Structural and functional studies of bacterial cellulose secretion

Wiem Abidi

université PARIS-SACLAY

ETUDES STRUCTURALES ET FONCTIONNELLES DE SECRETION DE CELLULOSE BACTERIENNE

STRUCTURAL AND FUNCTIONAL STUDIES OF BACTERIAL CELLULOSE SECRETION

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 569, Innovation thérapeutique : du fondamental à l'appliqué (ITFA)
Spécialité de doctorat : Biochimie et biologie structurale
Unité de recherche : Institut de Chimie & Biologie des Membranes & des Nano-objets
Référent : Faculté de pharmacie

Thèse présentée et soutenue à Bordeaux, le 02 décembre 2021, par

Wiem ABIDI

Composition du Jury

Nicolas BAYAN Professeur, Université Paris-Saclay	Président
Éric CASCALES Directeur de recherche, Université Aix-Marseille	Rapporteur & Examinateur
Andrea DESSEN Directrice de recherche, Institut de Biologie Structurale, Grenoble	Rapporteur & Examinatrice
Alain FILLOUX Professeur, Imperial College London	Examinateur
Rémi FRONZES Directeur de recherche, Université de Bordeaux	Examinateur
Esther MARZA Maîtresse de conférences, Université de Bordeaux	Examinatrice

Direction de la thèse

Petya V. KRASTEVA Directrice de recherche, CBMN, Université de Bordeaux	Directrice de thèse

Thèse de doctorat
NNT: 2021UPASQ041

إلى روح جدّي العزيز علي عبيدي

To the memory of my beloved grandfather Ali Abidi

Acknowledgments

First and foremost, I would like to express my sincere gratitude to the very inspiring supervisor **Dr. Petya KRASTEVA** for her guidance, motivation and support. Thank you for giving me the opportunity to join SBB Lab for my master's internship first and to stay as a PhD student. Being an experimentalist as well, she provided continuous and valuable guidance to experiments and methodology. I felt trust and encouragement to explore new tracks to achieve our scientific goals. From the first steps, she gave me the right balance between freedom and guidance allowing me to work independently, she even gave me a more general lab responsibility that contributed to building my confidence and constructed a valuable experience.

I would like next to thank my ex-colleague and friend **Dr. Samira ZOUHIR** for the enriching discussions and support. We started working hand in hand on the same project, which was a valuable and fruitful experience. Thank you for the day-to-day advice and for making the PhD a fun experience both inside and outside the lab. I would also like to thank my colleague and friend **Dr. Marion DECOSSAS**, from whom I learnt a lot both professionally and personally. Thank you for your great support and for many life lessons. I couldn't ask for a better person to share an office with. A big thanks to the friend and colleague **Lucía TORREZ-SANCHEZ** for supporting me and for the great discussions we had. My PhD experience would not have been the same without you. Thanks to the intern and ex-colleague **Lucie PUYGRENIER**, I was lucky to work with you. I would like to thank my colleague **Dr. Axel SIROY** for the valuable advice and discussions. Thanks to **Dr. Stéphane ROCHE** for being our host in our first lab in Gif-sur-Yvette Paris and for the constructive discussions about X-ray crystallography.

Big thanks to **Dr. Rémi FRONZES** and **Dr. Esther MARZA** for their help and support when we first moved our lab to the IECB and for their continuous guidance and advice. Many thanks to all group members working with **Dr. Rémi FRONZES**, **Dr. Axel INNIS** and **Dr. Yaser HASHEM**'s for the informative discussions about the cryo-EM data collection and processing. Thanks also to **Armel BEZAULT** for his training in grid preparation and data collection.

These last years of my life were marked by this life-changing PhD journey, during which many people contributed to make it easier and enjoyable. To name a few of such people, I would like

to give many thanks to **Laure BATAILLE**, **Myriam MEDERIC** and **Thierry DAKHLI** for the motivational and supportive discussions.

Thanks to the new friends I made in IECB **Coralie ROBERT**, **Julien MARQUEVIELLE** and **Pierre NOTTELET** for the scientific and non-scientific discussions and for the great moments that we shared together. I would also like to thank my friends **Cyrine** and **Olfa**. I want to thank you for the great moments, for the nights and days that we spent together. Thanks also to **Mayen**, **Abir**, **Marwa** and **Houwaida** for their support. I also want to thank my friends **Eva**, **Mathilde** and **Yasmina**, I was lucky that you were the first friends I made in France after coming from my native country Tunisia. From the time of the amphitheaters of Paris Saclay university, you were always there for me, and I hope to continue sharing each-others important life steps.

It is a real pleasure to thank my sister and best friend **Imene** and my brother-in-law and friend **Hamza**, to whom I owe a lot. Thank you for your unconditional love and support. Many thanks go to my brother **Ahmed** who taught me perseverance.

From the bottom of my heart, I thank my parents **Neila** and **Hassan**, to whom I owe everything.

I could not have completed this dissertation without the support of my husband **Ahmed**, I am grateful for having you in my life. Thank you for your love, encouragement and patience.

Table of contents

Table of figures

Abbreviations

σ - Sigma factor

Arg/R - Arginine

Asn/N - Asparagine

Asp/D - Aspartic acid

ATP - Adenosine triphosphate

Bcs - Bacterial cellulose synthesis

Ca - Calcium

Ccs - Clostridium cellulose synthase

c-di-GMP - Bis-(3′-5′)-cyclic dimeric guanosine monophosphate

Cryo-EM - Cryo-electron microscopy

CSC - Cellulose synthase complex

Cys/C - Cysteine

DGC - Diguanylate cyclase

DNA - Deoxyribonucleic acid

E. coli - Escherichia coli

EL - Extracellular Loop

EPS - Extracellular polymeric substances

G. hansenii - *Gluconacetobacter hansenii*

Glu/E - Glutamic acid

GMP - Guanosine monophosphate

GT - Glycosyl transferase

G. xylinus - *Gluconacetobacter xylinus*

IF - interface

IM - Inner Membrane

kDa - Kilodalton

Met/M - Methionine

Mg - Magnesium

Mn - Manganese

OM - Outer Mmebrane

PDB - Protein data bank

PDE - Phosphodiesterases

PE - Phostphatidylethanolamine

pEtN - phophoethanolamine

Phe/F - Phenylalanine

PNAG - poly-N-acetylglucosamine

R. sphaeroides - *Rhodobacter sphaeroides*

Ser/S - Serine

SIMIBI - Single-recognition particle, MinD and BioD

S. Typhimurium - *Salmonella enterica* Serovar Typhimurium

SEC-SAXS - Size-exclusion chromatography-coupled, small angle X-ray scattering

TAC - transcription antitermination complex

TC - terminal complex

Thr/T - Threonine

TM - transmembrane

TMD - transmembrane domain

Trp/W - Tryptophan

Tyr/Y - Tyrosine

TPR - Tetratricopeptide repeat

UDP - Uridine diphosphate

UDPGP - UDP-glucose pyrophosphorylase

UTI - Urinary Tract Infectious

Val/V - Valine

Zn - Zinc

1 Summary of the thesis in French

1.1 Contexte et motivations

Les biofilms bactériens sont des communautés collaboratives, protégés par une matrice extracellulaire et formés par la plupart des espèces bactériennes [1], [2]. C'est une forme de vie préférée par les espèces bactériennes qui est impliquée dans la tolérance aux antibiotiques, l'acquisition de résistances par des processus de transfert horizontal de gènes et le développement de maladies chroniques [2]–[4]. Les biofilms représentent un problème de santé majeur dans les hôpitaux et dans l'industrie agroalimentaire. Les bactéries vivant en communautés dans les biofilms sont protégées par une matrice extracellulaire qui est constituée majoritairement d'eau et d'EPS (Extracellular Polymeric Substances, soit polymères extracellulaires) [1], [2], [5]. Les EPS sont principalement composés d'acides nucléiques, de lipides, de protéines et de polysaccharides [1]. Parmi ces polysaccharides se trouve la cellulose, composant majeur de la matrice extracellulaire chez de multiples souches commensales ainsi que pathogènes [6], [7]. De par sa résistance mécanique, la cellulose confère aux bactéries une protection contre les environnements hostiles. Dans les biofilms bactériens, la sécrétion de cellulose par les bactéries à Gram-négatif nécessite la co-expression de gènes multiples dont les différentes sous-unités qui en résultent s'assemblent pour former un complexe de sécrétion traversant l'intégralité de la paroi bactérienne [8], [9]. Ce complexe stable possède une partie périplasmique, une partie que réside dans la membrane interne et une partie cytoplasmique. La structure de ce mégacomplexe a été caractérisée en microscopie électronique grâce à la technique dite de particule isolée en coloration négative par le Dr. P. Krasteva et ses collaborateurs [9]. Malgré l'apport important de cette structure, la faible résolution du complexe n'a pas permis la distinction des différents composants protéiques. De plus, plusieurs travaux intéressés par l'étude de la sécrétion de la cellulose se sont focalisés sur les deux sous unités catalytiques, négligeant les sous unités cytoplasmiques régulatrices qui sont ainsi longtemps restées énigmatiques [10]–[12].

1.2 État de l'art

La sécrétion de cellulose chez les bactéries à Gram-négatif nécessite la co-expression de gènes multiples qui sont conservés et répandus chez de multiples bactéries commensales et pathogènes [8]. Le système de gènes bactériens est typiquement organisé en un ou deux opérons appelés *bcs* (*bacterial cellulose synthesis*) [8]. Les protéines codées par les gènes *bcs* comprennent des protéines catalytiques comme BcsA et BcsB, qui mènent la polymérisation et l'export de la cellulose à travers la membrane interne ; des protéines pour la modification

enzymatique dans le périplasme et l'export à l'extérieur de la cellule comme BcsG, BcsZ et BcsC ; et plusieurs protéines régulatrices dont le mécanisme exact reste inconnu comme BcsR, BcsQ, BcsE et BcsF [8]–[12].

1.2.1 Les sous unités catalytiques

1.2.1.1 BcsA et BcsB

Le gène *bcsA* et le gène *bcsB* sont répandus dans presque tous les opérons de Bcs et se trouvent généralement en tandem [8]. Les structures des sous unités-catalytiques BcsAB ont été étudiées en détail chez *R. sphaeroides* [10]–[12]. BcsA est une protéine d'environ 100 kDa qui correspond à la cellulose synthase du système Bcs et qui est formée par une région transmembranaire N-terminale, sous laquelle se trouve un domaine cytoplasmique de glycosyl transférase suivie d'un domaine pilZ qui se lie au c-di-GMP (di-guanosine monophosphate cyclique) [10], [13]. La région transmembranaire est un domaine d'export de la cellulose formé de huit hélices transmembranaires (TM1-8) [10]–[13]. Quatre de ces hélices TM sont N-terminales et les quatre autres sont C-terminales. BcsB quant à elle est considérée comme le partenaire catalytique de BcsA. En effet, des études ont montré que l'hélice transmembranaire de BcsB liée à BcsA, est essentielle à la polymérisation et au passage de la cellulose à travers la membrane interne des bactéries. BcsB est formée par un module périplasmique contenant deux domaines de liaison aux glucides et deux autres domaines adoptants le repliement *jelly roll* et insérée dans la membrane interne via une hélice transmembranaire qui vient compléter les huit hélices TM1-8 de BcsA [10]–[13].

1.2.2 Les sous unités régulatrices

1.2.2.1 BcsR et BcsQ

BcsR et BcsQ sont des sous-unités cytoplasmiques essentielles à la sécrétion de cellulose *in vivo*. Il a notamment été montré par nos collaborateurs qu'au niveau des pôles de la bactérie où BcsQ a été localisée, l'adhésion intercellulaire y est initiée par la production de cellulose [14].

1.2.2.2 BcsE

BcsE est une protéine cytoplasmique nécessaire à la sécrétion maximale de cellulose et qui possède un domaine de liaison au c-di-GMP – une molécule clé pour la signalisation intracellulaire [9], [14]. La protéine BcsE se lie au c-di-GMP via le motif RxxD du domaine

GIL (GGDEF I-site-Like), situé à son extrémité C terminale [15]. Ce motif similaire au site I des domaines GGDEF des diguanylate cyclases, les enzymes qui produisent le c-di-GMP à partir de GTP, est crucial à la fonction de BcsE. Des études de notre groupe et en collaboration ont aussi montré que la présence de BcsE dans le complexe est nécessaire pour la stabilité de la sous-unité catalytique BcsA et que les deux protéines interagissent directement à travers leurs modules de liaison au c-di-GMP (domaine GIL de BcsE et domaine PilZ de BcsA) [9]. Ces résultats nous ont alors montré l'importance d'étudier cette protéine en détail afin de déterminer son rôle et sa contribution dans l'assemblage et la formation du complexe Bcs.

1.2.3 Le second messager c-di-GMP contrôle la formation des biofilms

Dans la plupart des organismes bactériens, le changement entre l'état motile et l'état stationnaire sous forme de biofilm est exercé par le c-di-GMP, un second messager intracellulaire qui a des effets multiples permettant d'inhiber, directement ou indirectement, la motilité flagellaire et de stimuler la sécrétion des composants de la matrice extracellulaire tels que les exopolysaccharides [1], [16], [17]. La présence intracellulaire du c-di-GMP est contrôlée par deux activités enzymatiques opposées : les diguanylate cyclases avec leur domaine GGDEF le synthétisent et des phosphodiestérases possédant un domaine EAL ou HD-GYP le dégradent [18]–[20].

1.2.4 Objectifs de la thèse

Dans ce contexte, j'ai combiné études structurales et fonctionnelles afin d'étudier la sécrétion de la cellulose et son rôle dans la persistance des biofilms bactériens. Pour ceci, j'ai entrepris d'étudier trois sous-unités clés pour l'assemblage et la fonction du système de sécrétion de cellulose chez la bactérie modèle *E. coli.* Dans un premier temps l'objectif était de déterminer les structures à haute résolution des protéines cytoplasmiques BcsR, BcsQ et BcsE, vu les rôles importants qu'elles pourraient jouer. D'une autre part, afin de déterminer la structure du complexe de Bcs avec une meilleure résolution, j'ai utilisé la cryo-EM (cryo-microscopie électronique) par la technique dite de « particules isolées ».

1.3 Résultats et discussion

1.3.1 Étude de la structure cristallographique de BcsRQ

L'un des premiers travaux que j'ai pu entamer dans le contexte de mes recherches était l'étude de l'interaction de BcsR et de BcsQ. Une fois le protocole de coexpression/copurification mis au point, j'ai procédé à l'optimisation des conditions de cristallisation des deux protéines. Dans ce cadre j'ai pu obtenir des cristaux qui diffractent et qui nous ont permis de résoudre la structure du complexe à haute résolution. Ces résultats sont importants non seulement parce que la structure du complexe BcsRQ a été résolue pour la première fois, mais aussi parce qu'ils représentent un outil indispensable pour l'étude des interactions possibles entre BcsRQ et leur partennaire BcsE. La structure de BcsRQ a montré une dimérisation du complexec'est-à-dire un dimère de deux hétérodimères BcsRQ avec une surface assez importante d'interaction. Comme la protéine BcsQ est prédite à fonctionner comme une ATPase (Adénosine TriPhosphatase), notre objectif était de collecter différents états conformationnels de la protéine qui pourraient exister. Pour cela une analyse de trois types de cristaux a été effectuée. La protéine BcsRQ sans nucléotides et la protéine purifiée et cristalisée en présence d'ADP (Adénosine DiPhosphate) et/ou d'AppCp (analogue non hydrolysable de l'ATP). A partir de la collecte de jeux complets de données de diffraction sur les différents cristaux, nous avons constaté qu'il n'existait qu'un seul état conformationnel, celui avec la protéine en complexe avec l'ATP et cela malgré l'ajout d'EDTA (Éthylènediaminetétraacétique).

1.3.2 Étude de la structure cristallographique de BcsRQ avec deux domaine C-terminaux de BcsE

Quand j'ai commencé le travail sur les sous unités cytoplasmiques, l'interaction entre BcsE et BcsRQ n'avait pas été démontré. Comme pour BcsRQ, le but dans un premier temps était de mettre au point un protocole pour la coexpression/copurification de BcsRQE. Ceci nous a permis de confirmer nos attentes que BcsE interagit avec BcsRQ en se liant à la partie C-terminale de BcsQ. D'autant plus, nos résultats confirment que BcsR et BcsQ sont localisées dans le cytoplasme car elles sont récupérées dans la fraction soluble. Toutefois quand BcsR, BcsQ et BcsE interagissent, leur extraction nécessite un protocole d'extraction semblable à un protocole d'extraction d'une protéine membranaire contrairement à BcsRQ seuls. En effet, pour extraire et purifier le complexe BcsRQE, nous sommes amenés à utiliser des détergents. Ceci rendait non seulement la procédure de cristallisation difficile voire impossible, mais en plus le complexe protéique BcsRQE n'était pas produit en quantité suffisante. Pour surmonter le

problème, j'ai généré des constructions de BcsE en gardant soit la partie C-terminale soit la partie N-terminale en se basant sur les prédictions de structures secondaires. Les résultats obtenus révèlent deux constructions qui sont copurifiés avec BcsRQ, à savoir $BcsE^{349-523}$ et $BcsE^{217-523}$. La seconde étape était alors l'étude de la structure et du mécanisme d'intéraction de BcsRQE, pour ceci j'ai opté pour la technique de cristallographie aux rayons X et j'ai pu obtenir des cristaux qui diffractent d'abord de $BcsRQE^{349-523}$ et de BcsRQ $^{R156E}E^{217-523}$. Les structures cristallographiques ont montré que BcsE se lie effectivement à la partie C-terminale de BcsQ. En plus de révéler les déterminants structuraux d'interaction avec BcsQ, les deux structures de BcsRQE ont fourni des informations inattendues quant à l'interaction avec le second messager c-di-GMP. Dans la structure $BcsRQE^{349-523}$ le c-di-GMP se trouve lié en tant que dimère intercalé à chaque domaine GGDEF. Cependant, dans le complexe BcsRQ $^{R156E}E^{217-523}$, BcsE subit un changement conformationnel par rapport à la structure de $BcsE^{217-523}$ seul (résolue par Samira Zouhir), contribuant ainsi à la découverte d'un motif d'interaction avec le c-di-GMP jamais identifié jusque-là.

1.3.3 Étude de la structure du complexe de sécrétion par cryo-EM

Dans l'optique d'étudier l'architecture et l'assemblage du système de sécrétion dans son intégralité, j'ai utilisé la technique d'analyse de particules isolées par cryo-EM. Ceci a permis non seulement de confirmer l'architecture non-canonique atypique de ce système de sécrétion mais aussi de déterminer la structure de la protéine BcsB. Nous avons montré que cette dernière fait partie du cœur catalytique et cette structure nous a également permis de comprendre l'architecture asymétrique particulière du système de sécrétion, dû à une polymérisation de sa partie périplasmique. Bien que la structure d'une protéine homologue eût déjà été résolue, nos travaux ont montré pour la première fois la propriété d'auto-assemblage du macrocomplexe de cette caractéristique particulière de ce système de sécrétion de cellulose.

1.4 Conclusion et perspectives

Les dernières années représentent une révolution dans l'étude de la cellulose. Les études antérieures se sont intéressées aux sous-unités catalytiques des systèmes bactériens de biogenèse de la cellulose, pendant que les sous-unités régulatrices sont longtemps restées énigmatiques. Notre groupe de recherche a largement marqué cette révolution de par l'identification des rôles mécanistiques de sous-unité accessoires et essentielles, exprimées par les bactéries dans les biofilms bactériens. Ces travaux de recherche ont concerné différents

niveaux, à savoir l'expression, l'assemblage et la régulation du système de sécrétion de la cellulose. De plus, ces travaux sont un élément clé pour mieux comprendre le comportement des bactéries dans les biofilms, ainsi que les liens mécanistiques aboutissant à l'acquisition du phénomène de résistance aux antibiotiques. Le système bactérien de sécrétion de la cellulose fait partie d'un ensemble de systèmes conservés chez les Eucaryotes et les Procaryotes. Les différents résultats de recherches cités ci-dessus peuvent ainsi contribuer à long terme au développement de thérapies de nouvelle génération, sans risques associés de forte pression évolutive et de développement de résistances.

Nos travaux présentent également un intérêt majeur dans un tout autre domaine. En effet, la cellulose possède un fort intérêt biotechnologique, en tant que nano-matériel inoffensif, biocompatible et à coût réduit.

2 Introduction

Bacteria can adopt two main modes of growth in nature, the first being the free-swimming planktonic lifestyle, while the second is the formation of sessile biofilms [1], [2]. Biofilms are bacterial cells living in a community and embedded in a self-produced matrix, usually attached to a biotic or abiotic surface [1], [2]. The biofilm matrix provides the embedded bacteria with tolerance, fitness and protection against the external world, among other privileges [1], [3], [4]. Thereby from an evolutionary point of view, the bacteria have more survival reasons to optimize mechanisms for the secretion of a biofilm matrix. Biofilms can be seen as an empire where bacterial cells live in a community surrounded by a fortress, the stronger the fortress walls, the more protected are the individual cells. Although the biofilm lifestyle is a common trait, the secreted biopolymers in the extracellular matrix, also known as the Extracellular Polymeric Substances (EPS), differ between the species by the type of matrix components, the underlying secretion machineries, as well as the external cues that orchestrate the mechanisms engaged in the EPS secretion [1], [2], [5]. The EPS typically includes polysaccharides, proteins, nucleic acids and lipids (Table 1) [2]. One thing is for sure, the key signaling molecule c-di-GMP (Bis-(3′-5′)-cyclic dimeric guanosine monophosphate), promotes the synthesis of many of the EPS components in biofilms [6], [21]–[23]. Usually through a c-di-GMP-guided pathway, as a key example, many bacteria produce cellulose as a major component of their extracellular matrix. The polymer has a large number of advantages, such as excellent mechanical resistance and water retention and, together with curli, builds up the biofilms of important enterobacterial species such as *Salmonella enterica* Serovar Typhimurium (*S. Typhimurium*) and *Escherichia coli* (*E. coli*), [6], [7].

In the present work, I decipher key components and mechanisms that drive cellulose secretion in the biofilm matrix. In this introduction part, I will provide the necessary background for better appreciation of our findings uncovering the mechanisms of bacterial cellulose secretion. I will provide an overview of cellulose as one of the major polysaccharides secreted by bacteria in the extracellular biofilm matrix. I will also discuss the conservation of cellulose secretion systems, as well as the components involved in this secretory mechanism with the second messenger c-di-GMP in the center of this sophisticated process. As unraveling all the Bcs (bacterial cellulose synthesis) systems is beyond the aim of this project, I rather spot an emphasis on cellulose secretion in Gram-negative bacteria and more precisely on the type II Bcs secretion system also known as the *E. coli*-like Bcs system.

2.1 The matrix of biofilms

2.1.1 Key steps for biofilm formation

The process of biofilm formation is cyclic and, although it can feature various degrees of complexity depending on external cues and the specific bacterial species, it can be summarized in three major steps (Figure 1). The overall process is typically under the control of intracellular c-di-GMP-dependent signaling events and comprises the attachment of planktonic cells to a surface, the multiplication and maturation into a multicellular three-dimensional structure and finally the dispersion of bacterial cells that can then colonize new surfaces [1], [16], [24]. Biofilms are truly collaborative multicellular communities, where the extracellular matrix provides not only mechanical protection but also a medium for exchange of nutrients, horizontal gene transfer (including genes for antibiotics resistance) and immune system escape, when in a eukaryotic host.

Regarding biofilm dispersion, on one hand it allows the bacteria to disseminate, on the other, the bacterial cells that leave the biofilm become exposed and unprotected. The dispersion takes place when remaining in the biofilm displays more drawbacks than advantages, for example upon development of oxygen or waste product gradients that can lead to cell lysis [25]. To tackle such toxicity and survive, many signals can trigger the activation of phosphodiesterases (PDE) that degrade the c-di-GMP second messenger and thereby inhibit EPS secretion and increase the cells' motility [17].

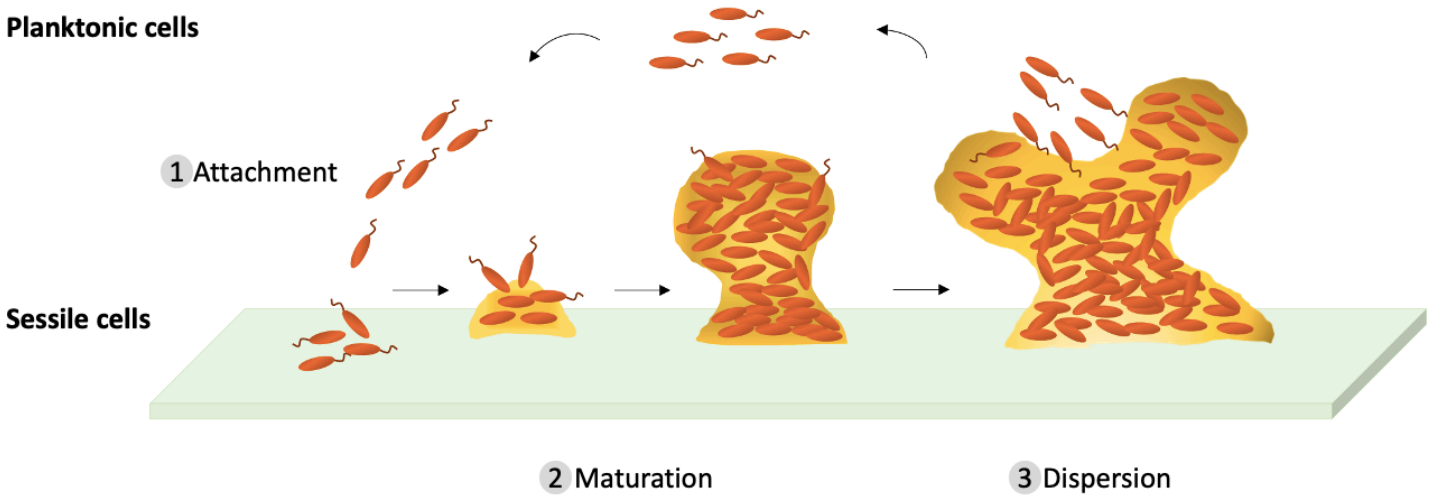


Figure 1. Key steps of the biofilm formation

Bacteria can reversibly attach to surfaces subsequently to the loss of motility and synthesis of extracellular adhesive substances providing strong anchors. Mature biofilms are then developed and are characterized by colony variance, complex architecture, and collaborative group behavior between functionally differentiated members. Finally, the dispersion leads to the release of highly motile planktonic cells, capable to colonize new surfaces [4], [25].

2.1.2 Cellulose, a major constituent of the bacterial biofilm maze, and more

The biofilm matrix is a hydrated mass that apart from intact cells contains a matrix of polysaccharides, proteins, nucleic acids and lipids [2]. The different secreted components play a plethora of functions that are relevant to the biofilm growth, some of which are summarized in Table 1.

Table 1. Examples of EPS and their functions in bacterial biofilms

EPS component	Function in biofilms
Polysaccharides	Adherence to biotic or abiotic surfaces Formation of the three-dimensional biofilm architecture Retention of water to keep the microenvironment hydrated Mechanical stability of biofilms
Proteins	Formation and stabilization of the polysaccharide matrix network Link the bacterial surface to the other EPS Electron donor or acceptor to ensure a redox activity in the biofilm matrix
DNA	Intercellular connection Bacterial cells adhesion Facilitate the exchange of genetic information in the biofilm matrix
Enzymes	Create pores to improve nutrients access Disrupt the biofilm matrix for dissemination

Adapted from [2]

Although it has been generally difficult to analyze the extracellular matrix components, technological developments have allowed for the characterization of many of them. One of these components is cellulose, which is the second major component of *S. Typhimurium* and *E. coli* biofilms. Together with curli proteins that upon secretion assemble into amyloid fibers, cellulose was discovered to provide agar-grown biofilms with the ability to form sophisticated

three-dimensional structures, such as the “rdar” (for red, dry, and rough) morphotype shown in figure 2 [7].

Figure 2. Rdar morphotypes of *E. coli* and *Salmonella* strains

Rdar morphotypes of *Salmonella* and *E. coli* strains grown on Congo red plates. Figure from [7]

Looking at a larger picture, cellulose has well-known reputation as the major constituent of the plants cell walls, making it the most abundant biopolymer on the planet Earth [8]. In addition to plants, however, cellulose is also produced by protists, fungi, animals, viruses, and of course bacteria, with Cyanobacteria being the first cellulose producers and the last common ancestor of cellulose biosynthesis genes [26], [27].

So, why do bacteria secrete cellulose? Simply put, for its exceptional water retention capacity, porosity, mechanical resistance and chemical simplicity, combined with its low antigenicity and ability to interact with additional components from the bacteria or their hosts [8].

In the present project, I will spot the light on the enterobacterial cellulose secretion system, as a widespread nanomachine for bacterial biofilm formation.

2.2 Structure and forms of bacterial cellulose

Cellulose is an unbranched homopolysaccharide of D-glucose molecules connected via acetal linkages joining the C1 and C4 carbons of the glucopyranose rings in a ß-configuration [28] (Figure 3A). The added glucose is rotated by approximatively 180° compared to its neighbor, as a result the repeating unit is the disaccharide cellobiose [29] (Figure 3A-B). Along the linear polysaccharide, each glucose moiety forms two hydrogen bonds with each of its neighbors. The C3 hydroxyl of a given glucose unit engages in a hydrogen bond with the ring oxygen of the glucose neighbor attached at the C4 carbon, while the C6 donates a hydrogen bond to the C2 hydroxyl of the same neighboring moiety (Figure 3A). On the other hand, the ring oxygen and the C2 hydroxyl accept a hydrogen bond from the C3 and C6 hydroxyls of preceding glucose unit, respectively, i.e. the glucose unit attached to the C1 carbon [28] (Figure 3A-B). These connections not only stabilize a coplanar conformation of the glucopyranose rings, moreover, the hydroxyl groups point away from the face of the pyranose rings resulting in a facility of engaging in different interactions with other cellulose polymers (Figure 3A). Even more, chemical modifications can occur on the building glucose units, such as the attachment phophoethanolamine (pEtN residues at the C6 position of the cellulose polymer produced by *E. coli* and *Salmonella*) [30] (Figure 3C). The chemical properties of the cellulose polymer all together bring the existence of different forms of bacterial cellulose. In bacteria, the cellulose is synthesized and extruded through the Bcs system and once at the cell surface, a subsequent assembly of the glucan chain can follow forming either a structured architecture also known as crystalline cellulose or more disordered structure known as amorphous cellulose.

