serrano, M., Mckenney, P.T., Chua, M.H., *et al.* (2012) Physical interaction between coat morphogenetic proteins SpoVID and CotE is necessary for spore encasement in *Bacillus subtilis*. *J Bacteriol* **194**: 4941–4950 https://pubmed.ncbi.nlm.nih.gov/22773792/. Accessed November 4, 2020.
Freitas, C., Plannic, J., Isticato, R., Pelosi, A., Zilhão, R., Serrano, M., *et al.* (2020) A protein phosphorylation module patterns the *Bacillus subtilis* spore outer coat. *Mol Microbiol* https://pubmed.ncbi.nlm.nih.gov/32592201/. Accessed November 4, 2020.
Fujita, M., González-Pastor, J.E., and Losick, R. (2005) High- and low-threshold genes in the Spo0A regulon of *Bacillus subtilis*. *J Bacteriol* **187**: 1357–1368 http://jb.asm.org/. Accessed September 17, 2020.
Fukushima, T., Ishikawa, S., Yamamoto, H., Ogasawara, N., and Sekiguchi, J. (2003) Transcriptional, functional and cytochemical analyses of the veg gene in *Bacillus subtilis*. *J Biochem* **133**: 475–483.
Grice, S.F.J. Le, Shih, C.C., Whipple, F., and Sonenshein, A.L. (1986) Separation and analysis of the RNA polymerase binding sites of a complex *Bacillus subtilis* promoter. *MGG Mol Gen Genet* **204**: 229–236 https://link.springer.com/article/10.1007/BF00425503. Accessed September 17, 2020.
Griffin, M., Casadio, R., and Bergamini, C.M. (2002) Transglutaminases: Nature's biological glues. *Biochem J* **368**: 377–396.
Hall-Stoodley, L., Costerton, J.W., and Stoodley, P. (2004) Bacterial biofilms: From the natural environment to infectious diseases. *Nat Rev Microbiol* **2**: 95–108 https://www.nature.com/articles/nrmicro821. Accessed September 16, 2020.
Hamon, M.A., Stanley, N.R., Britton, R.A., Grossman, A.D., and Lazazzera, B.A. (2004)

