Proteins with RBM(ring-building motif)-like domain involved in bactaria sporulation

Bowen Liu

Université Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : **BIOLOGIE STRUCTURALE ET NANOBIOLOGIE**

Arrêté ministériel : 25 mai 2016

Présentée par

Bowen LIU

Thèse dirigée par **Cécile MORLOT**
et co-encadrée par **Thierry VERNET**

préparée au sein de l'**Institut de Biologie Structurale**
dans **l'École Doctorale Chimie & Sciences du Vivant**

Protéines avec un domaine de type RBM impliquées dans la sporulation bactérienne

RBM-like proteins involved in bacterial sporulation

Thèse soutenue publiquement le **09 Novembre 2020**,
devant le jury composé de :

Mme Cécile BREYTON
Directrice de recherche CNRS, IBS, Grenoble (Présidente)
Mme. Véronique BROUSSOLLE
Directrice de recherche INRA, Avignon (Rapportrice)
M. Romé VOULHOUX
Directeur de recherche CNRS, LCB, Marseille (Rapporteur)
Mme. Stéphanie RAVAUD
Maître de conférence Université Lyon 1, MMSB, Lyon (Examinatrice)
M. Juan FONTECILLA-CAMPS
Ingénieur-chercheur CEA, IBS, Grenoble (Examinateur)
Mme Simonetta GRIBALDO
Directrice de recherche CNRS, Institut Pasteur, Paris (Examinatrice)
Mme Cécile MORLOT
Directrice de recherche CNRS, IBS, Grenoble (Directrice de thèse)

Acknowledgement

As I finished this thesis and approaching the next milestone in my life, I am really willing to give my sincere gratitude to those who have kept supporting me.

First of all, I would like to thank my supervisor Cécile Morlot with all my heart. Cécile is such an intellectual and serious scholar who always gives me a lot of instruction and inspiration about my project. She also checked and approved each part in my thesis elaborately with her utmost patience. I am deeply grateful for her help in completing my thesis.

I would also thanks to Caroline Mas (EMBL), Carlos CONTRERAS-MARTEL (Bacterial Pathogenesis Group, IBS), Emmanuelle Neumann and Daphna Fenel (EM platform, IBS), our collaborators Christopher Rodrigues (University of Sydney, Australia), Adriano Henriques (ITQB, Portugal), Laure Bellard, Anne-Marie Villard and all my other colleagues in pneumococcus group at IBS for the helpful suggestions and technical support of my project.

In addition, Cécile and my group leader Dr. Thierry Vernet give me a lot of concern and care, especially when I firstly arrived in France and when I was isolated at home due to the Covid-19 epidemic. That helps me through the most tough period. I would also thanks to all my colleagues in pneumococcus group at IBS. In the past three years, they have given me not only support of work but also their care about my daily life. It's my pleasure to study and work with you.

Thanks for Dr. Cécile BREYTON, Dr. Veronique BROUSSOLLE, Dr. Romé VOULHOUX, Dr. Stéphanie RAVAUD, Dr. Juan FONTECILLA-CAMPS and Dr. Simonetta GRIBALDO for accepting to participate in my thesis defense. I look forward to communicating with you in the thesis defense.

I owed the great appreciation for the financial support of China Scholarship during my Ph.D. This makes my dream of studying abroad come true.

Last but not least, thank my family and my girlfriend on the other side of the earth. Their constant encouragement and emotional support are the most powerful emotional anchor for me to complete my study.

Abstract

Specialized secretion systems found in Gram-negative bacteria allow the transport of molecules across their double-membrane cell envelope. Components of these nanomachines include ring-forming proteins from the PrgK and PrgH families, which are part of the inner membrane platform in Type-III secretion systems, or the InvG and GspD secretins from Type-III and Type-II secretion systems, respectively. Homo-oligomerization of these proteins involves a domain called RBM for "Ring-Building Motif". Despite low sequence identity, RBM domains display a conserved wedge-shaped fold composed of a three-stranded β-sheet packed against two α-helices.

Because the cell envelope of Gram-positive bacteria possess a single membrane, double-membrane spanning machineries are not necessary for secretion. During spore formation in Gram-positive bacteria however, the mother cell engulfs the developing spore, encasing it with a double membrane. Communication between the two cells involves a large multi-protein complex called the SpoIIIA-SpoIIQ complex. The global architecture and function of this new machinery remains mysterious but its components display structural similarities with essential constituents of specialized secretion systems. In particular, some of the SpoIIIA-SpoIIQ proteins possess RBM-like domains and one of them, called SpoIIIAG, forms large oligomeric rings that display remarkable similarities and differences with PrgK and PrgH rings from Type-III secretion systems. Ring formation by SpoIIIAG provides evidence that the SpoIIIA-SpoIIQ complex might serve as a secretion machinery between the mother cell and forespore but assembly of a transenvelope channel requires oligomerization of other SpoIIIA-SpoIIQ proteins.

To get further insights into the capacity of RBM-containing SpoIIIA-SpoIIQ proteins to form rings, I produced, purified and characterized full-length membrane and truncated soluble forms of these proteins. This part of my work showed that the RBM domain alone in SpoIIIAG is not sufficient to promote oligomerization in vitro, and that additional secondary structures observed in non-canonical RBM domain is not what prevents them from forming rings in vitro.

Intriguingly, RBM domains were also found in proteins that are not related to the SpoIIIA-SpoIIQ complex and raised the hypothesis that other putative secretion systems might form during sporulation. In order to investigate this, I studied the structure and oligomerization ability of one of these proteins, which is called YhcN and is likely involved in spore germination.

The crystal structure of YhcN revealed the presence of a non-canonical RBM domain and the protein did not show any oligomerization ability.

Altogether, my work questions the ring-forming function associated with RBM domains and suggests that some of these domains might have evolved to fulfill different roles.

Les systèmes de sécrétion spécialisés des bactéries Gram-négatives permettent le transport de molécules à travers leur enveloppe cellulaire à double membrane. Les composants de ces nanomachines comprennent des protéines de formation de cycle des familles PrgK et PrgH, qui font partie de la plate-forme de la membrane interne dans les systèmes de sécrétion de type III, ou les sécrétines InvG et GspD des systèmes de sécrétion de type III et de type II, respectivement. L'homo-oligomérisation de ces protéines implique un domaine appelé RBM pour "Ring-Building Motif". Malgré une faible identité de séquence, les domaines RBM présentent un pli cunéiforme conservé, composé d'une feuille β à trois brins et de deux hélices α.

Comme l'enveloppe cellulaire des bactéries Gram-positives possède une seule membrane, les machines à double membrane ne sont pas nécessaires pour la sécrétion. Cependant, lors de la formation des spores chez les bactéries Gram-positives, la cellule mère engloutit la spore en développement, l'enveloppant d'une double membrane. La communication entre les deux cellules implique un grand complexe multiprotéique appelé le complexe SpoIIIA-SpoIIQ. L'architecture globale et la fonction de cette nouvelle machinerie restent mystérieuses mais ses composants présentent des similitudes structurelles avec les constituants essentiels des systèmes de sécrétion spécialisés. En particulier, certaines des protéines SpoIIIA-SpoIIQ possèdent des domaines de type RBM et l'une d'entre elles, appelée SpoIIIAG, forme de grands anneaux oligomériques qui présentent des similarités et des différences remarquables avec les anneaux PrgK et PrgH des systèmes de sécrétion de type III. La formation d'anneaux par la SpoIIIAG fournit la preuve que le complexe SpoIIIA-SpoIIQ pourrait servir de mécanisme de sécrétion entre la cellule mère et le pré-spore, mais l'assemblage d'un canal trans-enveloppe nécessite l'oligomérisation d'autres protéines SpoIIIA-SpoIIQ.

Pour mieux comprendre la capacité des protéines SpoIIIA-SpoIIQ contenant la RBM à former des anneaux, j'ai produit, purifié et caractérisé des formes solubles de ces protéines sur toute la longueur de la membrane et sous forme tronquée. Cette partie de mon travail a montré que le domaine RBM seul dans la SpoIIIAG n'est pas suffisant pour promouvoir l'oligomérisation in vitro, et que les structures secondaires supplémentaires observées dans le domaine RBM non canonique ne sont pas ce qui les empêche de former des cycles in vitro.

De façon intrigante, des domaines RBM ont également été trouvés dans des protéines qui ne sont pas liées au complexe SpoIIIA-SpoIIQ et ont soulevé l'hypothèse que d'autres systèmes de sécrétion putatifs pourraient se former pendant la sporulation. Afin d'étudier cela, j'ai étudié la structure et la capacité d'oligomérisation de l'une de ces protéines, qui est appelée YhcN et est probablement impliquée dans la germination des spores. La structure cristallographique de YhcN a révélé la présence d'un domaine RBM non canonique et la protéine n'a montré aucune capacité d'oligomérisation.

Dans l'ensemble, mes travaux remettent en question la fonction de formation d'anneau associée aux domaines RBM et suggèrent que certains de ces domaines pourraient avoir évolué pour remplir différents rôles.

Acronyms and abbreviation

TXSS: type X secretion system

RBM: ring-building motif

OM: outer membrane

OMP: outer membrane platform

IM: inner membrane

IMP: inner membrane platform

A-Q complex: SpoIIIA-SpoIIQ complex

TMS: transmembrane-segment

EM: electron microscopy

CMC: critical micelle concentration

DDM: n-dodecyl β-D-maltoside

Ni-NTA: nickel-nitrilotriacetic acid

SUMO: small ubiquitin-like modifier

SEC-MALLS: size exclusion chromatography-multi angle laser light scattering

MST: microscale thermophoresis

PDB: protein data bank

RMSD: root-mean-square deviation

Table of contents

Introduction

I. Specialized secretion across bacterial membranes

A. Generalities about secretion through bacterial membranes

The transport of molecules out of the bacterial cell requires protein complexes assembling in the membrane(s) of bacteria, allowing the secretion of endogenous molecules (Green and Mecsas 2016). Transport across the bacterial membrane is involved in an array of processes such as development, movement, conjugation, adhesion to host cells and surfaces, virulence, host symbiosis, as well as bacterial competition.

Based on their structure, function and specificity, secretion systems have been categorized into several classes. Some of the transport systems are only found in a small number of species while others are conserved in many bacterial species. Table 1 summarizes some of the major features of the secretion systems that have been described so far.

Table1. Classes of bacterial specialized secretion systems.

Secretion Apparatus	Secretion Signal	Steps	Folded Substrates	Number of Membranes	Gram (+) or Gram (−)
T1SS	C-terminus	1	No	2	Gram (−)
T2SS	Unknown	2	Yes	1	Gram (−)
T3SS	N-terminus	1-2	No	2-3	Gram (−)
T3SS flagellar protein export apparatus	N-terminus	1	No	1	Both
T4SS	C-terminus	1	No	2-3	Gram (−)
T5SS	N-terminus	2	No	1	Gram (−)
T6SS	Unknown	1	Unknown	2-3	Gram (−)
T7SS	C-terminus	1	Yes	1-3	Gram (+)
Sec	N-terminus	1	No	1	Both
Tat	N-terminus	1	Yes	1	Both
Sortase	N-terminus (Sec) C-terimnus (cws)	2	Yes	1	Gram (+)
T9SS	C-terminus	2	Yes	2	Gram (−)

The selectivity for the transported molecule vary from one system to another. It can be based on the size and/or the nature of the molecule. Some systems will thus transport a broad array of substrates while others will be specific to only one or a few molecules.

Depending on the secretion system, the secreted substrate(s) can either remain associated with the bacterial membrane, or be released into the extracellular environment. In some cases, the substrate(s) will be injected into a eukaryotic or bacterial cell.

The transport happens across a single or a double bacterial membrane. In that regard, a major difference between Gram-negative and Gram-positive bacteria regarding the secretion of substances is the composition of their cell envelope. Gram-positive bacteria are surrounded by a plasma membrane and a thick layer of peptidoglycan (a polymer that is made of glycan chains cross-linked by peptide chains, and that confers the cell shape), while the cell envelope of Gram-negative bacteria contains an inner (plasma) membrane, a thin layer of peptidoglycan and an outer membrane. In Gram-negative bacteria, macromolecular systems allowing transport across the cell envelope will thus have to span the two membranes. This mechanism can happen in one or two steps (Green and Mecsas 2016, Costa et al. 2015).

The proteins that I have studied during my Ph.D. are membrane-anchored proteins specifically produced during bacterial sporulation. Although most spore-forming bacteria are Gram-positive, the developing spore is surrounded by two membranes (see chapter II), which are reminiscent of the inner and outer membranes found in Gram-negative bacteria. Intriguingly enough, my proteins of interest display weak sequence identity but obvious structural similarities with components of specialized secretion systems found in Gram-negative bacteria. For this reason, I have focused the following sections on these particular secretion systems.

B. Secretion systems in Gram-negative bacteria

In Gram-negative bacteria, transport across the inner and outer membranes is carried out by secretion systems that can be divided into three groups (Costa et al. 2015).

A first group includes protein complexes than span the inner membrane only. It includes the Sec system, which transports most of the secreted proteins across the inner membrane (Lycklama and Driessen 2012). Secretion across the cytoplasmic membrane can also be carried out by the Twin-arginine translocation (Tat) system, which allows fully folded proteins to pass through (Palmer and Berks 2012). The Sec and Tat pathways are the most highly conserved mechanisms for protein secretion and are present in all domains of life (bacteria, archaea and eukarya). Proteins delivered to the periplasmic space by these two systems can be subsequently transported across the outer membrane by another secretion system (such as the Type-II, Type-V and Type IX secretions systems) (Fig. 1).

A second group includes protein complexes that span the outer membrane only, such as the Type-V secretion system (Leo, Grin, and Linke 2012). By contrast with other secretion systems that span the inner membrane, T5SS does not require a transport energy source. This self-sufficient autotransporter transports proteins through the outer membrane through a β-barrel transmembrane machinery (Meuskens et al. 2019, Bernstein 2019) (Fig. 1).

Finally, a third group includes protein nanomachines that span both the inner and outer membranes. So far, five systems have been identified in this last group, and categorized as **Type-I**, **Type-II**, **Type-III, Type-IV** and **Type-VI secretion systems** (**T1SS**, **T2SS**, **T3SS**, **T4SS** and **T6SS**, respectively), based on their structural components and function (Costa et al. 2015, Green and Mecsas 2016). These double-membrane-spanning secretion systems are large multiprotein complexes allowing the secretion of specific molecules, including proteins and DNA, from the bacterial cytoplasm to the extracellular environment. Most systems exclusively secrete unfolded or partially unfolded proteins. Only T2SS and T6SS can secrete folded and partially folded proteins. So far, only T4SS has been shown to translocate DNA. Translocation, which is driven by ATP hydrolysis, happens in one (For T1SS, T3SS, T4SS and T6SS) or two steps (For T2SS). When proteins are translocated in two separate steps by T2SS, they are first

delivered to the cytoplasm through the Sec or the Tat pathway. Finally, the secreted substrate(s) is(are) either released in the extracellular space (For T1SS, T2SS and T4SS) (Fig. 1A) or injected into a target cell (For T3SS and T6SS).

The common structural feature of Type-I to Type-VI secretion systems is an oligomeric ring formed by outer membrane proteins, resulting in a β-barrel structure that allows secretion of substrates through the outer membrane. These systems otherwise display major differences in their structure, assembly and secretion mechanisms.

In 2008, bioinformatics analyses (HHPRED analyses) based on pairwise comparison of the primary sequence and on secondary structure predictions have identified a group of proteins produced during bacterial sporulation (The SpoIIIA proteins, see chapter II) that would have secondary structures organized similarly to those of components found in T1SS, T2SS, T3SS and T4SS (Camp and Losick 2008, Meisner et al. 2008). I will thus focus the next sections on these specific double-membrane-spanning nanomachines (Fig. 1), which are the most pertinent transport systems relative to my Ph.D. I will summarize the state-of-the-art regarding the structure and function of these systems, with a particular emphasis on components that share structural similarities with my proteins of interest.

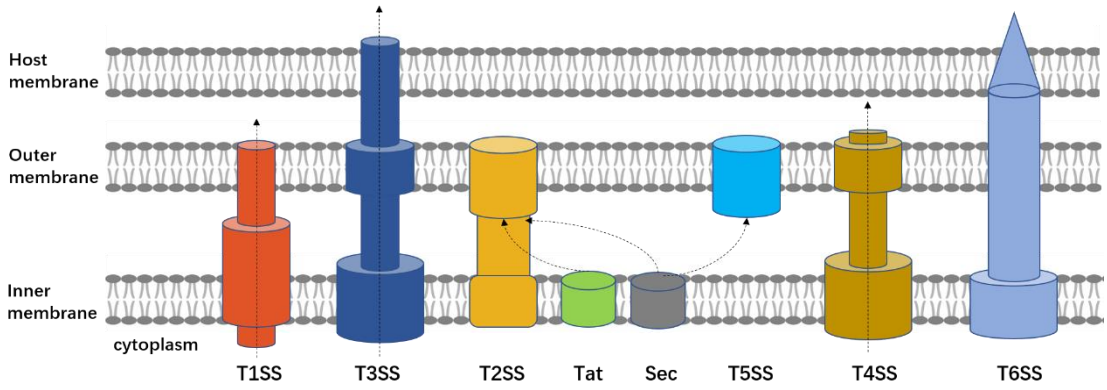


Figure 1. Schematic diagram of specialized secretion systems in Gram-negative bacteria. T1SS, T3SS, T4SS and T6SS transport the substrates through the inner membrane (IM) and outer membrane (OM) into the extracellular space in one step. T3SS and T6SS inject the substrate(s) into the host membrane. T2SS and T5SS rely on the Sec or Tat pathway to realize a two-step secretion.

1. Type-I secretion systems

Type-I Secretion Systems (T1SS) are present in a large number of Gram-negative bacteria, especially in pathogens of plants and animals. Since the discovery of the first T1SS substrate (the hemolysin A) in 1979, many other virulence factors have been shown to be secreted through T1SS (Noegel et al. 1979). The size of the substrates range from 10 kDa (e.g. the bacteriocins) to 1 MDa (e.g. the gigantic adhesins). Their function is also diverse, and include digestive enzymes, adhesins, heme-binding proteins and toxins. Bacteria can have several T1SS, specific of one or a few substrates.

T1SS are composed of three different membrane proteins. An **ATP-binding cassette** (**ABC**) transporter protein anchored in the inner membrane and an **outer membrane factor** (**OMF**) from the **TolC family** anchored in the outer membrane. Finally, a **membrane fusion protein** (**MFP**) crosses the periplasmic space and bridges the ABC transporter to the OMF (Fig. 2A) (Green and Mecsas 2016, Kanonenberg et al. 2018, Costa et al. 2015).

The ABC transporter is responsible for substrate recognition and ATP hydrolysis. The crystal structure of the ABC transporter PrtD from *Aquifex aeolicus* and MacB from *Escherichia coli* K-12 show that it forms a homodimer, each subunit containing 6 N-terminal transmembrane (TM) helices and a C-terminal nucleotide binding domain (NBD) (Fig. 2B, C) (Morgan, Acheson, and Zimmer 2017, Fitzpatrick et al. 2017). The N-terminal region of the homodimer forms an occluded channel spanning almost the entire TM region. The C-terminal NBDs locate in the cytoplasm and are responsible for ATP binding and hydrolysis.

In *Escherichia coli* K-12 T1SS, MacA (MFP) connect the MacB dimer (ATP-binding protein) and TolC (OMF), forms a hexamer with a central channel that allows substrate translocation (Fitzpatrick et al. 2017) (Fig. 2B). Finally, the TolC (OMF) trimer generates another channel that spans the outer membrane, through which the substrate passes. Most of the substrates possess a C-terminal signal that is recognized by the ABC transporter/MFP complex. The unfolded protein is then directly transported from the cytosol to the extracellular space, the secretion being energized by ATP hydrolysis performed by the ABC transporter. This is the

generally considered “one-step” translocation mechanism.

Interestingly, bioinformatic analyses performed with the CLUSTALW sequence alignment algorithm suggest that PrtD shares similarities with SpoIIIAE, which is involved in *Bacillus subtilis* sporulation. While the sequence identity between PrtD and SpoIIIAE is weak (13% sequence identity, 17% sequence similarity) (Fig. 2D), they both possess an equivalent number and organization of transmembrane helices (see section IIC). SpoIIIAE might thus share 3D structural and functional similarities with ABC transporters from T1SS.

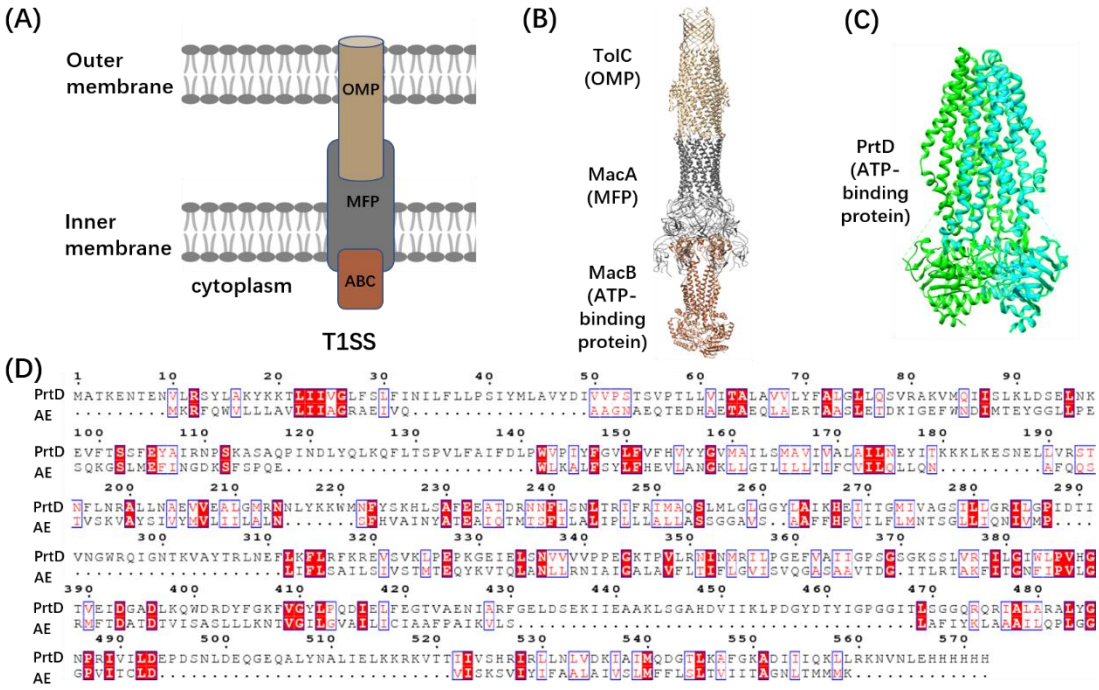


Figure 2. Structural model of T1SS. (A) Basic structure diagram of T1SS. **(B)** Ribbon representation of the structure of the MacA-MacB-TolC pump in *Escherichia coli* K-12 (PDB code: 5NIK). **(C)** Ribbon representation of the structure of the ABC transporter from *A. aeolicus* (PDB code 5L22). **(D)** Alignment of protein sequences of PrtD from *A. aeolicus* and SpoIIIAE (AE) from *B. subtilis*. Conserved residues are in red boxes; similar residues are shown by red letters boxed in blue.

2. Type-II secretion systems

Type-II secretion systems (T2SS) are widespread in Gram-negative bacteria and mediate extracellular delivery of a variety of protein substrates, some of which contribute to the virulence of bacterial pathogens and niche colonization. Most substrates are enzymes, such as proteases, lipases and phosphatases (Korotkov, Sandkvist, and Hol 2012).

T2SS substrates are first delivered by the Sec (Pugsley, Kornacker, and Poquet 1991) or the Tat system (Voulhoux et al. 2001) to the periplasmic space, where they will fold completely. They are then recognized (through still unresolved recognition determinants) by the T2SS apparatus, which will translocate them into the extracellular space via a "piston" mechanism based on the extension and retraction of a pseudopilus (Korotkov and Sandkvist 2019).

T2SS consists of 12-15 different proteins that belong to four subassemblies: the **outer-membrane complex** or **secretin complex** (**GspD/PulD** family of proteins), the **inner-membrane platform** (**GspC**, **GspF**, **GspL** and **GspM**), the **secretion ATPase** (**GspE**), and the **pseudo-pilus** (**GspG**, **GspH**, **GspI**, **GspJ**, **GspK** and **GspO**) (Fig. 3A) (Korotkov and Sandkvist 2019, Green and Mecsas 2016, Gu et al. 2017, Lopez-Castilla et al. 2017).

The GspE ATPase forms a hexamer that resides in the cytoplasm and connects GspF and GspL components of the inner membrane platform. The inner membrane platform is embedded in the plasma membrane and extends into the periplasm. So far, no structure of the inner membrane platform of a T2SS has been observed at high resolution, but by analogy with **electron cryo-tomography** (**ECT**) studies of T4P, it could be made of interconnected cytoplasmic and periplasmic rings. Together with the GspE ATPase, which provides energy to power the system, the inner membrane platform assembles the pseudopilus (Fig. 3A). This filamentous sub-structure is mainly composed of multimers of the major pseudopilin subunit GspG and possibly of minor pseudopilins GspH, I, J and K. GpsG precursors are inserted in the inner membrane through a short N-terminal pre-peptide anchor. After removal of this N-terminal membrane anchor by the GspO peptidase, the pilin subunit then remains associated with the membrane through its hydrophobic tail (Pugsley and Dupuy 1992, Dupuy et al. 1992).

During the assembly of the pseudopilus, the hydrophobic tails are gradually extracted from the inner membrane and pack together to form the inner core of the fiber, while the globular domains interact laterally to form the pseudopilus surface (Kohler et al. 2004, Campos et al. 2010, Nivaskumar and Francetic 2014) (Fig. 3A). Polymerization of the pseudopilus is the mechanism through which the folded substrate will be pushed through the pore formed by the outer membrane complex.

The outer-membrane component is the secretin. This protein forms a homomultimeric ring-like channel (Fig. 3B) through which folded substrates are translocated from the cytoplasm to the extracellular space. The secretin has a long N-terminal region made of two to four small variable domains called **N-subdomains** (**N_0** to **N_3**) (Yan et al. 2017). N-subdomains share a highly **conserved wedged arrangements** that are composed of a **two α-helices packed against a three-stranded anti-parallel β-sheet**. This small mixed α/β-modular domain, which was termed the "**ring-building motif**" (**RBM**), is also found in ring-forming proteins from T3SS (see section B3) (Spreter et al. 2009). Packing of N-subdomains from adjacent protomers form a highly stable periplasmic ring that connects the inner-membrane complex (Korotkov et al. 2011) (Fig. 3C). The central region of the protein is a secretin domain that forms a double β-barrel structure partially inserted in the outer membrane. Each β-barrel contains 48 (for dodecamers) to 60 (for pentadecamers) anti-parallel β-strands. The secretin channel has an outer diameter of 110 to 170 Å and a small pore at the center of the inner barrel (Nouwen et al. 2000, Chami et al. 2005, Tosi et al. 2014, Hay et al. 2018, Yin, Yan, and Li 2018, Yan et al. 2017, Chernyatina and Low 2019). At the C-terminus of the protein, the S domain seems to act as a hook by grabbing β-strands of the neighboring protomer, which likely enhances the stability of the outer β-barrel (Fig. 3B) (Yan et al. 2017). In the closed state, a central gate, and a possible additional cap gate, block translocation of the substrates (Fig. 3D). During secretion, the N3 domain might be pushed back by the substrate and/or the pseudopilus, then the central gate and cap gate can be pushed outwards to allow opening of the channel and substrate release (Fig. 3C, D).

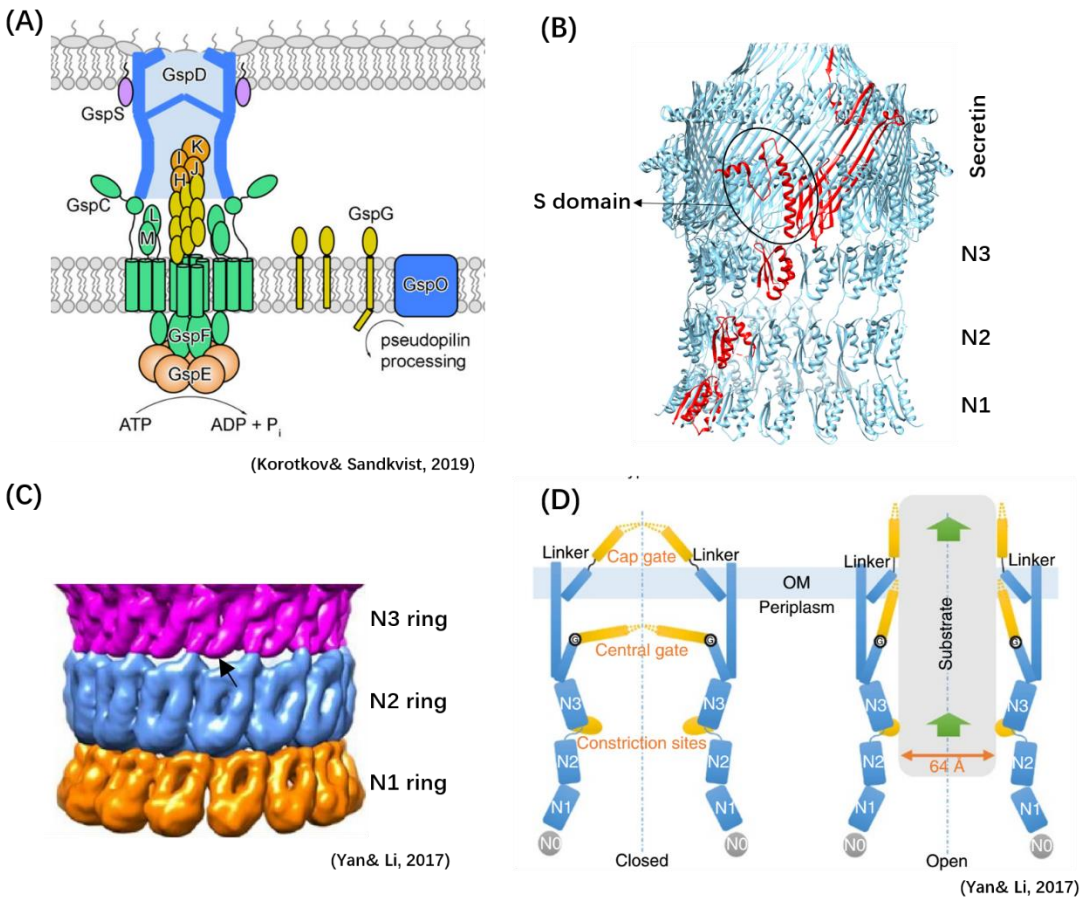


Figure 3. Structural modele of T2SS. (A) Schematic diagram of the topology and location of core components of T2SS. **(B)** Cryo-EM structure of the homo-multimeric GspD secretin complex from *V. cholera* (PDB code: 5WQ8). One of the 15 protomers is colored in red. The position of the rings made by the N1, N2, N3 and secretin domains of the protein are indicated, and the S domain is circled in black. **(C)** Low-resolution cryo-EM structure of the N-terminal rings of the GspD channel. The gaps between N3 and N2 rings show that the interaction between them is weak. **(D)** Schematic diagrams of the GspD channel in a closed state (left) and an open state (right). During secretion, the cap and central gates (in yellow) might open around the linker or glycine (marked as G) regions, and the N3 constriction sites might be pushed back in order to let the substrate pass through. Panel C-D are from Yan& Li et al., 2017.

3. Type-III secretion systems

Type-III secretion systems (T3SS) were discovered in 1993 by G.P. Salmond (Salmond and Reeves 1993). Then in the last few decades, T3SS were found in a large number of Gram-negative bacterial pathogens and have been called "**injectisomes**", or "needle and syringe"-like apparati because of their structure (Buttner 2012). They transport a wide variety of unfolded protein substrates (called "effectors") across the inner and outer membranes in a one-step mechanism, most likely through the inner conduit of their needle component. Many of the T3SS substrates are injected into a host cell to serve as virulence factors.

More than 20 conserved proteins constitute the T3SS apparatus, which is made of two main parts (Burkinshaw and Strynadka 2014, Abrusci et al. 2014) : a **double-membrane-spanning complex** composed of **stacked rings** and a **needle-like filament** that protrudes from the bacterial surface (Fig. 4A). **The transenvelope complex** (or **basal body**) includes the **export apparatus**, which is made of **inner-membrane** (**IM**) and **outer-membrane** (**OM**) **rings**, and an **ATPase complex** in the cytosol (Abrusci et al. 2014). The needle will deliver the toxin into the host (Fig. 4A).

Within the basal body, two homomeric rings are embedded in the inner membrane (Fig. 4B). These rings contain **24 protomers** from the **PrgK/EscJ** and **PrgH/EscD** protein families and display external diameters of 18 nm and 27 nm, respectively (Bergeron et al. 2015). PrgK/EscJ family members are made of two globular domains (Fig. 4C) and a C-terminal transmembrane segment. The two globular domains form two juxtaposed rings at the periplasmic surface of the inner membrane and despite low sequence identity between them or between orthologous domains (less than 30%), they share a conserved RBM fold (see section C). Members of the PrgH/EscD family proteins are composed of a N-terminal cytoplasmic globular domain, a TM segment and three periplasmic RBM domains (Fig. 4D). The three RBM domains of PrgH form homomeric rings that surround the PrgK ring (Spreter et al. 2009, Worrall et al. 2016).

Like T2SS, T3SS possess a secretin to allow the passage through the outer membrane. Secretins from T3SS belong to the **InvG/EscC** family of proteins, in which orthologues can

display rather high sequence identity (more than 60%). Like in T2SS, T3SS secretins form 12-to-15-mer rings, and possess a variable number of N-subdomains (displaying the RBM fold), a secretin and an S domain (Fig. 4E, F). In *Salmonella enterica* serovar Typhimurium pathogenicity island 1 (SPI-1), the InvG secretin contacts the third RBM domain of PrgH through its first N-subdomain (Fig. 4G). Although the asymmetry between the inner (PrgK and PrgH, 24-mers) and outer (InvG, 15-mer) rings appeared surprising at first, the near-atomic cryo-electron microscopy (cryo-EM) structure of the *Salmonella* SPI-1 injectisome obtained recently confirmed that five subunits of InvG are positioned on the top of height PrgK/PrgH subunits (Worrall et al. 2016). In the closed state, the internal diameter of the secretin measures about 15 Å while in the open state, it extends to 75 Å to allow the passage of the needle (Hu et al. 2018).

The needle complex is a helical filament emanating from the **rod complex** (**PrgJ**), which is itself encased at the center of the PrgK ring. The needle is made of more than 100 copies of **PrgI**, which adopts a helix-turn-helix motif that polymerizes along both helices. The PrgI needle is a tube that is 30-70-nm long and 8-nm wide, with an internal helical secretion channel that displays a maximal inner diameter of about 15 Å, allowing the passage of unfolded substrates (Hu et al. 2018). The polymerized needle extends through the secretin pore into the extracellular space. If the substrate must be delivered into a host cell, a translocon complex will assemble at the tip of the needle upon contact with the target cell. The translocon will form a pore in the host cell membrane, through which effectors will be injected.

Note : Cholesterol-dependent cytolysins (CDCs) such as streptolysin O and pneumolysin, which are secreted through the Sec apparatus, assemble into large oligomeric rings that form pores in cholesterol-containing membranes of target eukaryotic cells (Tilley et al. 2005). CDCs have sometimes been described as the "functional equivalent of T3SS in Gram-positive bacteria" or as "injectisomes in Gram-positive bacteria" although they share neither protein sequence similarities nor phylogenetic relationships with any components of T3SS. Furthermore, the sizes of the pores formed by CDCs (30 nm) and T3SS (2 nm) are different. More fundamentally, because effectors are first secreted across the bacterial cell envelope by the Sec apparatus,

CDCs are not strictly bacterial protein secretion systems but mechanisms for protein translocation across the eukaryotic plasma membrane. Finally, there is no evidence that a continuous protected channel that mediates protein transit from one cell to another (which is a fundamental feature of T3SS) is also found in CDCs. Therefore, the only common feature shared by T3SS and CDCs is that effectors are translocated into a eukaryotic host cell. So far, the only authentic Gram-positive type III secretion system is the flagellar protein export apparatus."

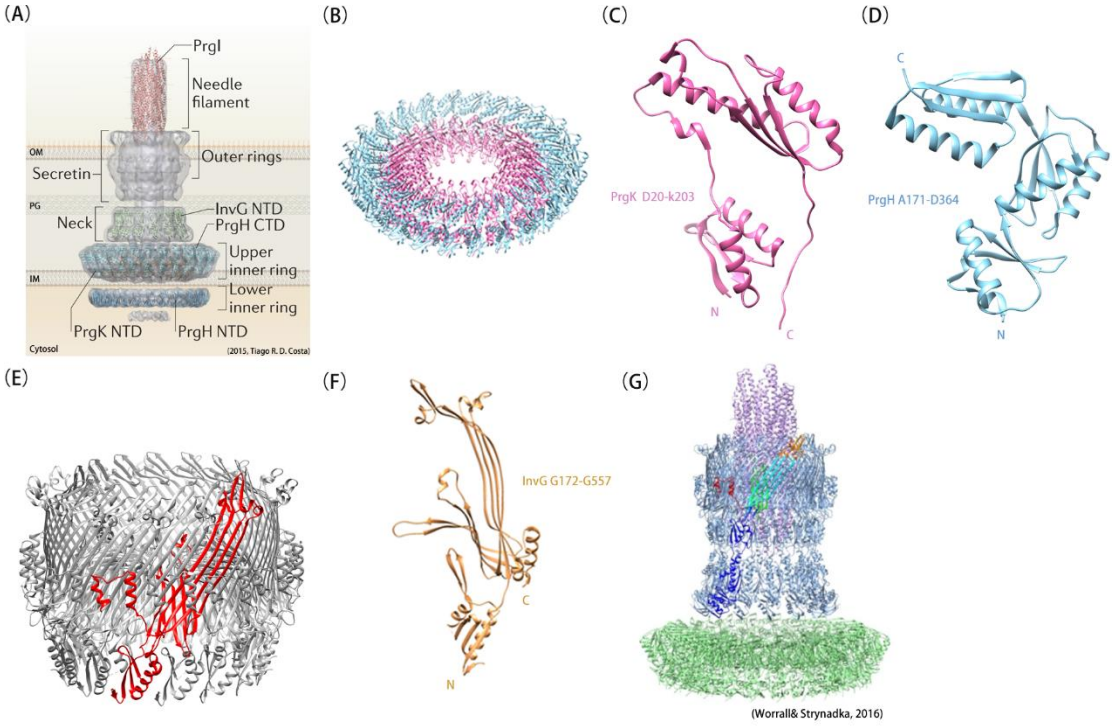


Figure 4. Structural model of T3SS. (A) Model of the major subassemblies of the T3SS. Adapted from Tiago R.D Costa et al., 2015. **(B)** Cryo-EM model of the two 24-mer rings made by the periplasmic domains of PrgK (in pink) and PrgH (in cyan) from *S. typhimurium* (PDB code: 5TC). **(C)** Ribbon representation of $PrgK_{D20\text{-}K203}$, showing the two RBM domains. **(D)** Ribbon representation of $PrgH_{A171\text{-}D364}$, showing the three periplasmic RBM domains. **(E)** Cryo-EM structure of the InvG OM ring from *S. typhimurium* (PDB code 5TCQ). One of the 15 protomers is colored in red. **(F)** Ribbon representation of a $InvG_{G171\text{-}G557}$ protomer showing the secretin and the S domain. **(G)** Cryo-EM model of $PrgH_{171\text{-}364}$ (in green), $PrgK_{20\text{-}203}$ (in green), $InvG_{34\text{-}557}$ (in blue and cyan), and $PrgI_{3\text{-}80}$ (in purple) components of T3SS from *S. typhimurium* (PDB code 6DUZ). One monomer encompassing $InvG_{34\text{-}557}$ is colored according to structural domains : N0-N3 domains (in blue), outer β-sheet (in cyan), inner β-sheet (in green), secretin domain lip (in orange) and S domain (in red) (Hu et al. 2018). The InvG secretin contacts the third RBM domain of PrgH through its first N-subdomain. Panel G is from Worrall& Strynadka et al., 2016.

4. Type-IV secretion systems

Widely present in both Gram-positive and Gram-negative bacteria, Type-IV secretion systems (T4SS) are versatile secretion systems that are related to the secretion of single protein, protein-protein, DNA and DNA-protein complexes into bacterial or eukaryotic target cells (Fronzes, Christie, and Waksman 2009, Cascales and Christie 2003). They have a unique ability among other secretion systems to translocate DNA and their main function is to mediate conjugation of plasmids, including some harboring antibiotic resistance genes (Costa et al. 2015).

In a canonical T4SS (based on the VirB/D system of *Agrobacterium tumefaciens*), 12 components (named as VirB1-11 and VirD4) assemble and power a channel that spans the inner membrane, the outer membrane and also the recipient cell membrane, mediating direct transfer of the substrate into the target cell cytoplasm (Sgro et al. 2018). T4SS are composed of two main assembly units: a **core-outer-membrane** (**Core-OM**) **complex** that spans both the inner and outer membranes, and an **inner-membrane** (**IM**) **complex** embedded in the inner membrane (Fig. 5A). These complexes assemble a **pilus** that will extend into the extracellular space. The pilus is made of a polymer of the major VirB2 pilin and the tip minor pilin VirB5 (Aly and Baron 2007).

Previous structural studies showed that the core-OM complex of canonical T4SS consist of 14 heterotrimers of **VirB7**, **VirB9** and **VirB10** proteins (Sgro et al. 2018). The core complex is located in the periplasm and is embedded into the inner membrane and outer membrane through the N- and C-termini of VirB10 (Sgro et al. 2018, Fronzes, Christie, and Waksman 2009, Fronzes et al. 2009). VirB10 forms the OM pore and is surrounded by VirB7 and VirB9 (Fig. 5B) (Sgro et al. 2018, Chandran et al. 2009). Interestingly, VirB7 lipoproteins from *Xanthomonales* contain an additional C-terminal globular domain, similar to N-subdomains found in T2SS secretins, and displaying the RBM fold (Sgro et al. 2019). This N_0/RBM1 domain wraps around the VirB9 and VirB10 layers (Fig. 5B).

The IM complex is composed of 12 copies of **VirB3**, 24 copies of **VirB6** and 12 copies of

VirB8, 14 copies of the **VirB10 N-terminus**, and 3 ATPases (**VirB4**, **VirB11** and **VirD4**). The ATPases are thought to provide energy for substrate unfolding and transfer through the T4SS channel. VirB4, which is the most conserved ATPase, localizes on the cytosolic side of the inner membrane. It forms two distinct barrel-like pedestals, each of them containing a VirB4 hexamer and interacting with other IM complex components (Fig. 5C) (Redzej et al. 2017).

The translocation mechanism through T4SS remains unclear. The VirB11 ATPase was suggested to act as a molecular switch between a pilus biogenesis mode and a secretion mode. In a first phase, binding of VirB11 to the VirB4 ATPase would promote pilus extension. Binding of the pilus tip pilin VirB5 to the target cell would lead to VirB11 release and its association with the VirD4 ATPase (Trokter et al. 2014). Then the DNA or protein substrate would first bind VirD4, which would act as a gate at the base of the IM complex. VirD4 would then transfer the substrate to VirB11, which would deliver the substrate the channel formed by the IM complex. The mechanism through which the substrate is transferred to the OM complex and is secreted remains unresolved. The pilus might serve as the conduit for substrate translocation or as a contacting device. In the latter case, it would just allow the two cells to be close enough to allow substrate secretion via a pilus-independent mechanism (Babic et al. 2008).

Bioinformatic analyses suggest that VirB4 from *Thermoanaerobacter pseudethanolicus* shares sequence similarity with the SpoIIIAA ATPase involved in *B. ubtilis* sporulation (12% sequence identity, 18% sequence similarity).

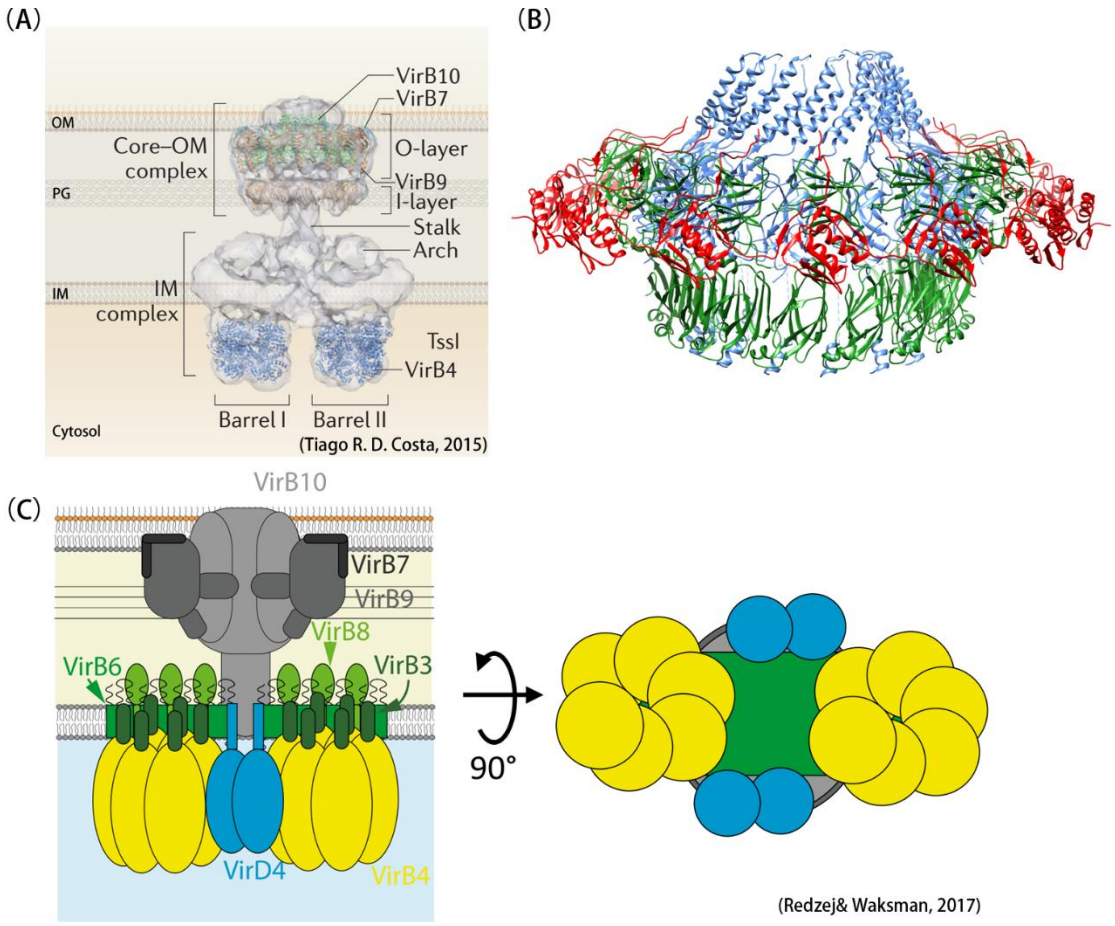


Figure 5. Structural model of T4SS. (A) Model of the major subassemblies of the T4S. Adapted from Tiago R.D Costa et al., 2015. **(B)** Cryo-EM structure of the OM core-complex from *Xanthomonas citri* T4SS (PDB: 6GYB). Fourteen protomers of the C-terminal N_0/RBM1 domain of VirB7 (in red) wrap around the 14-mer ring of VirB9 (in green), itself surrounding the 14-mer oligomer of VirB10 (in blue). **(C)** Schematic diagram of the T4SS structure in side (left) and bottom (right) views. The VirB4 (in yellow) and VirD4 (in blue) ATPases are respectively shown as hexamers and dimers connected to the VirB3 (in dark green), VirB6 (in green), and VirB8 (in light green) components of the inner membrane platform. The VirB7, VirB9 and VirB10 components of the core-OM complex are shown in different shades of grey. Adapted from Redzej& Waksman et al., 2017.

C. Ring-building motifs in specialized secretion systems

1. Presence of ring-building motifs in specialized secretion systems

Although specialized secretion systems are very different in their composition, architecture and assembly mechanism, some of their components share a common fold called the "ring-building motif" (RBM) (Spreter et al. 2009). This motif, as it is able to fold by itself (Bergeron et al. 2015), defines a new family of domains, which lacks detectable sequence identity but share similar arrangement of secondary structures : two helices stacked against a three-stranded β-sheet.

RBM domains are found in ring-forming proteins that located within IM or OM complexes of T2SS, T3SS and T4SS. (Fig. 6):

- the 1st and 2nd globular domains of PrgK/EscJ from the IM complex of T3SS. These domains will be called **PrgK RBM_1** and **RBM_2** in this manuscript (Fig. 6A) (Burkinshaw and Strynadka 2014, Schraidt and Marlovits 2011, Yip et al. 2005, Hu et al. 2018, Bergeron et al. 2015, Bergeron et al. 2013, Hu et al. 2019).
- the 2nd to 4th globular domains of PrgH from the IM complex of T3SS. These domains will be called **PrgH RBM_1** to **RBM_3** in this manuscript (Fig. 6A) (Spreter et al. 2009, Bergeron et al. 2013, Bergeron et al. 2015, Burkinshaw and Strynadka 2014, Hu et al. 2018, Hu et al. 2019).
- the 1st to 3rd N-subdomains of the InvG/EscC secretin in T3SS. These domains are usually called **N_1**, **N_2** and **N_3** in InvG; they will be called **InvG RBM_1** to **RBM_3** in this manuscript. These domains are usually called **N_1** and **N_2** in EscC; they will be called **EscC RBM_1** and **RBM_2** in this manuscript (Fig. 6B) (Spreter et al. 2009, Schraidt and Marlovits 2011, Bergeron et al. 2013, Worrall et al. 2016, Hu et al. 2018, Hu et al. 2019).
- the 1st to 4th N-subdomains of the GspD/PulD secretin in T2SS. These domains are usually called **N_0**, **N_1**, **N_2**, **N_3**; they will be called **GspD** or **PulD RBM_1** to **RBM_4** in this manuscript (Fig. 6C) (Yan et al. 2017, Hay et al. 2018, Chernyatina and Low

2019).

- an additional C-terminal domain found in Xanthomonales orthologues of the VirB7 component of the core-OM complex from T4SS. This domain is usually called **N_0**; it will be called **VirB7 RBM_1** in this manuscript (Fig. 6D) (Sgro et al. 2018).

The presence of RBM domains in various systems allowing transport through the inner and outer bacterial membranes could reflect a yet-unexplored evolutionary relationship between them (Souza et al., 2011).

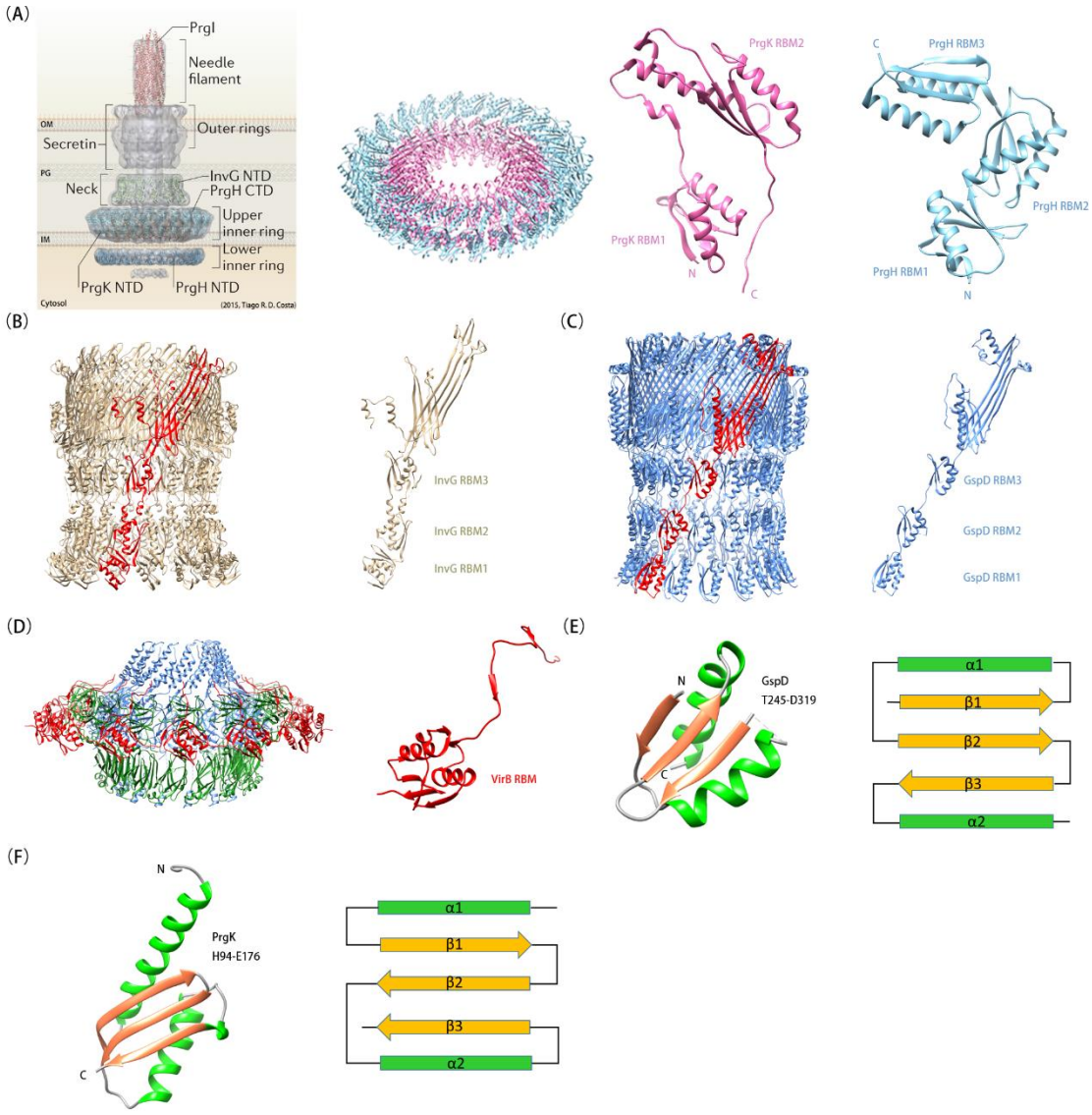


Figure 6. RBM domains in different secretion systems. (A) The left panel shows the cryo-EM structure of the PrgK-PrgH periplasmic rings from *S. typhimurium* T3SS (Prgk is in pink, PrgH is in light blue, PDB code: 5TCP). The right panels show the ribbon representation of $PrgK_{D20\text{-}K203}$ and $PrgH_{A171\text{-}D364}$, with the RBM domains labeled. **(B)** Cryo-EM structure of the InvG OM ring from *S. typhimurium* T3SS with one monomer shown in red (PDB code: 6DV3). The right panel shows the ribbon representation of $InvG_{G34\text{-}G557}$, with RBM1 to RBM3 domains labeled. **(C)** The left panel shows the cryo-EM structure of the GspD secretin channel from *V. cholerae* T2SS with one monomer shown in red (PDB code: 5WQ8). The right panel shows the ribbon representation of $GspD_{G97\text{-}M646}$, with RBM2 to RBM4 domains labeled. **(D)** The left panel shows the cryo-EM structure of the OM core-complex from *X. citri* T4SS (PDB code: 6GYB, VirB7 is in red, VirB9 is in green and VirB10 is in blue). The right panel shows the ribbon representation of the VirB7 C-terminal RBM domain. **(E)** Structure and topology model of the RBM3 domain of GspD from *V. cholerae* T2SS. α-helices are in green, β-sheets are in orange. **(F)** Structure and topology model of the RBM2 domain of PrgK from *S. typhimurium* T3SS.

Despite low primary sequence conservation (lower than 30%), RBM domains share an overall similar architecture made of two α-helices packing against a three-stranded antiparallel β-sheet. They are divided into two groups based on secondary structure connectivity: RBMs from the T3SS PrgK/PrgH group display an **α-β-β-α-β** arrangement (Fig. 6F) while RBMs from the N-subdomain secretin group display a **β-α-β-β-α** arrangement (Fig. 6E). Structural RBM models from the two groups have been obtained from X-ray crystallography or electron microscopy studies; the corresponding PDB entries, main characteristics and references are listed in Table 9 (see discussion). For the PrgK/PrgH group, they include EscJ RBM_{1-2} from *E. coli* T3SS, PrgK RBM_{1-2} and PrgH RBM_{1-3} from *S. typhimurium* T3SS (Yip et al. 2005, Spreter et al. 2009, Schraidt and Marlovits 2011, Bergeron et al. 2013, Bergeron et al. 2015, Worrall et al. 2016, Hu et al. 2018, Hu et al. 2019). Within the N-subdomain secretin group of T3SS, they include EscC RBM_{1-2} (N_{1-2}) from *E. coli* T3SS and InvG RBM_{1-3} (N_{1-3}) from *S. typhimurium* T3SS (Spreter et al. 2009, Schraidt and Marlovits 2011, Bergeron et al. 2013, Worrall et al. 2016, Hu et al. 2018, Hu et al. 2019). Within the N-subdomain secretin group of T2SS, they include GspD RBM_{2-4} (N_{1-3}) from *E. coli* T2SS, GspD RBM_{2-4} (N_{1-3}) from *V. cholerae*, PulD RBM_{1-4} (N_{0-3}) from *K. pneumoniae* T2SS, and XcpQ RBM_{3-4} (N_{2-3}) from *Pseudomonas aeruginosa* T2SS (Yan et al. 2017, Hay et al. 2018, Chernyatina and Low 2019). Within the N-subdomain secretin group of T4SS, they include VirB7 RBM_1 (N_0) from *X. citri* T4SS (Sgro et al. 2018).

2. ligomerization of ring-building motifs in specialized secretion systems

On the basis of molecular modelling, a broadly conserved ring-packing arrangement had been predicted for these α/β RBM domains (Yip et al. 2005), and the term RBM for "ring-building motif" had been proposed by the group of N. Strynadka (Spreter et al. 2009). The hypothesis that the wedge-shaped fold of RBM domains triggers ring-like oligomerization was later supported by studies in which PrgK and PrgH were located in purified T3SS from *S. typhimurium* (using immunogold labeling combined with single-particle EM). In these studies, **24-mer oligomeric ring** models of the two proteins (based on the crystal structure of EscJ for PrgK, and on the crystal structure of PrgH) were docked into corresponding cryo-EM maps (Spreter et al. 2009, Schraidt and Marlovits 2011, Schraidt et al. 2010).

Similar studies were performed with the N-subdomains of EscC from *E. coli* T3SS or InvG from *S. typhimurium* T3SS (Spreter et al. 2009, Schraidt et al. 2010, Schraidt and Marlovits 2011). These studies eventually revealed the **24:24:15 stoichiometry** of the **PrgK/PrgH/InvG complex** in T3SS (Schraidt and Marlovits 2011). The PrgK/PrgH/InvG ring model was later refined owing to new crystal or NMR structures of PrgK, PrgH and InvG, as well as higher-resolution cryo-EM maps of T3SS (Bergeron et al. 2013, Bergeron et al. 2015).