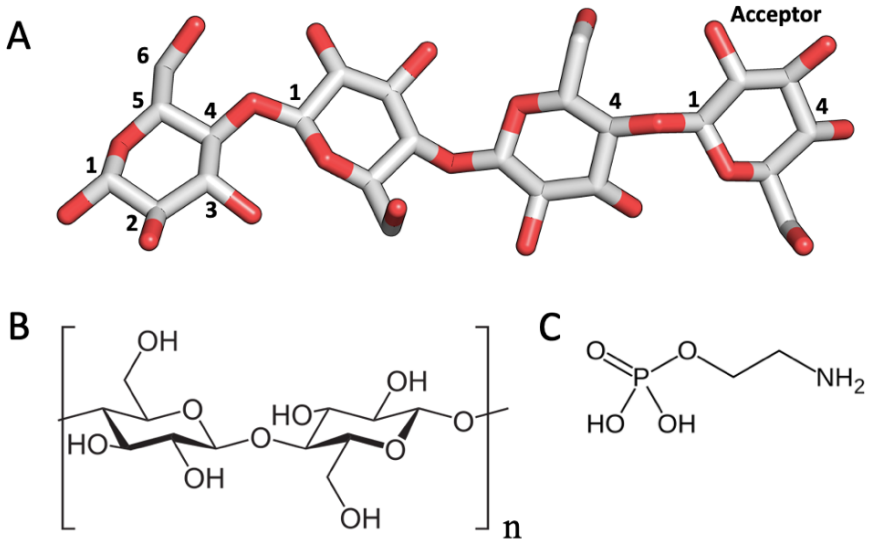


Figure 3. Cellulose structure

A. The structure of cellulose is shown in stick representation, carbon atoms are shown in gray, oxygen in red. The individual glucose units are connected via glycosidic ß-bonds between C1 and C4. Cellulose is elongated at its nonreducing end of the acceptor glucose and each glucose unit is rotated 180° compared to its neighbor. **B**. The cellobiose repeating unit is shown in Fischer projection. **C**. Structure of phosphoethanolamine.

2.2.1 Crystalline cellulose

Crystalline cellulose, has drawn the attention over the past decades due to its architecture reminding that of the plant cell walls. The different cellulose strands pack against each other forming hydrogen bonds between the equatorial hydroxyl groups and van der Waals forces between the glucopyranose rings (Figure 4A-B-C). Early studies using negative-stain, freeze-fracture electron microscopy revealed that in crystalline cellulose-secreting *Gluconacetobacter xylinus*, the cellulose synthase terminal complexes (TCs) form a line along the long axis of the cell surface which corresponds to the export sites of the elementary cellulose fibrils [31]–[33]. Once secreted in the extracellular environment, the elementary fibrils pack to form the cellulose ribbon, one per cell [31]–[33]. The cellulose biosynthesis involves a machinery encompassing the cell envelope, where a catalytic core composed of the BcsAB tandem is involved in the polymerization of the glucan chain using UDP (uridine diphosphate)-glucose as a substrate (Figure 6). In most systems of Gram-negative species, including crystalline cellulose secreting bacteria, the export of the nascent chain through the outer membrane (OM) involves an OM porin called BcsC. The structure of the latter subunit from pEtN-cellulose secreting *E. coli* has been solved recently and revealed that the protein adopts a ß-barrel fold at its C-terminus, however the pore diameter would be compatible with the extrusion of only a single-stranded cellulose chain [34]. Therefore, crystalline microfibril and ribbon formation must occur in the external environment, parallel to the longitudinal cell axis, however, it would be predetermined by the longitudinal TC microarray formation in the cell. Other factors, such as accessory subunits BcsHD thought to form individual cellulose chain conduits in the periplasm and a recently identified cytoskeletal element thought to stabilize the longitudinal TC array have also been shown to play key roles in cellulose crystallinity [35]–[37]. Nevertheless, the molecular mechanisms behind crystalline cellulose formation remain a missing piece in the puzzle and need to be better unraveled for the production of the bacterial cellulose for biotechnological applications.

2.2.2 Amorphous cellulose

Amorphous cellulose is disordered and non-structured cellulose that does not follow a pattern of regularly packed fibrils or layers (Figure 4B, bottom) and is generally believed to be easier to unravel when compared to the crystalline type. Interestingly, amorphous cellulose can also be commonly found in biofilms with high crystalline cellulose I content, marking the borders between large crystalline domains.

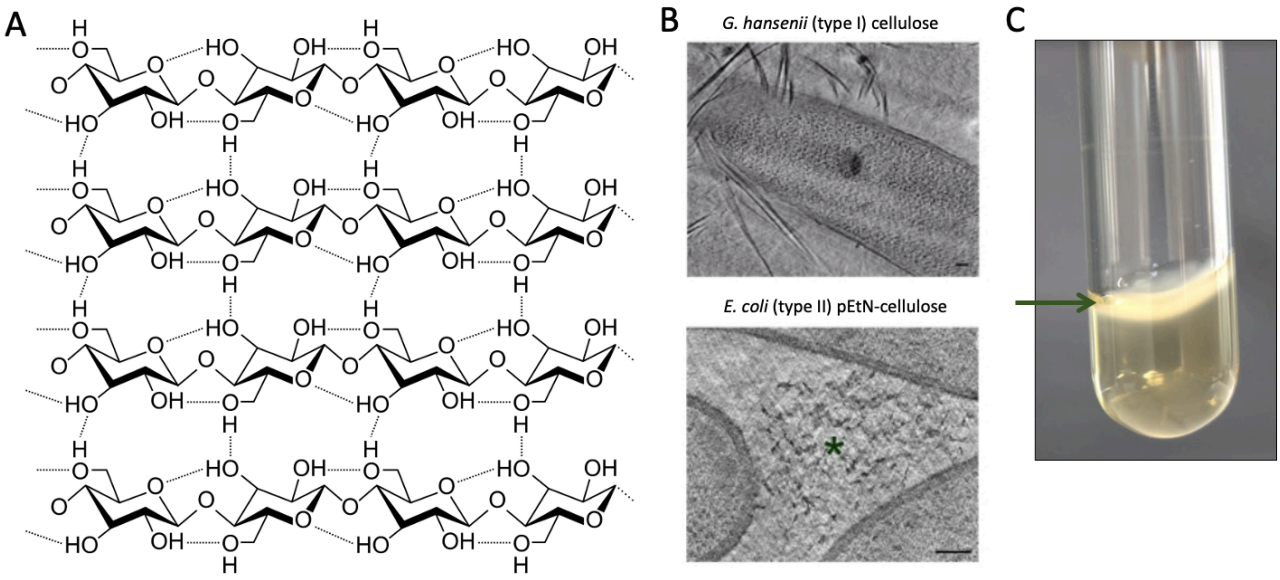


Figure 4. Bacterial cellulose structure

A. Structure of the crystalline cellulose formed through extensive inter and intramolecular hydrogen bonds and Van der Waals forces. Image by Luca Laghi, reproduced under license CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0/legalcode). **B**. Cryo-electron micrographs of secreted bacterial cellulose. Figures from [37] Above micrograph: a biofilm-embedded *Gluconacetobacter hansenii* cell surrounded by crystalline cellulose ribbons. Below micrograph: amorphous pEtN-cellulose, marked by an asterisk, secreted by the commensal *E. coli* 1094 strain. **C**. Crystalline cellulose secreted by *Gluconacetobacter xylinus*-containing medium. The cellulose pellicle formed at the water-air interface is shown by a green arrow.

2.3 Types of bacterial cellulose secreting systems

The mechanisms governing the secretion of bacterial cellulose have sparked intense interest since the polymer's discovery in bacteria at the end of the 19th century [38]. Many studies have since aimed at the elucidation of cellulose biosynthesis mechanism and the subunits involved in the process. Consequently, many nomenclatures have been attributed to the different subunits, so in 2015 Römling and Galperin revisited the classification of *bcs* operons and proposed a novel unified nomenclature for *bcs* proteins and their encoding genes [8]. The same

study highlighted the diversity among the different coding operons among bacterial species and clades and divided the majority of *bcs* loci into three major types [8] (Figure 5A-B) (see below for more details). The common point to all three systems is the catalytic subunit BcsA with its partner BcsB (Figures 5A-B-6). BcsA is the inner membrane protein composed of multiple transmembrane helices that stand above two cytosolic domains, a glycosyltransferase domain that catalyzes the polymerization of UDP-glucose into the β-1,4-linked glucan and a C-terminal c-di-GMP-sensing PilZ domain [10], [11], [39]. The majority of BcsB's amino acid sequence occupies a large periplasmic donut-shaped module, comprising two flavodoxin-like domains and two carbohydrate-binding domains, the latter of which are proposed guide the nascent polysaccharide to the outer membrane [10], [11], [40] (Figure 6). In addition, BcsB is anchored in the inner membrane by its C-terminal tail, composed of an amphipathic extracytosolic helix and a transmembrane anchor helix. Previous *in vitro* studies have demonstrated that, in both *E. coli* and *Rhodobacter sphaeroides (R. sphaeroides*), this C-terminal tail is indispensable for cellulose polymerization, thereby, BcsB is characterized as the co-catalytic subunit, or co-polymerase. In several cases, BcsA and BcsB genes are fused and likely post-translationally divided into two proteins [41]. In addition to these core catalytic subunits, most *bcs* operons in Gram-negative bacteria encode a periplasmic hydrolase BcsZ and an outer membrane (OM) secretory subunit BcsC composed of an N-terminal tetratricopeptide repeat (TPR)-rich periplasmic domain and a C-terminal OM porin domain.

Most cellulose secretion systems feature multiple additional Bcs subunits, many of which are indispensable for cellulose secretion. In addition, certain subunits typically co-occur with others, which prompted the recent classification of cellulose secretion systems based on the accessory to the catalytic BcsAB tandem subunits [8].

2.3.1 Type I cellulose secretion system

The type I cellulose secretion system is found in α, ß and γ subdivisions of proteobacteria. This type includes the core components genes *bcsABZC*, as well as the hallmark of this type of secretion system gene *bcsD* (Figure 5A-B). In *G. xylinus*, the encoded protein BcsD is proposed to localize to the periplasm and to contribute to the packing of the glucan chains into ordered cellulose ribbons also known as type I crystalline cellulose [8], [42]. Added to that, upstream of the *bcsD* gene is found *bcsH* (also known as "*ccpAx*" or "ORF2") whose protein product interacts with BcsD and is also suggested to play a role in crystalline fibril assembly [35]. Nevertheless, whereas the BcsH subunit seems limited to the *Komagataeibacter* lineage, many

types I *bcs* operons are found in non-crystalline cellulose-secreting bacteria, suggesting broader roles for the hallmark BcsD subunit. Additional Bcs components can also be found in these Type I systems, such as the BcsO, BcsP, and BcsQ subunits but their functions have not been characterized yet [8].

2.3.2 Type II cellulose secretion system

The type II cellulose secretion systems are widespread among ß- and γ-Proteobacteria and especially enterobacterial pathogens [8]. In *Salmonella* and *E. coli*, the *bcs* locus is divided into two divergently arranged operons, *bcsEFG* and *bcsRQABZC* [8] (Figure 5A-B). The discriminating components of this type are the cytosolic c-di-GMP-sensing subunit BcsE and the periplasmic membrane-anchored BcsG responsible for installing the pEtN modification in the secreted amorphous cellulose [30]. In addition, the BcsR, BcsQ and BcsF subunits, whose functions are discussed below in more details, are also encoded by the respective *bcs* operons.

2.3.3 Type III cellulose secretion system

The type III secretion systems are limited to some cyano- and α-Proteobacterial species [8]. They lack the three subunits BcsD, BcsE and BcsG, whereas BcsC subunit is replaced by another outer membrane TPR-rich subunit, BcsK, which lacks the C-terminal porin module [8] (Figure 5A-B).

2.3.4 Other types of cellulose secretion

There are additional types of cellulose secretion systems that don't fit into the standard picture of any of the types above or in some cases can feature a hybrid architecture. Examples include systems where the type I hallmark *bcsD* gene, can be encoded with the type II *bcsE(F)G* cluster. In other examples, such as in the cellulose-producing cyanobacterium *Thermosynechococcus vulcanus,* both *bcsB* and *bcsC* genes are missing and the BcsA-mediated cellulose polymerization and extrusion are assisted by an efflux pump-like tandem composed of an inner-membrane HlyD-like subunit and an outer membrane TolC-like exporter [43]. Further 'outsiders' involve *Pseudomonas fluorescens* SBW25, which is known to secrete acetylated cellulose and exhibits a hybrid *wss* operon, combining BcsQ- and BcsABZC-like components, together with an acetylation complex homologous to the alginate secretion system. In accordance with the omnipresence of cellulose secretion machineries among many diverse

organisms, some Gram-positive bacteria are also found to secrete cellulose. Studies on *Streptomyces coelicor* have shown that this actinomycete secretes cellulose in order to adapt to osmotic responses or growth of the hyphal tips [44], [45] and uses a yet different cellulose secretion system, some components of which appear to have a fungal origin.

Covering all the so far described systems of cellulose secretion is beyond this work, however this part highlights both the diversity and the complexity of the cellulose secreting systems. Moreover, with the ever-growing abundance of genomics data, the arborescence of Bcs secretion systems keeps growing and embraces now many Gram-positive bacteria, for whom many questions remain unanswered, especially concerning exopolysaccharide secretion through the thick peptidoglycan layer or the Gram-positive cell envelope devoid of an outer membrane [46].

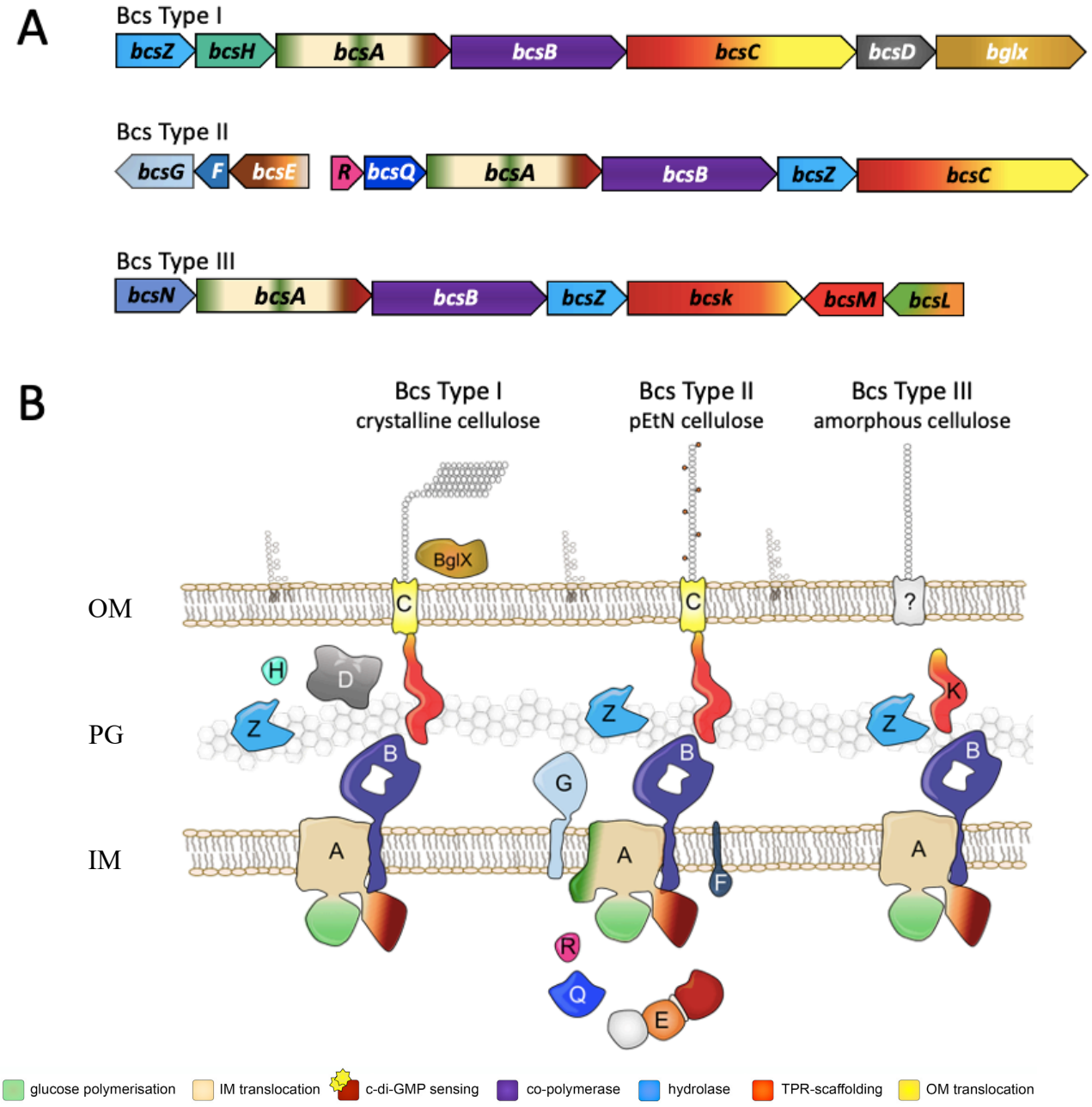


Figure 5. Prevalent types of Bcs systems

The three major types of the Bcs systems and their corresponding operons organizations as classified in [8], [47]. **A.** Examples of *bcs* operon organization and mosaicity for the three major types of cellulose secretion systems. **B.** Thumbnail representation of the topology of Bcs subunits relative to the bacterial envelope for the three types of Bcs systems. Color-coding as in A; the main conserved functional modules are annotated at the bottom. IM: inner membrane; PG: peptidoglycan; OM: outer membrane.

2.4 Mechanism for cellulose biosynthesis

Across the various organisms that secrete cellulose, a shared mechanism of cellulose polymerization is determined by an enzyme belonging to the glycosyl transferase family 2 [10], [48], [49]. The latter catalyzes the reaction of cellulose synthesis by the addition of one glucose at a time to the nascent polysaccharide using a preactivated nucleotide sugar substrate, UDP-glucose. UDP-glucose is made available by a dedicated enzyme, UDP-glucose pyrophosphorylase or UDPGP, which is essential for cellulose secretion [50]. Upon incorporation of the glucose moiety from the substrate into the nascent polysaccharide, the UDP product is then released back in the cytosol for its recycling. The formerly attached glucose through an α-configuration to the UDP, is inverted to be consequently attached to the glycosidic chain in a ß-configuration [10], [51]. The catalytic reaction is basically a SN2-like substitution reaction, in which the nucleophile C4-hydroxyl group (acceptor), attacks the C1 carbon of the donor glucose (Figure 3A). This reaction is facilitated by both a deprotonation of the nucleophile and a stabilization of the substrate's pyrophosphate group via a divalent metal cation, usually Mg^{2+} or Mn^{2+} [10], [48], [52]. Once the glycosidic transfer is completed at the non-reducing end in the cytosol, the glucan chain exits the bacterial cell with the reducing end pointing first to the periplasm and subsequently the extracellular milieu [10], [13].

2.5 Subunits for bacterial cellulose secretion

2.5.1 Polymerization and transport across the bacterial envelope

2.5.1.1 BcsAB tandem

The *bcsA* and *bcsB* genes are widespread in almost all cellulose synthase operons, and are usually found adjacently in a tandem [8] (Figure 5A). The structures of the catalytic BcsAB have been thoroughly studied in *R. sphaeroides,* whose Bcs system belongs to the type III cellulose secretion systems and secreted an amorphous polymer [8]. The BcsA subunit, or cellulose synthase protein, is roughly a 100 kDa-sized subunit formed by a non-continuous transmembrane region, a glycosyl transferase cytosolic domain and a C-terminal pilZ domain that binds the synthase-activating second messenger dinucleotide c-di-GMP (the mechanism underlying the c-di-GMP signaling is described in more details below) [9], [10]. The transmembrane region is an α-helical transmembrane cellulose export domain (TMD), which in *R. sphaeroides* is formed by eight transmembrane helices (TM1-8), four of which are N-terminal and the other four are C-terminal to the central GT domain. In addition, a single TM

helix of the partner BcsB protein completes the export domain and is indispensable for the catalytic mechanism [10] (Figure 6). Whereas BcsA's conserved glycosyl transferase (GT) domain is inserted between TM4 and TM5 [53], its active site is capped by a so-called 'gating loop' inserted between TM7 and and TM8 of the export domain and requires c-di-GMP binding to the protein to allow for substrate entry. The c-di-GMP sensing is carried out by a C-terminal cytosolic PilZ domain adopting a ß-barrel fold [10] (Figure 6). The GT domain interacts with the inner leaflet of the cytosolic membrane via three amphipathic interface helices (IF1-3). To date, all structures of the duo BcsAB co-crystallized with a translocating polysaccharide, which was apparently copurified with BcsAB as the polysaccharide was not added during the purification steps [10]–[12]. The structural studies revealed that BcsA's glycosyl transferase domain is able to accommodate 10 glucose units of the nascent polysaccharide, whose glucopyranose rings form van der Waals interactions with residues Met300, Phe301, Phe316 , Trp383 of IF2, Phe419, Phe426, Tyr433, Phe441,Val551, Val555, Trp558 [10]. The equatorial hydroxyl groups of the glucose units form hydrogen bonds with Tyr80, Asn118, His276, Asn412, Arg423, Glu439, Tyr455, Ser476, Glu477 [10]. While the van der Waals interactions of the translocation glucan chain encompass IF1, IF2, TM5, 5/6 loops, and TM8, the hydrogen bonds are made with TM3, TM4, GT, TM5, 5/6 loops, TM6, and IF3 [10]. The structure of BcsA's GT domain harbors a conserved motif formed by three non-sequential aspartic acids, as well as a Glutamine-Glutamine/Arginine-a random residue-Arginine-Tryptophan motif (or D, D, D, Q(Q/R)xRW) [10]. The latter signature is also characteristic in other ß-glycosyltransferases [10]–[12]. In *R. sphaeroides*, the first two Aspartic Acids (Asp179, Asp246) of the "D,D,D,Q(Q/R)xRW" are involved in UDP coordination, while the third aspartic acid (Asp 343) likely constitutes the catalytic base. This third astaptate is part of a strictly conserved TED (Threonine-Glutamate-Aspartate) motif of the so-called 'finger helix', which is also involved in gating loop relaxation and polysaccharide translocation [10]. It is conventional now that the polysaccharide extrusion through BcsA transmembrane channel is assisted by its membrane-anchored partner BcsB [10]. A study demonstrated that cellulose synthesis *in vitro* requires both BcsA and BcsB, and in particular BcsB's C-terminal tail of an amphipatic and a transmembrane helices [48]. This interdependence is further supported by the fact that in some species, *bcsB* gene is fused with *bcsA* [42]. The latter observation has also long supported the hypothesis of an equimolar assembly of BcsAB across the bacterial domain of life, however we will show here that such an assumption is over-simplistic and that at least in the *E. coli*-like Bcs biosynthetic platform, the BcsAB tandem can feature a markedly different stoichiometry allowing for additional functionalities within the secretion system.

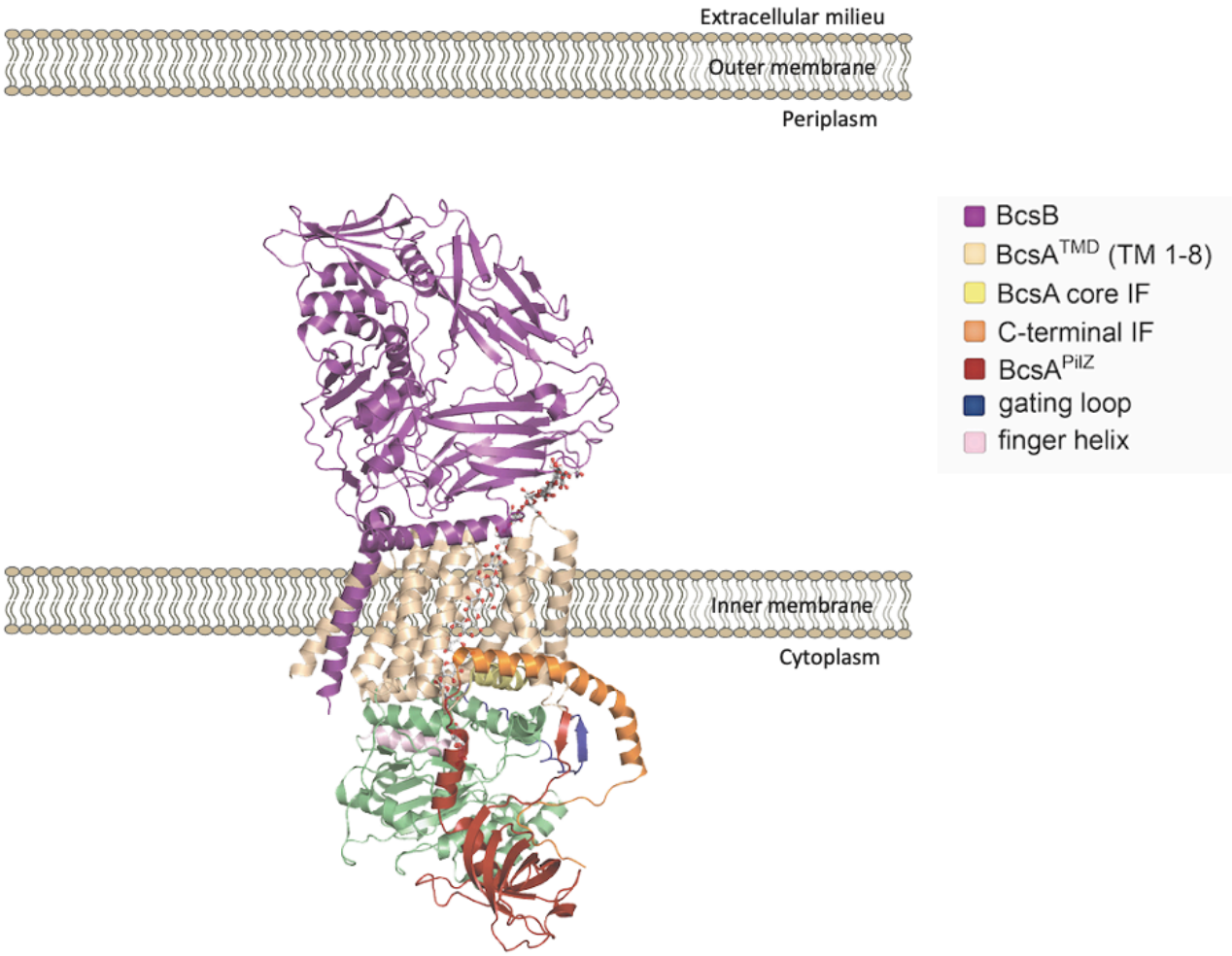


Figure 6. BcsAB cellulose synthase complex.

Crystal structure of the complex showing a 1:1 BcsAB assembly in cartoon representation and a co-purified cellulose polymer in sticks (from protein data bank entry 4p00) [10].

2.5.1.2 BcsC, the ß-barrel porin

BcsC is a ~130 kDa outer membrane protein whose amino acid sequence starts by a signal peptide. The mature protein, afterwards the signal peptide cleavage, adopts ~19-tetratricopeptide (TPR) repeats that most likely localize in the periplasm, guide the nascent polysaccharide and/or interact with the peptidoglycan, while the C-terminal ~400 residues adopt a ß-barrel conformation and function as an outer membrane porin for cellulose export (Figure 7A). Thus, BcsC is the second checkpoint export protein after the IM (Inner Membrane) BcsA synthase. More generally, the scheme of the porin-like domain preceded by periplasmic TPR-rich modules is a recurrent theme across exopolysaccharide secretion systems. In the alginate biosynthetic system, the secretion of alginate requires the AlgK-AlgE duo, where AlgK provides the periplasmic TPR scaffold, and AlgE functions as the outer membrane porin. In

addition, the PelE-PelB complex from the Pel system exhibits the same pattern with periplasmic TPR modules from (both PelE and PelB) and a porin domain (from PelB), as well as the PgaA protein from the PNAG (poly-N-acetylglucosamine) secretion machinery, which similarly to BcsC incorporates both TPR modules and a porin domain in the same polypeptide chain [54]. The crystal structure of a truncated versions of BcsC have been reported recently. One of these structures encompasses the last TPR motif and the C-terminal ß-barrel from the *E. coli* homolog. The porin is formed by 16 ß-strands whose connecting loops are longer on the extracellular side and form a dome-shaped structure on the cell surface [34]. The lumen of the ~15 Å-wide channel is highly electronegative and is suggested to facilitate the insertion of the hydrated zwitterionic pEtN-cellulose [34] (Figure 7B). The channel permeability is controlled by the extracellular loop connecting 8 ß-barrel strands 15 and 16 (also called the 'gating loop') and is stabilized by two conserved Tyrosines (Tyr1025 and Tyr1124) belonging to extracellular loop 6 and the gating loop, respectively [34]. Furthermore, both aromatic and hydrophilic residues line the lumen of the channel suggesting that facilitated cellulose diffusion relies on both aromatic stacking and hydrogen bonding with the polysaccharide [34]. Interestingly, the C-terminal 15 residues were found inserted as an elongated loop into the lumen of the porin domain, where it is stabilized by a number of interactions, most important of which is the stabilization of the conserved Y1157 residue midway across the channel [34]. Another recent study used SEC-SAXS (Size-exclusion chromatography-coupled, small angle X-ray scattering) on a construct covering most of the TPR repeats from another BcsC homolog and revealed an overall extended solenoid architecture typical to the TPR rich periplasmic scaffolds that can span most of the intermembrane width of the periplasm [55]. The same study also reported a crystal structure of five sequential TPR repeats from the N-terminus of the respective BcsC homolog. Interestingly, the five protomers in the asymmetric unit adopted three different conformations suggesting an overall flexible BcsC conformation that can adapt to variability in the intermembrane periplasmic distance [46], [55] (Figure 7A).

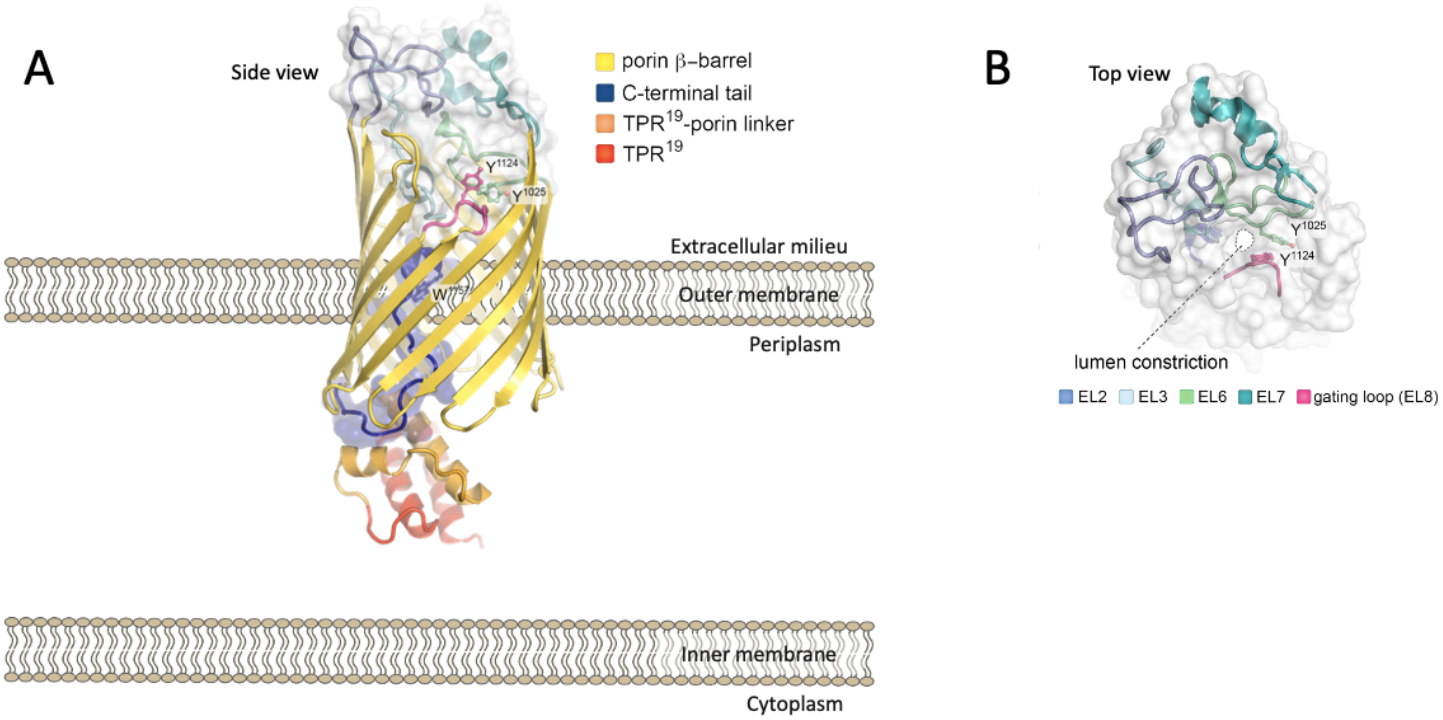


Figure 7. Architecture of the outer membrane pore BcsC

A-B. Crystal structure of *E. coli* BcsC$^{707\text{-}1157}$ (from protein data bank entry 6tzk) encompassing TPR19, a connecting linker region and the outer membrane porin domain [34]. The protein is shown in cartoon, the extracellular loops and luminal C-terminal tail are also shown as transparent surface. The lumen constriction proximal to the extracellular surface is seen in B, the gating π-stacking residues Y^{1025} and Y^{1124} from extracellular loops EL6 and EL8 are shown as sticks.