Identification of AbrB-regulated genes involved in biofilm formation by *Bacillus subtilis*. *Mol Microbiol* **52**: 847–860 /pmc/articles/PMC1409746/?report=abstract. Accessed September 17, 2020.
Harwood, C., and Cutting, S. (1990) *Molecular biological methods for Bacillus*. 2nd ed., Wiley, .
Henriques, A.O., and Moran, C.P. (2007) Structure, assembly, and function of the spore surface layers. *Annu Rev Microbiol* **61**: 555–588 http://www.annualreviews.org/doi/10.1146/annurev.micro.61.080706.093224. Accessed September 24, 2020.
Higgins, D., and Dworkin, J. (2012) Recent progress in *Bacillus subtilis* sporulation. *FEMS Microbiol Rev* **36**: 131–148 https://academic.oup.com/femsre/article/36/1/131/534503. Accessed September 19, 2020.
Hilbert, D.W., and Piggot, P.J. (2004) Compartmentalization of Gene Expression during *Bacillus subtilis* Spore Formation. *Microbiol Mol Biol Rev* **68**: 234–262 http://mmbr.asm.org/. Accessed September 21, 2020.
Hoon, M.J.L. De, Eichenberger, P., and Vitkup, D. (2010) Hierarchical evolution of the bacterial sporulation network. *Curr Biol* **20**.
Hullo, M.F., Moszer, I., Danchin, A., and Martin-Verstraete, I. (2001) CotA of *Bacillus subtilis* is a copper-dependent laccase. *J Bacteriol* **183**: 5426–5430 https://pubmed.ncbi.nlm.nih.gov/11514528/. Accessed September 23, 2020.
Isticato, R., Lanzilli, M., Petrillo, C., Donadio, G., Baccigalupi, L., and Ricca, E. (2020) *Bacillus subtilis* builds structurally and functionally different spores in response to the temperature of growth. *Environ Microbiol* **22**: 170–182 https://onlinelibrary.wiley.com/doi/abs/10.1111/1462-2920.14835. Accessed November 22, 2020.
Jiang, M., Shao, W., Perego, M., and Hoch, J.A. (2000) Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. *Mol Microbiol* **38**: 535–542 https://pubmed.ncbi.nlm.nih.gov/11069677/. Accessed September 17, 2020.
Jiang, S., Wan, Q., Krajcikova, D., Tang, J., Tzokov, S.B., Barak, I., and Bullough, P.A. (2015) Diverse supramolecular structures formed by self-assembling proteins of the *Bacillus subtilis* spore coat. *Mol Microbiol* **97**: 347–359 https://onlinelibrary.wiley.com/doi/10.1111/mmi.13030. Accessed November 17, 2020.
Kailas, L., Terry, C., Abbott, N., Taylor, R., Mullin, N., Tzokov, S.B., *et al.* (2011) Surface architecture of endospores of the *Bacillus* cereus/anthracis/ thuringiensis family at the subnanometer scale. *Proc Natl Acad Sci U S A* **108**: 16014–16019 www.pnas.org/cgi/doi/10.1073/pnas.1109419108. Accessed December 1, 2020.
Kobayashi, K. (2007) Gradual activation of the response regulator DegU controls serial expression of genes for flagellum formation and biofilm formation in *Bacillus subtilis*. *Mol Microbiol* **66**: 395–409 https://pubmed.ncbi.nlm.nih.gov/17850253/. Accessed September 17, 2020.
Kobayashi, K., Hashiguchi, K.I., Yokozeki, K., and Yamanaka, S. (1998) Molecular Cloning of the Transglutaminase Gene from *Bacillus subtilis* and Its Expression in Escherichi. *Biosci Biotechnol Biochem* **62**: 1109–1114 https://pubmed.ncbi.nlm.nih.gov/9692191/. Accessed November 21, 2020.
Kobayashi, K., and Iwano, M. (2012) BslA(YuaB) forms a hydrophobic layer on the surface of *Bacillus subtilis* biofilms. *Mol Microbiol* **85**: 51–66 https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2958.2012.08094.x. Accessed September 16, 2020.
Kobayashi, K., Kumazawa, Y., Miwa, K., and Yamanaka, S. (1996) ε-(γ-Glutamyl)lysine cross-links of spore coat proteins and transglutaminase activity in *Bacillus subtilis*. *FEMS Microbiol Lett* **144**: 157–160.
Kodama, T., Takamatsu, H., Asai, K., Ogasawara, N., Sadaie, Y., and Watabe, K. (2000) Synthesis and characterization of the spore proteins of *Bacillus subtilis* YdhD, YkuD, and YkvP, which carry a motif conserved among cell wall binding proteins. *J Biochem* **128**: 655–663.
Konovalova, A., Petters, T., and Søgaard-Andersen, L. (2010) Extracellular biology of Myxococcus xanthus. *FEMS Microbiol Rev* **34**: 89–106 https://pubmed.ncbi.nlm.nih.gov/19895646/. Accessed November 17, 2020.
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., *et al.* (1997) The

complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**: 249–256 https://pubmed.ncbi.nlm.nih.gov/9384377/. Accessed August 11, 2020.
Kuwana, R., Okuda, N., Takamatsu, H., and Watabe, K. (2006) Modification of GerQ reveals a functional relationship between Tgl and YabG in the coat of *Bacillus subtilis* spores. *J Biochem* **139**: 887–901 https://pubmed.ncbi.nlm.nih.gov/16751597/. Accessed September 29, 2020.
Lei, Y., Oshima, T., Ogasawara, N., and Ishikawa, S. (2013) Functional analysis of the protein veg, which stimulates biofilm formation in *Bacillus subtilis*. *J Bacteriol* **195**: 1697–1705.
López, D., Vlamakis, H., and Kolter, R. (2010) Biofilms. *Cold Spring Harb Perspect Biol* **2**: a000398 http://cshperspectives.cshlp.org/. Accessed September 16, 2020.
Lorand, L., and Graham, R.M. (2003) Transglutaminases: Crosslinking enzymes with pleiotropic functions. *Nat Rev Mol Cell Biol* **4**: 140–156.
Mah, T.F.C., and O'Toole, G.A. (2001) Mechanisms of biofilm resistance to antimicrobial agents. *Trends Microbiol* **9**: 34–39 http://tim.trends.com0966. Accessed September 16, 2020.
Marini, E. (2020) Assembly of the Clostridioides difficile spore surface layers. http://hdl.handle.net/10362/105568.
Mckenney, P.T., Driks, A., and Eichenberger, P. (2013) The *Bacillus subtilis* endospore: Assembly and functions of the multilayered coat. *Nat Rev Microbiol* **11**: 33–44.
McPherson, D.C., Kim, H., Hahn, M., Wang, R., Grabowski, P., Eichenberger, P., and Driks, A. (2005) Characterization of the *Bacillus subtilis* spore morphogenetic coat protein CotO. *J Bacteriol* **187**: 8278–8290 http://jb.asm.org/. Accessed September 27, 2020.
Mielich-Süss, B., and Lopez, D. (2015) Molecular mechanisms involved in *Bacillus subtilis* biofilm formation. *Environ Microbiol* **17**: 555–565 https://sfamjournals.onlinelibrary.wiley.com/doi/full/10.1111/1462-2920.12527. Accessed October 29, 2020.
Molle, V., Fujita, M., Jensen, S.T., Eichenberger, P., González-Pastor, J.E., Liu, J.S., and Losick, R. (2003) The Spo0A regulon of *Bacillus subtilis*. *Mol Microbiol* **50**: 1683–1701 https://pubmed.ncbi.nlm.nih.gov/14651647/. Accessed September 17, 2020.
Moran, C.P., Lang, N., LeGrice, S.F.J., Lee, G., Stephens, M., Sonenshein, A.L., *et al.* (1982) Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *MGG Mol Gen Genet* **186**: 339–346 https://link.springer.com/article/10.1007/BF00729452. Accessed November 13, 2020.
Mura, C., Randolph, P.S., Patterson, J., and Cozen, A.E. (2013) Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function. *RNA Biol* **10**: 636–651 https://www.tandfonline.com/doi/abs/10.4161/rna.24538. Accessed December 1, 2020.
Naclerio, G., Baccigalupi, I.L., Zilhao, R., Felice, M. De, and Ricca, E. (1996) *Bacillus subtilis* spore coat assembly requires cotH gene expression. *J Bacteriol* **178**: 4375–4380 http://jb.asm.org/. Accessed September 27, 2020.
Nagórska, K., Ostrowski, A., Hinc, K., Holland, I.B., and Obuchowski, M. (2010) Importance of eps genes from *Bacillus subtilis* in biofilm formation and swarming. *J Appl Genet* **51**: 369–381 https://pubmed.ncbi.nlm.nih.gov/20720312/. Accessed September 16, 2020.
Newman, J.A., Rodrigues, C., and Lewis, R.J. (2013) Molecular basis of the activity of SinR Protein, the master regulator of biofilm formation in *Bacillus subtilis*. *J Biol Chem* **288**: 10766–10778 https://pubmed.ncbi.nlm.nih.gov/23430750/. Accessed November 5, 2020.
Nguyen, K.B., Sreelatha, A., Durrant, E.S., Lopez-Garrido, J., Muszewska, A., Dudkiewicz, M., *et al.* (2016) Phosphorylation of spore coat proteins by a family of atypical protein kinases. *Proc Natl Acad Sci U S A* **113**: E3482–E3491 https://pubmed.ncbi.nlm.nih.gov/27185916/. Accessed November 18, 2020.
Nicholson, W.L., Munakata, N., Horneck, G., Melosh, H.J., and Setlow, P. (2000) Resistance of *Bacillus* Endospores to Extreme Terrestrial and Extraterrestrial Environments. *Microbiol Mol Biol Rev* **64**: 548–572 /pmc/articles/PMC99004/?report=abstract. Accessed September 20, 2020.
Nunes, F., Fernandes, C., Freitas, C., Marini, E., Serrano, M., Moran, C.P., *et al.* (2018) SpoVID functions as a non-competitive hub that connects the modules for assembly of the inner and outer spore coat layers in *Bacillus subtilis*. *Mol Microbiol* **110**: 576–595 https://europepmc.org/articles/PMC6282716. Accessed October 6, 2020.
Ollington, J.F., and Losick, R. (1981) A cloned gene that is turned on at an intermediate stage of