More recently, the rapid progress of cryo-EM remarkably pushed our knowledge of the T3SS architecture forward, providing models of the PrgK/PrgH/InvG complex at increasingly higher resolution, the best one so far reaching about 3.5 Å (Worrall et al. 2016, Hu et al. 2018). Additional elements of the T3SS get progressively incremented in the high-resolution cryo-EM models, such as the SpaP, SpaQ and SpaR export components or the PrgI/PrgJ needle complex (Hu et al. 2019, Guo et al. 2019).

First non-oligomeric structures of N-subdomains in secretins of T2SS (from *E. coli* ETEC GspD) were studied by Korotkov and co. using X-ray crystallography (Korotkov et al. 2009, Korotkov et al. 2011, Korotkov and Hol 2013). Later, the oligomeric status of these domains were revealed by cryo-EM studies performed on various orthologues of the GspD/PulD family, providing models of N-subdomain rings connected to rings of the secretin domain (Yan et al.

2017, Chernyatina and Low 2019, Burkinshaw and Strynadka 2014, Hay, Belousoff, and Lithgow 2017).

The variations observed in the stoichiometry (24 for PrgK and PrgH from T3SS, 15 for InvG from T3SS, 15 for GspD from T2SS, 14 for VirB7 from T4SS), dimensions and surface properties of those rings arise from variations in the primary sequence of the ring-forming proteins and likely allow this family of proteins to adapt to the specific architecture and assembly characteristics of the different secretion systems.

These high-resolution data now allow us to perform a more reliable analysis of the oligomerization interface of RBM domains. This analysis constituted part of my Ph.D. work and will be presented in the *Discussion* section.

Intriguingly, components of a putative secretion complex involved in bacterial sporulation, called the **SpoIIIA-SpoIIQ complex** (or **A-Q complex**) were shown to display weak sequence identities but obvious structural similarities with various components of specialized secretion systems (see chapter II). In particular, four of the A-Q proteins from *Bacillus subtilis* were shown to contain RBM domains and one of them (SpoIIIAG) was shown to form oligomeric rings resembling PrgK/PrgH RBM rings, suggesting that the A-Q complex might be a transport machinery (Rodrigues, Henry, et al. 2016, Zeytuni et al. 2017, Zeytuni et al. 2018a, Morlot and Rodrigues 2018). However, significant differences in the architecture and dimensions of the SpoIIIAG ring compared to RBM rings found in specialized secretion systems indicate that the A-Q complex should constitute a new type of secretion apparatus.

Other RBM-containing proteins were also recently identified with no apparent functional connection to the A-Q complex. During my Ph.D., I studied the structure of one of these proteins (called YhcN), and performed a structural analysis and comparison between RBM domains found in specialized secretion systems and in sporulation proteins.

II. The bacterial sporulation

A. Generalities about bacterial sporulation

As one of the oldest lives on earth, bacteria can thrive in multifarious environments, some can even survive in extreme conditions by adopting special strategies. The differentiation of bacteria into resistant spores, a process known as sporulation, is one of the most important strategies allowing certain bacterial species to survive in adverse conditions. The first spore was discovered by Robert Koch and Ferdinand Cohn in 1850 but despite more than 140 years of intensive studies, many aspects of the sporulation, spore dormancy and germination processes remain mysterious (Nicholson et al. 2000). As the most tenacious form of cell type in nature, spores can survive and stay dormant over long periods in many conventional sterilization methods such as high temperature, dehydration, radiation, detergent and some chemical solvents. Some spores can even reach several million years in earth core and fossils (Cano and Borucki 1995).

The capacity of bacteria to sporulate is mainly found in two genera from the Firmicute phylum, *Bacillus* and *Clostridium*, and it has many implications in industry and medicine: *Bacillus* species are used to produce many industrial enzymes and are also important biocontrol agents in agriculture. When spore-forming bacteria are also human pathogens, their ability to differentiate into resistant spores is an important pathogenesis factor (Traag et al. 2010, Ciccarelli et al. 2006). Famous pathogen spore-formers include *Clostridium tetani,* which causes tetanus (a disease characterized by muscle spasms), *Clostridium botulinum,* whose botulism toxin causes muscle failure and gastroenterological symptoms), *Clostridium perfringens,* responsible for tissue necrosis, and *Clostridium difficile*, which causes a toxin-mediated intestinal recurrent disease known as CDI (*C. difficile* infection). Finally, *Bacillus anthracis* is the well-known bioagent of anthrax, which can occur in skin, respiratory, intestinal or inflammatory forms. This bacterium is one of the most likely microorganisms to be used in a bioterrorist attack (source : www.cdc.gov).

In harsh conditions, some bacteria such as those cited above can stop their vegetative growth and enter a sporulation cycle. During this differentiation process, the developing spore and the

mother cell will undergo a series of transcriptional, morphological and biochemical changes that lead to the formation of a spore enveloped by protective layers called the cortex and the coat. When the spore is mature, the mother cell lyses and releases the endospore into the environment where it can remain dormant for a very long time. When the endospore encounters a suitable environment, it germinates and goes back to vegetative growth.

B. The sporulation process in *Bacillus subtilis*

The Gram-positive bacterium *Bacillus subtilis* has become the most studied spore-forming model because its natural competence, as well as the development of tools to manipulate its genetic content, have incredibly facilitated the creation of mutant strains to study cellular processes such as sporulation. Some differences exist in terms of gene conservation, activation mechanisms of the sporulation-specific transcription factors or composition of the spore envelope between the different spore-formers. However, most sporulation features are conserved and the knowledge obtained from studies performed on *B. subtilis* can be transposed to other species. Since all the sporulation proteins that I studied originate from *B. subtilis*, I will summarize in the next sections the knowledge regarding the sporulation cycle in this bacterium, and emphasize aspects that are directly related to my proteins of interest, the RBM-containing proteins.

Sporulation in *B. subtilis* is triggered by nutrient starvation and is easily inducible in laboratory conditions. At 37°C, it takes around 7-8 hours to form mature *B. subtilis* spores. The morphological differentiation process is orchestrated by four sporulation-specific transcription factors. It can be divided into seven main stages (**stages 0 to VI**, which are used to name many sporulation-specific genes), characterized by morphological and transcriptional landmark events (Fig. 7) that can be easily observed in the laboratory by observing samples every 30 minutes.

1. Morphological and biochemical sporulation landmarks

a. Asymmetric division and chromosome translocation

Upon starvation stress, the entry into sporulation is triggered by phosphorylation of the master transcription factor Spo0A (Tan and Ramamurthi 2014, Piggot and Hilbert 2004). Re-localization of the major division protein FtsZ (Erickson, Anderson, and Osawa 2010) at the quarter of the rod-shaped *B. subtilis* cell results in formation of a flat asymmetric septum that divides the cell into two unequally sized compartments. The big one is called the **mother cell** and the small one is called the **forespore** (Fig. 7). These two compartments are separated by two membranes that define an **intermembrane space**, which is about 25-nm thick and contains peptidoglycan, a polymer made of glycan chains cross-linked by peptide chains. Importantly for my thesis work, this double-membrane landscape is reminiscent of the cell envelope in Gram-negative bacteria.

The polar septum traps about 30% of one chromosome copy into the forespore. The remaining 70% of the chromosome is then transported from the mother cell cytoplasm to the forespore one by **SpoIIIE**, a membrane-anchored ATPase that assembles a translocation complex at the center of the septum (Besprozvannaya and Burton 2014). Chromosome translocation increases the turgor pressure into the forespore. At the same time, the asymmetric septum is thinned (to about 14 nm) by the peptidoglycan hydrolase complex **SpoIID-SpoIIM-SpoIIP** (Morlot et al. 2010, Khanna et al. 2019). As a consequence of those events, the forespore starts inflating into the mother cell (Lopez-Garrido et al. 2018).

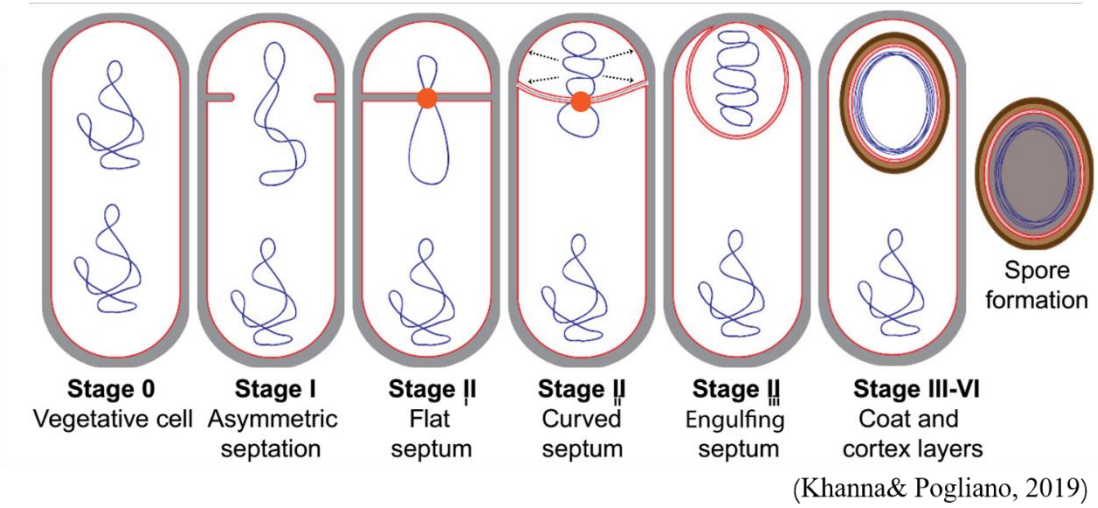


Figure 7. Schematic diagram of morphological changes and chromosome translocation into the forespore during the sporulation. Peptidoglycan is in grey, membranes are in red, chromosomes are in blue, SpoIIIE is in orange. The panel is from Khanna et al., 2019.

b. Engulfment

Shortly after asymmetric division and the beginning of chromosome translocation, the mother cell membrane starts migrating around the forespore in a phagocytic-like process called **engulfment**. During engulfment, the septal peptidoglycan is at least partially hydrolyzed and new peptidoglycan is synthesized at the leading edge of the engulfing membrane.

Peptidoglycan hydrolysis is carried out by the SpoIID-SpoIIM-SpoIIP complex, which includes two peptidoglycan hydrolases that perform processive degradation of the polymer (Fig. 8) (Morlot et al. 2010). Whether the SpoIID-SpoIIM-SpoIIP complex also degrades the new peptidoglycan synthesized during engulfment remains a matter of debate, as it is still not clear whether this new PG is synthesized ahead or behind the SpoIID-SpoIIM-SpoIIP machinery (Khanna et al. 2019). The activity of the SpoIID-SpoIIM-SpoIIP complex relies on the activation of the SpoIIP endopeptidase/amidase by SpoIID, and the conditional activity of SpoIID on naked glycan strands (devoid of stem peptides). Based on this enzymatic coordination, localization around the forespore and requirement for engulfment, the SpoIID-SpoIIM-SpoIIP complex was proposed to function as a motor pulling the mother cell membrane around the forespore (Fig. 8B, C).

At the end of engulfment, fission of the mother cell membranes, which involves the FisB protein but is not totally understood, releases the forespore into the mother cell cytoplasm (Doan et al. 2013). The forespore is thus eventually surrounded by two membranes: the **inner forespore membrane** and the **outer forespore membrane** (Fig. 7, 8C**)**.

Proper engulfment also requires the assembly of a large multi-protein complex into the inner and outer forespore membranes. This transenvelope complex called the **SpoIIIA-SpoIIQ complex** will be described in a separate section (section IIC) because it is one of the objects of my thesis.

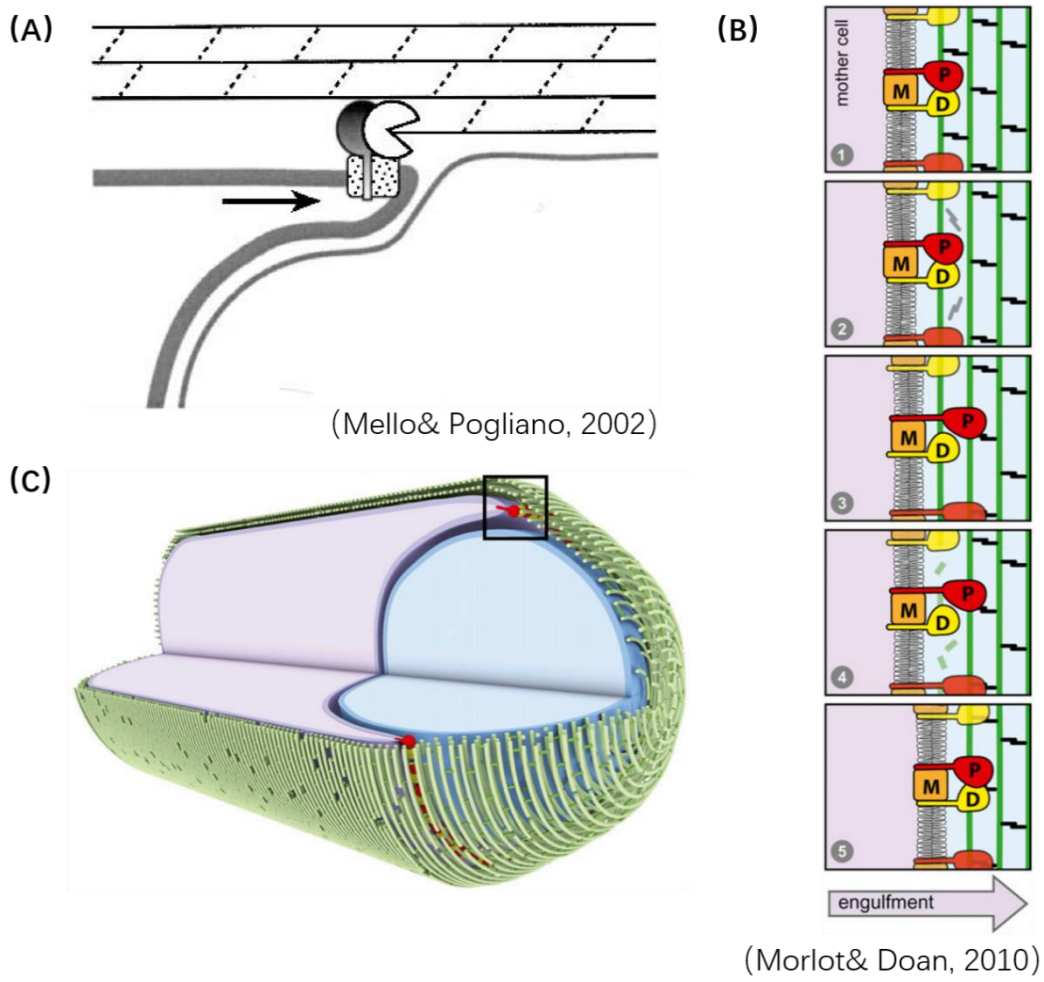


Figure 8. The SpoIID-SpoIIM-SpoIIP machinery. (A) Migration of the mother cell membrane (thick grey line) around the forespore membrane (thin grey line) is driven by peptidoglycan hydrolysis performed by the SpoIID (pacman) and SpoIIP (shaded lollipop) pair of hydrolases. SpoIIM (speckled box) anchors the peptidoglycan hydrolases at the leading edge of the engulfing mother cell membrane. The panel is from Mello et al., 2002. **(B)** Schematic diagram of the proposed catalytic cycle of the engulfment complex. (1) The SpoIID (D)-SpoIIM (M)-SpoIIP (P) complex binds the peptidoglycan. Glycan strands are in green, peptides are in black. (2) D stimulates the amidase activity of P, resulting in cleavage of the stem peptides cross-links and the release of P. (3) Released P rebinds at a nearby peptidoglycan site. (4) The denuded glycan strands are cleaved by D. (5) Released D rebinds to P at the nearby site. **(C)** Circumferentially distributed D-M-P engulfment complexes drive movement of the mother cell membrane (light purple) around the forespore membrane (blue). Panels (B) and (C) are from Morlot et al., 2010.

c. Assembly of the cortex and the coat

During and after engulfment, several protective layers are added around the forespore and will confer its resistance properties: the **cortex** and the **various coat layers** (Fig. 9) (McKenney, Driks, and Eichenberger 2013, Popham and Bernhards 2015).

The cortex is a modified peptidoglycan that is synthesized on top of the germ cell wall (unmodified peptidoglycan) in the intermembrane space. Genes involved in synthesis of the germ cell wall include the PG synthases genes *pbpF* and *pbpG* while genes involved in cortex synthesis include the PG synthases genes *spoVB*, *spoVD* and *spoVE*, as well as the PG hydrolase gene *lytH*, *dacA*, *dacB*, *dacF* and *cwlD* and the deacetylase gene *pdaA*.

Like vegetative peptidoglycan, the cortex contains sugar strands made of alternating *N*-acetylglucosamine (NAG) and *N*-acetylmuramic acid (NAM) attached to peptide chains. However in the cortex, about 50% of the peptide chains attached to NAM are removed, resulting in the formation of muramic-δ-lactam (McKenney, Driks, and Eichenberger 2013). In addition, 15 to 25% of the remaining peptide chains are shortened to single L-Ala residues. These peptide modifications result in a greatly reduced level of cross-linking of the spore peptidoglycan (~3% in cortex against ~40% in vegetative peptidoglycan).

The cortex maintains the dehydration state of the mature spore, which is essential for resistance to high temperatures and chemicals. The cortex feature that impacts most the spore resistance is the amount of cortex rather than its structural properties. Surprisingly enough, neither the presence of muramic-δ-lactam nor modified peptides affects the resistance of the cortex to vegetative peptidoglycan hydrolases. The spore is actually protected from these lytic enzymes by the several proteins layers of the coat. Instead, the muramic-δ-lactam, which is absent from the germ cell wall, is specifically recognized by germination hydrolases. This particularity would thus allow specific degradation of the cortex and protection of the germ cell wall during germination.

In *B. subtilis*, the spore coat is composed of three layers: a lamellar inner coat, a more coarsely layered outer coat and a crust layer (Fig. 9A). In recent years, coat proteins were identified through reverse genetics and immunogold-electron microscopy. Proteins playing a main role in coat morphogenesis include: SpoIVA, SpoVID and SpoVM, which play a role in anchoring the coat to the spore surface; SafA, which is necessary for inner coat assembly; and CotE, which is required for outer coat assembly (McKenney, Driks, and Eichenberger 2013). Another group of proteins including CotX, CotY and CotZ is essential to the assembly of the

crust layer (Fig. 9B) (McKenney, Driks, and Eichenberger 2013).

The spore coat is only permeable to molecules of 2 to 8 kDa and protects the spore from enzymatic aggressions, including peptidoglycan hydrolases.

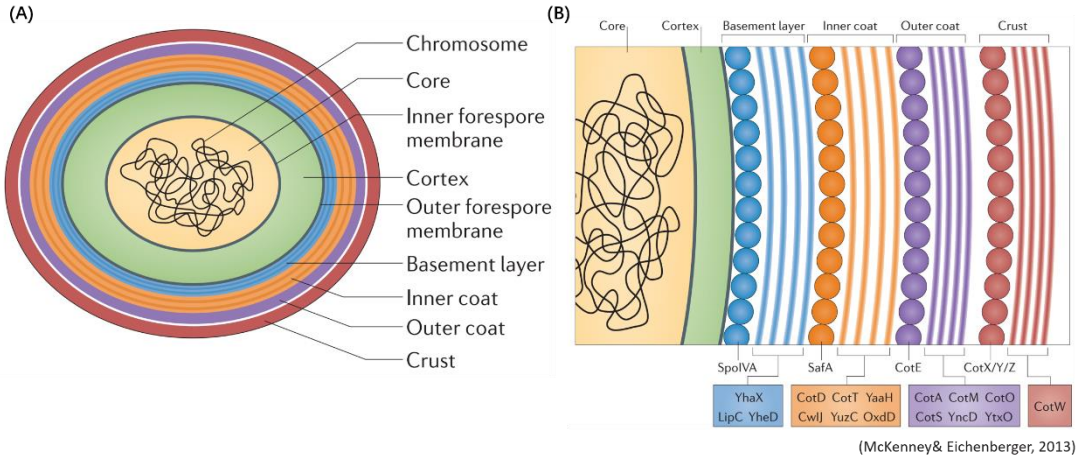


Figure 9. Protective layers found in the mature *B. subtilis* spore. (A) Schematic diagram of the mature *B. subtilis* spore**.** The chromosome is located in the dehydrated central core. The core is protected by multiple layers including the cortex (in green), the basement layer (in blue), the inner coat (in orange), the outer coat (in purple) and the crust (in red). **(B)** Composition of the different protein layers of the spore coat. The assembly of each layer may be driven by multimerization of the coat proteins. The interaction between each layer is still to be demonstrated. This figure is from McKenney et al., 2013.

d. Preparation for dormancy and mother cell lysis

To prepare for dormancy, the spore produces small DNA-binding proteins that compact the chromosome and protect it from irradiation and genotoxic stress.

In parallel, the mother cell produces the SpoVFA and SpoVFB enzymes that convert dihydroxydipicolinic acid into **dipicolinic acid** (**DPA**) (Daniel and Errington 1993). DPA is transported across the outer forespore membrane by the SpoVV transporter and then imported into the forespore by the SpoVA proteins (SpoVAC, SpoVAD, SpoVAEb and SpoVAF) (Ramirez-Guadiana et al. 2017). Accumulation of DPA in complex with Ca^{2+} (CaDPA) into the forespore leads to dehydration of the spore core, providing heat resistance to the mature spore.

At the end of the sporulation process, the mother cell needs to lyse to release the mature spore into the environment. This lysis involves the amidases CwlC, LytC and YqeE, which cleave the bonds between the sugar and peptide chains of the mother-cell peptidoglycan (Smith, Blackman, and Foster 2000).

e. **Germination**

Germination of the spore consists in uptake of water and core expansion, leading to the loss of spore resistance properties. It is essentially a biophysical process that remarkably happens without the need for *de novo* synthesis of biological macromolecules. Many aspects of the germination process remain unknown and I have summarized below the key steps in this process (Setlow 2014).

Germinants such as single amino acids (L-Ala, L-Val, L-Asn), sugars or purine nucleosides, penetrate the spore coat, the outer forespore membrane and cortex to bind protein complexes located in the inner forespore membrane. Upon binding of germinants to germinant receptors (usually called GRs, including GerAA, GerAB and GerAC, and requiring GerD for assembly), change in the permeability of the inner membrane allows monovalents cations to move out of the spore. CaDPA is then released through channels formed by the SpoVA proteins (Setlow 2014).

The cortex is next cleaved by the lytic transglycosylases CwlJ and SleB, which specifically recognize the muramic-δ-lactam component, leaving the germ cell wall intact. The mother cell protein CwlJ is known to require the coat proteins GerQ, CotE and SafA to accumulate in the spore coat. It also requires the Ca^{2+}-DPA complex and SwsB for its hydrolytic function. SleB is a forespore protein localized near the inner membrane. Its activation mechanism involves the YpeB protein but remains unknown.

Degradation of the cortex leads to expansion of the germ cell wall and the entry of potassium ions and water. Rehydration of the spore core eventually allows metabolism resumption, breakage of the coat and outgrowth of the germinated spore.

2. The sporulation-specific transcription factors

The different stages of the forespore development are governed by sequential and compartmentalized expression of hundreds of genes in the mother cell or in the forespore (Fig. 10). The chain of the different gene expression programs is orchestrated by the sequential activation of sporulation-specific transcription factors (sigma factors, Sig or σ) in the mother cell and in the forespore. Cell-cell signaling pathways ensure that gene expression in one cell is coordinated with gene expression in the other (Errington 2003).

The different morphological landmark events of sporulation are mainly regulated by 4 sigma factors, σ^F, σ^E, σ^G and σ^K. Activation of σ^F, σ^E and σ^K is reviewed in (Errington 2003, Tan and Ramamurthi 2014) and will not be described here because they are not directly linked to my thesis work. I will describe more the activation of σ^G, as it involved the SpoIIIA-SpoIIQ complex, which I have studied during my PhD.

Activation of σG in the forespore

During the engulfment, the third sigma factors, σ^G is activated in the forespore. The activation of this spore specific sigma factor depends on both σ^F and σ^E.

At the beginning of engulfment, the first sigma factor, σ^F is activated in forespore, and it is then required for the activation of sE in the mother cell. σ^E E is responsible for the activation of sG in the forespore. σ^F can direct control the transcription of the gene encoding σ^G. However, the synthesized is keep inactive because of inhibitor protein Gin/CsfB (also under σ^F control) until the completion of engulfment(Karmazyn-Campelli et al. 2008, Mearls et al. 2018, Doan et al. 2009). From then on, σ^F is deactivated and “replaced” by the active σ^G. Once σ^G is activated, it would recognizes its own promoter and further promote its activation (Sun, Cabrera-Martinez, and Setlow 1991). After activated, the maintenance of subsequent activity of σ^G requires a SpoIIIA-SpoIIQ putative secretion protein complex (A-Q complex) that localized around the forespore membrane. Lacking any of the components from the A-Q complex would result in the collapse of the forespore after engulfment (Doan et al. 2009). The role that A-Q complex plays in B. subtilis sporulation is still not clear and will be described

below.

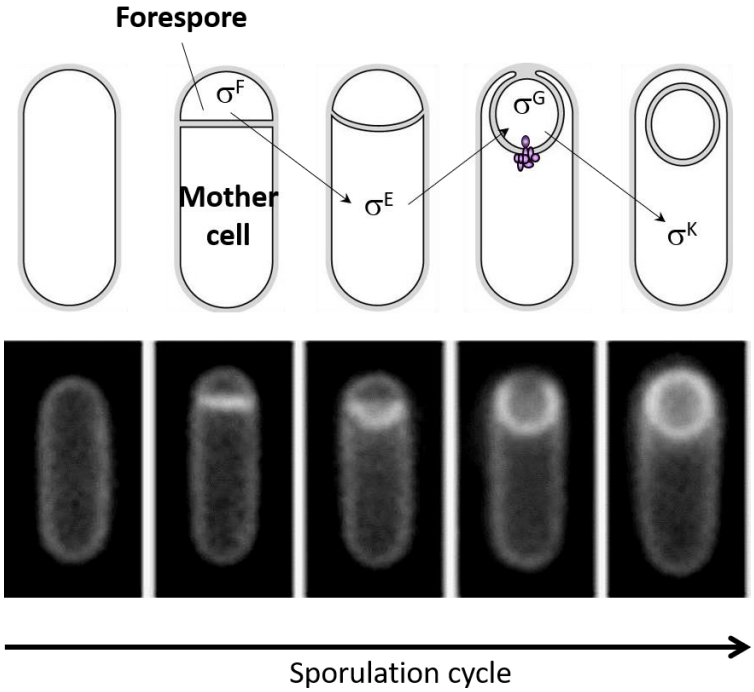


Figure 10. The sequential activation of sporulation-specific transcription factors. The different morphological stages observed during sporulation, as well as the activation sequence of the σ transcription factors, are represented with diagrams (top panels). For the engulfment stage, components of the SpoIIIA-SpoIIQ complex are represented with purple spheres. For each developmental stage, a fluorescence image of a sporulating cell in which the membranes are stained with the FM4-64 dye is shown.

C. The SpoIIIA-SpoIIQ complex in *Bacillus subtilis*

During engulfment, the mother cell and forespore assemble a multimeric complex that spans the double membrane between them. Discovered in the 70's during the screening of genetic mutants defective for sporulation, this transenvelope complex (called the **SpoIIIA-SpoIIQ complex** or **A-Q complex**) includes eight mother cell proteins (SpoIIIAA to SpoIIIAH) encoded in the *spoIIIA* operon, the mother-cell lipoprotein GerM, and the forespore protein SpoIIQ (Fig. 11A).

The A-Q complex is essential to the morphogenesis and maturation of the forespore. In the absence of any member of the complex, σG activation is impaired, forespores are smaller, they tend to collapse and display membrane deformities (Fig. 11B, 12B, 12C and 17B).

Between 2008 and 2009, bioinformatic analyses of the A-Q proteins identified weak homologies with components found in specialized secretion systems (Meisner et al. 2008, Camp and Losick 2008, Doan et al. 2009). These predictions were later confirmed when the structures of several A-Q proteins were solved and revealed that some of them indeed possess an RBM-like domain (see the next sections), aligning well with those involved in the oligomerization of secretion system components. Based on these observations, the A-Q complex was suggested to be a new type of transporter, allowing the secretion or passage of an unknown molecule from the mother cell to the forespore, required for σG activity.

In this chapter, I will summarize the structural and functional knowledge concerning the A-Q proteins, highlighting their similarities with ring-forming proteins from specialized secretion systems.

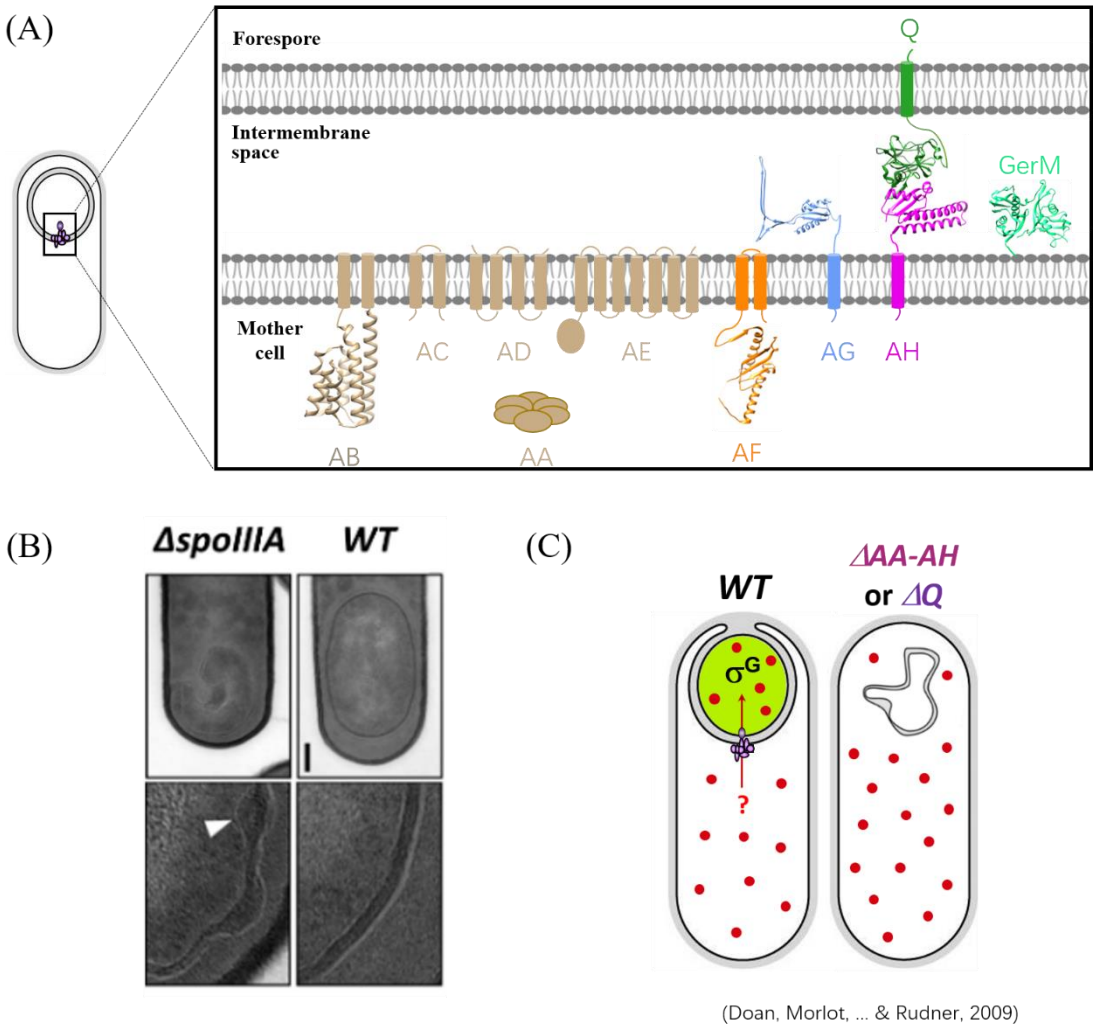


Figure 11. Schematic diagrams of the components of the SpoIIIA-SpoIIQ complex in *B. subtilis*. **(A)** Localization, topology and structure of mother cell and forespore proteins from the A-Q complex. SpoIIIAA (AA, the ATPase, represented as an hexamer), SpoIIIAB (AB), SpoIIIAC (AC), SpoIIIAD (AD), SpoIIIAE (AE), SpoIIIAF (AF), SpoIIIAG (AG), SpoIIIAH (AH), GerM and SpoIIQ (Q) are shown with an arbitrary stoichiometry. **(B)** Spore morphology in wild-type (*WT*) cells and in the absence of the SpoIIIA proteins (*ΔspoIIIA*) observed by EM. Scale bar, 200 nm. The caret highlights membrane defects. **(C)** Model describing the putative role of the A-Q complex (violet circles and ellipses) in *B. subtilis sporulation*. Passive transport or active secretion of an unknown metabolite/osmolyte (red circle) into the forespore would maintain forespore integrity and σG activity (green forespore) in wild-type cells (WT). In the absence of the A-Q proteins (ΔAA-AH or ΔQ), the forespore loses metabolic potential, collapses and σG activity cannot be maintained. Adapted from Doan et al., 2009.

1. The SpoIIIAA ATPase

SpoIIIAA (**AA**) is a soluble protein containing 307 amino acids, and sharing about 18% sequence identity with ATPases found in T2SS (GspE) and T4SS (VirB11) (Doan et al. 2009, Zeytuni and Strynadka 2019). The conserved motifs consist of the Walker A and B boxes, Aspartate box and Histidine box, that are responsible for nucleotide and NTP binding or hydrolysis (Fig. 12A).

In a previous study, the Rudner laboratory built a series of *B. subtilis* mutant strains harboring amino-acid substitutions in these conserved motifs. They explored σG activity using a σG-dependent fluorescent reporter (P_{sspE}-*cfp*) and a σG-dependent *sspB-lacZ* fusion (Doan et al. 2009). The results indicated that AA, and in particular the Walker B box, plays an essential role in σG activity (Fig. 12B). Even though the ATPase activity of AA could not be reproduced *in vitro*, these data strongly support the idea that AA is an ATPase homologous to those found in T2SS and T4SS. By analogy with these proteins, AA may form a hexameric structure and be involved in substrate export or biogenesis of a pseudo-pilus (Yamagata and Tainer 2007, Mancl et al. 2016, Lu et al. 2013).

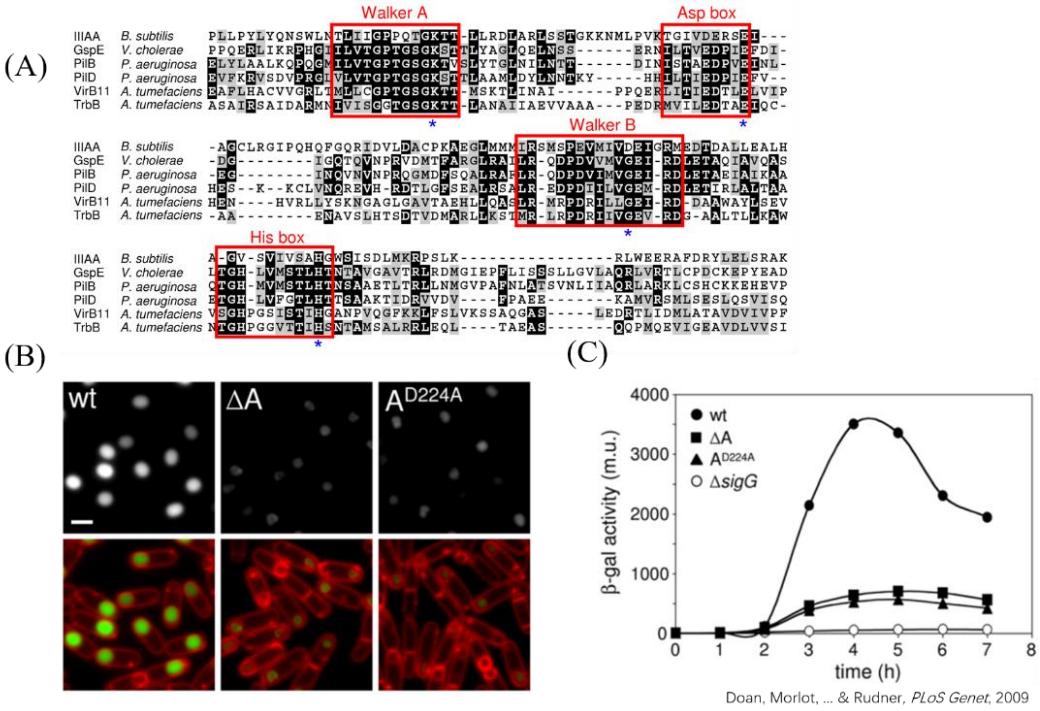


Figure. 12 SpoIIIAA is required for the activity of σG. (A) SpoIIIAA (IIIAA) shares sequence similarities with GspE from T2SS, PilB, PilD, VirB11 and TrbB from T4SS, in four conserved motifs (red boxes). **(B)** The activity of σG in the forespore is observed using a fluorescent reporter (P_{sspE}-*cfp*). Compared to the wild-type background (wt), the CFP fluorescence signal is strongly reduced in a strain lacking SpoIIIAA (ΔA) or expressing a SpoIIIAA^{D224A} Walker B box mutant (A^{D224A}). **(C)** Expression of a σG-dependent *sspB-lacZ* translational fusion was monitored by measuring the activity of the β-galactosidase. Compared to the wild-type background (wt), the β-galactosidase production is totally impaired in a strain lacking σG (*ΔsigG*), and strongly reduced in strains lacking SpoIIIAA (*ΔA*) or expressing a SpoIIIAA^{D224A} Walker B box mutant (A^{D224A}). The figure is from Doan et al., 2009.

2. SpoIIIAB : the putative ATPase anchor

SpoIIIAB (**AB**) is a bitopic membrane protein harboring two transmembrane segments and a soluble domain predicted to be localized on the mother cell side (Fig. 13A).

Previous studies showed that AB shares sequence similarities with GspF/PilC proteins found in T2SS and TadB/C proteins from T4SS. Homologous regions reach a maximum of 28% sequence identity (Fig. 13B) (Zeytuni et al. 2018b, Zeytuni and Strynadka 2019). The crystal structure of AB, published in 2018 by the Strynadka's group, showed that its soluble region ($AB_{27\text{-}153}$) superposes well onto EpsF from *Vibrio cholerae*, PilC from *Thermus thermophilus* and TcpE from *V. cholerae*, with overall Cα RMSDs (root mean square deviation) of 1.89, 1.97 and 1.77 Å (Fig. 13C). Bioinformatic analyses that $AB_{27\text{-}153}$ also shares structural similarity with the C subunit of the V-ATPase from *T. thermophilus*.

The fact that AB shares striking structural similarity with ATPase anchors located in the inner membrane platform of specialized secretion systems suggests that AB could also anchor the AA ATPase to the transmembrane A-Q complex.

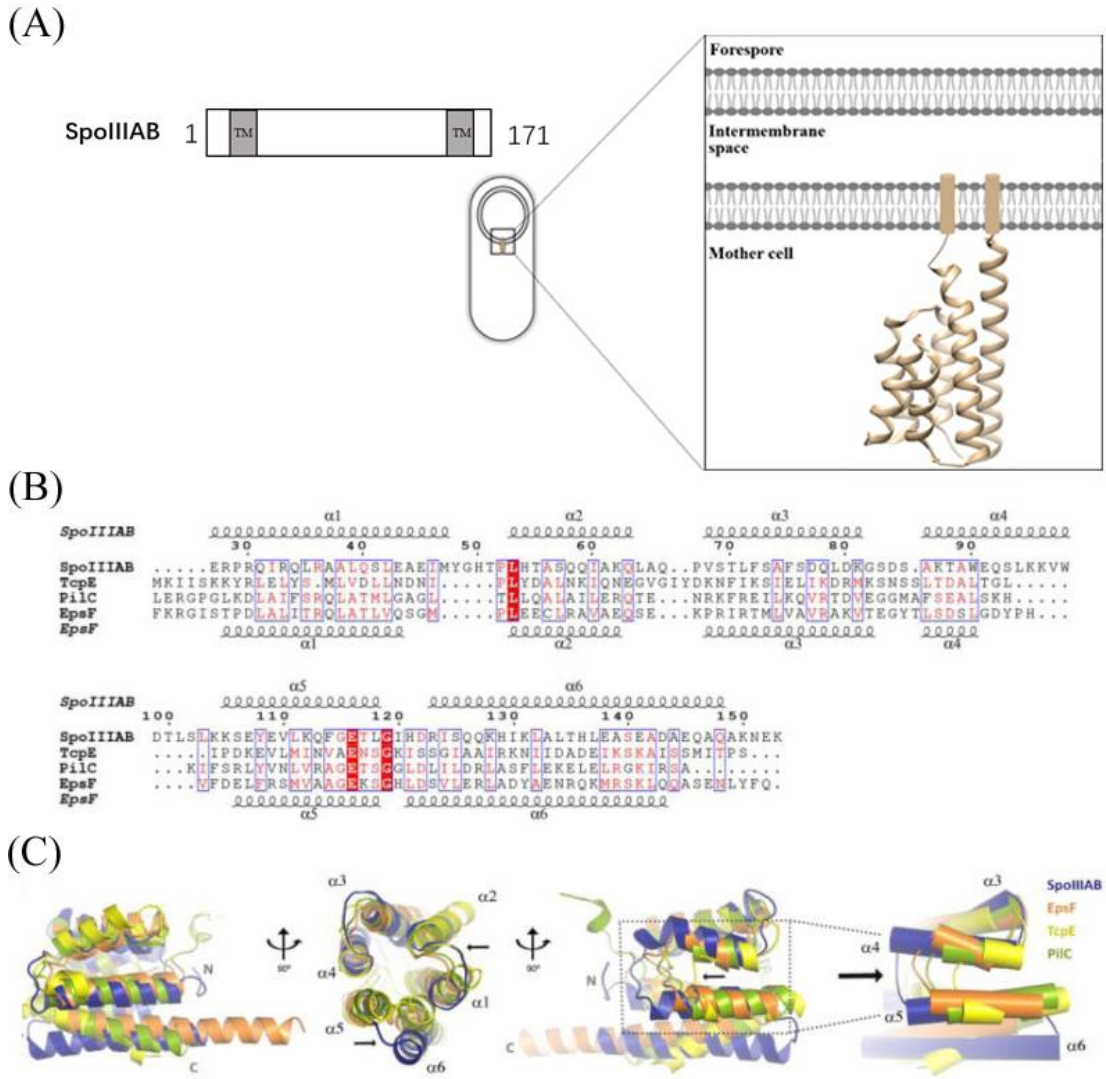


Figure. 13 SpoIIIAB shares structural similarity with T2SS and T4SS proteins. (A) Topology of SpoIIIAB. The structure of SpoIIIAB is shown in ribbon representation. **(B)** Amino acid sequence alignment of SpoIIIAB against homologues from specialized secretion systems. The secondary structures of SpoIIIAB and EpsF are shown on the top and bottom of the aligned sequences, respectively. Residue numbering is for SPoIIIAB. Conserved residues are boxed in red. Similar residues are boxed in blue. **(C)** Structural overlay of $AB_{27\text{-}153}$ (in blue, PDB code 6BS9) with $EpsF_{56\text{-}171}$ (in orange, PDB code 3C1Q), $PilC_{53\text{-}168}$ (in green, PDB code 4HHX) and $TcpE_{1\text{-}102}$ (in yellow, PDB code 2WHN). The two SpoIIIAB regions displaying structural divergence with its homologues (angle of helix α6, dimensions of helices α4 and α5) are pointed with black arrows. TM, transmembrane segment. Panel B-C are adapted from Zeytuni et al., 2018.

3. SpoIIIAC and SpoIIIAD: putative pilins or export components?

Because they are mainly made of transmembrane segments (Fig. 14A), **SpoIIIAC** (**AC**) and **SpoIIIAD** (**AD**) could be pilin-like proteins forming oligomeric structures in the intermembrane space. However, this hypothesis is so far not supported by any structural or functional data.

On the other hand, the topology of AC and AD is also similar to FliQ/SpaQ and FliP/SpaP proteins from flagella and T3SS, although the orientation of the N- and C-termini of FliQ is inverted compared to the predicted ones of AC (Fig. 14B) (Kuhlen et al. 2018).

A recent structural study of the T3SS export apparatus in *S. typhimurium* revealed that FliP, FliQ and another membrane protein (FliR) form a periplasmic export gate called the PQR complex. This complex is part of the export apparatus (made of FliP, FliQ, FliR, FlhA and FlhB), which is required for the translocation of substrates across the bacterial envelope (Fabiani et al. 2017, Wagner et al. 2010, Fukumura et al. 2017). The PQR complex was proposed to adopt two different conformations. A closed conformation in the absence of rod components would guarantee that the complex does not form holes in the inner membrane before the whole structure of the secretion system is assembled. This closed conformation is the one recently obtained by cryo-EM at 4.2 Å (Fig. 14C) (Kuhlen et al. 2018). An open conformation would allow the passage of the substrate(s) into the secretion channel. In support of this idea, an open conformation of the PQR complex was modeled from the closed one and could be fit in a low-resolution cryo-EM map of a fully assembled T3SS (Fig. 14C) (Kuhlen et al. 2018).

In view of the topological similarities between AC and AD and components of the PQR complex, it is tempting to propose that these proteins (as well as SpoIIIE, as discussed in the next paragraph) may form an export apparatus within the A-Q complex. They might play a role in gating the complex during its assembly. Once the A-Q complex would be mature, they would then allow substrate secretion through a putative conduit traversing the A-Q complex.

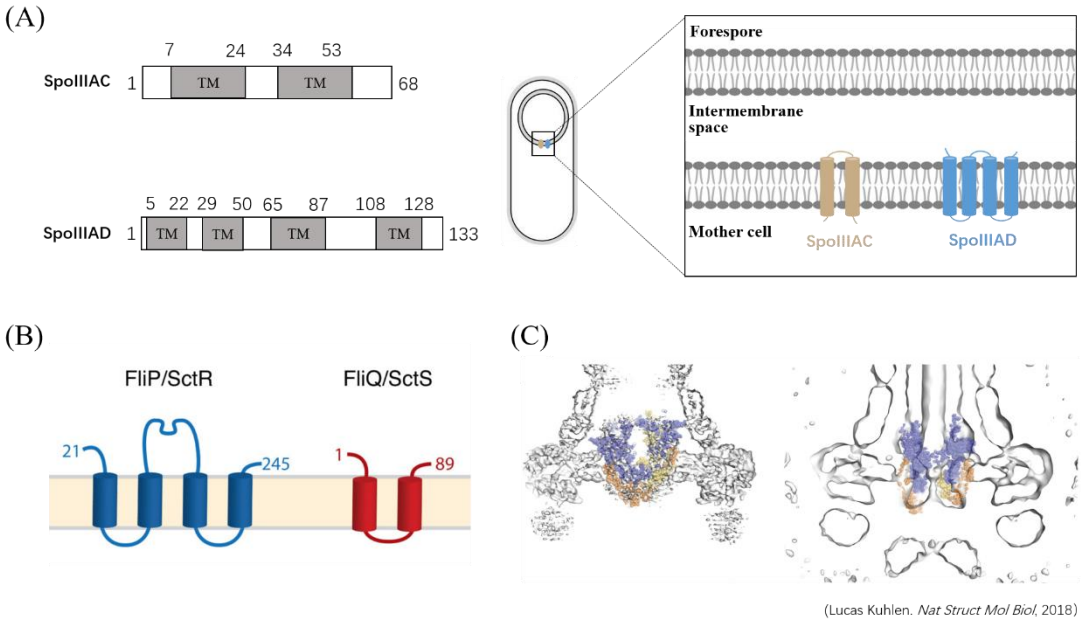


Figure 14. SpoIIIAC and SpoIIIAD share similar topology with Flip/SctR, FliQ/SctS proteins from the export apparatus of T3SS. (A) Topology of SpoIIIAC and SpoIIIAD. **(B)** Topology of Flip/SctR and FliQ/SctS from *S. typhimurium*. **(C)** Models of the closed and open conformations of the PQR complex. The PQR complex in its closed conformation (as obtained experimentally by single-particle cryo-EM) fits in the cellular cryo-EM map obtained from a basal body in the absence of rod components (left panel). The PQR complex in its open conformation (obtained by *in silico* modeling) fits in the cellular cryo-EM map obtained from a basal body in the presence of rod components (right panel). TM, transmembrane segment. Panel B-C are adapted from Kuhlen et al., 2018.

4. SpoIIIAE: permease or export component?

SpoIIIAE (**AE**) is an integral membrane protein with 7 predicted transmembrane helices that cover around 3/4 of the full-length sequence (Fig. 15A). Insertion of AE in the mother cell membrane requires its interaction with the membrane protein translocase SpoIIIJ (Serrano et al. 2008, Camp and Losick 2008).

Sequence analyses shows weak sequence identity (11%) with the permease domain of ATP-binding cassette transporters from T1SS (Zeytuni and Strynadka 2019). The latter domain is inserted in the cytoplasmic membrane and is important for the transmembrane transport carried out by T1SS (Green and Mecsas 2016).

On the other hand, the N-terminal domain of AE also shares low sequence similarity (~8% sequence similarity) with subdomain 3 of FlhA proteins from the export apparatus of T3SS. This subdomain contributes to the interaction of FlhA with the substrate(s) in complex with their chaperones, the conserved ATPase and its regulators (Worrall, Vuckovic, and Strynadka 2010, Minamino et al. 2011).

Finally as a third alternative, if AC and AD are homologues of FliQ and FliP, it would make sense that AE be the homologue of FliR (Fig. 15B).

Taken together, these observations suggest that AE could participate to a gating platform and/or export apparatus (made of AC, AD and AE) within the A-Q complex.

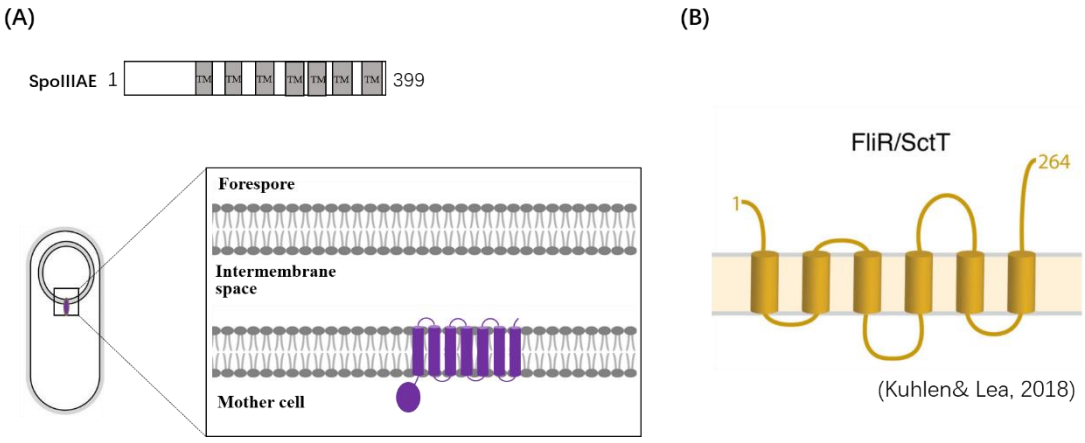


Figure 15. Topology of SpoIIIAE (A) and FliR/SctT (B) from flagellar/injectisome machineries. **(A)** 7 predicted TMs cover around 3/4 of the SpoIIIAE sequence**. (B)** The FliR/SctT component of the flagellar/injectisome export complex contains 6 TMs distributed all along the protein sequence. Panel B is adapted from Kuhlen et al., 2018.

5. SpoIIIAF: a putative ring-forming protein

SpoIIIAF (**AF**) is a bitopic membrane protein harboring two N-terminal transmembrane helices and a soluble domain predicted to be exposed in the mother cell cytoplasm or in the intermembrane space with equivalent probabilities (unpublished data) (Fig. 16A). The Strynadka's group solved the crystal structure of a soluble construct of AF (residues 85 to 206, 2.7 Å of resolution) in 2018 (Zeytuni et al., 2018). It revealed that AF has an RBM-like domain that aligns well (Cα backbone RMSD < 2 Å) with RBMs of the ring-forming proteins EscJ and PrgK from T3SS. This structural similarity will be further described in the *Discussion* part of this manuscript.

In the same study, the authors performed size-exclusion chromatography on two different recombinant constructs of AF (AF_{60-206} and AF_{85-206}) and found out that the longest one (AF_{60-206}) could form high molecular-weight species in solution (Fig. 16B). Transmission EM analysis of purified AF_{60-206} showed the presence of scarce circular structures (Fig. 16C) (Zeytuni et al. 2018a). However, no structure of AF rings has been published so far, likely due to the heterogeneity of the sample shown in Fig. 16C.

During my thesis, I have purified a soluble construct of AF encompassing residues 55 to 206, and another student purified the full-length membrane construct. None of them oligomerized in solution, as described in the *Results* part. In my opinion, the oligomerization of AF thus remains fully hypothetical. However, the presence of the RBM-like domain in AF supports the idea that the A-Q complex has structural similarities with specialized secretion systems.

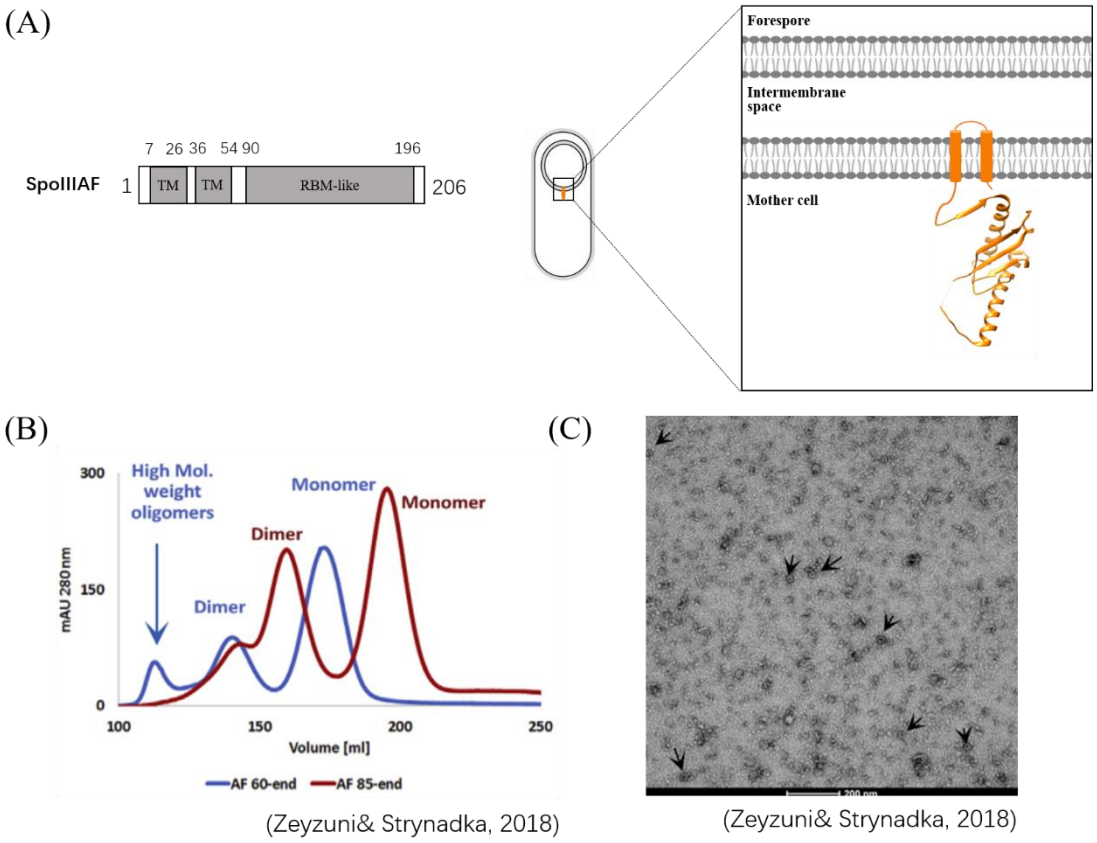


Figure 16. Topology and structure of SpoIIIAF. (A) Topology of SpoIIIAF. TM, transmembrane segment. The orientation of SpoIIIAF relative to the mother-cell membrane was chosen arbitrarily. The structure of AF_{85-206} (PBD code 6DCS) is shown in ribbon representation. **(B)** Size-exclusion chromatograms of AF_{60-206} (in blue) and AF_{85-206} (in red). **(C)** Micrograph of the purified AF_{60-206} construct observed by negative-stain TEM. Suspected ring structures are pointed with black arrows. Panels (B) and (C) are from Zeytuni et al., 2018.

6. SpoIIIAG: THE ring-forming component

SpoIIIAG (**AG**) is a bitopic membrane protein containing a short N-terminal cytosolic portion, a transmembrane segment and a soluble domain. The latter is made of a N-terminal part displaying low sequence conservation and no predicted secondary structure (residues 50 to 89, called the D1 domain), and a longer C-terminal part displaying high sequence conservation (residues 90 to 229, called the D2 domain) (Fig. 17A). The soluble domain is made of a N-terminal (residues 90 to 123) and C-terminal (residues 181 to 229) region that share about 15% sequence identity with canonical RBM domains from the EscJ/PrgK family of proteins. These RBM-homologous regions sandwich a **long insertion segment** (residues 124 to 180) that does not show similarity with any known protein structure but that is also found in the third RBM domain (residues 228 to 439) of the ring-forming protein FliF from flagellar T3SS (Bergeron 2016).

Fluorescence microscopy experiments performed with a GFP-AG fusion showed that AG is enriched around the forespore and delocalizes to the lateral mother-cell membrane in the absence of Q, GerM and AH (Fig. 17B) (Rodrigues, Ramirez-Guadiana, et al. 2016, Doan et al. 2009). In addition, AG can be co-immunoprecipitated with AD, AE and AF (Doan et al. 2009). Altogether, these data indicate that AG interacts directly or indirectly with AD, AE, AF, AH, Q and GerM.