2.5.2 Accessory subunits

2.5.2.1 BcsG and the phosphoethanolamine modification

Together with BcsE and BcsF, BcsG is specific to the type II *E. coli*-like secretion systems. In *E. coli* it is associated with the *bcsEFG* operon, that is localized next to the *bcsRQABZC* [8]. BcsG is roughly 60 kDa and its large C-terminal domain had been first predicted to lie in the periplasm and belong to the phosphatase alkaline family [56]. The role of this subunit had remained elusive until recent studies showed that BcsG is responsible for adding phosphoethanolamine (pEtN) groups to the nascent polysaccharide, in both *Salmonella* and *E. coli* [30], [57] (Figure 3C). The pEtN groups are likely added from the phostphatidylethanolamine (PE) lipids of the inner membrane. Solid-state nuclear magnetic resonance spectroscopy analysis revealed that a pEtN group attach to the C6 of the glucose building blocks and initial quantification analyses estimated that up to half of the residues can carry the pEtN modification in the secreted polysaccharide [30]. The pEtN transferase activity

is indeed carried out by the periplasmic C-terminal domain whereas the N-terminal part for the protein in embedded in the inner membrane. In line with biofilm formation, by interacting with curli fibers, pEtN cellulose was also shown to play a key role in promoting the adhesion of uropathogenic *E. coli* cells to bladder cell *in vitro* [57]. The adhesion of cells to a surface is a crucial step of biofilm formation [2]. While curli are required for cell adhesion, pEtN cellulose strengthens the adhesion of curli to the host cell surface [57]. By consequence, pEtN cellulose together with curli was proposed to enhance Urinary Tract Infectious (UTI) *E. coli* pathogenesis [57]. Lately, high resolution BcsG structures of the C-terminal domain from *Salmonella enterica* and *E. coli* have been reported [58], [59]. The crystallographic snapshots from both studies were in a full accordance, and both C-terminal domains adopt a typical phosphatase alkaline fold, with 7 stranded ß-sheets sandwiched between α-helices (Figure 8). Both structures exhibit almost identical folds (0.4Å R.M.S.D) and contain a Zn^{2+} ion interacting with different moieties of amino acid Cys243, Ser278, Glu442, and His443, altogether form the active site (Figure 8) [58], [59]. The function of BcsG is not fully understood yet, as a BcsG deletion in *Salmonella typhimurium* severely decreased not only the efficiency of cellulose secretion, but also affected BcsA insertion and/or stability in the membrane [58]. This observation stands in line with previous works where two-hybrid assay showed that BcsG interacts with its operon neighbors BcsF and BcsE, as well as with the synthase BcsA likely having structural and not only enzymatic roles in secretion system assembly and cellulose biogenesis [9], [30].

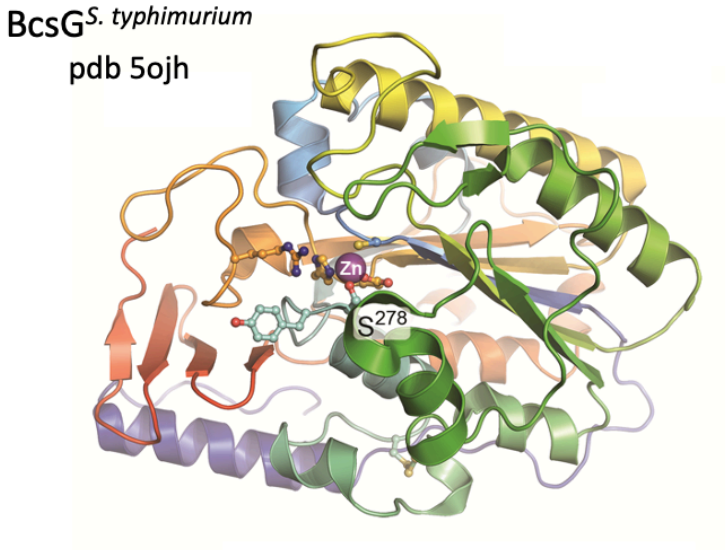


Figure 8. Crystal structure of the BcsG periplasmic domain

The overall fold of the catalytic module is shown in cartoon representation. The Zn^{++} coordination in the active with the key residues involved in cation coordination and/or essential for the pEtN modification are shown as sticks (Figure generated in PyMoL, Schrodinger LLC).

2.5.2.2 BcsZ

BcsZ is ubiquitous among Bcs systems and part of the conserved 3-4 component core, which also includes the catalytic BcsAB tandem and, in Gram-negative species, the OM pore BcsC [8]. This peculiar subunit has been shown to exhibit an endo-β-1,4-glucanase activity, meaning that it can hydrolyze the nascent polysaccharide. This feature of having a polysaccharide-degrading enzyme as part of the cognate secretion system is not unique to cellulose secretion, as other EPS secreting systems such as the PNAG, Pel and alginate machineries encode similarly an EPS-digesting enzyme within the same gene clusters as the other core components [54]. Similarly to the endocellulase BcsZ, PgaB and PelA exhibit both a hydrolase activity while AlgL displays a lyase function [54]. Moreover, this feature is also preserved in higher kingdoms, as higher plants include an endoglucanase named Korrigan (KOR) required for the correct assembly of the elongating cell walls [60].

Although this pattern is frequently observed in other systems, the biological relevance to have a glucanase around the nascent glucan chain has sparked many questions that remain to be answered. Hydrolases such as BcsZ in *S. typhimurium* has been shown to promote virulence, while *in vitro* experiments have showed that PslA and PelG disrupt the biofilm formed by *P. aeruginosa* [61], [62]. Furthermore, BcsZ could play a role in the packing of the cellulose microfibrils in *G. xylinus*. In fact, the deletion of *bcsZ* gene in *G. xylinus* resulted in irregular and twisted packing of the *de novo* synthesized fibrils suggesting that BcsZ could play a crucial role in cleaving the entanglements that can occur between secreted cellulose fibrils and thus BcsZ ensures the formation of a well-structured crystallized cellulose [63]. The latter observation opens the door to another question regarding amorphous cellulose where wild-type levels of BcsZ are also required for optimal cellulose secretion.

BcsZ structures from both *E. coli* and *G. xylinus* adopt a $(\alpha/\alpha)_6$-barrel fold forming a single domain, in which six pairs of antiparallel α -helices are found to form an inner and outer ring [64], [65] (Figure 9A-B). Structural analysis on the *E. coli* homolog revealed that the connecting loops extend on one side of the barrel forming antiparallel ß-sheets, as a result the structure adopts a deep “sickle shaped” groove on one side where the two catalytic residues D243 and E55 are embedded deep down the groove [65]. Both BcsZ structures from *E. coli* and *G. xylinus* fit completely in the GH-8 family of enzymes and adopt an overall similar architecture, with a slight difference regarding the substrate polysaccharide [64], [65]. BcsZ from *G. xylinus* needs

cellopentaose or longer polysaccharide for its activity, as opposingly to its homolog BcsZ from *E. coli*, which requires at least an hexasaccharide [65], [66]. Thus, a catalytically inactive *E.coli* BcsZ mutant solved in complex with cellopentaose is suggested to be in the post-hydrolysis state in which the protein contacts glucan chain at the non-reducing end as the reducing end has already left the catalytic center [65]. The common part between the BcsZ structures and other homologs from GH-8 family enzymes revealed that the interactions with the glucan chain are basically hydrogen bonds and stacking interactions with conserved aromatic residues. Nevertheless, all studies on enterobacterial BcsZ were carried out before the discovery of the pEtN modification, so substrate specificity for the modified cellulose remains to be characterized.

BcsZ is not limited to Gram-negative bacteria, as a recent study identified the CcsZ (*Clostridium cellulose* synthase Z) protein and functional BcsZ homologs have been described even in the atypical efflux-type cellulose secretion systems of cyanobacteria [43], [67]. It is true that there is no direct sequence homology between BcsZ and CcsZ and added to that CcsZ is a membrane tethered endo-β-1,4-glucanase that belongs to the GH-5 family instead of the GH-8 family of enzymes. Nevertheless, CcsZ has specificity for cellulosic materials and is believed to act in a similar manner in its cognate secretion systems [67]. The discoveries around BcsZ together point toward a putative role in limited hydrolytic action on the cellulose chains in the periplasm or at the cell surface thus preventing aggregative microfibrils formation before the exopolysaccharide reaches the extracellular environment. Interestingly, whereas in most studies BcsZ is positioned in the periplasmic compartment, other studies have suggested fully extracellular role for the subunit [64], [68]. Indeed, more studies are necessary to better characterize the fine functional balance exerted by BcsZ homologs as they can be suitable targets to prevent biofilm formation or, alternatively, play a role in biotechnological applications to increase cellulose production.

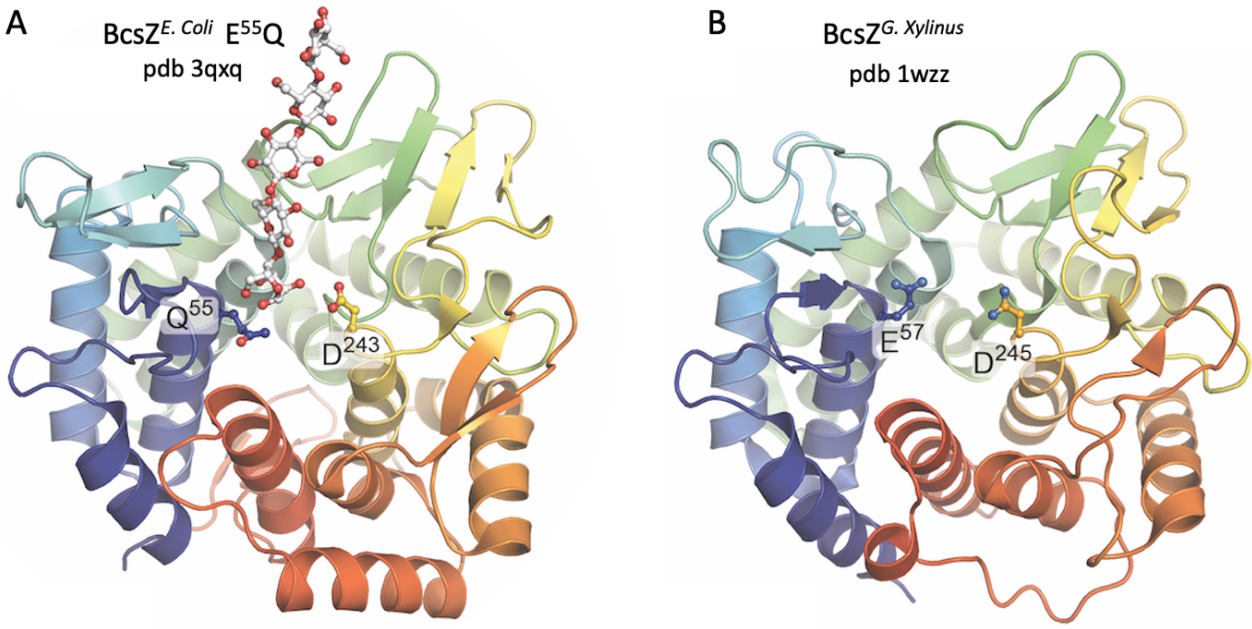


Figure 9. BcsZ homologs in *E. coli* and *G. xylinus*

A. Crystal structure of a cellopentaose-bound catalytically inactive mutant of E. coli BcsZ [65]. The oligosaccharide and catalytic dyad are shown as sticks. **B**. Crystal structure of the *G. xylinus* BcsZ homolog [64].

2.5.2.3 BcsQ and BcsR

The type II cellulose secretion system of *E. coli* is characterized by two additional genes *bcsR* (formerly *yhjR*) and *bcsQ* (formerly *yhjQ*) leading the operon that contains the four core secretory componens BcsA, BcsB, BcsZ and BcsC (Figure 5A) [8]. Whereas BcsR is a short ~7 kDa peptide without reliable structural models until our work, BcsQ belongs to the SIMIBI (Single-recognition particle, MinD and BioD) ATPases family conserved in both prokaryotes and Eukaryotes. Members of the latter family comprise MinD, responsible for the positioning the divisome in bacterial cells, FlhG/FlhF involved in bacterial flagellum assembly and SRP54-SR and Get3 NTPases responsible for membrane protein sorting [69], [70]. Regarding cellulose secretion, the closest homologs to BcsQ are WssA/WssJ proteins that are part of the acetylated cellulose secretion machinery in *P. fluorescens* SBW25 [14], [71]. BcsR and BcsQ share predicted cytosolic localization and partake in interactions with several Bcs components [9]. Interestingly, BcsQ localizes at the cell pole, which corresponds to the site of cellulose secretion, as well as initial cell-cell-adhesion in *E. coli* [14]. Remarkably, both BcsR and BcsQ are essential for cellulose secretion in *E. coli* but are absent in many Type I and Type III cellulose secretion systems suggesting different regulatory inputs for the enterobacterial Bcs

machinery [9]. These secretion system-specific effects and the essentiality of the two subunits for pEtN cellulose biogenesis motivated us to decipher the roles of both BcsR and BcsQ and many mechanistic insights regarding the structures and functions of both components can be found in the Results chapter.

2.5.2.4 BcsE, the c-di-GMP sensor

Before the identification of BcsE as a c-di-GMP receptor in the Bcs system, the PilZ domain-containing BcsA was the first described to require the interaction with c-di-GMP to perform its glycosyltransferase activity [10], [72], [73]. Later BcsE joined the group of c-di-GMP receptors upon the characterization of a conserved C-terminal domain containing a c-di-GMP-binging RxxD motif, which – when found on c-di-GMP-producing GGDEF domain-containing diguanylate cyclases – can serve as an autoinhibitory, or I-, site for allosteric regulation. Conversely, BcsE was dubbed as a GGDEF I-site Like, or GIL-, domain protein by Fang and coworkers [15]. Although BcsE is not absolutely essential for cellulose secretion it plays a crucial role in maintaining the stability of the Bcs complex and contributes to optimal polysaccharide production [9], [15]. The *bcsE* gene leads the operon *bcsEFG*, which is hallmark for *E. coli*-like Type II cellulose secretion systems. The pEtN-cellulose modifications by BcsG on one hand, and the proximity of the *bcsEFG* genes to the *bcsABZC* cluster on the other, suggests that this second Bcs cluster not simply contributes to secretion and chemical modifications of the cellulose but rather plays key roles in the macroscopic architecture of the mature bacterial biofilms [9], [30].

2.5.2.5 BcsF

BcsF is a small subunit anchored to the inner membrane by a single transmembrane helix. The *bcsF* gene is not common to all proteobacteria, and where annotated is usually a downstream operon neighbor of the *bcsE* gene (Figure 5A) [8]. It has been shown that BcsF, interacts with BcsA, BcsE and BcsG *in vivo,* however the mechnisms by which the interactions occur and the exact role of BcsF in cellulose secretion, needed to be decrypted. Here we examined further the role of this subunit and demonstrate that it is key to Bcs macrocomplex assembly integrity via membrane recruitment of the regulatory BcsRQE components (see Results section) [9], [30].

2.6 C-di-GMP, a versatile second messenger

Cyclic diguanylate or c-di-GMP is a monocyclic RNA dinucleotide second messenger that is almost ubiquitous in the prokaryotic world and has also been observed in eukaryotic social amoeba from the *Dictyostelium* clade [74]. The discovery of c-di-GMP goes back to almost half a century now, when Benziman and coworkers first described that the *in vitro* cellulose synthase activity of *G. xylinus* was promoted by GTP and an unknown GTP-converting protein factor [75]. The latter hypothesis was supported in a successive study, where digitonin solubilization showed that the GTP-converting protein associates with the membrane-bound enzyme and is promoted by Ca^{2+} ions [76]. Shortly after, the GTP-converting enzyme was isolated using agarose-conjugated GTP as an affinity matrix and its synthase-activating product identified as *bis*(3′ ,5′)-cyclic diguanylic acid, or c-di-GMP [72]. Almost two decades later with the revolution of DNA (Deoxyribonucleic acid) sequencing methods and ever-growing data on genome assemblies, c-di-GMP has revealed itself as a universal second messenger in the center of a plethora of mechanisms in the bacterial world [22], [54], [77]. Often through direct control, c-di-GMP orchestrates many biological processes including motility, virulence gene expression and secretion of extracellular polysaccharides [22], [54], [77]. Many studies thoroughly reviewed the different roles of this small molecule and highlighted the layers of physiological responses controlled by c-di-GMP covering the transcriptional, translational and posttranslational levels [21], [22], [54], [77]. Moreover, in order to adapt to the large spectrum of mechanisms, c-di-GMP is able to adopt variable stable conformations when bound to proteins which makes it difficult to predict its new targets [23] (Figure 10). In this part of the present study, I will discuss targets and regulation of the near-ubiquitous second-messenger c-di-GMP. I also contribute to the understanding of c-di-GMP signaling in relation with cellulose secretion by the discovery of new dinucleotide-sensing mechanisms specific to the *E. coli*-like type II cellulose secretions in the Results chapter.

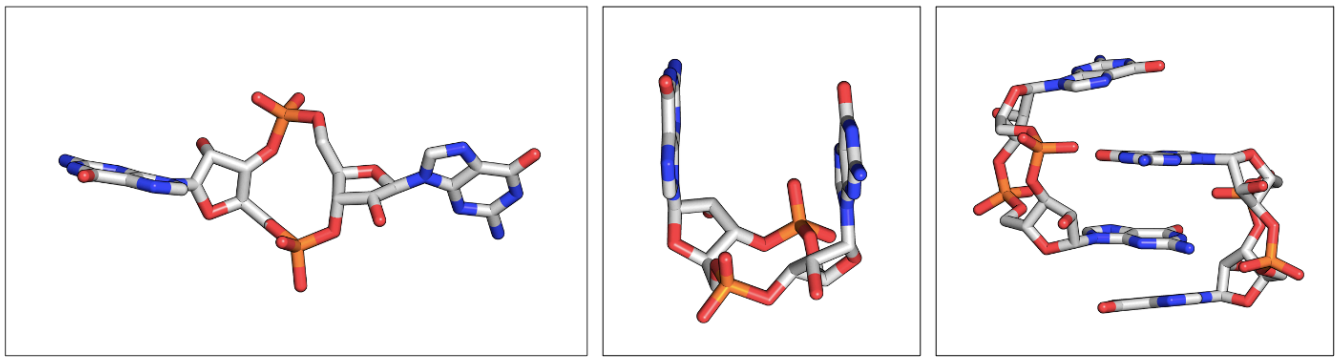


Figure 10. Examples highlighting the conformational adaptability of c-di-GMP

Based on the structural studies, c-di-GMP can adopt a variety of conformations when bound to proteins in order to adapt and confer specificity to the different targets.

2.6.1 Synthesis and degradation of c-di-GMP

C-di-GMP is synthesized from 2 molecules of GTP by diguanylate cyclases (DGCs) and degraded by phosphodiesterases (PDEs) (Figure 11). Multiple-sequence analyses across bacterial genomes have led to the characterization of the domains that confer the activity of the c-di-GMP-metabolizing enzymes, either the diguanylate cyclase activity or the phosphodiesterase. DGCs contain GGDEF domains, while PDEs are characterized by EAL or HD-GYP domains, where the domain names come from conserved amino acid motifs [18]–[20]. Interestingly, many proteins feature both GGDEF and EAL domains, for that reason it is difficult to assign a specific function to such hybrid proteins from domain architecture and sequence conservation alone. The first structurally characterized protein featuring a GGDEF domain, PleD$^{C.\ crescentus}$, not only allowed for the experimental confirmation of the diguanylate cyclase function, but also revealed an allosteric binding site, called I-site (from Inhibitory site) which is represented by an RxxD motif (i.e. an arginine and an aspartate spaced by two amino acids) and is involved in a negative feedback control [18]. The degradation of c-di-GMP, on the other hand, can be achieved in two steps, with first phosphodiester cleavage leading to a linear pGpG product and a subsequent hydrolytic event producing 2 moieties of GMP. While HD-GYP domains typically break the c-di-GMP fully to GMP, some EAL domains can lead to the accumulation of pGpG (Figure 11). In Gram-negative bacteria, the HD-GYP proteins, are less common, whereas GGDEF and/or EAL domain-containing proteins are widespread among Gram-negative bacteria and a bit less in Gram-positive bacteria. Many bacteria feature multiple

proteins containing c-di-GMP-metabolizing domains and oftentimes these modules can be 'degenerate', or catalytically inactive, yet important in c-di-GMP-mediated signal transduction. For example, the *E. coli* K-12 genome encodes 29 proteins containing both GGDEF and/or EAL domains, and 4 of those exhibit degenerate domains [78]. Among the studies of ever-increasing number of GGDEF/EAL proteins, different nomenclatures have been used, and a recent consensus nomenclature was proposed to name genes encoding catalytically active GGDEF and EAL domains as *dgc* and *pde*, respectively [78]. Surprisingly, c-di-GMP-metabolizing proteins are generally not redundant and despite the presence of multiple such actors in the cell, their function appears to be restricted to specific signaling pathways and physiological effects. For example, among the identified DGCs in cellulose-secreting *E. coli* K12 strains mostly DgcC is responsible for activating pEtN cellulose secretion, whereas in the *E. coli* 1094 strain studied in this work this role is taken by DgcQ enzyme [79], [80]. Bacterial two-hybrid assays have showed that cellulose-promoting DgcC interacts with the phosphodiesterase PdeK, and both enzymes interact with the co-catalytic partner BcsB [79]. To coordinate cellulose biogenesis, a model is thus proposed where interactions between the co-polymerase BcsB, the c-di-GMP "source" DgcC and the c-di-GMP "sink" PdeK secure physical proximity to fine-tune cellulose biogenesis [79]. The idea of having a source and sink in physical proximity is consistent with a mathematical model from the same study where c-di-GMP molecules produced locally by the DgcC enzyme can dramatically increase the probability of c-di-GMP capture by the Bcs secretion machinery**,** and thus lead to cellulose secretion activation *in vivo*, where – as we show in the Results section – activating dinucleotide retention will be further secured by the multi-site c-di-GMP complexation by the cytosolic BcsERQ vestibule-forming complex and the PilZ module of the synthase BcsA [79], [81]. Finally, the co-localization of PdeK with the secretion machinery can be seen as a security valve that controls the flow of the intracellular c-di-GMP and consequently adapts the cellulose biogenesis to the cadence of the external stimuli and cellular metabolism.

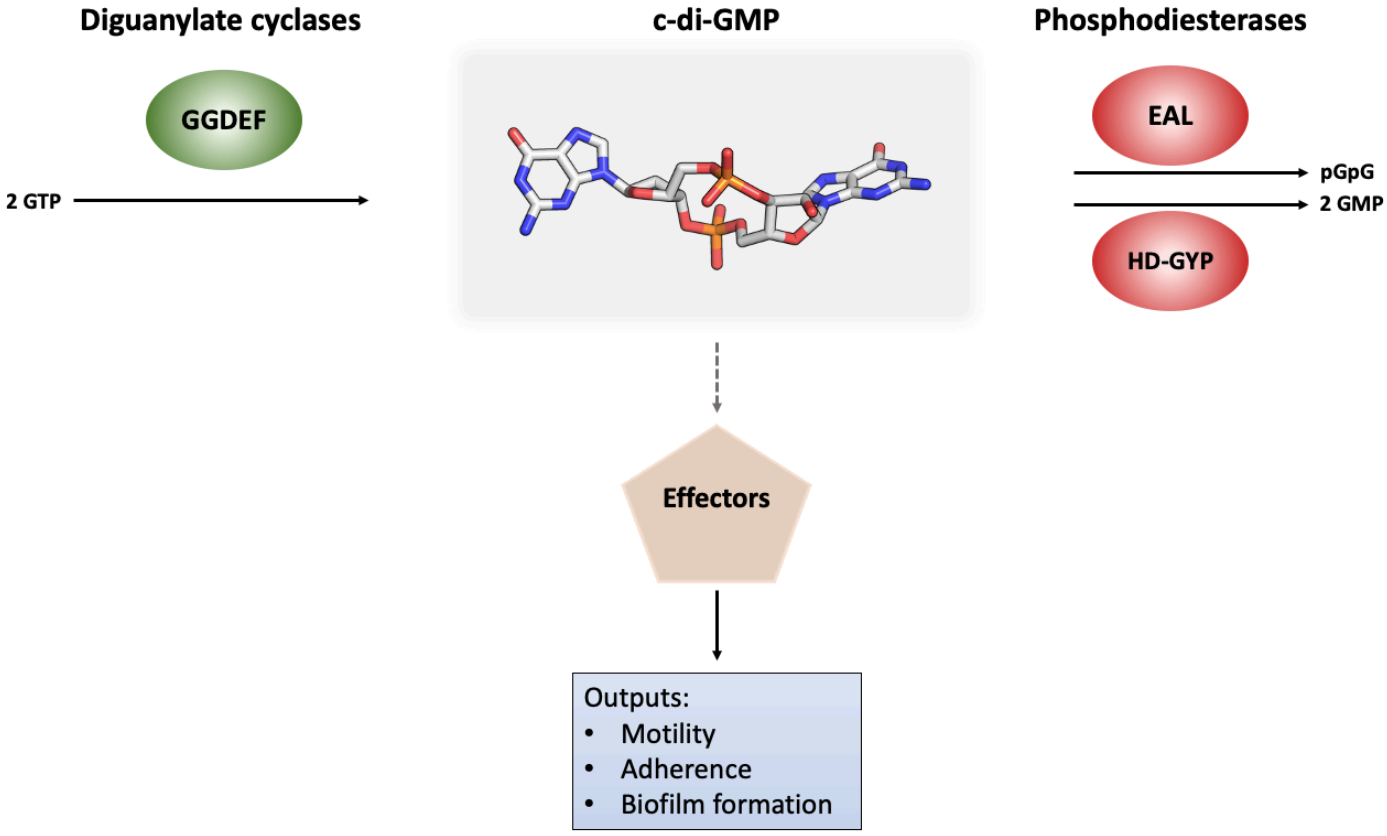


Figure 11. Biochemistry of c-di-GMP

The three-dimensional structure of c-di-GMP is formed by two GMP moieties that are linked via 5'-3' macrocyclic ring. The synthesis of c-di-GMP is catalyzed by diguanylate cyclases containing GGDEF domains. c-di-GMP is degraded by proteins containing either EAL or HD-GYP domains. yielding pGpG or 2GMP molecules. The three-dimensional structure of c-di-GMP is shown in the middle of the reaction in stick representation, carbon atoms are shown in gray, oxygen in red, nitrogen in blue and phosphorus in orange (from protein data bank entry 3hv8). Upon binding to effector molecules, c-di-GMP regulates diverse cellular processes generally leading to loss of motility, cell adhesion and persistence of biofilm forming communities [17]–[20].

2.6.2 Multileveled signaling of c-di-GMP in Bcs system

In order to orchestrate cellulose biogenesis in *E. coli*, c-di-GMP typically employs multilevel signal transduction mechanisms, which culminate with BcsA activation upon the binding of the c-di-GMP to the PilZ domain [10] (more details are in the section below). In pEtN-cellulose secreting *E. coli* 1094 and *Salmonella*, the signaling cascade starts with expression of the stationary phase sigma factor RpoS, or σ^S, in response to external stimuli [80] (Figure 12). Under physiological conditions, the actual c-di-GMP levels in the cytosol are quite low, mainly due to efficient degradation of the dinucleotide by the phosphodiesterase PdeH [82]. Stationary phase RpoS leads to the expression of a specific diguanylate cyclase DgcE, which counteract

the effects of the phosphodiesterase PdeH and releases the diguanylate cyclase DgcM from inhibitory interactions with the phosphodiesterase PdeR. DgcM thus can engage in an interaction with the MlrA transcription activator and activate the transcription of CsgD [82]–[84]. CsgD in in turn activates the expression of the rest of the *csg* genes involved in curli secretion and the diguanylate cyclase DgcC which ultimately produces the cellulose synthase-activating c-di-GMP [85], [86].

Adding to the complexity of c-di-GMP signaling, studies on a collection of *E. coli* isolates showed that CsgD-/DgcC-dependent cellulose secretion is not the only mechanism for c-di-GMP control of the BcsA cellulose synthase [80]. For example, in the commensal *E. coli* 1094 strain used as a model organism in this work, cellulose secretion is specifically activated by a different diguanylate cyclase – the enzyme DgcQ – and appears to circumvent the CsgD-dependent signaling cascade [80] (Figure 12). Both signaling pathways, CsgD-/DgcC-dependent or independent, share in common the RpoS-dependent initiation, however it is unclear what the signaling mediators are for the downstream DgcQ activation [80].