spore formation in *Bacillus subtilis*. *J Bacteriol* **147**: 443–451 /pmc/articles/PMC216063/?report=abstract. Accessed September 17, 2020.
Ozin, A.J., Costa, T., Henriques, A.O., and Moran, J. (2001a) Alternative translation initiation produces a short form of a spore coat protein in *Bacillus subtilis*. *J Bacteriol* **183**: 2032–2040.
Ozin, A.J., Henriques, A.O., Yi, H., and Moran, C.P. (2000) Morphogenetic proteins SpoVID and SafA form a complex during assembly of the *Bacillus subtilis* spore coat. *J Bacteriol* **182**: 1828–1833.
Ozin, A.J., Samford, C.S., Henriques, A.O., and Moran, C.P. (2001b) SpoVID guides SafA to the spore coat in *Bacillus subtilis*. *J Bacteriol* **183**: 3041–3049 http://jb.asm.org/. Accessed November 13, 2020.
Pereira, F.C., Nunes, F., Cruz, F., Fernandes, C., Isidro, A.L., Lousa, D., *et al.* (2019) A LysM Domain intervenes in sequential protein-protein and protein-peptidoglycan interactions important for spore coat assembly in *Bacillus subtilis*. *J Bacteriol* **201** http://jb.asm.org/. Accessed August 30, 2020.
Plácido, D., Fernandes, C.G., Isidro, A., Carrondo, M.A., Henriques, A.O., and Archer, M. (2008) Auto-induction and purification of a *Bacillus subtilis* transglutaminase (Tgl) and its preliminary crystallographic characterization. *Protein Expr Purif* **59**: 1–8.
Portinha, I. (2015) Exploring the evolutionary link between biofilms and spores formation in spore-formers. .
Price, K.D., and Losick, R. (1999) A four-dimensional view of assembly of a morphogenetic protein during sporulation in *Bacillus subtilis*. *J Bacteriol* **181**: 781–790 http://jb.asm.org/. Accessed September 25, 2020.
Ragkousi, K., and Setlow, P. (2004) Transglutaminase-mediated cross-linking of GerQ in the coats of *Bacillus subtilis* spores. *J Bacteriol* **186**: 5567–5575 /pmc/articles/PMC516844/?report=abstract. Accessed September 29, 2020.
Ramamurthi, K.S., Clapham, K.R., and Losick, R. (2006) Peptide anchoring spore coat assembly to the outer forespore membrane in *Bacillus subtilis*. *Mol Microbiol* **62**: 1547–1557 http://doi.wiley.com/10.1111/j.1365-2958.2006.05468.x. Accessed September 25, 2020.
Roels, S., Driks, A., and Losick, R. (1992) Characterization of spoIVA, a sporulation gene involved in coat morphogenesis in *Bacillus subtilis*. *J Bacteriol* **174**: 575–585 /pmc/articles/PMC205752/?report=abstract. Accessed September 26, 2020.
Romero, D., Vlamakis, H., Losick, R., and Kolter, R. (2011) An accessory protein required for anchoring and assembly of amyloid fibres in *B. subtilis* biofilms. *Mol Microbiol* **80**: 1155–1168 https://pubmed.ncbi.nlm.nih.gov/21477127/. Accessed September 16, 2020.
Sambrook, J., Fritsch, E.F., and Maniatis, T. (2012) *Molecular cloning: a laboratory manual.* 4th ed., Cold Spring Harbor Laboratory Press, .
Sauer, E. (2013) Structure and RNA-binding properties of the bacterial LSm protein Hfq. *RNA Biol* **10**: 610–618 https://www.tandfonline.com/doi/abs/10.4161/rna.24201. Accessed December 1, 2020.
Serrano, M., Zilhao, R., Ricca, E., Ozin, A.J., Moran, C.P., and Henriques, A.O. (1999) A *Bacillus subtilis* secreted protein with a role in endospore coat assembly and function. *J Bacteriol* **181**: 3632–3643 http://jb.asm.org/. Accessed November 4, 2020.
Setlow, P. (2006) Spores of *Bacillus subtilis*: Their resistance to and killing by radiation, heat and chemicals. In *Journal of Applied Microbiology*. John Wiley & Sons, Ltd, pp. 514–525 https://sfamjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2672.2005.02736.x. Accessed September 20, 2020.
Setlow, P. (2007) I will survive: DNA protection in bacterial spores. *Trends Microbiol* **15**: 172–180 https://pubmed.ncbi.nlm.nih.gov/17336071/. Accessed September 23, 2020.
Setlow, P., Wang, S., and Li, Y.Q. (2017) Germination of Spores of the Orders Bacillales and Clostridiales. *Annu Rev Microbiol* **71**: 459–477 https://pubmed.ncbi.nlm.nih.gov/28697670/. Accessed September 21, 2020.
Shafikhani, S.H., Mandic-Mulec, I., Strauch, M.A., Smith, I., and Leighton, T. (2002) Postexponential regulation of sin operon expression in *Bacillus subtilis*. *J Bacteriol* **184**: 564–571 https://pubmed.ncbi.nlm.nih.gov/11751836/. Accessed November 5, 2020.
Shuster, B., Khemmani, M., Abe, K., Huang, X., Nakaya, Y., Maryn, N., *et al.* (2019)