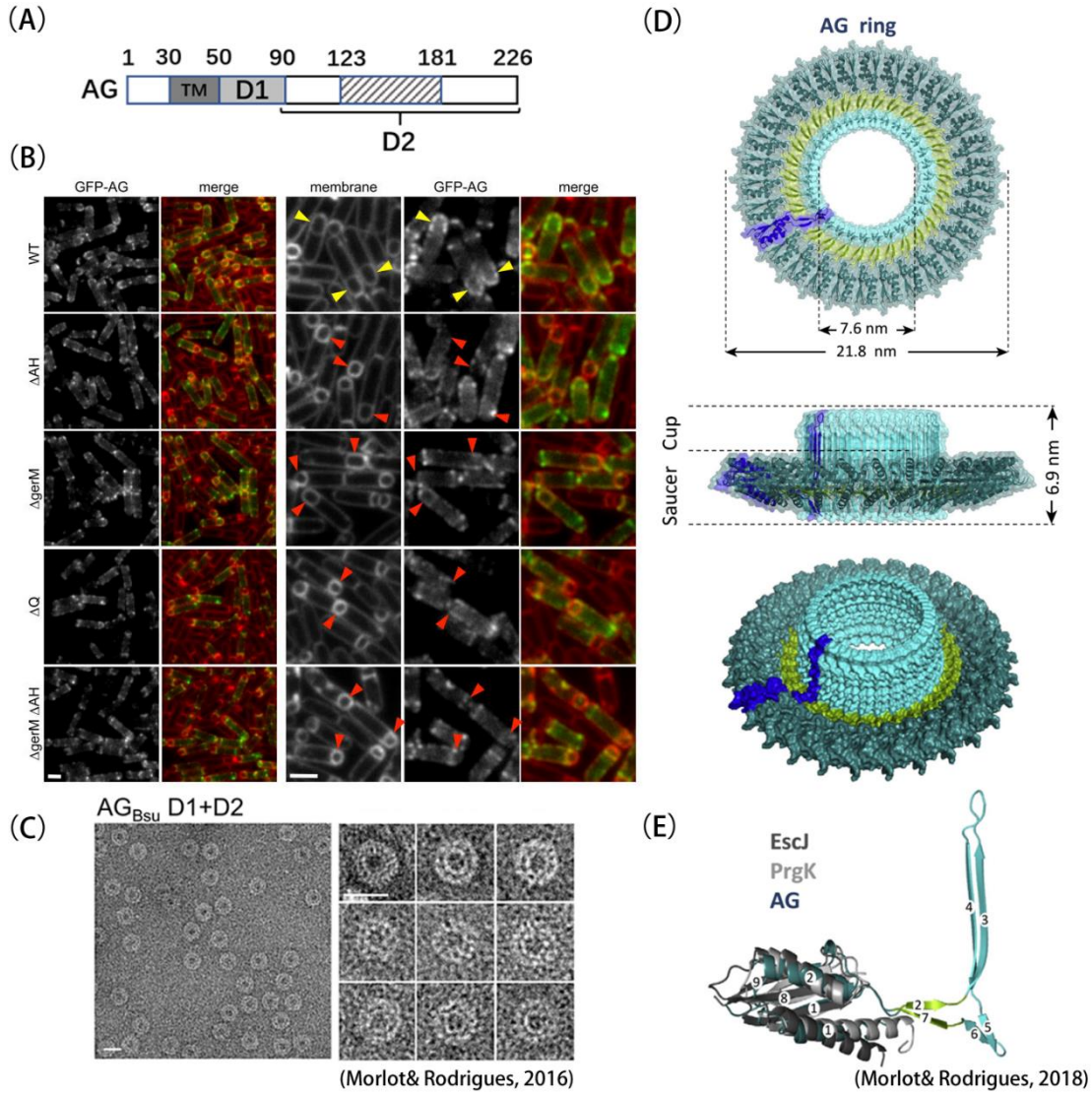


Figure 17. The ring-forming component SpoIIIAG of the A-Q complex. (A) Topology of AG showing the transmembrane segment (TM) and the extracytoplasmic D1 and D2 domains. Numbering refers to the *B. subtilis* AG sequence. The hatched area represents the region inserted within the RBM-homologous sequence and displaying the β-triangle fold. **(B)** AG localization around the forespore requires AH, Q and GerM. Conventional fluorescence microscopy images obtained with a strain producing a GFP-AG fusion in a wild-type (WT), *ΔspoIIIAH* (ΔAH), *ΔgerM* (ΔGerM), *ΔspoIIQ* (ΔQ) and *ΔspoIIIAH ΔgerM* (ΔAH ΔGerM) background. The GFP signal is false-colored green in the merge image. Membranes are stained with TMA-DPH (false-colored red in the merge image). Enrichment of AG around the forespore is reduced in the absence of GerM, Q, and to a lesser extent in the absence of AH. From Rodrigues et al., 2016a. **(C)** Negative-stained EM images of rings obtained with a $AG_{S51\text{-}S229}$ construct. The left panel shows a large field of homogeneous rings. The right panel shows enlargements of individual rings. Scale bar, 20 nm. From Rodrigues et al., 2016b. **(D)** Ribbon and surface representation of an $AG_{89\text{-}227}$ oligomeric ring displayed in top, side, and tilted views from top to bottom. An AG monomer is labeled in blue. **(E)** Structural overlay of $AG_{89\text{-}227}$ (PDB code 5WC3) and $EscJ_{96\text{-}186}$ (PDB code 1YJ7, in dark grey), $PrgK_{94\text{-}176}$ (PDB code 4OYC, in light grey). Panels D and E are from Morlot et al., 2018.

In 2016, my hosting lab purified a recombinant construct of AG encompassing domains D1 and D2 ($AG_{S51-S229}$). They discovered that this construct forms homogenous rings with a large central pore (Fig. 17C) and obtained a 3D cryo-EM map at 35 Å that they used to model the oligomeric AG ring (Rodrigues, Henry, et al. 2016). In collaboration with the Rudner lab (Harvard Medical School, Boston, USA), they further showed that residues predicted to be localized at the AG-AG interface are required for AG oligomerization *in vitro* and for σG activity *in vivo* (Rodrigues, Henry, et al. 2016). The year after, the cryo-EM structure of the AG rings was further refined to 3.5 Å by the Strynadka's laboratory (Zeytuni et al. 2017).

AG forms 30-mer rings displaying a "cup-and-saucer" architecture with a 6-Å central pore. The external diameter of the rings is 21.8 nm, close to the dimensions of the 24-mer EscJ/PrgK rings from T3SS (Worrall et al. 2016). The height of the AG rings are 6.9 nm while the eighth of the flagellar FliF ring is also about 7 nm (Schraidt et al. 2010) (Fig. 17D).

The AG monomer is made of three distinct regions. A first region (residues 90-123 and 181-229, shown in deepteal in figure 17D-E is made of the RBM-homologous moieties and folds into the canonical RBM fold. Upon oligomerization, it forms the saucer region in the AG rings (Fig. 17D, E). This region can be superimposed well onto the second EscJ/PrgK RBM domain (Cα backbone rmsd <2 Å) (see the *Discussion* part). The second and third regions (residues 124-180, shown in green and cyan, respectively, in Fig. 17D, E) are made of the sequence inserted into the canonical RBM and form a planar β-triangle. The third region contains two long and two short β-strands that form a vertical 60-stranded anti-parallel β-barrel upon oligomerization. This unique structure, which makes up the cup region of the AG rings, appears as a hollow channel with an inner diameter of 7.6 nm. Oligomerization of the second region forms an horizontal β-ring that connects the cup and saucer regions.

AG oligomerization results from the packing of helix α1 from one monomer against the three-stranded β-sheet from the adjacent monomer. This interface is very similar to the canonical homo-holigomerization interface of the second RBM domain in EscJ/PrgK rings (see the *Discussion* part). Both polar interactions and hydrophobic forces stabilize the RBM interface in the saucer region of the AG rings while the cup region involves hydrogen bonds

that typically stabilizes β-barrels.

The discovery of the AG oligomeric rings provides direct evidence that the A-Q complex assembles a channel between the mother cell and the forespore. It is worth noting that AG oligomerization is not unique to *B. subtilis*. Indeed, published and unpublished studies from my laboratory also showed oligomerization of AG orthologs from *Geobacillus thermodenitrificans*, *Geobacillus stearothermophilus* and *Acetonema longum* (Rodrigues, Henry, et al. 2016). The conserved oligomerization ability of AG suggests that if the A-Q complex contains a transport or secretion conduit that connects mother cell and forespore, AG is likely an essential component of such channel.

7. SpoIIIAH and SpoIIQ: the bridging proteins

SpoIIIAH (**AH**) is anchored in the mother cell membrane through a single N-terminal transmembrane domain and its harbors a soluble domain localized in the intermembrane space (Fig. 18A). The X-ray crystal structure of AH from *B. subtilis* (Levdikov et al. 2012, Meisner et al. 2012) showed that it contains a C-terminal RBM-like domain which aligns well (Rmsd < 2 Å) with the second RBM domains of EscJ (PDB code 1YJ7, residues 98-186) and PrgK (PDB code 4OYC, residues 98-200) despite low sequence identity (< 15%). A major difference with canonical αββαβ RBMs is that AH RBM possesses an additional α-helix called α0 (Fig. 18B). Residues 26 to 103 in the soluble domain of AH are not visible in the electron density map and are thus likely disordered.

SpoIIQ (**Q**) is anchored in the forespore membrane through a single N-terminal transmembrane segment and its harbors a soluble domain localized in the intermembrane space (Fig. 18A). The X-ray crystal structure of Q from *B. subtilis* (Meisner et al. 2012, Levdikov et al. 2012) showed that it contains a LytM metalloendopeptidase domain (20% sequence identity with LytM from *Staphylococcus aureus*) but lacks two of the catalytic residues in the active site pocket. Consistent with this, no cell wall hydrolytic activity has been detected for the Q orthologue from *B. subtilis*. Compared to LytM, Q contains an additional region (a short helix and a β-hairpin) engaged into the interaction with AH.

The crystal structure of the AH-Q heterodimer shows that AH-Q interface is formed by the pairing of the three-stranded β-sheet of AH with the Q β-hairpin, resulting in a five-stranded composite β-sheet (Fig. 18B) (Meisner et al. 2012, Levdikov et al. 2012).

Although neither AH nor the AH-Q dimer oligomerize in solution or in the crystal, the fact that AH contains an RBM-like domain strongly suggests that it could form rings. As AH interacts with Q, the two proteins might thus form two homo-oligomeric rings piled onto each other to form part of a channel connecting the mother cell and forespore compartments. Several models of the AH-Q rings have been generated *in silico* (Levdikov et al. 2012, Meisner et al. 2012). In a 12-mer model, the AH-Q channel would display an inner diameter of 6 to 10 nm. In

a 15-mer model, the inner diameter of the AH-Q channel would range from 8 to 13 nm, and from 11 to 15 nm in an 18-mer model. These models might however be still far from the physiological arrangement of the putative AH-Q rings because the AH-AH oligomerization interfaces are all very different from those of EscJ or PrgK.

Similar to AG, AH and Q localize at the mother cell-forespore interface (Fig. 18C). The localization of AH requires Q while the localization of Q requires either AH or GerM (see section II.C.8) (Fredlund et al. 2013). Finally, in agreement with the idea that A-Q proteins form a membrane complex, Q can be co-immunoprecipitated with AH, AE and AF (Fig. 18D).

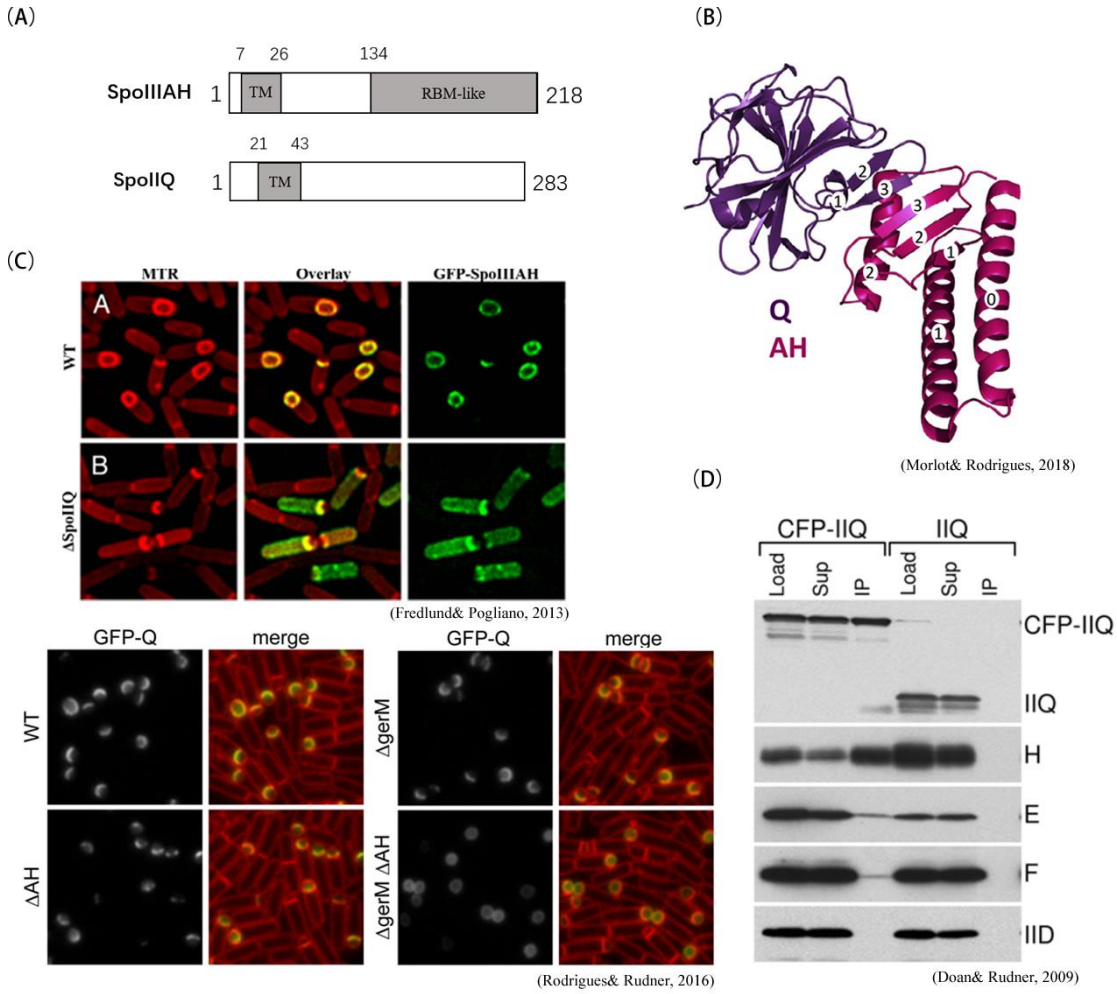


Figure 18. The SpoIIIA-SpoIIQ dimer. (A) Topology of SpoIIIAH and SpoIIQ. **(B)** Ribbon representation of the SpoIIIAH -SpoIIQ dimer showing the intermolecular composite β-sheet. Helix α1 and strands β2 and β3 from Q are annotated. α-helices and β-strands of AH are numbered 0-1 and 1-3, respectively. This panel is from Morlot and Rodrigues, 2018. **(C)** The top panels show conventional fluorescence microscopy images obtained with a strain producing a GFP-SpoIIIAH fusion in a wild-type (WT) or *ΔspoIIQ* (ΔSpoIIQ) background. The GFP signal is false-colored green and membranes are stained with MTR, which is false-colored red. Enrichment of SpoIIIAH around the forespore is reduced in the absence of Q. The bottom panels show conventional fluorescence microscopy images obtained with a strain producing a GFP-SpoIIQ (GFP-Q) fusion in wild-type (WT), *ΔspoIIIAH* (ΔAH), *ΔgerM* (ΔGerM) or *ΔspoIIIAH ΔgerM* (ΔAH ΔgerM) backgrounds. The GFP signal is false-colored green in the merge image. Membranes are stained with TMA-DPH (false-colored red in the merge image). Enrichment of Q at the forespore-mother cell interface is reduced in the absence of both AH and GerM. These panels are from Fredlund et al., 2013, as well as from Rodrigues et al., 2016a. **(D)** Immunoprecipitations performed on detergent-solubilized membrane fractions from *B. subtilis* sporulating cells harvested 2.5 h after the onset of sporulation. Membrane samples were obtained from a *ΔspoIIQ* mutant harboring a cfp-spoIIQ fusion at a non-essential locus (CFP-IIQ) and from a strain expressing the native *spoIIQ* gene (IIQ). The load (L), supernatant (Sup), and immunoprecipitated (IP) fractions were subjected to immunoblot analysis using anti-GFP, anti-SpoIIIAH (H), anti-SpoIIIAF (F), anti-SpoIIIAE (E), and anti-SpoIID (IID) sera. This panel is from Doan et al., 2009.

8. GerM

GerM is a lipoprotein with a N-terminal lipobox and two GerMN domains (Fig. 19A). GerM was originally identified through a genetic screen and was shown to be related to spore germination (Sammons, Slynn, and Smith 1987, Slynn et al. 1994).

Recent studies revealed that GerM is also involved in spore formation. Deletion of the *gerM* gene results in defects similar to those observed in *ΔspoIIIA* or *ΔspoIIQ* mutant strains (collapsed forespores, membrane deformities, reduced σG activity and heat-resistant spores) (Rodrigues, Ramirez-Guadiana, et al. 2016). GerM is localized around the forespore during the engulfment stage (Fig. 19B). GerM mis-localizes to the whole mother cell membrane in the absence of Q (which also cause the mis-localization of AH). To further dissect the localization dependency of GerM, the Rudner's laboratory constructed a *B. subtilis* strain expressing Q168A and D123A variants of Q. Residues Q168 and D123 are located in the degenerated active site groove of Q and their mutation into Ala do not cause AH delocalization. In the $spoIIQ^{Q168A}$ and $spoIIQ^{D123A}$ backgrounds, AH is required to maintain GerM localization around the forespore (Fig. 19B). These data thus suggest that GerM localization requires interaction with residues Q168 and D123 of Q or interaction with AH. As mentioned in the preceding section, GerM in turn is required for the localization of Q in the absence of AH (Fig. 19C). Therefore, GerM, Q and AH appear as a tripartite complex whose members have interdependent localization requirements.

The GerMN domains exists in many major bacterial phyla, such as *Firmicutes, Cyanobacteria, Deinococcus-Thermus, Spirochaetes, Actinobacteria*, and *Proteobacteria* groups (Rigden and Galperin 2008). Tandem copies of GerMN domains (called GerMN1 and GerMN2) are only found in the sporulation protein GerM while isolated GerMN domains are present in spore-forming and non-sporogenous bacterial species. However so far, the function of GerMN domains is not clear.

At the beginning of my thesis, the crystal structure of the two GerMN domains from *B. subtilis* GerM was obtained by Jennyfer Trouvé in my hosting laboratory and I contributed to

their functional characterization and structural analysis as described in the *Results* part.

The two GerMN domains fold against each other in a butterfly-like arrangement stabilized by a series of hydrophobic interactions, salt bridges and multiple hydrogens bonds. Interestingly, the core of both GerMN domains resembles RBM folds (αββαβ fold) (Trouve et al. 2018). Despite low sequence identity (< 15%) with RBM domains, the RBM-like core of GerMN1 and GerMN2 superimposes well with RBMs from EscJ, PrgK, AH or AG, as described in the *Results* section. However, there is no clear evidence indicating that GerM could oligomerize through this RBM-like domain. The asymmetric unit of the crystal contains a tetramer whose crystallographic symmetries form a protein filament in which each protomer interacts with the adjacent one through a conserved dimerization interface (Trouve et al. 2018). This dimerization interface is however very different from those observed in AG, EscJ/PrgK or PrgH rings. Moreover, GerM oligomers are not detected in solution, suggesting that the GerM oligomer observed in the crystal is not a physiological assembly.

In summary, GerM is another component of the A-Q complex containing an RBM-like domain. It forms a tripartite complex with AH and Q but its function so far remains limited to the localization of these proteins and/or stability of the A-Q complex.

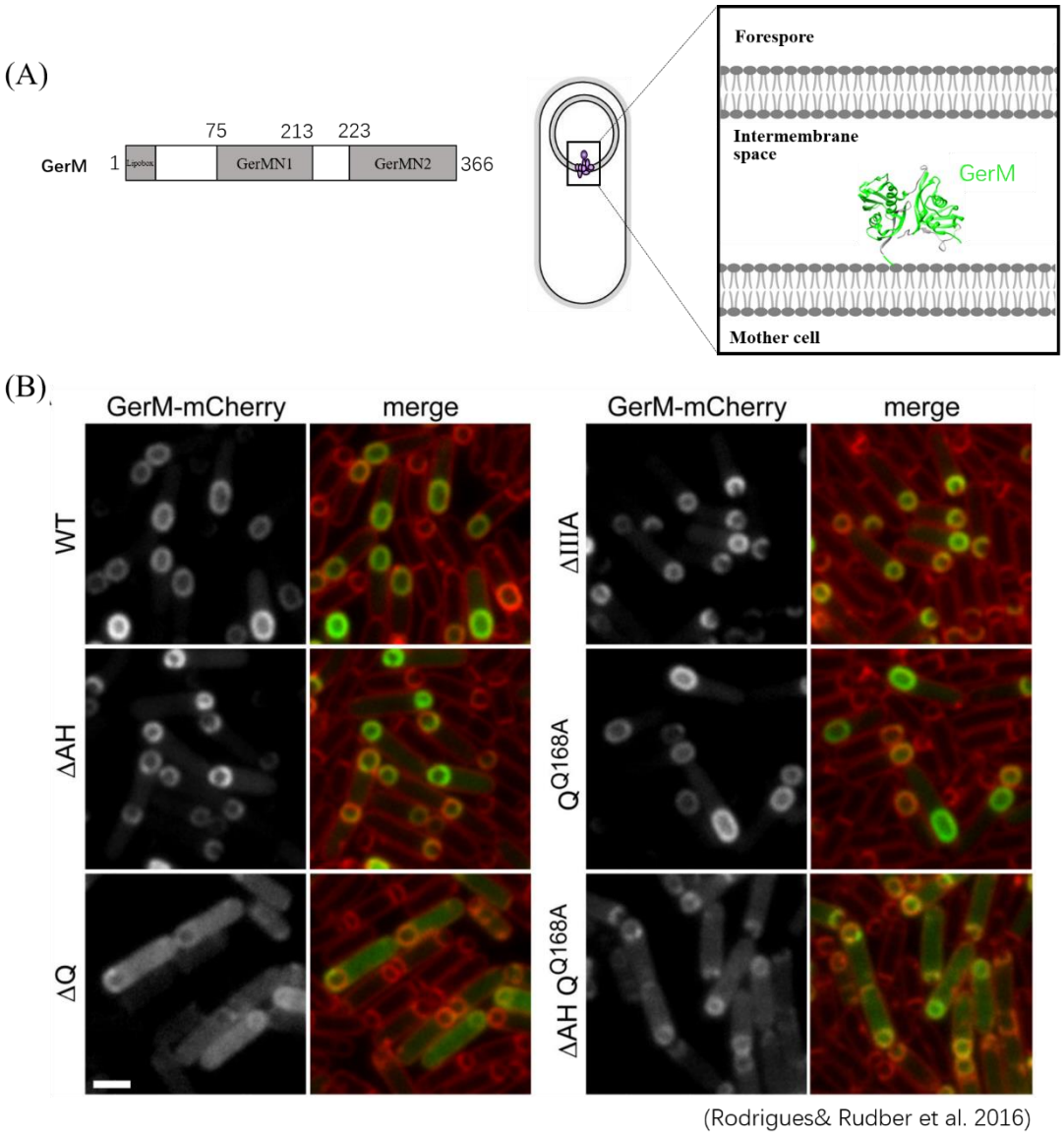


Figure 19. GerM displays an RBM-like domain and is functionally linked to the AH-Q dimer. (A) Topology of GerM showing the N-terminal lipobox and the two RBM-like domains called GerMN1 and GerMN2 (left panel). On the right panel, the structure of the soluble region of GerM ($GerM_{F26\text{-}F366}$) is shown in ribbon representation with an arbitrary orientation in the intermembrane space. From Trouve al., 2018. **(B)** Conventional fluorescence microscopy images of GerM-mCherry localization 2.5 hours after the onset of sporulation. Images were obtained from wild-type (WT), *ΔspoIIIAH* (ΔAH), *ΔspoIIQ* (ΔQ), *ΔspoIIIA* (ΔA), *spoIIQ(Q168A)* (Q^{Q168A}) and *ΔspoIIIAH spoIIQ(Q168A)* (ΔAH Q^{Q168A}) strains. The mCherry signal and the TMA-DPH-stained membranes are respectively false-colored green and red in the merge image. Scale bars, 2 µm. From Rodrigues& Rudner et al., 2016.

9. Function of the A-Q complex

As soon as they were predicted, the structural similarities between A-Q proteins and components of specialized secretion systems led to the hypothesis that the A-Q complex might assemble into a large macromolecular complex transporting a yet-to-be discovered molecule from the mother cell to the forespore compartment, to activate σG or maintain global forespore physiology (Fig. 20A) (Meisner et al. 2008, Camp and Losick 2008, 2009, Doan et al. 2009)

A first model proposes that the A-Q complex would form an apparatus resembling endotoxin secretion systems (Fig. 20B))(Morlot and Rodrigues 2018). In this **secretion system model**, the A-Q complex would actively secrete protein(s) from the mother cell to the forespore, possibly using ATP hydrolysis as a source of energy for secretion. If AG is an inner component of the A-Q complex, its 6 Å inner diameter is consistent with the idea that the complex could transport proteins whose molecular weight could reach 150 kDa. However, no potential secreted substrate has been identified so far.

In addition, a study carried out by the Losick laboratory has shown that a σG-independent DNA polymerase (the T7 polymerase) is also inactive in the absence of the A-Q proteins (Camp and Losick 2009). This observation led to a second model in which the A-Q complex would act as a gap-like junction between the mother cell and forespore. In this **"feeding-tube" model**, the A-Q complex would passively transport small molecules required for global forespore physiology, such as nucleotides, electrolytes, or other metabolic compounds (Camp and Losick 2009). The A-Q complex is synthesized during engulfment, which progressively isolates the forespore from the outside environment. Feeding by the A-Q complex would thus allow the forespore maintaining its metabolic potential. This model is however not consistent with the fact that engulfment-defective strains (in which the forespore has always access to the extracellular environment) still depend on the A-Q complex for σG activity.

Finally, we cannot exclude that the A-Q complex might neither secrete nor transport anything but serve as a macromolecular pilar that would maintain the cohesion between the two membranes surrounding the forespore. This **piliation model** is supported by electron

micrographs that show highly variable distance between the inner and outer forespore membranes in the absence of the A-Q complex (Fig. 11B). However, it seems unlikely that such a large molecular structure would be needed to fulfill this function.

In conclusion, there is still much to learn about the A-Q complex to understand its role during sporulation. Since the identification of potentially transported substrates has failed so far, determining the global architecture of the complex appears to be a consistent approach to better understand the function of the A-Q complex. It will also provide insights into the evolutionary relationship between the A-Q complex and specialized secretion systems.

D. The SpoIIIA-SpoIIQ complex in *Clostridium difficile*

The A-Q complex is conserved in all spore-forming bacteria. It has been intensively studied in *B. subtilis* and to a lesser extend in the human pathogen *C. difficile*. In these two species, the A-Q complex displays interesting similarities and differences.

In both *B. subtilis* and *C. difficile*, the A-Q complex includes eight mother cell proteins (AA to AH) encoded within the *spoIIIA* operon and the forespore protein Q. Two other proteins (the mother cell lipoprotein GerM and the forespore protein SpoIIIL) interact functionally and/or physically with the A-Q complex in *B. subtilis* but are absent in *C. difficile* (Morlot and Rodrigues 2018).

In both species, the complex is localized at the interface between the mother cell and forespore, and AH directly interacts with Q (Doan, Marquis, and Rudner 2005, Serrano et al. 2016). The localization dependency between AH and Q however differs in *B. subtilis* and *C. difficile*. In *B. subtilis,* the localization of Q requires the presence of AH or GerM: in the absence of AH or GerM, Q remains localized at the mother cell-forespore interface provided that GerM or AH, respectively, is present (Rodrigues, Ramirez-Guadiana, et al. 2016). In addition, Q localization depends on the degradation of the septal peptidoglycan by the SpoIID and SpoIIP hydrolases (Rodrigues, Ramirez-Guadiana, et al. 2016, Rodrigues et al. 2013). By contrast in *C. difficile*, GerM is not conserved and Q localization requires AH, and it does not depend on peptidoglycan degradation by the SpoIID-SpoIIP hydrolase machinery(Serrano et al. 2016). Finally in *B. subtilis,* AH localization strongly depends on Q (Doan, Marquis, and Rudner 2005) but in *C. difficile*, AH localization is only mildly impaired in the absence of Q(Serrano et al. 2016). In the later species, AH interacts directly with the SpoIID-SpoIIP machinery and its localization might thus involve these proteins (Dembek et al. 2018, Kelly and Salgado 2019).

Another interesting difference between the two bacterial species resides in the globular domain of Q. The soluble region of Q contains a LytM endopeptidase domain found in several peptidoglycan hydrolases. In *C. difficile*, Q harbors an intact LytM domain. It carries all the catalytic residues, including the critical histidine required for Zn^{2+} coordination and formation of the catalytic site, suggesting that Q might degrade or bind peptidoglycan (Crawshaw et al.

2014). By contrast in *B. subtilis*, Q harbors a degenerated LytM domain, in which the Zn-binding histidine is substituted by a serine (Meisner et al. 2012, Levdikov et al. 2012). This divergence might be related to the fact that C. *difficile* cells lacking Q show a more severe phenotype than *ΔspoIIQ* mutants of *B. subtilis* (Serrano et al. 2016). In *ΔspoIIQ* mutants of *C. difficile*, engulfment is impaired at earlier stages than in *B. subtilis*, and the forespore membranes bulge into the mother cell (Serrano et al. 2016), a phenomenon possibly due to defect in peptidoglycan degradation. Intriguingly enough though, the bulging phenotype is not observed in a *C. difficile* strain expressing a H120S variant. This variant cannot bind zinc and should thus not be able to degrade peptidoglycan. Besides, the Q(H120S) variant still interact with AH *in vitro* and *in vivo* but the complex seems to be less stable(Serrano et al. 2016). Since neither peptidoglycan binding nor peptidoglycan degradation has been reported so far for the *C. difficile* Q orthologue, it is thus possible that zinc coordination is only required to enhance the folding of the protein and/or to stabilize the AH-Q interface(Kelly and Salgado 2019). In further support of this idea, secondary structure predictions indicate that the region of Q involved in the interface is less structured in *C. difficile* than in *B. subtilis(Serrano et al. 2016)*.

A final intriguing observation regarding AH and Q is that in *C. difficile*, the two proteins are important for both σ^G activity in the forespore and σ^K activity in the mother cell, although σ^K activity in *C. difficile* is partially independent of σ^G activity (Serrano et al. 2016, Fimlaid et al. 2015). In *B. subtilis*, proteolysis of pro-σ^K into active σ^K requires the production of SpoIVB (Campo and Rudner 2006), whose gene is under the control of σ^G. By contrast, the σ^K protein in *C. difficile* lacks a pro-sequence and its activity does not fully depend on σ^G(Fimlaid et al. 2013). One hypothesis for the requirement of AH and Q for σ^K activity in *C. difficile* might be that the A-Q complex transports an inhibitor of σ^K from the mother cell to the forespore, or the transport of a molecule required for σ^K activity from the forespore to the mother cell (Serrano et al. 2016). Alternatively, AH and Q might be directly involved in σ^K activity in *C. difficile*, for example through a physical interaction between AH and σ^K.

The other members of the A-Q complex have only been poorly studied in *C. difficile*. Lack of the AA ATPase in *C. difficile* causes defects in coat assembly and bulging of the forespore

membranes in a fraction of cells (Fimlaid et al. 2015), while in AA mutants of *B. subtilis*, the forespore collapses (Doan et al. 2009).

In *B. subtilis*, two proteins were shown to form ring-like structures *in vitro*: AG forms homogenous 30-mer rings while AF forms heterogenous annular structures (Rodrigues, Henry, et al. 2016, Zeytuni et al. 2017, Zeytuni et al. 2018a). Intriguingly despite a rather good sequence identity (about 26%) between the AG orthologues in *B. subtilis* and *C. difficile*, the later orthologue does not oligomerize *in vitro* (unpublished data from my PhD project).

The two AF orthologues display a weak sequence identity (about 16%) and they do not cluster when analyzed by homology mapping approach(Ramos-Silva, Serrano, and Henriques 2019). The high sequence divergence between the two AF orthologues suggests that they might have different functions and/or binding partners. As the heterogeneous and sparse ring-like structures formed by AF from *B. subtilis* do not provide strong evidence that the protein is able to form rings, it would be interesting to investigate the ability of the *C. difficile* AF orthologue to oligomerize.

To conclude, further studies of the A-Q complex in various species should unravel crucial differences that should help unraveling its structure and function.

III. Objectives and rationale of the experimental approaches

So far, the most common structural elements shared by the specialized secretion systems and the A-Q complex are the RBM domains. An intriguing observation is that very few RBM-containing proteins are able to spontaneously form rings *in vitro*. So far, only secretins from T2SS and T3SS, as well as AG from the A-Q complex oligomerize on their own when isolated from the cellular context (Rodrigues, Henry, et al. 2016, Tosi et al. 2014, Hay, Belousoff, and Lithgow 2017, Hay et al. 2018, Nouwen et al. 2000, Chami et al. 2005, Yan et al. 2017, Howard et al. 2019). In this context, I have dedicated part of my Ph.D. work to investigate the structural determinants of RBM oligomerization in sporulation proteins. Because some of the RBM-containing sporulation proteins harbor additional secondary structures that pack against the RBM core in proteins that do not oligomerize *in vitro* (AH and GerM), I aimed to determine whether these regions impair the oligomerization ability of the RBM domains. On the other hand, I investigated which region(s) in AG is(are) required for oligomerization. The experimental approach used to address these questions was based on recombinant protein production and purification, coupled to SEC-MALLS (Size-exclusion chromatography and multi-angle laser light scattering) and negative-stain transmission electron microscopy (TEM) analysis of high-molecular weight species. I applied this approach to full-length and truncated forms of the RBM-containing soluble domains of AG, AH and YhcN.

In parallel, I have pushed the structural characterization of AG, which is so far the only ring-forming component of the A-Q complex. The D1 domain is not visible in cryo-EM reconstructions of the soluble region of AG (containing the D1 and D2 domains) (Rodrigues, Henry, et al. 2016, Zeytuni et al. 2017) and is thus likely flexible in this construct. However, I showed that the D1 domain is required for AG oligomerization in the first part of my project (described in the above paragraph). My first objective here was thus to study the structure of the full-length membrane form of AG ($AG_{M1\text{-}S229}$) by cryo-EM. Obtaining a cryo-EM model of $AG_{M1\text{-}S229}$ would provide information regarding the structure and oligomerization ability of the D1 domain. In addition, getting the dimensions of full-length AG rings would shed light on their positioning relative to the inner forespore membrane. My second objective was to

reconstitute part of the A-Q complex *in vitro* using the full-length $AG_{M1\text{-}S229}$ construct and full-length membrane forms of other RBM-containing components: AF and the AH-Q heterodimer. To do so, I have produced and purified full-length membrane forms of AF, AG, AH and Q and tested whether they could form rings in the presence of detergents.

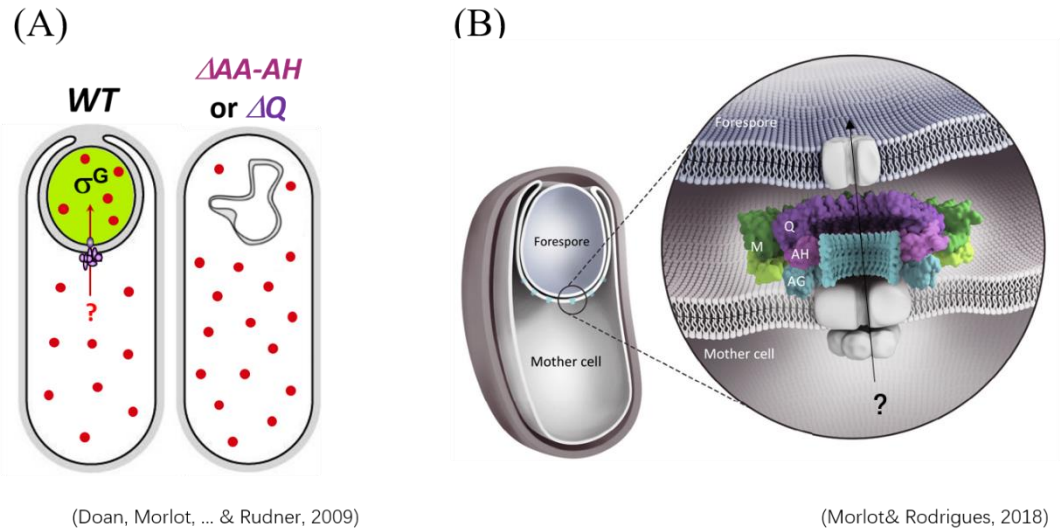


Figure 20. (A) Model proposing a putative role of the A-Q complex during *B. subtilis* sporulation. Passive transport or active secretion of an unknown metabolite/osmolyte (red circles) into the forespore would maintain forespore integrity and σG activity (green forespore) in wild-type cells (WT). In the absence of the A-Q proteins (ΔAA-AH or ΔQ), the forespore loses metabolic potential, collapses and σG activity cannot be maintained. Panel A is from Doan et al., 2009. **(B)** Schematic diagram showing a potential architecture of the A-Q complex, in which stacked rings of AG (in cyan), AH (pink), GerM (green) and Q (purple) form a channel in the intermembrane space. Other components of the A-Q complex are shown in grey. The AA ATPase might form a hexamer in the mother-cell cytoplasm. Panel B is from Morlot and Rodrigues, 2018.

Methods and Materials

Methods and Materials

A. Plasmids and oligonucleotides used during my PhD

Plasmids and oligonucleotides used in my work are listed in Table 2. All *B. subtilis* genes were cloned from strain 168 as translational fusions to a poly-histidine (His_6) tag followed by the cleavage site (SUMO tag) of the Ulp1 protease (Uehara et al. 2010). All plasmids carry an ampicillin resistance gene for selection of the transformed bacteria.

Table 2. Plasmids and oligonucleotides used in this study.

Construct	Abbreviation/Description/5'-3' sequence	Source
Plasmids		
SpoIIIAF		
pCR 93	HS-AF$_{55\text{-}206}$ / *His$_6$-SUMO-spoIIIAF*$_{K55\text{-}E206}$	This work (CR)
pCR 291	HS-AF$_{1\text{-}206}$ / *His$_6$-SUMO-spoIIIAF*$_{M1\text{-}E206}$	This work (CR)
SpoIIIAG		
pCR 94	HS-AG$_{51\text{-}229}$ / *His$_6$-SUMO-spoIIIAG*$_{S51\text{-}S229}$	Rodrigues et al., 2016
pCR 284	HS-AG$_{1\text{-}229}$ / *His6-SUMO-spoIIIAG*$_{M1\text{-}S229}$	This work (CR)
pCM 124	HS-AG$_{51\text{-}229}$Δ(128-180) / *His$_6$-SUMO-spoIIIAG*$_{S51\text{-}S229}$Δ(L128-K180)	This work (BL)
pCM 129	HS-AG$_{126\text{-}181}$ / *His$_6$-SUMO-spoIIIAG*$_{T126\text{-}K181}$	This work (BL)
SpoIIIAH		
pCR 105	HS-AH$_{32\text{-}218}$ / *His$_6$-SUMO-spoIIIAH*$_{A32\text{-}K218}$	Rodrigues et al., 2016
pCR 290	HS-AH$_{1\text{-}218}$ / *His$_6$-SUMO-spoIIIAH*$_{M1\text{-}K218}$	This work (CR)
pCM 127	HS-AH$_{32\text{-}218}$Δ(105-128) / *His$_6$-SUMO-spoIIIAH*$_{A32\text{-}K218}$Δ(L105-V128)	This work (BL)
pCM 128	HS-AH$_{129\text{-}218}$ / *His$_6$-SUMO-spoIIIAH*$_{S129\text{-}K218}$	This work (BL)
SpoIIQ		
pCR 106	HS-Q$_{43\text{-}283}$ / *His$_6$-SUMO-spoIIQ*$_{Q43\text{-}S283}$	Rodrigues et al., 2016
pCR 289	HS-Q$_{1\text{-}283}$ / *His$_6$-SUMO-spoIIQ*$_{M1\text{-}S283}$	This work (CR)
GerM		
pCR 280	HS-GerM$_{26\text{-}366}$ / *His$_6$-SUMO-gerM*$_{F26\text{-}F366}$	Trouve et al., 2018
pCR 282	HS-GerMN1 / *His$_6$-SUMO-gerM*$_{T76\text{-}E213}$	Trouve et al., 2018
pCR 283	HS-GerMN2 / *His$_6$-SUMO-gerM*$_{T223\text{-}F366}$	Trouve et al., 2018
YhcN		
pCM 122	HS-YhcN$_{24\text{-}189}$ / *His$_6$-SUMO-yhcN*$_{A24\text{-}E189}$	This work (AH)
pCM 123	HS-YhcN$_{78\text{-}189}$ / *His$_6$-SUMO-yhcN*$_{D78\text{-}E189}$	This work (AH)
pCM 125	HS-YhcN$_{24\text{-}150}$ / *His$_6$-SUMO-yhcN*$_{A24\text{-}N150}$	This work (BL)
pCM 126	HS-YhcN$_{78\text{-}150}$ / *His$_6$-SUMO-yhcN*$_{D78\text{-}N150}$	This work (BL)

Oligo.		
pCM 124	AGΔ(L128-K180)$_{for}$ / gatgcaacgtcgaaacctgatatacgc	This work (BL)
	AGΔ(L128-K180)$_{rev}$ / gcgtatatcaggtttcgacgttgcatc	This work (BL)
pCM 125	YhcN$_{24\text{-}150\text{-}for}$ / **gtttacgtttcagctaac**taataagcttgtcgagcacc	This work (BL)
	YhcN$_{24\text{-}150\text{-}rev}$ / ggtgctcgacaagcttatta**gttagctgaaacgtaaac**	This work (BL)
pCM 126	YhcN$_{78\text{-}150\text{-}for}$ / **gtttacgtttcagctaac**taataagcttgtcgagcacc	This work (BL)
	YhcN$_{78\text{-}150\text{-}rev}$ / ggtgctcgacaagcttatta**gttagctgaaacgtaaac**	This work (BL)
pCM 127	AHΔ(L105-V128)$_{for}$ / **gtttacgtttcagctaac**taataagcttgtcgagcacc	This work (BL)
	AHΔ(L105-V128)$_{rev}$ / ggtgctcgacaagcttatta**gttagctgaaacgtaaa**	This work (BL)
pCM 128	AH$_{129\text{-}218\text{-}for}$ / **gaacagattggtggt**tcaagcgatgatgc	This work (BL)
	AH$_{129\text{-}218\text{-}rev}$ / Gcatcatcgcttga**accaccaatctgttc**	This work (BL)
pCM 129	AG$_{126\text{-}181\text{-}for}$ / **cgca**ggatccacgtcgctga	This work (BL)
	AG$_{126\text{-}181\text{-}rev}$ / **ccga**ctcgagtttttcgtttgg	This work (BL)

The name of the persons who have cloned the genes are indicated into brackets : BL, Bowen Liu; CR, Christopher Rodrigues; AH, Adriano Henriques.

B. RBM-like plasmid construction

pCR291 [*His$_6$-SUMO-spoIIIAF*$_{M1\text{-}E206}$], **pCR284** [*His$_6$-SUMO-spoIIIAG*$_{M1\text{-}S229}$], **pCR290** [*His$_6$-SUMO-spoIIIAH*$_{M1\text{-}K218}$], **pCR289** [*His$_6$-SUMO-spoIIQ*$_{M1\text{-}S283}$], **pCM122** [*His$_6$-SUMO-yhcN*$_{A24\text{-}E189}$], **pCM123** [*His$_6$-SUMO-yhcN*$_{D78\text{-}E189}$] and **pCM129** [*His$_6$-SUMO-spoIIIAG*$_{T126\text{-}K181}$] were generated by ligating a PCR product containing the relevant DNA segment into pCR134 [*His$_6$-SUMO-spoIIIAG(Alo)$_{66\text{-}201}$*] cut with *Bam*HI/*Xho*I (Rodrigues, Henry, et al. 2016).

pCM124 [*His$_6$-SUMO-spoIIIAG*$_{S51\text{-}S229}$Δ(L128-K180)], **pCM125** [*His$_6$-SUMO-yhcN*$_{A24\text{-}E189}$], **pCM126** [*His$_6$-SUMO-yhcN*$_{D78\text{-}E189}$], **pCM127** [*His$_6$-SUMO- spoIIIAH*$_{A32\text{-}K218}$Δ(L105-V128)] and **pCM128 [***His$_6$-SUMO-spoIIIAH*$_{S129\text{-}K218}$] were generated by site-directed mutagenesis using appropriate oligonucleotide primers and plasmid templates : AGΔ(L128-K180)$_{for/rev}$ on pCR94, YhcN$_{24\text{-}150\text{-}for/rev}$ on pCM122 or pCM123, AHΔ(L105-V128)$_{for/rev}$ or AH$_{129\text{-}218\text{-}for/rev}$ on pCR105.

PCR products were cleaned up using the NucleoSpin Gel and PCR clean-up kit. For site-directed mutagenesis, the *Dpn*I restriction enzyme was used for digestion of the methylated plasmid template. DNA preparations were transformed into *E. coli* DH5α competent cells and

grown on LB-agar-Amp plates at 37℃, overnight. Single colonies were inoculated into LB-Amp medium for an overnight culture at 37℃. Plasmid DNA was extracted and sent for DNA sequencing to the GENEWIZ Ltd company.

C. Production of recombinant proteins

Plasmids for expression of recombinant proteins were transformed into competent *E. coli* BL21 (*DE3*) STAR cells. Cells were grown in 2-L baffled flasks containing 500-mL of Terrific Broth (TB) medium supplemented with ampicillin (100 μg/mL) at 37°C under agitation (180 rpm) until OD_{600nm} reached ~ 0.8. Cells were cooled to 25°C (for soluble proteins) or 16°C (for membrane proteins) until the OD_{600nm} reached ~ 1.5. Expression of the gene of interest was then induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and cultures were further incubated at 25℃ or 16°C for ~ 18 h. Cells were harvested by centrifugation at 3,500 × g, at 4℃ for 30 min. After been resuspended in $1/50^{th}$ volume of buffer A (Tris-HCl 50 mM pH 8.0, NaCl 500 mM, imidazole 25 mM, glycerol 10% vol/vol) supplemented with the Complete™ EDTA-free cocktail of protease inhibitors (Roche), cells from a total of 2 L of culture were either flash-frozen into liquid nitrogen for storage at -80°C, or directly used for protein purification.

D. Purification of soluble recombinant proteins

Cells were lysed with a cell disruptor (Microfluidics) at 15 kpsi for 5 passages. Cell debris were pelleted by centrifugation (40,000 × g, 4℃, 30 min) and the supernatant was loaded onto a 8-mL Ni-NTA agarose resin (Qiagen). After extensive wash with 10 column volumes of buffer A, His-SUMO tagged proteins were eluted with a 0-100% linear gradient of buffer B (Tris-HCl 50 mM pH 8.0, NaCl 300 mM, imidazole 500 mM, glycerol 10% vol/vol) over 10 column volumes.

Fractions of the main elution peak were pooled and incubated with the Ulp1 SUMO protease

fused to a poly-His tag at a 1:200 ratio (in molar concentration), for 1 h at room temperature and overnight at 4°C. During the overnight incubation with Up1, the protein mix (usually about 20 mL) was dialyzed against 2 L of buffer C (Tris-HCl 50 mM pH 8.0, NaCl 300 mM, glycerol 10% vol/vol) using a porous membrane tubing system. The dialyzed sample was loaded onto a 8-mL Ni-NTA agarose resin. The cleaved recombinant protein was recovered in the flow-through fractions while the uncleaved protein and the His_6-SUMO tag were retained by the resin. Elution was carried out with a 0-100% linear gradient of buffer B over 5 column volumes. Flow-through fractions were pooled and concentrated by centrifugation (4,000 × g, 4°C) to 0.5-2 mL using an Amicon Ultra Centrifugal Unit (Millipore) with a 10-kDa cutoff. 500-µL protein samples were purified by size-exclusion chromatography (SEC) using an ENrich™ SEC650 10x300 column (Biorad) in buffer D (Tris-HCl 25 mM pH 8.0, NaCl 150 mM). The main elution fractions were pooled and concentrated with a 10-kDa Amicon Ultra Centrifugal filter. Finally, the protein concentration was measured through its absorbance at 280 nm and 25-µL aliquots were stored at -80℃.

The theoretical molecular weight, PI and extinction coefficient were predicted using the ProtParam tool from the ExPASy bioinformatics resource portal (https://web.expasy.org/protparam/).

E. Screening of detergents for solubilization and purification of membrane proteins

Detergent screening was performed in collaboration with Anne-Marie Villard (RoBioMol platform, hosted by the Pneumococcus Group at the IBS) using a Microlab STAR Hamilton automation robot.

Cells from 2-L cultures were lysed as described in the previous section. Following low-speed centrifugation, the supernatant was further ultra-centrifuged at 250,000 × g, 4℃ for 1 h. The membrane pellet was resuspended in 1/1000th volume (2 mL for membranes prepared from 2-L cultures) of buffer A containing a detergent at the appropriate concentration (see Table 3). Thirteen different detergents were screened : 3-[(3-cholamidopropyl)dimethylammonio]-2-

hydroxy-1-propanesulfonate (CHAPS); 5-Cyclohexylpentyl β-D-maltoside (cymal5); n-dodecylphosphocholine (FC12); Octyl-beta-Glucoside (β-OG); Lauryldimethylamine oxide (LDAO); 3-Dodecylamido-N,N′-dimethylpropyl amine oxide (LAPAO); Octaethylene Glycol Monododecyl Ether (C12E8); N-Dodecanoyl-N-methylglycine sodium salt (NLS); Triton X-100; n-Decyl-β-D-Maltopyranoside (DM); n-Dodecyl β-D-maltoside (DDM); n-Undecyl β-maltoside (UDM) and Decyl Maltose Neopentyl Glycol (DMNG).

Detergent solubilization was carried out overnight on a rotating wheel placed at 4°C. The detergent-solubilized membrane protein samples were loaded onto 50-μL nickel-affinity resins, washed and eluted with buffer A and B containing the appropriate detergent concentration (see Table 3). The eluted samples were concentrated using 10-kDa Amicon Ultra Centrifugal filter centrifuged at 1,000 × g, 4℃ and analyzed by negative-stain EM.

Table 3. Detergent screening conditions

Condition number	Detergent name	Detergent concentration for membrane solubilisation (1) (mM)	Detergent concentration for washing of the nickel resin (2)	Detergent concentration for elution from the nickel resin (2)
1	CHAPS	28 (CMC + 20 mM)	10 (CMC + 2 mM)	10 (CMC + 2 mM)
2	Cymal 5	22,4 (CMC + 20 mM)	4.4 (CMC + 2 mM)	4.4 (CMC + 2 mM)
3	FC12	21,5 (CMC + 20mM)	3.5 (CMC + 2 mM)	3.5 (CMC + 2 mM)
4	B-OG	23 (CMC + 5 mM)	20 (CMC + 2 mM)	20 (CMC + 2 mM)
5	LDAO	21 (CMC + 20 mM)	3 (CMC + 2 mM)	3 (CMC + 2 mM)
6	LAPAO	21.6 (CMC + 20 mM)	3.6 (CMC + 2 mM)	3.6 (CMC + 2 mM)
7	C12E8	20.1 (CMC + 20 mM)	2.1 (CMC + 2 mM)	2.1 (CMC + 2 mM)
8	NLS	34.6 (CMC + 20 mM)	16.6 (CMC + 2 mM)	16.6 (CMC + 2 mM)
9	TX100	20.2 (CMC + 20 mM)	2.2 (CMC + 2 mM)	2.2 (CMC + 2 mM)
10	DM	21.8 (CMC + 20 mM)	3.8 (CMC + 2 mM)	3.8 (CMC + 2 mM)
11	UDM	20.6 (CMC + 20 mM)	2.6 (CMC + 2 mM)	2.6 (CMC + 2 mM)
12	DMNG	0.034% (CMC x 10)	2 (CMC + 2 mM)	2 (CMC + 2 mM)
13	DDM / DDM	20.17 (CMC + 20 mM)	2.2 (CMC + 2 mM)	2.2 (CMC + 2 mM)
14	DDM / UDM	.17 (CMC + 20 mM)	2.6 (CMC + 2 mM)	2.6 (CMC + 2 mM)
15	DDM / DMNG	20.17 (CMC + 20 mM)	2 (CMC + 2 mM)	2 (CMC + 2 mM)
16	DDM / DDM	20.17 (CMC + 20 mM)	2.2 (CMC + 2 mM)	2.2 (CMC + 2 mM)
17	DDM / UDM	20.17 (CMC + 20 mM)	2.6 (CMC + 2 mM)	2.6 (CMC + 2 mM)
18	DDM / DMNG	20.17 (CMC + 20 mM)	2 (CMC + 2 mM)	2 (CMC + 2 mM)

CMC, critical micelle concentration

F. Large-scale purification of membrane recombinant proteins

Following cell lysis, the low-speed centrifugation supernatant was further ultra-centrifuged

at 250,000 × g, 4℃ for 1 h. The membrane pellet was resuspended in 1/1000th of buffer A containing 10 mM of the nonionic detergent DDM (2 mL of buffer A-DDM for membranes prepared from 2-L cultures). Detergent solubilization was carried out overnight on a rotating wheel placed at 4°C. After ultra-centrifugation (250,000 × g, 4℃, 1 h), the supernatant containing the DDM-solubilized membrane protein was used for the two-step Ni-NTA purification as described in the previous section, except that washing, elution and dialysis buffers contained 0.4 mM DDM. Cleavage of the His_6-SUMO tag was performed with a 1:100 ratio of Ulp1 protease. Before size-exclusion chromatography, the membrane protein was concentrated using 10-kDa Amicon Ultra Centrifugal filter centrifuged at 1,000 × g, 4℃. The concentrated sample was then injected onto an ENrich™ superdex 200 equilibrated with buffer D supplemented with 0.4 mM DDM. The elution fractions were concentrated before analysis by negative-stain EM.

G. SEC-MALLS analysis

The SEC-MALLS analyses were performed in collaboration with Caroline Mas, who is responsible for the biophysical platform of the Grenoble Instruct-ERIC center (ISBG, PSB - Grenoble Partnership for Structural Biology). The ENrich™ SEC650 10x300 column (Biorad) was equilibrated with buffer D for 10 column volumes before the experiment. A 50-µL protein sample was loaded onto an ENrich™ SEC650 10x300 column (Biorad) in buffer D at 0.5 mL/min, and analyzed by a Multi-Angle static Laser Light Scattering (Wyatt Dawn Heleos II) and a UV Detector (Hitachi Elite LaChrom L-2400). Accurate weight-averaged molar masses (MW) determination was performed with the Astra software (Wyatt Technologies).

H. Negative-stain EM analysis

Negative-stain EM analyses were performed in collaboration with Emmanuelle Neumann and Daphna Fenel from the Electron Microscopy and Methods group at the IBS. The

concentration of the protein sample was adjusted to 0.05 mg/mL before negative-stain Mica-carbon Flotation Technique (MFT) was used to prepare samples. In brief, samples were absorbed on the clean side of a carbon film on mica, stained with 2% (wt/vol) uranyl acetate. Samples were then transferred to a 400-mesh copper grid, which was subsequently air-dried. Images were taken under low dose conditions (< 10 $e^{-}/Å^{2}$) with defocus values between 1.2 and 2.5 µm on a Technai 12 FEI LaB6 electron microscope at 120 kV accelerating voltage. Image acquisition was performed with calibrated nominal magnifications ranging from 1,000 to 30,000, using a CCD Gatan ORIUS SC1000 camera (Gatan, Inc., Pleasanton, CA).

I. Mass spectrometry analysis

Liquid Chromatography Electrospray Ionization Mass Spectrometry (LC/ESI-MS) was used to assess potential proteolytic degradation of the recombinant proteins. The equipment included a 6210 LC/ESI-TOF mass spectrometer interfaced with an HPLC binary pump system (Agilent Technologies). Mass spectra were recorded in the 300-3200 *m/z* range, in the positive ion mode with spectra in the profile mode. MS spectra were acquired and data were processed with the MassHunter workstation software (v. B.02.00, Agilent Technologies) and with the GPMAW software (v. 7.00b2, Lighthouse Data, Denmark).

Just before analysis, protein samples were diluted into acidic denaturing conditions to a final concentration of 5 µM with solution E (0.03% TFA in water). Samples were cooled to 10°C in the autosampler and the analysis was run by injecting 4 µL of each sample. Samples were first trapped and desalted on a reverse phase-C8 cartridge (Zorbax 300SB-C8, 5 µm, 300 µm ID×5 mm, Agilent Technologies) for 3 min at a flow rate of 50 µL/min with 100% solution E and then eluted with 70% solvent F (95% acetonitrile-5% water-0.03% TFA) at flow rate of 50 µL/min for MS detection.

J. Microscale thermophoresis

Binding experiments were carried out with a Monolith NT.115 Series instrument (Nano Temper Technologies GMBH). GerM constructs were labeled with the red dye NT-647. Four μL of sample containing 100 nM of labeled GerM and increasing concentrations of AG, AH and/or Q (from 7 nM to 235 μM) or BSA (negative control, from 5 nM to 360 μM) were loaded on K003 Monolith NT.115 hydrophilic treated silicon capillaries and thermophoresis was measured for 30 sec. Each measurement was made in triplicates. Experiments were carried out at 25°C in MST optimized buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM MgCl2, 0.05% Tween-20). Analysis was performed with the Monolith software. Affinity KD was quantified by analyzing the change in normalized fluorescence (Fnorm = fluorescence after thermophoresis/initial fluorescence) as a function of the concentration of the protein partner.

K. Protein crystallization and X-ray diffraction data collection

A freshly purified $YhcN_{24\text{-}189}$ sample concentrated to 40 mg/mL in buffer D was used to screen initial crystallization conditions using a Cartesian PixSys 4200 nano-crystallization robot available through the High-Throughput Crystallization platform (HTX lab, EMBL - European Molecular Biology Laboratory, Grenoble). Three sitting drops containing 100 nL of protein (at 40, 20 or 10 mg/mL) and 100 nL of reservoir solution (from commercial kits) were set up in 96-well Crystal Quick plates (Greiner) and incubated at 20°C. Several crystal hits were then reproduced and refined manually through the hanging drop method. Three drops containing 1 μL of protein (at 10, 20 or 40 mg/mL) and 1 μL of reservoir solution were set up onto a siliconized glass cover slide (Hampton Research) hanging on top of 24-well Crystal Quick plates containing 500 μL reservoir solution in each well. (Hanging drop crystallization conditions are shown in Table 4). The plates were incubated at 20℃ and needle-shaped crystals were obtained in 0.1 M HEPES pH 7.5, 20% (w/vol) PEG 8000. Crystals appeared within 7 days and kept growing until around 15 days to reach dimensions of about 20 x 20 x 100 μm. Crystals were soaked briefly (a few seconds) into a cryo-protecting solution containing 0.1 M

HEPES pH 7.5, 21% (w/vol) PEG 8000, 15% (vol/vol) glycerol or 15% (vol/vol) ethylene glycol before flash freezing into liquid nitrogen. X-ray diffraction data were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble), on the ID30A-3 (MASSIF-3) beamline. The experimental beamline parameter settings and initial acquired image quality assessment were performed with MxCuBE2 (Gabadinho et al. 2010). Statistics on data collection are summarized in Table 5 in the *Results* section.

Table 4. Hanging drop crystallization screening conditions.