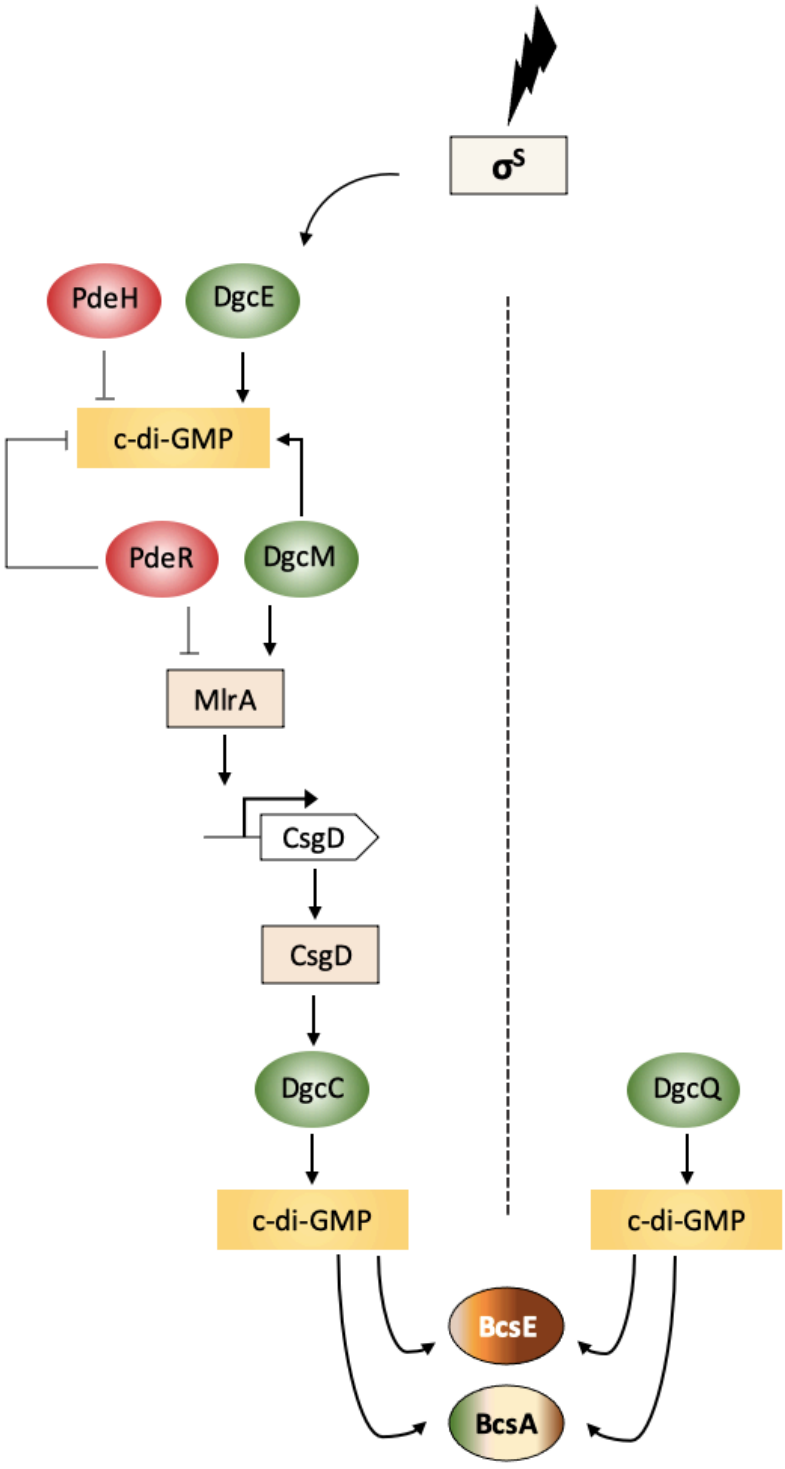


Figure 12. Cellulose biosynthesis regulation in *E. coli*

The figure illustrates the two pathways regulating the pEtN-cellulose production in enterobacteria. On the left is the CsgD-/DgcC-dependent pathway, on the right is the CsgD-independent/DgcQ-dependent pathway. In both cases, the production of c-di-GMP is required for the cellulose biosynthesis. In particular, cellulose polymerization and extrusion requires the binding of the dinucleotide to the PilZ module of the synthase BcsA, which in turn is facilitated by the multi-site c-di-GMP complexation of the cytosolic regulatory complex BcsERQ (see Results section) [81], [87]. Diguanylate cyclases (DGC) are shown in green, phosphodiesterases (PDE) are shown in pink [17], [80].

2.6.3 C-di-GMP and the cellulose synthase BcsA

In most Bcs systems, cellulose secretion is initiated upon the activation of the synthase BcsA by the second messenger c-di-GMP. BcsA harbors the three modules, namely the transmembrane export domain, the family-2 glycosyltransferase responsible for the catalytic reaction, and the well-studied c-di-GMP-binding C-terminal PilZ domain [10]. Different crystallographic snapshots of the BcsA$^{R.\ sphaeroides}$ with its co-catalytic partner BcsB, have been studied in great details. Comparisons between the different states reveal that cyclic dinucleotide binding induces several important conformational changes. First, in the c-di-GMP bound state, the PilZ ß-barrel rotates by approximatively 20° around a so-called 'hinge' helix, which is sandwiched at the interface between the β-barrel and the glycosyltransferase domain [10], [12]. An additional conformational change that characterizes the c-di-GMP-bound state involves Arg580 situated from the conserved R580xxxR motif that precedes the PilZ β-barrel and partakes in c-di-GMP complexation [10]. In the c-di-GMP-free state the synthase is in an auto-inhibitory state, where the so-called 'gating loop' caps the active site entry in a resting-state conformation stabilized by a salt bridge interaction between the Arg580 and an equally conserved Glu371 from the gating loop [11]. Upon c-di-GMP binding, Arg580 rotates by approximatively 180° and is involved in a new interaction with the dinucleotide, which frees the gating loop and creates a large 22.5 x 12.4 Å opening to the active site, wide enough to allow for entry of the substrate UDP-glucose. Further increase in c-di-GMP concentrations do not increase the affinity of BcsA for UDP-glucose *in vitro* [48] consistent with c-di-GMP-based control of active site entry as evidenced by the crystal structures.

Nevertheless, control of active site accessibility is not the only function of the BcsA's gating loop. Some of the resolved c-di-GMP-bound crystal structures contain either a homolog of the substrate UDP-glucose, or the product UDP. In fact, in both latter states, the gating loop inserts deeply into the catalytic pocket, where many of the loop's residues and especially the conserved FxVTxK motif coordinate the uracil base and the pyrophosphate of the UDP moiety [11] (Figure 13). Moreover, a crystallographic snapshot of a c-di-GMP-bound state in the pre-translocation conformation features a largely disordered gating loop with an overall trajectory reminding that of the UDP-bound but c-di-GMP-free resting state [10], [12]. Therefore, it is conceivable that both the release of c-di-GMP or of the product UDP can initiate gating loop relaxation. Furthermore, different studies have reported markedly different binding affinities between BcsA and c-di-GMP, more consistently in the micromolar range, which in a

physiological context would suggest a more dynamic sampling rather than strong complexation [48], [88], [89]. One may thus think that the c-di-GMP binding to the PilZ domain is highly dynamic, and that release of c-di-GMP in-between catalytic cycles can be even required to allow for efficient gating loop relaxation, product release and polysaccharide translocation, before c-di-GMP re-complexation causing the initiation of a new cycle. Substrate entry and insertion of the gating loop would allow the initiation of the catalysis, and upon the glycosyl transfer the gating loop would relax, retracting from the catalytic pocket to allow for the recycling of UDP to UDP-glucose in a mechanism depending simultaneously on the local concentrations of c-di-GMP, UDP-glucose, and UDP (Figure 13). As we show in this work, such dynamic process of activating dinucleotide recycling is likely greatly facilitated in the *E. coli*-like Bcs system by the formation of a synthase-proximal c-di-GMP pool by multi-site complexation to the regulator BcsERQ vestibule surrounding the PilZ module of the synthase (see Results section).

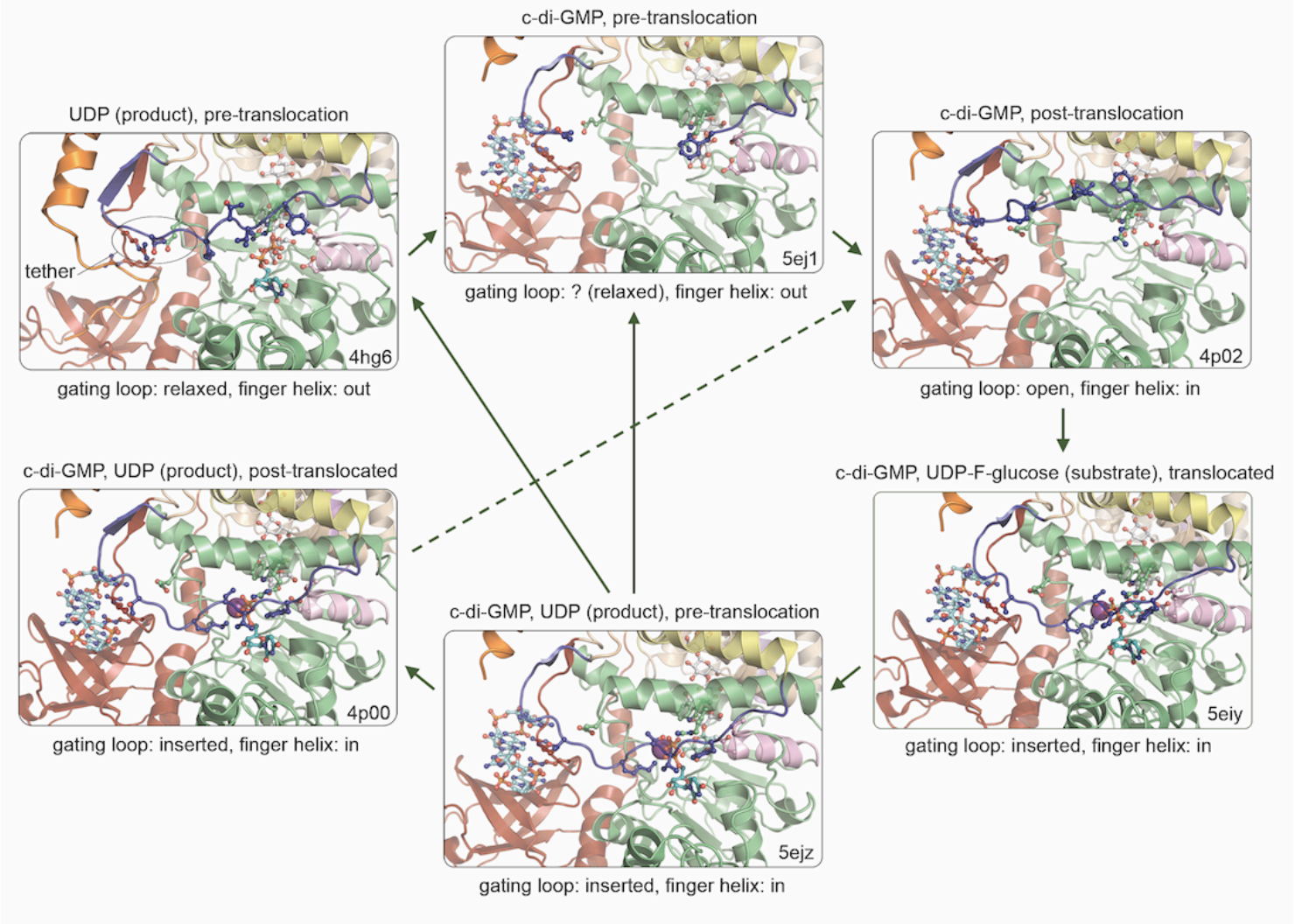


Figure 13. Zoom-in of the active site of synthase BcsA

The different catalytic states are captured in the crystal structures of the *R. sphaeroides* BcsAB tandem from crystals grown or soaked with different ligands. Protein data bank accession numbers are indicated in each panel; presence of substrate homologs, products, c-di-GMP, translocation state of the polymer and gating loop conformation are indicated for each state [10]–[12]. Green arrows indicate putative conformation transition pathways integrating all captured conformational states. Alternative pathways would depend on local variations of c-di-GMP, product and substrate molecules, as well as finger helix-mediated or spontaneous translocation. The cellulosic polymer, substrate homolog, UDP product, c-di-GMP and key residues from the gating loop, c-di-GMP coordinating PilZ-proximal linker, finger helix and conserved QRGRW motif are shown as sticks.

2.6.4 Cellulose synthase core within kingdoms

Plants are more famous for the processive synthesis of cellulose as a primary building material of their cell walls, thus making cellulose the most abundant biopolymer on the planet Earth. However, the identification of plant cellulose synthases relied on gene homology to their bacterial *bcsA* homologs. Later, it became evident that the genes for cellulose synthesis in plants have likely evolved from cyanobacteria via lateral gene transfers during multiple ancient endosymbiotic events [26], [27]. Plants build cellulose through CesA variants, which belong to the glycosyl transferase 2 family like their bacterial counterparts BcsA [26]. Globally, both cellulose synthases catalyze the glucose transfer from the UDP-glucose to the C4 hydroxyl of the nascent cellulose polymer through a process coupling the polymerization of the cellulose polymer and its extrusion through the plasma membrane (Figure 3A). Nevertheless, in plants the synthase complexes enrolled in polysaccharide secretion have been typically shown to form hexameric macroassemblies called ‘rosettes’ or CSCs (cellulose synthase complexes). Recent studies revealed that the rosettes likely accommodate six CesA trimers, or 18 CesA protomers, that play a major role in securing the packing of the secreted glucan chains into fundamental microfibrils composed of 18 glucan chains in a mechanism similar to the crystalline cellulose ribbon secretion in *G. xylinus* [90] (Figure 5A). A recent study using *in situ* cryo-electron tomography, revealed another similarity between cellulose synthesis in crystalline cellulose-secreting bacteria and plants by identifying a novel cytoplasmic structure in *G. xylinus* and *G. hansenii* dubbed “the cortical belt”, which spatially correlates with the extracellular cellulose ribbon and is proposed to tether the synthase terminal complexes in a manner similar to cortical microtubule tethering of plant CSCs [37] (Figure 5B-C).

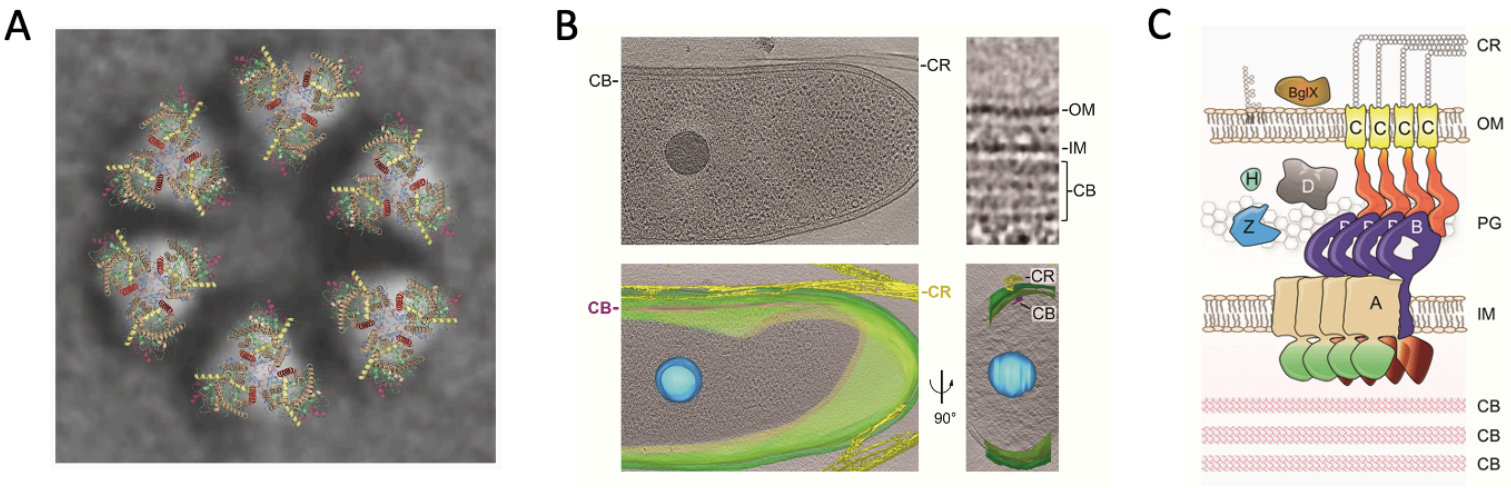


Figure 14. The similarity between cellulose synthesis in plants and bacteria

A. Proposed structure of a CesA rosette complexes based on [90], where multiple CesA protomers will secure crystalline microfibril bundling upon exit of the nascent cellulose strains from the cell reproduced under the CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/legalcode) **B**. Cryo-electron tomography visualization of the cellulose ribbon and the underlying cortical belt proposed to organize the linear assembly of Bcs terminal complexes for crystalline cellulose secretion in the *Komagataeibacter* lineage. Data from [37] CB, cortical belt; IM, inner membrane; PG, peptidoglycan; OM, outer membrane; CR, crystalline cellulose ribbon. **C**. Schematic representation of the *G. xylinus* / *G. hansenii* type I cellulose secretion system.

2.7 Applications of bacterial cellulose

Since its discovery, bacterial cellulose is increasingly stealing the spotlight from plant cellulose, especially the crystalline product secreted by bacteria from the *Komagataeibacter* lineage. Although both plant and bacterial celluloses share similarities in their overall structure, the bacterial nanocellulose is much purer due to the lack of co-produced hemicelluloses and lignin. Purity is only one of the many outstanding properties of bacterial cellulose, such as its excellent water retention capacity, high crystallinity, transparency, biodegradability and low antigenicity. Moreover, nanocellulose exhibits tremendous thermal and chemical stability allowing efficient non-denaturing sterilization, as well as valuable mechanical properties, such as tensile strength and elasticity. The almost endless list of desirable properties from a materials science perspective, makes bacterial cellulose an iconic biopolymer in the center of many applications in diverse fields such as the food, cosmetics and medical industry (See table 2 for examples of applications). To date, the major limiting factor for use of bacterial cellulose at the industrial scale, is perceptibly its cost. To tackle this economic problem, we need to decode and optimize the molecular mechanisms governing cellulose biogenesis by the Bcs nanomachineries in nature.

Table 2. Examples of bacterial cellulose applications

Field	Application
Food	Food thickening [91] Emulsification [92] Water-binding [92] Immobilization of probiotics [93] Active food packaging [94]
Paper	Higher quality papers [95] Flame retardation material [96]
Electronics	Flexible displays [97] Electronic paper displays [98] Organic light emitting diodes [99]
Cosmetics	Texturing agent [100] Emulsion stabilizer [101] Carrier of skin active substances [101]
Medicine	Suture biomaterials [102] Wound dressing [103], [104] Drug delivery [105] Tissue engineering [106] Artificial cornea [107] Retinal pigment epithelium [108] Blood vessels [109]

3 Objectives of the thesis

This PhD aims to provide mechanistic insights into the cellulose secretion machinery involved in bacterial biofilm formation. We know now that bacterial cellulose can be produced by three major types of secretion systems sharing as a common feature the catalytic core BcsAB [8]. This conserved assembly formed by the accessory BcsB co-polymerase and the glycosyl transferase, c-di-GMP-sensing PilZ and inner membrane export module of BcsA has been studied in great details [2–4]. Added to the catalytic subunits BcsAB, additional Bcs components as well as the chemical properties of the secreted exopolysaccharide delineate the separate types of cellulose secretion [8], [13]. Several works spot the light on some of these additional subunits, such as the structural characterization of the OM membrane porin domain of BcsC and its role in the delivery of the polysaccharide to the extracellular compartment [34]; the structural characterization of the BcsG subunit and its role in the modification of the polysaccharide by the addition of pEtN groups [30], [58]; or the ever-increasing evidence of the crucial roles of c-di-GMP in the regulation of cellulose secretion and biofilm formation [22], [54], [77]. What is missing in the big picture, are the structures and functions of the multiple additional Bcs subunits, some of which are indispensable for cellulose secretion, as well as the mechanism of assembly and signaling within the system in its integrity.

The *E. coli*-like cellulose secretion system is encoded by at least two operons, one of which encodes a second c-di-GMP sensing protein, BcsE, which is important for maximal cellulose secretion, yet little more was known about this regulator subunit [8]. In addition, this operon encodes two more proteins whose disruption drastically affects cellulose production [9] the BcsG subunit, whose role as a pEtN transferases was only recently uncovered, as well as a short transmembrane peptide of unknown mechanism of function, BcsF. On the other side, the second operon starts with genes encoding the essential for cellulose secretion BcsR and BcsQ, followed by those for the core components BcsABZC.

A study by Krasteva and colleagues has demonstrated recently that in *E. coli* most of the inner membrane and cytosolic Bcs subunits assemble to form a stable, megadalton-sized secretory nanomachine with unique asymmetric architecture (BcsRQABEF or Bcs macrocomplex here-in) [9]. Additionally, the latter study unveiled the functions of many of the accessory subunits, and in particular the *in vivo* contributions of each Bcs component to cellulose secretion, as well as protein partners that partake in binary protein-protein interactions [9]. Yet, the structural reconstruction of the Bcs macrocomplex, based on low resolution electron microscopy data on a sample embedded in negative-stain, did not allow for further mechanistic insights. In this

context, the goal of the thesis was to study the structure of key components and multiprotein subcomplexes of the *E. coli* Bcs secretion system in higher resolution. To this aim, we resorted to cryo-EM to study the structure of the assembled Bcs macrocomplex, studies performed primarily by my colleague Samira Zouhir. Personally, I focused on studies started by a short-term intern in the group, Meryem Caleechurn, aimed at deciphering the structure and function of the BcsRQ and BcsERQ regulatory complexes by X-ray crystallography. My work involved significant protein engineering efforts, screening and optimization of expression construct, purification protocols and crystallization conditions, and, following successful crystallization and structure solving, the design and execution of multidisciplinary functional studies in validating structure-derived hypotheses on the subunits' function. On the other hand, I also provided key evidence on the self-polymerization mechanisms driving Bcs macrocomplex assembly in the inner membrane by solving the cryo-EM structure of isolated BcsB multimers, consistent with the assymetric Bcs macrocomplex assembly and non-canonical BcsA:BcsB stoichiometry.

4 Results

4.1 Article 1

4.1.1 Introduction to article 1

In many gram-negative bacteria, biofilm formation goes in hand with cellulose secretion. As discussed above, the latter usually requires the co-expression of a c-di-GMP-activatable inner membrane BcsAB tandem, an outer membrane porin BcsC and multiple accessory subunits which allow for the classification of Bcs systems [8], [47]. The Type II *E. coli*-like cellulose secretion system is distinguished by two separate operons, namely *bcsEFG* and *bcsRQABZC*, where *bcsEFG* operon is the hallmark of this type of secretion system [8]. Previous studies demonstrated that BcsR and BcsQ are essential for cellulose secretion, while BcsE is required for maximal cellulose production and has been shown to also bind cyclic-di-GMP [14]. Interestingly, the latter cytosolic subunits, together with the additional partner BcsF were suggested to interact in different combinations as evidenced by cell-based bacterial two-hybrid complementation assays [9]. The aim of the study described in the article 1 was to further explore the interactions between the different cytosolic subunits, identify stable protein-protein interactions, design stable protein expression constructs, overexpress and purify the target proteins and multisubunit assemblies and study the structural and functional determinants of these subunits both *in vitro* and *in cellulo*.

To this end, we performed multiple recombinant (co-)expression/(co-)purification experiments on the BcsR, BcsQ and BcsE components, alone and in different combitions. We established point mutants and truncated variants that were instrumental for the stabilization of the protein complexes or subunits and their subsequent crystallization and structure determination. We reported the structure of the c-di-GMP-binding module of BcsE and showed that the previously postulated GIL domain is in fact a tandem of degenerate receiver (REC*) and diguanylate cyclase (GDDEF*) domains that are incompetent for phosphotransfer or c-di-GMP biogenesis, respectively. We further showed that BcsE senses both BcsQ and c-di-GMP through non-overlapping interfaces on the degenerate GGDEF* module and that the remaining N-terminal domain is likely a degenerated P-loop ATPase module, which is responsible for membrane targeting of the BcsERQ regulatory complex via interactions with inner-membrane BcsF. Furthermore, the expression and purification of BcsE and this N-terminal module, led to the identification of a new high-affinity binding partner, which was identified with the help of mass-spectrometry analysis on as the S10 or NusE protein, a key component of the conserved cellular transcription antitermination complex (TAC) [110]. We validated the latter discovery

by *in cellulo* two-hybrid functional complementation assays and protein co-purification experiments and revealed that BcsE competes with a second TAC component, NusB, for S10 complexation. Although the exact role of these interactions with the TAC machinery remain to be further investigated, the *in silico* detection of putative intrinsic terminators in the *bcsR* and *bcsQ* loci leading the second Bcs operon, led us to speculate that BcsE might not only contribute to c-di-GMP complexation and recruitment of the essential BcsRQ complex to the inner membrane but also be potentially targeted to the site of BcsRQ expression via the TAC machinery for efficient co-translational BcsERQ complex assembly.

AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE
Molecular Biology and Physiology

Structure and Multitasking of the c-di-GMP-Sensing Cellulose Secretion Regulator BcsE

Samira Zouhir,[a,b,c] Wiem Abidi,[a,b,c] Meryem Caleechurn,[a] Petya Violinova Krasteva[a,b,c]

[a]Structural Biology of Biofilms Group, Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Paris-Sud University, Gif-sur-Yvette, France
[b]Structural Biology of Biofilms Group, European Institute of Chemistry and Biology (IECB), Pessac, France
[c]CBMN UMR 5248 CNRS, University of Bordeaux, Pessac, France

Samira Zouhir and Wiem Abidi contributed equally to this work. Author order was determined alphabetically on first-name basis.

ABSTRACT Most bacteria respond to surfaces by biogenesis of intracellular c-di-GMP, which inhibits motility and induces secretion of biofilm-promoting adherence factors. Bacterial cellulose is a widespread biofilm component whose secretion in Gram-negative species requires an inner membrane, c-di-GMP-dependent synthase tandem (BcsAB), an outer membrane porin (BcsC), and various accessory subunits that regulate synthase assembly and function as well as the exopolysaccharide's chemical composition and mechanical properties. We recently showed that in *Escherichia coli*, most Bcs proteins form a megadalton-sized secretory nanomachine, but the role and structure of individual regulatory components remained enigmatic. Here, we demonstrate that essential-for-secretion BcsR and BcsQ regulate each other's folding and stability and are recruited to the inner membrane via c-di-GMP-sensing BcsE and its intraoperon partner BcsF. Crystallographic and solution-based data show that BcsE's predicted GIL domain is a degenerate receiver-GGDEF domain tandem ($BcsE^{REC*-GGDEF*}$), where the divergent diguanylate cyclase module binds both dimeric c-di-GMP and BcsQ through mutually independent interfaces. In addition, we reveal that a third N-terminal domain ($BcsE^{NTD}$) determines the protein's homooligomerization and targeting of BcsERQ to the membrane as well as previously unreported interactions with transcription antitermination complex components. Together, the data suggest that BcsE acts on multiple levels to fine-tune bacterial cellulose secretion, from the early stages of secretion system assembly to the maintenance of a membrane-proximal pool of dimeric c-di-GMP for processive synthase activation.

IMPORTANCE Bacterial cellulose is a widespread biofilm component that can modulate microbial fitness and virulence both in the environment and infected hosts. Whereas its secretion generally involves an inner membrane c-di-GMP-dependent synthase tandem (BcsAB) across the bacterial domain of life, enterobacteria feature sophisticated *Escherichia coli*-like Bcs secretion systems, where multiple additional subunits are either required for secretion or contribute to the maximal production of the polysaccharide *in vivo*. Here, we demonstrate that essential-for-secretion BcsR and BcsQ regulate each other's folding and stability and are recruited to the inner membrane via c-di-GMP-sensing BcsE and its intraoperon partner, BcsF. Crystallographic and functional data reveal that BcsE features unexpected domain architecture and likely acts on multiple levels to fine-tune bacterial cellulose production, from the early stages of secretion system assembly to the maintenance of a membrane-proximal pool of dimeric c-di-GMP for processive synthase activation.

KEYWORDS biofilm formation, c-di-GMP signaling, cellulose secretion, structural biology

Citation Zouhir S, Abidi W, Caleechurn M, Krasteva PV. 2020. Structure and multitasking of the c-di-GMP-sensing cellulose secretion regulator BcsE. mBio 11:e01303-20. https://doi.org/10.1128/mBio.01303-20.

Editor Caroline S. Harwood, University of Washington

Address correspondence to Petya Violinova Krasteva, pv.krasteva@iecb.u-bordeaux.fr.

Received 15 May 2020
Accepted 9 July 2020
Published 11 August 2020

Bacterial biofilm formation is a ubiquitous adaptational strategy that provides fitness and resistance advantages to both free-living and clinically important species (1). In most motile bacteria, the switch from planktonic to biofilm life styles is orchestrated by an intracellular second messenger, c-di-GMP, that acts at the transcriptional, translational and posttranslational levels to inhibit flagellar motility and induce the secretion of extracellular matrix components (2, 3). Bacterial cellulose is a widespread biofilm exopolysaccharide that typically requires an inner membrane, c-di-GMP-dependent synthase tandem for glucose polymerization and inner membrane transport (BcsAB), and in Gram-negative species, an outer membrane porin with peptidoglycan-binding scaffolding motifs (BcsC) (4). Depending on the type of core and accessory subunits, four major types of cellulose secretion systems are generally recognized among bacteria (5). Many *Betaproteobacteria* and *Gammaproteobacteria* feature sophisticated *Escherichia coli*-like systems for cellulose biogenesis, where multiple additional subunits are either essential for secretion or contribute to the maximal production of the polysaccharide *in vivo* (5, 6).

In particular, the *E. coli bcsEFG* and *bcsRQABZC* operons encode a total of nine subunits that span from the cytosol to the surface of the cell (5, 6) (Fig. 1). The processive glucose polymerization reaction is carried out by a glycosyl transferase domain on BcsA ($BcsA^{GT}$), whose active site is made accessible by the recurrent binding of dimeric intercalated c-di-GMP to an adjacent PilZ β-barrel domain on the protein ($BcsA^{PilZ}$) and the displacement of a so-called gating loop capping the substrate-binding pocket (7, 8). Transport is coupled to polymerization, and the nascent polysaccharide chain is extruded, one molecule at a time, through the inner membrane transport domain of BcsA completed by the C-terminal tail-anchor of the cocatalytic subunit BcsB ($BcsB^{TA}$) (4, 7). We showed earlier that in *E. coli*, most of the inner membrane and cytosolic Bcs components interact stably to form a megadalton-sized Bcs macrocomplex with a seashell-like, layered, and asymmetric architecture (Fig. 1) (6). In it, multiple copies of BcsB arrange in a fan-like assembly, or "crown," in the periplasm, which is proposed to lead the outcoming cellulose toward the outer membrane secretory component BcsC (6). En route, the synthesized cellulose can undergo enzymatic modifications through the addition of phosphoethanolamine residues by BcsG or limited hydrolysis by BcsZ (4, 5, 9).

Interestingly, *E. coli*-like cellulose secretion *in vivo* is absolutely dependent on the expression of two small cytosolic proteins, BcsR and BcsQ (6), whose genes precede those for the membrane-embedded secretory components in their respective *bcs* operon. We showed earlier that deletion of the BcsB periplasmic modules ($BcsB^{peri}$) did not abolish Bcs macrocomplex assembly (6), indicating that membrane targeting of the cytosolic components likely precedes the multimerization of BcsB protomers in the crown. BcsR is a short 7-kDa polypeptide with unknown structure and function, whereas BcsQ is predicted to belong to the ancient SIMIBI (signal recognition particle, MinD and BioD) superfamily of NTPases (10, 11). Members of the latter are key to a large variety of cellular processes, including bacterial flagellar secretion (FlhG and FlhF) and membrane protein sorting in both prokaryotes and higher organisms (SRP54-SR and Get3) (10, 11). This, together with our earlier observations that BcsQ affects detection of the downstream BcsA synthase in the membrane (6), suggests that BcsQ might play a role in the early stages of cellulose secretion system assembly. A third cytosolic protein, BcsE, has been shown to significantly boost cellulose secretion *in vivo* and to present a second c-di-GMP binding module in addition to the $BcsA^{PilZ}$ domain (6, 12). Previous work has defined BcsE as a GGDEF-I-site-like (GIL) domain-containing protein due to c-di-GMP recognition by a conserved RXXD sequence, which, when found on diguanylate cyclases, can serve as a product-sensing regulatory motif called "I-site" (12). Finally, the Bcs secretion system is completed by a short membrane-embedded polypeptide, BcsF, that is also necessary for maximal cellulose production *in vivo* through an as-yet-unknown mechanism (6).