Contributions of crust proteins to spore surface properties in *Bacillus subtilis*. *Mol Microbiol* **111**: 825–843 /pmc/articles/PMC6417949/?report=abstract. Accessed November 5, 2020.
Sonenshein, A.L., Hoch, J.A., and Losick, R. (2001) *Bacillus subtilis*: From Cells to Genes and from Genes to Cells. In *Bacillus subtilis and Its Closest Relatives*. Sonenshein, A.L., Hoch, J.A., and Losick, R. (eds). ASM Press, Washington, DC, USA. pp. 1–5 http://doi.wiley.com/10.1128/9781555817992.ch1. Accessed November 5, 2020.
Stevens, C.M., Daniel, R., Illing, N., and Errington, J. (1992) Characterization of a sporulation gene, spoIVA, involved in spore coat morphogenesis in *Bacillus subtilis*. *J Bacteriol* **174**: 586–594 /pmc/articles/PMC205753/?report=abstract. Accessed September 26, 2020.
Stöver, A.G., and Driks, A. (1999) Secretion, localization, and antibacterial activity of TasA, a *Bacillus subtilis* spore-associated protein. *J Bacteriol* **181**: 1664–1672 http://jb.asm.org/. Accessed November 4, 2020.
Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**: 207–234.
Sullivan, M.A., Yasbin, R.E., and Young, F.E. (1984) New shuttle vectors for *Bacillus subtilis* and Escherichia coli which allow rapid detection of inserted fragments. *Gene* **29**: 21–26.
Takamatsu, H., Imamura, D., Kuwana, R., and Watabe, K. (2009) Expression of YeeK during *Bacillus subtilis* sporulation and localization of YeeK to the inner spore coat using fluorescence microscopy. *J Bacteriol* **191**: 1220–1229.
Takamatsu, H., Kodama, T., Imamura, A., Asai, K., Kobayashi, K., Nakayama, T., *et al.* (2000) The *Bacillus subtilis* yabG gene is transcribed by SigK RNA polymerase during sporulation, and yabG mutant spores have altered coat protein composition. *J Bacteriol* **182**: 1883–1888 https://pubmed.ncbi.nlm.nih.gov/10714992/. Accessed July 16, 2020.
Takamatsu, H., Kodama, T., Nakayama, T., and Watabe, K. (1999) Characterization of the yrbA gene of *Bacillus subtilis*, involved in resistance and germination of spores. *J Bacteriol* **181**: 4986–4994 http://www.ncbi.nlm.nih.gov/pubmed/10438771. Accessed December 4, 2019.
Terra, R., Stanley-Wall, N.R., Cao, G., and Lazazzera, B.A. (2012) Identification of *Bacillus subtilis* sipw as a bifunctional signal peptidase that controls surface-adhered biofilm formation. *J Bacteriol* **194**: 2781–2790 http://jb.asm.org/. Accessed September 16, 2020.
Veening, J.W., Hamoen, L.W., and Kuipers, O.P. (2005) Phosphatases modulate the bistable sporulation gene expression pattern in *Bacillus subtilis*. *Mol Microbiol* **56**: 1481–1494 https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2958.2005.04659.x. Accessed November 3, 2020.
Verhamme, D.T., Murray, E.J., and Stanley-Wall, N.R. (2009) DegU and Spo0A jointly control transcription of two loci required for complex colony development by *Bacillus subtilis*. *J Bacteriol* **91**: 100–108 http://jb.asm.org/. Accessed September 17, 2020.
Vlamakis, H., Chai, Y., Beauregard, P., Losick, R., and Kolter, R. (2013) Sticking together: Building a biofilm the *Bacillus subtilis* way. *Nat Rev Microbiol* **11**: 157–168 https://pubmed.ncbi.nlm.nih.gov/23353768/. Accessed August 2, 2020.
Wang, K.H., Isidro, A.L., Domingues, L., Eskandarian, H.A., McKenney, P.T., Drew, K., *et al.* (2009) The coat morphogenetic protein SpoVID is necessary for spore encasement in *Bacillus subtilis*. *Mol Microbiol* **74**: 634–649 http://doi.wiley.com/10.1111/j.1365-2958.2009.06886.x. Accessed November 18, 2020.
Warth, A.D., Ohye, D.F., and Murrell, W.G. (1963) The composition and structure of bacterial spores. *J Cell Biol* **16**: 579–592 /pmc/articles/PMC2106236/?report=abstract. Accessed September 23, 2020.
Winkelman, J.T., Bree, A.C., Bate, A.R., Eichenberger, P., Gourse, R.L., and Kearns, D.B. (2013) RemA is a DNA-binding protein that activates biofilm matrix gene expression in *Bacillus subtilis*. *Mol Microbiol* **88**: 984–997 https://pubmed.ncbi.nlm.nih.gov/23646920/. Accessed November 5, 2020.
Zheng, L.B., Donovan, W.P., Fitz-James, P.C., and Losick, R. (1988) Gene encoding a morphogenic protein required in the assembly of the outer coat of the *Bacillus subtilis* endospore. *Genes Dev* **2**: 1047–1054 https://pubmed.ncbi.nlm.nih.gov/3139490/. Accessed September 26, 2020.
Zilhão, R., Isticato, R., Martins, L.O., Steil, L., Völker, U., Ricca, E., *et al.* (2005) Assembly and