HEPES 0. 1 M pH 7.0 PEG 8k 20% H_2O	HEPES 0.1 M pH 7.2 PEG 8k 20% H_2O	HEPES 0.1 M pH 7.5 PEG 8k 20% H_2O	HEPES 0.1 M pH 7.6 PEG 8k 20% H_2O	HEPES 0.1 M pH 7.8 PEG 8k 20% H_2O	
HEPES 0.1 M pH 7.5 PEG 8k 16% H_2O	HEPES 0.1 M pH 7.5 PEG 8k 18% H_2O	HEPES 0.1 M pH 7.5 PEG 8k 20% H_2O	HEPES 0.1 M pH 7.5 PEG 8k 22% H_2O	HEPES 0.1 M pH 7.5 PEG 8k 24% H_2O	
MES 0.1 M pH 6.0 PEG 10k 20% H_2O	MES 0.1 M pH 6.2 PEG 10k 20% H_2O	MES 0.1 M pH 6.4 PEG 10k 20% H_2O	MES 0.1 M pH 6.6 PEG 10k 20% H_2O	MES 0.1 M pH 6.7 PEG 10k 20% H_2O	
MES 0.1 M pH 6.5 PEG 10k 16% H_2O	MES 0.1 M pH 6.5 PEG 10k 18% H_2O	MES 0.1 M pH 6.5 PEG 10k 20% H_2O	MES 0.1 M pH 6.5 PEG 10k 22% H_2O	MES 0.1 M pH 6.5 PEG 10k 24% H_2O	
MES 0.1 M pH 6.0 PEG 20k 15% H_2O	MES 0.1 M pH 6.2 PEG 20k 15% H_2O	MES 0.1 M pH 6.4 PEG 20k 15% H_2O	MES 0.1 M pH 6.6 PEG 20k 15% H_2O	MES 0.1 M pH 6.7 PEG 20k 15% H_2O	
MES 0.1 M pH 6.5 PEG 20k 10% H_2O	MES 0.1 M pH 6.5 PEG 20k 13% H_2O	MES 0.1 M pH 6.5 PEG 20k 15% H_2O	MES 0.1 M pH 6.5 PEG 20k 17% H_2O	MES 0.1 M pH 6.5 PEG 20k 20% H_2O	
Na/K phosphate 0.8 M pH 5.0 PEG 3350 20% H_2O	Na/K phosphate 0.8 M pH 5.6 PEG 3350 20% H_2O	Na/K phosphate 0.8 M pH 6.3 PEG 3350 20% H_2O	Na/K phosphate 0.8 M pH 6.9 PEG 3350 20% H_2O	Na/K phosphate 0.8 M pH 7.5 PEG 3350 20% H_2O	Na/K phosphate 0.8 M pH 8.2 PEG 3350 20% H_2O

In the last screening lane, the Na/K phosphate buffers were from the Quick Screen kit (Hampton).

L. Structure determination and refinement

X-ray diffraction data were collected on beamline ID30-A3 of the European Synchrotron for Radiation Facility in Grenoble. The mxCube2 interface was used to set up the experimental parameters, including the data collection strategy. Data were collected at 0.9677 Å over 250°, with oscillation angles of 0.1°. The reflection data were indexed and scaled using the X-ray Detector Software (XDS) (Kabsch 2010). $YhcN_{24\text{-}189}$ crystals belong to the orthorhombic space group $P2_12_12_1$, with unit cell dimensions of 31.38 x 52.12 x 133.72 Å and two molecules per asymmetric unit. *Ab initio* phase determination was achieved using the ARCIMBOLDO_LITE program (Sammito et al. 2013), which is based on the location of predicted α-helices with PHASER (McCoy et al. 2007) and density modification with SHELXE (Thorn and Sheldrick 2013). Based on secondary structure predictions performed with the JPRED4 server (http://www.compbio.dundee.ac.uk/jpred/), we searched for two α-helices containing 13 or 12 residues, using machines of 4 cores or 8 cores, respectively. The high quality of the electron density map obtained with this strategy allowed automated building of a $YhcN_{24\text{-}189}$ model from residue 79 to187 using Phenix (Terwilliger et al. 2008). The initial model was completed and refined by cycles of manual building using Crystallographic Object-Oriented Toolkit (COOT) (Emsley and Cowtan 2004), addition of water molecules with ARP/wARP (Langer et al. 2008) and refinement with REFMAC5 (Murshudov et al. 2011). Stereochemical verification was performed with PROCHECK (Laskowski et al. 1993) and secondary structure assignment was verified with DSSP (Kabsch and Sander 1983). Figures were generated with PymMol (http://www.pymol.org) or Chimera (Pettersen et al. 2004). Statistics on data refinement are summarized in Table 5 in the *Results* section.

Result

Result

A. Biophysical and structural characterization of GerM

1. Context and personal contribution to the GerM study.

GerM is a lipoprotein initially identified by a genetic screen as a protein involved in the germination of the spore (Rigden and Galperin 2008). However, a more recent study performed by the group of David Rudner (Harvard Med. School, Boston) in *B. subtilis* has shown that it has an earlier role in the development of the spore (Rodrigues et al. 2013, Rodrigues, Ramirez-Guadiana, et al. 2016). Like the A-Q proteins, GerM is required for the activation of σG and for the integrity of the forespore. In addition, it is involved in the localization of the A-Q complex. The study of GerM in my hosting laboratory started when the first author of these studies (Christopher Rodrigues) started his own group at the University of Sydney, Australia.

Q is anchored in the forespore membrane and AH is anchored in the mother cell membrane. Because the soluble domains of Q and AH interact directly in the intermembrane space (Doan, Marquis, and Rudner 2005, Camp and Losick 2008), this interaction drives the localization of AH and Q at the interface between the forespore and the mother cell (Fig. 21A). The AH-Q dimer was shown to act as a molecular zipper that tethers the outer and inner forespore membranes during engulfment, and Q as a key player for the localization of many mother-cell proteins (Doan, Marquis, and Rudner 2005, Broder and Pogliano 2006). In the absence of Q, AH delocalizes to the whole mother-cell membrane while in the absence of AH, Q remains localized at the forespore-mother cell interface (Rodrigues et al. 2013). Using fluorescence localization experiments in diverse genetic backgrounds, GerM was shown to participate to Q localization. In the presence of AH, the deletion of GerM does not have any impact on Q localization (Fig. 21A). By contrast in the absence of AH, GerM is required for the localization of Q around the forespore (Fig. 21A) (Rodrigues, Ramirez-Guadiana, et al. 2016). In the double *ΔgerM ΔAH* mutant, Q displays a diffuse localization in the whole inner forespore membrane. This observation suggests that GerM interacts directly or indirectly with Q. On the other hand, AH requires Q for its localization but does not require GerM. GerM localization requires Q but

does not require AH (Fig. 21B). These observations suggest that Q, AH and GerM form a tripartite complex, in with AH and GerM both contribute to Q localization and Q anchors both AH and GerM at the mother cell-forespore interface. Since Q and AH are core members of the A-Q complex, GerM is thus considered to be part of the A-Q complex, through direct or indirect interaction with Q.

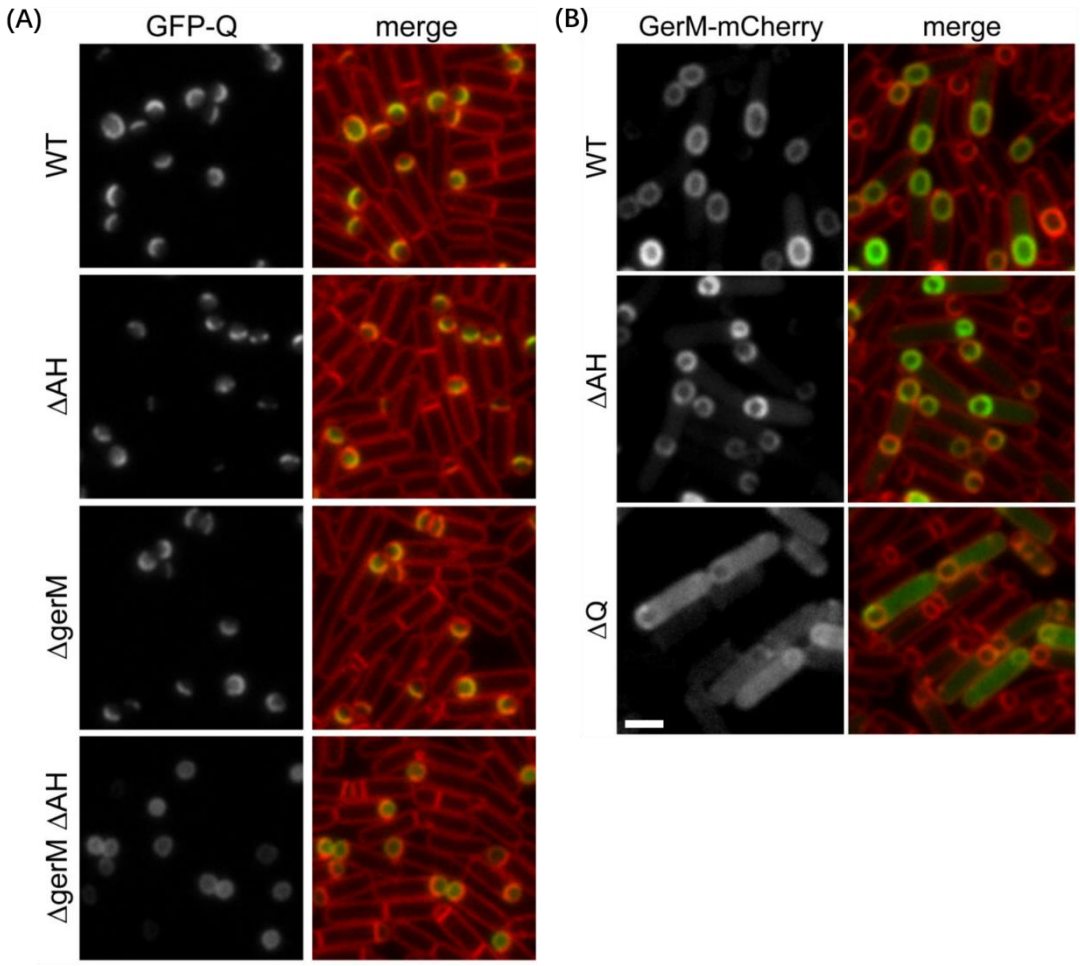


Figure 21. Interdependent localizations of SpoIIQ, SpoIIIAH and GerM. (A) GFP-SpoIIQ (GFP-Q) localization requires GerM or SpoIIIAH (AH). Conventional fluorescence microscopy images of GFP-Q localization in sporulating cells at hour 2 of sporulation. Images are from wild-type (WT), *ΔspoIIIAH* (ΔAH), *ΔgerM* and *ΔspoIIIAH ΔgerM* (ΔgerM ΔAH) strains. **(B)** GerM-mCherry localizes to septal membrane in a manner that depends on Q. Conventional fluorescence microscopy images of GerM-mCherry localization in sporulating cells at hour 2.5 of sporulation. Images are from wild-type (WT), *ΔspoIIIAH* (ΔAH) and *ΔspoIIQ* (ΔQ) strains. Scale bar, 2 μm. The images are from Rodrigues et al. 2016a.

To understand how GerM contributes to Q localization, my hosting laboratory initiated a project in collaboration with the group of Christopher Rodrigues, aiming at solving the structure

of GerM and dissecting its interactions with AH and Q.

The project started before I joined the group, and was mainly accomplished by Jennyfer Trouve, who solved the structure of truncated and full-length forms of GerM from *B. subtilis*. GerM is composed of two GerMN domains (called GerMN1 and GerMN2), found in spore-formers and non sporogenous bacteria (Rigden and Galperin 2008). So far the function of these domains remains unknown. J. Trouve first solved the crystal structures of GerMN1 *ab initio* at the resolution of 1.0 Å with the help of C. Contreras-Martel from the Dessen group at the IBS. In addition, she solved the structure of the full-length soluble form of GerM, containing the tandem GerMN1-GerMN2 domains, at 2.1 Å using the molecular replacement method. The structure of these domains revealed an unexpected structural similarity between GerMN domains and ring-building motifs (RBMs) found in ring-forming proteins of specialized secretion systems (EscJ/PrgK) and in SpoIIIA proteins (AF, AG and AH) (Fig. 22B-D). More precisely, the core of GerMN domains display the αββαβ RBM arrangement ($\alpha_1\beta_3\beta_4\alpha_2\beta_5$) but GerMN domains contain two additional, N-terminal, anti-parallel β-strands ($\beta_1\beta_2$) that fold against the $\alpha_1\beta_3\beta_4\alpha_2\beta_5$ core in the crystal form (Fig. 22A)

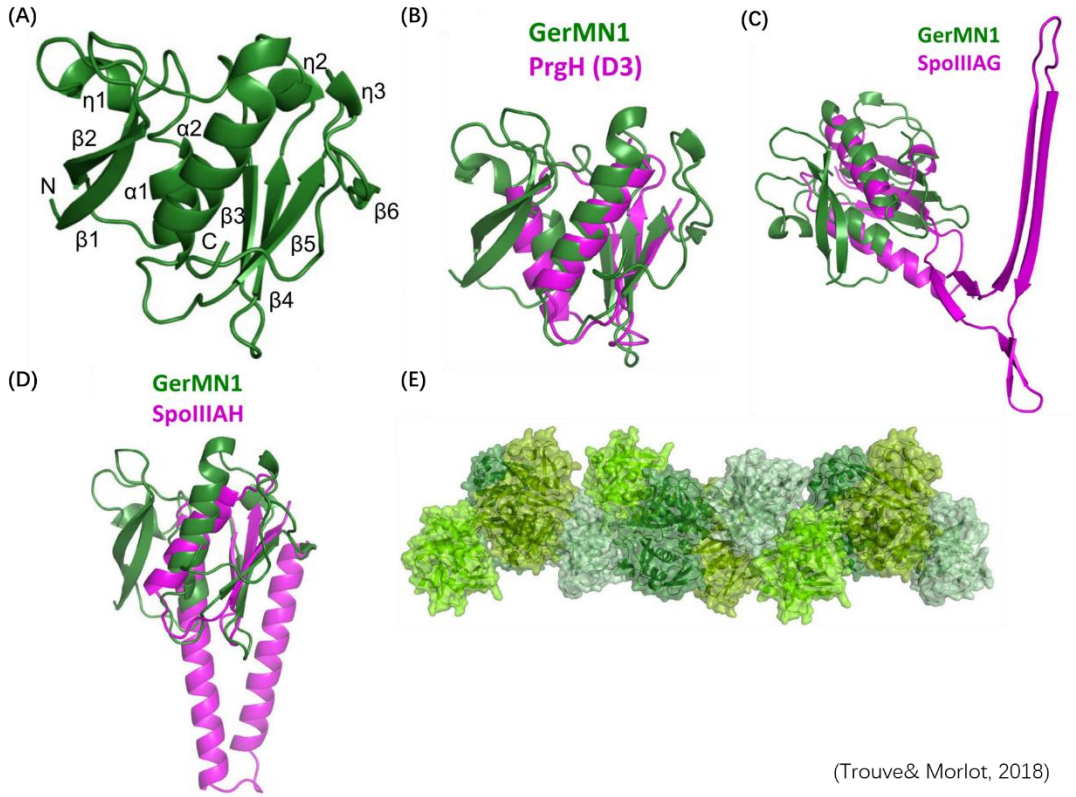


Figure 22. The crystal structure of GerM. (A) Ribbon representation of the GerMN1 domain of GerM. α-helices, β-strands, 3_{10}-helices (η) are marked. **(B)** Structural overlay of GerMN1 and the second RBM domain (D3) of PrgH from *S. typhimurium* (PDB code 4G11). **(C)** Structural overlay of GerMN1 and AG from *B. subtilis* (PDB code 5WC3). **(D)** Structural overlay of GerMN1 and AH from *B. subtilis* (PDB code 3UZ0). (**E)** The $GerM_{26\text{-}366}$ protofilament resulting from the continued protein oligomerization in the crystal. The four molecules in the asymmetric unit are colored in different shades of green and are shown with surface and ribbon representations. This figure is from Trouve et al., 2018.

In terms of oligomerization, crystals of the GerMN1-GerMN2 construct ($GerM_{26\text{-}366}$) contain a fiber resulting from contacts between crystallographic symmetries of the $GerM_{26\text{-}366}$ tetramer present in the asymmetric unit (Fig. 22E). However, the interface between $GerM_{26\text{-}366}$ protomers is mainly established by residues carried by loops (see the *Discussion* section) and is thus completely different from interfaces observed between protomers in RBM rings. In addition, GerM does not form any oligomer in solution. Interestingly however, it seems that the two additional β-strands ($\beta_1\beta_2$) display some flexibility in solution. Even though the oligomerization of GerM *in vitro* and *in vivo* could not be demonstrated, it is thus possible that in the presence of its partners, strands β_1 and β_2 move away from the RBM core to allow assembly of a GerM ring. One additional hypothesis is that assembly of the GerM ring is promoted and/or stabilized

by its Q and AH partners, or other partners from the A-Q complex.

At the beginning of my thesis, my objective in this project was to test the direct interaction of GerM with the AH and Q components of the A-Q complex. This study has not been conclusive enough to be included in the article about GerM structure (see section I.C of the *Results* chapter) but I have described my results in section I.B. In addition, I produced and purified GerM samples for SAXS (Small-Angle X-ray Scattering) analyses, which are reported in the Trouve et al. article. My authorship in the article from Trouve et al. thus results from my contribution to the biophysical study of GerM (test of GerM interaction with protein partners and preparation of SAXS samples).

2. Biophysical study of the interaction between GerM and other A-Q proteins

The functional interactions between GerM, AH and Q (co-dependent localizations) indicate that GerM is part of the basal platform of the A-Q complex (Rodrigues, Ramirez-Guadiana, et al. 2016). The core components of this platform include AG, which forms large homo-oligomeric rings (Rodrigues, Henry, et al. 2016), as well as AH and Q, which establish a ratchet-like interaction between the outer and inner forespore membranes (Broder and Pogliano 2006). To investigate whether GerM interacts directly with core components of this platform, (AG, AH and Q), and describe this interaction at the molecular level, I produced and purified the soluble domains of those proteins (Fig. 23).

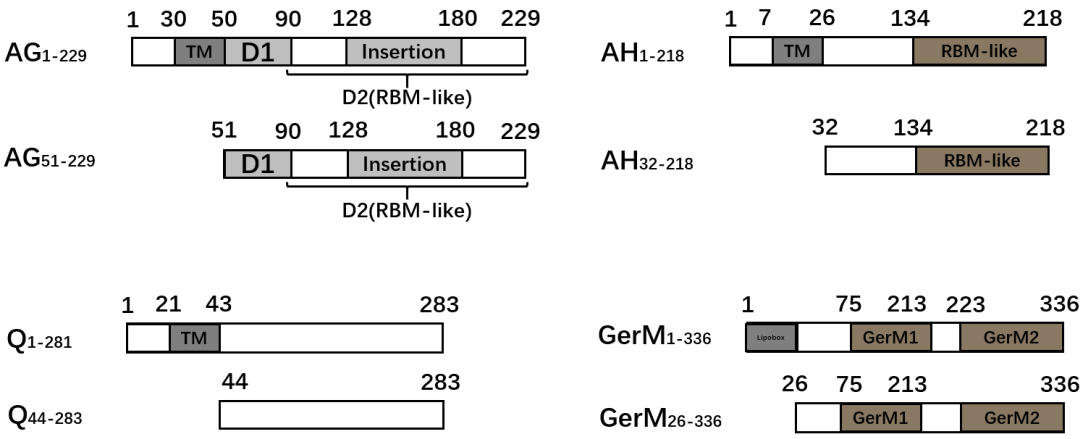


Figure 23. Topology of SpoIIIAG (AG), SpoIIIAH (AH), SpoIIQ (Q) and GerM, and constructs purified for this study. The constructs purified for the protein-protein interaction assays are shown below the schematics describing the domain boundaries of the A-Q proteins. Residue numbering is for the A-Q proteins from *B. subtilis* 168. Transmembrane domains (TM), RBM-like domains, as well as the D1 and D2 domains of AG, and the insertion region within the D2 domain (see section IIC6 of the *Introduction* chapter) are labeled.

Protein-protein interactions assays included microscale thermophoresis (MST), size-exclusion chromatography, His-tag pull downs and isothermal titration calorimetry (ITC).

a. Microscale thermophoresis experiments

To test the interaction of GerM with AH, Q and AG in solution, I labeled GerM and titrated it with increasing concentrations of AH, Q or AG (see the *Material and Methods* chapter). Interestingly, we could detect a direct interaction between GerM and Q (K_D = 1.2 ± 0.2 µM) and between GerM and AH (K_D = 0.17 ± 0.02 µM) using MST (Fig. 24A).

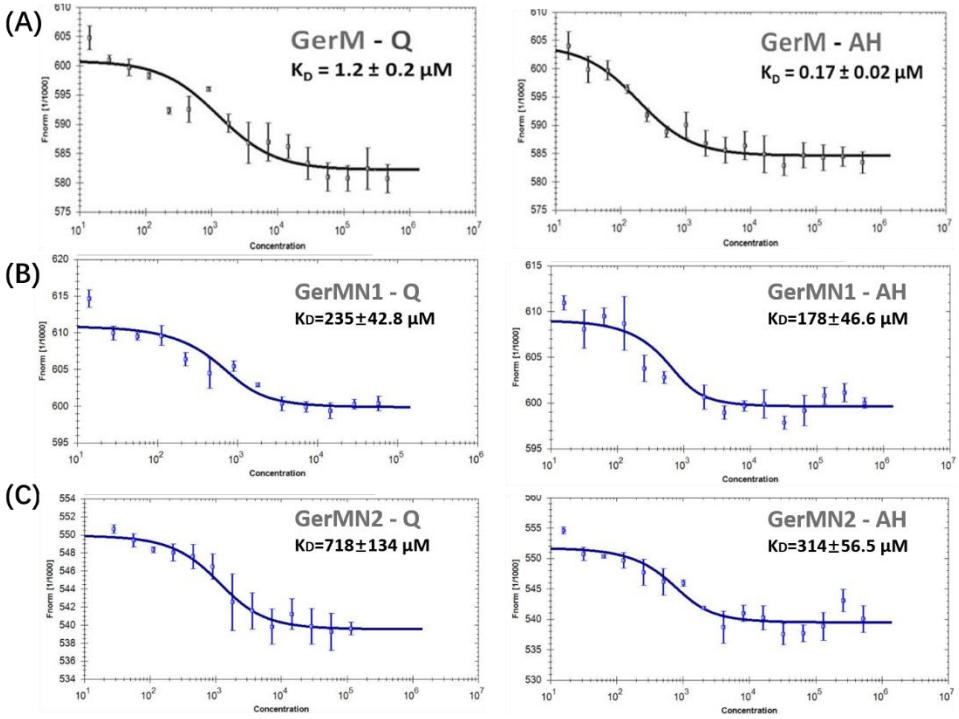


Figure 24. Affinity measurements by microscale thermophoresis (MST) of GerM against other proteins from A-Q complex. (A) Labeled GerM binds to increasing concentrations of Q (left panel) or AH (right panel). Fnorm (normalized fluorescence = fluorescence after thermophoresis/initial fluorescence) is plotted against Q or AH concentrations. Measures are represented by dots and the fitted curve by a line. **(B)** Labeled GerMN1 binds weakly to increasing concentrations of Q and AH. **(C)** Labeled GerMN1 binds weakly to increasing concentrations of Q and AH.

In order to determine which GerMN domain interacts with AH and/or Q, I repeated those tests with purified recombinant constructs of GerMN1 and GerMN2. Both domains interacted with AH and Q although with a weaker binding constant than the full-length GerMN1-GerMN2 soluble domain (Fig. 24B-C). This result suggests that both domains contribute to the interaction of GerM with Q and AH. To make sure that these interactions were specific, I sought to get an internal MST control experiment.

We first tested whether GerM interacts with the BSA (Bovine Serum Albumin), a protein that is unrelated to the A-Q complex. GerM did not bind the BSA, suggesting that its interaction with AH and/or Q is specific. The best control for a protein-protein assay is however to identify a mutation that disrupts the interaction. So far, two mutations in Q were reported to disrupt its interaction with GerM. Q has evolved from an endopeptidase called LytM, which cleaves peptide bonds in *Staphylococcus aureus* peptidoglycan (Meisner et al., 2012; Levdikov et al.,

2012). One of the catalytic residues has been lost in Q and the protein displays neither hydrolase activity, nor affinity for the peptidoglycan. Intriguingly, two positions located in the degenerated active site of Q (Q168 and D123) were shown to be required for Q localization in the absence of AH : GFP-Q(Q168A) and GFP-Q(D123A) constructs fail to localize at the forespore-mother cell interface in a *ΔspoIIIAH* background (Rodrigues, Ramirez-Guadiana, et al. 2016). So in other words the Q(Q168A) and Q(D123A) mutations phenocopy the deletion of GerM. Since GerM was shown to be sufficient to localize Q in the absence of AH, this result suggests that positions Q168 and D123 are important for the interaction between Q and GerM.

When MST experiments were performed with GerM and the Q(Q168A) mutant, the two proteins kept interacting, with a K_D of 15.4 ± 4.34 µM (data not shown). When we introduced a charged and bulkier residue at position Q168 (mutation Q168R) or introduced a Q(D123R) mutation, the affinity for Q was only decreased by about ten fold (K_D of 74.9 ± 9.9 µM for Q(Q168R) and 85.6 ± 13.4 µM for Q(D123R)). If the direct interaction detected between GerM and Q by MST is specific, then these results indicate that they do not strongly depend on positions Q168 and D123. Because this experiment failed to validate the specificity of the GerM-Q interaction, we thus sought to demonstrate this interaction using other approaches.

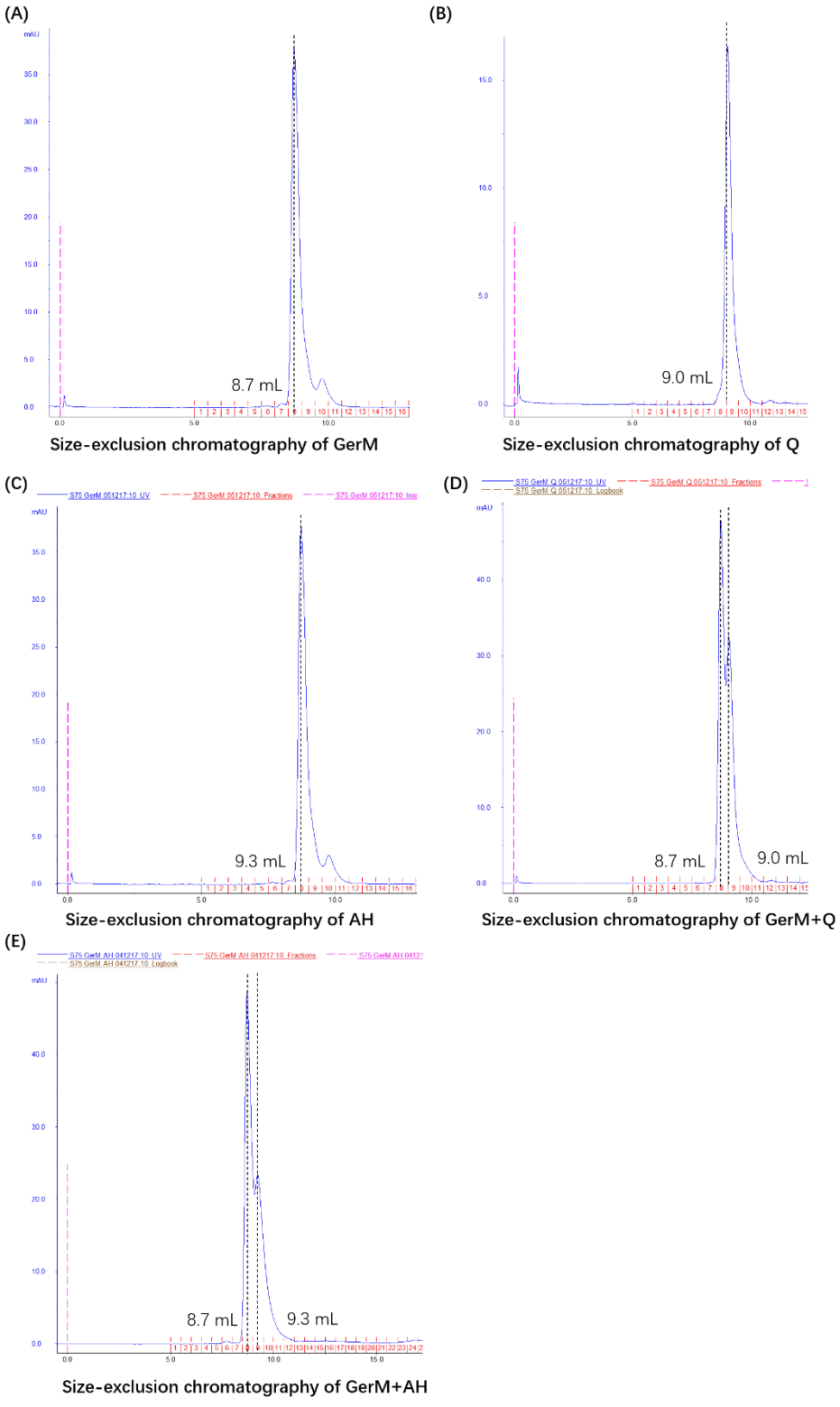


Figure 25. (A-C) Size-exclusion chromatography (SEC) profiles of individual proteins show the elution of GerM (**A**) 8.7 mL, Q (**B**) at 9.0 mL and AH (**C**) at 9.3 mL. **(D)** The SEC profile of a pre-incubated mix containing GerM and Q shows two peaks corresponding to the separated elution of GerM and Q at 8.7 and 9.0 mL, respectively. **(E)** The SEC profile of a pre-incubated mix containing GerM and AH proteins shows two peaks corresponding to the separated elution of GerM and AH at 8.7 and 9.3 mL, respectively. UV absorbance chromatogram is in blue, black dotted lines mark the elution volumes.

b. Alternative GerM-Q interaction assays

To detect a direct interaction between GerM and Q or GerM and AH, we used three other methods. First, we performed a size-exclusion chromatography on a sample containing a concentrated mix of GerM and Q or AH, with each protein at about 1 mM in a low-salt buffer (Tris-HCl 25 mM pH 8.0, NaCl 150 mM). In each experiment, the proteins (GerM and Q or GerM and AH) eluted at two different volumes, indicating that no GerM-Q complex was present in solution (Fig. 25).

We also performed a dot-blot assay in which different proteins (Q, AH, AG and unrelated proteins such as the BSA, StkP from *Streptococcus pneumoniae* or EnvC from *E. coli*) were loaded onto a PVDF membrane. Following saturation with milk proteins, the membrane was incubated with GerM overnight and after extensive washing, the presence of GerM was revealed using anti-GerM antibodies (Fig. 26). Both AH and Q, as well as AG, EnvC and StkP retained GerM on the membrane. This experiment thus indicates that GerM is able to bind surfaces coated with AH, Q or AG, but also surfaces coated with proteins unrelated to the A-Q complex.

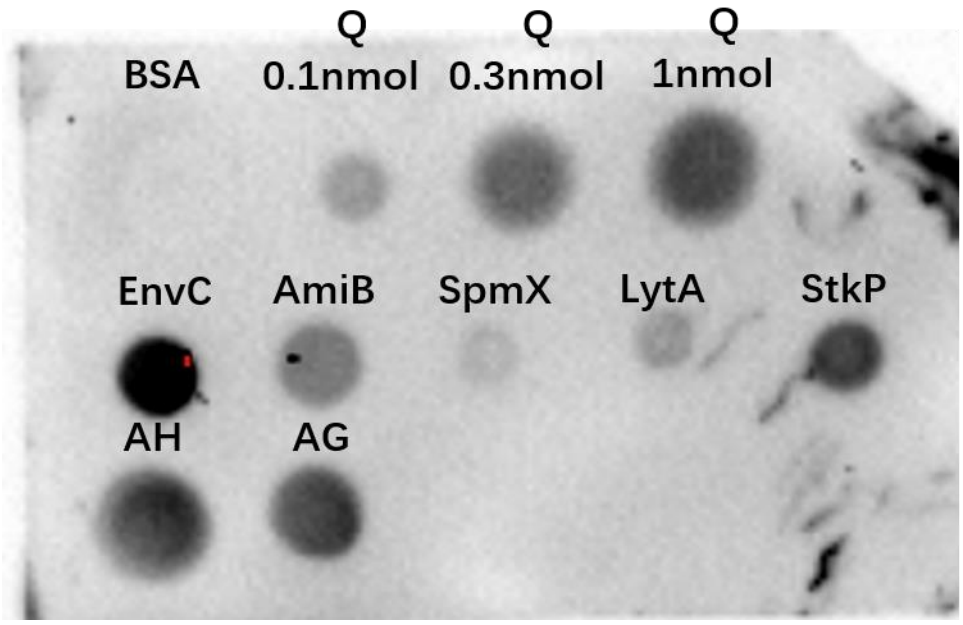


Figure 26. Dot-blot assay of GerM binding to different proteins. 0.3 nmol of purified proteins were loaded on a PVDF membrane, except for Q, for which 0.1, 0.3 and 1 nmol of protein were loaded. The PVDF membrane was then incubated with the purified GerM sample at 1 mM overnight at 4°C. There is no binding of BSA with GerM. GerM showed a strong binding with AH, AG and an increasing binding with an increasing concentration of Q. GerM also shows weak binding with unrelated proteins like AmiB and StkP, and strongest binding with EnvC. SpoIID and LytA also show weak binding with GerM.

We also performed His-tag pull-downs in which purified AH or Q were passed through a Ni-NTA resin saturated with His_6-GerM. Following extensive wash, neither AH nor Q were retained by the His_6-GerM construct.

Finally, we tried to detect GerM interaction with Q and AH using ITC. Even though very high concentrations of protein were used (> 1 mM), no binding could be detected using this technique.

In conclusion, although direct interaction was detected between GerM and Q and between GerM and AH using MST, these interactions remain to be supported by strong MST controls (mutations of GerM, AH and Q that totally disrupt the interaction) or by other protein-protein interaction assays.

The article on the structure of GerM needed to be published as soon as possible because we were aware of a dangerous competition with the group of Nathalie Strynadka in Vancouver, Canada. After two or three months of unconclusive trials to demonstrate the specificity of the GerM-Q and GerM-AH interaction, we thus decided to publish the structure of GerM without any protein-protein interaction assays. As of today, further biophysical evidence is still required to validate the existence of direct GerM-AH or GerM-Q complexes.

B. Structural characterization of the sporulation protein GerM from Bacillus subtilis. (published in 2018)

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Structural characterization of the sporulation protein GerM from *Bacillus subtilis*

Jennyfer Trouve[a], Ahmed Mohamed[b], Francisco Leisico[c], Carlos Contreras-Martel[a], Bowen Liu[a], Caroline Mas[a], David Z. Rudner[d], Christopher D.A. Rodrigues[b,*], Cecile Morlot[a,*]

[a] Institut de Biologie Structurale (IBS), Univ. Grenoble Alpes, CEA, CNRS, 38044 Grenoble, France
[b] The ithree institute, University of Technology, Sydney, Broadway NSW 2007, Australia
[c] UCIBIO, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal
[d] Department of Microbiology and Immunobiology, Harvard Medical School, Boston MA 02115, USA

ARTICLE INFO

Keywords:
Specialized secretion systems
Sporulation
Endospores
GerM
GerMN domain
SigG

ABSTRACT

The Gram-positive bacterium *Bacillus subtilis* responds to starvation by entering a morphological differentiation process leading to the formation of a highly resistant spore. Early in the sporulation process, the cell asymmetrically divides into a large compartment (the mother cell) and a smaller one (the forespore), which will maturate into a resistant spore. Proper development of the forespore requires the assembly of a multiprotein complex called the SpoIIIA-SpoIIQ complex or "A-Q complex". This complex involves the forespore protein SpoIIQ and eight mother cell proteins (SpoIIIAA to SpoIIIAH), many of which share structural similarities with components of specialized secretion systems and flagella found in Gram-negative bacteria. The assembly of the A-Q complex across the two membranes that separate the mother cell and forespore was recently shown to require GerM. GerM is a lipoprotein composed of two GerMN domains, a family of domains with unknown function. Here, we report X-ray crystallographic structures of the first GerMN domain of GerM at 1.0 Å resolution, and of the soluble domain of GerM (the tandem of GerMN domains) at 2.1 Å resolution. These structures reveal that GerMN domains can adopt distinct conformations and that the core of these domains display structural similarities with ring-building motifs found in components of specialized secretion system and in SpoIIIA proteins. This work provides an additional piece towards the structural characterization of the A-Q complex.

1. Introduction

To transport molecules (small molecules, proteins or DNA) across their cell envelope, bacteria assemble large multi-protein complexes called specialized secretion systems (Chandran, 2013; Costa et al., 2015; Portaliou et al., 2016). Depending on their function and the nature of the bacterial cell envelope (Gram-negative or Gram-positive), these nanomachines span one or two lipid bilayers. Double-membrane-spanning secretion systems include the type 1 secretion system (T1SS), T2SS, T3SS, T4SS and T6SS found in Gram-negative bacteria (Costa et al., 2015), as well as a new putative member: the SpoIIIA-SpoIIQ complex (called the A-Q complex henceforth) required for endospore development (Morlot and Rodrigues, 2018). Although most sporulating bacteria are Gram-positive and are thus enveloped by a single lipid bilayer, the developing endospore (the forespore) gets surrounded by two lipid bilayers, one derived from the forespore and the other one from the mother cell (Tan and Ramamurthi, 2014). The A-Q complex assembles in this double membrane around the forespore and is composed of proteins displaying similarities with core components of distinct secretion systems, suggesting that it might transport molecule(s) between the mother cell and forespore cytosols (Camp and Losick, 2008, 2009; Doan et al., 2009; Meisner et al., 2008). The most significant evidence for this model is the recent discovery that the A-Q complex contains an oligomeric ring (formed by SpoIIIAG) that resembles those found in T3SS and that might participate in a transenvelope conduit connecting the mother cell and forespore (Rodrigues et al., 2016a; Zeytuni et al., 2017). In addition, three other A-Q proteins (SpoIIIAB, SpoIIIAF and SpoIIIAH) were shown to display structural similarities with components of T2SS and T3SS (Levdikov et al., 2012; Meisner et al., 2012; Zeytuni et al., 2018a,b). Despite this recent progress, the assembly mechanism and global architecture of the complex remain mysterious (Morlot and Rodrigues, 2018). Here, we provide

* Corresponding authors.
E-mail addresses: christopher.rodrigues@uts.edu.au (C.D.A. Rodrigues), cecile.morlot@ibs.fr (C. Morlot).

https://doi.org/10.1016/j.jsb.2018.09.010
Received 5 July 2018; Received in revised form 21 September 2018; Accepted 24 September 2018
Available online 25 September 2018

structural insights into GerM, a protein recently shown to be involved in the assembly of the A-Q complex (Rodrigues et al., 2016b).

In response to nutrient depletion, sporulating bacteria enter a differentiation process called sporulation that leads to the development of dormant and stress-resistant cells called spores (Tan and Ramamurthi, 2014). The different steps in sporulation are governed by transcriptional programs based on the sequential activation of sporulation-specific sigma transcription factors (sig or σ) in the mother cell or in the forespore (Higgins and Dworkin, 2012; Hilbert and Piggot, 2004). After formation of an asymmetric septum, two cells of unequal size (a large one called the mother cell and a small one called the forespore) lie side by side. Remodeling of the septal cell wall then allows migration of the mother cell membrane around the forespore, a phagocytic-like process called engulfment. Ultimately, the forespore is released within the mother cell and is surrounded by two membranes (called the outer and inner forespore membranes) that define a periplasmic-like space called the intermembrane space. The A-Q complex assembles across the two forespore membranes during engulfment and is required for σG activation in the forespore, maintenance of the forespore physiological potential and spore development (Camp and Losick, 2009; Doan et al., 2009; Rodrigues et al., 2016b). This complex is composed of the forespore protein SpoIIQ and 8 mother cell proteins encoded by the *spoIIIA* operon (SpoIIIAA to SpoIIIAH, called AA to AH henceforth) (Guillot and Moran, 2007; Londoño-Vallejo et al., 1997). All the A-Q proteins are membrane proteins except AA, which is a soluble ATPase (Doan et al., 2009). Most of the SpoIIIA proteins share similarities to components of specialized secretion systems: AA resembles T4SS ATPases, AB shares a common fold with EpsF/GspF/PilC that tether the ATPase to the inner membrane complex in T2SS, the multi-transmembrane topology of AE is reminiscent of permease domains found in T1SS, and three other A-Q components (AF, AG and AH) possess ring-building motifs (RBMs) found in ring-forming proteins of T3SS, suggesting that the A-Q complex has been cobbled from different parts of well-defined secretion systems (Doan et al., 2009; Levdikov et al., 2012; Meisner et al., 2012; Rodrigues et al., 2016a; Zeytuni et al., 2017, 2018a,b). The hybrid nature of the A-Q complex might even extend beyond the family of secretion systems since AB also displays structural similarities with the component (the C-subunit) connecting cytosolic and membrane domains in bacterial V-ATPases, which are involved in aerobic ATP synthesis (Yokoyama et al., 1998; Zeytuni et al., 2018a).

Recently, a new protein named GerM has been implicated in the A-Q complex (Rodrigues et al., 2016b). This lipoprotein contains two GerMN (Germane) domains, a family of approximately 100 residue-long domains (Pfam10646) present in all major bacterial phyla, including *Firmicutes*, *Actinobacteria*, *Deinococcus-Thermus*, *Cyanobacteria*, *Spirochaetes* and *Proteobacteria* groups. Tandem copies (called GerMN1 and GerMN2 henceforth, Fig. 1A) of the GerMN domain are present in the sporulation and germination protein GerM found in the Firmicute phylum (Sammons et al., 1987; Slynn et al., 1994), while isolated GerMN domains are found in spore-formers and non-sporogenous Firmicutes, as well as in the other bacterial phyla listed above (Rigden and Galperin, 2008). The function of the GerMN domains is unknown so far. Similar to the *spoIIIA* and *spoIIQ* genes, deletion of the *gerM* gene results in collapsed forespores phenotype, deficient σG activity and failure to produce heat-resistant spores (Doan et al., 2009; Rodrigues et al., 2016b). In addition, GerM was shown to be required for AG localization around the forespore and to partially compensate for the absence of AH in the localization of Q, indicating that GerM is involved in the assembly of the A-Q complex and might potentially be part of it (Morlot and Rodrigues, 2018; Rodrigues et al., 2016b). The structure of GerM and its exact role in the assembly of the A-Q complex remain undetermined.

Here, we report the *ab initio* crystal structure of the first GerMN domain of GerM at 1 Å resolution and the structure of the soluble full-length region of GerM at 2.1 Å resolution. These structures provide evidence that GerMN domains can adopt different conformations, and reveal an interesting similarity between the core of the GerMN domains and RBMs found in ring-forming proteins from T3SS and from the A-Q complex. RBMs are structural arrangements composed of two α-helices folding against a three-stranded β-sheet; they establish homomeric interfaces required for the oligomerization of ring-forming proteins such as PrgK and PrgH from T3SS or AG from the A-Q complex (Bergeron et al., 2015; Rodrigues et al., 2016a; Schraidt and Marlovits, 2011; Worrall et al., 2016; Zeytuni et al., 2017). Interestingly, GerM was found to form an oligomer in the crystals but we could not provide evidence that GerM oligomerizes *in vivo* using structure-guided GerM mutants. We discuss these observations and propose hypotheses for the functions of GerM and GerMN domains.

2. Results and discussion

2.1. Ab initio crystallographic structure of the first GerMN domain of GerM

The purified construct encompassing the first GerMN domain of GerM (GerMN1, residues Thr76 to Glu213) formed large needle-shaped crystals which remarkably diffracted to a resolution of 1.0 Å (Fig. S1A and Table 1). We first tried to solve this structure by molecular replacement using the coordinates of a single GerMN domain available in the Protein Data Bank (PDB entry 5J7R). This GerMN domain is the sole folded region (residues 44–182) of a 184 residue-long uncharacterized lipoprotein (Uniprot ID A0A0H2YS46, called $GerMN_{Cp}$ henceforth) from *Clostridium perfringens*. As expected given the low sequence identity (20% identity over 165 residues) between $GerMN_{Cp}$ and GerMN1, molecular replacement was unsuccessful. We thus took advantage of the particularly high diffraction power of the GerMN1 crystals to determine *ab initio* phases using the ARCIMBOLDO_LITE program (http://chango.ibmb.csic.es/) (Sammito et al., 2013), which combines the location of model fragments like small α-helices with Phaser (McCoy et al., 2007) and density modification with SHELXE (Thorn and Sheldrick, 2013) (see the *Methods* section). This strategy yielded an electron density map from the correct positioning of two helices and the high quality of this map allowed building of the model of the GerMN1 domain from Thr76 until Asp208, present as a single copy in the asymmetric unit. The core of GerMN1 contains two helices (α1 and α2) sandwiched between a three-stranded β-sheet (β3β4β5) and a twisted two-stranded antiparallel β-sheet (β1β2) (Fig. 1B).

A search for structural homologues using the DALI server (Holm and Rosenström, 2010) confirmed the close similarity of GerMN1 with $GerMN_{Cp}$ from *C. perfringens* (PDB entry 5J7R, rmsd (root mean square deviation) of 2.2 Å over 143 residues, Z-score = 12.0). The major difference between the two structures is the "open" versus "closed" conformation of the β1β2 sheet in $GerMN_{Cp}$ and GerMN1, respectively. In $GerMN_{Cp}$, the β1β2 sheet positions away from the globular core of the protein, while in GerMN1 the β1β2 sheet is packed against helices α1 and α2 (Fig. 1C). However, analysis of the asymmetric unit content in the $GerMN_{Cp}$ crystal shows that the β1β2 sheet of one molecule packs against the globular core of the neighboring molecule, mimicking the intramolecular packing of β1β2 in GerMN1 (Fig. S1B) and involving many interface residues equivalent in position and nature (Fig. S2A). Altogether, these observations suggest that the two conformations might exist in GerMN domains. In an attempt to detect the "open" conformation in GerMN1, we characterized its soluble form using SAXS (Small Angle X-ray Scattering).

The GerMN1 experimental scattering profile (Fig. 1D) presents linear ranges in the Guinier plot with *Rg* of 1.46 nm, pointing to a homogenous sample in solution composed of GerMN1 monomers with a $MM_{I(0)}$ of 13.3 kDa. In order to compare the GerMN1 structure in solution and in the crystal, the low-resolution SAXS envelope of GerMN1 in solution was generated using DAMMIF, and the scattering curve from GerMN1 crystal structure was computed using the CRYSOL software. Both experimental and computed scattering curves fit with a χ^2 of 2.1 (Fig. S1C), suggesting a similar GerMN1 structure in crystal and in

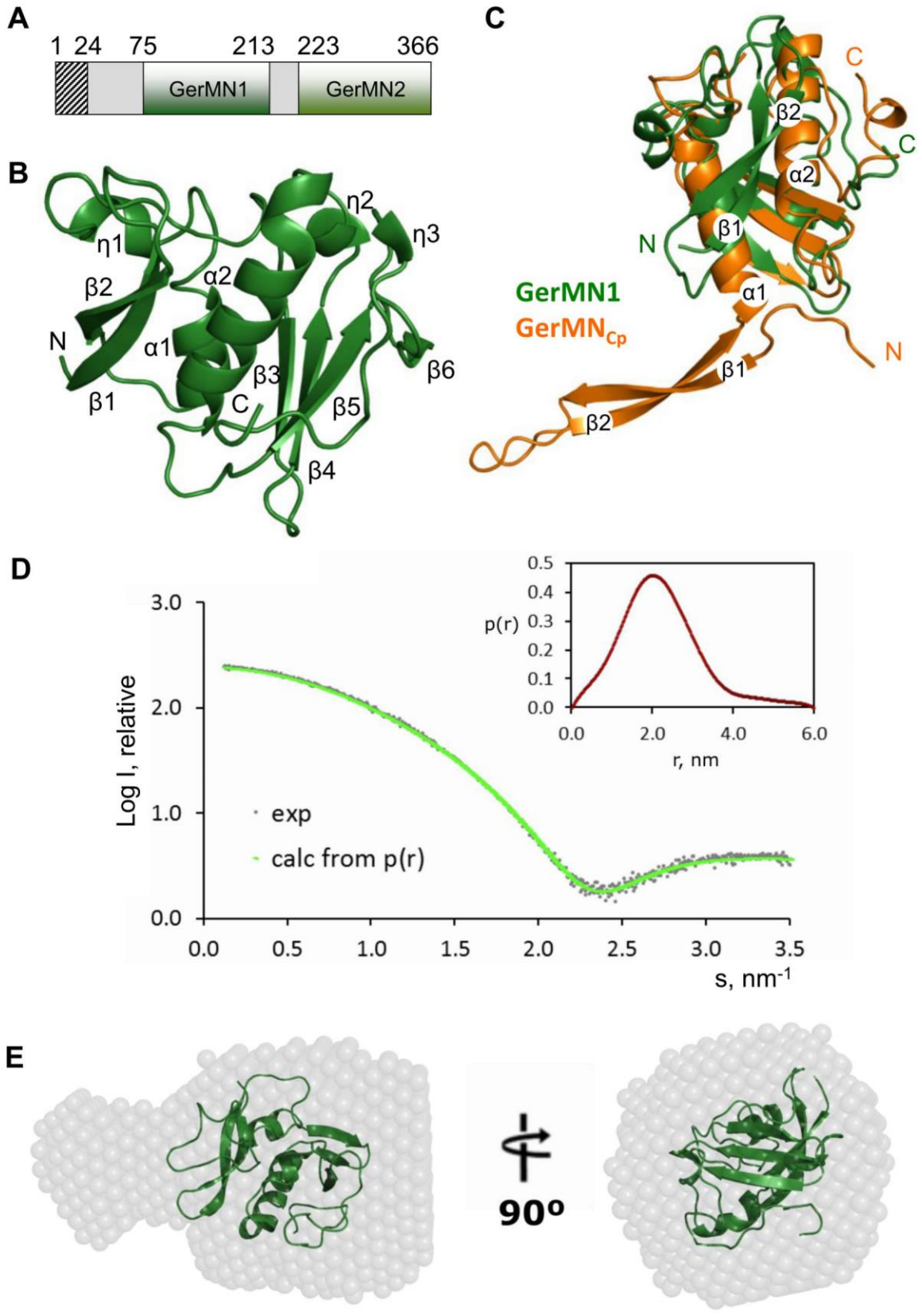


Fig. 1. *Ab initio* crystal structure of the first GerMN domain of GerM from *B. subtilis*. A. Domain structure of GerM showing the lipobox (hatched area), the first and second GerMN domains (GerMN1 in dark green and GerMN2 in light green). B. Ribbon representation of the GerMN1 domain of GerM. β-strands, α-helices, 3_{10}-helices (η), N- and C-termini are labeled. C. Overlay of the GerMN1 domain (in green) from *B. subtilis* GerM with GerMN$_{Cp}$ from *Clostridium perfringens* (in orange, PDB entry 5J7R) showing the "open" (in GerMN$_{Cp}$) and "closed" (in GerMN1) conformations of the β1β2 sheet. Strands β1 and β2, helices α1 and α2, N- and C-termini are labeled. D. GerMN1 scattering data and distance distribution function. Fitting of the scattering curve computed from the $p(r)$ function (colored green) to the experimental scattering data of GerMN1 (colored grey). The insert shows the distance distribution function $p(r)$. E. GerMN1 crystal structure (green cartoon) fitted into the envelope of GerMN1 in solution, generated with DAMMIF (grey spheres) from the experimental scattering curve.

Table 1
Data collection and refinement statistics.

Data collection		
Name of dataset	4GerMN1_C2_1	5GerM1_B2_4
X-ray source	ID23eh1 (ESRF)	ID29 (ESRF)
Wavelength (Å)	0.97242	1.07252
Scan range (°)	102	160
Oscillation (°)	0.15	0.05
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters		
a, Å	38.36	90.36
b, Å	53.60	103.94
c, Å	64.86	162.49
α, °	90.00	90.00
β, °	90.00	90.00
γ, °	90.00	90.00
Number of molecules in ASU	1	4
Resolution (last shell), Å	1.0 (1.06–1.0)	2.1 (2.22–2.1)
Completeness, %	94.5 (90.4)	99.8 (99.4)
I/σ(I)	11.22 (2.95)	17.83 (3.14)
Rsym[†], %	6.1 (35.6)	5.9 (53.8)
Unique reflections	68,958 (10428)	89,726 (13589)
Observed reflections [I/σ(I) > 1]	250,905 (34925)	536,084 (79503)
Wilson B factor, ($Å^2$)	10.5	45.5
Refinement and model statistics		
Resolution (last shell), Å	1.00 (1.00–1.026)	2.1 (2.155–2.1)
R-factor[‡], R-free[§]	0.135, 0.154	0.204, 0.236
rmsd from target[i]		
Bond lengths, Å	0.008	0.004
Bond angle, °	1.33	0.740
Mean B factor ($Å^2$)	14.3	50.1
Ramachandran plot[**]		
Core, %	98.4	97.0
Allowed, %	1.6	2.9
Disallowed, %	0	0.1

Values in parentheses are for the outermost shell of data.

[†] Rsym = (Σ(ABS(I(h,i) − (I(h))))/(Σ(I(h,i))).

[‡] R-factor = ΣjjFoj − jFcjj/ΣjFoj where Fo and Fc are the observed and calculated structure factor amplitudes, respectively.

[§] R-free is the R-factor calculated with 5% of the reflections chosen at random and omitted from refinement.

[i] rmsd of bond lengths and bond angles from ideal geometry.

[**] Performed by Procheck.

solution. However, fitting of GerMN1 crystal structure into the SAXS envelope showed unoccupied volume extending from the main globular region (Fig. 1E). In addition, the linker between strand β2 and helix α1 positioned near this unoccupied volume, and fitting of $GerMN_{Cp}$ in the SAXS envelope showed that the unoccupied extended region could accommodate a β1β2 sheet displaying an "open" conformation (Fig. S1D). These observations thus suggest that when GerMN1 is in solution, the β1β2 sheet might display some flexibility and occupy the elongated region of the low-resolution envelope. In agreement with this hypothesis, the distance distribution function (Fig. 1E, insert) yields a maximum particle dimension (D_{max}) of 6 nm, which is close to the length of the long axis of $GerMN_{Cp}$.

Interestingly, the DALI program also detected structural similarities between the α1β3β4α2β5 core of the GerMN1 domain and RBMs found in PrgK/PrgH components of the inner membrane platform in T3SS and in AF, AG and AH components of the A-Q complex (Fig. 2A–C) (Bergeron et al., 2015; Levdikov et al., 2012; Meisner et al., 2012; Rodrigues et al., 2016a; Zeytuni et al., 2017, 2018b). Although the sequence identity between GerMN domains and RBMs is very weak (lower than 15%), the αββαβ organization of secondary structures is conserved (Fig. 2D). GerMN1 superimposes onto the second RBM of PrgH from *Salmonella typhimurium* (PDB code 4G1I, residues Glu225-Asp295), and RBMs of AG (PDB code 5WC3) and AH (PDB code 3UZ0) from *B. subtilis* with rmsds of 2.57, 3.06 and 3.02 Å, respectively (over 65, 65 and 64 residues, respectively) (Fig. 2A–C). These similarities suggest that similar to RBMs, GerMN domains might have an oligomerization function. GerMN1, however, was only found to form monomers in solution using SAXS ($MM_{I(0)}$ of 13.3 kDa). To investigate whether oligomers would form in the presence of the second GerMN domain, we then sought to characterize the full-length soluble region of GerM.

2.2. Structural characterization of GerM

A recombinant construct lacking the lipobox and encompassing the two GerMN domains ($GerM_{26\text{-}366}$, residues Phe26 to Phe366) was purified and crystallized as rectangle-shaped crystals that diffracted to a resolution of 2.1 Å (Table 1). $GerM_{26\text{-}366}$ structure was solved by molecular replacement using the atomic coordinates of GerMN1 as a template for the two GerMN domains. The main chains of the four molecules in the asymmetric unit could be easily traced from Ser74 to Phe366 but no electron density was visible for residues Phe26 to Ala73. Mass spectrometry analyses performed on the purified protein and on dissolved crystals provided experimental molecular weights of 37,173.4 Da and 37,173.1 Da, respectively (Table S1). These values are very close to the expected molecular weight value (37,172.5 Da), indicating that the protein had not suffered from proteolytic degradation, and that absence of electron density corresponding to residues Phe26 to Ala73 is due to flexibility of this N-terminal region.

The two GerMN domains fold against each other in a compact arrangement displaying a butterfly shape (Fig. 3A). Each GerMN domain forms a wing and the two domains are connected by the C-terminal loop of GerMN1 and a short linker. The GerMN1 and GerMN2 domains, which display 20% sequence identity, superimpose with rmsd of 2.54 Å over 127 residues (Fig. 3B). The interface between the two GerMN domains buries a surface of 1093 $Å^2$ and involves many hydrophobic interactions, which include Y89, V91, L95, P96, I120, F125 and I210 from GerMN1, and Y239, V241, P242, L272 and F366 from GerMN2 (Fig. 3C and S2). The interface also involves a salt bridge established by the side chains of E213 and K245, and multiple hydrogens bonds including the side chains of D85, Y89, T94 and N123 from GerMN1, and of N234, E235, Y239, T244 and N362 from GerMN2 (Fig. 3D and S2).

As expected given the similar fold of the two GerMN domains, the α3β9β10α4β11 core of GerMN2 also resembles RBMs of PrgH, AG and AH onto which it superimposes with rmsds of 2.47, 3.24 and 2.61 Å, respectively (over 64, 66 and 63 residues, respectively) (Fig. S3). Despite the fact that no oligomer could be detected in solution by SEC-MALLS analysis (molecular weight of ~35 kDa, corresponding to a $GerM_{26\text{-}366}$ monomer, Fig. S4), $GerM_{26\text{-}366}$ forms a tetramer in the asymmetric unit of the crystal. If we call the four molecules by their respective chain identifiers, the dimers formed by chains A and B, B and C, and C and D superimpose with rmsds ranging from 0.42 to 0.66 $Å^2$ over 284 residues (Fig. 4A). As the dimer formed by chains D and A′ (the crystallographic symmetry of chain A) is also similar to the dimers observed in the asymmetric unit, the crystallographic symmetries extend the $GerM_{26\text{-}366}$ tetramer into a protein protofilament in which each molecule interacts with the adjacent one through a conserved dimerization interface (Fig. 4B). This interface involves hydrophobic interactions between P199 and P225 from one molecule, and L273 and V269 from the adjacent molecule (Fig. 4C). In addition, seven salt bridges are established by the side chains of K163, K268, D275 and E278 from one chain, and the side chains of K182, K191, D203, D208 and D219, from the neighboring chain. Finally, a dozen of hydrogen bonds are established and include residues N87, G88, K163, T214, G216, K268, V269 and E278 from one chain and residues D202, D203, L220, T221, R246, N249 and K252 from the adjacent molecule (Fig. 4C).