We showed earlier that both BcsF and the cytosolic BcsERQ components assemble stably with the inner membrane BcsAB biosynthetic platform to form the seashell-like

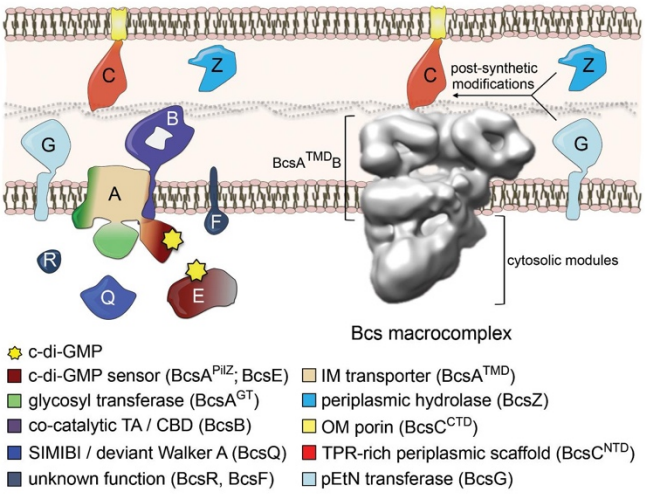


FIG 1 *E. coli*-like cellulose secretion systems. (Left) Thumbnail representation and proposed topology of the nine Bcs proteins. (Right) Electron microscopy-based three-dimensional reconstruction of the Bcs macrocomplex, encompassing most of the inner membrane and cytosolic subunits. Known and proposed roles for the different subunits and/or protein domains are color coded and annotated at the bottom. GT, glycosyl transferase domain; TA, tail-anchor; CBD, carbohydrate-binding domains; SIMIBI, signal recognition particle, MinD and BioD superfamily; TMD, transmembrane domain(s); IM, inner membrane; OM, outer membrane; NTD, N-terminal domain; CTD, C-terminal domain; TPR, tetratricopeptide repeats; pEtN, phosphoethanolamine.

Bcs macrocomplex visualized by single-particle electron microscopy (Fig. 1) (6). The low resolution of these structural data, however, precluded us from gaining specific insights into individual regulatory components or the interdependence of Bcs subunit interactions. Here, we determine that essential-for-secretion BcsR and BcsQ determine each other's folding and stability and that their membrane targeting is facilitated by high-affinity interactions with the c-di-GMP sensor BcsE. To unravel the latter's structure and function, we solved the crystal structure of a stable, N-terminally truncated BcsE variant ($BcsE^{217-523}$) and reveal that the previously postulated GIL domain is in fact a degenerate receiver-GGDEF domain tandem (REC*-GGDEF*). We further show that the catalytically incompetent diguanylate cyclase module senses through separate interfaces both BcsQ and c-di-GMP and that the dinucleotide likely adopts a dimeric conformation in solution, such as the one necessary for processive BcsA gating loop displacement and glucose polymerization (8). We also present evidence that although BcsQ is recruited by the C-terminal $BcsE^{GGDEF*}$ domain, efficient BcsERQ membrane targeting requires the remaining N-terminal module ($BcsE^{NTD}$) and that membrane partitioning is largely triggered by inner membrane BcsF. Finally, we determine that BcsE further uses its N-terminal domain to both homooligomerize and interact with transcription antitermination complex (TAC) components and discuss a putative physiological role for these unexpected interactions. Together, the data presented here suggest that BcsE and BcsF proteins might have evolved in *E. coli*-like cellulose secretion systems to boost exopolysaccharide production through actions at multiple levels: from high-affinity sequestration and membrane targeting of essential-for-secretion components to the maintenance of a membrane-proximal pool of dimeric c-di-GMP for processive synthase activation.

RESULTS AND DISCUSSION

BcsR and BcsQ interdependence and heterocomplex formation. In cellulose-producing enterobacteria, *bcs* genes are typically arranged in two separate operons,

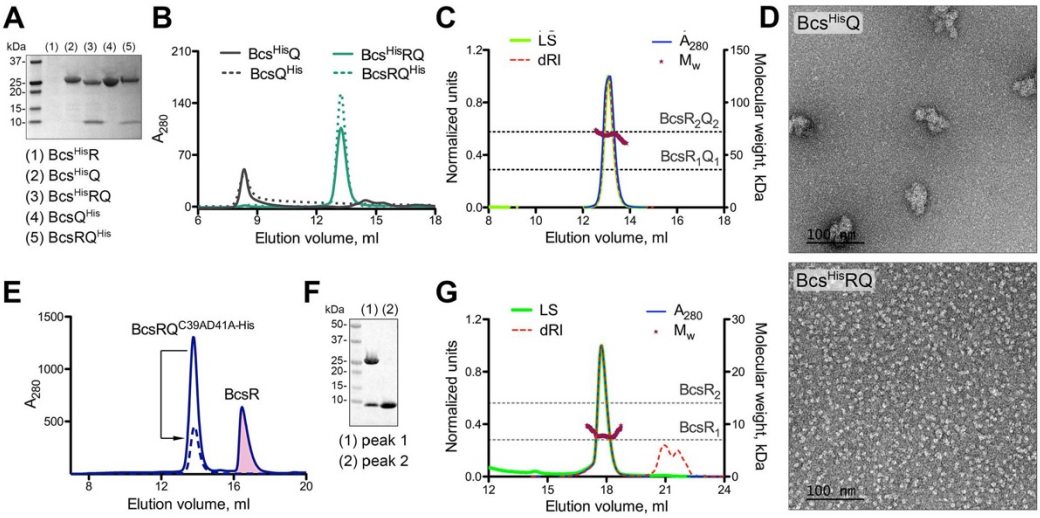


FIG 2 BcsR and BcsQ interdependence and complex formation. (A) IMAC elution fractions upon expression of Bcs^{His}R (pProExHTB-Bcs^{His}R), Bcs^{His}Q (pProExHTB-Bcs^{His}Q), Bcs^{His}RQ (pProExHTB-Bcs^{His}RQ), BcsQ^{His} (pET21b-BcsQ^{His}), and BcsRQ^{His} (pET21b-BcsRQ^{His}). (B) Size exclusion chromatography (SEC) profiles of the purified proteins from panel A (2 to 5) using a Superdex 200 Increase 10/30 GL column. (C) SEC-coupled multiangle light scattering (SEC-MALS) of purified tag-free BcsRQ complex. Normalized experimental traces for the light scattering (LS), differential refractive index (dRI), UV absorbance at 280 nm (A_{280}) and calculated molecular weight (M_w) are annotated at the top, theoretical molecular weights for the BcsRQ complex at 1:1 and 2:2 stoichiometries are shown as dashed lines. (D) Electron micrographs in negative stain of the purified Bcs^{His}Q (top) and Bcs^{His}RQ complex (bottom). (E) Purification of folded noncomplexed BcsR upon coexpression with the BcsQ$^{C39AD41A\text{-}His}$ mutant following IMAC and SEC. The BcsR peak is colored in pink. A dashed line shows the SEC profile of the mutant BcsRQ$^{C39AD41A\text{-}His}$ complex upon reinjection. (F) SDS-PAGE analysis of fractions corresponding to the two peaks in panel E. (G) SEC-MALS of the purified BcsR protein with experimental and theoretical traces as described above.

with hallmark *bcsRQ* and *bcsE* genes featuring promoter-proximal positions in each (5, 6). Different efforts to purify BcsR and BcsQ constructs on their own were not successful, with the proteins failing to express stably (Bcs^{His}R) or aggregating upon purification (Bcs^{His}Q and BcsQ^{His}) (Fig. 2; see also Table S1 in the supplemental material). Coexpression of the two subunits, however, led to the stable expression and purification of a homogeneous heterotetrameric BcsRQ complex with apparent 2:2 stoichiometry in solution, where BcsR-dependent BcsQ stabilization appeared independent of the presence or position of epitope tags on the subunits (Bcs^{His}RQ, BcsRQ^{His}, and BcsRQ) (Fig. 2A to D). Interestingly, while individual expression of BcsR did not yield detectable levels of purified protein, we identified empirically BcsQ variants that, when coexpressed under the same promoter with BcsR (BcsRQ$^{C39AD41A\text{-}His}$ and BcsRQ$^{C39AD41AL43D\text{-}His}$), yielded an excess of purified BcsR protein, which remained relatively stable in monomeric form in solution (Fig. 2E to G). These data indicate that the two proteins likely exhibit chaperone-like functions toward each other, where BcsR stabilizes BcsQ to form monodisperse heterotetramers in solution, while BcsQ itself might play a role in the folding and subsequent stability of BcsR.

Although bacterial operons have now been described for more than half a century (13), only recently have mechanistic insights into the role of operon organization begun to emerge. In particular, not only are proteins that function together through the assembly of heteromeric complexes likely to be encoded by genes in the same or adjacent operons, but operon gene order has also been reported as generally optimized for the order of protein complex assembly itself (14). This appears to be especially true for low-copy systems, as are typically the energetically costly secretion systems, where expression-coupled protein-protein interactions would minimize the stochasticity of heterocomplex formation (14). Nevertheless, protein folding in the context of multiprotein assemblies, as well as intraoperon partners remains enigmatic.

Studies on native and engineered proteins have shown that charged or intrinsically disordered N-terminal domains and protein tails can act as so-called "entropic bristles" with protein folding helper effects that stabilize fused downstream modules by minimizing their intrinsic aggregation propensity (15–17). We propose here that BcsRQ complex formation represents a paradigm of similar folding helper effects at the intraoperon level where upstream expression of an initially disordered BcsR minimizes the aggregation of its intraoperon partner BcsQ. The sequential expression of the two proteins could therefore not only limit the stochasticity of complex assembly within a low-copy cellulose secretion system but also couple the inhibition of intermolecular BcsQ aggregation with the intramolecular folding of BcsR to secure a stable stoichiometric assembly. Moreover, maintenance of separate polypeptides versus the evolution of genetically fused modules could present further advantages of operon organization, such as additional regulatory inputs or possible stoichiometry and symmetry variations upon secretion system assembly and function. In support for this model, recent work from our group has revealed that even in the context of the stable $BcsR_2Q_2$ heterocomplex, BcsR features a highly flexible and partly disordered N-terminal region that can partake in nonsymmetric protein-protein interactions, whereas the C-terminal domain adopts an α-helical fold at the interface of two BcsQ protomers (W. Abidi, S. Zouhir, M. Caleechurn, S. Roche, and P. V. Krasteva, unpublished).

BcsERQ complex formation and membrane targeting. We previously showed that although predicted as hydrophilic cytosolic proteins, BcsRQ associate stably with pelleted membranes in cell fractionation experiments and subsequently copurify with the detergent-extracted Bcs macrocomplex (6). Based on sequence conservation and putative fold recognition, BcsQ belongs to the ancient family of SIMIBI NTPases, many of which are involved in membrane-mediated processes such as division septum inhibition (MinD), flagellar assembly (FlhG and FlhF), protein secretion, and membrane protein sorting (Srp54-SR and Get3), among others (10, 11). While some of these proteins are targeted to the membrane via specific protein-protein interactions, others, such as MinD and FlhG homologs, have intrinsic membrane-targeting sequences (MTS) that are proposed to adopt an amphipathic α-helical fold upon contact with membrane lipids (Fig. 3A) (18–20). Comparative sequence analysis shows that a conserved basic residue midway in the MTSs of MinD and FlhG homologs is replaced by a proline in the corresponding 10-residue-long C-terminal tail of BcsQ ($BcsQ^{C10}$) (Fig. 3A). Although proline is generally a potent breaker of both α-helical and β-strand secondary structures in aqueous environments, it is often found in putative transmembrane protein helices and has been shown to protect α-helical conformations in hydrophobic milieus (21). To determine the potential role of $BcsQ^{C10}$ in membrane targeting, we performed cell-based phenotypic and *in vitro* lipid-binding assays. Interestingly, deletion of the $BcsQ^{C10}$ region had no significant effect on cellulose secretion in a functional complementation assay *in vivo* (Fig. 3A), and purified BcsRQ failed to partition with the lipid-enriched fractions in liposome flotation experiments *in vitro* (Fig. 3B). As these data favor protein-based membrane targeting of the essential-for-secretion BcsRQ complex, we proceeded to determine the nature and sequence of downstream BcsRQ interactions. We started by probing putative interactions with the third cytosolic component, BcsE, and after testing different strategies for BcsERQ recombinant coexpression (see Materials and Methods), we were able to purify a stable BcsERQ heterocomplex with equimolar 2:2:2 stoichiometry in solution (Fig. 3C to E).

BcsE occupies a leader position in its respective *bcsEFG* operon, which is consistent with a role in the early stages of Bcs macrocomplex assembly (14). Residues 164 to 474 of the protein were previously defined as a conserved GGDEF I-site-like (GIL) domain based on the identification of a c-di-GMP binding RXXD motif, similar to the I-site regulatory sequence often found on diguanylate cyclases (12) (Fig. 3F). Interestingly, fold recognition programs predict that the N-terminal BcsE region, which features significantly lower overall sequence conservation (see Fig. S1A), adopts a RecA-like ATPase fold whose boundaries significantly overlap those of the postulated C-terminal

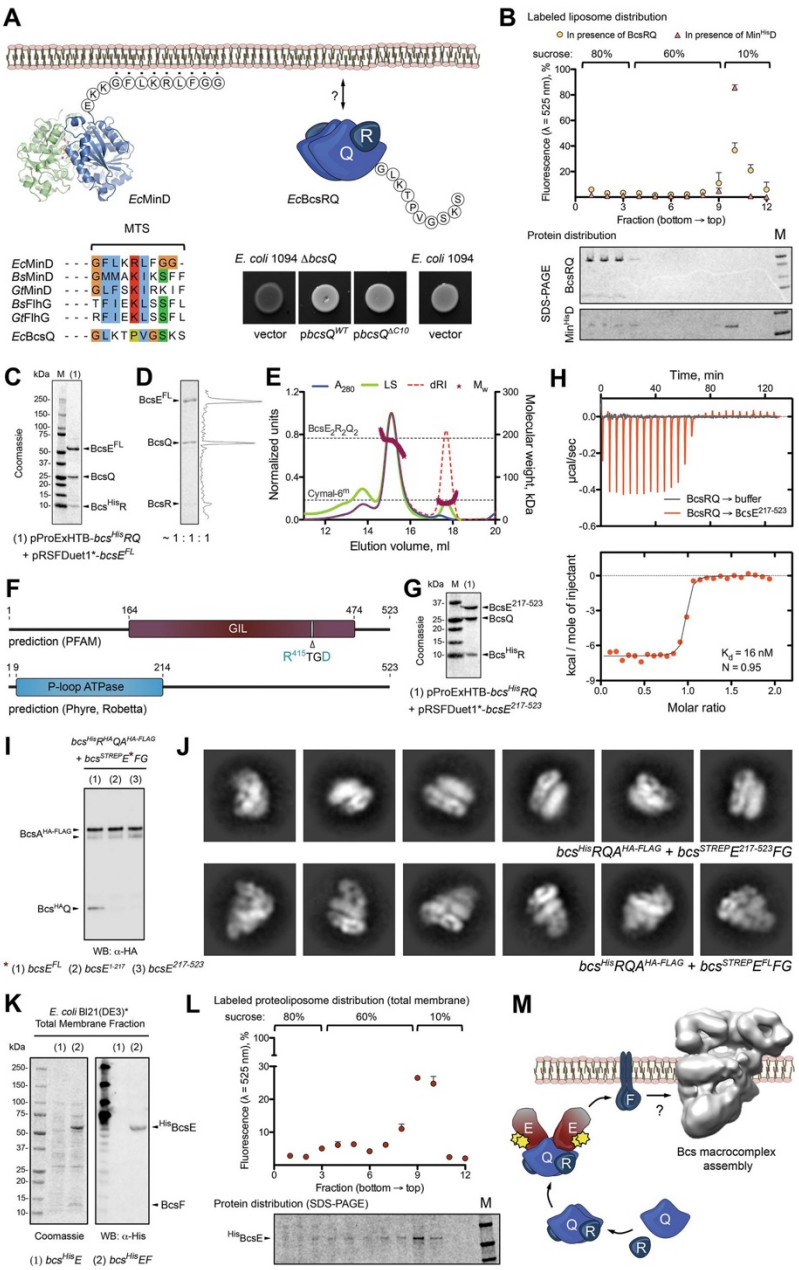


FIG 3 BcsRQ membrane targeting and BcsERQ complex formation. (A) Testing the presence and role of a putative membrane targeting sequence (MTS). (Top left) crystal structure and MTS sequence of the *E. coli* MinD protein. (Bottom left) MTS conservation among representative SIMIBI proteins and comparison with the corresponding BcsQ C-terminal tail (C10). (Top right) Thumbnail representation of the BcsRQ complex and C10 sequence. (Bottom right) Calcofluor binding assay for

(Continued on next page)

GIL domain of the protein (Fig. 3F) (22, 23). To identify stable BcsE modules and examine their role in secretion system assembly, we used both sequence conservation criteria and predicted three-dimensional fold models to create a series of N- and C-terminally truncated BcsE variants for recombinant coexpression. From these, we identified a construct, $BcsE^{217-523}$, that copurifies with BcsRQ similarly as full-length BcsE (Fig. 3G). When individually expressed, the truncated variant featured higher purity, stability, and protein yields than $BcsE^{FL}$, which allowed us to obtain a thermodynamic profile of the BcsRQ → BcsE interaction and reveal a dissociation constant in the low nanomolar range ($K_d \approx 16$ nM) (Fig. 3H). Considering that the binding affinity is likely even higher in the crowded high-viscosity environment of the bacterial cytosol and that the typical volume of an *E. coli* cell is in the low femtoliter range (24), these data indicate that as soon as the first copies of folded BcsRQ heterocomplex are formed, they will be bound and sequestered by their BcsE partners *in vivo*.

Interestingly, biochemical and electron microscopy data show that Bcs complexes purified via a C-terminal FLAG tag on BcsA fail to efficiently incorporate cytosolic Bcs components upon deletion of either the N-terminal $BcsE^{1-217}$ or the C-terminal $BcsE^{217-523}$ regions (Fig. 3I and J). While the latter can be explained by disrupted BcsERQ complex formation through deletion of the BcsRQ binding module, the effects of $BcsE^{1-217}$ deletion indicate that this N-terminal domain remains virtually indispensable for BcsERQ membrane targeting and its stable incorporation into the native Bcs macrocomplex.

We previously showed that deletion of BcsE intraoperon partners BcsF and BcsG have similar effects of incomplete macrocomplex assembly as the deletion of $BcsE^{1-217}$ shown here (Fig. 3J) (6). Though both BcsF and BcsG are inner membrane proteins, BcsG is involved in covalent modifications of the secreted cellulose in the periplasm and does not purify stably with the assembled Bcs macrocomplex (6, 9). We therefore hypothesized that of the two, BcsF is more likely to act at the early stages of Bcs macrocomplex assembly as a membrane triggering factor. To test this, we examined BcsE membrane partitioning in the presence or absence of BcsF ($Bcs^{His}E^{FL}$ versus $Bcs^{His}E^{FL}F$ expression). Indeed, BcsEF coexpression led to enrichment of BcsE in the pelleted total membrane fraction (Fig. 3K), and the protein partitioned with membrane-derived proteoliposomes upon flotation, effectively ruling out potential aggregation in the coexpression context (Fig. 3L). These data provide further support for coordinated subunit expression and protein complex assembly, where BcsRQ-bound BcsE is subsequently recruited to the inner membrane by its immediate downstream operon neighbor, BcsF.

Interestingly, in *Pseudomonas putida*, the BcsF gene is preceded by two putative open reading frames (PP_2629 and PP_2630), each of which shares conservation with

FIG 3 Legend (Continued)
cellulose secretion in wild-type (positive control) and mutant (Δ*bcsQ*) *E. coli* 1094 upon complementation with wild-type or C10-truncated BcsQ. Transformation with an empty pAM238 vector in the Δ*bcsQ* background was used as a negative control. (B) Liposome flotation assay of potential BcsRQ-lipid interactions. (Top) Relative fluorescence of sucrose gradient fractions after NBD-PE-labeled liposome flotation in the presence of $Min^{His}D$ (positive control) or BcsRQ. (Bottom) Representative SDS-PAGE analysis of protein distribution along the gradient fractions. Migrated proteins were stained with Coomassie. (C) SDS-PAGE analysis of the IMAC elution fraction upon coexpression of a $Bcs^{His}R$, BcsQ, and $BcsE^{FL}$ complex (pProExHTB-Bcs^{His}RQ plus pRSFDuet1*-$BcsE^{FL}$ coexpression strategy). (D) Calculated protein ratio in the purified tag-free $BcsERQ^{FL}$ complex based on densitometric analysis of the SDS-PAGE migrated bands. (E) SEC-MALS of the purified $BcsERQ^{FL}$ complex. Experimental and theoretical traces (as described above) are shown for both protein and detergent micelle (Cymal-6^m) peaks. (F) Conserved domain detection using sequence alignment and fold prediction tools. (G) SDS-PAGE analysis of the IMAC elution fraction of upon coexpression of a $Bcs^{His}R$, BcsQ, and $BcsE^{217-523}$ complex (pProExHTB-Bcs^{His}RQ plus pRSFDuet1*-$BcsE^{217-523}$ coexpression strategy). (H) Isothermal titration calorimetry (ITC) profile of the $BcsE^{217-523}$ → BcsRQ interaction. (I) Western blot analysis of the Bcs^{HA}Q integration into anti-FLAG tag-purified Bcs macrocomplex ($Bcs^{His}R^{HA}QA^{HA\text{-}FLAG}B^{\text{-}STREP}EFG$ coexpression) in the context of $BcsE^{FL}$, $BcsE^{1-217}$, and $BcsE^{217-523}$. (J) Representative views (class averages) of Bcs macrocomplex carrying $BcsE^{217-523}$ (top) versus $BcsE^{FL}$ (bottom, control). (K) $BcsE^{FL}$ membrane targeting in the context of BcsF coexpression (pProExHTB-$Bcs^{His}E^{FL}$ versus pProExHTB-$Bcs^{His}E^{FL}F$ expression strategies). (Left) SDS-PAGE analysis of the total membrane fractions; (right) Western blot detection of $Bcs^{His}E^{FL}$ in the corresponding fractions. (L) Liposome flotation experiments using NBD-PE-labeled total membrane proteoliposomes from cells coexpressing $Bcs^{His}E^{FL}$ and BcsF (pProExHTB-$Bcs^{His}E^{FL}F$ coexpression). (Top) Relative fluorescence of the gradient fractions indicating proteoliposome distribution; (bottom) SDS-PAGE analysis of $Bcs^{His}E^{FL}$ distribution across the corresponding gradient fractions. (M) Results summary showing proposed membrane targeting and macrocomplex integration of the essential for secretion subunits BscR and BcsQ via cytosolic BcsE and membrane-embedded BcsF.

the BcsE$^{1-217}$ or Bcs$^{217-523}$ fragments empirically characterized here (see Fig. S2A and B). This, together with the significant difference in sequence conservation between the two BcsE fragments, points toward the evolution of multidomain enterobacterial BcsE from the genetic fusion of smaller protein subunits to secure not only c-di-GMP recognition as known for the so-called I-site RXXD motif but also efficient BcsRQ complexation and subsequent delivery to the inner membrane biosynthetic platform via high-affinity BcsE-BcsF interactions (Fig. 3M).

Crystal structure of BcsE$^{217-523}$. To gain further insights into BcsE structure and function, we pursued crystallization of the stable C-terminal BcsE$^{217-523}$ construct, which encompasses most of the postulated GIL domain module. Purified untagged BcsE$^{217-523}$ crystallized in the presence of c-di-GMP, and its structure was determined to 2.2 Å using single-wavelength anomalous dispersion (SAD) phasing on crystals grown from selenomethionine-derivatized protein (Table S2). The protein packed in the $P4_12_12$ space group with two BcsE$^{217-523}$ molecules per asymmetric unit adopting virtually identical conformations with root mean square deviation (RMSD) of 0.835 Å over all atoms. A single c-di-GMP is found splayed in a symmetrical conformation between the two protomers, and nucleotide recognition involves four residues of each subunit, namely, R^{415} and D^{418} from the conserved I site-like motif as well as the side chain of H^{445} and the peptide carbonyl of S^{432} (Fig. 4 and 5). Unexpectedly, the construct adopts a dual-domain fold with an apparent N-proximal module connecting via an interstitial helix (α_6) to a C-terminal domain, in which the last ~40 residues (C-terminal tail) remain unresolved in the structure (Fig. 4A and B).

A search for three-dimensional (3D) structural homologs using the fold recognition server DALI (25) revealed that the N-proximal domain adopts a receiver (REC) domain-like $(\beta\alpha)_5$ fold (26), where the central five-stranded parallel β-sheet is flanked by 4 α-helices, while the canonical α_1 is mostly unfolded in an extended conformation by a stretch of proline and other small uncharged amino acids (Fig. 4C). Canonical REC domains are typically found in tandem with DNA-binding modules in response regulator proteins, which use phosphoryl transfer from upstream kinases as input signals for transcription regulation (26). Structural and sequence alignments of BcsE^{REC*} with phosphorylation-competent receiver domains, however, show significant deviation from the amino acid consensus of key functional motifs, indicating that the module is unlikely to function in phosphotransfer-dependent signal transduction (Fig. 4D).

Similar DALI search using the resolved C-proximal domain as an input revealed the closest structural homolog as the cytosolic C-terminal domain of *Pseudomonas aeruginosa* PelD. Interestingly, the latter is itself a c-di-GMP-binding protein responsible for the activation of synthase-dependent exopolysaccharide secretion, the Pel system in pseudomonads, and has been characterized as a degenerate GGDEF domain-containing protein where c-di-GMP sensing is carried out by the conserved I-site motif (27–29). Indeed, both BcsE^{GGDEF*} and PelD^{GGDEF*} show severe degeneration of the consensus $\beta\alpha\alpha\beta\beta\alpha\beta$ catalytic core shared by diguanylate and adenylate cyclases (30), with the substrate-coordinating P-loop and α_1-helix completely missing and catalytic residues, including those from the signature GGDEF motif, showing significant divergence (Fig. 4E to H; see also Fig. S3). Nevertheless, I-site-dependent c-di-GMP complexation remains virtually unchanged from that of active diguanylate cyclases, with the dinucleotide participating in both polar and π-stacking interactions with the side chains of the conserved arginine (R^{415}) and aspartate (D^{418}) residues (Fig. 4E to H and 5A; Fig. S3). Taken together, these results classify BcsE as a member of a growing superfamily of c-di-GMP-sensing proteins, in which canonical signaling (REC, PAS, etc.) or enzymatic (GGDEF, EAL, etc.) modules have been repurposed to serve c-di-GMP-dependent signal transduction (2, 3).