function of a spore coat-associated transglutaminase of *Bacillus subtilis*. *J Bacteriol* **187**: 7753–7764.

Zilhão, R., Naclerio, G., Henriques, A.O., Baccigalupi, L., Moran, C.P., and Ricca, E. (1999) Assembly requirements and role of CotH during spore coat formation in *Bacillus subtilis*. *J Bacteriol* **181**: 2631–2633 http://jb.asm.org/. Accessed September 17, 2020.

Zilhão, R., Serrano, M., Isticato, R., Ricca, E., Moran, C.P., and Henriques, A.O. (2004) Interactions among CotB, CotG, and CotH during Assembly of the *Bacillus subtilis* Spore Coat. In *Journal of Bacteriology*. American Society for Microbiology Journals, pp. 1110–1119 http://jb.asm.org/. Accessed September 22, 2020.

Appendices

Table 1. E. coli and *B. subtilis* strains used in this study

Strain	Genotype	Origin
E. coli		
DH5α	F- φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_K^-, m_K^+) *pho*A *sup*E44 λ^- *thi*-1 *gyr*A96 *rel*A1	GIBCO/BRL
BL21(DE3)	F- *ompT gal dcm lon hsdSB*(*rB–mB–*) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*+]K-12(λS)	Novagen
AH4611	BL21(DE3) pLOM4	(Plácido *et al.*, 2008)
AHEC1100	BL21(DE3) pBG1	This work
AHEC1101	BL21(DE3) pBG2	“
AH10317	BL21(DE3) pCF68	(Fernandes *et al.*, 2019)
AH10326	BL21(DE3) pCF70	Lab stock
AH10529	BL21(DE3) pCF138	“
AH10530	BL21(DE3) pCF146	“
AH10531	BL21(DE3) pCF174	“
B. subtilis		
MB24	*trpC2 metC3*, wild-type	Laboratory stock
MSB180	*trpC2 metC3 Δveg::erm*	This work
MSB182	*trpC2 metC3 Δveg::erm* Δ*amyE*:: P_{gerE}-*veg*	“
MSB183	*trpC2 metC3* Δ*amyE*:: P_{gerE}-*veg*	“
MSB184	*trpC2 metC3 pBG7*	“
MSB185	*trpC2 metC3 pMK3*	“
MSB186	*trpC2 metC3 Δveg::erm pBG7*	“
MSB187	*trpC2 metC3 Δveg::erm pMK3*	“