The fact that $GerM_{26\text{-}366}$ oligomers are not detected in solution suggests that GerM's self-association may be weak, requiring the stability and high protein concentration of a crystalline environment to be visualized. GerM oligomers might, however, form *in vivo* when the protein reaches high local concentrations in the intermembrane space and/or when GerM interacts with A-Q components. To investigate this idea, we introduced mutations in residues located at the $GerM_{26\text{-}366}$

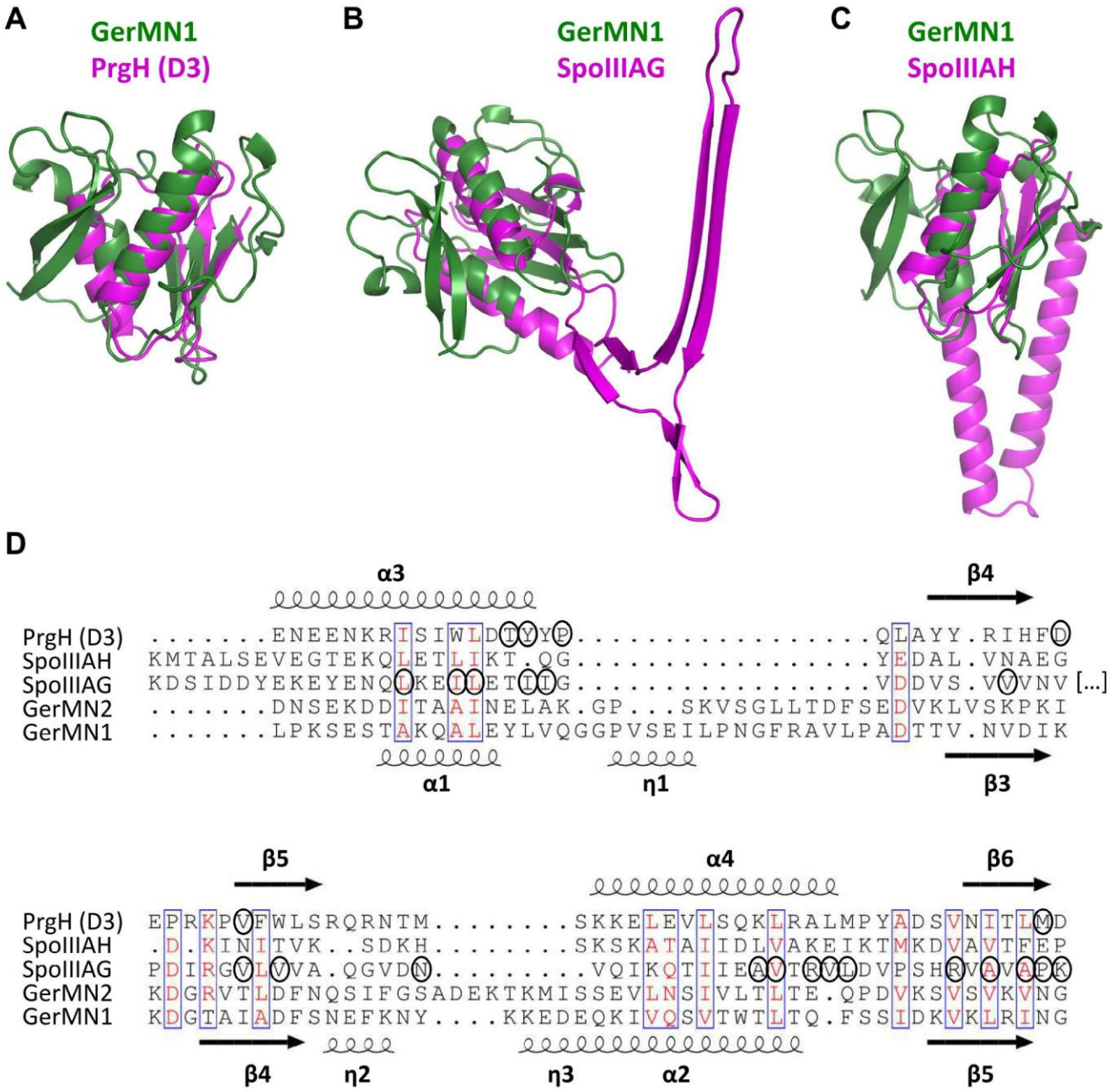


Fig. 2. Comparison of GerMN1 with RBM domains. A-C. Ribbon representations of GerMN1 (in dark green) superimposed onto the second periplasmic RBM of PrgH from *S. typhimurium* (A, PDB code 4G1I, in magenta), onto SpoIIIAG (B, PDB code 5WC3, in magenta) and onto SpoIIIAH (C, PDB code 3UZ0, in magenta) from *B. subtilis*. D. Sequence alignment of RBMs found in PrgH from *S. typhimurium* (second periplasmic domain, residues Glu225-Asp295, PrgH (D3)), SpoIIIAH and SpoIIIAG from *B. subtilis*, with the first (GerMN1) and second (GerMN2) GerMN domains of GerM from *B. subtilis*. The insertion region (Asp124 to Lys181) in the RBM of SpoIIIAG was omitted from the sequence alignment and represented with dots into brackets ([…]). Similar residues are shown by red letters boxed in blue. The secondary structures of PrgH (D3) and GerMN1 are indicated above and below the sequence alignment, respectively. Arrows indicate β-strands; α, α-helices; η, 3_{10}-helices. Residues at the RBM oligomerization interfaces are indicated by black circles.

dimerization interface observed in the crystal (Fig. S2B). When introduced in the recombinant $GerM_{26\text{-}366}$ construct, four of these mutations (K268E, D219R, R246E and K252E) resulted in partial proteolysis of the protein (Table S1). Three mutations (D275R, D203R-D202R and E278R) did not affect the stability of $GerM_{26\text{-}366}$ and were thus introduced *in vivo* in the context of the full-length *gerM* gene. All these mutations were able to rescue a null mutation of *gerM*, in the presence or in the absence of *spoIIIAH* (Table 2). Although it remains possible that a potential GerM oligomer might still be able to form despite these mutations, these data strongly suggest that the GerM homodimerization interface observed in the crystal is not physiological.

3. Conclusions

The structural characterization of GerM reported here reveals that the N-terminal β-sheet (β1β2) of GerMN domains can display different conformations. While the β1β2 sheet of the GerMN1 and GerMN2 domains of GerM display a "closed" conformation (packed against the core of the domain) in the crystals, SAXS analysis of GerMN1 is consistent with the hypothesis that the β1β2 sheet might display an "open" conformation in solution. The "closed" and "open" conformations of the β1β2 sheet might represent different functional states of the GerMN domains and might constitute a way to regulate their function in

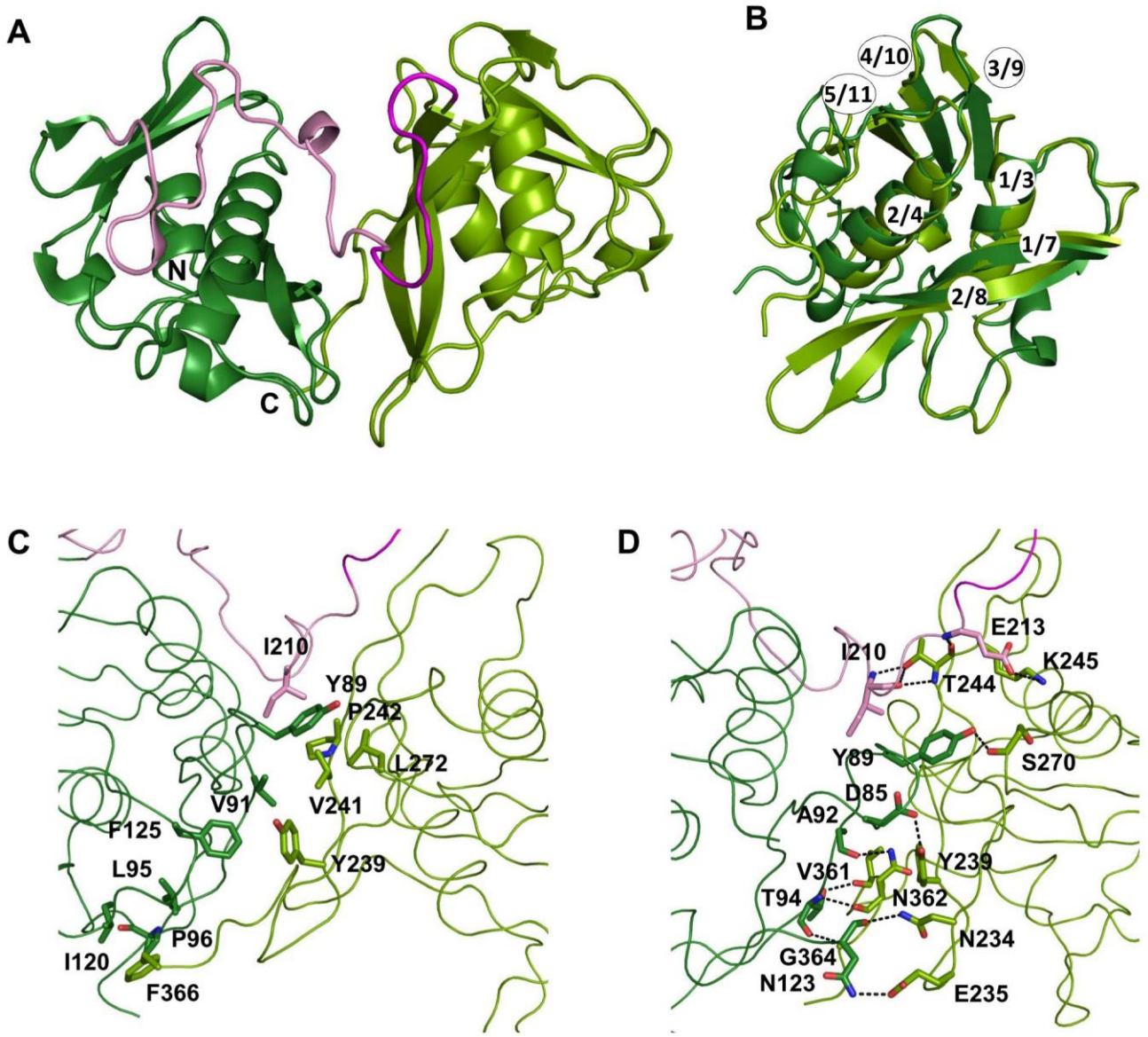


Fig. 3. Structure of $GerM_{26\text{-}366}$ from *B. subtilis*. A. Ribbon representation of $GerM_{26\text{-}366}$, with the GerMN1 and GerMN2 domains colored in dark and light green, respectively. The part of the GerMN1 domain belonging to the loop connecting GerMN1 and GerMN2 is colored in pink and the linker region in magenta. N- and C-termini are labeled. B. Superimposition of GerMN1 (in dark green) onto GerMN2 (in light green). N- and C-termini are labeled. Numbering of β-strands and α-helices are indicated for GerMN1/GerMN2. C–D. Views of the GerMN1-GerMN2 interface with residues involved in hydrophobic (C) or electrostatic (D) interactions labeled and shown as atom-colored sticks. The main chain of the protein is shown as a loop.

different environments (presence of protein partners, cellular location). This hypothesis will be investigated in the future.

In addition, our GerM structures reveal that the αββαβ core of GerMN domains resembles RBMs found in ring-forming proteins such as PrgH from T3SS and AG, which is a central building block of the A-Q complex. In the $GerM_{26\text{-}366}$ crystal, we observe an oligomerization interface which is very different from the oligomerization interfaces observed for PrgH or AG RBMs, and we could not provide evidence that this crystalline interface is physiological. However, it remains possible that GerM oligomerizes *in vivo* through an interface that would resemble those of RBMs and require environmental conditions (interaction with the membrane and/or protein partners) that are missing in our *in vitro* experiments. AH was also proposed to oligomerize through its RBM domain. Redundancy of putative AH and GerM rings in the formation of a conduit at the core of the A-Q complex might thus explain why AH and GerM partially compensate each other in forespore development. This remains a highly speculative hypothesis because no oligomerization of AH or GerM has been demonstrated so far. In addition, oligomerization of RBMs is difficult to predict because interface residues are very poorly conserved between different ring-forming proteins (Fig. 2D). Alternatively, although the core of GerMN domains display similarities with RBMs, the presence of additional N- and C-terminal elements (the β1β2 sheet and the C-terminal loop) might prevent RBM-like oligomerization and provide a totally different function for this family of domains. Consistent with this idea, the RBM domain of AH also contains an additional N-terminal element (the α-helix formed by residues Leu105 to Ser129) when compared to canonical RBMs, and no oligomerization was reported so far for AH. In addition, AH and GerM partially compensate each other for the localization of Q. Based on these observations, one might speculate that a function of non-canonical RBMs such as those present in AH and GerM might be to localize proteins.

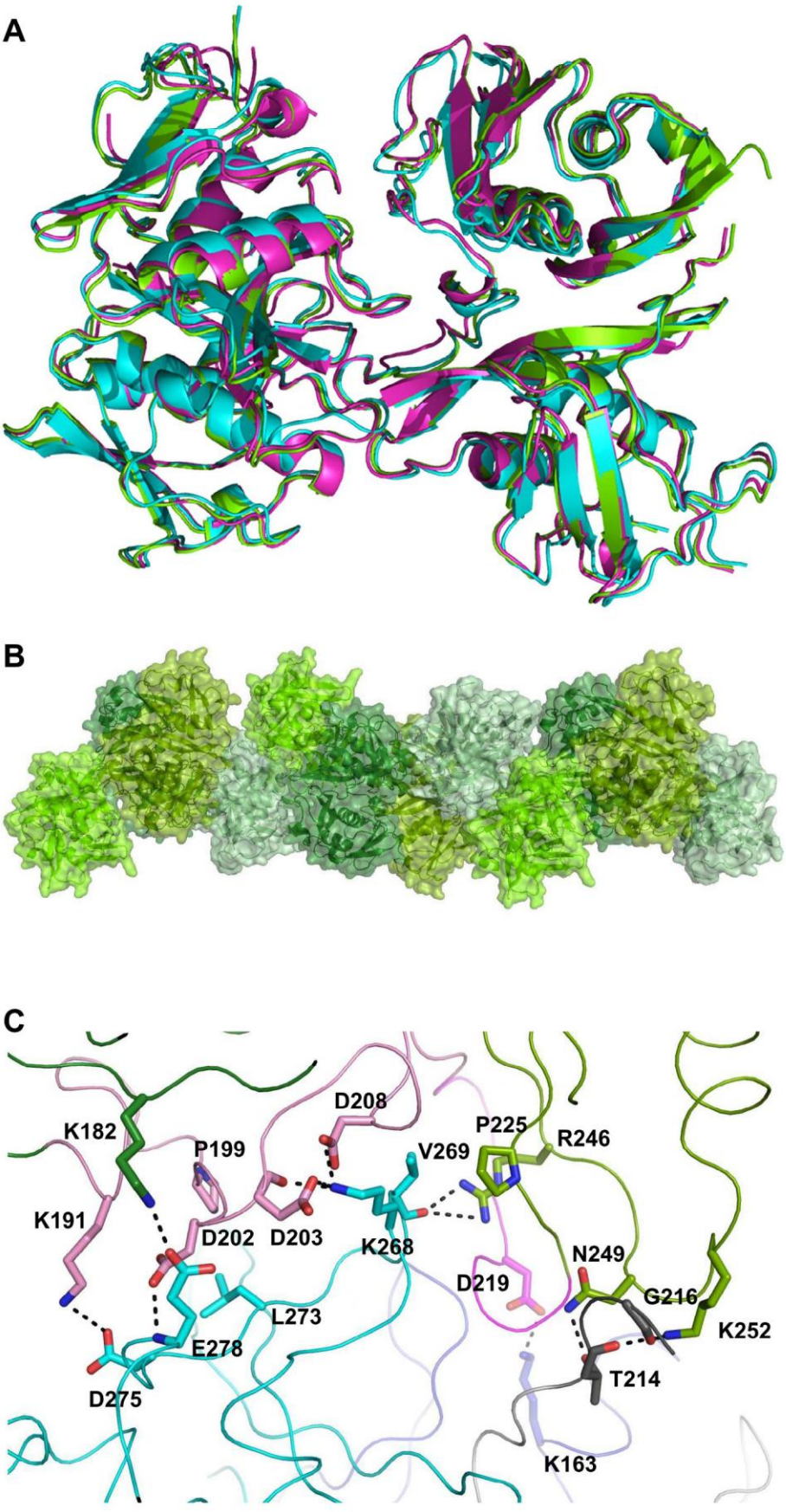


Fig. 4. GerM oligomerization in the crystal. A. Overlay of the $GerM_{26\text{-}366}$ dimers formed by chains A/B (in green), B/C (in cyan) and C/D (in magenta) in the asymmetric unit. B. $GerM_{26\text{-}366}$ protofilament resulting from the continued dimerization of $GerM_{26\text{-}366}$ molecules in the crystal. The protofilament is visualized upon display of the crystallographic symmetries of the asymmetric unit content. The four $GerM_{26\text{-}366}$ molecules in each asymmetric unit are colored in different shades of green and shown with ribbon and surface representations. C. View of the dimerization interface between two $GerM_{26\text{-}366}$ molecules in the crystal structure. Residues involved in hydrophobic or electrostatic interactions are labeled and shown as atom-colored sticks. The main chain of the protein is shown as a loop.

Table 2
Sporulation efficiency of *B. subtilis* strains expressing GerM interface mutants.

Strain	Genotype	Sporulation efficiency (%)
PY79	WT	100
BCR1237	Δ*gerM*	2.1
BCR287	Δ*spoIIIAH*	2.5
BAT0044	Δ*gerM* Δ*spoIIIAH*	0.002
BAT0075	Δ*gerM gerM* (D202R-D203R)	85.6
BAT0076	Δ*gerM* Δ*spoIIIAH gerM* (D202R-D203R)	0.9
BAT0047	Δ*gerM gerM* (D275R)	89.4
BAT0048	Δ*gerM* Δ*spoIIIAH gerM* (D275R)	1.1
BAT0094	Δ*gerM gerM* (E278R)	73.3
BAT0095	Δ*gerM* Δ*spoIIIAH gerM* (E278R)	2.2

4. Materials and Methods

4.1. Bacterial strains and plasmids

Strains, plasmids and oligonucleotides used in this study are listed in Table S2. All *B. subtilis* strains were derived from the prototrophic strain PY79 (Youngman et al., 1983).

4.2. Protein production and purification

All recombinant constructs were fused to a hexahistidine tag followed by the SUMO cleavage site of the Ulp1 protease (His-SUMO tag) (Marblestone et al., 2006) and overexpressed in *E. coli* BL21 (DE3) RIL cells. Cells were grown at 37 °C in 2 l of Terrific Broth (BD Biosciences) supplemented with ampicillin (100 µg/ml) until the OD_{600nm} reached 0.8. Production of recombinant proteins was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to 0.5 mM after the cultures were cooled to 25 °C. Cell growth was continued overnight at 25 °C, and cells were harvested by centrifugation. Cell pellets were resuspended in 1/20th volume of buffer A (50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 25 mM imidazole, 10% (vol/vol) glycerol) containing the Complete™ cocktail of protease inhibitors (Roche). Cells were lysed by six passages through a cell disruptor (Constant Systems Limited) at 20 kPsi, and cell debris were pelleted by centrifugation at 40,000*g* for 30 min at 4 °C. The centrifugation supernatant was loaded on a Ni-NTA agarose resin (Qiagen) equilibrated with buffer A. After extensive washing with buffer A, the fusion protein was eluted with a linear 0–100% gradient of buffer B (50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 500 mM imidazole, 10% (vol/vol) glycerol) over 10 column volumes. Peak fractions were pooled, mixed with a 1:100 dilution of a His-tagged Ulp1 (SUMO) protease preparation (Uehara et al., 2010) and dialyzed overnight at 4 °C in buffer C (50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 10% (vol/vol) glycerol). Cleavage reactions were passed through Ni-NTA resin to remove free His-SUMO tag and His-Ulp1, and untagged protein was collected in the flow through. Flow-through fractions were concentrated with Amicon Ultra Centrifugal filter units with a molecular weight cutoff of 10 kDa (Millipore) and were injected onto an ENrich™ SEC650 10x300 gel-filtration column (Biorad). Proteins were eluted with buffer D (25 mM Tris-HCl (pH 8.0), 150 mM NaCl) and again concentrated with Amicon Ultra Centrifugal filter units. Protein concentration was measured using absorbance at 280 nm.

4.3. Mass spectrometry analyses

Liquid Chromatography Electrospray Ionization Mass Spectrometry (LC/ESI-MS) was applied for quality control of intact proteins by using a 6210 LC/ESI-TOF mass spectrometer interfaced with an HPLC binary pump system (Agilent Technologies). Mass spectra were recorded in the 300–3200 *m/z* range, in the positive ion mode with spectra in the profile mode. The MS spectra were acquired and the data processed with MassHunter workstation software (v. B.02.00, Agilent Technologies) and with GPMAW software (v. 7.00b2, Lighthouse Data, Denmark).

Just before analysis the protein samples were diluted in acidic denaturing conditions to a final concentration of 5 µM with solution E (0.03% TFA in water). Samples were cooled to 10 °C in the autosampler and the analysis was run by injecting 4 µl of each sample. They were first trapped and desalted on a reverse phase-C8 cartridge (Zorbax 300SB-C8, 5um, 300 µm ID × 5 mm, Agilent Technologies) for 3 min at a flow rate of 50 µl/min with 100% solvent E and then eluted with 70% solvent F (95% acetonitrile-5% water-0.03% TFA) at flow rate of 50 µl/min for MS detection.

4.4. SEC-MALLS analyses

SEC-MALLS runs were performed using a ENrich™ SEC650 10 × 300 gel-filtration column (Biorad) connected to an analytic system including a L2130 pump (Hitachi), a L-2400 UV detector (Hitachi), an Optilab T-rEX refractometer (Wyatt technologies) and a DAWN HELEOS-II multi angle light scattering detector (Wyatt technologies). Prior to injection, columns and systems were equilibrated in 5 to 10 column volumes of buffer D. Fifty-µl of protein samples concentrated at a minimum of 2 mg/ml were injected with a constant flow rate of 0.5 ml/min. Protein concentration was quantified online by measuring the differential refractive index and using an averaged refractive index increment *dn/dc* of 0.185 ml/g. Accurate weight-averaged molar masses (MW) determination was performed with the Astra 6 software (Wyatt Technologies) and curves were represented with Excel (Microsoft office 2013).

4.5. Protein crystallization and X-ray data collection

High-throughput crystallization trials were performed with a Cartesian PixSys 4200 crystallization robot (Genomic Solutions, U.K.). Hanging drops containing 100 nl of protein (40, 20 or 10 mg/ml) and 100 nl of reservoir solution were set up in 96-well Crystal Quick plates (Greiner) and incubated at 20 °C. Initial crystal hits were refined manually by setting up hanging drops containing 1 µl of protein (40 or 20 mg/ml) and 1 µl of reservoir solution in 24-well plates (Molecular Dimensions) incubated at 20 °C. Large needle-shaped crystals (dimensions of about 40 × 40 x 400 µm) were finally obtained for GerMN1 (first GerMN domain) in 200 mM Na acetate, 21% (w/vol) PEG 3350. GerMN1 crystals appeared within 2 days and reached their full size within 7 days. The full-length soluble construct of GerM ($GerM_{26\text{-}366}$) yielded short 3D needles in 100 mM Hepes pH 7.2, 2.6 M NH_4SO_4. $GerM_{26\text{-}366}$ crystals appeared within 4 days and reached their full size within 2 weeks. Before X-ray diffraction data collection, crystals were soaked for < 5 min in the appropriate reservoir solution containing 10% (vol/vol) glycerol and 10% (vol/vol) ethylene glycol, and flash-frozen in liquid nitrogen. X-ray diffraction data were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble, France), on the ID23eh1 and ID29 beamlines for GerMN1 and $GerM_{26\text{-}366}$ crystals, respectively. Experimental setup of the beamline and data quality of the collected images were monitored with MxCuBE (Gabadinho et al., 2010). Statistics on data collection and refinement are summarized in Table 1.

4.6. Structure determination and refinement

Diffraction data were indexed and scaled using the XDS program suite (Kabsch, 2010). GerMN1 crystals belong to the orthorhombic space group $P2_12_12_1$, with unit cell dimensions of 38.36 x 53.60 x 64.84 Å and a single molecule per asymmetric unit. *Ab initio* phase determination for GerMN1 was achieved using the ARCIMBOLDO_LITE program (Sammito et al., 2013), which combines the location of model fragments like small α-helices with PHASER (McCoy et al., 2007) and density modification with SHELXE (Thorn and Sheldrick, 2013). We

first unsuccessfully started the structure search by using two 10-residue long polyalanine α-helices (default size), using a machine of 4 cores. Based on secondary structure predictions performed by the JPRED4 server (http://www.compbio.dundee.ac.uk/jpred/), we then searched for two α-helices containing 13 or 12 residues, using machines of 4 cores or 8 cores, respectively. Both strategies yielded an electron density map from the correct positioning of the two α-helices. The high quality of this map allowed automated building of the GerMN1 model (from T76 until D208) using Phenix (Terwilliger et al., 2008).

GerM1 crystals belong to the orthorhombic space group $P2_12_12_1$, with unit cell dimensions of 90.36 × 103.94 × 162.49 Å and four molecules per asymmetric unit. Phase determination was carried out by the molecular replacement method with PHASER, using as a search model for the two GerMN domains the structure of GerMN1. The molecular replacement solution model was rebuilt de novo using PHENIX to prevent bias from the model.

The structures of GerMN1 and $GerM_{26\text{-}366}$ were completed by cycles of manual building with COOT (Emsley and Cowtan, 2004) and addition of water molecules with ARP/wARP (Langer et al., 2008). Several cycles of manual building and refinement with REFMAC (Murshudov et al., 2011), as implemented in the CCP4 program suite, were performed until R_{work} and R_{free} converged (Brünger, 1992). Stereochemical verification was performed with PROCHECK (Laskowski et al., 1993). The secondary structure assignment was verified with DSSP (Kabsch and Sander, 1983). Figures were generated with PyMol (http://www.pymol.org). Coordinates of the final refined models were deposited at the Protein Data Bank (PDB, http://www.rcsb.org) and were assigned PDB entry codes 6GZ8 and 6GZB for GerMN1 and $GerM_{26\text{-}366}$, respectively.

4.7. SAXS analyses

SAXS experiments were carried out at the BM29 beamline (ESRF, Grenoble, France) using a robotic sample changer and a Pilatus 1 M detector (Dectris, Baden-Daettwill, Switzerland) with a detector distance of 2.867 m. The scattering intensity, $I(s)$, was recorded in the range of the momentum transfer $0.03 < s < 4.92\,nm^{-1}$, where $s = (4\pi\sin\theta)/\lambda$, 2θ is the scattering angle and $\lambda = 1.0$ Å the X-ray wavelength (Pernot et al., 2013). The measurements were performed with the sample in buffer D at 20 °C using continuous flow operation over a total exposure time of 0.5 s divided into 10 individual frames to monitor for potential radiation damage. GerMN1 samples at five protein concentrations ranging from 10.4 to 0.63 mg/ml were measured to account for interparticle interactions. Since all samples measured for the different GerMN1 concentrations showed similar SAXS data for the s angles and no repulsive interparticle interference, the SAXS data at low angles from the 1.25 g/L sample were merged with $s > 1.44\,nm^{-1}$ intensities of the 10.4 g/L sample of GerMN1 using PRIMUS to produce the final GerMN1 experimental scattering profile.

The data were processed with ATSAS package (https://www.embl-hamburg.de/biosaxs) (Franke et al., 2017) using standard procedures, corrected for buffer contribution, and extrapolated to infinite dilution using the program PRIMUS. The forward scattering $I(0)$ and the radii of gyration R_g were evaluated using the Guinier approximation assuming that at very small angles ($sR_g < 1.3$) the intensity is represented as $I(s) = I(0)\exp(-(sR_g)^2/3)$. Both parameters were combined with protein concentration relative to bovine serum albumin standard to assess the GerMN molecular mass $MM_{I(0)}$. The entire scattering pattern was also computed using the indirect transform package GNOM that provides the maximum dimension of the particle D_{max} and the distance distribution function $p(r)$. The excluded volume of the hydrated particle (the Porod volume, V_p) was computed using the Porod invariant (Petoukhov et al., 2012). The GerMN molecular weight in solution was also calculated from the concentration-independent excluded Porod volume (MM_{porod}) assuming the 1.7 empirical ratio between V_p and the MM of a protein. The program CRYSOL was used to compute the scattering from the GerMN1 crystal structure. The *ab initio* modeling program DAMMIF was employed for low-resolution shape generation, and 20 models were calculated in the slow mode, using standard settings. The program DAMAVER was used to superimpose individual structures, and to determine the averaged and the most probable reconstruction. The *ab initio* model was superimposed with the high-resolution crystal structure of GerMN1 using SUPCOMB.

Acknowledgments

We thanks members of the Vernet laboratory for support, advice and encouragement. We thank Luca Signor for mass spectrometry analysis, Christoph Mueller (ESRF, beamline ID23eh1) and Nicolas Foos (ESRF, beamline ID29) for support in data collection. This work was supported by start-up funds provided by University of Technology Sydney to Christopher Rodrigues, and by funds provided by the French National Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02) to Cecile Morlot. We acknowledge the platforms of the Grenoble Instruct center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL) supported by the French Infrastructure for Integrated Structural Biology Initiative FRISBI (ANR-10-INSB-05-02) and by the Grenoble Alliance for Integrated Structural Cell Biology GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB).

Author contributions

J.T., A.M., F.L., D.Z.R., C.D.A.R. and C.Mo. designed the research; J.T., A.M., F.L., C.C-M., B.L., C.M., C.D.A.R. and C.Mo performed the research; J.T., A.M., F.L., C.C-M., C.M., C.D.A.R. and C.Mo. analyzed the data; C.D.A.R. and C.Mo. wrote the paper; J.T., A.M., F.L., C.C-M., B.L., C.M., and D.Z.R. helped revise the paper.

Conflict of interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsb.2018.09.010.

References

Bergeron, J.R.C., Worrall, L.J., De, S., Sgourakis, N.G., Cheung, A.H., Lameignere, E., Okon, M., Wasney, G.A., Baker, D., McIntosh, L.P., Strynadka, N.C.J., 2015. The modular structure of the inner-membrane ring component PrgK facilitates assembly of the Type III secretion system basal body. Structure 23, 161–172.

Brünger, A.T., 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 355, 472–475.

Camp, A.H., Losick, R., 2008. A novel pathway of intercellular signalling in Bacillus subtilis involves a protein with similarity to a component of type III secretion channels. Mol. Microbiol. 69, 402–417.

Camp, A.H., Losick, R., 2009. A feeding tube model for activation of a cell-specific transcription factor during sporulation in *Bacillus subtilis*. Genes Dev. 23, 1014–1024.

Chandran, V., 2013. Type IV secretion machinery: molecular architecture and function. Biochem. Soc. Trans. 41, 17–28.

Costa, T.R.D., Felisberto-Rodrigues, C., Meir, A., Prevost, M.S., Redzej, A., Trokter, M., Waksman, G., 2015. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nat. Rev. Microbiol. 13, 343–359.

Doan, T., Morlot, C., Meisner, J., Serrano, M., Henriques, A.O., Moran, J.C.P., Rudner, D.Z., 2009. Novel secretion apparatus maintains spore integrity and developmental gene expression in Bacillus subtilis. PLoS Genet. 5, e1000566.

Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132.

Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., Svergun, D.I., 2017. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J. Appl. Crystallogr. 50, 1212–1225.

Gabadinho, J., Beteva, A., Guijarro, M., Rey-Bakaikoa, V., Spruce, D., Bowler, M.W., Brockhauser, S., Flot, D., Gordon, E.J., Hall, D.R., Lavault, B., McCarthy, A.A., McCarthy, J., Mitchell, E., Monaco, S., Mueller-Dieckmann, C., Nurizzo, D., Ravelli,

R.B.G., Thibault, X., Walsh, M.A., Leonard, G.A., McSweeney, S.M., 2010. MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments. J. Synchrotron Radiat. 17, 700–707.
Guillot, C., Moran, C.P., 2007. Essential Internal Promoter in the spoIIIA Locus of *Bacillus subtilis*. J. Bacteriol. 189, 7181–7189.
Higgins, D., Dworkin, J., 2012. Recent progress in *Bacillus subtilis* sporulation. FEMS Microbiol. Rev. 36, 131–148.
Hilbert, D.W., Piggot, P.J., 2004. Compartmentalization of gene expression during *Bacillus subtilis* spore formation. Microbiol. Mol. Biol. Rev. 68, 234–262.
Holm, L., Rosenström, P., 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res. 38, W545–W549.
Kabsch, W., 2010. XDS. Acta Crystallogr. D Biol. Crystallogr. 66, 125–132.
Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637.
Langer, G.G., Cohen, S.X., Lamzin, V.S., Perrakis, A., 2008. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. 3, 1171–1179.
Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291.
Levdikov, V.M., Blagova, E.V., McFeat, A., Fogg, M.J., Wilson, K.S., Wilkinson, A.J., 2012. Structure of components of an intercellular channel complex in sporulating *Bacillus subtilis*. Proc. Natl. Acad. Sci. 109, 5441–5445.
Londoño-Vallejo, J.-A., Fréhel, C., Stragier, P., 1997. spoIIQ, a forespore-expressed gene required for engulfment in *Bacillus subtilis*. Mol. Microbiol. 24, 29–39.
Marblestone, J.G., Edavettal, S.C., Lim, Y., Lim, P., Zuo, X., Butt, T.R., 2006. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. Protein Sci. Publ. Protein Soc. 15, 182–189.
McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674.
Meisner, J., Wang, X., Serrano, M., Henriques, A.O., Moran, C.P., 2008. A channel connecting the mother cell and forespore during bacterial endospore formation. Proc. Natl. Acad. Sci. 105, 15100–15105.
Meisner, J., Maehigashi, T., André, I., Dunham, C.M., Moran, C.P., 2012. Structure of the basal components of a bacterial transporter. Proc. Natl. Acad. Sci. 109, 5446–5451.
Morlot, C., Rodrigues, C.D.A., 2018. The new kid on the block: a specialized secretion system during bacterial sporulation. Trends Microbiol. 26, 663–676.
Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A., 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. 67, 355–367.
Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., Morawe, C., Mueller-Dieckmann, C., Ohlsson, S., Schmid, W., Surr, J., Theveneau, P., Zerrad, L., McSweeney, S., 2013. Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. J. Synchrotron Radiat. 20, 660–664.
Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V., Svergun, D.I., 2012. New developments in the ATSAS program package for small-angle scattering data analysis. J. Appl. Crystallogr. 45, 342–350.
Portaliou, A.G., Tsolis, K.C., Loos, M.S., Zorzini, V., Economou, A., 2016. Type III secretion: building and operating a remarkable nanomachine. Trends Biochem. Sci. 41, 175–189.
Rigden, D.J., Galperin, M.Y., 2008. Sequence analysis of GerM and SpoVS, uncharacterized bacterial "sporulation" proteins with widespread phylogenetic distribution. Bioinforma. Oxf. Engl. 24, 1793–1797.
Rodrigues, C.D.A., Ramírez-Guadiana, F.H., Meeske, A.J., Wang, X., Rudner, D.Z., 2016b. GerM is required to assemble the basal platform of the SpoIIIA–SpoIIQ transenvelope complex during sporulation in *Bacillus subtilis*. Mol. Microbiol. 102, 260–273.
Rodrigues, C.D.A., Henry, X., Neumann, E., Kurauskas, V., Bellard, L., Fichou, Y., Schanda, P., Schoehn, G., Rudner, D.Z., Morlot, C., 2016a. A ring-shaped conduit connects the mother cell and forespore during sporulation in *Bacillus subtilis*. Proc. Natl. Acad. Sci. 113, 11585–11590.
Sammito, M., Millán, C., Rodríguez, D.D., de Ilarduya, I.M., Meindl, K., De Marino, I., Petrillo, G., Buey, R.M., de Pereda, J.M., Zeth, K., Sheldrick, G.M., Uson, I., 2013. Exploiting tertiary structure through local folds for crystallographic phasing. Nat. Methods 10, 1099–1101.
Sammons, R.L., Slynn, G.M., Smith, D.A., 1987. Genetical and molecular studies on gerM, a new developmental locus of Bacillus subtilis. J. Gen. Microbiol. 133, 3299–3312.
Schraidt, O., Marlovits, T.C., 2011. Three-Dimensional model of Salmonella's needle complex at subnanometer resolution. Science 331, 1192–1195.
Slynn, G.M., Sammons, R.L., Smith, D.A., Moir, A., Corfe, B.M., 1994. Molecular genetical and phenotypical analysis of the *gerM* spore germination gene of *Bacillus subtilis* 168. FEMS Microbiol. Lett. 121, 315–320.
Tan, I.S., Ramamurthi, K.S., 2014. Spore formation in *Bacillus subtilis*. Environ. Microbiol. Rep. 6, 212–225.
Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.-W., Read, R.J., Adams, P.D., 2008. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallogr. D Biol. Crystallogr. 64, 61–69.
Thorn, A., Sheldrick, G.M., 2013. Extending molecular-replacement solutions with SHELXE. Acta Crystallogr. D Biol. Crystallogr. 69, 2251–2256.
Uehara, T., Parzych, K.R., Dinh, T., Bernhardt, T.G., 2010. Daughter cell separation is controlled by cytokinetic ring-activated cell wall hydrolysis. EMBO J. 29, 1412–1422.
Worrall, L.J., Hong, C., Vuckovic, M., Deng, W., Bergeron, J.R.C., Majewski, D.D., Huang, R.K., Spreter, T., Finlay, B.B., Yu, Z., Strynadka, N.C.J., 2016. Near-atomic-resolution cryo-EM analysis of the Salmonella T3S injectisome basal body. Nature 540, 597–601.
Yokoyama, K., Muneyuki, E., Amano, T., Mizutani, S., Yoshida, M., Ishida, M., Ohkuma, S., 1998. V-ATPase of Thermus thermophilus Is Inactivated during ATP Hydrolysis but Can Synthesize ATP. J. Biol. Chem. 273, 20504–20510.
Youngman, P.J., Perkins, J.B., Losick, R., 1983. Genetic transposition and insertional mutagenesis in *Bacillus subtilis* with *Streptococcus faecalis* transposon Tn917. Proc. Natl. Acad. Sci. 80, 2305–2309.
Zeytuni, N., Hong, C., Flanagan, K.A., Worrall, L.J., Theiltges, K.A., Vuckovic, M., Huang, R.K., Massoni, S.C., Camp, A.H., Yu, Z., Strynadka, N.C.J., 2017. Near-atomic resolution cryoelectron microscopy structure of the 30-fold homooligomeric SpoIIIAG channel essential to spore formation in *Bacillus subtilis*. Proc. Natl. Acad. Sci. U. S. A. 114, E7073–E7081.
Zeytuni, N., Flanagan, K.A., Worrall, L.J., Massoni, S.C., Camp, A.H., Strynadka, N.C.J., 2018b. Structural and biochemical characterization of SpoIIIAF, a component of a sporulation-essential channel in *Bacillus subtilis*. J. Struct. Biol. 204, 1–8.
Zeytuni, N., Flanagan, K.A., Worrall, L.J., Massoni, S.C., Camp, A.H., Strynadka, N.C.J., 2018a. Structural characterization of SpoIIIAB sporulation-essential protein in *Bacillus subtilis*. J. Struct. Biol. 202, 105–112.

C. Determination of structural determinants for the (putative) oligomerization of RBM-containing proteins from the A-Q complex

Except GerM, which is a lipoprotein, all the RBM-containing proteins of the A-Q complex (AF, AG and AH) have a transmembrane (TM) segment. AF has a canonical RBM domain but was not shown to form ring so far. AG forms 30-mer rings but has a non-canonical RBM domain in which a unique β-strand structure (called a β-triangle) is inserted within the core RBM domain (Fig. 23). In addition, AG possesses a N-terminal unfolded region. Finally, AH possesses a non-canonical RBM domain containing an additional N-terminal helix, as well as a N-terminal unfolded region (Fig. 23).

The "RBM" term suggests that this domain is sufficient to promote ring-like oligomerization and indeed, some RBM-containing proteins were shown to form rings *in vivo* (PrgK, PrgH, InvG, GspD and AG) and those rings were observed *in vitro* for AG. However, whether the RBM fold is sufficient to promote oligomerization, whether ring-formation is initiated, stabilized or inhibited by other regions, and whether all RBM-containing proteins are able to forms rings remain open questions. In order to investigate this, we characterized the biophysical behavior of different constructs of A-Q proteins containing RBM-like domains. Beyond a better comprehension of the oligomerization capacity of RBM-containing proteins, our objective was to get evidence that AF and AH also form rings.

1. Role of the transmembrane segments

The constructs used in all the previous *in vitro* studies of AF, AG and AH only contained the soluble domain of the proteins: the $AF_{K55\text{-}E206}$ construct lacked the 54 N-terminal residues (including 2 TM segments), the $AG_{S51\text{-}S229}$ construct lacked the 50 N-terminal residues (including 1 TM segment), and the $AH_{A32\text{-}K218}$ construct lacked the 31 N-terminal residues (including 1 TM segment). In order to determine whether the TM segment(s) of AF and AH are required for protein oligomerization, we purified full-length (FL) recombinant constructs of those proteins. As a positive control for our capacity to purify ring-forming membrane proteins, we also purified the full-length membrane form of AG. Finally, we also purified the full-length

membrane form of Q to test whether it would help the oligomerization of AH.

a. Purification of full-length membrane forms of AF, AG, AH and Q

- **Small-scale purification of full-length membrane AG ($AG_{M1-S229}$)**

In order to identify a detergent with optimal solubilization and purification yields, 18 different detergents were tested, including ionic detergents (FC12, NLS), non-ionic detergents (cymal-5, B-OG, C12E8, TX-100, DM, UDM, DMNG) and zwitter-ionic detergents (CHAPS, LDAO, LAPAO) (See the *Materials and Methods* chapter and Fig. 27A). In collaboration with Anne-Marie Villard from my hosting laboratory, we performed small-scale solubilization and purification assays of His-SUMO-$AG_{M1-S229}$. Coomassie-stained SDS-PAGE gel showed that His-SUMO-$AG_{M1-S229}$ purification efficiency was the highest in Cymal 5 (5-Cyclohexyl-1-Pentyl-β-D-Maltoside), FC12 (Fos-Choline 12) and DDM (n-Dodecyl-β-D-Maltoside) (Fig. 27B).

(A)

	Detergent	Conc solubilization (mM/%)	Conc purification(mM)
1	CHAPS	28	10
2	Cymal 5	22,4	4,4
3	FC12	21,5	3,5
4	B-OG	23	20
5	LDAO	21	3
6	LAPAO	21,6	3,6
7	C12E8	20,1	2,1
8	NLS	34,6	16,6
9	TX100	20,2	2,2
10	DM	21,8	3,8
11	UDM	20,6	2,6
12	DMNG	0,034%	2
13	DDM / DDM	20,17	2,2
14	DDM / UDM	20,17	2,6
15	DDM / DMNG	20,17	2
16	DDM / DDM	20,17	2,2
17	DDM / UDM	20,17	2,6
18	DDM / DMNG	20,17	2

(B)

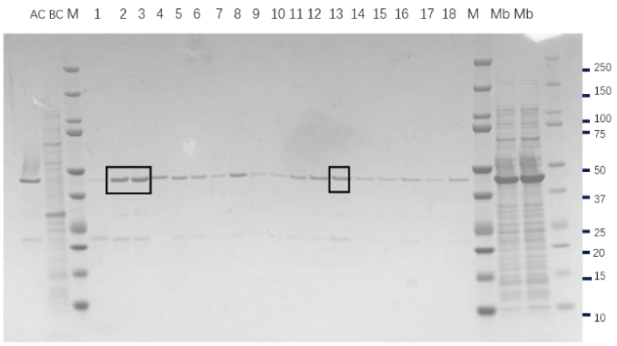


Figure 27. Detergent screening. (A) List of the different detergents used in the small-scale solubilization and purification assays of His-SUMO-$AG_{M1-S229}$. The concentrations used for the solubilization of membrane proteins (Conc. solubilization) and for the Ni-NTA purification (Conc. purification) are indicated. **(B)** 12.5% SDS-PAGE gel stained by Coomassie blue, showing the output of small-scale Ni-NTA purification of His-SUMO-$AG_{M1-S229}$ using different detergents. Numbers on top of each lane refer to the detergents listed in panel A. His-SUMO-$AG_{M1-S229}$ was produced in *E. coli* BL21 (DE3) Star, membrane proteins in groups 1-15 were solubilized in Tris-HCl 50 mM pH 8.0, NaCl 500 mM, containing an appropriate concentration of detergent (see panel A), proteins in groups 16-18 were solubilized in Tris-HCl 50 mM pH 8.0, NaCl 300 mM, containing an appropriate concentration of detergent , and purified using 50-µL Ni-NTA resin. M, molecular weight markers; Mb, membranes. The protein bands squared with black boxes indicate the conditions that provided the highest purification yields.

Normally, non-ionic detergents are considered to be milder than ionic detergents and preserve protein-protein interaction (Stetsenko and Guskov 2017). Besides, the concentration of detergents should reach their critical micelle concentration (CMC) to disrupt the membrane of membrane proteins in order to solubilize them, but the overconcentration of detergents can denature the membrane proteins. Therefore, we chose DDM (non-ionic detergent, CMC ~ 0.17 mM) for further membrane protein purification to minimize the detergent concentration, rather than cymal 5 (non-ionic detergent, CMC ~ 2.5 mM) and FC12 (ionic detergent, CMC ~ 1.5 mM). DDM was used to purify AG, but also AF, AH and Q, in order to be able to eventually mix all these proteins together and reconstitute part of the A-Q complex.

- **Large-scale purification of full-length membrane AG, AH, AF and Q**

Before affinity chromatography, the detergent concentration should be sufficient to solubilize all membrane proteins contained in the crude membrane extract from *E. coli*. During elution from the affinity resin, the detergent concentration can be decreased because all the contaminant *E. coli* membrane proteins have been eliminated during the washing step, but it should still be sufficient to solubilize the TM segments of the purified protein. In addition, since the A-Q proteins are supposed to form rings that might involve hydrophobic interactions (as it is indeed the case for the AG rings), we wanted the detergent concentrations to be low enough to preserve such interactions. The concentration of DMM used for the large-scale purification of His-SUMO-$AG_{M1-S229}$ was therefore further optimized in order to use the lowest possible concentration.

In the end, His-SUMO-$AG_{M1-S229}$ was solubilized in 10 mM DDM (60 times the CMC), and eluted from the Ni-NTA resin using 0.4 mM DDM (twice the CMC) (see the *Material and Methods* section). Cleavage of the His-SUMO tag by the Ulp1 protease resulted in aggregation of the protein. Large-scale purification of His-SUMO-$AG_{M1-S229}$ was thus performed using a Ni-NTA affinity chromatography followed by size-exclusion chromatography in Tris-HCl 25 mM pH 8.0, NaCl 150 mM, DDM 0.4 mM, and the protein was concentrated to 1-2 mg/mL using 10-kDa concentration units.

As shown in figure 28, we obtained a homogenous sample of His-SUMO-$AG_{M1-S229}$. When

analyzed by SDS-PAGE, the protein migrated at the expected position, corresponding to a molecular weight of about 30 kDa, and it showed no sign of proteolytic degradation. Larger species were also observed on the top part of the gel. Large species corresponding to protein rings had also been observed with a previously purified soluble form of AG ($AG_{S51-S229}$) (Rodrigues, Henry, et al. 2016). Indeed, the AG rings are so stable that they can resist SDS denaturation. Here again, negative-stain electron microscopy analysis of the purified membrane form of AG showed that these large species correspond to SDS-resistant rings of His-SUMO-$AG_{M1-S229}$ (Fig. 28A).

His-SUMO-$AF_{M1-E206,}$ His-SUMO-$AH_{M1-K218}$ and His-SUMO-$Q_{M1-S283}$ were solubilized and purified using the protocol established for His-SUMO-$AG_{M1-S229.}$

Following cleavage of the His-SUMO tag and size-exclusion chromatography, the purified $Q_{M1-S283}$ sample migrated as expected on a SDS-PAGE gel given its theoretical molecular weight (43.7 kDa) (Fig. 28B). This likely intact species represents the major content of the sample. However, a species of about 22 kDa could also be observed. This species is too large to correspond to the His-SUMO tag (~ 15 kDa), it thus likely corresponds to a proteolytic degradation product of $Q_{M1-S283}$.

Following cleavage of the His-SUMO tag and size-exclusion chromatography, SDS-PAGE analysis of the purified $AH_{M1-K218}$ sample showed three different species, in roughly equivalent proportions. A first species likely corresponds to the intact $AH_{M1-K218}$ protein (theoretical molecular weight of 38.3 kDa) while two species of lower molecular weight, migrating around 30 kDa and 22 kDa, likely correspond to proteolytic degradation products of $AH_{M1-K218}$ (Fig. 28B).

The His-SUMO-$AF_{M1-E206}$ construct was purified by Elda Bauda, a M1 student in my laboratory (Fig. 28C). Following affinity purification, the His-SUMO-$AF_{M1-E206}$ protein only displayed a slight proteolytic degradation, as the major part of the sample migrated at the expected position (theoretical molecular weight of about 36 kDa). However, the His-SUMO tag could not be cleaved by the Ulp1 protease, suggesting that the protein was aggregated or

that the tag was not accessible to the protease.

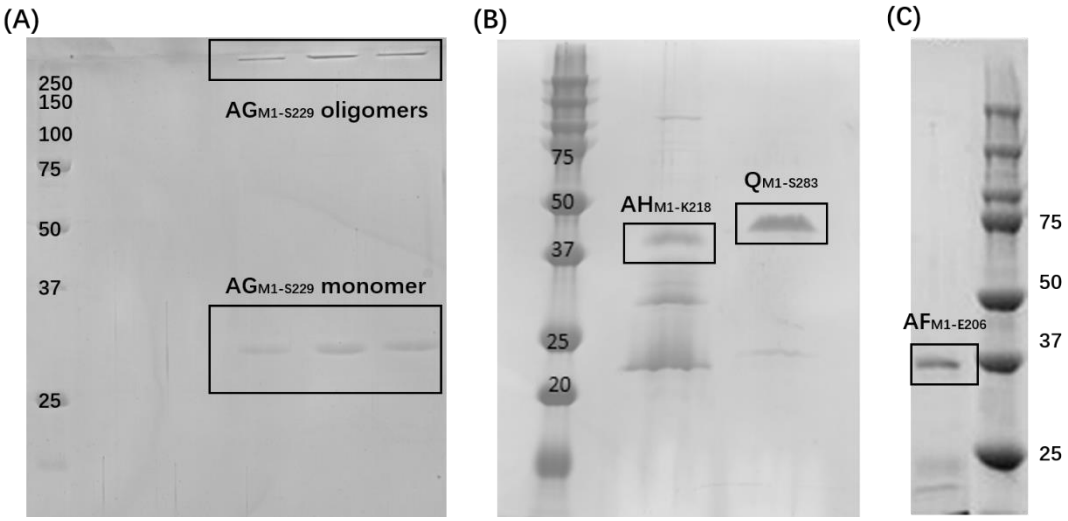


Figure 28. Purification of full-length RBM-containing proteins from *B. subtilis*. (A) 12.5% SDS-PAGE gel stained with Coomassie blue, showing elution fractions of size-exclusion chromatography purification of His-SUMO-$AG_{M1-S229}$. The band migrating around 30 kDa corresponds to monomers of His-SUMO-$AG_{M1-S229}$ while the band migrating above 250 kDa corresponds to rings of His-SUMO-$AG_{M1-S229}$. **(B)** Coomassie blue-stained gel showing the protein content of the purified samples of $AH_{M1-K218}$ and $Q_{M1-S283}$ following size-exclusion chromatography and protein concentration. Both $AH_{M1-K218}$ and $Q_{M1-S283}$ migrate around 40 kDa. **(C)** Coomassie blue-stained gel showing the concentrated sample of His-SUMO-$AF_{M1-E206}$ after nickel-affinity chromatography.

b. EM analysis of full-length membrane AF, AG and AH

To investigate whether these full-length membrane RBM-containing proteins from the A-Q complex could form oligomers *in vitro*, we analyzed the purified samples by negative-stain EM. These analyses were done in collaboration with Emmanuelle Neumann and Daphna Fenel from the Electron Microscopy and Methods group at the IBS.

We could not observe any organized oligomer of His-SUMO-$AF_{M1-E206}$ or $AH_{M1-K218}$ by negative-stain EM. Given that AH and Q interact directly, we wondered whether Q would be required for the oligomerization of AH. Unfortunately, mixing the two $AH_{M1-K218}$ and $Q_{M1-S283}$ samples did not result in the oligomerization of AH. By contrast, His-SUMO-$AG_{M1-S229}$ displayed an unexpected macromolecular assembly.

The cryo-EM structure of the soluble domain of AG ($AG_{S51-S229}$) had been previously studied by my hosting lab and the group of N. Strynadka. They had discovered that $AG_{S51-S229}$ forms homo-oligomeric rings composed of 30 protomers and displaying a "saucer-and-cup" architecture with a large central pore (Fig. 17 from the *Introduction* chapter) (Rodrigues, Henry, et al. 2016, Zeytuni et al. 2017). When observed by negative-stain EM, those rings lye on a preferred orientation on the microscopy grid and are therefore mostly viewed from the top (Fig. 29A). A few rings can however be observed from the side, and in this orientation, the saucer region can be distinguished from the cup region (Fig. 29A).

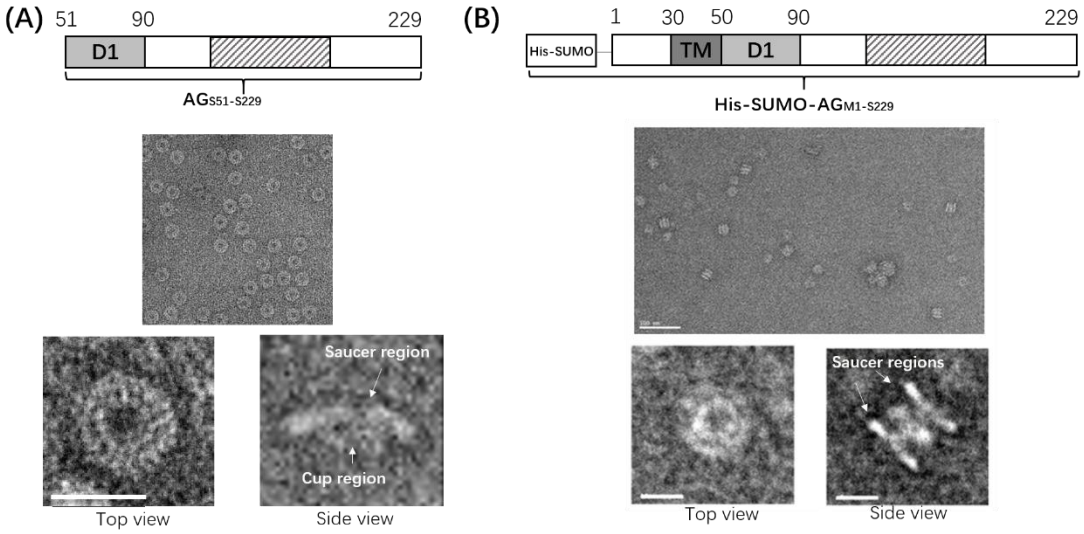


Figure 29. Rings formed by soluble and membrane AG constructs. (A) The top panel shows domain boundaries of the soluble $AG_{S51-S229}$ construct. The middle panel shows a large field of rings formed by the purified $AG_{S51-S229}$ sample and observed by negative-stain EM. The lower panels show top and side view of a single ring formed by $AG_{S51-S229}$. The white arrows point at the saucer and the cup regions in the side view panel. Scale bar is 20 nm. EM micrographs are from a previous study from my hosting lab (Rodrigues, Henry, et al. 2016). **(B)** The top panel shows domain boundaries of the membrane His-SUMO-$AG_{M1-S229}$ construct. The middle panel shows a large field of double rings formed by the DDM-solubilized His-SUMO-$AG_{M1-S229}$ sample and observed by negative-stain EM. Scale bar is 100 nm. The lower panels show top and side views of a double ring formed by His-SUMO-$AG_{M1-S229}$. The white arrows point at the putative saucer regions in the side view panel. Scale bar is 10 nm.

Surprisingly, negative-stain EM analysis of the His-SUMO-$AG_{M1-S229}$ construct revealed the presence of double rings that were never reported before (Fig. 29B). These double rings displayed a majority of side orientations on the EM grid (Fig. 29B), allowing to distinguish

three regions: two dense and thin bands flanking a central one that is shorter and thicker. Based on their dimensions, we predict that the two external bands correspond to saucer regions while the inner part might result from the association of the TM segments or from the association of the cup region. I did not have time to push the EM study of the His-SUMO-$AG_{M1-S229}$ sample further so the hypotheses regarding the orientation of the two rings are discussed in section IIA of the *Discussion* chapter.

Intriguingly, cleavage of the His-SUMO tag for this construct resulted in the aggregation of the cleaved protein, suggesting that the His-SUMO tag stabilizes the full-length membrane protein. A few double rings could still be observed after cleavage (not shown here) but we cannot exclude that they correspond to uncleaved traces of His-SUMO-$AG_{M1-S229}$. This observation contrast with the fact that the soluble His-SUMO-$AG_{S51-S229}$ construct required cleavage of the His-SUMO tag to form rings (Rodrigues, Henry, et al. 2016).

In order to determine whether His-SUMO-$AG_{M1-S229}$ rings would promote the oligomerization of His-SUMO-$AF_{M1-E206}$, $AH_{M1-K218}$ or the oligomerization of a mix containing $AH_{M1-K218}$ and $Q_{M1-S283}$, His-SUMO-AG_{M1-S2} was mixed with His-SUMO-$AF_{M1-E206}$, $AH_{M1-K218}$ or $AH_{M1-K218}$-$Q_{M1-S283}$ after purification and observed by negative-stain EM. In parallel, experiments were performed in which membranes were co-solubilized with DDM and proteins were co-purified before EM analysis. In a "last chance" experiment, we even mixed all these proteins together (His-SUMO-$AG_{M1-S229}$, His-SUMO-$AF_{M1-E206}$, $AH_{M1-K218}$ and $Q_{M1-S283}$) for EM observation. If the presence of His-SUMO-$AG_{M1-S229}$ rings would have promoted the oligomerization of other A-Q proteins, we would have observed additional density around the AG double rings. Unfortunately, negative-stain EM analysis of all these samples failed to reveal the presence of His-SUMO-$AF_{M1-E206}$, $AH_{M1-K218}$ or $AH_{M1-K218}$-$Q_{M1-S283}$ rings, or even the presence of additional material around the AG rings.

In conclusion, we managed to produce and purify large quantities of recombinant, full-length membrane forms of AF, AG, AH and Q (about 500 µL at 1-2 mg/mL), which was *per se* a real challenge. Unfortunately, we did not detect any ring-like structures for AF, AH or for the AH-Q complex, when the proteins were observed alone or in combination with A-Q partners,

including the AG rings. These results seriously question the existence of rings made by AF and AH and this aspect is discussed in the *Discussion* section IIB. A potentially exciting result was obtained on the other hand, with the observation of the $AG_{M1\text{-}S229}$ double rings.

2. Role of the soluble regions

AF, AG and AH all contain an RBM-like domain that shares secondary structure similarities with RBM domains found in the PrgK and PrgH family of T3SS components. While AG was indeed shown to form rings resembling those of PrgK or PrgH, we have so far no evidence that AF and AH can oligomerize. In the previous section, I have shown that neither the presence of TM segments of AF and AH, nor the presence of their partners is sufficient to promote their oligomerization *in vitro*. The soluble region of AF is only made of the RBM domain and no further investigation regarding the oligomerization of this protein could be performed *in vitro*.