C-di-GMP and BcsQ binding by the BcsE^{GGDEF*} domain. Although the primary sequence of the resolved BcsE$^{REC*-GGDEF*}$ modules is overall highly conserved, surface mapping of the amino acid conservation reveals distinct conserved residue clusters on both the degenerate receiver and diguanylate cyclase modules, which might be

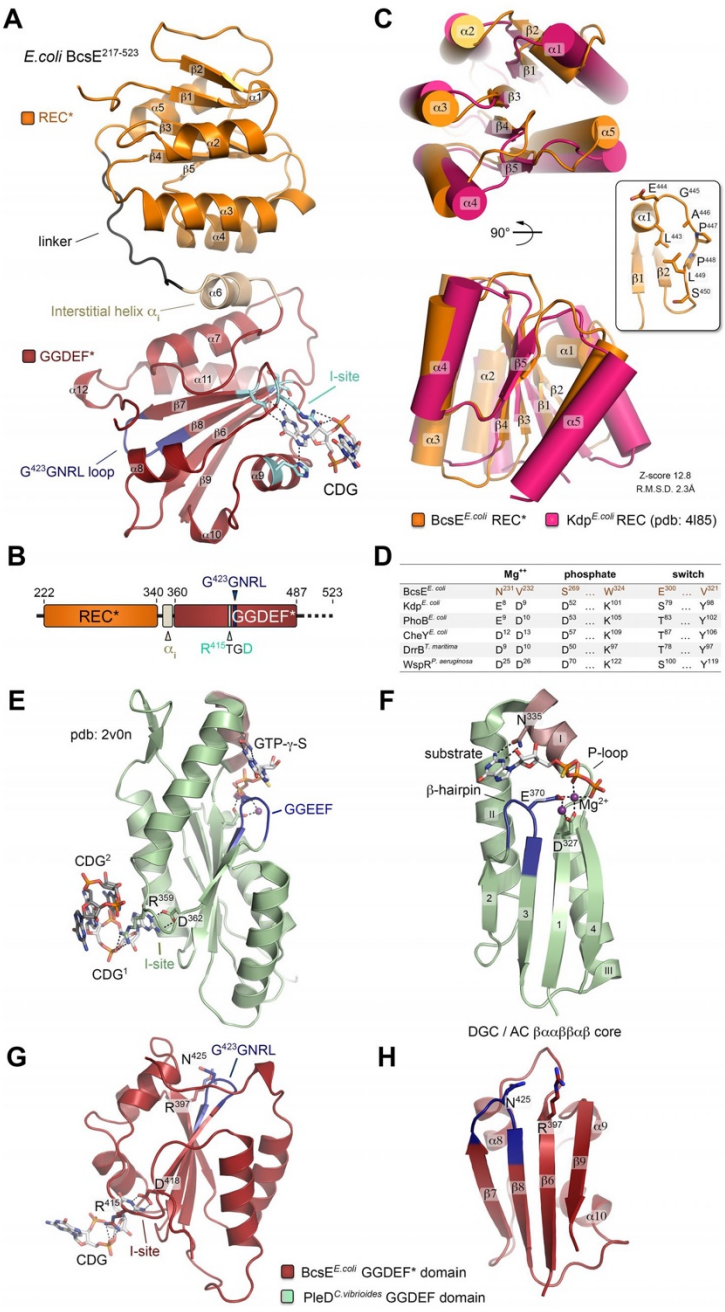


FIG 4 BcsE$^{217-523}$ crystal structure and domain organization. (A) Crystal structure of BcsE$^{217-523}$. The degenerate receiver (REC*) and diguanylate cyclase (GGDEF*) domains are colored orange and red, respectively, and

(Continued on next page)

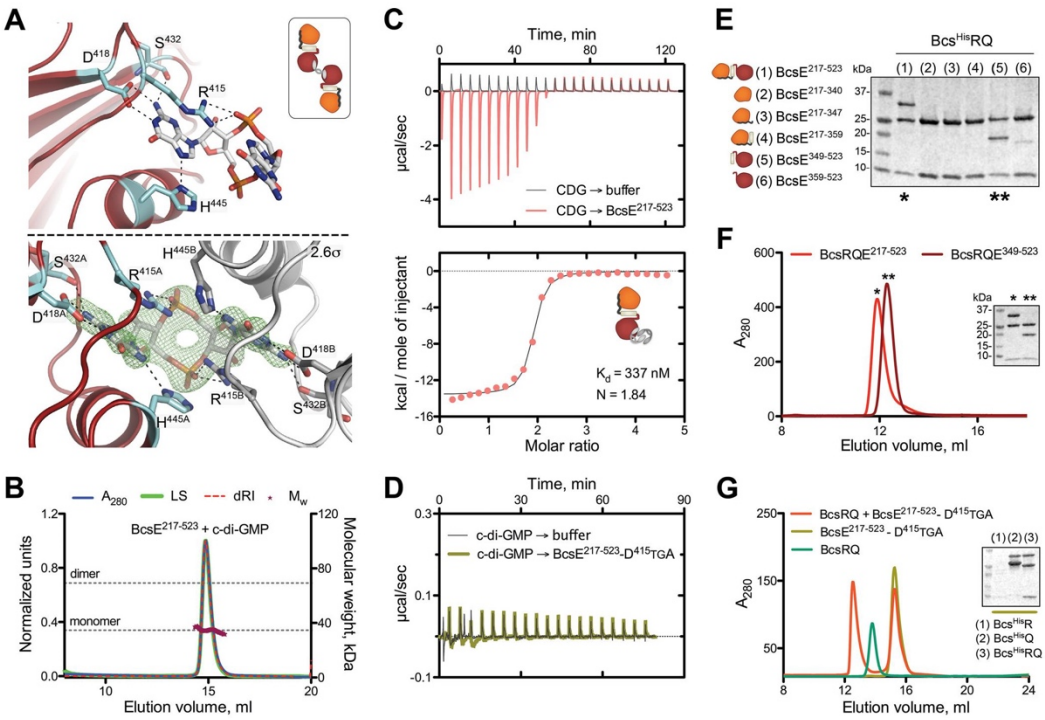


FIG 5 c-di-GMP binding and BcsE-BcsQ interactions. (A) c-di-GMP binding to the conserved I-site. (Top) Stick representation of c-di-GMP and the coordinating residues with only one protein molecule shown. (Inset) Thumbnail representation of the 2:1 protein-to-dinucleotide complexation observed in the crystals. (Bottom) An (|Fo|-|Fc|) partial electron density map calculated from a model prior to inclusion of the dinucleotide and contoured at 2.6σ with both coordinating protomers shown in red/cyan and gray. (B) SEC-MALS of BcsE$^{217-523}$ in the presence of excess c-di-GMP with experimental and theoretical traces as described above. (C) ITC profile of the c-di-GMP → BcsE$^{217-523}$ interaction and thumbnail representation of the calculated binding stoichiometry (1:2, protein to dinucleotide). (D) Control ITC titration of c-di-GMP to the I-site-defective BcsE$^{217-523}$-D^{415}TGA mutant. (E) SDS-PAGE analysis of IMAC elution fractions testing BcsERQ complex formation upon Bcs^{His}RQ coexpression with various truncated BcsE variants (pProExHTB-Bcs^{His}RQ plus pRSFDuet1*-BcsE^{trunc} coexpression). (F) SEC profiles of the purified BcsERQ^{GGDEF*} complex and BcsERQ$^{217-523}$. (G) SEC profile of purified BcsRQ preincubated with excess BcsE$^{217-523}$-D^{415}TGA compared to profiles for separate injections of the individual components. (Inset) SDS-PAGE analysis of IMAC elution fractions assaying BcsE$^{217-523}$-D^{415}TGA copurification upon coexpression with Bcs^{His}R, Bcs^{His}Q, or Bcs^{His}RQ.

indicative of oligomerization or protein-protein interaction interfaces (Fig. S1). To determine the protein's homooligomerization propensity in solution, we performed solution-based light-scattering experiments and determined that the BcsE$^{217-523}$ construct remains monomeric even in the presence of saturating c-di-GMP (Fig. 5B). This is particularly surprising considering the intrinsic dimerization propensity of REC domains in general (26), the binding stoichiometry of the BcsE$^{217-523}$RQ complex in solution (N $\approx$ 0.95, consistent with 2:2:2 binding) (Fig. 3H), and the symmetrical c-di-GMP conformation in the crystal structure (Fig. 5A, bottom and inset; Fig. S3A and B), where

FIG 4 Legend (Continued)

key motifs are highlighted. Secondary structure elements are numbered without accounting for the missing N-terminal domain. (B) Summary of the resolved domain architecture for the previously predicted GIL domain. (C) Overlay of the *E. coli* BcsE^{REC*} domain and a canonical receiver domain (*E. coli* Kdp^{REC}) in two different views. Inset, unfolding of the canonical α1 helix into a P-rich loop. Structural alignment scores are calculated in DALI. (D) Comparison of key conserved residues in phosphotransfer-competent response regulators with corresponding residues in BcsE^{REC*}. (E) GGDEF domain, I-site-mediated c-di-GMP binding, and substrate homologue coordination of the catalytically active diguanylate cyclase PleD$^{C\ vibrioides}$. (F) Conserved $\beta\alpha\alpha\beta\beta\alpha\beta$ catalytic core shared among adenylate and diguanylate cyclases. Key residues involved in substrate and Mg^{2+} coordination are shown as sticks. Crystal structures of the corresponding BcsE^{GGDEF*} domain (G) and α/β core (H).

the dinucleotide bridges two separate BcsE protomers by identical interactions. Furthermore, thermodynamic characterization of the c-di-GMP → $BcsE^{217-523}$ interaction reveals a binding stoichiometry consistent with two c-di-GMP molecules binding to a single BcsE I-site rather than the apparent inverse stoichiometry observed in the crystals (Fig. 5C versus Fig. 5A; Fig. S3A and B). These results are consistent with both the propensity of c-di-GMP to adopt diverse conformations, including intercalated dimers in solution (2, 3), and the capability of GGDEF I-sites to coordinate both monomeric and dimeric ligands as shown for *P. aeruginosa* PelD (28, 29) (Fig. S3C and D). Importantly, the dimeric c-di-GMP conformation derived from the solution-based data is also consistent with the reported c-di-GMP conformation necessary for $BcsA^{PilZ}$ domain binding and gating loop displacement during each step of UDP-glucose coordination and cellulose incorporation of the sugar moiety (Fig. S3E and F) (8). We therefore propose that Bcs macrocomplex-bound BcsE could secure the maintenance of a secretion system-proximal pool of c-di-GMP in a BcsA-activating conformation, thus limiting dinucleotide diffusion and boosting processive glucose polymerization.

In line with the monomeric state of $BcsE^{REC*\text{-}GGDEF*}$ in solution, equimolar 2:2:2 $BcsE^{217-523}$RQ heterocomplex assembly appears to be driven by the BcsRQ interactions rather than the BcsE variant itself. To determine how BcsE binds each half of the BcsRQ complex, we designed a series of shorter BcsE variants for copurification assays and identified a GGDEF* domain construct, $BcsE^{349-523}$, covering the interstitial helix α_i, the GGDEF* domain, and the unstructured C-terminal tail that partakes in stable equimolar interactions with BcsRQ (Fig. 5E and F). We also found that BcsRQ binding is independent of c-di-GMP complexation, as BcsERQ complex reconstitution can be carried out in the absence of dinucleotide and with an I-site mutant incapable of c-di-GMP complexation (R^{415}TGD → D^{415}TGA). Finally, we show that $BcsE^{GGDEF*}$ interacts with BcsQ rather than BcsR, as shown in copurification experiments using individual Bcs^{His}R or Bcs^{His}Q proteins as baits (Fig. 5F and G). Indeed, a stretch of highly conserved residues distinct from the c-di-GMP binding I-site is found on one side of the $BcsE^{GGDEF*}$ module that could have evolved for high-affinity BcsQ binding (Fig. S1B). Importantly, nonoverlapping sites for c-di-GMP and BcsQ complexation would allow both stable assembly of BcsRQ within the Bcs macrocomplex and the possibility of c-di-GMP to migrate in and out of the I-site in a model where the dinucleotide processively switches between the $BcsE^{GGDEF*}$ module and BcsA's PilZ domain for cocatalytic synthase regulation.

$BcsE^{NTD}$-dependent homooligomerization and binding of conserved Nus antitermination complex components. As mentioned above, the N-terminal region of BcsE ($BcsE^{1-217}$) is predicted to adopt a conserved RecA-like ATPase fold (Fig. 6A). RecA-like motor ATPases are a large family of proteins that use the energy of nucleotide binding and hydrolysis to oligomerize and perform mechanical work in a variety of cellular functions, such as the transport or hydrolysis of proteins (e.g., ABC transporters and proteases) or the binding and remodeling of nucleic acid substrates (e.g., helicases and recombinases) (31).

Structural and sequence alignments of $BcsE^{NTD}$ with catalytically active RecA-like ATPases show severe divergence of key functional motifs (e.g., the ATP/Mg^{2+}-coordinating Walker A motif) (Fig. 6A), indicating that BcsE is likely incapable of ATP binding and hydrolysis. Nevertheless, bacterial two-hybrid assays based on split adenylate cyclase (AC) functional reconstitution (32) suggest that $BcsE^{FL}$ is prone to oligomerization and, consistent with the monomeric state of the $BcsE^{REC*\text{-}GGDEF*}$ tandem described above (Fig. 5B), that these interactions are $BcsE^{NTD}$ dependent (Fig. 6B). Interestingly, blue colony growth indicative of $BcsE^{FL}$ and $BcsE^{NTD}$ homooligomerization was only observed in cases when coexpressed AC fragments were fused to different BcsE termini, regardless of their specific type (T25 or T18) (32) or location (N or C terminus) in the fusion constructs. These data suggest that the homotypic BcsE interactions involve different surface regions among the interacting BcsE protomers and are thus consistent with head-to-tail oligomerization mechanisms that are frequently observed in biologically active RecA-like ATPases (Fig. 6B). However, whether

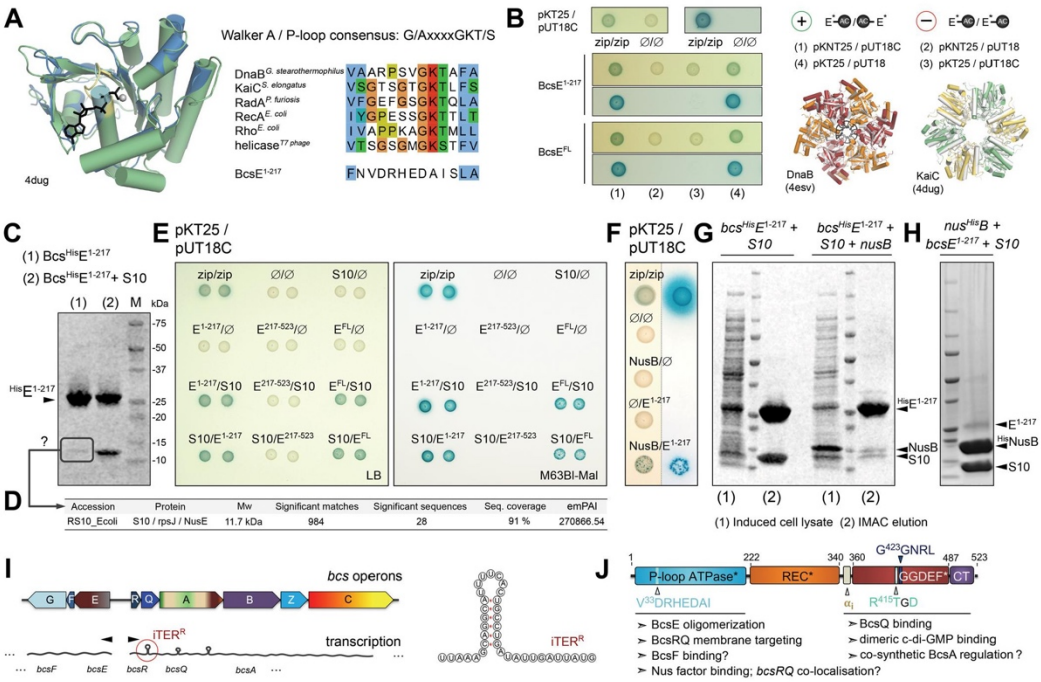


FIG 6 BcsE oligomerization and binding of conserved Nus antitermination complex components. (A) Predicted BcsE$^{1-217}$ RecA-like ATPase fold. (Left) overlay of the prediction model onto *Synechococcus elongatus* KaiC. (Right) Walker A conservation shown in representative RecA-like ATPases and compared to the corresponding region of BcsE$^{1-217}$. (B) BcsE$^{1-217}$ and BcsE^{FL} homooligomerization. (Left and top right) Bacterial two-hybrid assays using different coexpression strategies for the BcsE-AC fragment fusions. (Bottom right) Overlay of modeled BcsE$^{1-217}$ copies (in gray) with head-to-tail oligomers of RecA-like ATPases in superhelical (DnaB) or ring (KaiC) oligomeric states. (C) IMAC elution fractions of BcsE$^{His1-217}$ when expressed individually or coexpressed with *E. coli* NusE/S10 protein. (D) Mass spectrometry-based protein identification of a consistently copurifying low-molecular-weight band from an individually expressed, purified, and SDS-PAGE-migrated BcsE$^{1-217}$. (E) Bacterial two-hybrid assay of interactions between BcsE domains and S10/NusE based on plasmid-based adenylate cyclase functional reconstitution in a *cya*-defective *E. coli* strain (BTH101). The positive zip/zip control is based on coexpressed adenylate cyclase fragments each fused to a homodimerizing leucine zipper region of the yeast protein GCN4. Interactions were evaluated by the growth of blue colonies on X-Gal-supplemented LB (left) or M63BI (right) agar plates. (F) Bacterial two-hybrid assays of BcsE$^{1-217}$-NusB interactions. (G) SDS-PAGE analysis of induced cell lysates, IMAC elution fractions, and S10 copurification upon BcsE$^{His1-217}$ and S10 coexpression in the absence (left) or presence (right) of NusB. (H) IMAC elution fraction upon Nus^{His}B, BcsE$^{1-217}$, and S10 coexpression. (I) Organization of the two *bcs* operons and schematic representation of bioinformatically detected potential intrinsic terminators. (Right) Representation of a putative intrinsic terminator in *bcsR*. (J) Results summary showing BcsE domain architecture and proposed functional roles for the identified structural modules.

BcsE^{NTD} homooligomerization plays a functional role in the above-described targeting of the BcsERQ complex to the inner membrane (Fig. 3) or is involved in additional regulatory processes (see below), remains to be further examined.

Attempts to recombinantly purify the BcsE$^{1-217}$ construct consistently led to the copurification of a second protein species, even in elevated imidazole and salt concentrations in the immobilized-metal affinity chromatography (IMAC) purification buffer (Fig. 6C). Mass spectrometric analyses identified the copurifying species as small ribosomal protein S10, also known as NusE or RpsJ (Fig. 6D). Interestingly, apart from associating with the small ribosomal subunit during protein translation, S10 is also known to moonlight as a key component of the Nus transcription antitermination complex. Nus factors NusA, NusB, S10/NusE, NusG, and SuhB are mostly essential, highly conserved bacterial proteins that are known to associate with and reprogram the transcription apparatus in order to overcome elongation complex dissociation at certain intrinsic and Rho-dependent transcription terminators (33, 34). Although the best-studied examples include N protein-dependent antitermination at early λ phage

genes and the regulation of bacterial ribosomal (*rrn*) gene expression (34, 35), a recent study identified conserved NusB-S10 binding sites upstream of additional genes in diverse bacterial species (36). Nevertheless, characterized S10 interactions at the protein-protein level have remained limited to the context of assembled ribosomes or extensively studied transcription antitermination subcomplexes (e.g., see references 35 and 37).

Intrigued by this, we proceeded to assay the putative BcsE-S10 interaction by recombinant coexpression/copurification and cell-based bacterial two-hybrid experiments. We observed that $BcsE^{NTD}$ copurified at an equimolar ratio with overexpressed S10 (pProExHTB-*bcsE*^{NTD} plus pRSFDuet1*-*s10* coexpression) and that BcsE likely interacts with S10 *in cellulo* as observed by blue colony growth in the context of both the truncated $BcsE^{1-217}$ construct ($BcsE^{NTD}$) and the full-length BcsE protein ($BcsE^{FL}$) (Fig. 6C and E). We further assayed putative interactions of $BcsE^{NTD}$ with a second Nus factor, NusB, known to interact directly with S10 in the early steps of TAC assembly onto the target mRNA. Whereas bacterial two-hybrid experiments were indicative of weak $BcsE^{NTD}$-NusB interactions *in cellulo*, recombinant coexpression of $Bcs^{His}E^{NTD}$ with tag-free S10 and NusB led to the purification of excess $Bcs^{His}E^{NTD}$ and only trace amounts of copurifying Nus factors, even if S10 expression levels appeared virtually unchanged (Fig. 6F and G). Conversely, recombinant coexpression of $Nus^{His}B$ with tag-free S10 and $BcsE^{NTD}$ led to the purification of an equimolar amount of NusB-S10 complex and trace amounts of a third species, whose molecular weight corresponds to that of $BcsE^{NTD}$ (Fig. 6H). Together, these data indicate that S10 likely uses similar surface regions to interact with its BcsE and NusB partners, whereas higher-affinity NusB complexation could cause competitive remodeling of the equimolar $BcsE^{NTD}$-S10 assemblies and subsequent release of free BcsE.

Although the physiological significance of the observed BcsE-Nus factor interactions remains enigmatic, our findings suggest a possibly broader role for the conserved Nus antitermination machinery than in the well-studied examples of ribosomal or viral gene expression. Interestingly, *in silico* prediction tools (38, 39) detect putative intrinsic terminators within both the *bcsR* and *bcsQ* coding regions. If these potential regulatory elements indeed function as predicted, then the protein-protein interactions observed here could serve to target (via S10 complexation upon exiting the ribosome) and subsequently release (via downstream recruitment of NusB upon antitermination complex assembly) BcsE at the site of *bcsRQ* expression, prior to binding the newly synthesized BcsRQ complex directly and delivering it to the inner membrane for downstream Bcs macrocomplex assembly. Such a hypothesis, however, remains to be experimentally tested.

Concluding remarks. Bacteria have evolved complex secretion machineries to deliver large molecules to the cell envelope, external milieu, or host cell targets. Although these systems are typically not essential for bacterial physiology *per se*, they could often provide significant advantages in interspecies competition or be key to a pathogen's infection cycle. Bacterial exopolysaccharide secretion shares many similarities with the various types of protein secretion systems in that it typically involves intricate signal transduction events to induce the expression and assembly of multiple subunits in order to provide the biosynthetic activities, physical conduit, and energetics for biopolymer extrusion through the complex bacterial envelope (40).

Often, secretion systems are viewed as such at the level of assembled macrocomplexes and substrate extrusion, whereas the initial steps of subunit expression and sequential protein-protein interactions remain largely overlooked. Here, we present the *E. coli*-like Bcs system as a new candidate paradigm for concerted secretion system assembly and function. We demonstrate that essential-for-secretion BcsR and BcsQ regulate each other's folding and stability, whereas BcsE packs a subtle but diverse toolkit to fine-tune enterobacterial cellulose production (Fig. 6J). We provide structural and functional data that reveal the protein's multidomain evolution, fold conservation, and complexation of synthase-activating intercalated c-di-GMP on one hand, together

with high-affinity BcsRQ recruitment and facilitated membrane targeting through BcsF interactions on the other. Although more research is needed to uncover physiological roles for the observed BcsE-Nus factor interactions or how the essential BcsRQ subunits control assembly and function of the inner membrane biosynthetic platform, this work lays an important milestone toward more comprehensive models of operon-encoded synthase-dependent polysaccharide secretion in bacterial biofilms.

MATERIALS AND METHODS

The experiments were not randomized, and the investigators were not blinded during experimental design, execution, or outcome assessment. However, most experiments were reproduced independently by different investigators, including crystallographic, biochemical, biophysical, and phenotypic functional assays.

Bacterial strains. Plasmids for recombinant protein expression (see below) were propagated in and isolated from *E. coli* DH5α cells. All recombinant protein expression for structural and *in vitro* biochemical studies was carried out in BL21(DE3) Star cells, including the expression of selenomethionine-derivatized protein. An *E. coli* 1094 Δ*bcsQ* strain was used for the complementation phenotypic assays with BcsQ variants expressed from a low-copy-number isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible vector (pAM238; see below). Finally, bacterial two-hybrid experiments were performed using chemically competent BTH101 cells and the IPTG-inducible pKT(N)25 and pUT18(C) expression plasmids with custom-modified multiple cloning sites (see below). All bacterial strains and plasmids used in this study are available upon request.

Recombinant DNA techniques. DNA manipulations were carried out using standard protocols for PCR, molecular cloning, transformation, and DNA analyses. Coding regions for BcsR, BcsQ, BcsRQ, BcsE, MinDE, S10, and NusB variants were amplified using *E. coli* 1094 genomic DNA as a template and a high-fidelity DNA polymerase (Phusion; New England BioLabs) and inserted via digestion/ligation cloning into IPTG-inducible expression vectors with custom-modified multiple-cloning sites (MCS). Point mutations, insertion of stop codons, MCS modifications, and domain deletions within previously reported and newly generated expression constructs were performed using inverse PCR-based protocols and mutation-specific oligonucleotides as primers. All recombinant vectors and introduced mutations were verified by DNA sequencing and, where applicable, IPTG-inducible protein expression.

Protein expression and purification. All pProExHTB-encoded constructs ($Bcs^{His}R$, $Bcs^{His}Q$, $Bcs^{His}RQ$, $Bcs^{His}E^{NTD}$, $Nus^{His}B$, and $Min^{His}DE$) were expressed as IPTG-inducible variants carrying N-terminal hexahistidine tags cleavable by the human rhinovirus (HRV) 3c protease. BcsQ was also cloned in a standard pET21b vector yielding a C-terminally hexahistidine-tagged protein. As all BcsR ($Bcs^{His}R$) and BcsQ ($Bcs^{His}Q$ and $BcsQ^{His}$) constructs failed to yield stable proteins, the coding region corresponding to the BcsRQ tandem was subsequently amplified and cloned into both the pProExHTB and pET21b expression vectors, adding a cleavable N-terminal or noncleavable C-terminal hexahistidine tag to BcsR (pProExHTB-^{His}RQ) or BcsQ (pET21b-RQ^{His}), respectively. For coexpression studies, the coding region corresponding to full-length tag-free BcsE was cloned into custom-modified pRSFDuet1* expression vector under the control of the first T7 promoter (pRSFDuet1*-$BcsE^{FL}$) (see Table S1 in the supplemental material). Full-length BcsE was also cloned in pProExHTB and pET-^{His}SUMO (see below) vectors for standalone expression, but the purified proteins were judged insufficiently stable or pure for structural studies. Based on sequence conservation (PFAM [41]) and predicted tertiary structure (Phyre2, Robetta [22, 23]), several N- and C-terminal deletions of the pRSFDuet1*-$BcsE^{FL}$ construct were tested for expression and copurification with BcsRQ. The coding region for the interacting $BcsE^{217-523}$ construct was subsequently cloned for standalone expression into a modified pET-^{His}SUMO plasmid, yielding a hexahistidine-tagged Ulp1-cleavable SUMO moiety fused to the N terminus of the protein of interest (pET-^{His}SUMO-$BcsE^{217-523}$). Based on the resulting crystal structure of the $BcsE^{217-523}$ construct, an additional construct corresponding to the C-terminal GGDEF* domain was designed ($BcsE^{349-523}$), and its coding sequence was cloned into the pRSFDuet1* and pET-^{His}SUMO expression vectors as described above. The $BcsE^{1-217}$ construct corresponding to the protein's N-terminal domain was cloned into both pProExHTB and pRSFDuet1* (site 1) vectors. For coexpression of $BcsE^{1-217}$ (in pProExHTB) with S10 and NusB, the coding sequences for the latter were cloned in the first and second sites, respectively, of custom-modified pRSFDuet1* vectors (pProExHTB-$Bcs^{His}E^{1-217}$ plus pRSFDuet1*-$S10^{(site\ 1)}$-$NusB^{(site\ 2)}$ coexpression; the pRSFDuet1* vector was further modified at the second promoter to introduce unique XhoI and HindIII restriction sites) (Table S1). In addition, a pProExHTB-$Nus^{His}B$ plus pRSFDuet1*-$BcsE^{1\text{-}217(site\ 1)}$-$S10^{(site\ 2)}$ was also employed. For control liposome flotation studies, the coding region for the MinDE tandem was PCR-amplified and cloned into a pProExHTB vector, and MinD was purified as a partner-free protein ($Min^{His}D$) from the clarified cytosolic fraction. Finally, $Bcs^{His}E^{FL}$ and $Bcs^{His}E^{FL}F$ cloned into pProExHTB vectors were used for examining the membrane-targeting role of BcsF. Protein constructs used in the bacterial two-hybrid studies are described separately.

For protein purification, all expression vectors were (co)transformed into chemically competent *E. coli* BL21(DE3) Star cells. For the expression of native proteins, cells were grown at 37°C under aerobic conditions in terrific broth (TB) medium supplemented with appropriate antibiotics (100 μg/ml ampicillin, 40 μg/ml kanamycin, or a combination of 70 μg/ml ampicillin plus 30 μg/ml kanamycin for coexpressed vectors). At a cell optical density corresponding to an optical density at 600 nm (OD_{600}) of 0.8 to 1.0, the cells were moved to 17°C, and overnight protein expression was induced by the addition of IPTG at a final concentration of 0.7 mM. For the expression of selenomethionine-derivatized proteins, 4 liters of cells was initially grown at 37°C in LB medium to an OD_{600} of 0.5 to 0.6. Cells were then pelleted by

centrifugation (4,000 × *g*, 15 min, 20°C), gently washed with 200 ml 1× SelenoMet medium base (Molecular Dimensions), collected again, and resuspended in 1 liter complete SelenoMet medium (Molecular Dimensions) supplemented with 40 mg/liter L-selenomethionine and the appropriate antibiotic. Cells were then grown for an additional 1 h at 37°C, transferred to 17°C, and induced with IPTG as described above.

After 16 h, cells were harvested by centrifugation, resuspended in lysis buffer, and flash-frozen in liquid nitrogen. The composition of the lysis buffer was 20 mM HEPES (pH 8.0), 120 mM NaCl, 19 mM imidazole (pH 8.0,) 2 mM β-mercaptoethanol, and 1 tablet/50 ml cOmplete protease inhibitors (Roche) for the $Bcs^{His}R$, $Bcs^{His}Q$, $BcsQ^{His}$, $Bcs^{His}RQ$, $Min^{His}DE$, and $Bcs^{His}E^{217-523}$ constructs. For the $Bcs^{His}RQ$-$BcsE^{FL}$, $Bcs^{His}RQ$-$BcsE^{217-523}$, and $Bcs^{His}RQ$-$BcsE^{349-523}$ complexes, the IMAC buffer was also supplemented with 0.5 μM c-di-GMP (Jena Bioscience or Sigma-Aldrich), 2 μM AppCp (Jena Bioscience), 5 mM $MgCl_2$, and 10% glycerol. For the expression of $Bcs^{His}E^{1-217}$ and the $Bcs^{His}E^{1-217}$-S10, $Bcs^{His}E^{1-217}$-S10-NusB, and $Nus^{His}B$-S10-$BcsE^{1-217}$ complexes, the concentration of salt in the lysis buffer was increased to 750 mM NaCl.

For all cytosolic protein purifications, cells were thawed and lysed using an Emulsiflex-C3 high-pressure homogenizer (Avestin). Cell debris was removed by centrifugation (1 h at 50,000 × *g* and 4°C), and the cleared lysates were loaded onto buffer-washed Talon Superflow resin (GE Healthcare) at approximately 0.5 to 1 ml of resin per liter of culture. The resin was subsequently washed with more than 20 volumes of IMAC buffer A (protease inhibitor-free lysis buffer as described above), and bound proteins were eluted in a single step with IMAC buffer A supplemented with 200 mM imidazole (pH 8.0) (IMAC buffer B).

For purification of tag-free BcsRQ, eluted protein $Bcs^{His}RQ$ protein was supplemented with 15 mM EDTA (pH 8.0) and homemade HRV3c protease at 4°C, concentrated to 2.5 ml using an Amicon Ultra centrifugal filter (30-kDa cutoff; Millipore), desalted using a disposable PD-10 desalting column (GE Healthcare), and incubated overnight for tag removal. The cleaved tag and protease were removed by inverse IMAC, concentrated, and subjected to size exclusion chromatography on a Superdex 200 Increase 10/300 GL column (GE Healthcare) equilibrated in gel filtration buffer (20 mM HEPES [pH 8.0], 120 mM NaCl, and 2 mM dithiothreitol [DTT]). Collected protein fractions were analyzed for purity by SDS-PAGE, pooled, concentrated, flash-frozen in liquid nitrogen, and stored at −80°C.

For purification of tag-free $BcsE^{217-523}$, the eluted ^{His}SUMO-fused protein was mixed with homemade yeast protease Ulp1, concentrated to 2.5 ml, desalted on a disposable PD-10 column, and incubated for ^{His}SUMO cleavage at 4°C overnight. Cleaved protein was collected in the flowthrough fraction during reverse IMAC on the following day, concentrated, and subjected to size exclusion chromatography on a Superdex 200 Increase 10/300 GL column equilibrated with gel filtration buffer (20 mM HEPES [pH 8.0], 100 mM NaCl, and 2 mM DTT). Collected protein fractions were analyzed for purity, concentrated, aliquoted, and flash frozen for storage at −80°C.

Complexes BcsRQ-$BcsE^{217-523}$ and BcsRQ-$BcsE^{349-523}$ were purified in a similar 2-step IMAC procedure. Eluted proteins were incubated with the viral HRV3c protease for cleavage of the N-terminal hexahistidine tag on BcsR. Imidazole concentrations were lowered via desalting on a disposable PD-10 column, and after overnight incubation at 4°C, the proteins were subjected to size exclusion chromatography using a Superdex 200 Increase 10/300 GL column and gel filtration buffer composed of 20 mM HEPES (pH 8.0), 120 mM NaCl, 5 mM $MgCl_2$, 0.5 μM c-di-GMP, 2 μM AppCp, 2 mM DTT, and 10% glycerol. Collected protein fractions were analyzed for purity and stoichiometric complex assembly, concentrated, and flash frozen in liquid nitrogen for storage at −80°C.

To characterize the complex formation and stoichiometry of interaction between BcsRQ and $BcsE^{FL}$, a ternary complex was coexpressed and purified using a similar protocol. However, as the complex appeared to be stabilized by the presence of detergents, after cell lysis, the cell debris was pelleted by slower centrifugation (12,000 × *g*, 15 min, 4°C), and the remaining supernatant was incubated with 0.25% *n*-dodecyl-β-D-maltopyranoside (β-DDM; Anatrace) for 1 h at 4°C. The lysates were then cleared by high-speed centrifugation, and the ternary BcsRQ-$BcsE^{FL}$ complex was purified as the rest of the BcsERQ complexes while keeping a low concentration of detergent (0.06% Cymal-6; Anatrace) in all buffers. For size exclusion chromatography, the Superdex 200 Increase 10/300 GL column was replaced by a Superose 6 Increase 10/300 GL column.