Table 2. Growth media

Medium	Composition
Luria broth	10 g/L tryptone, 5 g/L yeast extract, 0.5 g/L NaCl
Luria agar	Tryptone 10 g/L, yeast extract 5 g/L, NaCl 0.5 g/L, agar 16 g/L
Auto-induction medium	$MgSO_4$ 1 mM, 5052 1 X, NPS 1 X, LB medium 92.9% (v/v) (Studier, 2005)
Difco sporulation medium (DSM)	Nutrient broth 8 g/L, KCl 1 g/L, $MgSO_4.7H_2O$ 0.25 g/L
Growth medium 1 (GM1)	B&W salts 96% (v/v), $MgSO_4$ 1mM, glucose 0.5% (w/v), yeast extract 0.1% (w/v), B&W amino acids 0.05 X
Growth medium 2 (GM2)	GM1 96.5% (v/v), $CaCl_2$ 0.25 mM, $MgCl_2$ 25 mM

Table 3. Antibiotics

Antibiotic	Stock concentration (mg/ml)	Solvent	Medium concentration (µg/ml)
Ampicillin	100	ddH_2O	100
Chloramphenicol	30	95% ethanol	20
Kanamycin	30	ddH_2O	10
Erythromycin	10	95% ethanol	1
Neomycin	10	ddH_2O	1 (liquid) 3 (solid)

Table 4. Plasmids

Plasmid	Description	Origin
pLOM4	Plasmid derived from pET30a carrying the tgl^{WT} gene with the His-tag, for overproduction of Tgl-His_6; Km^R	(Zilhão *et al.*, 2005)
pCF68	Plasmid derived from pET28a carrying the 3' end of the *safA* gene with the His-tag, for overproduction of $SafA^{C30}$-His_6; Km^R	(Fernandes, 2014)
pCF70	Plasmid derived from pET28a carrying the *safA* gene with the His-tag, for overproduction of $SafA^{FL}$-His_6; Km^R	Laboratory stock
pCF138	Plasmid derived from pCF68 carrying the 3' end of the *safA* gene with the His-tag, for overproduction of $SafA^{C30(K318A)}$-His_6; Km^R	"
pCF146	Plasmid derived from pCF68 carrying the 3' end of the *safA* gene with the His-tag, for overproduction of $SafA^{C30(K177A)}$-His_6; Km^R	"
pCF174	Plasmid derived from pCF173 carrying the 3' end of the *safA* gene with the His-tag, for overproduction of $SafA^{C30(K177/318A)}$-His_6; Km^R	"
pET16b	Amp^R	Novagen
pBG1	Plasmid derived from pET16b carrying the *veg* gene (from *Clostridioides difficile*) with the His-tag, for overproduction of Veg; Amp^R	This work
pBG2	Plasmid derived from pET16b carrying the 5' end of the *safA* gene with the His-tag, for overproduction of N21-His_6; Amp^R	"
pMK3	Neo^R	(Sullivan *et al.*, 1984)
pBG7	Plasmid derived from pMK3 carrying the *veg* gene with the promoter from the *gerE* gene (P_{gerE}*::veg* from *B. subtilis*); Neo^R	This work
pDG364	Cm^R	(Harwood and Cutting, 1990)
pBG6	Plasmid derived from pDG364 inserted with the *veg* gene with the promoter from the *gerE* gene (P_{gerE}*::veg* from *B. subtilis*); Cm^R	This work