By contrast, the structural organization of AG and AH is more complex. AG is made of a N-terminal disordered region (residues S51 to P89) followed by the RBM-like domain (residues K90 to S229). The RBM-like domain is split in two regions homologous to canonical RBM domains (residues K90 to S127 and residues K181 to S229, corresponding to the αββαβ fold forming the saucer region in the AG rings) and flanking a central β-triangle domain (residues L128 to K180; Fig. 30A). The saucer domain of AG aligns onto PrgK from *S. typhimurium* with a rmsd (root-mean-square deviation) of 1.075 Å (Fig. 30B). Similar to the EscJ/PrgK rings, oligomerization of the RBM core domain of AG involves hydrophobic and electrostatic interactions allowed by the packing of two helices from one RBM against the β-sheet of an adjacent RBM. The cup domain has no structural homologue so far and forms a unique 60-stranded antiparallel β-barrel upon oligomerization Fig. 30B and 17D).

AH displays a N-terminal unfolded region (residues A32 to D103) followed by an RBM-like domain that contains an additional N-terminal helix called the α0 helix (Fig. 30C). When AH is superposed onto one protomer of a PrgK dimer, the α0 helix clashes with the adjacent PrgK protomer (Fig. 30D). Therefore, the presence of the α0 helix in AH might prevent the formation of a PrgK-like oligomer.

In this section, I will present data performed on soluble constructs of AG and AH, aiming at determining which region(s) in AG promote(s) its oligomerization *in vitro*, and whether the N-terminal unfolded region of AH and/or the additional α0 helix of its RBM-like domain impair

oligomerization *in vitro*.

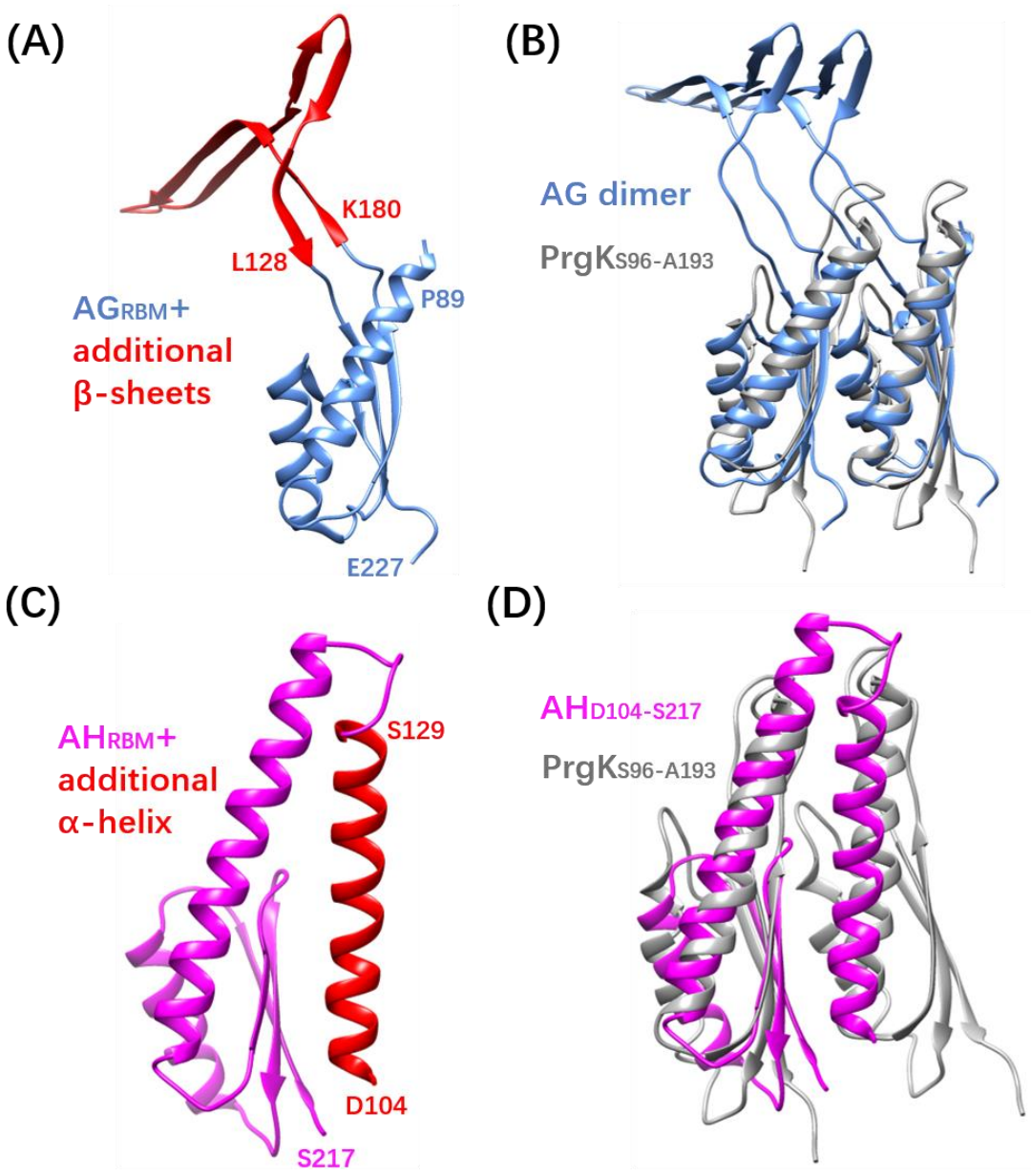


Figure 30. The RBM-like domains of AG and AH contain additional secondary structures compared to canonical RBMs. (A) Ribbon representation of the structure of the $AG_{S51\text{-}S229}$ construct, showing the RBM core in blue and the β-triangle insertion in red. **(B)** Superimposition of an AG dimer (PDB code 5WC3, in blue) onto a PrgK dimer (PDB code 6DUZ, in grey). **(C)** Ribbon representation of the structure of the $AH_{A24\text{-}K218}$ construct, showing the RBM core in magenta and the additional α0 helix in red (PDB code 3UZ0). **(D)** Superimposition of AH (PDB code 3UZ0, in magenta) onto a PrgK dimer (PDB code 6DUZ, in grey).

a. Purification and characterization of truncated soluble constructs of AG

To determine the role of the different soluble regions of AG in its oligomerization, we produced the following constructs:

- His-SUMO-$AG_{S511\text{-}S229(\Delta L128\text{-}K180)}$ (saucer region = soluble domain deleted from the cup region)

- His-SUMO-$AG_{L128\text{-}K180}$ (cup region = soluble domain deleted from the saucer region)

- His-SUMO-$AG_{K90\text{-}S229}$ (soluble domain deleted from the disordered N-terminal region)

As far as I am concerned, so far I only had time to produce, purify and characterize His-SUMO-$AG_{S51\text{-}S229(\Delta L128\text{-}K180)}$. Following cleavage of the His-SUMO tag, size-exclusion chromatography separated the $AG_{S51\text{-}S229(\Delta L128\text{-}K180)}$ sample into a large molecular weight species (eluting at about 10 mL) and several low molecular weight species (eluting at about 15-18 mL) (Fig. 31A). SDS-PAGE analysis of the high molecular weight species showed the presence of a single band below the 15-kDa molecular weight marker, corresponding to the monomeric form of $AG_{S51\text{-}S229(\Delta L128\text{-}K180)}$ denaturated by SDS (expected molecular weight of 13.7 kDa). By contrast with the $AG_{S51\text{-}S229}$ or His-SUMO-$AG_{M1\text{-}S229}$ constructs (Fig. 28A), no high-molecular weight species corresponding to oligomeric rings could be observed on the gel. In agreement with this observation, negative-stain EM analysis of the $AG_{S51\text{-}S229(\Delta L128\text{-}K180)}$ sample shows the presence of protein aggregates but no ring (Fig. 31C). The fractions eluting at higher volumes from the size-exclusion column contained multiple bands that migrated higher than the expected molecular weight of $AG_{S51\text{-}S229(\Delta L128\text{-}K180)}$, suggesting that they could be oligomeric forms of the protein (Fig. 31B). However, negative-stain analysis of these fractions did not show the presence of rings or small oligomers (data not shown). These multiple bands are therefore likely *E. coli* contaminants.

The His-SUMO-$AG_{K90\text{-}S229}$ construct was produced and purified a few years ago by the technician in our lab (Laure Bellard). Following cleavage of the His-SUMO tag, part of the $AG_{K90\text{-}S229}$ sample eluted in the void volume of the size-exclusion chromatography, suggesting

the presence of oligomeric rings or protein aggregates. Part of the protein eluted at higher volumes, indicating the presence of small molecular weight species. Negative-stain EM analysis of all these $AG_{K90\text{-}S229}$ samples did not show the presence of protein rings (Fig. 31D).

So far, the results thus indicate that the unfolded region of AG, as well as the cup region, are either required for folding of the soluble domain or for oligomerization of the RBM-like domain.

The His-SUMO-$AG_{L128\text{-}K180}$ construct will be produced and purified soon to determine whether the β-triangle region of AG is able to fold on its own and to form β-barrels in the absence of the RBM-homologous region.

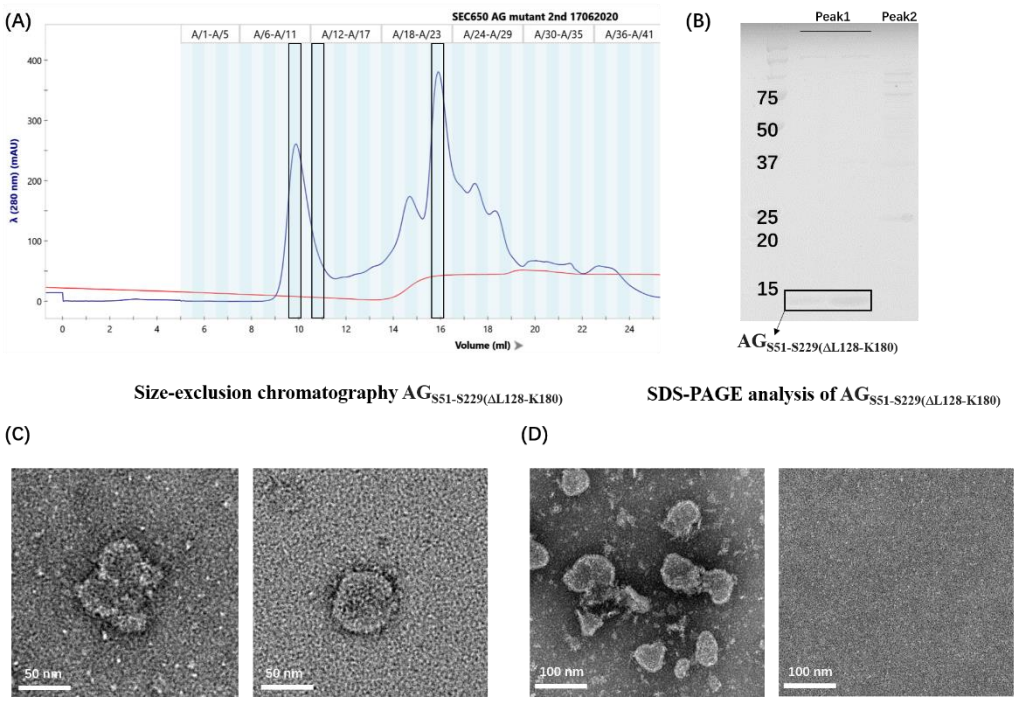


Figure 31. Analysis of recombinant truncated forms of AG. (A) Size-exclusion chromatography of $AG_{S51\text{-}S229(\Delta L128\text{-}K180)}$. A first peak (Peak 1) eluting at around 10 mL corresponds to large molecular weight species. Several other peaks eluting between 15 and 18 mL correspond to smaller molecular weight species. Fractions in boxes correspond to those analyzed by SDS-PAGE in panel B. **(B)** SDS-PAGE analysis of elution fractions from the size-exclusion chromatography. Peak 1 contains a single species migrating below the 15-kDa molecular weight marker (expected molecular weight of 13.7 kDa). Peak 2 (the most concentrated fraction eluting at about 16 mL) contains several species. **(C)** Negative-stain EM analysis of $AG_{S51\text{-}S229(\Delta L128\text{-}K180)}$ of the content of Peak 1 shows the presence of protein aggregates. No ring-shaped object was observed. Scale bar is 50 nm. **(D)** Negative-stain EM analysis of $AG_{K90\text{-}S229}$. The fraction eluted in the void volume of the size-exclusion chromatography column contains protein aggregates (left panel). The fraction eluted at around 15 mL contains small molecular weight species. No oligomeric ring was observed in both samples. Scale bar is 100 nm.

b. Purification and characterization of truncated soluble constructs of AH

To determine whether the disordered N-terminal region of AH or the additional α0 helix of the RBM-like domain prevent oligomerization of the protein *in vitro*, we produced the following constructs:

- His-SUMO-$AH_{S129\text{-}K218}$ (RBM core domain = soluble domain deleted from the N-terminal disordered region and the α0 helix)

- His-SUMO-$AH_{A24\text{-}S229(\Delta 24\text{-}128)}$ (soluble domain deleted from the α0 helix)

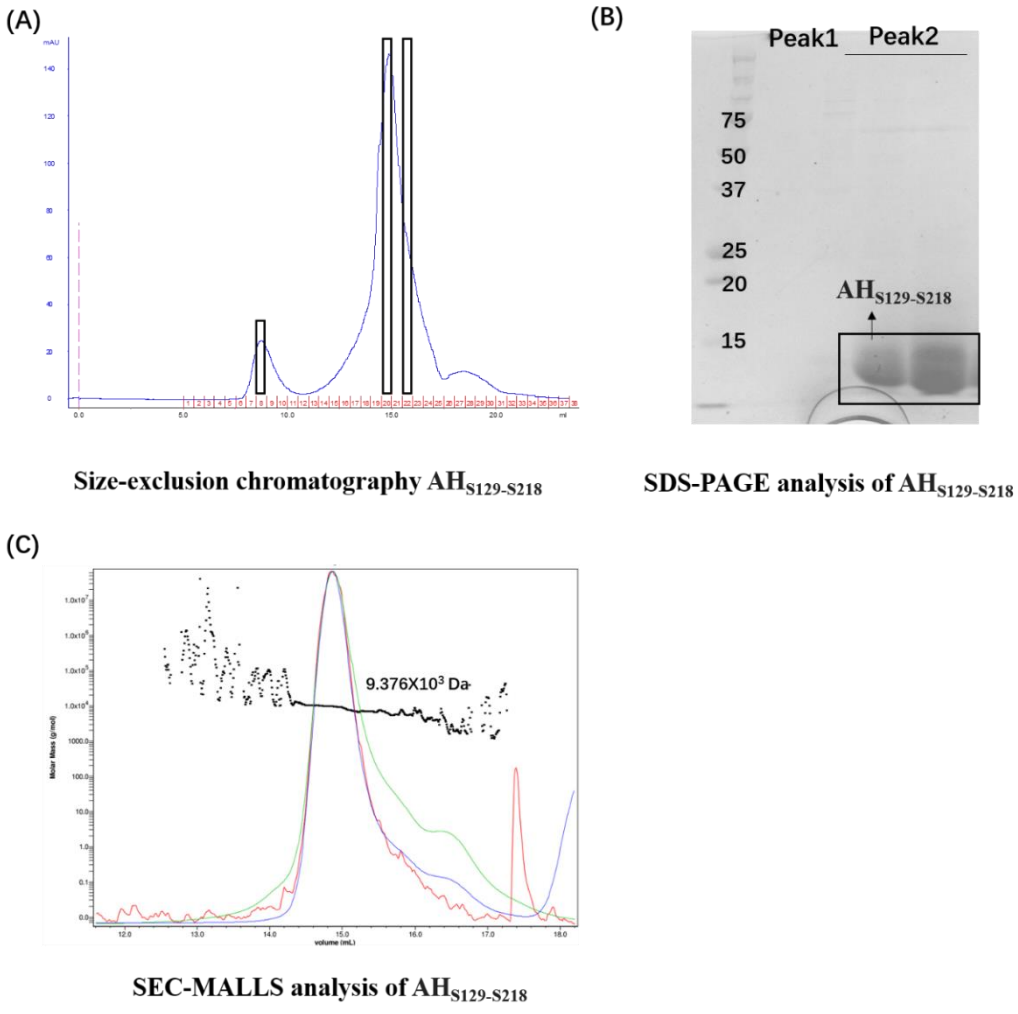


Figure 32. Analysis of truncated forms of AH. (A) The size-exclusion chromatography profile of $AH_{S129\text{-}S218}$ shows a first peak (Peak 1) eluting at around 8 mL (large molecular weight specie) and another peak eluting at about 15 mL (small molecular weight species). Fractions in boxes correspond to those analyzed by SDS-PAGE in panel **(B)**. **(C)** SEC-MALLS analysis of the oligomerization status of $AH_{S129\text{-}K218}$ in solution. Chromatograms are displayed with the absorbance at 280 nm as a blue line, the refractive index as a red line and the UV as a green line. Molecular weight estimation is displayed as a black line with values (in Da) displayed on the left axis.

Following cleavage of the His-SUMO tag, the $AH_{S129-K218}$ sample could be separated into a first species eluting in the void volume (about 8 mL) using a size-exclusion chromatography and a second species eluting at about 15 mL. (Fig. 32A). Content of the second elution peak migrates to the expected position below the 10-kDa molecular weight marker (theoretical molecular weight of 9.8 kDa) (Fig. 32B). SEC-MALLS analysis of the purified sample indicated that $AH_{S129-K218}$ is present as a monomer in solution (Fig. 32C) or as soluble aggregates. Indeed, the detected molecular weight of the fraction eluting at around 15 mL is 9.4 kDa (±7.1%), which is close to the theoretical molecular weight of 9.8 kDa. In support of this observation, no oligomeric ring species was detected when this fraction was concentrated and analyzed by negative-stain EM (data not shown). It thus seems that the isolated RBM core in AH is not able to form rings.

I will perform soon the purification and characterization of the $AH_{A24-S229(\Delta L105-V128)}$ construct.

D. Biophysical and structural study of YhcN

1. Context and personal contribution to the YhcN study.

Using the HHPRED server at the beginning of my Ph.D., Adriano Henriques from the ITQB in Portugal, had identified sporulation proteins other than A-Q proteins that possess predicted RBM domains. These proteins belong to the YhcN-YlaJ family, which so far includes the *yhcN*, *ylaJ*, *yutC* and *yrbB* (*coxA*) genes. The *yhcN* and *ylaJ* genes are under the control of the transcription factor σG and encode lipoproteins that are anchored in the inner forespore membrane and are involved in spore germination (Bagyan et al. 1998). In addition, both proteins have been detected in the spore outer layers (Kuwana et al. 2002) and a processed form of YhcN (encompassing residues N40 to E189) has been detected in material released from germinating *B. subtilis* spores(Chirakkal et al. 2002).

In the absence of *yhcN* or *ylaJ*, spores of *B. subtilis* germinate more slowly than the wild-type strain (Johnson and Moir 2017). This defect is even more pronounced in the double *ΔyhcN ΔylaJ* mutant. A closer examination of the germination defects indicated that loss of heat resistance (an early event in spore germination), as well as rehydration of the spore coat and

DPA release (two late germination events) were impaired in single and double mutants.

In this study, the authors used a *cotE* mutant (which displays defective coat assembly and increased permeability to spore germinants) to determine whether the germination defects of the *yhcN* and *ylaJ* mutants were due to poor permeability of the spore to germinants. Defective permeability of the spore was not the cause of impaired germination in *yhcN* and *ylaJ* mutants, but deletion of CotE in the absence of YhcN or YlaJ caused a further decrease in spore germination (Johnson and Moir 2017). Since CotE is required for proper coat assembly, itself required for proper localization of the cortex hydrolase CwlJ, the authors investigated the relationship between YhcN, YlaJ and the two cortex hydrolases, CwlJ and SleB.

Germination requires the hydrolytic activity of the two cortex lytic enzymes CwlJ and SleB (Setlow, Melly, and Setlow 2001). The study from Johnson and Moir showed that in the absence of CwlJ, either YlaJ or YhcN is required for spore germination. Indeed, double *ΔcwlJ ΔyhcN* and *ΔcwlJ ΔylaJ* mutants showed only slightly affected spore germination efficiency while the triple *cwlJ yhcN ylaJ* deletion reduced spore germination by 80% (Johnson and Moir 2017). Therefore, one of the two homologues, either YhcN or YlaJ, seems to be required for SleB function. In a *ΔcotE* mutant, in which deficient outer spore coat impairs the activity of CwlJ (which binds the outer spore coat), the loss of either YhcN or YlaJ is sufficient to decrease spore germination by 80%. In other words, *ΔcotE ΔyhcN* and *ΔcotE ΔylaJ* double mutants phenocopy the triple *ΔcwlJ ΔyhcN ΔylaJ* mutant. Importantly, SleB stability was not affected in all these mutants. YhcN and YlaJ thus have a role in SleB function, or in other words in cortex hydrolysis. However, since their inactivation also affect the loss of heat resistance and DPA release, these proteins also appear to have a role in earlier events of the germination process.

The exact function of YhcN and YlaJ, as well as the function or their two other homologues in *B. subtilis* (YutC and CoxA/YrbB), remain to be deciphered. As they are all predicted to possess an RBM domain, we wondered whether they might form some oligomeric ring-like structure involved in the transport of molecules from the forespore. To get first insights into the structure of these proteins, we studied the oligomeric behavior and structure of YhcN from *B. subtilis*.

I performed all the experiments described in the following sections, except those mentioned explicitly (mass spectrometry analysis by L. Signor, phase determination by C. Contreras-Martel and purification of His-SUMO-AF$_{M1-E206}$ by E. Bauda).

2. Purification and crystallization of YhcN

The 189-residue YhcN protein contains a N-terminal lipoprotein signal peptide (residues M1 to V23), followed by a region predicted to have no secondary structure (residues A24 to N77), the RBM-like domain (residues D78 to N150) and two more predicted C-terminal α-helices (from P151 to E189) (Fig. 33A).

To improve my chances to crystallize YhcN, I designed two different constructs. A His-SUMO-YhcN$_{A24-E189}$ construct contained the whole soluble region of the protein. If this construct would crystallize, it would allow getting the most complete structural information regarding this protein (Fig. 33B). In parallel, I designed a His-SUMO-YhcN$_{D78-E189}$ construct that would have better chances to crystallize if the A24-V23 region would indeed be disordered. Adriano Henriques (ITQB, Lisbon, Portugal) cloned these constructs and I took in charge all the rest of this project.

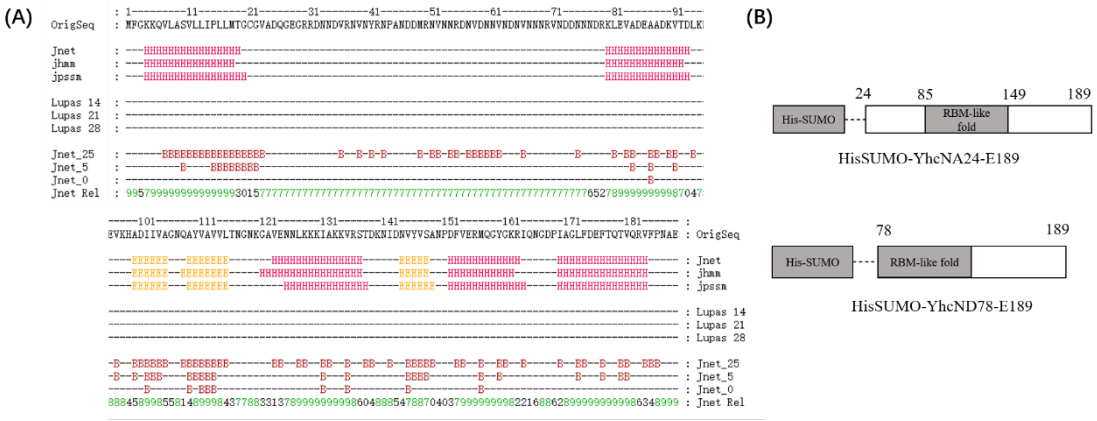
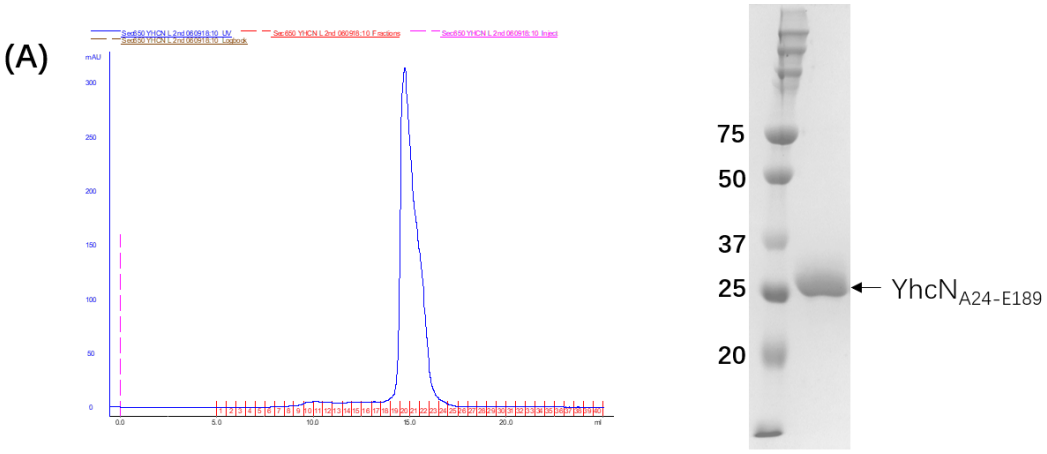
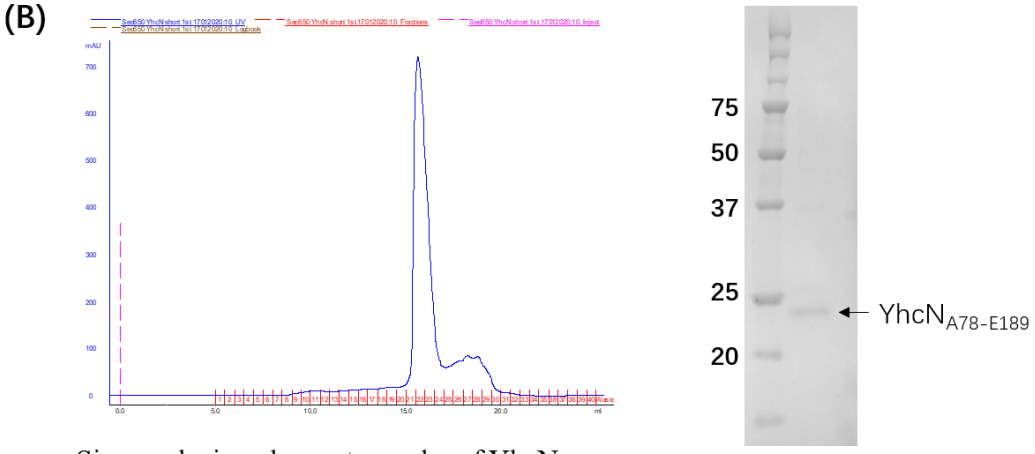


Figure 33. Design of YhcN constructs. (A) Secondary structure prediction of YhcN performed with the Jpred4 server (http://www.compbio.dundee.ac.uk/jpred/). The predicted α-helices are indicated with red "H" letters and the predicted β-strands are shown as yellow "E" letters. **(B)** Domain boundaries of the $YhcN_{A24-E189}$ (top scheme) and $YhcN_{D78-E189}$ (lower scheme) constructs.

Both constructs were purified through nickel affinity chromatography. The His-SUMO tag was then cleaved using the Ulp1 SUMO protease and the protein was passed again through a Ni-NTA resin, which retained the uncleaved fraction, the His-SUMO tag and the Ulp1 protease. The cleaved protein, recovered in the flow-through fractions, was finally purified by size-exclusion chromatography. The elution profiles of both $YhcN_{A24-E189}$ and $YhcN_{D78-E189}$ display a single peak, eluting at around 15 mL and 16 mL, respectively, and suggesting that both constructs are monomeric in solution (Fig. 34A, B). SDS-PAGE analysis showed that the purity of the sample was higher than 95% (Fig. 34A, B). Although the two proteins migrated lower than expected on a polyacrylamide gel (theoretical molecular weights are is 18.7 kDa and 12.4 kDa for $YhcN_{A24-E189}$ and $YhcN_{D78-E189}$, respectively), the experimental molecular weight determined for $YhcN_{A24-E189}$ by mass spectrometry (18.7 kDa, analysis performed by Lucas Signor from the MS platform facility at the IBS) showed that the major species did not suffer from proteolytic degradation (Fig. 34C).



Size-exclusion chromatography of YhcN$_{A24-E189}$



Size-exclusion chromatography of YhcN$_{A78-E189}$

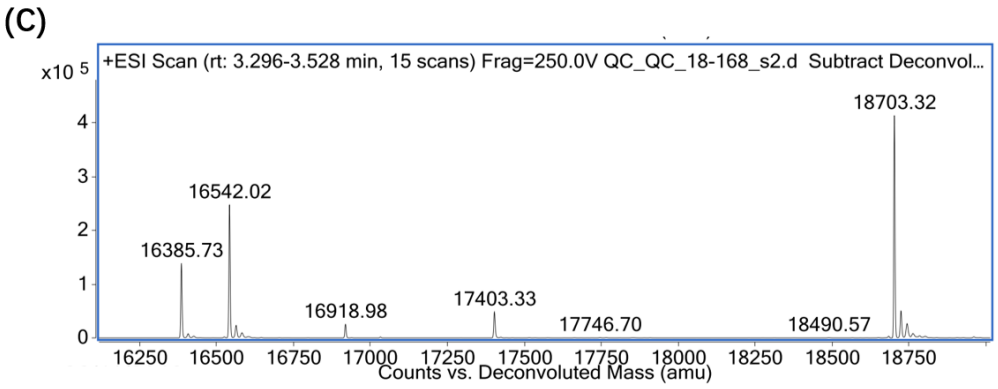
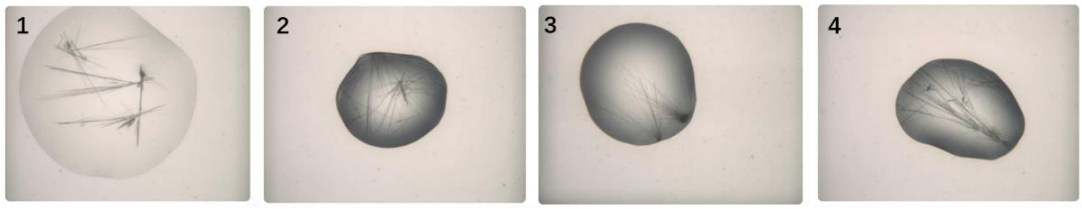


Figure 34. Purification of recombinant YhcN constructs. (A, B) Elution profiles of the YhcN$_{A24-E189}$ **(A)** or the YhcN$_{D78-E189}$ **(B)** construct from a Sec650 gel filtration column. The elution volume of the proteins suggest that they are monomeric in solution. On the right of these panels, Coomassie-stained SDS-PAGE gel shows the presence of a single species in each sample. **(C)** Purified YhcN $_{A24-E189}$ sample was analyzed by liquid chromatography-mass spectrometry (LC-MS), which showed a major species of 18.7 kDa.

Three different concentrations of a purified recombinant sample of $YhcN_{A24-E189}$ (12 mg/mL, 24 mg/mL and 48 mg/mL) were used to screen initial crystallization conditions using the high-throughput nano-crystallization platform of the EMBL (HTX lab, EMBL, Grenoble). Needle-shaped crystals appeared within 5 days in 14 different conditions, all containing a PEG polymer as the precipitant. Most crystals reached their full size after 15 days. I chose 4 conditions containing different PEG polymers (PEG 8 000, PEG 10 000, PEG 20 000 and PEG 3 350) and in which the crystals had reached a large size to reproduce and improve them myself (Fig. 35).

(A)



(B)

Group	Precipitant/ concentration	Buffer/ concentration	pH	Salt/ concentration	Protein concentration
1	PEG 8000/ 20% w/v	HEPES/ 0.1 M	7.5		24 mg/ml
2	PEG 10000/ 20% w/v	MES/ 0.1 M	6.5		24 mg/ml
3	PEG 20000/ 15% w/v	MES/ 0.1 M	6.5		12 mg/ml
4	PEG 3350/ 20% w/v			Sodium, Potassium phosphate/ 0.02M	24 mg/ml

Figure 35. Initial protein crystallization hits. (A) Images of four conditions in which $YhcN_{A24-E189}$ nano-crystals were obtained with different PEG polymers as the precipitation agent. Numbers correspond to the composition of the crystallization buffer indicated in panel (B). **(B)** Composition of the crystallization buffer corresponding to images shown in panel (A).

The crystals were reproduced manually using the hanging drop method and larger (µL scale) volumes (see *Material and Methods*). These 4 conditions were refined to improve the size of the crystals by tuning the pH of the buffer, the concentration of the precipitant, the concentration of the salt solution when appropriate and the volume ratio of the protein sample versus the crystallization solution (see *Material and Methods*).

The largest needle-shaped crystals we obtained after 14 days of incubation at 20°C, using HEPES 0.1 M, pH 7.5; 16 % (w/vol) PEG 8000 as the crystallization condition. Single YhcN crystals were then harvested, cryo-protected, flash-frozen into liquid nitrogen and tested for X-ray diffraction. Glycerol was used as cryo-protectant in a solution containing HEPES 0.1 M, pH 7.5 and 17 % (w/vol) PEG 8000.

3. Structural characterization of YhcN$_{A24-E189}$

X-ray diffraction data were collected to 1.4 Å resolution on beamline ID30A-3 at the ESRF in Grenoble. Following processing with the XDS program package (statistics provided in Table 5), we tried phasing using the molecular replacement method. YhcN$_{A24-E189}$ templates were selected according to the HHPRED predictions, including SpoIIIAH from *B. subtilis* (PDB entry 3UZ0, 18% sequence identity, 16% sequence similarity), SpoIIIAG from *B. subtilis* (PDB entry 5WC3, 12% sequence identity, 28% similarity), the second RBM domain of EscJ from *E. coli EPEC* T3SS (PDB entry 1YJ7, 10% sequence identity, 22% similarity), the second RBM domain of PrgK from *Salmonella typhimurium* (PDB entry 4OYC, 11% sequence identity, 18% similarity) and the EscU protease from *E. coli* T3SS (PBD entry 3BZS, 11% sequence identity, 18% similarity). Unfortunately, none of these templates provided a solution, likely due to poor sequence identity with YhcN.

Given the high resolution of our data and the presence of four predicted α-helices in YhcN structure, we sought to perform *ab initio* phase determination using the ARCIMBOLDO LITE program, which combines the location of model fragments like small α-helices with PHASER (McCoy et al. 2007) and density modification with SHELXE (Thorn and Sheldrick 2013). *Ab initio* phasing was achieved by Carlos Contreras-Martel from the Dessen lab, from the correct positioning of four helices. Using the *ab initio* phases, I performed semi-automated model building and refinement of YhcN$_{A24-E189}$ as described in the *Material and Methods* section. Statistics on model refinement are summarized in Table 5.

The high quality of the electron density map allowed building of the YhcN$_{A24-E189}$ model, which contains two molecules in the asymmetric unit. The main chain and side chains could be built from R79 to A179 but no electron density was visible for residues A24 to D78 (Fig. 36A), even when the non-crystallographic symmetries were used to build the model. Since mass spectrometry analysis of the purified protein showed no sign of proteolytic degradation, the absence of electron density for residues A24 to D78 indicates that this region is disordered. This observation is in agreement with the YhcN secondary structure prediction (Fig. 33A).

TABLE 5. Data collection and refinement statistics.

Data collection	
Name of dataset	4YhcN_A2_1
X-ray source	ID30A-3 (ESRF)
Wavelength (Å)	0.9677
Scan range (°)	250
Oscillation (°)	0.1
Space group	$P2_12_12_1$
Unit-cell parameters	
a, Å	31.48
b, Å	52.28
c, Å	134.13
α, °	90.00
β, °	90.00
γ, °	90.00
Number of molecules in ASU	2
Resolution (last shell), Å	1.44 (1.44-1.53)
Completeness, %	96.8 (87.5)
I/σ(I)	8.77 (0.64)
Rsym†, %	7.3 (168.3)
Unique reflections	39,371 (5,820)
Observed reflections [I/σ(I) > 1]	177,413 (20,052)
Wilson B factor, (Å²)	31.12
Refinement and model statistics	
Resolution (last shell), Å	1.77 (1.77-1.87)
R-factor‡, R-free§	0.207, 0.233
rmsd from target[i]	
Bond lengths, Å	0.009
Bond angle, °	1.54
Mean B factor (Å²)	41.5
Ramachandran plot**	
Core, %	94.3
Allowed, %	5.7
Disallowed, %	0

Values in parentheses are for the outermost shell of data.

†Rsym = (Σ(ABS(I(h,i)-(I(h)))) / (Σ(I(h,i))).

‡R-factor = ΣjjFoj - jFcjj/ΣjFoj where Fo and Fc are the observed and calculated structure factor amplitudes, respectively.

§R-free is the R-factor calculated with 5% of the reflections chosen at random and omitted from refinement.

[i]rmsd of bond lengths and bond angles from ideal geometry.

**Performed by Procheck.

YhcN is made of a three-stranded β-sheet (β1β2β3) sandwiched by two α-helices on one side (α1 and α2), and two other α-helices on the other side (α3 and α4) (Fig. 36C). Therefore, part of the globular domain of YhcN contains the founding elements of the RBM domains (α1β1β2α2β3) but two additional α-helices are present (α3 and α4). The YhcN crystal contained two molecules in the asymmetric unit. These two molecules align with an rmsd of 0.48 Å and no divergence was observed in the orientation of the two C-terminal α-helices or in the position of the loop connecting strand β3 to helix α3 (Fig. 36D). In addition, the B-factors of the atoms present in the loop connecting strand β3 to helix α3 are low in the two molecules of the asymmetric unit, indicating that this region is quite rigid. On the other hand, the B-factors of the atoms present in the loop connecting helix α3 to helix α4 are high, indicating a slight flexibility in the orientation of helix α4 (Fig. 36E). Altogether, these observations indicate that the position of helix α3 in the crystal is stable but that helix α4 displays more flexibility. In the cellular context, movement of helix α4 upon interaction with yet-to-be identified partners could induce the release of helix α3 and expose the RBM core domain. This hypothesis is discussed in the *Discussion* section IB.

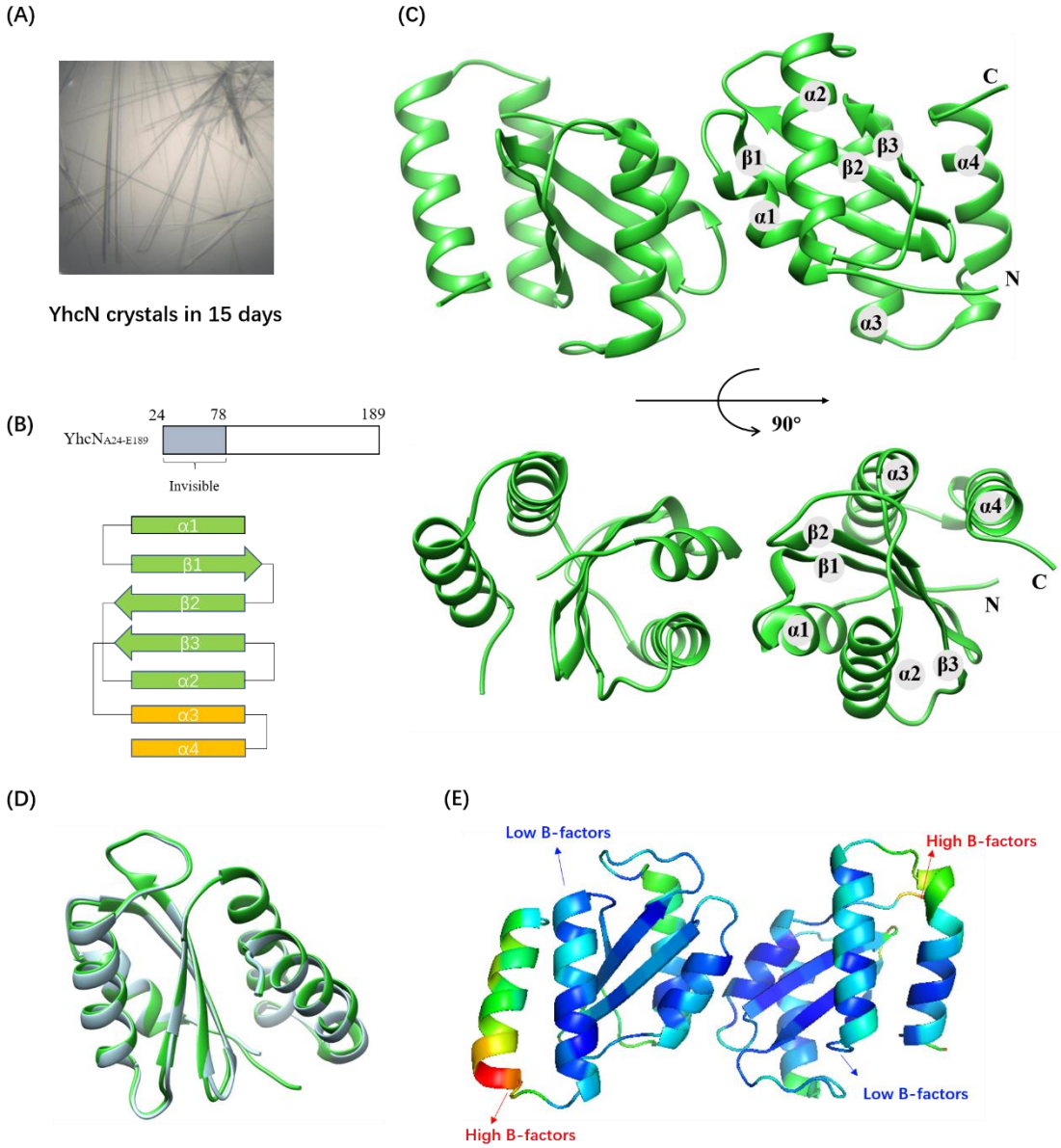


Figure 36. YhcN structure model. (A) Full size YhcN crystals obtained in 15 days in 100 mM HEPES pH 7.5; 16% (w/vol) PEG 8000. **(B)** Topology of $YhcN_{A24-E189}$. Residues A24 to D78 are invisible in electron density map. The rest of the protein displays a α1β1β2α2β3α3α4 topology, which includes an RBM core region (in green) and two additional C-terminal α-helices in (orange). **(C)** Refined YhcN crystal structure. Figures were generated with Chimera in two different orientations. **(D)** Overlay of Chain A (green) and Chain B (blue) of $YhcN_{A24-E189}$. **(E)** B-factor color-coded ribbon diagrams of the $YhcN_{A24-E189}$ dimer present in the asymmetric unit. The ribbon color changes from blue (rigid) to red (flexible) to represent a low to high residue flexibility. The low B-factors of α1β1β2α2β3α3 suggest that these domains are stable. The high B-factors at the beginning of helix α4 suggest that this helix displays more flexibility in the crystal, and could therefore be flexible in the cellular context.

The two YhcN molecules make hydrogen bonds listed in Tables 6 and 7, as well as other ionic interactions listed in Table 8 (analyzed through Protein Interaction Calculator (PIC) server,

http://pic.mbu.iisc.ernet.in/). The residues involved in the dimer interface are shown in figure 37A. With these contacts, the two molecules form a buried interface of 391.5 $Å^2$ (analyzed through the PISA service at the European Bioinformatics Institute.). The absence of hydrophobic interactions, as well as the small interface area suggest that this dimer results from crystal packing. In support of this hypothesis, SEC-MALLS analysis of the purified $YhcN_{A24-E189}$ sample (in collaboration with Caroline Mas from the ISBG biophysical platform, Grenoble) showed that the protein is monomeric in solution. Indeed, SEC-MALLS analysis provided an apparent MW of 21.0 ± 1.2 kDa, close to the experimental MW determined by mass spectrometry (18.7 kDa) (Fig. 37B). This analysis indicates that YhcN does not dimerize in solution and thus that the dimer observed in the crystal is likely not physiological.

Table 6. Main chain-side chain hydrogen bonds between YhcN chain A and chain B

Donor					Acceptor			
Position	Chain	Residue	Atom	Distance	Position	Chain	Residue	Atom
89	A	ASP	NZ	2.83	134	B	LYS	O
92	A	THR	NZ	2.97	90	B	LYS	O
94	A	LEU	NZ	3.01	90	B	LYS	O
97	A	VAL	NZ	2.73	90	B	LYS	O

Table 7. Side chain-side chain hydrogen bonds between YhcN chain A and chain B

Donor					Acceptor			
Position	Chain	Residue	Atom	Distance	Position	Chain	Residue	Atom
89	A	ASP	NZ	3.41	130	B	LYS	OD1
89	A	ASP	NZ	2.89	130	B	LYS	OD2
89	A	ASP	NZ	2.86	134	B	LYS	OD2
93	A	ASP	NZ	3.38	133	B	LYS	OD1
93	A	ASP	NZ	2.79	134	B	LYS	OD1

Table 8. Ionic interactions between YhcN chain A and chain B

Position	Chain	Residue	Position	Chain	Residue
89	A	ASP	130	B	LYS
89	A	ASP	133	B	LYS
89	A	ASP	134	B	LYS
93	A	ASP	133	B	LYS
93	A	ASP	134	B	LYS
98	A	LYS	89	B	ASP
99	A	HIS	93	B	ASP

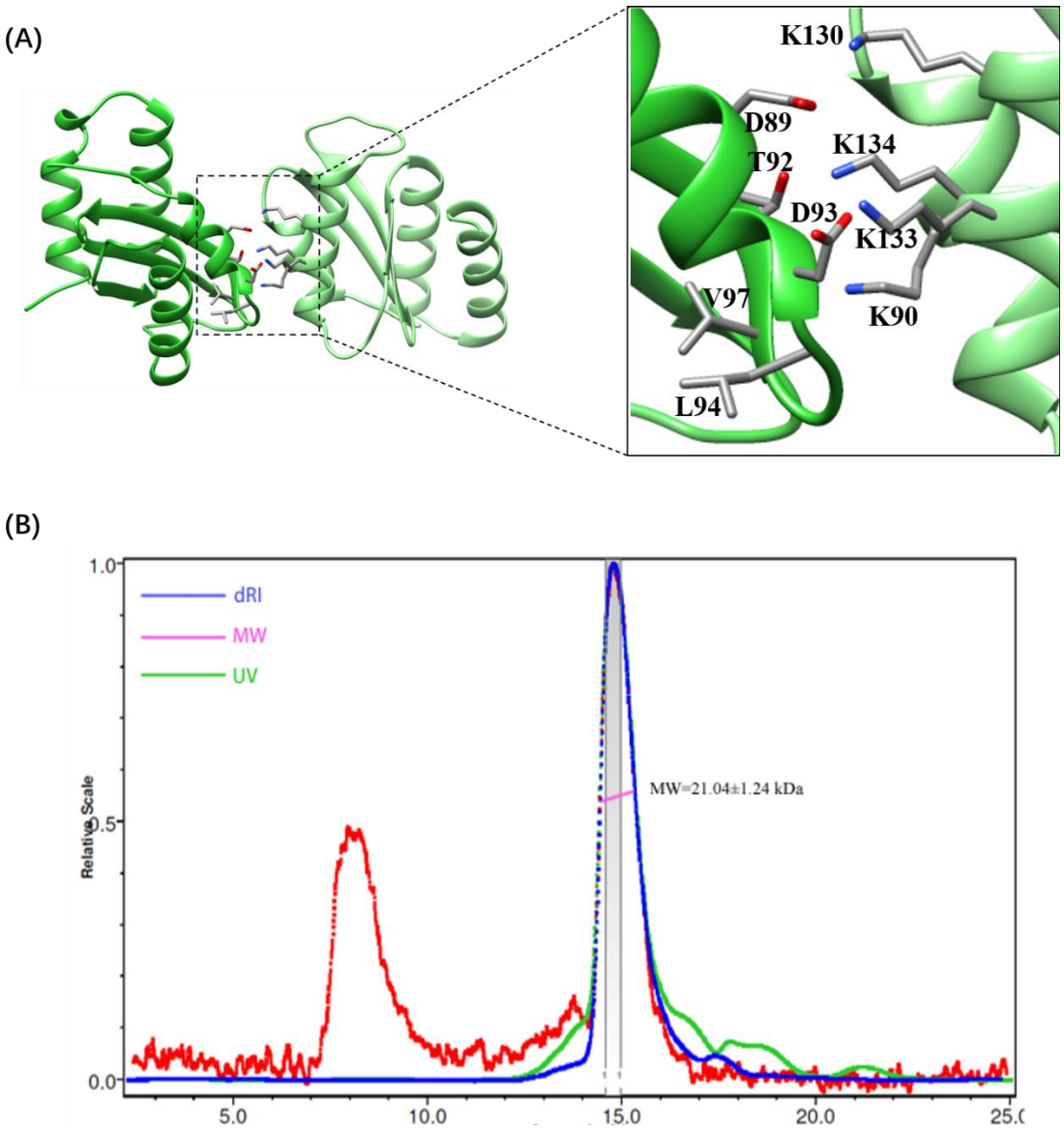


Figure 37. Analyses of the YhcN crystal dimer. Views of the YhcN chain A (in dark green) and chain B (in light green) interface with residues involved in hydrogen bonds labeled and shown as atom-colored sticks. **(B)** SEC-MALLS analysis of the oligomerization status of $YhcN_{A24\text{-}E189}$ in solution. Chromatograms are displayed with the absorbance at 280 nm as a green line (UV), the refractive index as a blue line (dRI) and arbitrary units displayed on the left axis. Molecular weight estimation is displayed as a pink line with the value (in kDa). SEC-MALLS analysis provided an apparent MW of 21.04 ± 1.24 kDa, close to the experimental MW determined by mass spectrometry (18.7 kDa).

4. Potential YhcN structural homologues

Interestingly, a search for structural homologues using the DALI server (Holm and Rosenstrom 2010) identified PrgK, EscJ, AH and AG as potential homologues of YhcN (Fig. 38A). Although the sequence identity between those proteins is very weak (11% for PrgK(RBM2) from *S. thyphimirium*, 10% for EscJ(RBM2) from *E. coli EPEC*, 18% for AH and 12% for AG from *B. subtilis*), the α1β1β2α2β3 RBM core of YhcN superimposes well onto the RBM core of those protein, with rmsds of 2.8 Å, 3.2 Å, 3.0Å and 3.0 Å, respectively (Fig. 38B-E).

Despite these intriguing structural similarities, there is so far no evidence indicating that YhcN is able to oligomerize like EscJ/PrgK or AG. First, the residues involved in the dimerization interface of PrgK or AG are not conserved in YhcN (Fig. 38B and 38E). In addition, when the YhcN model is superimposed onto one protomer of PrgK or AG dimers, the two C-terminal α-helices clash with the adjacent protomer (Fig. 38F-G). Therefore if YhcN forms rings *in vivo* in the conformation observed in the current crystal form, the oligomerization interface will likely diverge from the one observed in PrgK or AG rings. Alternatively, if the two C-terminal α-helices can move away from the RBM core, for example in the presence of yet-to-be identified YhcN partner(s), it remains possible that YhcN might oligomerize through a canonical RBM-RBM interface.

Interestingly, the DALI server also found structural similarities between YhcN and the magnetosome protein MamM (Fig. 38A, C). MamM belongs to the family of Cation Diffusion Facilitators (CDF), which are involved in the cellular homeostasis of metal cations (Zeytuni et al. 2014). MamM dimerizes to form a stable, V-shaped homodimer that is essential to its regulatory function regarding metal cations. The current hypothesis is that the cations is transported through the V-shaped MamM homodimer (Zeytuni et al. 2014). Although the MamM protomer resembles RBM domains, the β-sheet possesses an additional strand and the topology of the secondary structures differs from those observed in RBM domains and in YhcN. In addition, the structural alignment of YhcN dimer onto the MamM dimer shows that they share different dimerization interfaces (Fig. 38H). Furthermore, alignment of a YhcN monomer

onto the MamM dimer shows that the two C-terminal α-helices of YhcN clash with the dimerization interface of MamM. Altogether, these observations suggest that MamM is unlikely to be a homologue of YhcN.

(A)

Chain	Z-score	Rmsd	PDB identity	Description
4OYC-A	8.4	2.8	13%	PrgK from Salmonella enterica subsp T3SS
3W8G-A	7.6	3.3	13%	Manetosome protein MAMM
1YJ7-C	7.5	3.2	13%	EscJ from EPEC T3SS
3UZ0-A	6.5	3.0	13%	SpoIII AH of *B.subtilis*
5WC3-2	5.5	3.0	12%	SpoIII AG of *B.subtilis*

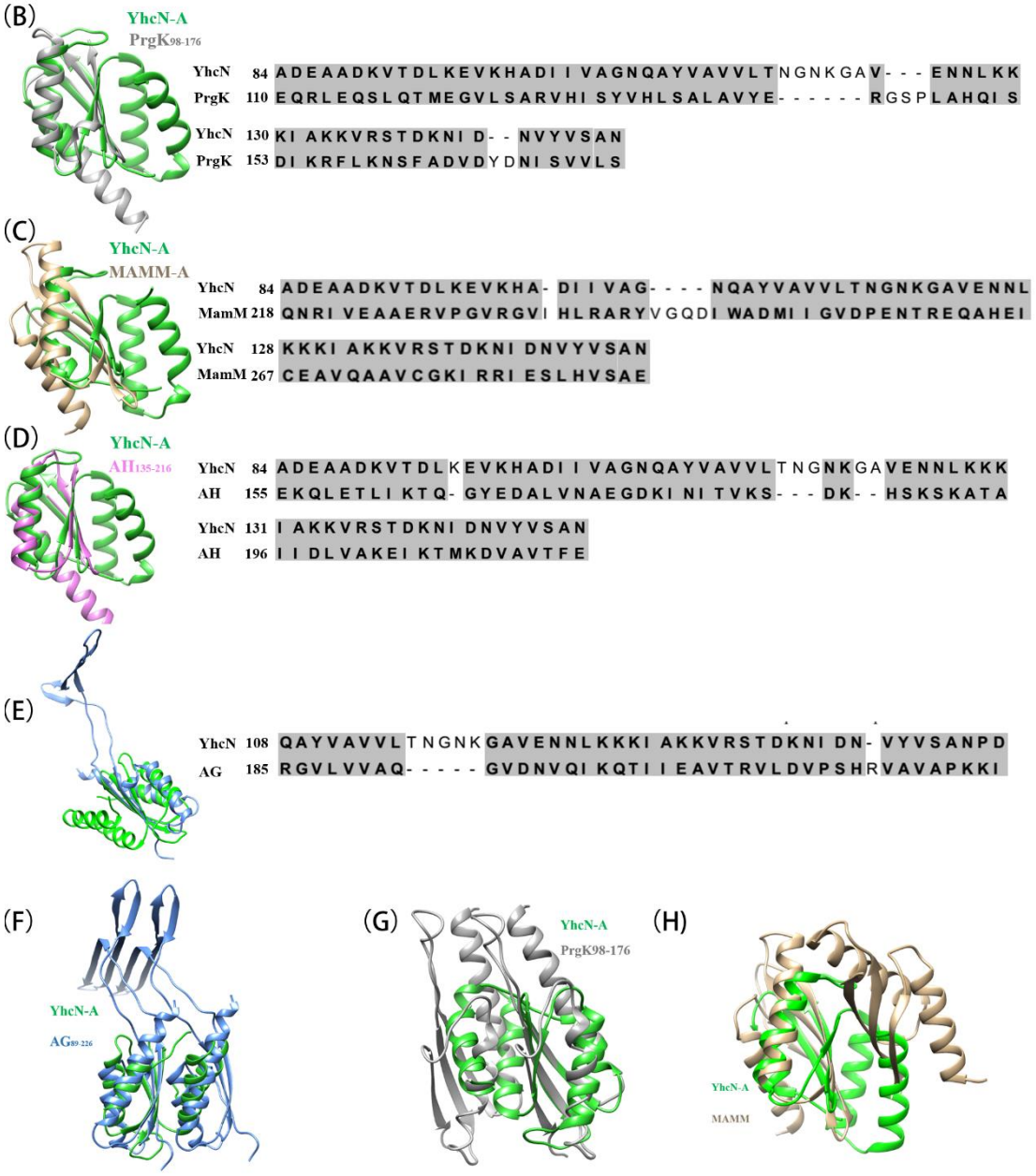


Figure 38. Analysis of potential YhcN homologues. (A) Potential homologues of YhcN identified by the Dali server. The PDB entry, Z-score, rmsd (in Å) and sequence identity are provided. **(B-E)** Left

panels show structural alignments of YhcN with potential homologues listed in panel A. Sequence alignments resulting from the structural overlays are shown on the right. Overlay regions are back-colored in grey. **(F)** Structural overlay of YhcN chain A against the AG dimer. The RBM-like domains of YhcN superpose well onto AG, but the two C-terminal α-helices clash with the AG oligomerization interface. (**G)** Structural overlay of YhcN chain A against a PrgK dimer. The N-terminus domains of YhcN superpose well onto PrgK, even if the first α-helix of PrgK is much longer than in YhcN. The two C-terminal α-helices of YhcN clash the PrgK dimerization interface. **(H)** Structural comparison of YhcN chain A against the MamM dimer. The RBM-like domain of YhcN superpose well onto one of the MamM molecule, but the two C-terminal α-helices clash with the MamM dimerization interface.

5. Biophysical characterization of truncated YhcN constructs

To determine whether the N-terminal disordered region of YhcN and the additional secondary structure elements of the RBM-like domain prevent YhcN oligomerization *in vitro*, I produced the following constructs:

- His-SUMO-YhcN$_{D78\text{-}N150}$ (RBM core domain = soluble domain deleted from the N-terminal disordered region and the C-terminal α-helices)

- His-SUMO-YhcN$_{A24\text{-}N150}$ (soluble domain deleted from the C-terminal α-helices)

The primers were obtained before the Covid-19 confinement. The two truncated constructs were then built when the lab opened again. Purification of these constructs is now under progress.

Discussion

Discussion

A. RBM-containing proteins unrelated to the A-Q complex

1. RBM domains found in specialized secretion systems

A thorough structural comparison of RBM domains found in specialized secretion systems has never been reported so far. Although the structure of the RBM domains discussed below are not the result of my own benchwork, it seemed important to me to compare these structures to point their similarities and differences. In addition, this structural analysis has allowed me to place my analysis of the RBM-like domains found in sporulation proteins (on which I performed benchwork) in a broader context.

The first structure of a ring-building motif (RBM) was solved in the early 2000's and was then associated with oligomerization of the EscJ/PrgK family of ring-forming proteins from the inner membrane platform of T3SS (Yip et al. 2005). These domains of about 100 residues are composed of two α-helices folding against a three-stranded β-sheet. A few years later, similar domains were discovered in the EscD/PrgH component of the IM platform and in the EscC/InvG components of the OM platform of T3SS (Spreter et al. 2009). In addition, they were identified in the GspD/PulD secretin component of T2SS and their structure was solved the same year (Korotkov et al. 2009, Spreter et al. 2009). Based on their structural similarity and their association with ring-forming proteins, these domains were then named ring-building motifs (Spreter et al. 2009). Since then, many structures of RBMs from T2SS or T3SS were solved by X-ray crystallography or cellular cryo-electron microscopy (cryo-EM) (Table 9) and there is no doubt today that these domains are indeed found in ring-forming proteins from specialized secretion systems.