MinD was purified from clarified cytosolic fraction using a single-step metal-affinity purification (IMAC buffer A with 20 mM HEPES [pH 8.0], 120 mM NaCl, and 19 mM imidazole), followed by size exclusion chromatography on a Superdex 200 Increase 10/300 GL column equilibrated with gel filtration buffer (20 mM HEPES [pH 8.0], 100 mM NaCl, and 2 mM DTT). Clean protein fractions were concentrated and flash-frozen for storage at −80°C.

Cells expressing $Bcs^{His}E^{1-217}$, $Bcs^{His}E^{1-217}$-S10, $Bcs^{His}E^{1-217}$-S10-NusB, and $Nus^{His}B$-S10-$BcsE^{1-217}$ were resuspended in high-salt lysis buffer (same as described above but with 750 mM NaCl), and the proteins were purified by a single-step IMAC. The high-salt conditions (750 mM NaCl) were maintained in all buffers.

Finally, expression and purification of the Bcs macrocomplex (pCDFDuet1-$Bcs^{His}RQA^{HA\text{-}FLAG}B$ plus pRSFDuet1*-^{Strep}EFG) with various BcsE ($Bcs^{Strep}E^{FL}$, $Bcs^{Strep}E^{1-217}$, or $Bcs^{Strep}E^{217-523}$) and BcsQ (BcsQ or $Bcs^{HA}Q$) variants were performed as reported previously (6). pRSFDuet1*-$Bcs^{Strep}E^{1-217}$FG was generated from pRSFDuet1*-$Bcs^{Strep}E^{FL}$FG via inverse PCR using two different strategies which yielded consistent results: (i) an insertion of a 4-letter STOP codon following BcsE residue A^{217} (TAAT in DNA) and (ii) a deletion of the REC*-GGDEF* tandem while preserving the ribosome-binding site for *bcsF* to avoid polar effects). pRSFDuet1*-$Bcs^{Strep}E^{217-523}$FG was generated by standard restriction/ligase subcloning. Insertion of a hemagglutinin (HA) tag at the N terminus of BcsQ was also conducted by inverse PCR. After

expression vector cotransformation, culture growth, and overnight expression induction, cells were pelleted by centrifugation and resuspended in ice-cold lysis buffer containing 20 mM HEPES (pH 8.0), 120 mM NaCl, 10% glycerol, 5 mM $MgCl_2$, 10 μM AppCp, 2 μM c-di-GMP, 250 μM cellobiose, 0.5 mg/ml *Aspergillus niger* cellulase (Sigma-Aldrich), 100 μg/ml lysozyme, and 1 tablet/50 ml cOmplete EDTA-free protease inhibitors (Roche). After lysis (Emulsiflex-C3), cell debris was removed by low-speed centrifugation (12,000 $\times$ *g*, 15 min, 4°C), and the membranes were pelleted by ultracentrifugation using an SW 28 Ti Beckman rotor (26,500 rpm, or up to 126,000 $\times$ *g*, for 1 h at 4°C). After removal of the supernatant, the membrane fraction was resuspended in solubilization buffer containing all lysis buffer components except lysozyme and cellulase, as well as a mix of detergents at the following final concentrations: 0.4% (wt/vol) digitonin (Sigma-Aldrich), 0.4% (wt/vol) *n*-dodecyl-β-D-maltopyranoside (anagrade β-DDM; Anatrace), 0.4% (wt/vol) decyl maltose neopentyl glycol (DM-NPG; Anatrace), and 0.2% lauryl maltose neopentyl glycol (LM-NPG; Anatrace). After a 60- to 90-min-long incubation at 20°C and under mild agitation, the solubilized membrane fraction was cleared by a second high-speed centrifugation step as described above. The supernatant was then incubated with anti-FLAG M2 affinity gel (50 μl packed resin per liter of induced culture; Sigma-Aldrich) with mild agitation at 4°C for 1 h. After gravity elution of the nonbound fraction, the resin was washed extensively (>30 column bed volumes) with binding buffer containing all lysis buffer components except lysozyme and cellulase, as well as 0.008% (wt/vol) LM-NPG. The bound complexes were then eluted using 4 column bed volumes of elution buffer (affinity buffer supplemented with 3$\times$ FLAG peptide at 100 μg/ml) and concentrated on a 100-kDa cutoff Amicon Ultra (MerckMillipore) centrifugal filter.

SDS-PAGE and Western blot analyses. Protein fractions were analyzed by standard denaturing SDS-PAGE using 4% to 20% gradient mini-gels (Bio-Rad), Expedeon InstantBlue Coomassie stain, and a Li-Cor Odyssey Fc system for Coomassie visualization (700-nm channel). For Western blot analyses, SDS-PAGE-migrated proteins were directly transferred using a standard mini-gel transfer protocol, 0.2-μm polyvinylidene difluoride (PVDF) membranes, and a Trans-blot Turbo transfer system (Bio-Rad). Blocking and antibody incubations were in the presence of 5% skim milk in Tris-phosphate-buffered saline (TPBS); all washes between and after antibody incubations were with 1$\times$ TPBS buffer. Rabbit anti-His_6 (dilution 1:1,000, ab200537; Abcam) and mouse anti-HA (dilution 1:1,000, number 26183; Thermo Fisher Scientific) antibodies were used as primary antibodies; Alexa Fluor 680-conjugated goat ant-rabbit (dilution 1:10,000, ab175773; Abcam) and donkey anti-mouse (dilution 1:10,000, ab175774; Abcam) were used as secondary antibodies. The Alexa Fluor 680 signal was detected using a Li-Cor Odyssey Fc system in the 700-nm channel.

Crystallization, data collection, and structure determination. Crystals were obtained by sitting or hanging-drop vapor diffusion by mixing equal volumes of protein (1.5 to 6 mg/ml) and reservoir solution followed by incubation at 4°C. $BcsE^{217-523}$ crystals also appeared within 3 to 14 days under multiple conditions, with diffracting data sets collected on crystals grown in 100 mM morpholineethanesulfonic acid (MES; pH 6.0), 4% polyethylene glycol 4000 (PEG 4000), 200 mM $MgCl_2$, 5% glycerol, and 50 μM c-di-GMP. For cryoprotection, crystals were soaked in reservoir solution supplemented with 25% to 30% glycerol, 1 mM DTT, and 50 μM c-di-GMP. Cryopreserved crystals were flash frozen and stored in liquid nitrogen. Data were collected on frozen crystals at 100 K using synchrotron radiation at beamlines PX1 and PX2 at the Soleil synchrotron.

Data reduction was carried out with the software package XDS (42). Experimental phases were obtained by single-wavelength anomalous diffraction (SAD) experiments on crystals grown from selenomethionine-derivatized protein and with wavelengths corresponding to the experimentally determined selenium K-edge. Initial $BcsE^{217-523}$ models were obtained using the automated model building tools of PHENIX and Buccaneer (43, 44). Reiterative refinements in PHENIX, COOT, and BUSTER yielded the final refined model (43, 45, 46). Data collection and refinement statistics are summarized in Table S2. For illustration purposes, all crystal structures were displayed with the PyMol Molecular Graphics System (Schrödinger, LLC) or UCSF Chimera (47). The latter was also used for displaying the 3D reconstructions of the assembled Bcs macrocomplex.

Single-particle electron microscopy. Negative-stain single-particle electron microscopy was used for visualization of various Bcs proteins and protein complexes. Briefly, 5 μl of eluted samples (concentration, ~0.01 to 0.05 mg/ml) were spotted on glow-discharged carbon-coated copper grids (Agar Scientific). After a 1-min incubation, the extra liquid was blotted off, and the grids were passed sequentially through three drops of 2% (wt/vol) uranyl acetate solution, with a second incubation in the last drop before blotting and air drying. Micrographs were taken on a Thermo Fisher Scientific T12 Tecnai electron microscope operated at 100 kV accelerating voltage and equipped with a LaB6 filament and a K2 Base direct electron detector. For the protein complexes purified from cells expressing $Bcs^{His}RQA^{HA\text{-}FLAG}B$ plus $Bcs^{Strep}E^{FL}FG$ or $Bcs^{His}RQA^{HA\text{-}FLAG}B$ plus $Bcs^{Strep}E^{217-523}FG$, particles were autopicked in EMAN2 (48), saved as .box coordinates, and converted into a .star particle stack in Relion2 (49). Micrograph contrast transfer function (CTF) correction and two-dimensional (2D) classification were performed in cryoSPARC v2 after particle reextraction using the Relion2-generated .star file as metadata input (50). A total of 3,810 particles were classified for the Bcs macrocomplex carrying full-length BcsE (control) and 6,242 particles for the complex purified from cells expressing the $BcsE^{217-523}$ truncated variant.

Protein identification by mass spectrometry. Coomassie-stained gel bands were excised and subjected to in-gel enzymatic digestion. Briefly, the bands were extensively washed with acetonitrile and 100 mM NH_4HCO_3, dried, and treated with 10 mM DTT at 56°C for 30 min. After DTT removal, cysteine carbamidomethylation was performed at room temperature for 30 min by the addition of 55 mM iodoacetamide. The washing procedure was then repeated, the gel slices were dried, and the proteins were digested overnight at room temperature by the addition of 20 μl/band of 10 ng/μl Porcine Gold

trypsin (Promega) diluted in 50 mM NH_3HCO_3. Peptides were extracted in two steps: first, with 20 μl of 50% acetonitrile-0.1% formic acid solution and, second, with 20 μl of 100% acetonitrile. Peptides were vacuum dried and resuspended in 5% acetonitrile-0.1% trifluoroacetic acid (TFA) prior to nanoscale liquid chromatography-tandem mass spectrometry (nanoLC-MS/MS) analyses. The latter were performed with a TripleTOF 4600 mass spectrometer (Sciex) coupled to an UltiMate 3000 RSLCnano system (Thermo Fisher Scientific). Peptides were first desalted on an Acclaim Pepmap 100 C_{18} reverse-phase precolumn (3 μm, 100 Å, 75-μm inside diameter [i.d.], 2-cm length) using a loading buffer containing 2% acetonitrile and 0.05% TFA in water and a flow rate of 5 μl/min. A second Acclaim Pepmap 100 C_{18} column (2 μm, 100 Å, 75-μm i.d., 50-cm length) was then used as an analytical column, and bound peptides were eluted from the reverse phase using a 5% to 35% solvent B gradient for 40 min at a flow rate of 300 nl/min (solvent A, 0.1% formic acid in water; solvent B, 0.1% formic acid in 100% acetonitrile). nanoLC-MS/MS experiments were conducted using data-dependent acquisition by selecting the 20 most intense precursors for collision-induced dissociation (CID) fragmentation with the Q1 quadrupole set to low resolution for increased sensitivity. Raw data were processed using proprietary MS data converter software (Sciex), and protein identification was performed using the Mascot search engine (Matrix Science) against the *E. coli* taxon in the Swiss-Prot database and with carbamidomethylation of cysteines set as fixed modification. Oxidation of methionines was set as variable modifications. Peptide and fragment tolerances were set at 25 ppm and 0.05 Da, respectively. Only peptides with a Mascot score higher than the identity threshold (30) at less than 1% of false-positive discovery rate were considered.

Size exclusion chromatography coupled with static multiangle light scattering. For size exclusion chromatography coupled with static multiangle light scattering (SEC-MALS), purified proteins or protein complexes were subjected to gel filtration using a Superose 6 Increase 10/300 GL column (GE Healthcare) for the BcsRQ-$BcsE^{FL}$ complex or a Superdex 200 Increase 10/300 GL column for all other samples. The columns were preequilibrated with the respective gel filtration buffers. For the $BcsE^{217-523}$ construct, the protein was analyzed in its apo form, as well as following incubation with 2-fold excess c-di-GMP. The high-performance liquid chromatography (HPLC) system (Shimadzu) was coupled to a 3-angle light-scattering detector (miniDAWN TREOS) and a refractive index detector (Optilab rEX) (Wyatt technology). For each experiment, the system was preequilibrated overnight with gel filtration buffer and at the desired flow rate. Data were collected every second at a flow rate of 0.5 ml/min and were analyzed using the ASTRA software to obtain the molar mass and polydispersity of the sample across the protein elution peaks. For detector normalization and data quality control, purified delipidated bovine serum albumin (BSA; Sigma) was used prior to each experiment.

Protein complex reconstitution using purified proteins. Purified BcsRQ complex was incubated with excess $BcsE^{217-523}$-D^{415}TGA, and after a 15-min incubation on ice, the proteins were subjected to size exclusion chromatography using a Superdex 200 Increase 10/300 GL column and using c-di-GMP-free gel filtration buffer (20 mM HEPES [pH 8.0], 100 mM NaCl, and 2 mM DTT). The purified BcsRQ and $BcsE^{217-523}$-D^{415}TGA proteins were also injected on their own in separate chromatography runs. BcsRQ-$BcsE^{217-523}$-D^{415}TGA complex formation was detected by depletion of the BcsRQ peak and appearance of a new A_{280} peak shifted toward the front of the column.

Isothermal titration calorimetry. Apparent dissociation constants (K_d) and stoichiometry of interactions (N) were measured by isothermal titration calorimetry (ITC) using a Microcal VP-ITC calorimeter from Malvern Panalytical at 20°C. For c-di-GMP binding studies, 0.8 to 1 mM c-di-GMP was used as a ligand in the syringe, and 50 μM purified protein was added to the cuvette. The proteins and ligand were purified/diluted in the exact same buffer to minimize nonspecific dilution heat effects. Protein concentrations were determined by a combination of methods, including a reducing agent-compatible colorimetric assay (RC DC; Bio-Rad) and 280-nm absorbance measurements under denaturing conditions (A_{280}; 6 M guanidinium chloride), while accounting for potential scattering contributions (A_{330}). For BcsRQ-$BcsE^{217-523}$ complex formation, tag-free BcsRQ and $BcsE^{217-523}$ were purified in the same buffer (20 mM HEPES [pH 8.0], 100 mM NaCl, and 2 mM DTT), and 10-fold more concentrated BcsRQ (180 μM) was titrated from the syringe into $BcsE^{217-523}$ (18 μM) in the cuvette. All ITC data were analyzed by integrating the injection heat effects, normalized to the amount of ligand and protein present, and curve fitting based on a single-site binding model using the Origin software package for Microcal. For all titrations, titrations of the ligand into buffer were performed to account for heat dilution effects, and the latter were subtracted during the ligand binding analysis. The apparent dissociation constants (K_d) and stoichiometries of interaction (N) were derived from the data by using standard procedures, and the graphs were replotted using GraphPad Prism software.

Calcofluor-binding cellulose secretion assay. To test for the functional effects of the C-terminal BcsQ deletion ($BcsQ^{\Delta C10}$), chemically competent cells were prepared from an *E. coli* 1094 Δ*bcsQ* deletion strain (6). The latter was transformed with a low-copy-number plasmid (pAM-238) carrying wild-type or mutant *bcsQ* genes and plated on LB agar plates (Miller) supplemented with the appropriate antibiotics (60 μg/ml streptomycin and 15 μg/ml chloramphenicol). Single colonies were inoculated in 5 ml LB medium with antibiotics and left to grow overnight at 37°C with agitation. On the following morning, 5 μl of each culture was spotted onto low-salt LB agar plates (1.5 g/liter NaCl) supplemented with the antibiotics, 0.1 mM IPTG, and 0.02% calcofluor (fluorescent brightener 28; Sigma-Aldrich). The spots were allowed to air dry, and the plates were incubated at 30°C. After 24 h, the plates were photographed under brief illumination with long-wave UV light (365 nm).

Bacterial two-hybrid assay. We used the adenylate cyclase two-hybrid complementation assay to probe protein-protein interactions (32). We first custom modified the standard expression vectors to introduce BamHI and KpnI cloning sites in the respective MCS while optimizing the number and type of exogenous amino acids to be added to the recombinant hybrids by PCR amplification and restriction

digestion of the products (Table S1). An intrinsic KpnI site in the BcsE^{FL} coding region was also modified by introducing a silent mutation in the pRSF-Duet-BcsE^{FL} construct through inverse PCR. Coding regions for BcsE, S10, and NusB full-length proteins or truncated variants were then PCR amplified with primers carrying the corresponding restriction sites, digested, and ligated into the modified vectors. All recombinant constructs were amplified in DH5α cells and verified by DNA sequencing.

The bacterial two-hybrid assay was performed using standard protocols (32). Briefly, chemically competent *E. coli* BTH101 cells were cotransformed with derivatives of the pUT18(C) and pK(N)T25 vectors and plated on LB Miller agar supplemented with 100 μg/ml ampicillin and 40 μg/ml kanamycin. Individual cotransformant colonies were picked and grown overnight at 37°C in liquid antibiotic-supplemented LB medium. The next morning, 4 μl of saturated culture was spotted onto LB Miller agar (supplemented with 100 μg/ml ampicillin, 40 μg/ml kanamycin, 0.1 mM IPTG, and 40 μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside [X-Gal]) or M63BI agar (supplemented with 50 μg/ml ampicillin, 25 μg/ml kanamycin, 0.1 mM IPTG, 40 μg/ml X-Gal, and 0.2% maltose) plates. Protein interactions were evaluated after approximately 30 h of incubation at 30°C by blue colony color in the case of LB Miller agar plates and by both colony growth and blue color in the case of M63BI agar plates. pUT18(C) and pK(N)T25 vectors carrying only the AC fragment coding sequences were used in cotransformations as negative controls, whereas cotransformants expressing pKT25-zip and pUT18C-zip vectors were used as positive controls. The latter vectors are derivatives of the pK(N)T25 and pUT18(C) vectors in which the leucine zipper of Gcn4 is genetically fused in frame to the T25 and T18 adenylate cyclase fragments, respectively. The results are representative of at least 3 independent experiments and 6 biological replicates.

Liposome flotation experiments. We used liposome flotation to monitor the correlation between the distributions of fluorescently labeled liposomes and Coomassie-stained proteins across sucrose density gradients. For BcsRQ membrane binding studies (Fig. 3B), we used extruded liposomes prepared from commercially available *E. coli* total membrane lipids (Avanti Polar Lipids, Inc.) at stock concentration of 10 mg/ml in buffer containing 120 mM NaCl and 20 mM HEPES (pH 8.0). For lipid detection, 1% (vol/vol) of egg L-α-phosphatidylethanolamine-*N*-(7-nitro-2-1,3-benzoxadiazol-4-yl) (NBD-PE; Avanti Polar Lipids, Inc.) was added to the mix prior to liposome preparation. For this, powdered lipids were hydrated with buffer and subjected to multiple cycles of flash freezing in liquid nitrogen, thawing at 50°C, sonication, and 0.1-μm filter extrusion. Liposomes and purified protein (BcsRQ or, as a positive control, MinD) were mixed in 100 μl of the same buffer at final concentration of 1 mg/ml each, and this volume was then premixed with the heavy fraction of a sucrose density gradient (3 ml of 80% sucrose in buffer) in 13.2-ml ultracentrifuge tubes. The heavy fraction was then layered with 6 ml 60% sucrose in buffer, which in turn was layered with 3 ml 10% sucrose in buffer without mixing the gradient layers. The samples were then subjected to ultracentrifugation in a Beckman Coulter SW 41 Ti rotor for 16 h at 35,000 rpm and 4°C. One-milliliter aliquots were then gently removed from the top, and the NBD fluorescence for each fraction was measured with excitation at 465 nm and emission detection at 535 nm using a Perkin Elmer LS-50B luminescence spectrometer. In parallel, protein distribution along the gradient aliquots was visualized by denaturing SDS-PAGE. Additions of ADP-Mg^{2+}, ATP-Mg^{2+}, AppCp-Mg^{2+}, and cardiolipin were tested but not found to affect BcsRQ distribution along the density gradients. Results are representative of at least 3 experiments with 2 technical replicates each.

For Bcs^{His}E^{FL} versus Bcs^{His}E^{FL}F localization (Fig. 3K), *E. coli* Bl21(DE3) Star cells overexpressing the respective constructs from pProExHTB expression vectors were lysed using an Emulsiflex-C3 homogenizer in buffer containing 20 mM HEPES (pH 8.0), 120 mM NaCl, 10% glycerol, 1 μM c-di-GMP, and 1 tablet/50 ml cOmplete Protease inhibitor cocktail. Cell debris was removed by centrifugation at 12,000 $\times$ *g* and 4°C for 10 min, and the membrane fraction was then pelleted from the supernatant by ultracentrifugation at 35,000 rpm for 75 min at 4°C using a Beckman Coulter SW 41 Ti rotor. After removing the supernatant, the membranes were resuspended with a Potter-Elvehjem homogenizer in the same buffer, and after a second ultracentrifugation step, the pelleted membranes were weighed and resuspended at 20 mg/ml final concentration, and 1% (vol/vol) egg NBD-PE was added to trace the lipid distribution. The total membrane protein content was analyzed by SDS-PAGE and Western blotting, leading to identification of the Bcs^{His}E^{FL} protein band enriched in the membrane fractions of BcsF-coexpressing cells (pProExHTB-^{His}E^{FL}F coexpression). These membrane fractions were subjected to several cycles of low-intensity sonication for liposome generation, and 100 μl of each replicate was mixed with the heavy fraction of a sucrose density gradient (3 ml 80% sucrose in buffer) and subjected to gradient ultracentrifugation as described above. Native proteoliposome distribution was visualized via detection of NBD fluorescence along the gradient aliquots, whereas Bcs^{His}E^{FL} distribution was visualized by SDS-PAGE and Coomassie staining (Fig. 3L). Results for all flotation experiments are representative of at least 3 independent experiments with 2 technical replicates each.

Protein and RNA structure prediction tools. Protein conserved domain detection and tertiary structure prediction were carried out using the NCBI BLASTP suite (https://blast.ncbi.nlm.nih.gov/Blast.cgi), PFAM database (https://pfam.xfam.org), Phyre2 (http://www.sbg.bio.ic.ac.uk), and Robetta servers (http://new.robetta.org) (22, 23, 41, 51). A search for structural homologues in the Protein Data Bank was carried out using the DALI server (http://ekhidna2.biocenter.helsinki.fi/dali/) (25). Potential intrinsic terminators in mRNA were predicted using the ARNold (http://rssf.i2bc.paris-saclay.fr/toolbox/arnold/) and iTERM-PseKNC (http://lin-group.cn/server/iTerm-PseKNC/) tools (38, 39).

Data availability. Crystallographic structure factors and coordinates have been deposited in the RCSB Protein Data Bank (https://www.rcsb.org/) with accession code 6TJ0.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 2.8 MB.

FIG S2, PDF file, 2.3 MB.

FIG S3, PDF file, 2.8 MB.

TABLE S1, PDF file, 0.1 MB.

TABLE S2, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We thank all current and former members of the Structural Biology of Biofilms group and especially Agnès Thomas-Collignon, Marius Poulain, Gaurav Shajepal, and Lucia Torres-Sanchez for technical assistance and/or work peripheral to the project, Stéphane Roche for X-ray crystallography advice and insightful discussions, the I2BC Virology department for provided equipment and laboratory space, Yoshiharu Yamaichi, Virginia Vicky Lioy, Jean-Marc Ghigo, and Holger Sondermann for providing expression vectors and bacterial strains, and Rémi Fronzes, Holger Sondermann, and Axel Innis for critical reading of the manuscript prior to submission.

P.V.K. conceived the project; P.V.K., S.Z., W.A., and M.C. designed, performed, and optimized the experimental procedures; P.V.K., S.Z., and W.A. analyzed the data; and P.V.K. wrote the paper with feedback from all authors.

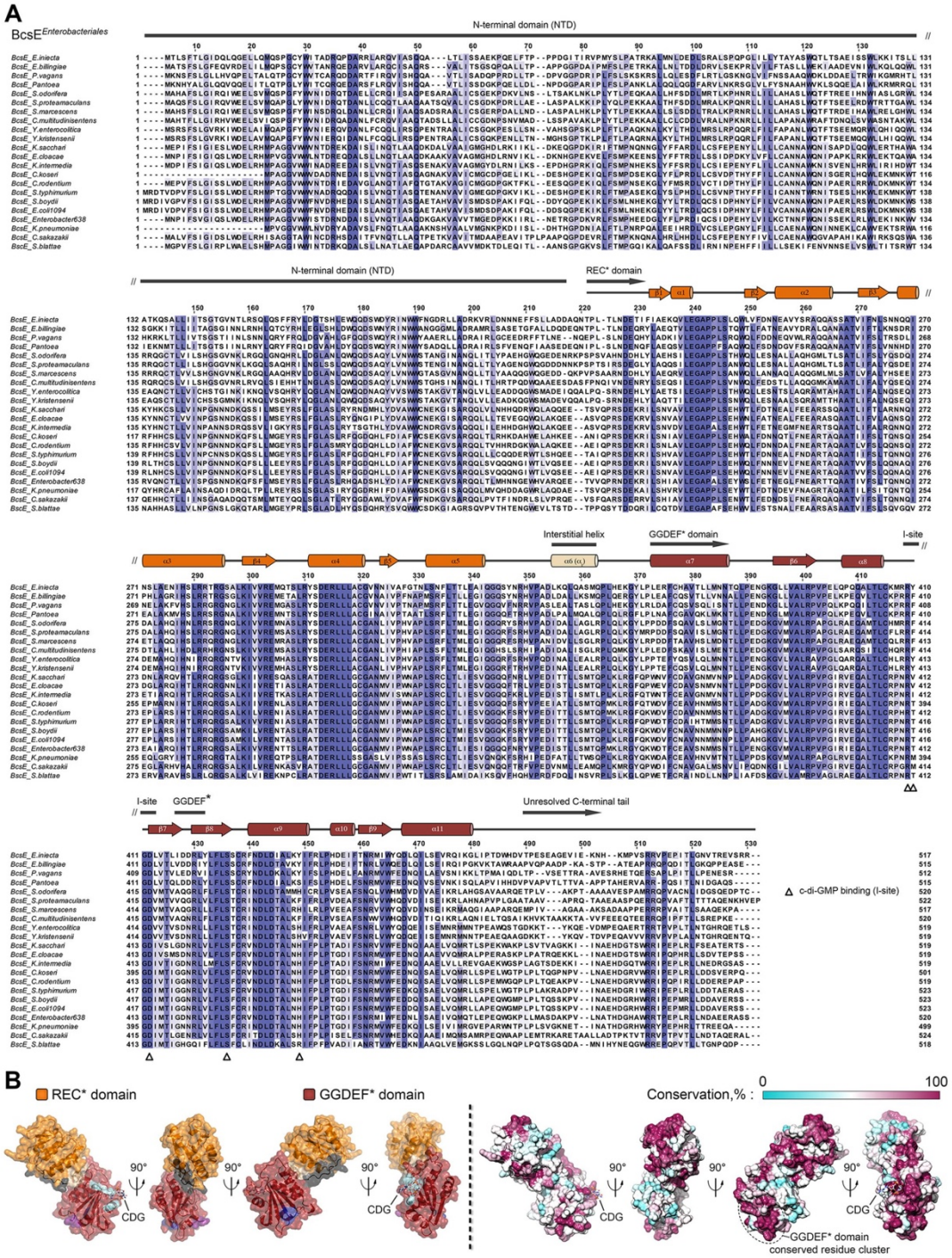
This project was supported by the Institute for Integrative Biology of the Cell (I2BC), the European Institute of Chemistry and Biology (IECB), the Centre National de la Recherche Scientifique (CNRS), an ATIP-Avenir starting grant (to P.V.K.), and a starting grant from the European Research Council (ERC StG BioMatrix to P.V.K.). The work has also benefited from the I2BC proteomics platform SICaPS, supported by IBiSA, Ile de France Region, Plan Cancer, CNRS and Paris-Sud University, as well as from the I2BC platforms PIM and Cryo-EM, which are supported by the French Infrastructure for Integrated Structural Biology (FRISBI; ANR-10-INBS-05).

REFERENCES

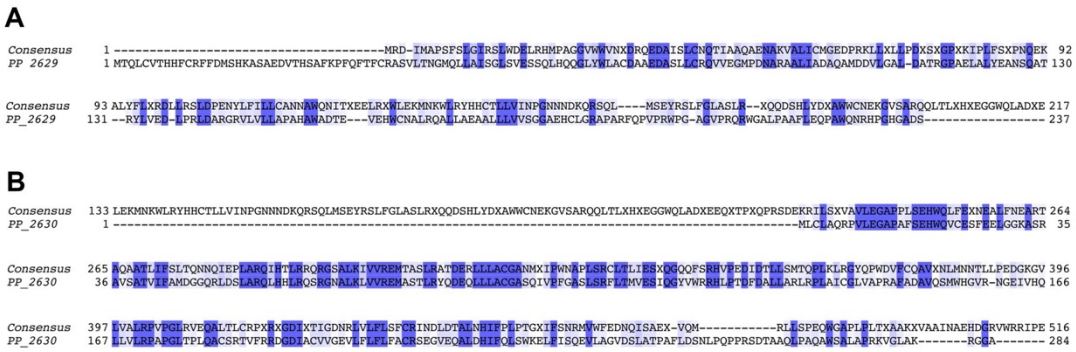
1. Hall-Stoodley L, Costerton JW, Stoodley P. 2004. Bacterial biofilms: from the natural environment to infectious diseases. Nat Rev Microbiol 2:95–108. https://doi.org/10.1038/nrmicro821.
2. Krasteva PV, Giglio KM, Sondermann H. 2012. Sensing the messenger: the diverse ways that bacteria signal through c-di-GMP. Protein Sci 21:929–948. https://doi.org/10.1002/pro.2093.
3. Krasteva PV, Sondermann H. 2017. Versatile modes of cellular regulation via cyclic dinucleotides. Nat Chem Biol 13:350–359. https://doi.org/10.1038/nchembio.2337.
4. McNamara JT, Morgan JL, Zimmer J. 2015. A molecular description of cellulose biosynthesis. Annu Rev Biochem 84:895–921. https://doi.org/10.1146/annurev-biochem-060614-033930.
5. Römling U, Galperin MY. 2015. Bacterial cellulose biosynthesis: diversity of operons, subunits, products, and functions. Trends Microbiol 23:545–557. https://doi.org/10.1016/j.tim.2015.05.005.
6. Krasteva PV, Bernal-Bayard J, Travier L, Martin FA, Kaminski P-A, Karimova G, Fronzes R, Ghigo J-M. 2017. Insights into the structure and assembly of a bacterial cellulose secretion system. Nat Commun 8:2065. https://doi.org/10.1038/s41467-017-01523-2.
7. Morgan JLW, McNamara JT, Fischer M, Rich J, Chen H-M, Withers SG, Zimmer J. 2016. Observing cellulose biosynthesis and membrane translocation *in crystallo*. Nature 531:329–334. https://doi.org/10.1038/nature16966.
8. Morgan JL, McNamara JT, Zimmer J. 2014. Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. Nat Struct Mol Biol 21:489–496. https://doi.org/10.1038/nsmb.2803.
9. Thongsomboon W, Serra DO, Possling A, Hadjineophytou C, Hengge R, Cegelski L. 2018. Phosphoethanolamine cellulose: a naturally produced chemically modified cellulose. Science 359:334–338. https://doi.org/10.1126/science.aao4096.
10. Shan SO. 2016. ATPase and GTPase tangos drive intracellular protein transport. Trends Biochem Sci 41:1050–1060. https://doi.org/10.1016/j.tibs.2016.08.012.
11. Bange G, Sinning I. 2013. SIMIBI twins in protein targeting and localization. Nat Struct Mol Biol 20:776–780. https://doi.org/10.1038/nsmb.2605.
12. Fang X, Ahmad I, Blanka A, Schottkowski M, Cimdins A, Galperin MY, Römling U, Gomelsky M. 2014. GIL, a new c-di-GMP-binding protein domain involved in regulation of cellulose synthesis in enterobacteria. Mol Microbiol 93:439–452. https://doi.org/10.1111/mmi.12672.
13. Jacob F, Perrin D, Sanchez C, Monod J. 1960. Operon: a group of genes with the expression coordinated by an operator. C R Hebd Seances Acad Sci 250:1727–1729. (In French.)
14. Wells JN, Bergendahl LT, Marsh JA. 2016. Operon gene order is optimized for ordered protein complex assembly. Cell Rep 14:679–685. https://doi.org/10.1016/j.celrep.2015.12.085.
15. Kim CW, Han KS, Ryu K-S, Kim BH, Kim K-H, Choi SI, Seong BL. 2007. N-Terminal domains of native multidomain proteins have the potential to assist de novo folding of their downstream domains *in vivo* by acting as solubility enhancers. Protein Sci 16:635–643. https://doi.org/10.1110/ps.062330907.
16. Santner AA, Croy CH, Vasanwala FH, Uversky VN, Van YY, Dunker AK. 2012. Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. Biochemistry 51:7250–7262. https://doi.org/10.1021/bi300653m.
17. Graña-Montes R, Marinelli P, Reverter D, Ventura S. 2014. N-Terminal protein tails act as aggregation protective entropic bristles: the SUMO case. Biomacromolecules 15:1194–1203. https://doi.org/10.1021/bm401776z.
18. Szeto TH, Rowland SL, Rothfield LI, King GF. 2002. Membrane localization of MinD is mediated by a C-terminal motif that is conserved across eubacteria, archaea, and chloroplasts. Proc Natl Acad Sci U S A 99:15693–15698. https://doi.org/10.1073/pnas.232590599.
19. Szeto TH, Rowland SL, Habrukowich CL, King GF. 2003. The MinD membrane targeting sequence is a transplantable lipid-binding helix. J Biol Chem 278:40050–40056. https://doi.org/10.1074/jbc.M306876200.
20. Schuhmacher JS, Rossmann F, Dempwolff F, Knauer C, Altegoer F,

Steinchen W, Dörrich AK, Klingl A, Stephan M, Linne U, Thormann KM, Bange G. 2015. MinD-like ATPase FlhG effects location and number of bacterial flagella during C-ring assembly. Proc Natl Acad Sci U S A 112:3092–3097. https://doi.org/10.1073/pnas.1419388112.