Table 5. Primers used in this study

Primer	Oligonucleotide sequence
safA490Fw	5’ CATGCCATGGCCTTGAAAATCCATATCG 3’
safA980Rv	5’ CCGCTCGAGTTAGTGATGATGATGATGATGCATAGCCTCCTGTTGTGG 3’
Veg600Fwd	5’ CATGCCATGGCCGTGGCTACTGTTCAAACTC 3’
Veg764Rv	5’ CCGCTCGAGTTAGTGATGATGATGATGATGACTTATTTGTAATTTATCTTG 3’
vegbs500fwd	5’ ATGGCGAAGACGTTGTCCG 3’
vegbs768fwd	5’ CGCGGATCCAAATGCCACTGAGCTTGC 3’
ger1D	5’ CGTGAAGCTTCAAGTCTCAAAGGAAAG 3’
ger480revvegbs	5’ GGACAACGTCTTCGCCATGTATTGTAACCCTCCTTGC 3’

Table 6. Solutions and buffers

Solution	Composition
Buffer A	2mM CAPS pH 10.0, 0.5M NaCl, 10mM imidazole
Lysis buffer	50mM NaH_2PO_4, 0.5M NaCl, 10mM imidazole, pH 8.0
French Press buffer	10 mM imidazole, 1 X Phosphate buffer, 1 mM PMSF, pH 7.4
Elution buffer	1 X Phosphate buffer, pH 7.4, variable imidazole concentrations
Dialysis buffer	0.15M NaCl, 0.1M Tris-HCl, pH 8.0; with 2mM of DTT
SDS-PAGE gels	10% resolving: 48% distilled water, 25% 4x lower Tris buffer, 10% acrylamide:bis-acrylamide 29:1, 0.1% SDS, 0.1% ammonium peroxydisulphate, 0.05% tetramethylethylenediamine; 12.5% resolving: 40.3% distilled water, 25% 4x lower Tris buffer, 12.5% acrylamide:bis-acrylamide 29:1, 0.1% SDS, 0.1% ammonium peroxydisulphate, 0.05% tetramethylethylenediamine; 4.5% stacking: 61.1% distilled water, 25.5% 4x lower Tris buffer, 4.5% acrylamide:bis-acrylamide 29:1, 0.1% SDS, 0.1% ammonium peroxydisulphate, 0.1% tetramethylethylenediamine
SDS-PAGE reducing loading buffer 10 X	Tris-HCl 62.5 mM pH6.8, SDS 2% (w/v), glycerol 25% (v/v), 0.5mM DTT, 2-mercaptoethanol 0.5% (v/v) bromophenol blue 0.02% (w/v)
SDS-PAGE non-reducing loading buffer 10 X	Tris-HCl 62.5 mM pH6.8, SDS 2% (w/v), glycerol 25% (v/v), bromophenol blue 0.02% (w/v)
SDS-PAGE loading buffer 2 X	Upper Tris 1 X, glycerol 5% (v/v), 2% (w/v) SDS, 0.5 mM DTT, 5% 2-mercaptoethanol, 0.025% (w/v) bromophenol blue
Coomassie solution	0.5g/ml Coomassie Brilliant Blue R-250, 80% absolute ethanol, 20% acetic acid
Destaining solution	30% absolute ethanol, 10% acetic acid
TAE1X	40 mM Tris, 20 mM acetic acid, 1 mM EDTA
STET buffer	8% sucrose, 0.5% TritonX-100, 50 mM EDTA, 10 mM Tris-HCl pH 8.0
Orange G loading buffer	0.2% Orange G, 0.37% EDTA, 50% glycerol
Transfer buffer	14.4 g/L glycine, 3.02 g/L Tris base, 10% ethanol
PBS-T	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, Tween® 20 detergent 0.1% (w/v)
Iodine solution	1% (w/v) I_2, 10% (w/v) KI
Bott and Wilson salts (B&W salts)	K_2HPO_4 12.4 g/L, KH_2PO_4 7.6 g/L, tri-sodium citrate 1 g/L, $(NH_4)_2SO_4$ 6 g/L, pH 6.7
Bott and Wilson amino acids (B&W amino acids)	Trp, arg, lys, gly, met, his, val, thr, and asp 2.5 mg/ml per amino acid