Table 9. Reported structures of RBM domains from specialized secretion systems

PDB entry (oligo. status)	Species	Protein name (SS localization)	Construct	RBM content	Method	Resol.
1YJ7 (4-mer)	*E. coli* EPEC	EscJ (T3SS IMP)	20-190	RBM1-2	X-ray	1.8
3GR0 (4 molecules)	*S. typhi*	PrgH (T3SS IMP)	177-362	RBM1-3	X-ray	2.3
3GR1 (8 molecules)	*S. typhi*	PrgH (T3SS IMP)	170-392	RBM1-3	X-ray	2.8
3GR5 (monomer)	*E. coli* EPEC	EscC (T3SS secretin)	21-173	RBM 1-2	X-ray	2.05
2Y9J (24-mer ring)	*S. typhi*	PrgK-PrgH (T3SS IMP)	PrgK:21-190 PrgH:177-362	RBM1-2 RBM1-2	EM	6.4

PDB entry (oligo. status)	**Species**	**Protein name (SS localization)**	**Construct**	**RBM content**	**Method**	**Resol.**
2Y9K (15-mer ring)	*S. typhi*	InvG (T3SS secretin)	34-170	RBM1-2	EM	8.3
4G1I (2 molecules)	*S. typhi*	PrgH (T3SS IMP)	172-369	RBM1-3	X-ray	1.85
3J1W (24-mer ring)	*S. typhi*	PrgH (T3SS IMP)	14-119	CTD	EM	11.7
3J1X (24-mer ring)	*S. typhi*	PrgH (T3SS IMP)	173-363	RBM1-3	EM	11.7
4G2S (6-mer)	*S. typhi*	PrgH (T3SS IMP)	14-119	CTD	X-ray	1.86
3J1V (15-mer ring)	*S. typhi*	InvG (T3SS secretin)	34-173	RBM1-2	EM	11.7
4G08 (monomer)	*S. typhi*	InvG (T3SS secretin)	34-173	RBM1-2	X-ray	1.8
2MKY (monomer)	*S. typhi*	PrgK (T3SS IMP)	19-76	RBM1	NMR	
4OYC (2 molecules)	*S. typhi*	PrgK (T3SS IMP)	94-176	RBM2	X-ray	2.6
4W4M (14 molecules)	*S. typhi*	PrgK (T3SS IMP)	19-79	RBM1	X-ray	3.2
3J6D (24-mer ring)	*S. typhi*	PrgK (T3SS IMP) PrgH	PrgK:19-189 PrgH:173-363	RBM1-2 RBM1-3	EM	11.7
5TCP (24-mer ring)	*S. typhi*	PrgK (T3SS IMP) PrgH	PrgK:20-203 PrgH:171-364	RBM1-2 RBM1-3	EM	4.3
5TCQ (15-mer ring)	*S. typhi*	InvG (closed) (T3SS secretin)	172-557	RBM3 + secretin	EM	3.6
5TCR (15-mer ring)	*S. typhi*	PrgK (T3SS IMP) PrgH InvG (closed) (T3SS secretin)	PrgK:20-203 PrgH:171-364 InvG:172-557	RBM1-2 RBM1-3 RBM3 + secretin	EM	6.3
6DV3 (15-mer ring)	*S. typhi*	InvG (open) (T3SS secretin)	34-557	RBM1-3 + secretin	EM	4.1
6DV6 (15-mer ring)	*S. typhi*	InvG (T3SS secretin)	176-557	RBM3 + secretin	EM	3.9
6DUZ (24-mer ring)	*S. typhi*	PrgK (T3SS IMP) PrgH	PrgK:20-203 PrgH:171-364	RBM1-2 RBM1-3	EM	3.6
6PEM (24-mer ring)	*S. typhi*	PrgK (T3SS IMP) PrgH Spa PQR InvG (T3SS secretin)		RBM1-2 RBM1-3 RBM1-2	EM	3.5

PDB entry	Organism	Protein name	Construct	RBM content	Method	Resolution
6HCG (15-mer ring)	*Klebsiella pneumoniae*	PulD (T2SS secretin)	27-652	RBM1-4	EM	4.3
6GYB (14-mer ring)	*Xantho. citri*	VirB7 (T4SS)	22-132		EM	3.3
5W68 (15-mer ring)	*E. coli*	GspD (T2SS secretin)	282-668	RBM4 + secretin	EM	3.3
5WLN (15-mer ring)	*Pseudomonas aeruginosa*	XcpQ (T2SS secretin)	174-613	RBM3-4	EM	3.04
5WQ7 (15-mer ring)	*E. coli* K12	GspD (T2SS secretin)	99-617	RBM2-4 + secretin	EM	3.04
5WQ8 (15-mer ring)	*Vibrio cholerae*	GspD (T2SS secretin)	97-646	RBM2-4 + secretin	EM	3.26
5ZDH (15-mer ring)	*E. coli* ETEC	GspD (T2SS secretin)	100-643	RBM2-4	EM	3.2
3OSS (monomer)	*E. coli* ETEC	GspD (T2SS secretin)	3-165	RBM1-2	X-ray	2.63
3EZJ (2 dimers)	*E. coli* ETEC	GspD (T2SS secretin)	3-235	RBM1-3	X-ray	2.8
4JTM (dimer)	*E. coli* ETEC	GspD (T2SS secretin)	1-80	RBM1	X-ray	1.4
6I1Y (15-mer ring)	*Vibrio vulnificus*	EpsD (T2SS secretin)	97-649	RBM2-4	EM	3.4
6I1X (15-mer ring)	*Aeromonas hydrophila*	ExeD (T2SS secretin)	97-620	RBM2-4	EM	3.7

SS, secretion system; *S. thyphi*, Salmonella typhimurium; *Xantho. citri*, *Xanthomonas citri*; IMP, inner membrane platform; OMP, outer membrane platform. The oligomeric status of the supra-molecular assemblies are indicated as "XX-mer". When molecule association results from crystal packing, the number of molecules in the asymmetric unit is indicated as "XX molecules".

In specialized secretion systems all the RBM domains characterized so far are engaged in macromolecular ring assemblies *in vivo* but their oligomerization features display a great diversity. First, the interacting residues are so poorly conserved that it is impossible to predict which ones are involved in the oligomerization interface. Therefore, their identification requires solving the structure of the rings that they form. Second, the number of RBM domains vary from one family of protein to another. Ring-forming proteins from specialized secretion systems contain one to four RBM domains. Here again the conservation of interacting residues is poor between RBMs of the same protein. Third, the orientation of one RBM protomer relative to the adjacent protomer varies between different RBM rings, even within the same protein. Many RBM interfaces are established by contacts between the two α-helices from one protomer and the three-stranded β-sheet from the adjacent molecule and this interaction conformation can thus be considered as the canonical one (**αα/βββ conformation**, Fig. 39, 40 and 41). However, other RBM domains interact through different secondary structures (Fig. 39, 40 and 41). Finally, the number of protomers present in rings formed by RBM domains is variable, and ranges so far from 15 protomers in InvG rings to 24 protomers in PrgK or PrgH rings (Hu et al. 2018, Bergeron et al. 2015, Zeytuni et al. 2017).

a. RBM domains from PrgK and PrgH from T3SS

The two stacked rings formed by PrgK and PrgH in the IM platform of *S. typhimurium* T3SS provides a good example of the diversity displayed by the interfaces of RBM oligomeric rings (Fig. 39A-C). Indeed, PrgK is composed of two RBM domains while PrgH consists of three (Fig. 4C, 4D and Fig. 39D,39H). In addition, RBM protomers display different orientations in the different rings : in PrgH, oligomerization of RBM1 is mediated by helix α1 of one protomer interacting with helix α2 and strand β3 of the neighboring protomer, oligomerization of RBM2 involves helix α1 of one protomer and the β-sheet of the adjacent protomer, while RBM3 oligomerizes through the packing of the two α-helices of one protomer against the β-sheet of the adjacent molecule (**αα/βββ conformation**, Fig. 39E-G). In PrgK, oligomerization of RBM1 is mediated by helix α1 of one protomer interacting with helix α2 of the neighboring protomer while RBM2 oligomerizes through contacts between the two α-helices of one protomer and the β-sheet of the adjacent molecule (Fig. 39I-J), similar to the oligomerization of RBM3 from PrgH. In other words, the only similar oligomerization interfaces are the RBM3-RBM3 interface from PrgH and the RBM2-RBM2 interface from PrgK. This observation suggests that these **αα/βββ interfaces** might be the most stable ones.

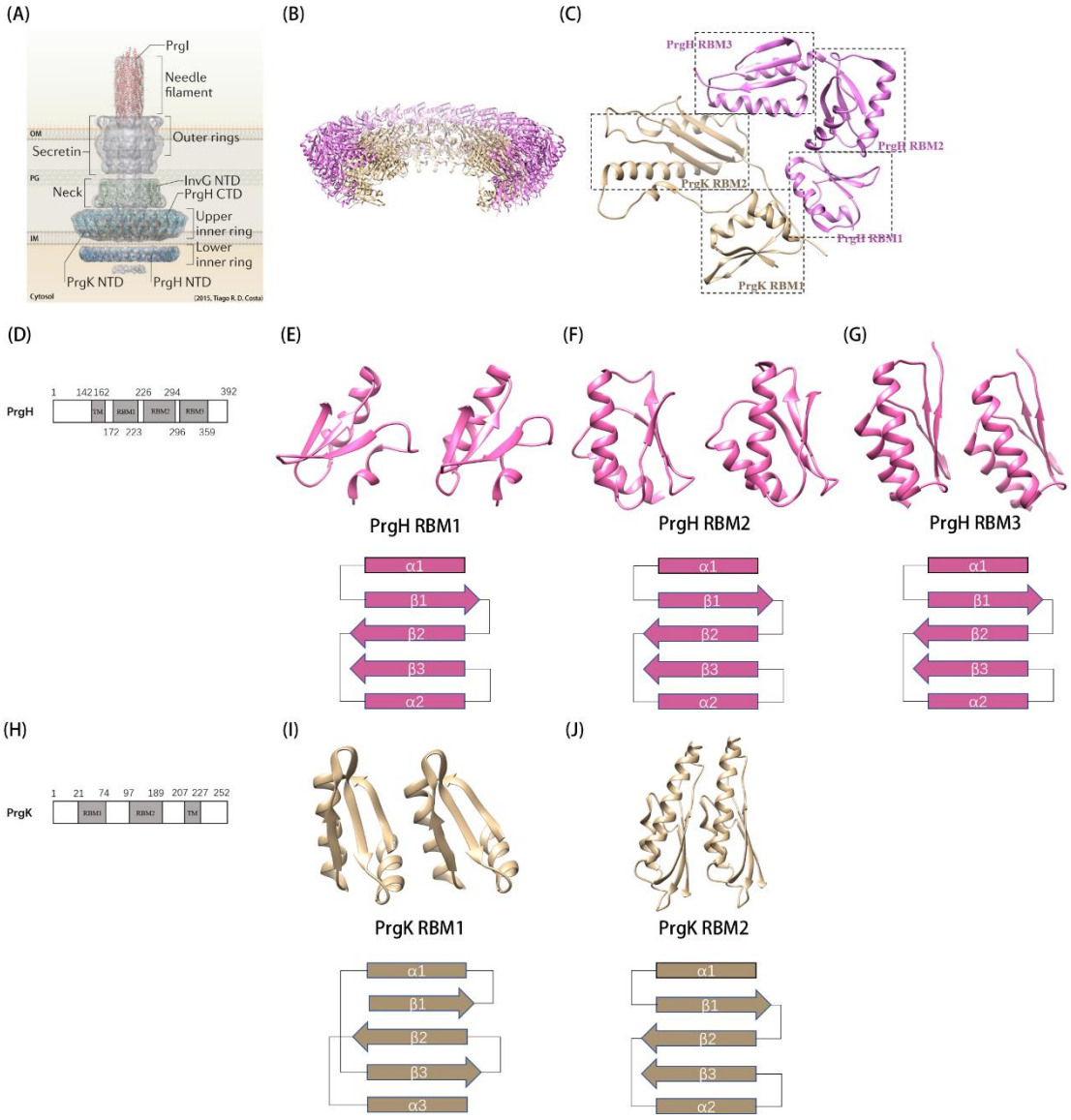


Figure 39. RBM domains and their oligomerization interfaces in PrgK-PrgH rings of the inner membrane platform in T3SS. (A) Schematic diagram of T3SS. Adapted from Costa et al., 2015. **(B)** Ribbon representation of the stacked 24-mer rings of PrgK (in light brown) and PrgH (in pink) in the T3SS from *S. typhimurium* (PDB code 6DUZ). **(C)** Ribbon representation of a ring unit containing a PrgK and PrgH protomer. RBM domains are individually boxed with dotted lines and labeled. **(D)** Scheme of the domain composition and boundaries in PrgH from *S. typhimurium*. **(E-G)** Adjacent protomers in rings made by RBM1 (E), RBM2 (F) and RBM3 (G) of PrgH display variable orientations but a similar αββαβ arrangement. **(H)** Scheme of the domain composition and boundaries in PrgK from *S. typhimurium*. **(I-J)** Adjacent protomers in rings made by RBM1 (I) and RBM2 (J) of PrgK display different orientations and different arrangements. RBM1 displays a βαβαβ topology while RBM2 shares the αββαβ topology of RBM1-3 from PrgH.

b. RBM domains from the InvG secretin from T3SS

The InvG secretin family of T3SS contains three periplasmic RBM domains in the N-terminal region of the protein while the secretin domain anchored in the OM is C-terminal (Fig. 40B). By contrast with RBM1-2 from PrgK and RBM1-3 from PrgH, which display a kinked conformation, RBM1-3 of InvG display a rather straight conformation that allows them to extend through the periplasmic space (Fig. 40A) (Hu et al. 2018). In InvG, RBM2-RBM2 and RBM3-RBM3 dimers interact through the packing of two α-helices of one protomer against the β-sheet of the neighboring molecule (αα/βββ conformation, Fig. 40D-E). This oligomerization packing is similar to those observed in the RBM3 ring from PrgH and the RBM2 ring from PrgK. In the InvG RBM1 rings, the protomers show a different orientation and the interface is mainly established by interactions between the α-helices of neighboring protomers (Fig. 40C). Very interestingly regarding my work on sporulation proteins, the RBM1 domain has an additional two-stranded β-sheet packed against the two α-helices (Fig. 40C). These additional secondary structures might be the reason why RBM1 domains orient differently within the rings and display different interfaces. In RBM3, the two additional β-strands are positioned away from the RBM core and therefore do not shield the canonical αα/βββ interface (Fig. 40E).

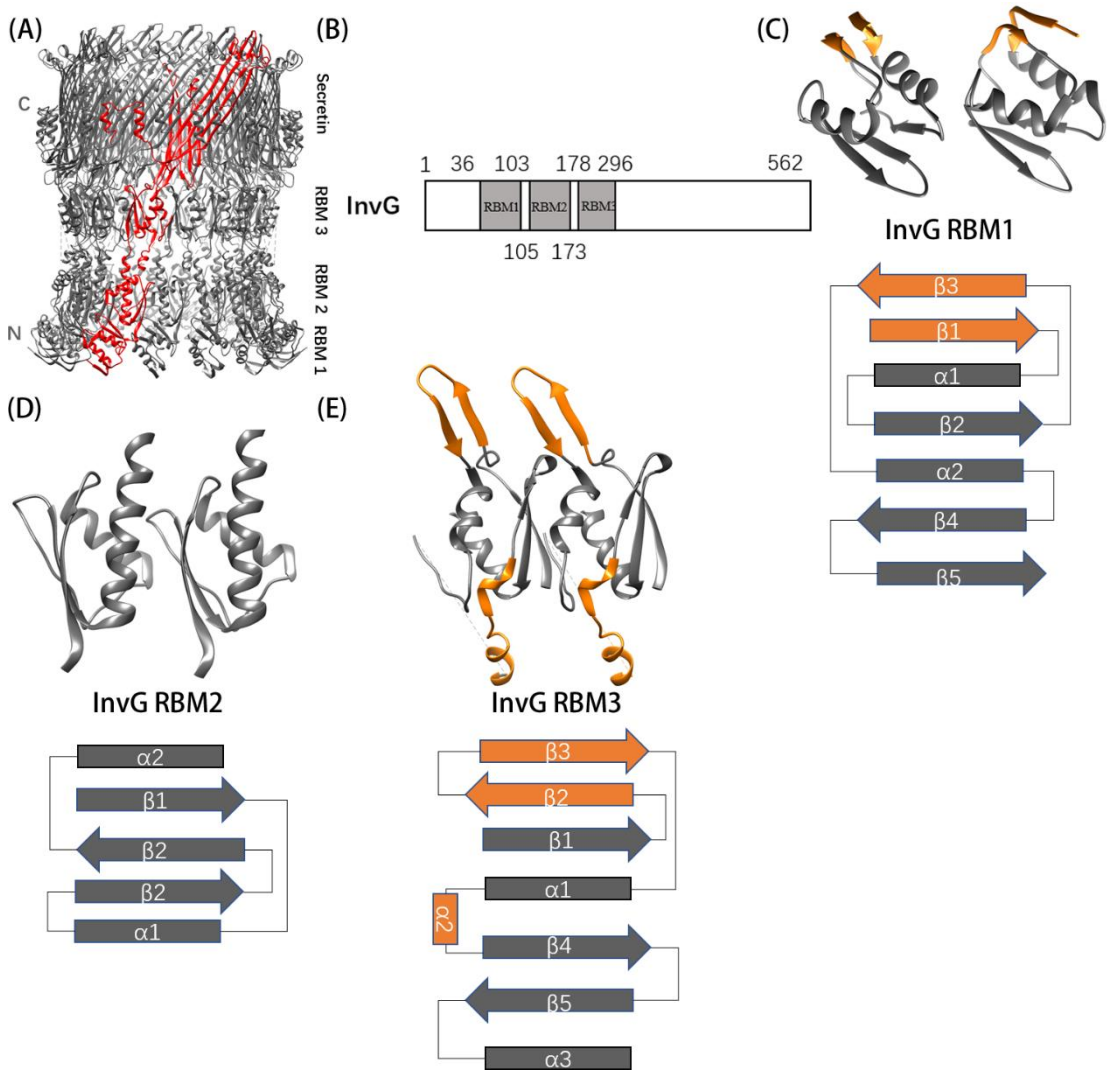


Figure 40. RBM domains and their oligomerization interfaces in the InvG component of the outer membrane platform in T3SS. (A) The InvG component of T3SS from *S. typhimurium* forms a cylinder-like assembly containing 15 protomers (PDB code 6DV3). One protomer is labeled in red. **(B)** Domain composition and boundaries in InvG from *S. typhimurium*. **(C-E)** Adjacent protomers in rings made by RBM1 (C), RBM2 (D) and RBM3 (E) of InvG. RBM2 and RBM3 rings display the canonical αα/βββ interface while the RBM1 ring displays an α/α interface.

c. RBM domains from the GspD secretin from T2SS

Similar to the InvG secretin ring of T3SS, the GspD secretin ring of T2SS also contains 15 protomers (Yan et al. 2017). The two proteins contain a C-terminal secretin domain but GspD possesses four N-terminal RBM domains (Fig. 41C). Like in InvG, the conformation of the RBM domains is rather linear, allowing them to extend in the periplasm (Fig. 41A, B). RBM1 domains interact through hydrophobic contacts between strand β2 of one protomer and strand β3 of the neighboring molecule (Fig. 41D) (Korotkov, Delarosa, and Hol 2013). Interestingly, the RBM1 domain contains two additional β-strands (β4 and β5) compared to RBM2-4 (Fig. 41D). These additional secondary structures might be the reason why this domain does not oligomerize through the canonical αα/βββ interface. The RBM2-4 domains of GspD have a βαββα topology and their oligomerization interfaces display the canonical αα/βββ orientation found in RBM2 from PrgK, RBM3 from PrgH and RBM2-3 from InvG (Fig. 31E-G).

Like in all the characterized canonical αα/βββ RBM interfaces so far, the RBM4-RBM4 interface in GspD mainly involves hydrophobic interactions between the two α-helices from one protomer and the three-stranded β-sheet from the adjacent one (Fig. 41H). However, the RBM4-RBM4 interface is additionally stabilized by hydrogen bonds established between loops from the RBM4 domain and from the secretin β-barrel (Fig. 41H).

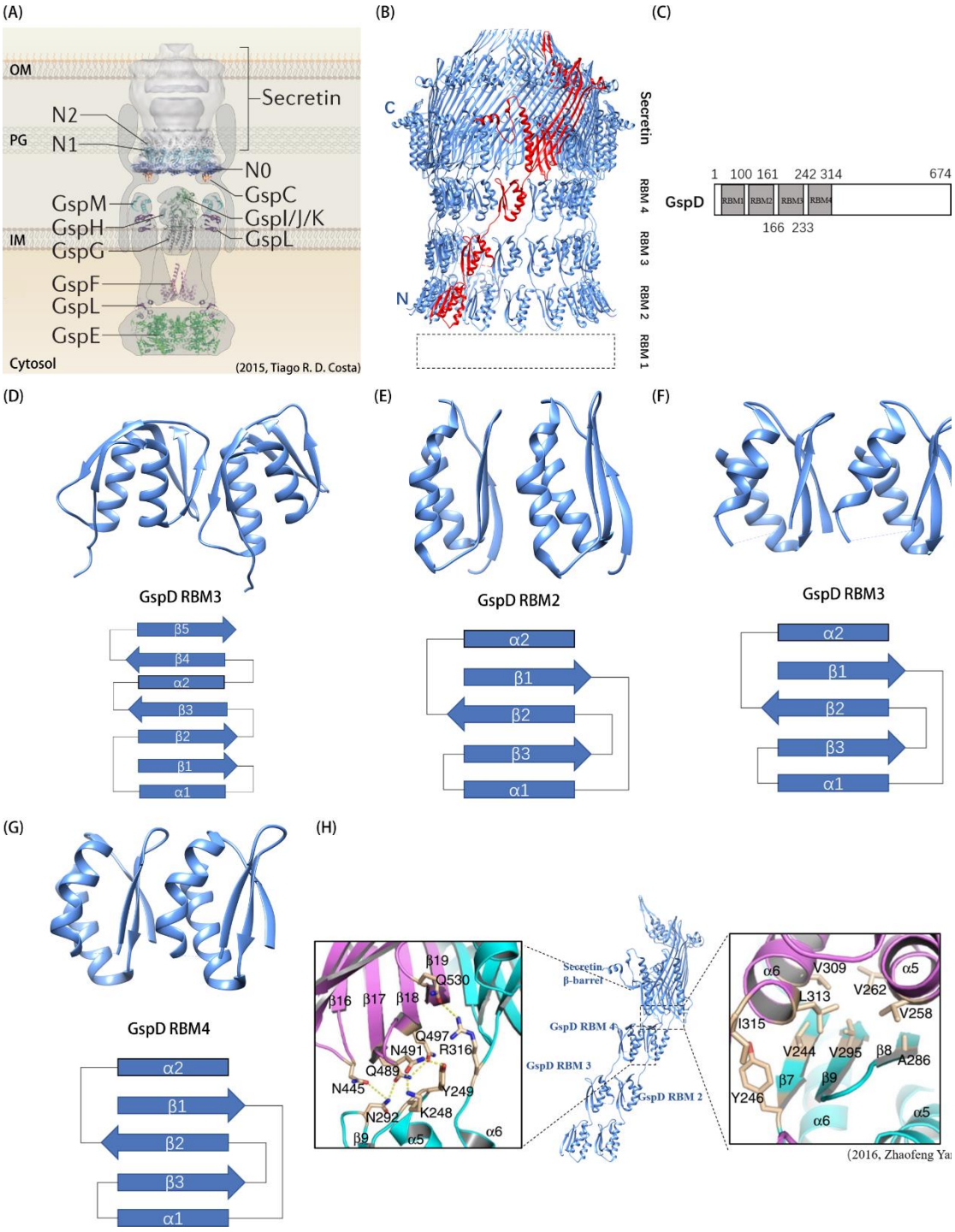


Figure 41. RBM domains and their oligomerization interfaces in GspD component of the outer membrane platform in T2SS. (A) Scheme diagram of the T2SS. Adapted from Costa et al., 2015. **(B)** The GspD component of the T2SS from *V. cholerae* forms a cylinder-like assembly containing 15 protomers (PDB code 5WQ8). One protomer is labeled in red. **(C)** Scheme of the domain composition and boundaries in GspD from *V. cholerae*. **(D-G)** Adjacent protomers in rings made by RBM1 (D, PDB code 4JTM), and by RBM2 (E), RBM3 (F) and RBM4 (G) (PDB code 5WQ8) of InvG. RBM2, RBM3 and RBM4 rings display the canonical αα/βββ interface while the RBM1 ring displays a β/β interface. **(H)** Oligomerization interfaces between RBM4 protomers (left) and RBM4-secretin β-barrel (right). Residues involved in the interfaces are shown as sticks in the boxes. Adapted from Yan et al., 2016.

The variety described above in the oligomerization interface of RBM domains could reflect a yet-unexplored evolutionary relationship between them. In my thesis project, I contributed to enriching the structural knowledge of RBM-like domains through the structural characterization of sporulation proteins that contain RBM-like domains that diverge from those found in specialized secretion systems. A first example is discussed below with the sporulation protein YhcN.

2. The RBM-like domain of YhcN

YhcN is wide spread in *Bacilli* and *Clostridia* species. In *B. subtilis*, the *yhcN* gene was initially identified in 1996 (Noback et al. 1996). However, there is no significant sequence similarity between YhcN and any other protein with known function. Through secondary structure prediction, we found out that YhcN shows similarity with RBM-containing proteins such as the A-Q components AF, AG and AH and the EscJ/PrgK family of proteins from T3SS.

The structure of YhcN has confirmed the presence of a region similar to the canonical RBM core (αββαβ fold), but the RBM-like domain of this protein contains two additional C-terminal α-helices (see section **IV** of the *Results* chapter). The asymetric unit in the crystal contains two molecules, which interact through a few electrostatic contacts. However in solution, no YhcN dimer was observed by SEC-MALLS. Therefore, the interactions observed between the two YhcN molecules present in the asymmetric unit likely result from crystal packing. To date, we have thus no evidence that YhcN can oligomerize like other RBM-containing proteins.

We cannot exclude that oligomerization of YhcN requires the two additional α-helices to move away from the RBM core. This idea will be tested through the characterization of the His-SUMO-YhcN$_{A24\text{-}N150}$ construct. Such scenario could happen *in vivo* upon interaction of YhcN with putative partners.

Alternatively, it is possible that the αββαβ fold in YhcN shares structural similarity with RBMs found in components of T2SS and T333 but have a different function. For example, it could be involved in protein-protein interactions with yet-to-be identified partner(s).

The functional characterization of YhcN, which is currently performed by Adriano Henriques (ITQB, Portugal) and Christopher Rodrigues (University of Sydney, Australia), will help unraveling the function of the RBM-like domain of this protein.

So far, deletion of the *yhcN* gene has not revealed any defect in heat resistance (heat-kills) morphology of the spore, coat assembly (localization of CotE, SafA, SpoIVD, SpoIVA and SpoVM) and activation of σ factors.

The YhcN gene was detected in both the forespore inner membrane and outer layers (Kuwana et al. 2002) but its localization around the forespore remains unknown. To investigate this, we will localize YhcN using a C-terminal mCherry fusion protein. Localization of truncated variants of YhcN (missing the unfolded N-terminal region and/or the additional C-terminal α helices) will also performed to determine which regions is involved in its localization.

A previous study has shown that YhcN can influence spore germination by affecting the activity of the cortex hydrolase SleB (Johnson and Moir 2017, Bagyan et al. 1998), but this mechanism is still unknown. There are two partially redundant cortex hydrolases in *B. subtilis*, SleB and CwlJ, and in the absence of the coat protein CotE, the activity of CwlJ is deficient. Germination assays will soon be performed on a *ΔcotE ΔyhcN* double mutant strain. If the role of YhcN in germination identified in the article from Johnson and Moir is confirmed (Johnson and Moir 2017), we will repeat these experiments with strains expressing truncated variants of YhcN to determine which region(s) of the protein (the unfolded N-terminal region, the RBM core and the additional C-terminal α helices) is(are) required for YhcN function.

Although these works did not go well because of the Covid-19 epidemic, they will be continued after my graduation in collaboration with Christopher Rodrigues and Adriano Henriques.

B. RBM-containing proteins within the A-Q complex

B. subtilis is the best known spore-forming bacterium. The several stages of its sporulation cycle have been studied for decades but many aspects of this sophisticated developmental process still remain mysterious. One of the most exciting gaps of knowledge in the field is the function and structure of the macromolecular multi-protein SpoIIIA-SpoIIQ (A-Q) complex, which assembles in the two membranes separating the mother cell and the forespore throughout the engulfment stage. Four of the A-Q proteins (AF, AG, AH and GerM) contain a ring-building motif (RBM) that is also present in ring-forming proteins from specialized secretion systems. By analogy with these systems, the A-Q complex could thus also contain stacks of rings spanning the intermembrane space to transport molecules between the mother cell and the forespore. During my Ph.D., I studied the capacity of AF, AG, AH and GerM to oligomerize but so far, AG is the only one that was shown to form oligomeric rings. In this context, I discuss below the different models regarding the structure and function of the A-Q complex.

1. Do SpoIIIAG form a single or a double ring?

The structure of the 30-mer ring formed by the soluble region of AG was characterized by my hosting lab in 2016 and by the Strynadka's group at a higher resolution in 2017 (Rodrigues, Henry, et al. 2016, Zeytuni et al. 2017). An AG monomer is composed of an RBM core (the building unit of the saucer region in the AG ring) domain, which superposes onto canonical EscJ/PrgK RBM domains, and an unprecedented β-triangle arrangement (the building unit of the cup region in the AG ring). The dimer of the RBM core of AG superposes well onto EscJ/PrgK dimers and also involves hydrophobic and electrostatic interactions allowed by the packing of two helices from one RBM against the β-sheet of an adjacent RBM (Fig. 42A). However, the residues involved in the oligomerization interface diverge between AG and PrgK (Fig. 42C).

I have shown that *in vitro*, the RBM core alone is not able to form rings. It requires the presence of the unfolded N-terminal soluble region (from residue S51 to K90) of AG, as well as the β-triangle arrangement. In the absence of these regions, the protein remains monomeric and the only large molecular weight species that it forms are soluble aggregates.

In the AG rings, the β-triangle regions form a hollow channel that has no structural homologue so far. Whether or not this region can be considered as a domain (i.e. a protein segment that has the capacity to fold by itself) still requires the structural characterization of the His-SUMO-$AG_{L128\text{-}K180}$ recombinant construct, which only contains the β-triangle.

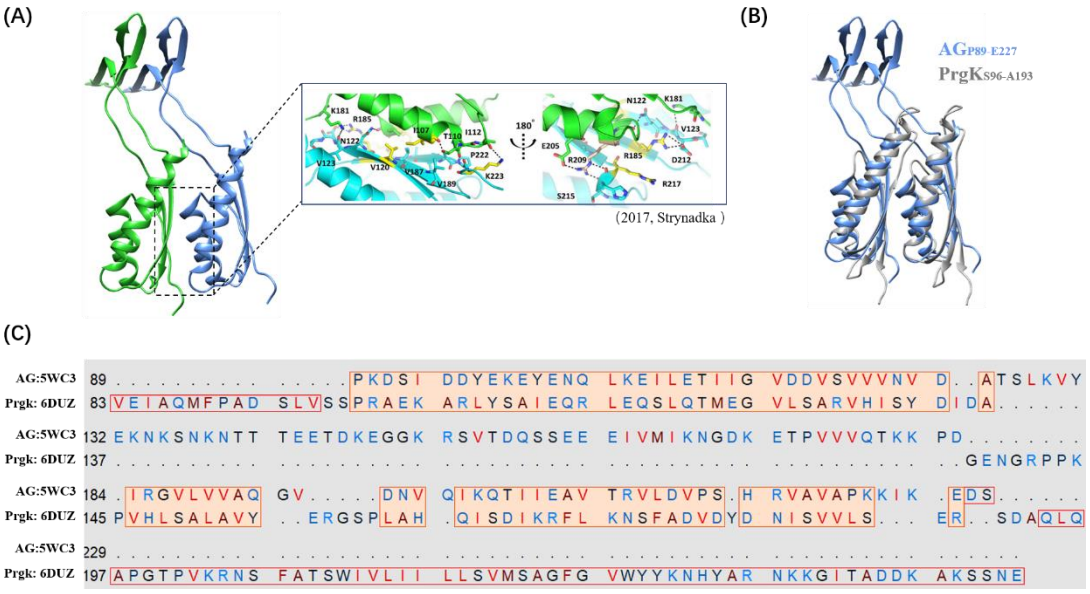


Figure 42. Oligomerization of AG. (A) Ribbon representation of an AG dimer that composes the AG ring and zoom on the dimerization interface. Two protomers are labeled in green and blue (PDB code 5WC3). In the box, AG residues involved in hydrophobic and electrostatic interactions are labeled and shown as atom-colored sticks. **(B)** Structural overlay of an AG dimer (PDB code 5WC3, in blue) onto a dimer of the second RBM domain of PrgK (PDB code 6DUZ, in grey). **(C)** Sequence alignment between the RBM-like domain of AG and the second RBM domain of PrgK, based on the structural alignment shown in panel (B). Residues involved in the oligomerization interface are labeled with a black star. Overlayed domains are boxed in orange.

The discovery of the soluble AG rings provided the first structural evidence that the A-Q complex might assemble into a transport nanomachine that would connect the mother cell and forespore cytoplasms. In addition, the robustness of the AG rings and the presence of a large inner pore (diameter of about 8 nm) suggests that AG is a main piece of the A-Q complex and might be part of a potential conduit. However, the height of the AG ring (about 8 nm) is not large enough to span the entire intermembrane space, which was shown to measure from 25 (before peptidoglycan thinning) to 14 nm (after peptidoglycan thinning) (Khanna et al. 2019). Therefore to span the two membranes around the forespore, the putative conduit must include other components than the AG ring.

Interestingly, the negative-stain EM observation of the full-length, membrane form of AG (His-SUMO-$AG_{M1-S229}$) has revealed the presence of double rings that were never reported before. The resolution of these negative-stain EM images of the double AG rings is too low to determine whether the individual rings interact through the cup region or through the TM segment and two alternative hypotheses can therefore be proposed and are discussed below:

1. Two His-SUMO-$AG_{M1-S229}$ rings interact through their TM segments. Placed in a cellular

context, this hypothesis predicts that the two rings of AG will position on both sides of the mother cell membrane (Fig. 43A).

The double AG rings observed by negative-stain EM show three main parts, which appear as whitish bands on the EM micrographs (see Fig. 29 in the *Result* section). If the central one is indeed made of the TM segments, then the region connecting the central band to the outer bands is likely made of the protein segment encompassing residues S51 to K90 (called the D1 domain). The height of the connecting region, supposedly made by the D1 domain, is about 4 nm on the EM micrographs. If this region makes a ring, then the total height of the AG channel would thus reach about 12 nm. It might thus still be not enough to span the intermembrane space. The AG channel would thus allow the passage of a molecule through the outer forespore membrane but transport across the intermembrane space and inner forespore membrane would require other channel-forming proteins (Fig. 43B). In this model, one might question the positioning of the AA ATPase, which might form heaxamers like its homologues. AA hexamer would be too large to accommodate in the pore formed by the cup region, it could thus only associate to the outer surface of the AG ring.

A model in which AG would form a double ring on both sides of the mother cell membrane is not supported by the protease digestion susceptibility assay reported in (Rodrigues, Henry, et al. 2016). Indeed in this assay, sporulating cells were protoplasted (peptidoglycan digestion) and treated with trypsin. Proteins exposed in the intermembrane space are fully degraded by the trypsin while proteins located in the mother cell cytoplasm are protected. If AG would be present on both sides of the mother cell membrane, the fraction of the protein located in the cytoplasm would be partially protected from the trypsin. In this experiment AG was fully degraded (Rodrigues, Henry, et al. 2016), indicating that it is not exposed in the mother cell cytoplasm. This result thus contradicts the first double ring model.

2. Two His-SUMO-$AG_{M1\text{-}S229}$ rings interact through their cup region. Placed in a cellular context, this hypothesis predicts that one ring of AG would assemble in the mother cell membrane while the other one would assemble in the forespore membrane (Fig. 43C). Given the height of one AG ring (~ 8 nm), this transenvelope double AG ring would form a channel that could span the entire intermembrane space in the region where it measures less than 16 nm.

   This model seems very appealing because so far, AG is the only A-Q protein that was

shown to form rings. A transenvelope double ring of AG could thus be sufficient to make a conduit connecting the mother cell and forespore and there would be no need for other rings formed by other A-Q proteins. Nonetheless if AF and AH oligomerize, they could play a role in stabilizing the channel by making rings that surround the AG channel. Alternatively, the AG channel could also be stabilized by other A-Q components that do not form rings. In particular the AH-Q dimer could make discrete transenvelope pillars all around the AG rings (Fig. 43D). GerM on the other hand, if it forms fibers *in vivo*, could weave some kind of web to connect the A-Q complexes all around the forespore (Fig. 43D). Such a molecular net would help the forespore resisting the turgor pressure exerted by the packing of the DNA inside the forespore (Lopez-Garrido et al. 2018). Other A-Q proteins could have an export function, such as AC, AD and AE.

This model however presents two main inconsistencies. A first one is that interaction between the cup regions of AG was never observed with the soluble $AG_{S51\text{-}S229}$ construct. It seems unlikely that such interaction requires the presence of the TM segments since the cup region is localized on the opposite side of the N-terminus in the soluble AG rings (Rodrigues, Henry, et al. 2016). A second contradiction comes from the fact that AG is encoded within the *spoIIIA* operon, which is under the control of the mother-cell specific σE transcription factor. AG is thus produced in the mother cell and there is so far no evidence that it can also be produced in the forespore. In addition, the presence of its α-helical transmembrane segment makes it very unlikely that following its production by the mother cell, the protein traverses the outer forespore membrane and intermembrane space to be inserted in the inner forespore membrane. To achieve this model, the *spoIIIAG* gene should also be transcribed by a forespore-specific promoter.

In conclusion, interaction of the AG rings through the cup region seems unlikely *in vitro* and interaction through the TM segments seems the most probable hypothesis. *In vivo* however, both models present inconsistencies with the reported literature. In my opinion, these double rings are most likely recombinant artefacts. Nonetheless, even if the double AG rings are not physiological, solving their structure by cryo-EM would allow determining whether the D1 domain forms a ring in the presence of the TM segments and whether the TM segments form a membrane pore. This structure would thus provide more information regarding the architecture and dimensions of the AG channel. Detergent purification of the His-SUMO-$AG_{M1\text{-}S229}$ sample is currently being optimized to increase the concentration of the protein and prepare cryo-EM grids. If the presence of the DDM detergent causes problems to prepare the grids, the full-length AG rings could be reconstituted in nanodiscs.

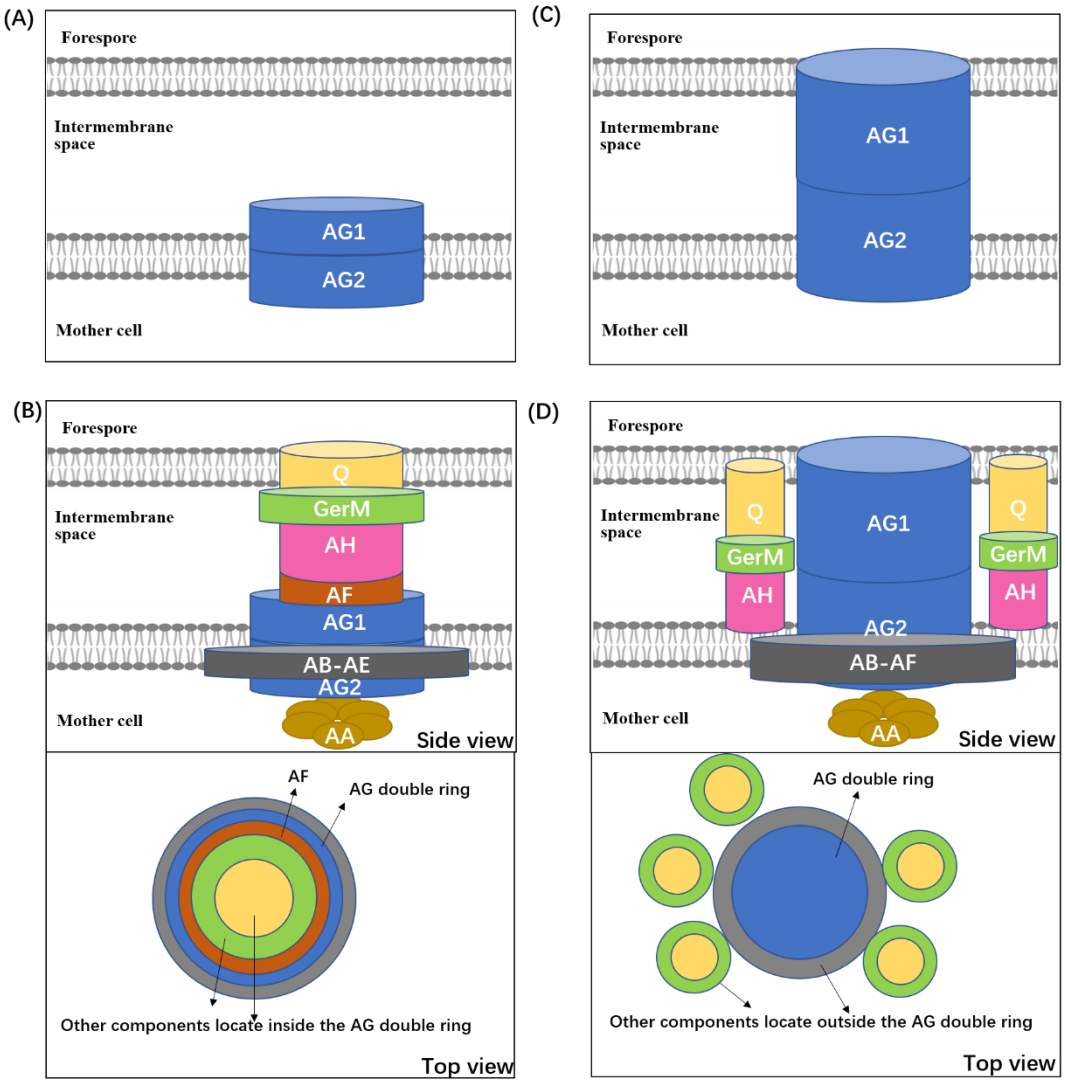


Figure 43. Models for the assembly of the AG oligomer as a double ring. (A) The two AG rings position on both sides of the mother cell membrane. **(B)** Model in which two AG rings anchored in the mother-cell membrane, together with rings of AF, AH and Q, make the trans-enveloppe conduit. AF and the AH-Q-GerM trimer are shown as embedded in the AG ring but could alternatively surround the AG ring in the intermembrane space. **(C)** Two AG rings assemble in both the mother cell and forespore membranes and span the intermembrane space. **(D)** Model in which a transenvelope AG channel is stabilized and connected by AH, Q and GerM. AA is shown as an hexameric ATPase, and AB to AE as components of a platform anchored in the mother-cell membrane.

2. Do SpoIIIAF, SpoIIIAH and GerM oligomerize?

The structure of $AF_{V60-E206}$ was characterized by the Strynadka's group in 2018 while I was screening for crystallization conditions of a $AF_{K55-E206}$ construct at that time. The study by Strynadka and co. showed the presence of a high molecular-weight species by size-exclusion chromatography (Fig. 44A) (Zeytuni et al. 2018a). Some $AF_{V60-E206}$ rings were observed by EM but they were very heterogeneous, irregular, and the presence of a central pore was not obvious (Fig. 44B). In my opinion, some of them are more likely to be sunk aggregates and I am not convinced by the interpretation of the authors regarding these particles. Besides, AF does not form a large oligomer in the crystal; the asymmetric unit only contains a dimer. AF shares an RBM core (αββαβ) as in AG and PrgK RBMs, but the orientation of AF's second α-helix is different from those in AG and PrgK (Fig. 44C-E). The structural overlay of AF and PrgK dimer shows that AF possesses an RBM domain that aligns well with the second RBM domain of PrgK. However, when I superpose AF onto a PrgK dimer, the second α-helix of $AF_{V60-E206}$ collides with the adjacent PrgK protomer (Fig. 44F). Similarly, alignment of $AF_{V60-E206}$ onto an AG dimer shows that the second α-helix of AF could clash with the AG oligomerization interface (Fig. 44G).

In conclusion, even if AF could oligomerize, it would do so through an interface that would slightly diverge from the canonical αα/βββ interface observed in rings formed by AG, PrgK RBM2, PrgH RBM3, InvG RBM2-3 and GspD RBM2-4.

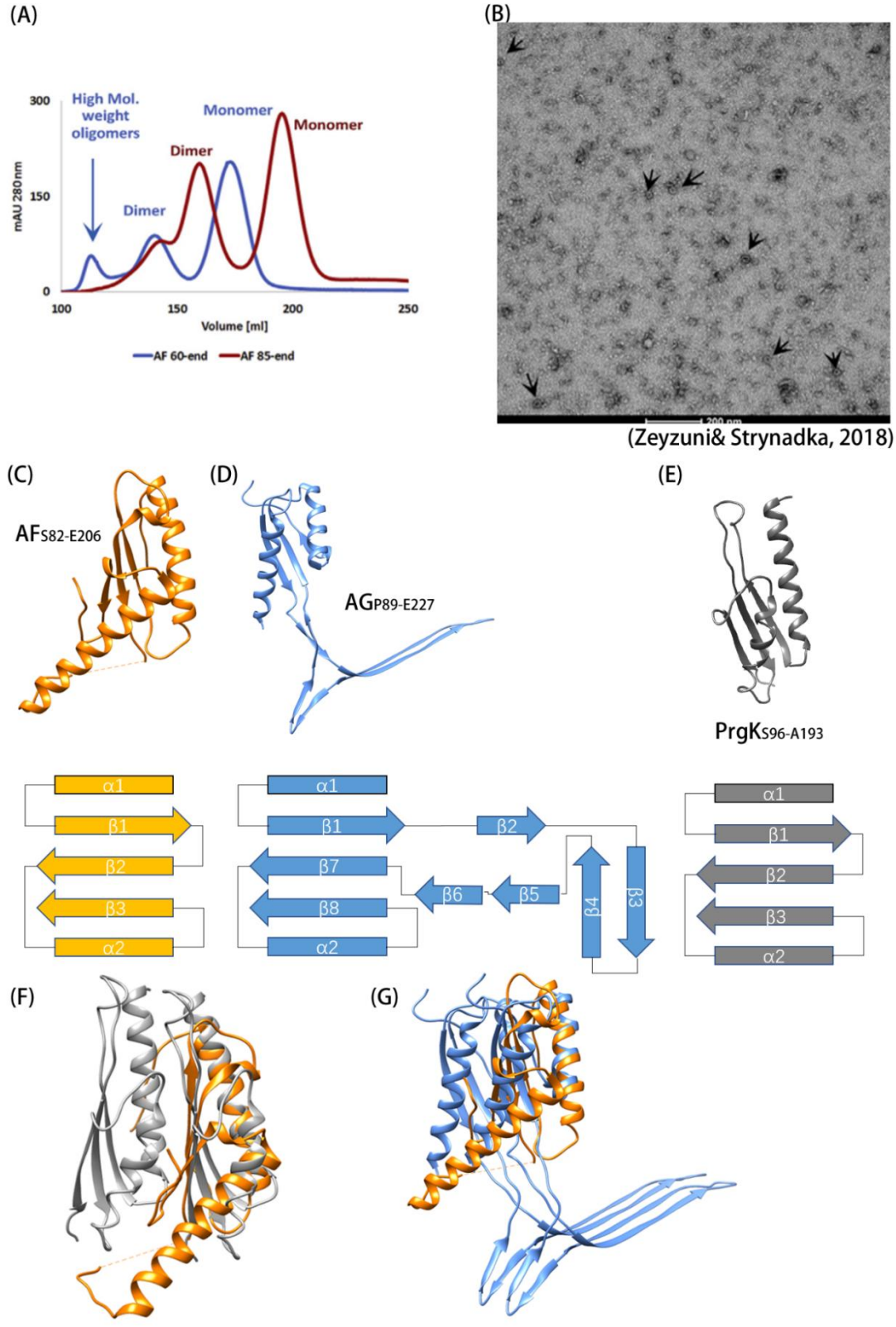


Figure 44. Structural analysis of the RBM-containing protein SpoIIIAF. (A) Size-exclusion chromatograms of $AF_{V60\text{-}E206}$ showing the presence of a monomers, dimers and large oligomers. **(B)** Negative-stain EM analysis of $AF_{V60\text{-}E206}$. Black arrows point to ring-like particles. From Zeytuni et al., 2018. **(C)** Structure and topology of $AF_{S82\text{-}E206}$ (PDB code 6DCS). **(D)** Structure and topology of $AG_{P89\text{-}E227}$ (PDB code 5WC3). **(E)** Structure and topology of $PrgK_{S96\text{-}A193}$ (PDB code 6DUZ). **(F)** Structural overlay of $AF_{S82\text{-}E206}$ (in orange) onto a dimer made by PrgK RBM2 (in grey). (**G**) Structural overlay of $AF_{S82\text{-}E206}$ (in orange) onto a dimer made by AG (in blue).

The C-terminal RBM-like domain of AH is exposed in the intermembrane space. Based on the similarity between the core of the RBM-like domain of AH and PrgK RBM2 domain, and its direct interaction with Q, the AH-Q dimer was previously modeled into compact rings containing 12-25 protomers (Meisner et al. 2012, Levdikov et al. 2012). AH however was never shown to oligomerize: neither the soluble construct (in previous studies) nor the full-length membrane construct that I purified during my Ph.D.

AH possesses a αββαβ RBM core and an additional N-terminal α-helix (Fig. 45A). Similar to $AF_{V60\text{-}E206}$, the RBM core of $AH_{D104\text{-}S218}$ aligns well with the second RBM domain of PrgK (Fig. 45C), but the additional N-terminal α-helix interferes with the PrgK oligomerization interface (Fig. 45C). This observation was not reported in any of the structural studies of AH published previously (Levdikov et al. 2012, Meisner et al. 2012) but it might prevent AH from oligomerizing. To test this idea, I have characterized an $AH_{S129\text{-}K218}$ construct lacking the α0 helix. Unfortunately, this construct does not oligomerize and so far, we have therefore still no proof that AH is able to form rings. If it does oligomerize *in vivo*, it will involve an interface that will allow accommodation of the α0 helix and thus will be different from the PrgK RBM2-RBM2 and AG-AG ones.

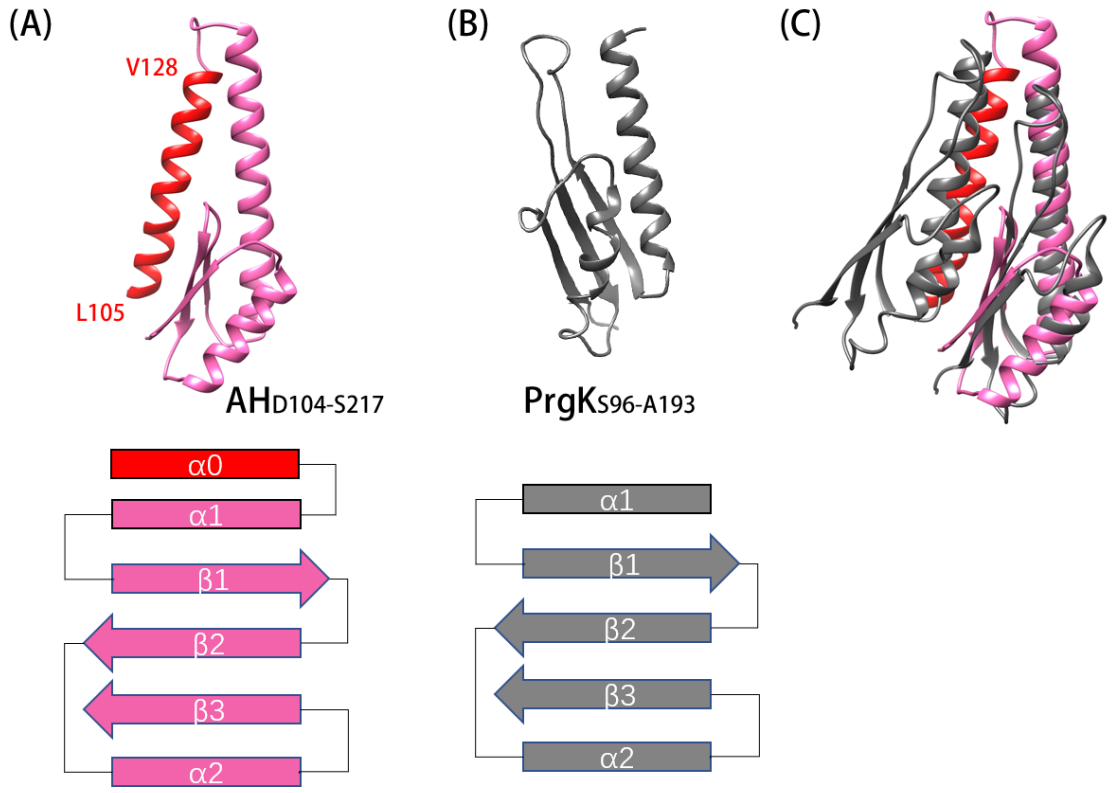


Figure 45. Structural analysis of the RBM-containing protein SpoIIIAH. (A) Structure and topology of $AHD_{104\text{-}S217}$ (PDB code 3UZ0). The RBM core domains is colored pink, the additional α0 helix is in red. **(B)** Structure and topology of $PrgK_{S96\text{-}A193}$ from *S. typhimurium* (PDB code 6DUZ). **(C)** Structural overlay of $AF_{S82\text{-}E206}$ onto a dimer made by the second RBM domain of PrgK (PDB code 6DUZ, in grey).

The case of GerM is different from AF and AH since it contains two RBM-like domains, called GerMN1 and GerMN2 (Fig. 46A). The two domains are connected with a loop and stabilized with hydrophobic and electrostatic interactions (Fig. 46A). The interface between GerMN1 and GerMN2 is very different from those of PrgK RBM2-RBM2 and AG-AG dimers (Fig. 46B).

Each of the domains display a αββαβ RBM core but they also contain an additional N-terminal two-stranded β-sheet (Fig. 46C). The existence of a GerM ring remains to be demonstrated and the tandem of RBM-like domains, together with the additional β-sheets make it unlikely that GerM can form a ring resembling those of PrgK or AG.

A GerM fiber was observed in the crystal (Fig. 46D) but no evidence that the oligomerization interface exists *in vivo* could be provided so far (Trouve et al. 2018). The only oligomer that GerM is able to form so far is therefore a straight fiber, which might not be physiological. If they exist, GerM fibers might not only stabilize the AH-Q complex, but also connect different A-Q complexes around the forespore.

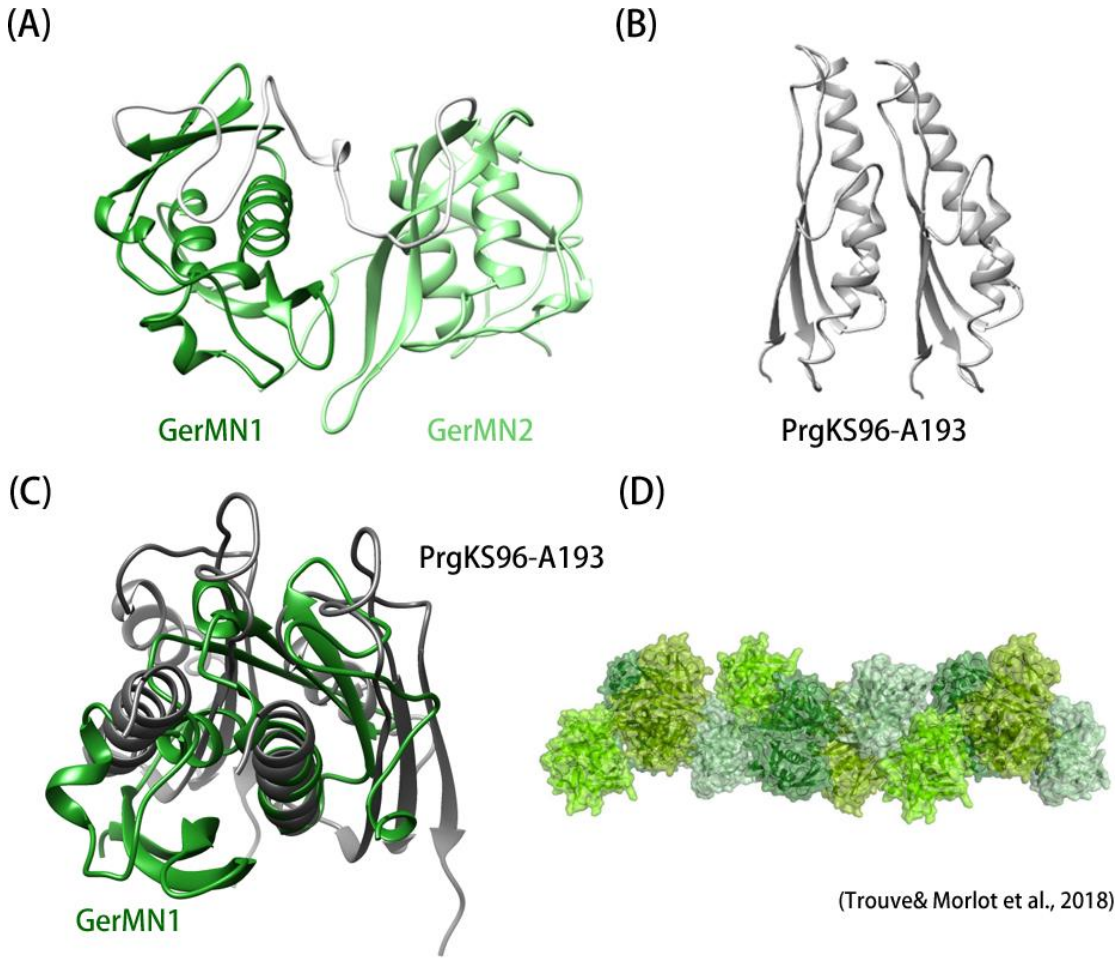


Figure 46. Structure of GerM$_{F26-F366}$. (A) Ribbon representation of GerM$_{F26-F366}$ (PDB code 6GZB). GerM$_{F26-F366}$ contains two RBM-like domains, called GerMN1 (dark green) and GerMN2 (light green). **(B)** Structure of a PrgK$_{S96-A193}$ dimer from *S. typhimurium* (PDB code 6DUZ), in which two α-helices fold against the three anti-parallel β-sheet, forming the canonical αα/βββ RBM oligomerization interface. **(C)** Structural overlay of GerMN1 onto a dimer made by the second RBM domain of PrgK. The orientation of one of the GerM's α-helix is different with the corresponding one in PrgK. The additional N-terminal two-stranded β-sheet interferes with the PrgK oligomerization interface. **(D)** Ribbon and surface representation of the GerM fiber observed in the crystal structure. The protofilament results from the continued dimerization of GerM26-366 molecules in the crystal. The fiber is visualized upon display of the crystallographic symmetries of the asymmetric unit content. The four GerM molecules present in the asymmetric unit are colored with different shades of green.

C. Hypotheses regarding the capacity of RBM-containing proteins to oligomerize.

Based on the structural analysis of the RBM-like proteins involved in bacterial sporulation and the canonical RBM domains found in specialized secretion systems, I propose the following hypotheses.

- The presence of an RBM core domain (consisting of two α-helices folding against an anti-parallel three-stranded β-sheet) might be necessary but not sufficient to promote *in vitro* oligomerization of a protein.

- If the canonical αα/βββ RBM interface (established by the β-sheet of one molecule and the two α-helices from the adjacent protomer) must be exposed to allow ring formation, then the additional secondary structures observed in the RBM-like domains of AH, GerM and YhcN might display another conformation *in vivo* to allow oligomerization of the proteins. This alternative conformation might be induced by the presence of protein partners.

- If the conformation of the additional secondary structures observed in the crystal structure of the RBM-like domains of AH, GerM and YhcN do not prevent ring formation, then these proteins oligomerize through an interface that is different from the canonical αα/βββ interface observed in PrgK RBM2-RBM2 and AG rings.

- The structure of AF, AH, GerM and YhcN might have structurally evolved from canonical RBM domains to fulfill different functions, not related to protein oligomerization. The acquisition of additional secondary structures might for example have allowed these proteins to establish interactions with yet-to-be identified partner(s). Regarding YhcN, which has so far not been involved in a transport machinery, we cannot exclude that its αββαβ fold is a convergent structural evolution. In other words, the YhcN function might be totally unrelated to the ring-forming function of secretion proteins but its fold might have converged towards the RBM-like fold, which seems a rather simple fold to make.

The structural study of sporulation RBM-containing proteins such as AF, AH, GerM and YhcN in their physiological context should allow determining whether they form rings. Such study on A-Q proteins will be performed by a new Ph.D. student (Elda Bauda) in our laboratory, starting in October 2020. More precisely, she will determine the 3D structure of the A-Q complex using cellular cryo-electron microcopy on thin sections of sporulating *B. subtilis* cells. Beyond revealing whether AF, AH and GerM are able to form rings, this study will also help

discriminating between the different hypotheses regarding the function of the A-Q complex. These hypotheses are discussed in the next section.

D. Hypotheses regarding the function of the A-Q complex

During sporulation, over 500 genes are expressed under the control of mother cell- and forespore-specific σ transcription factors. The activation of these σ factors depends on complex mother cell-forespore signaling pathways that have been studied for decades. While the activation of σF, σE and σK is rather well understood, the activation of σG in the forespore remains unclear. It was shown to require the assembly of the transenvelope A-Q complex but the role of this complex remains mysterious. There are three main models for the function of the A-Q complex that are discussed below:

1. The A-Q complex works as a T3SS-like machinery that allows the transport of a protein between the mother cell and forespore. This model is mainly supported by the structural similarities between the AG ring and PrgK rings found in T3SS. By analogy with T3SS, it was proposed that AG, AF, AH, Q and possibly GerM would form stacked rings in the intermembrane space (Fig. 47A) (Camp and Losick 2008; Meisner et al. 2008, 2012; Rodrigues et al. 2016b; Morlot and Rodrigues 2018). In this model, the AA ATPase would provide energy for secretion of a yet-to-be defined molecule while AB would connect AA to the inner-membrane platform made of AC to AH. AC, AD and AE would form an export platform that would gate the A-Q channel and select and/or load the secreted substrate. In this model, the global architecture of the A-Q complex would (at least partially) resemble the architecture of a T3SS (Fig. 47B).

2. In an alternative model, stacked rings of AF, AG, AH, Q and GerM would form a channel that would accommodate a pilus made of AC and AD. The AA ATPase would provide energy for polymerization of the AC-AD pilus that would push the secreted substrate through the A-Q channel (Fig. 47C). Compared to the previous one, this hypothesis is mainly supported by the presence of multi transmembrane segments in AC and AD. Equivalent numbers of TM segments are however also found between AD, AC and AE and the FliP, FliQ and FliR components of flagellar T3SS, respectively (see section **IIC** of the *Introduction* chapter). In the T4P model, the global architecture of the A-Q complex would (at least partially) resemble the architecture of a T4 pilus (Fig. 47D)

3. The A-Q complex works as a feeding tube responsible for the passive transport of small molecules that are necessary for the forespore physiology (ions, nucleotides, …) (Camp and Losick 2009). In this model, the A-Q complex would serve as an umbilical cord that would provide nutrients to the forespore, while it is progressively isolated from the extracellular environment during engulfment (Camp and Losick 2008). Like in the two previous models, AG, AH, Q and GerM would also form stacked rings spanning the

intermembrane space, but the AA ATPase would only provide energy for the assembly of the A-Q complex, not for the transport of the molecule(s) from the mother cell to the forespore cytoplasm (Fig. 47E). Here again, AB would connect AA to the inner membrane platform while AC-AD-AE would serve as a gating platform. In this model, the global architecture of the A-Q complex could (at least partially) resemble the architecture of gap junctions previously observed in cyanobacteria (Fig. 47F) (Weiss et al. 2019).

4. In a last model, the A-Q components would not have a function related to substrate secretion but function related to the integrity and adhesion of the two membranes surrounding the forespore. There might be two different complexes in this case : a first one made of the AH-Q dimer, and a second one made of AA to AG. The AH-Q dimer would function as a molecular "zipper" or "ratchet" facilitating migration of the mother cell membrane around the forespore during engulfment and tethering the outer and inner forespore membranes. AA to AG on the other hand would form a complex that would rigidify the outer forespore membrane and counteract the pressure exerted by the nascent coat. In the absence of the AA-AG complex, synthesis of the coat on the mother cell side might deform the outer forespore membrane and provoke fusion of the two membranes. In this model, fibers of GerM might connect the AH-Q complexes to the AA-AG complexes and form a molecular net that would rigidify the outer forespore membrane (Fig. 47G).

I am more inclined to the first (T3SS) and second (T4P) models. Given the large pore (8 nm) of the AG ring, a channel of this size would be big enough for the passage of globular proteins weighing up to 150 kDa. In addition, the fact that AA to AH are encoded within a single operon suggest that they work as a single machinery and contradicts the fourth model (two distinct complexes involved in membrane adhesion and integrity). As argued in the review from Morlot and Rodrigues, against the third model (feeding tube) is the fact that even when the forespore has access to external nutrient (before being completely engulfed), its development still requires the A-Q complex.

Interestingly, potential homologues of all the A-Q proteins are found in several different secretion systems: AA with GspE in T2SS; AB with GspF/PliC in T2SS and T4SS; AC, AD and AE with FliQ, FlipP and FliR in T3SS, and AF, AG and AH with PrgK ring-forming proteins of T3SS. Whether or not the A-Q complex represents a common ancestor from which the different specialized secretion systems have derived is an exciting question.

In conclusion, there are still many questions that remain to be answered regarding the

structure and function of the A-Q complex.

What is the global architecture of the complex?

What is the nature of secreted molecule(s) if the A-Q complex transports something?

What is the secretion mechanism if any?

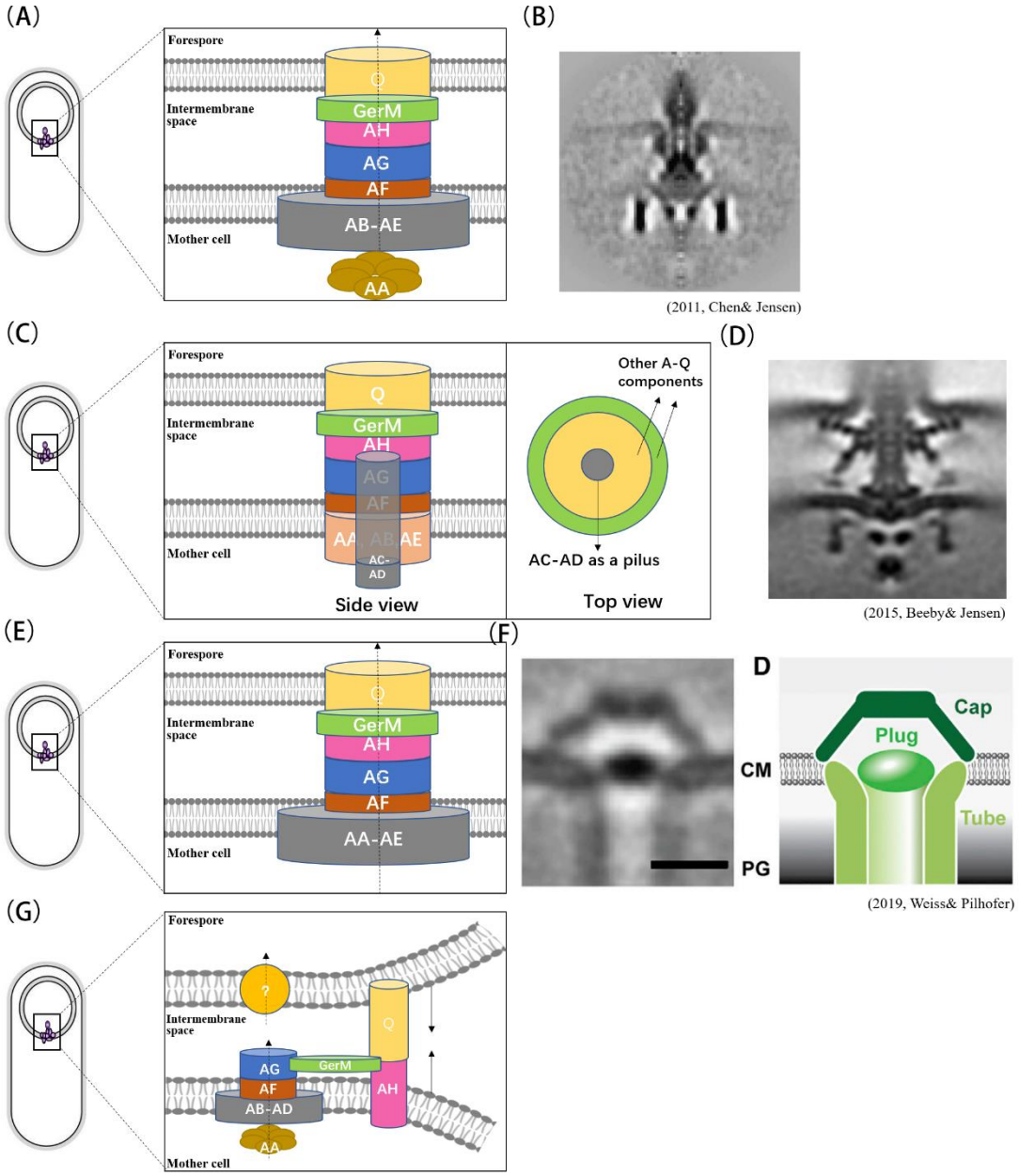


Figure 47. Models of for the structure and function of the A-Q complex. (A) The A-Q complex forms a T3SS-like machine. AF, AG, AH and Q assemble as stacked rings that form a transenvelope channel connecting the mother cell and forespore comportments. AA provides energy for substrate secretion. AB connects AA to the export/gating platform made of AC-AD-AE. **(B)** Cellular cryo-EM image of *Salmonella enterica* T3SS. Adapted from Chen et al., 2011. **(C)** The A-Q complex forms a T4P-like machine. AF, AG, AH and Q assemble as stacked rings that form a transenvelope channel connecting the mother cell and forespore comportments. AA provides energy for polymerization of a pilus made of AC and AD, that pushes the secreted substrate through the AF-AG-AH-Q channel. **(D)** Cellular cryo-EM image of *Thermus thermophilus* T4P machinery. Adapted from Beeby et al., 2015. **(E)** The A-Q complex forms a feeding tube. AF, AG, AH and Q assemble as stacked rings that form a transenvelope channel connecting the mother cell and forespore comportments. AA provides energy for assembly of the AF-AG-AH-Q channel. AC-AD-AE form a gating platform. **(F)** Cellular cryo-EM image (left) and schematic representation (right) of gap-like channel in anabaena filament. The gated channel consists of three main components: cap, plug and tube. Adapted from Weiss et al., 2019. **(G)** The A-Q components form two complexes involved in membrane integrity and tethering. AA to AG would form a complex that would rigidify the outer forespore membrane and counteract the pressure exerted by the growing coat on the mother cell side. The A-Q dimer would tether the outer and inner forespore membranes during engulfment. GerM would form a net connecting the AA-AG and AH-Q complexes to further rigidify the two membranes surrounding the forespore.

In my project, *in vitro* reconstitution of putative stacked rings of the complex based on purified recombinant full-length membrane components has failed. As mentioned earlier, reconstructing the structure of the A-Q complex in its native environment using cryo-cellular electron microscopy appears as the most relevant and promising strategy to unravel its global architecture and provide insights into its function. In support of this idea, the most ground-breaking discoveries in the field of specialized secretion systems were recently achieved using such approach (Worrall et al., 2016). Our graduating master student, Elda Bauda, will tackle this challenging project in the next three years, in close collaboration with the Methods and Electron Microscopy group at the IBS. To do so, she will use ultramicrotomy, which is a technique that allows cutting of biological samples into nanometer-thick lamellae. In these sections, the A-Q complex will be labeled with antibodies coupled to gold beads to allow its localization in EM micrographs. This approach will allow to precisely localize the A-Q complex in the intermembrane space of sporulating cells, in order to identify specific structural elements that will facilitate that will facilitate further cellular cryo-tomography studies. Through these cutting-edge EM approaches, we expect to determine the structure of the A-Q complex at high resolution in its native state. Reconstructing the structure of the A-Q complex will determine if the complex actually **presents structural elements necessary for the transport of molecules**, such as the presence of a membrane pore between the mother cell and the forespore, or export components. The diameter and architecture of the conduit will serve as a first indication regarding the nature of the transport and will facilitate the identification of protein-protein interaction interfaces to potentially develop, in the long term, inhibitors of the sporulation process. Moreover, the architecture of the A-Q complex will reveal similarities and differences with specialized secretion systems found in Gram-negative bacteria, allowing to understand their evolutionary history.

Valorization

Valorization

The A-Q complex involved in *B. subtilis* sporulation has been considered as a potential novel transport system for the last 12 years. However, the architecture, assembly mechanism and the putative secreted substrate are still unknown. In this project, we focused on the A-Q components that harbor a "ring-building motif" (RBM), a conserved fold that was shown to be present in ring-forming components of specialized secretion systems. The data generated during my PhD have been or will be communicated through various supports, including research articles published in peer-reviewed journals, posters and oral presentations during symposia.

In the first part of my project, I joined a structure-function study of GerM, one of the A-Q components that was shown to play an important role in the localization of A-Q proteins and in σG activation. I performed a biophysical study of GerM, which showed that a direct interaction with Q and AH can be detected by MST. However, no good internal control could be obtained and the interactions could not be demonstrated using another technique. Due to the urge to publish the structure of GerM, we could not include the MST data in the article. I was however listed as a co-author (Trouve, Mohamed, Leisico, Contreras-Martel, Liu et al., *J. Struct. Biol.*, 2018) because my experiments were mentioned in the article and because I produced the protein that was used for SAXS experiments in this study.

One of my projects consisted in testing the ability of full-length membrane constructs of AF, AG, AH and Q to form oligomeric rings. I discovered that AG can form a double ring that has never been reported before. Although this double ring might not be physiological, solving its structure should allow unraveling the structure of regions that were absent (transmembrane segment) or invisible (D1 domain) in the soluble construct of AG, whose structure was previously published(Rodrigues, Henry, et al. 2016). This study will be performed after the end of my PhD by electron microscopy, in collaboration with the group of Guy Schoehn at the IBS. Besides, if the double AG ring could form a hollow channel, it might be used for biotechnological applications such as a nano-pores.

In the third part of my project, I studied YhcN, another RBM-containing protein involved

in *B. subtilis* sporulation. I solved the crystal structure of YhcN but the protein showed no sign of ring-like oligomerization. To try to understand why some RBM domains do not oligomerize *in vitro* (such as AF, AH, GerM and YhcN), I purified truncated RBM-like domains of AG, AH and YhcN, and characterized their ability to form rings. This work suggests that all or at least some RBM domains are not sufficient to trigger ring oligomerization. So far, only RBM domains associated with β-barrels were shown to oligomerize *in vitro*. It is thus possible that RBM domains are rather ring-stabilizing motifs than ring-building motifs. This study will be communicated in a research article that is currently being prepared and will be untitled "Structural insights into ring-building motifs involved in bacterial sporulation".

During my PhD, this project named ‘Proteins with RBM (ring-building motif)-like domains involved in bacterial sporulation’ has been presented as a poster at the "PSB student day" in 2019, as a flash talk at the "PSB student day" in 2020, and also as an oral presentation at the "Bacto-Gre" symposium in 2020.

References

References

Abrusci, P., M. A. McDowell, S. M. Lea, and S. Johnson. 2014. "Building a secreting nanomachine: a structural overview of the T3SS." *Curr Opin Struct Biol* no. 25:111-7. doi: 10.1016/j.sbi.2013.11.001.

Aly, K. A., and C. Baron. 2007. "The VirB5 protein localizes to the T-pilus tips in Agrobacterium tumefaciens." *Microbiology (Reading)* no. 153 (Pt 11):3766-3775. doi: 10.1099/mic.0.2007/010462-0.

Babic, A., A. B. Lindner, M. Vulic, E. J. Stewart, and M. Radman. 2008. "Direct visualization of horizontal gene transfer." *Science* no. 319 (5869):1533-6. doi: 10.1126/science.1153498.

Bagyan, I., M. Noback, S. Bron, M. Paidhungat, and P. Setlow. 1998. "Characterization of yhcN, a new forespore-specific gene of Bacillus subtilis." *Gene* no. 212 (2):179-88. doi: 10.1016/s0378-1119(98)00172-3.

Bergeron, J. R. 2016. "Structural modeling of the flagellum MS ring protein FliF reveals similarities to the type III secretion system and sporulation complex." *PeerJ* no. 4:e1718. doi: 10.7717/peerj.1718.

Bergeron, J. R. C., L. J. Worrall, S. De, N. G. Sgourakis, A. H. Cheung, E. Lameignere, M. Okon, G. A. Wasney, D. Baker, L. P. McIntosh, and N. C. J. Strynadka. 2015. "The modular structure of the inner-membrane ring component PrgK facilitates assembly of the type III secretion system basal body." *Structure* no. 23 (1):161-172. doi: 10.1016/j.str.2014.10.021.

Bergeron, J. R., L. J. Worrall, N. G. Sgourakis, F. DiMaio, R. A. Pfuetzner, H. B. Felise, M. Vuckovic, A. C. Yu, S. I. Miller, D. Baker, and N. C. Strynadka. 2013. "A refined model of the prototypical Salmonella SPI-1 T3SS basal body reveals the molecular basis for its assembly." *PLoS Pathog* no. 9 (4):e1003307. doi: 10.1371/journal.ppat.1003307.

Bernstein, H. D. 2019. "Type V Secretion in Gram-Negative Bacteria." *EcoSal Plus* no. 8 (2). doi: 10.1128/ecosalplus.ESP-0031-2018.

Besprozvannaya, M., and B. M. Burton. 2014. "Do the same traffic rules apply? Directional chromosome segregation by SpoIIIE and FtsK." *Mol Microbiol* no. 93 (4):599-608. doi: 10.1111/mmi.12708.

Broder, D. H., and K. Pogliano. 2006. "Forespore engulfment mediated by a ratchet-like mechanism." *Cell* no. 126 (5):917-28. doi: 10.1016/j.cell.2006.06.053.

Burkinshaw, B. J., and N. C. Strynadka. 2014. "Assembly and structure of the T3SS." *Biochim Biophys Acta* no. 1843 (8):1649-63. doi: 10.1016/j.bbamcr.2014.01.035.

Buttner, D. 2012. "Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria." *Microbiol Mol Biol Rev* no. 76 (2):262-310. doi: 10.1128/MMBR.05017-11.

Camp, A. H., and R. Losick. 2008. "A novel pathway of intercellular signalling in Bacillus subtilis involves a protein with similarity to a component of type III secretion channels." *Mol Microbiol* no. 69 (2):402-17. doi: 10.1111/j.1365-2958.2008.06289.x.

Camp, A. H., and R. Losick. 2009. "A feeding tube model for activation of a cell-specific transcription factor during sporulation in Bacillus subtilis." *Genes Dev* no. 23 (8):1014-24. doi: 10.1101/gad.1781709.

Campo, N., and D. Z. Rudner. 2006. "A branched pathway governing the activation of a developmental transcription factor by regulated intramembrane proteolysis." *Mol Cell* no. 23 (1):25-35. doi: 10.1016/j.molcel.2006.05.019.

Campos, M., M. Nilges, D. A. Cisneros, and O. Francetic. 2010. "Detailed structural and assembly model of the type II secretion pilus from sparse data." *Proc Natl Acad Sci U S A* no. 107 (29):13081-6. doi: 10.1073/pnas.1001703107.

Cano, R. J., and M. K. Borucki. 1995. "Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber." *Science* no. 268 (5213):1060-4. doi: 10.1126/science.7538699.

Cascales, E., and P. J. Christie. 2003. "The versatile bacterial type IV secretion systems." *Nat Rev Microbiol* no. 1 (2):137-49. doi: 10.1038/nrmicro753.

Chami, M., I. Guilvout, M. Gregorini, H. W. Remigy, S. A. Muller, M. Valerio, A. Engel, A. P. Pugsley, and N. Bayan. 2005. "Structural insights into the secretin PulD and its trypsin-resistant core." *J Biol Chem* no. 280 (45):37732-41. doi: 10.1074/jbc.M504463200.

Chandran, V., R. Fronzes, S. Duquerroy, N. Cronin, J. Navaza, and G. Waksman. 2009. "Structure of the outer membrane complex of a type IV secretion system." *Nature* no. 462 (7276):1011-5. doi: 10.1038/nature08588.

Chernyatina, A. A., and H. H. Low. 2019. "Core architecture of a bacterial type II secretion system." *Nat Commun* no. 10 (1):5437. doi: 10.1038/s41467-019-13301-3.

Chirakkal, H., M. O'Rourke, A. Atrih, S. J. Foster, and A. Moir. 2002. "Analysis of spore cortex lytic enzymes and related proteins in Bacillus subtilis endospore germination." *Microbiology (Reading)* no. 148 (Pt 8):2383-2392. doi: 10.1099/00221287-148-8-2383.

Ciccarelli, F. D., T. Doerks, C. von Mering, C. J. Creevey, B. Snel, and P. Bork. 2006. "Toward automatic reconstruction of a highly resolved tree of life." *Science* no. 311 (5765):1283-7. doi: 10.1126/science.1123061.

Costa, T. R., C. Felisberto-Rodrigues, A. Meir, M. S. Prevost, A. Redzej, M. Trokter, and G. Waksman. 2015. "Secretion systems in Gram-negative bacteria: structural and mechanistic insights." *Nat Rev Microbiol* no. 13 (6):343-59. doi: 10.1038/nrmicro3456.

Crawshaw, A. D., M. Serrano, W. A. Stanley, A. O. Henriques, and P. S. Salgado. 2014. "A mother cell-to-forespore channel: current understanding and future challenges." *FEMS Microbiol Lett* no. 358 (2):129-36. doi: 10.1111/1574-6968.12554.

Daniel, R. A., and J. Errington. 1993. "Cloning, DNA sequence, functional analysis and transcriptional regulation of the genes encoding dipicolinic acid synthetase required for sporulation in Bacillus subtilis." *J Mol Biol* no. 232 (2):468-83. doi: 10.1006/jmbi.1993.1403.

Dembek, M., A. Kelly, A. Barwinska-Sendra, E. Tarrant, W. A. Stanley, D. Vollmer, J. Biboy, J. Gray, W. Vollmer, and P. S. Salgado. 2018. "Peptidoglycan degradation machinery in Clostridium difficile forespore engulfment." *Mol Microbiol* no. 110 (3):390-410. doi: 10.1111/mmi.14091.

Doan, T., J. Coleman, K. A. Marquis, A. J. Meeske, B. M. Burton, E. Karatekin, and D. Z. Rudner. 2013. "FisB mediates membrane fission during sporulation in Bacillus subtilis." *Genes Dev* no. 27 (3):322-34. doi: 10.1101/gad.209049.112.

Doan, T., K. A. Marquis, and D. Z. Rudner. 2005. "Subcellular localization of a sporulation membrane protein is achieved through a network of interactions along and across the septum." *Mol Microbiol* no. 55 (6):1767-81. doi: 10.1111/j.1365-2958.2005.04501.x.

Doan, T., C. Morlot, J. Meisner, M. Serrano, A. O. Henriques, C. P. Moran, Jr., and D. Z. Rudner. 2009. "Novel secretion apparatus maintains spore integrity and developmental gene expression in Bacillus subtilis." *PLoS Genet* no. 5 (7):e1000566. doi:

10.1371/journal.pgen.1000566.

Dupuy, B., M. K. Taha, O. Possot, C. Marchal, and A. P. Pugsley. 1992. "PulO, a component of the pullulanase secretion pathway of Klebsiella oxytoca, correctly and efficiently processes gonococcal type IV prepilin in Escherichia coli." *Mol Microbiol* no. 6 (14):1887-94. doi: 10.1111/j.1365-2958.1992.tb01361.x.

Emsley, P., and K. Cowtan. 2004. "Coot: model-building tools for molecular graphics." *Acta Crystallogr D Biol Crystallogr* no. 60 (Pt 12 Pt 1):2126-32. doi: 10.1107/S0907444904019158.

Erickson, H. P., D. E. Anderson, and M. Osawa. 2010. "FtsZ in bacterial cytokinesis: cytoskeleton and force generator all in one." *Microbiol Mol Biol Rev* no. 74 (4):504-28. doi: 10.1128/MMBR.00021-10.

Errington, J. 2003. "Regulation of endospore formation in Bacillus subtilis." *Nat Rev Microbiol* no. 1 (2):117-26. doi: 10.1038/nrmicro750.

Fabiani, F. D., T. T. Renault, B. Peters, T. Dietsche, E. J. C. Galvez, A. Guse, K. Freier, E. Charpentier, T. Strowig, M. Franz-Wachtel, B. Macek, S. Wagner, M. Hensel, and M. Erhardt. 2017. "A flagellum-specific chaperone facilitates assembly of the core type III export apparatus of the bacterial flagellum." *PLoS Biol* no. 15 (8):e2002267. doi: 10.1371/journal.pbio.2002267.

Fimlaid, K. A., J. P. Bond, K. C. Schutz, E. E. Putnam, J. M. Leung, T. D. Lawley, and A. Shen. 2013. "Global analysis of the sporulation pathway of Clostridium difficile." *PLoS Genet* no. 9 (8):e1003660. doi: 10.1371/journal.pgen.1003660.

Fimlaid, K. A., O. Jensen, M. L. Donnelly, M. S. Siegrist, and A. Shen. 2015. "Regulation of Clostridium difficile Spore Formation by the SpoIIQ and SpoIIIA Proteins." *PLoS Genet* no. 11 (10):e1005562. doi: 10.1371/journal.pgen.1005562.

Fitzpatrick, A. W. P., S. Llabres, A. Neuberger, J. N. Blaza, X. C. Bai, U. Okada, S. Murakami, H. W. van Veen, U. Zachariae, S. H. W. Scheres, B. F. Luisi, and D. Du. 2017. "Structure of the MacAB-TolC ABC-type tripartite multidrug efflux pump." *Nat Microbiol* no. 2:17070. doi: 10.1038/nmicrobiol.2017.70.

Fredlund, J., D. Broder, T. Fleming, C. Claussin, and K. Pogliano. 2013. "The SpoIIQ landmark protein has different requirements for septal localization and immobilization." *Mol Microbiol* no. 89 (6):1053-68. doi: 10.1111/mmi.12333.

Fronzes, R., P. J. Christie, and G. Waksman. 2009. "The structural biology of type IV secretion systems." *Nat Rev Microbiol* no. 7 (10):703-14. doi: 10.1038/nrmicro2218.

Fronzes, R., E. Schafer, L. Wang, H. R. Saibil, E. V. Orlova, and G. Waksman. 2009. "Structure of a type IV secretion system core complex." *Science* no. 323 (5911):266-8. doi: 10.1126/science.1166101.

Fukumura, T., F. Makino, T. Dietsche, M. Kinoshita, T. Kato, S. Wagner, K. Namba, K. Imada, and T. Minamino. 2017. "Assembly and stoichiometry of the core structure of the bacterial flagellar type III export gate complex." *PLoS Biol* no. 15 (8):e2002281. doi: 10.1371/journal.pbio.2002281.

Gabadinho, J., A. Beteva, M. Guijarro, V. Rey-Bakaikoa, D. Spruce, M. W. Bowler, S. Brockhauser, D. Flot, E. J. Gordon, D. R. Hall, B. Lavault, A. A. McCarthy, J. McCarthy, E. Mitchell, S. Monaco, C. Mueller-Dieckmann, D. Nurizzo, R. B. Ravelli, X. Thibault, M. A. Walsh, G. A. Leonard, and S. M. McSweeney. 2010. "MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments." *J Synchrotron Radiat* no.

17 (5):700-7. doi: 10.1107/S0909049510020005.

Green, E. R., and J. Mecsas. 2016. "Bacterial Secretion Systems: An Overview." *Microbiol Spectr* no. 4 (1). doi: 10.1128/microbiolspec.VMBF-0012-2015.

Gu, S., V. E. Shevchik, R. Shaw, R. W. Pickersgill, and J. A. Garnett. 2017. "The role of intrinsic disorder and dynamics in the assembly and function of the type II secretion system." *Biochim Biophys Acta Proteins Proteom* no. 1865 (10):1255-1266. doi: 10.1016/j.bbapap.2017.07.006.

Guo, E. Z., D. C. Desrosiers, J. Zalesak, J. Tolchard, M. Berbon, B. Habenstein, T. Marlovits, A. Loquet, and J. E. Galan. 2019. "A polymorphic helix of a Salmonella needle protein relays signals defining distinct steps in type III secretion." *PLoS Biol* no. 17 (7):e3000351. doi: 10.1371/journal.pbio.3000351.

Hay, I. D., M. J. Belousoff, R. A. Dunstan, R. S. Bamert, and T. Lithgow. 2018. "Structure and Membrane Topography of the Vibrio-Type Secretin Complex from the Type 2 Secretion System of Enteropathogenic Escherichia coli." *J Bacteriol* no. 200 (5). doi: 10.1128/JB.00521-17.

Hay, I. D., M. J. Belousoff, and T. Lithgow. 2017. "Structural Basis of Type 2 Secretion System Engagement between the Inner and Outer Bacterial Membranes." *mBio* no. 8 (5). doi: 10.1128/mBio.01344-17.

Holm, L., and P. Rosenstrom. 2010. "Dali server: conservation mapping in 3D." *Nucleic Acids Res* no. 38 (Web Server issue):W545-9. doi: 10.1093/nar/gkq366.

Howard, S. P., L. F. Estrozi, Q. Bertrand, C. Contreras-Martel, T. Strozen, V. Job, A. Martins, D. Fenel, G. Schoehn, and A. Dessen. 2019. "Structure and assembly of pilotin-dependent and -independent secretins of the type II secretion system." *PLoS Pathog* no. 15 (5):e1007731. doi: 10.1371/journal.ppat.1007731.

Hu, J., L. J. Worrall, C. Hong, M. Vuckovic, C. E. Atkinson, N. Caveney, Z. Yu, and N. C. J. Strynadka. 2018. "Cryo-EM analysis of the T3S injectisome reveals the structure of the needle and open secretin." *Nat Commun* no. 9 (1):3840. doi: 10.1038/s41467-018-06298-8.

Hu, J., L. J. Worrall, M. Vuckovic, C. Hong, W. Deng, C. E. Atkinson, B. Brett Finlay, Z. Yu, and N. C. J. Strynadka. 2019. "T3S injectisome needle complex structures in four distinct states reveal the basis of membrane coupling and assembly." *Nat Microbiol* no. 4 (11):2010-2019. doi: 10.1038/s41564-019-0545-z.

Johnson, C. L., and A. Moir. 2017. "Proteins YlaJ and YhcN contribute to the efficiency of spore germination in Bacillus subtilis." *FEMS Microbiol Lett* no. 364 (7). doi: 10.1093/femsle/fnx047.

Kabsch, W. 2010. "Xds." *Acta Crystallogr D Biol Crystallogr* no. 66 (Pt 2):125-32. doi: 10.1107/S0907444909047337.

Kabsch, W., and C. Sander. 1983. "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features." *Biopolymers* no. 22 (12):2577-637. doi: 10.1002/bip.360221211.

Kanonenberg, K., O. Spitz, I. N. Erenburg, T. Beer, and L. Schmitt. 2018. "Type I secretion system-it takes three and a substrate." *FEMS Microbiol Lett* no. 365 (11). doi: 10.1093/femsle/fny094.

Karmazyn-Campelli, C., L. Rhayat, R. Carballido-Lopez, S. Duperrier, N. Frandsen, and P. Stragier. 2008. "How the early sporulation sigma factor sigmaF delays the switch to late

development in Bacillus subtilis." *Mol Microbiol* no. 67 (5):1169-80. doi: 10.1111/j.1365-2958.2008.06121.x.

Kelly, A., and P. S. Salgado. 2019. "The engulfasome in C. difficile: Variations on protein machineries." *Anaerobe* no. 60:102091. doi: 10.1016/j.anaerobe.2019.102091.

Khanna, K., J. Lopez-Garrido, Z. Zhao, R. Watanabe, Y. Yuan, J. Sugie, K. Pogliano, and E. Villa. 2019. "The molecular architecture of engulfment during Bacillus subtilis sporulation." *Elife* no. 8. doi: 10.7554/eLife.45257.

Kohler, R., K. Schafer, S. Muller, G. Vignon, K. Diederichs, A. Philippsen, P. Ringler, A. P. Pugsley, A. Engel, and W. Welte. 2004. "Structure and assembly of the pseudopilin PulG." *Mol Microbiol* no. 54 (3):647-64. doi: 10.1111/j.1365-2958.2004.04307.x.

Korotkov, K. V., J. R. Delarosa, and W. G. J. Hol. 2013. "A dodecameric ring-like structure of the N0 domain of the type II secretin from enterotoxigenic Escherichia coli." *J Struct Biol* no. 183 (3):354-362. doi: 10.1016/j.jsb.2013.06.013.

Korotkov, K. V., and W. G. Hol. 2013. "Crystal structure of the pilotin from the enterohemorrhagic Escherichia coli type II secretion system." *J Struct Biol* no. 182 (2):186-91. doi: 10.1016/j.jsb.2013.02.013.

Korotkov, K. V., T. L. Johnson, M. G. Jobling, J. Pruneda, E. Pardon, A. Heroux, S. Turley, J. Steyaert, R. K. Holmes, M. Sandkvist, and W. G. Hol. 2011. "Structural and functional studies on the interaction of GspC and GspD in the type II secretion system." *PLoS Pathog* no. 7 (9):e1002228. doi: 10.1371/journal.ppat.1002228.

Korotkov, K. V., E. Pardon, J. Steyaert, and W. G. Hol. 2009. "Crystal structure of the N-terminal domain of the secretin GspD from ETEC determined with the assistance of a nanobody." *Structure* no. 17 (2):255-65. doi: 10.1016/j.str.2008.11.011.

Korotkov, K. V., and M. Sandkvist. 2019. "Architecture, Function, and Substrates of the Type II Secretion System." *EcoSal Plus* no. 8 (2). doi: 10.1128/ecosalplus.ESP-0034-2018.

Korotkov, K. V., M. Sandkvist, and W. G. Hol. 2012. "The type II secretion system: biogenesis, molecular architecture and mechanism." *Nat Rev Microbiol* no. 10 (5):336-51. doi: 10.1038/nrmicro2762.

Kuhlen, L., P. Abrusci, S. Johnson, J. Gault, J. Deme, J. Caesar, T. Dietsche, M. T. Mebrhatu, T. Ganief, B. Macek, S. Wagner, C. V. Robinson, and S. M. Lea. 2018. "Structure of the core of the type III secretion system export apparatus." *Nat Struct Mol Biol* no. 25 (7):583-590. doi: 10.1038/s41594-018-0086-9.

Kuwana, R., Y. Kasahara, M. Fujibayashi, H. Takamatsu, N. Ogasawara, and K. Watabe. 2002. "Proteomics characterization of novel spore proteins of Bacillus subtilis." *Microbiology* no. 148 (Pt 12):3971-3982. doi: 10.1099/00221287-148-12-3971.

Langer, G., S. X. Cohen, V. S. Lamzin, and A. Perrakis. 2008. "Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7." *Nat Protoc* no. 3 (7):1171-9. doi: 10.1038/nprot.2008.91.

Laskowski, Roman A, Malcolm W MacArthur, David S Moss, and Janet M Thornton. 1993. "PROCHECK: a program to check the stereochemical quality of protein structures." *Journal of applied crystallography* no. 26 (2):283-291.

Leo, J. C., I. Grin, and D. Linke. 2012. "Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane." *Philos Trans R Soc Lond B Biol Sci* no. 367 (1592):1088-101. doi: 10.1098/rstb.2011.0208.

Levdikov, V. M., E. V. Blagova, A. McFeat, M. J. Fogg, K. S. Wilson, and A. J. Wilkinson. 2012. "Structure of components of an intercellular channel complex in sporulating Bacillus subtilis." *Proc Natl Acad Sci U S A* no. 109 (14):5441-5. doi: 10.1073/pnas.1120087109.

Lopez-Castilla, A., J. L. Thomassin, B. Bardiaux, W. Zheng, M. Nivaskumar, X. Yu, M. Nilges, E. H. Egelman, N. Izadi-Pruneyre, and O. Francetic. 2017. "Structure of the calcium-dependent type 2 secretion pseudopilus." *Nat Microbiol* no. 2 (12):1686-1695. doi: 10.1038/s41564-017-0041-2.

Lopez-Garrido, J., N. Ojkic, K. Khanna, F. R. Wagner, E. Villa, R. G. Endres, and K. Pogliano. 2018. "Chromosome Translocation Inflates Bacillus Forespores and Impacts Cellular Morphology." *Cell* no. 172 (4):758-770 e14. doi: 10.1016/j.cell.2018.01.027.

Lu, C., S. Turley, S. T. Marionni, Y. J. Park, K. K. Lee, M. Patrick, R. Shah, M. Sandkvist, M. F. Bush, and W. G. Hol. 2013. "Hexamers of the type II secretion ATPase GspE from Vibrio cholerae with increased ATPase activity." *Structure* no. 21 (9):1707-17. doi: 10.1016/j.str.2013.06.027.

Lycklama, A. Nijeholt J. A., and A. J. Driessen. 2012. "The bacterial Sec-translocase: structure and mechanism." *Philos Trans R Soc Lond B Biol Sci* no. 367 (1592):1016-28. doi: 10.1098/rstb.2011.0201.

Mancl, J. M., W. P. Black, H. Robinson, Z. Yang, and F. D. Schubot. 2016. "Crystal Structure of a Type IV Pilus Assembly ATPase: Insights into the Molecular Mechanism of PilB from Thermus thermophilus." *Structure* no. 24 (11):1886-1897. doi: 10.1016/j.str.2016.08.010.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. 2007. "Phaser crystallographic software." *J Appl Crystallogr* no. 40 (Pt 4):658-674. doi: 10.1107/S0021889807021206.

McKenney, P. T., A. Driks, and P. Eichenberger. 2013. "The Bacillus subtilis endospore: assembly and functions of the multilayered coat." *Nat Rev Microbiol* no. 11 (1):33-44. doi: 10.1038/nrmicro2921.

Mearls, E. B., J. Jackter, J. M. Colquhoun, V. Farmer, A. J. Matthews, L. S. Murphy, C. Fenton, and A. H. Camp. 2018. "Transcription and translation of the sigG gene is tuned for proper execution of the switch from early to late gene expression in the developing Bacillus subtilis spore." *PLoS Genet* no. 14 (4):e1007350. doi: 10.1371/journal.pgen.1007350.

Meisner, J., T. Maehigashi, I. Andre, C. M. Dunham, and C. P. Moran, Jr. 2012. "Structure of the basal components of a bacterial transporter." *Proc Natl Acad Sci U S A* no. 109 (14):5446-51. doi: 10.1073/pnas.1120113109.

Meisner, J., X. Wang, M. Serrano, A. O. Henriques, and C. P. Moran, Jr. 2008. "A channel connecting the mother cell and forespore during bacterial endospore formation." *Proc Natl Acad Sci U S A* no. 105 (39):15100-5. doi: 10.1073/pnas.0806301105.

Meuskens, I., A. Saragliadis, J. C. Leo, and D. Linke. 2019. "Type V Secretion Systems: An Overview of Passenger Domain Functions." *Front Microbiol* no. 10:1163. doi: 10.3389/fmicb.2019.01163.

Minamino, T., Y. V. Morimoto, N. Hara, and K. Namba. 2011. "An energy transduction mechanism used in bacterial flagellar type III protein export." *Nat Commun* no. 2:475. doi: 10.1038/ncomms1488.

Morgan, J. L. W., J. F. Acheson, and J. Zimmer. 2017. "Structure of a Type-1 Secretion System ABC Transporter." *Structure* no. 25 (3):522-529. doi: 10.1016/j.str.2017.01.010.

Morlot, C., and C. D. A. Rodrigues. 2018. "The New Kid on the Block: A Specialized Secretion System during Bacterial Sporulation." *Trends Microbiol* no. 26 (8):663-676. doi: 10.1016/j.tim.2018.01.001.

Morlot, C., T. Uehara, K. A. Marquis, T. G. Bernhardt, and D. Z. Rudner. 2010. "A highly coordinated cell wall degradation machine governs spore morphogenesis in Bacillus subtilis." *Genes Dev* no. 24 (4):411-22. doi: 10.1101/gad.1878110.

Murshudov, G. N., P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, and A. A. Vagin. 2011. "REFMAC5 for the refinement of macromolecular crystal structures." *Acta Crystallogr D Biol Crystallogr* no. 67 (Pt 4):355-67. doi: 10.1107/S0907444911001314.

Nicholson, W. L., N. Munakata, G. Horneck, H. J. Melosh, and P. Setlow. 2000. "Resistance of Bacillus endospores to extreme terrestrial and extraterrestrial environments." *Microbiol Mol Biol Rev* no. 64 (3):548-72. doi: 10.1128/mmbr.64.3.548-572.2000.

Nivaskumar, M., and O. Francetic. 2014. "Type II secretion system: a magic beanstalk or a protein escalator." *Biochim Biophys Acta* no. 1843 (8):1568-77. doi: 10.1016/j.bbamcr.2013.12.020.

Noback, Michiel A, Peter Terpstra, Siger Holsappel, Gerard Venema, and Sierd Bron. 1996. "A 22 kb DNA sequence in the cspB-glpPFKD region at 75° on the Bacillus subtilis chromosome." *Microbiology* no. 142 (11):3021-3026.

Noegel, A., U. Rdest, W. Springer, and W. Goebel. 1979. "Plasmid cistrons controlling synthesis and excretion of the exotoxin alpha-haemolysin of Escherichia coli." *Mol Gen Genet* no. 175 (3):343-50. doi: 10.1007/bf00397234.

Nouwen, N., H. Stahlberg, A. P. Pugsley, and A. Engel. 2000. "Domain structure of secretin PulD revealed by limited proteolysis and electron microscopy." *EMBO J* no. 19 (10):2229-36. doi: 10.1093/emboj/19.10.2229.

Palmer, T., and B. C. Berks. 2012. "The twin-arginine translocation (Tat) protein export pathway." *Nat Rev Microbiol* no. 10 (7):483-96. doi: 10.1038/nrmicro2814.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. 2004. "UCSF Chimera--a visualization system for exploratory research and analysis." *J Comput Chem* no. 25 (13):1605-12. doi: 10.1002/jcc.20084.

Piggot, P. J., and D. W. Hilbert. 2004. "Sporulation of Bacillus subtilis." *Curr Opin Microbiol* no. 7 (6):579-86. doi: 10.1016/j.mib.2004.10.001.

Popham, D. L., and C. B. Bernhards. 2015. "Spore Peptidoglycan." *Microbiol Spectr* no. 3 (6). doi: 10.1128/microbiolspec.TBS-0005-2012.

Pugsley, A. P., and B. Dupuy. 1992. "An enzyme with type IV prepilin peptidase activity is required to process components of the general extracellular protein secretion pathway of Klebsiella oxytoca." *Mol Microbiol* no. 6 (6):751-60. doi: 10.1111/j.1365-2958.1992.tb01525.x.

Pugsley, A. P., M. G. Kornacker, and I. Poquet. 1991. "The general protein-export pathway is directly required for extracellular pullulanase secretion in Escherichia coli K12." *Mol Microbiol* no. 5 (2):343-52. doi: 10.1111/j.1365-2958.1991.tb02115.x.

Ramirez-Guadiana, F. H., A. J. Meeske, X. Wang, C. D. A. Rodrigues, and D. Z. Rudner. 2017. "The Bacillus subtilis germinant receptor GerA triggers premature germination in response to morphological defects during sporulation." *Mol Microbiol* no. 105 (5):689-704. doi: 10.1111/mmi.13728.

Ramos-Silva, P., M. Serrano, and A. O. Henriques. 2019. "From Root to Tips: Sporulation Evolution

and Specialization in Bacillus subtilis and the Intestinal Pathogen Clostridioides difficile." *Mol Biol Evol* no. 36 (12):2714-2736. doi: 10.1093/molbev/msz175.

Redzej, A., M. Ukleja, S. Connery, M. Trokter, C. Felisberto-Rodrigues, A. Cryar, K. Thalassinos, R. D. Hayward, E. V. Orlova, and G. Waksman. 2017. "Structure of a VirD4 coupling protein bound to a VirB type IV secretion machinery." *EMBO J* no. 36 (20):3080-3095. doi: 10.15252/embj.201796629.

Rigden, D. J., and M. Y. Galperin. 2008. "Sequence analysis of GerM and SpoVS, uncharacterized bacterial 'sporulation' proteins with widespread phylogenetic distribution." *Bioinformatics* no. 24 (16):1793-7. doi: 10.1093/bioinformatics/btn314.

Rodrigues, C. D., X. Henry, E. Neumann, V. Kurauskas, L. Bellard, Y. Fichou, P. Schanda, G. Schoehn, D. Z. Rudner, and C. Morlot. 2016. "A ring-shaped conduit connects the mother cell and forespore during sporulation in Bacillus subtilis." *Proc Natl Acad Sci U S A* no. 113 (41):11585-11590. doi: 10.1073/pnas.1609604113.

Rodrigues, C. D., K. A. Marquis, J. Meisner, and D. Z. Rudner. 2013. "Peptidoglycan hydrolysis is required for assembly and activity of the transenvelope secretion complex during sporulation in Bacillus subtilis." *Mol Microbiol* no. 89 (6):1039-52. doi: 10.1111/mmi.12322.

Rodrigues, C. D., F. H. Ramirez-Guadiana, A. J. Meeske, X. Wang, and D. Z. Rudner. 2016. "GerM is required to assemble the basal platform of the SpoIIIA-SpoIIQ transenvelope complex during sporulation in Bacillus subtilis." *Mol Microbiol* no. 102 (2):260-273. doi: 10.1111/mmi.13457.

Salmond, G. P., and P. J. Reeves. 1993. "Membrane traffic wardens and protein secretion in gram-negative bacteria." *Trends Biochem Sci* no. 18 (1):7-12. doi: 10.1016/0968-0004(93)90080-7.

Sammito, M., C. Millan, D. D. Rodriguez, I. M. de Ilarduya, K. Meindl, I. De Marino, G. Petrillo, R. M. Buey, J. M. de Pereda, K. Zeth, G. M. Sheldrick, and I. Uson. 2013. "Exploiting tertiary structure through local folds for crystallographic phasing." *Nat Methods* no. 10 (11):1099-101. doi: 10.1038/nmeth.2644.

Sammons, R. L., G. M. Slynn, and D. A. Smith. 1987. "Genetical and molecular studies on gerM, a new developmental locus of Bacillus subtilis." *J Gen Microbiol* no. 133 (12):3299-312. doi: 10.1099/00221287-133-12-3299.

Schraidt, O., M. D. Lefebre, M. J. Brunner, W. H. Schmied, A. Schmidt, J. Radics, K. Mechtler, J. E. Galan, and T. C. Marlovits. 2010. "Topology and organization of the Salmonella typhimurium type III secretion needle complex components." *PLoS Pathog* no. 6 (4):e1000824. doi: 10.1371/journal.ppat.1000824.

Schraidt, O., and T. C. Marlovits. 2011. "Three-dimensional model of Salmonella's needle complex at subnanometer resolution." *Science* no. 331 (6021):1192-5. doi: 10.1126/science.1199358.

Serrano, M., A. D. Crawshaw, M. Dembek, J. M. Monteiro, F. C. Pereira, M. G. Pinho, N. F. Fairweather, P. S. Salgado, and A. O. Henriques. 2016. "The SpoIIQ-SpoIIIAH complex of Clostridium difficile controls forespore engulfment and late stages of gene expression and spore morphogenesis." *Mol Microbiol* no. 100 (1):204-28. doi: 10.1111/mmi.13311.

Serrano, M., F. Vieira, C. P. Moran, Jr., and A. O. Henriques. 2008. "Processing of a membrane protein required for cell-to-cell signaling during endospore formation in Bacillus subtilis." *J Bacteriol* no. 190 (23):7786-96. doi: 10.1128/JB.00715-08.

Setlow, B., E. Melly, and P. Setlow. 2001. "Properties of spores of Bacillus subtilis blocked at an intermediate stage in spore germination." *J Bacteriol* no. 183 (16):4894-9. doi: 10.1128/JB.183.16.4894-4899.2001.

Setlow, P. 2014. "Spore Resistance Properties." *Microbiol Spectr* no. 2 (5). doi: 10.1128/microbiolspec.TBS-0003-2012.

Sgro, G. G., T. R. D. Costa, W. Cenens, D. P. Souza, A. Cassago, L. Coutinho de Oliveira, R. K. Salinas, R. V. Portugal, C. S. Farah, and G. Waksman. 2018. "Cryo-EM structure of the bacteria-killing type IV secretion system core complex from Xanthomonas citri." *Nat Microbiol* no. 3 (12):1429-1440. doi: 10.1038/s41564-018-0262-z.

Sgro, G. G., G. U. Oka, D. P. Souza, W. Cenens, E. Bayer-Santos, B. Y. Matsuyama, N. F. Bueno, T. R. Dos Santos, C. E. Alvarez-Martinez, R. K. Salinas, and C. S. Farah. 2019. "Bacteria-Killing Type IV Secretion Systems." *Front Microbiol* no. 10:1078. doi: 10.3389/fmicb.2019.01078.

Slynn, G. M., R. L. Sammons, D. A. Smith, A. Moir, and B. M. Corfe. 1994. "Molecular genetical and phenotypical analysis of the gerM spore germination gene of Bacillus subtilis 168." *FEMS Microbiol Lett* no. 121 (3):315-20. doi: 10.1111/j.1574-6968.1994.tb07119.x.

Smith, T. J., S. A. Blackman, and S. J. Foster. 2000. "Autolysins of Bacillus subtilis: multiple enzymes with multiple functions." *Microbiology (Reading)* no. 146 (Pt 2):249-262. doi: 10.1099/00221287-146-2-249.

Spreter, T., C. K. Yip, S. Sanowar, I. Andre, T. G. Kimbrough, M. Vuckovic, R. A. Pfuetzner, W. Deng, A. C. Yu, B. B. Finlay, D. Baker, S. I. Miller, and N. C. Strynadka. 2009. "A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system." *Nat Struct Mol Biol* no. 16 (5):468-76. doi: 10.1038/nsmb.1603.

Stetsenko, Artem, and Albert Guskov. 2017. "An overview of the top ten detergents used for membrane protein crystallization." *Crystals* no. 7 (7):197.

Sun, D. X., R. M. Cabrera-Martinez, and P. Setlow. 1991. "Control of transcription of the Bacillus subtilis spoIIIG gene, which codes for the forespore-specific transcription factor sigma G." *J Bacteriol* no. 173 (9):2977-84. doi: 10.1128/jb.173.9.2977-2984.1991.

Tan, I. S., and K. S. Ramamurthi. 2014. "Spore formation in Bacillus subtilis." *Environ Microbiol Rep* no. 6 (3):212-25. doi: 10.1111/1758-2229.12130.

Terwilliger, T. C., R. W. Grosse-Kunstleve, P. V. Afonine, N. W. Moriarty, P. H. Zwart, L. W. Hung, R. J. Read, and P. D. Adams. 2008. "Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard." *Acta Crystallogr D Biol Crystallogr* no. 64 (Pt 1):61-9. doi: 10.1107/S090744490705024X.

Thorn, A., and G. M. Sheldrick. 2013. "Extending molecular-replacement solutions with SHELXE." *Acta Crystallogr D Biol Crystallogr* no. 69 (Pt 11):2251-6. doi: 10.1107/S0907444913027534.

Tosi, T., L. F. Estrozi, V. Job, I. Guilvout, A. P. Pugsley, G. Schoehn, and A. Dessen. 2014. "Structural similarity of secretins from type II and type III secretion systems." *Structure* no. 22 (9):1348-1355. doi: 10.1016/j.str.2014.07.005.

Traag, B. A., A. Driks, P. Stragier, W. Bitter, G. Broussard, G. Hatfull, F. Chu, K. N. Adams, L. Ramakrishnan, and R. Losick. 2010. "Do mycobacteria produce endospores?" *Proc Natl Acad Sci U S A* no. 107 (2):878-81. doi: 10.1073/pnas.0911299107.

Trokter, M., C. Felisberto-Rodrigues, P. J. Christie, and G. Waksman. 2014. "Recent advances in the structural and molecular biology of type IV secretion systems." *Curr Opin Struct Biol* no.

27:16-23. doi: 10.1016/j.sbi.2014.02.006.

Trouve, J., A. Mohamed, F. Leisico, C. Contreras-Martel, B. Liu, C. Mas, D. Z. Rudner, C. D. A. Rodrigues, and C. Morlot. 2018. "Structural characterization of the sporulation protein GerM from Bacillus subtilis." *J Struct Biol* no. 204 (3):481-490. doi: 10.1016/j.jsb.2018.09.010.

Uehara, T., K. R. Parzych, T. Dinh, and T. G. Bernhardt. 2010. "Daughter cell separation is controlled by cytokinetic ring-activated cell wall hydrolysis." *EMBO J* no. 29 (8):1412-22. doi: 10.1038/emboj.2010.36.

Voulhoux, R., G. Ball, B. Ize, M. L. Vasil, A. Lazdunski, L. F. Wu, and A. Filloux. 2001. "Involvement of the twin-arginine translocation system in protein secretion via the type II pathway." *EMBO J* no. 20 (23):6735-41. doi: 10.1093/emboj/20.23.6735.

Wagner, S., L. Konigsmaier, M. Lara-Tejero, M. Lefebre, T. C. Marlovits, and J. E. Galan. 2010. "Organization and coordinated assembly of the type III secretion export apparatus." *Proc Natl Acad Sci U S A* no. 107 (41):17745-50. doi: 10.1073/pnas.1008053107.

Weiss, G. L., A. K. Kieninger, I. Maldener, K. Forchhammer, and M. Pilhofer. 2019. "Structure and Function of a Bacterial Gap Junction Analog." *Cell* no. 178 (2):374-384 e15. doi: 10.1016/j.cell.2019.05.055.

Worrall, L. J., C. Hong, M. Vuckovic, W. Deng, J. R. C. Bergeron, D. D. Majewski, R. K. Huang, T. Spreter, B. B. Finlay, Z. Yu, and N. C. J. Strynadka. 2016. "Near-atomic-resolution cryo-EM analysis of the Salmonella T3S injectisome basal body." *Nature* no. 540 (7634):597-601. doi: 10.1038/nature20576.

Worrall, L. J., M. Vuckovic, and N. C. Strynadka. 2010. "Crystal structure of the C-terminal domain of the Salmonella type III secretion system export apparatus protein InvA." *Protein Sci* no. 19 (5):1091-6. doi: 10.1002/pro.382.

Yamagata, A., and J. A. Tainer. 2007. "Hexameric structures of the archaeal secretion ATPase GspE and implications for a universal secretion mechanism." *EMBO J* no. 26 (3):878-90. doi: 10.1038/sj.emboj.7601544.

Yan, Z., M. Yin, D. Xu, Y. Zhu, and X. Li. 2017. "Structural insights into the secretin translocation channel in the type II secretion system." *Nat Struct Mol Biol* no. 24 (2):177-183. doi: 10.1038/nsmb.3350.

Yin, M., Z. Yan, and X. Li. 2018. "Structural insight into the assembly of the type II secretion system pilotin-secretin complex from enterotoxigenic Escherichia coli." *Nat Microbiol* no. 3 (5):581-587. doi: 10.1038/s41564-018-0148-0.

Yip, C. K., T. G. Kimbrough, H. B. Felise, M. Vuckovic, N. A. Thomas, R. A. Pfuetzner, E. A. Frey, B. B. Finlay, S. I. Miller, and N. C. Strynadka. 2005. "Structural characterization of the molecular platform for type III secretion system assembly." *Nature* no. 435 (7042):702-7. doi: 10.1038/nature03554.

Zeytuni, N., K. A. Flanagan, L. J. Worrall, S. C. Massoni, A. H. Camp, and N. C. J. Strynadka. 2018a. "Structural and biochemical characterization of SpoIIIAF, a component of a sporulation-essential channel in Bacillus subtilis." *J Struct Biol* no. 204 (1):1-8. doi: 10.1016/j.jsb.2018.06.002.

Zeytuni, N., K. A. Flanagan, L. J. Worrall, S. C. Massoni, A. H. Camp, and N. C. J. Strynadka. 2018b. "Structural characterization of SpoIIIAB sporulation-essential protein in Bacillus subtilis." *J Struct Biol* no. 202 (2):105-112. doi: 10.1016/j.jsb.2017.12.009.

Zeytuni, N., C. Hong, K. A. Flanagan, L. J. Worrall, K. A. Theiltges, M. Vuckovic, R. K. Huang, S. C. Massoni, A. H. Camp, Z. Yu, and N. C. Strynadka. 2017. "Near-atomic resolution cryoelectron microscopy structure of the 30-fold homooligomeric SpoIIIAG channel essential to spore formation in Bacillus subtilis." *Proc Natl Acad Sci U S A* no. 114 (34):E7073-E7081. doi: 10.1073/pnas.1704310114.

Zeytuni, N., and N. C. J. Strynadka. 2019. "A Hybrid Secretion System Facilitates Bacterial Sporulation: A Structural Perspective." *Microbiol Spectr* no. 7 (1). doi: 10.1128/microbiolspec.PSIB-0013-2018.

Zeytuni, N., R. Uebe, M. Maes, G. Davidov, M. Baram, O. Raschdorf, A. Friedler, Y. Miller, D. Schuler, and R. Zarivach. 2014. "Bacterial magnetosome biomineralization--a novel platform to study molecular mechanisms of human CDF-related Type-II diabetes." *PLoS One* no. 9 (5):e97154. doi: 10.1371/journal.pone.0097154.