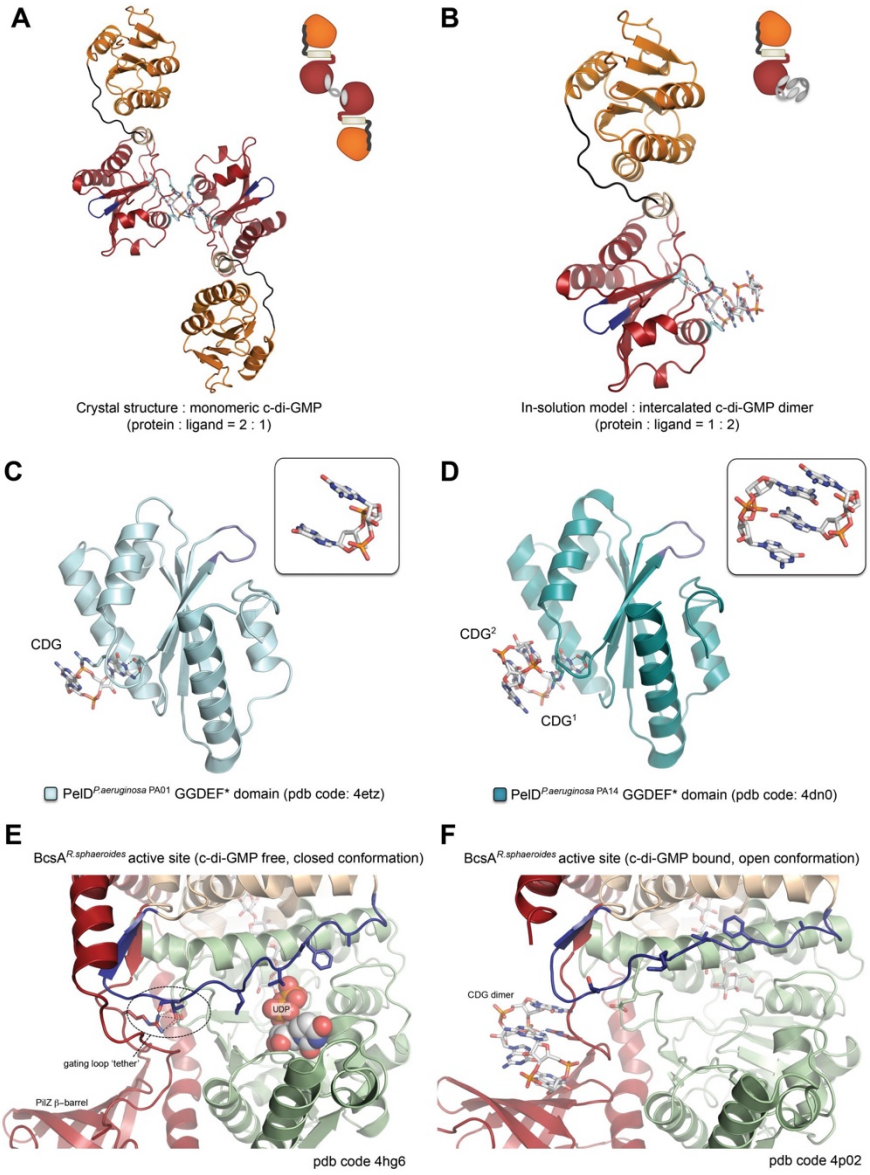
21. Li SC, Goto NK, Williams KA, Deber CM. 1996. Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. Proc Natl Acad Sci U S A 93:6676–6681. https://doi.org/10.1073/pnas.93.13.6676.
22. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858. https://doi.org/10.1038/nprot.2015.053.
23. Kim DE, Chivian D, Baker D. 2004. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32:W526–W531. https://doi.org/10.1093/nar/gkh468.
24. Volkmer B, Heinemann M. 2011. Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. PLoS One 6:e23126. https://doi.org/10.1371/journal.pone.0023126.
25. Holm L. 2019. Benchmarking fold detection by DaliLite v.5. Bioinformatics 35:5326–5327. https://doi.org/10.1093/bioinformatics/btz536.
26. Gao R, Stock AM. 2009. Biological insights from structures of two-component proteins. Annu Rev Microbiol 63:133–154. https://doi.org/10.1146/annurev.micro.091208.073214.
27. Lee VT, Matewish JM, Kessler JL, Hyodo M, Hayakawa Y, Lory S. 2007. A cyclic-di-GMP receptor required for bacterial exopolysaccharide production. Mol Microbiol 65:1474–1484. https://doi.org/10.1111/j.1365-2958.2007.05879.x.
28. Whitney JC, Colvin KM, Marmont LS, Robinson H, Parsek MR, Howell PL. 2012. Structure of the cytoplasmic region of PelD, a degenerate diguanylate cyclase receptor that regulates exopolysaccharide production in *Pseudomonas aeruginosa*. J Biol Chem 287:23582–23593. https://doi.org/10.1074/jbc.M112.375378.
29. Li Z, Chen JH, Hao Y, Nair SK. 2012. Structures of the PelD cyclic diguanylate effector involved in pellicle formation in *Pseudomonas aeruginosa* PAO1. J Biol Chem 287:30191–30204. https://doi.org/10.1074/jbc.M112.378273.
30. Schirmer T. 2016. C-di-GMP synthesis: structural aspects of evolution, catalysis and regulation. J Mol Biol 428:3683–3701. https://doi.org/10.1016/j.jmb.2016.07.023.
31. Ye J, Osborne AR, Groll M, Rapoport TA. 2004. RecA-like motor ATPases–lessons from structures. Biochim Biophys Acta 1659:1–18. https://doi.org/10.1016/j.bbabio.2004.06.003.
32. Karimova G, Pidoux J, Ullmann A, Ladant D. 1998. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc Natl Acad Sci U S A 95:5752–5756. https://doi.org/10.1073/pnas.95.10.5752.
33. Bubunenko M, Baker T, Court DL. 2007. Essentiality of ribosomal and transcription antitermination proteins analyzed by systematic gene replacement in *Escherichia coli*. J Bacteriol 189:2844–2853. https://doi.org/10.1128/JB.01713-06.
34. Santangelo TJ, Artsimovitch I. 2011. Termination and antitermination: RNA polymerase runs a stop sign. Nat Rev Microbiol 9:319–329. https://doi.org/10.1038/nrmicro2560.
35. Krupp F, Said N, Huang Y-H, Loll B, Bürger J, Mielke T, Spahn CMT, Wahl MC. 2019. Structural basis for the action of an all-purpose transcription anti-termination factor. Mol Cell 74:143.e5–157.e5. https://doi.org/10.1016/j.molcel.2019.01.016.
36. Baniulyte G, Singh N, Benoit C, Johnson R, Ferguson R, Paramo M, Stringer AM, Scott A, Lapierre P, Wade JT. 2017. Identification of regulatory targets for the bacterial Nus factor complex. Nat Commun 8:2027. https://doi.org/10.1038/s41467-017-02124-9.
37. Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JHD. 2005. Structures of the bacterial ribosome at 3.5 Å resolution. Science 310:827–834. https://doi.org/10.1126/science.1117230.
38. Naville M, Ghuillot-Gaudeffroy A, Marchais A, Gautheret D. 2011. ARNold: a web tool for the prediction of Rho-independent transcription terminators. RNA Biol 8:11–13. https://doi.org/10.4161/rna.8.1.13346.
39. Feng C-Q, Zhang Z-Y, Zhu X-J, Lin Y, Chen W, Tang H, Lin H. 2019. iTerm-PseKNC: a sequence-based tool for predicting bacterial transcriptional terminators. Bioinformatics 35:1469–1477. https://doi.org/10.1093/bioinformatics/bty827.
40. Costa TRD, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M, Waksman G. 2015. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nat Rev Microbiol 13:343–359. https://doi.org/10.1038/nrmicro3456.
41. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. Nucleic Acids Res 47:D427–D432. https://doi.org/10.1093/nar/gky995.
42. Kabsch W. 2010. XDS. Acta Crystallogr D Biol Crystallogr 66:125–132. https://doi.org/10.1107/S0907444909047337.
43. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66:213–221. https://doi.org/10.1107/S0907444909052925.
44. Cowtan K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr 62:1002–1011. https://doi.org/10.1107/S0907444906022116.
45. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66:486–501. https://doi.org/10.1107/S0907444910007493.
46. Smart OS, Womack TO, Flensburg C, Keller P, Paciorek W, Sharff A, Vonrhein C, Bricogne G. 2012. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. Acta Crystallogr D Biol Crystallogr 68:368–380. https://doi.org/10.1107/S0907444911056058.
47. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. https://doi.org/10.1002/jcc.20084.
48. Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ. 2007. EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol 157:38–46. https://doi.org/10.1016/j.jsb.2006.05.009.
49. Kimanius D, Forsberg BO, Scheres SH, Lindahl E. 2016. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. Elife 5:18722. https://doi.org/10.7554/eLife.18722.
50. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14:290–296. https://doi.org/10.1038/nmeth.4169.
51. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.



Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3

Expression constructs	Proteins	Primers	Ref.
1. pProExHTB-*bcs*^{His}*R*	Bcs^{His}R: MSYYHHHHHHDYDIPTTLEVLFQ-⋏-GPMGSM^{1}… ; 10.4 kDa (with tag)	s: CATGTAGGATCCATGAATAACAATGAACCAGATACTCTGCCTGATCCC (BamHI) as: CTAGATGCGGCCGCCTACTTTTGTTGCGCAAACTCTGCCAG (NotI)	This study
2. pProExHTB-*bcs*^{His}*Q*	Bcs^{His}Q: MSYYHHHHHHDYDIPTTLEVLFQ-⋏-GPMGSM^{1}… ; 31.3 kDa (with tag)	s: CATATGGGATCCATGGCCGTACTGGGATTGCAG (BamHI) as: CTTGATGCGGCCGCTCATGATTTACTCCCGACTGGCGTTTTCAGC (NotI)	This study
3. pProExHTB-*bcs*^{His}*RQ*	Bcs^{His}R: MSYYHHHHHHDYDIPTTLEVLFQ-⋏-GPMGSM^{1}… ; 10.4 kDa (with tag) BcsQ: wild-type, full-length, tag-free protein ; 27.9 kDa	s: CATGTAGGATCCATGAATAACAATGAACCAGATACTCTGCCTGATCCC (BamHI) as: CTTGATGCGGCCGCTCATGATTTACTCCCGACTGGCGTTTTCAGC (NotI)	This study
4. pET21b-*bcsQ*^{His}	BcsQ^{His}: M^{1} … TPVGSKS^{250}AAALEHHHHHH ; 29.2 kDa (with tag)	s: GTATACCATATGGCCGTACTGGGATTGCAGGG (NdeI) as: CTTGATGCCG GCC GCTGATTTACTCCCGACTGGCGTTTTCAGC (NotI)	This study
5. pET21b-*bcsRQ*^{His}	BcsR: wild-type, full-length protein, tag-free protein ; 7 kDa BcsQ^{His}: M^{1} … TPVGSKS^{250}AAALEHHHHHH ; 29.2 kDa (with tag)	s: GGATCCCATATGAATAACAATGAACCAGATACTCTGCC (NdeI) as: CTTGATGCCG GCC GCTGATTTACTCCCGACTGGCGTTTTCAGC (NotI)	This study
6. pET21b-*bcsRQ*$^{C39AD41A\text{-}His}$	same as BcsRQ^{His} ; point mutations C39AD41A in BcsQ: … V^{36}DAAPANLLRLSFN …	s: GGCCAACTTGTTGCGCCTGTCATTTAACGTTGATTTTAC as: GGGGCGGCATCGACCACCAGGACATTTTCTCC (inverse PCR on pET21b-*bcsRQ*^{His})	This study
7. pET21b-*bcsRQ*$^{C39AD41AL43D\text{-}His}$	same as BcsRQ^{His} ; point mutations C39AD41AL43D in BcsQ: … V^{36}DAAPANDLRLSFN …	s: CCAACGACTTGCGCCTGTCATTTAACGTTGATTTTACCC as: CCGGGGCGGCATCGACCACCAGGACATTTTCTCCCAACATTTG (inverse PCR as in 6.)	This study
8. pAM238-*bcsQ*	BcsQ: M^{1}…S^{250} ; 27.9 kDa	pAM238 vector amplification: s: GATTGGTGAAGCTTACTGGCCGTCGTTTTACAACG (HindIII) as: CCCTATGAATTCGGCGTAATCATGGTCATAGCTG (EcoRI) *bcsQ* cloning (with own ribosome-binding site): s: CATATGGAATTCAAGTAGGGGATTGGTGAATGGCCGTACTGGGATTGCAG (EcoRI) as: CTCGAGAAGCTTTCATGATTTACTCCCGACTGGCGTTTTCAGC (HindIII)	This study
9. pAM238-*bcsQ*$^{\Delta C10}$	Bcs$^{\Delta C10}$Q: M^{1}…S^{240} ; 27 kDa	Strategy 1. STOP codon insertion after S^{240} (inverse PCR on 8.): s: GGCTGAAAACGCCAGTCGGGAGTAAATCATGA as: CTCAGGAATAGTTCAACAGGCACCAGTTCGCCAG Strategy 2. STOP codon insertion and *bcsQ*^{C10} deletion (inverse PCR on 8.) as: GGAATAGTTCAACAGGCACCAGTTCGCC s: TGAAGCTTACTGGCCGTCGTTTTACAACGTC	This study
10. pProExHTB-*bcs*^{His}*E*^{FL}	Bcs^{His}E^{FL}: MSYYHHHHHHDYDIPTTLEVLFQ-⋏-GPMGSM^{1}… ; 62.8 kDa (with tag)	s: CATATGGGATCCATGAGGGACATTGTGGACCCTGTATTC (BamHI) as: CTCGAGGCGGCCGCTCATGATGAGCGCTCCACAGCATC (NotI)	This study
11. pProExHTB-*bcs*^{His}*E*^{FL}*F*	Bcs^{His}E^{FL}: same as in 10. BcsF: wild-type, full-length, untagged protein ; 7.4 kDa	s: CATATGGGATCCATGAGGGACATTGTGGACCCTGTATTC (BamHI) as: CTCGAGGCGGCCGCTCATTTTTTGGTTGCCCTGGCTTTTTCCGTG (NotI)	This study
12. pET-^{His}SUMO-*bcsE*^{FL}	MGSSHHHHHH-SUMOtag-GG-⋏-SM^{1}… ; 73 kDa (with tag), 59.5 kDa after cleavage	s: CATATGGGATCCATGAGGGACATTGTGGACCCTGTATTC (BamHI) as: CTCGAGGCGGCCGCTCATGATGAGCGCTCCACAGCATC (NotI)	This study
13. pRSFDuet1*-*bcsE*^{FL}	MGSM^{1} … ; full-length, untagged protein, 59.7 kDa	Vector MCS1 redesign (His-tag removal, redesign of BamHI/NotI sites):: s: CTCGAGGCGGCCGCATAATGCTTAAGTCGAACAGA (NotI) as: CATATGGGATCCCATGGTATATCTCCTTATTAAAG (BamHI) *bcsE*^{FL} cloning: s: CATATGGGATCCATGAGGGACATTGTGGACCCTGTATTC (BamHI) as: CTCGAGGCGGCCGCTCATGATGAGCGCTCCACAGCATC (NotI)	This study
14. pRSFDuet1*-*bcsE*$^{1\text{-}217}$	MGSM^{1}… A^{217} ; untagged BcsE^{NTD}, 25.1 kDa	Vector MCS1 redesign: as in 13. *bcsE*$^{1\text{-}217}$ cloning: s: CATATGGGATCCATGAGGGACATTGTGGACCCTGTATTC (BamHI) as: CTCGAGGCGGCCGCCTACGCCTCTTCGCTTTGAACTAATGTC (NotI)	This study
15. pRSFDuet1*-*bcsE*$^{217\text{-}523}$	MGSA^{217}… ; untagged BcsE$^{REC^{*}\text{-}GGDEF^{*}}$, 35 kDa	Vector MCS1 redesign: as in 13. *bcsE*$^{217\text{-}523}$ cloning: s: ACCATGGGATCCGCGGAGATCCAACCACGCAGC (BamHI) as: CTCGAGGCGGCCGCCTACGCCTCTTCGCTTTGAACTAATGTC (NotI)	This study
16. pRSFDuet1*-*bcsE*$^{217\text{-}523}$-*D*^{415}*TGA*	same as 14. ; I-site targeted mutagenesis R^{415}TGD→D^{415}TGA	s: GGCGCTATCATGACCATTGGCGGTAATC as: GGTATCGTTAGGGCGACACAGCGTCAG (inverse PCR)	This study
17. pRSFDuet1*-*bcsE*$^{217\text{-}340}$	MGSA^{217}… G^{340} ; untagged, 14.1 kDa	Vector MCS1 redesign: as in 13. *bcsE*$^{217\text{-}340}$ cloning: s: ACCATGGGATCCGCGGAGATCCAACCACGCAGC (BamHI) as: CTCGAGGCGGCCGCCTACCCTTGCACGCTTTCGATCATCGTC (NotI)	This study
18. pRSFDuet1*-*bcsE*$^{217\text{-}347}$	MGSA^{217}… V^{347}; untagged, 15 kDa	Vector MCS1 redesign: as in 13. *bcsE*$^{217\text{-}347}$ cloning: s: ACCATGGGATCCGCGGAGATCCAACCACGCAGC (BamHI) as: CTCGAGGCGGCCGCCTACACATAGCGACTAAACTTCTGCCCTTGC (NotI)	This study
19. pRSFDuet1*-*bcsE*$^{217\text{-}359}$	MGSA^{217}… V^{359}; untagged, 16.3 kDa	Vector MCS1 redesign: as in 13. *bcsE*$^{217\text{-}359}$ cloning: s: ACCATGGGATCCGCGGAGATCCAACCACGCAGC (BamHI) as: CTCGAGGCGGCCGCCTACTGGGTCATTGACAGCAAGGTAGTGATATC (NotI)	This study
20. pRSFDuet1*-*bcsE*$^{349\text{-}523}$	MGSE^{349}… ; untagged BcsE$^{GGDEF^{*}}$, 20.1 kDa	Vector MCS1 redesign: as in 13. *bcsE*$^{349\text{-}523}$ cloning: s: ACCATGGGATCCGAAGATATCACTACCTTGCTGTCAATGACCCAG (BamHI) as: CTCGAGGCGGCCGCTCATGATGAGCGCTCCACAGCATC (NotI)	This study
21. pRSFDuet1*-*bcsE*$^{359\text{-}523}$	MGSQ^{359}…; untagged, 19 kDa	Vector MCS1 redesign: as in 13. *bcsE*$^{359\text{-}523}$ cloning: s: ACCATGGGATCCCAGCCGCTCAAACTGCGTGGTTTC (BamHI) as: CTCGAGGCGGCCGCTCATGATGAGCGCTCCACAGCATC (NotI)	This study

22. pET-^{His}SUMO-*bcsE*$^{217-523}$	MGSSHHHHHH-SUMOtag-GG-λ←-SA^{217}… ; 48 kDa (with tag), 34.7 after cleavage	s: ACCATGGGATCCGCGGAGATCCAACCACGCAGC (BamHI) as: CTCGAGGCCGGCCGCCTACGCCTCTTCGCTTTGAACTAATGTC (NotI)	This study
23. pET-^{His}SUMO-*bcsE*$^{217-523}$-*D*^{415}*TGA*	same as 21. ; I-site targeted mutagenesis R^{415}TGD→D^{415}TGA	same as in 16.	This study
24. pCDFDuet1-*bcs*^{His}*RQA*$^{HA-FLAG}$*B*	Bcs^{His}R: MGSSHHHHHHHAAGSN^{2}NNE… ; 8.6 kDa (with tag) BcsQ: wild-type, full-length, untagged BcsQ ; 27.9 kDa BcsA$^{HA-FLAG}$: … Q^{872}GSARSSGRTGLEFEEFYPYDVPDYAADYKDDDDKRS ; 103.9 kDa (with tags) BcsB: wild-type, full-length, untagged BcsB ; 86.1 kDa	Vector MCS1 redesign for PstI/NotI restriction/ligation cloning s: GCGCAACAAAAGTAGGCGGCCGCATAATGCTTAAGTCGAACAGA (NotI) as: TCGATCCTGCAGCGTGATGGTGGTGATGATGGTGATGGCTGCTGCCCATG (PstI) Template DNA: *E. coli* 1094 *bcsA*$^{HA-FLAG}$ genomic DNA *bcs*^{His}*RQA*$^{HA-FLAG}$*B* cloning: s: CATCACGCTGCAGGATCGAATAACAATGAACCAGATACTC (PstI) as: CGACTTAAGCATTATGCGGCCGCTTACTCGTTATCCGGGTTAAGACG (NotI)	6
25. pCDFDuet1-*bcs*^{His}*R*^{HA}*QA*$^{HA-FLAG}$*B*	Bcs^{His}R, BcsA$^{HA-FLAG}$, BcsB: as in 24. Bcs^{HA}Q: M^{1}YPYDVPDYAGSGAGSGTGA^{2}VL… ; 29.6 kDa (with tag)	s: GTAGCGGTGCCGGTAGCGGTACAGGTGCCGTACTGGGATTGCAGG as: CTGCATAATCCGGAACATCATACGGATACATTCACCAATCCCCTACTTTTGTTGC (inverse PCR on 24.)	This study
26. pRSFDuet1*-*^{STREP}*E*^{FL}*FG*	Bcs^{STREP}E^{FL}: MASWSHPQFEKGSM^{1}… BcsF: wild-type, full-length, untagged protein ; 7.4 kDa BcsG: wild-type, full-length, untagged protein ; 59.6 kDa	*bcsE*^{FL}*FG* cloning in pRSFDuet1* (vector redesign as in 13.) s: CATATGGGATCCATGAGGGACATTGTGGACCCTGTATTC (BamHI) as: CTCGAGGCCGGCCGCCTACTGCATTTGAGTTCTCGGAGAC (NotI) STREP II tag insertion (inverse PCR): s: CCGCAGTTCGAAAAAGGATCCATGAGGGACATTGTGGACCCTGTATTC as: GTGGCTCCAGCTAGCCATGGTATATCTCCTTATTAAAG	6
27. pRSFDuet1*-*^{STREP}*E*$^{217-523}$*FG*	Bcs^{STREP}E$^{217-523}$: MASWSHPQFEKGSA^{217}EIQPRS… ; 36.1 kDa	Insert replacement in 26. s: ACCATGGGATCCGCGGAGATCCAACCACGCAGC (BamHI) as: CTCGAGGCCGGCCGCCTACTGCATTTGAGTTCTCGGAGAC (NotI)	This study
28. pRSFDuet1*-*^{STREP}*E*$^{1-217}$*FG*	Bcs^{STREP}E$^{1-217}$: MASWSHPQFEKGSM^{1}…A^{217} ; 26.3 kDa	Strategy 1. 4-letter STOP codon insertion after A^{217} (inverse PCR on 26.): s: ATGCGGAGATCCAACCACGCAGCGACG as: TAAGCCTCTTCGCTTTGAACTAATGTCCAGATACCATTTTGTTGC Strategy 2. STOP codon insertion and *BcsE*$^{REC'-GGDEF'}$ deletion (inverse PCR on 26.): s: ATGAATACCAGAACCCATGCGACTGTTAGATGATGCTG as: TAAGCCTCTTCGCTTTGAACTAATGTCCAGATACCATTTTGTTGC	This study
29. pKT25-*zip*	T25-GCN4^{ZIP} (…IQRMKQLEDKVEELLSKNYHLENEVARLKKLVGER…)	Vector control used as commercially available (Euromedex)	32
30. pKT18C-*zip*	T18-GCN4^{ZIP} (…IQRMKQLEDKVEELLSKNYHLENEVARLKKLVGER…)	Vector control used as commercially available (Euromedex)	32
31. pKT25-*bcsE*$^{1-217}$	T25-GSTLEGS-M^{1}… A^{217} ; 49.9 kDa	Vector MCS redesign: s: ACTATGGGTACCCTAAGAATTCGGCCGTCGTTTTACAACGTCG (KpnI) as: CATATGGGATCCCTCTAGAGTCGACCCTGCAGCC (BamHI) *bcsE*$^{1-217}$ cloning: s: CATATGGGATCCATGAGGGACATTGTGGACCCTGTATTC (BamHI) as: ACTATGGGTACCATTACGCCTCTTCGCTTTGAACTAATGTCCAGATACC (KpnI)	This study
32. pKNT25-*bcsE*$^{1-217}$	MGSM^{1}… A^{217}-GTSNSMT-T25 ; 50.3 kDa	Vector MCS redesign: s: ACTATGGGTACCTCGAATTCAATGACCATGCAGCAATCGCATC (KpnI) as: CATATGGGATCCCATAGCTGTTTCCTGTGTGAAATTGTTATCC (BamHI) *bcsE*$^{1-217}$ *cloning:* s: CATATGGGATCCATGAGGGACATTGTGGACCCTGTATTC (BamHI) as: ACTATGGGTACCCGCCTCTTCGCTTTGAACTAATGTCCAGATACC (KpnI)	This study
33. pUT18-*bcsE*$^{1-217}$	MGSM^{1}… A^{217}-GTSNS-T18 ; 44.9 kDa	Vector MCS redesign: s: ACTATGGGTACCTCGAATTCAGCCGCCAGCGAGG (KpnI) as: CATATGGGATCCCATAGCTGTTTCCTGTGTGAAATTGTTATCC (BamHI) *bcsE*$^{1-217}$ *cloning:* same as 32.	This study
34. pUT18C-*bcsE*$^{1-217}$	T18-HCRSTGS-M^{1}… ; 45.4 kDa	Vector MCS redesign: s: ACTATGGGTACCGTGCACTCTCAGTACAATCTGCTCTGATGC (KpnI) as: CATATGGGATCCAGTCGACCTGCAGTGGCGTTCC (BamHI) *bcsE*$^{1-217}$ cloning: same as 31.	This study
35. pKT25-*bcsE*^{FL}	T25-GSTLEGS-M^{1}… ; 84.4 kDa	Vector MCS redesign: as in 31. *bcsE*^{FL} cloning: s: CATATGGGATCCATGAGGGACATTGTGGACCCTGTATTC (BamHI) as: ACTATGGGTACCATTATGATGAGCGCTCCACAGCATCATCTAACAGTC (KpnI)	This study
36. pKNT25-*bcsE*^{FL}	MGSM^{1}… S^{523}-GTSNSMT-T25 ; 84.8 kDa	Vector MCS redesign: as in 32. *bcsE*^{FL} cloning: s: CATATGGGATCCATGAGGGACATTGTGGACCCTGTATTC (BamHI) as: ACTATGGGTACCTGATGAGCGCTCCACAGCATCATCTAACAGTC (KpnI)	This study
37. pUT18-*bcsE*^{FL}	MGSM^{1}… S^{523}-GTSNS-T18 ; 79.4 kDa	Vector MCS redesign: as in 33. *bcsE*^{FL} cloning: same as 36.	This study
38. pUT18C-*bcsE*^{FL}	T18-HCRSTGS-M^{1}… ; 80 kDa	Vector MCS redesign: as in 34. *bcsE*^{FL} cloning: same as 35.	This study
39. pKT25-*bcsE*$^{217-523}$	T25-GSTLEGS-A^{217}… ; 59.7 kDa	Vector MCS redesign: as in 31. *bcsE*$^{217-523}$ cloning: s: ACCATGGGATCCGCGGAGATCCAACCACGCAGC (BamHI) as: ACTATGGGTACCATTATGATGAGCGCTCCACAGCATCATCTAACAGTC (KpnI)	This study
40. pUT18C-*bcsE*$^{217-523}$	T18-HCRSTGS-A^{217}… ; 55.2 kDa	Vector MCS redesign: as in 34. *bcsE*$^{217-523}$ cloning: same as 39.	This study
41. pKT25-*s10*	T25-GSTLEGS-M^{1}… ; 36.7 kDa	Vector MCS redesign: as in 31. s: CATATGGGATCCCAGAACCAAAGAATCCGTATCCGCCTGAAAG (BamHI) as: ACTATGGGTACCATTAACCCAGGCTGATCTGCACGTCTAC (KpnI)	This study

42. pUT18C-*s10*	T18-HCRSTGS-Q^{2}...	same as in 34. (vector) and 41. (insert)	This study
43. pKT25-*nusB*	T25-GSTLEGS-M^{1}... ; 40.6 kDa	Vector MCS redesign: as in 31. s: CATGTAGGATCCATGAAACCTGCTGCTCGTCGCC (BamHI) as: ACT GTA GGT ACC ATCA CTT TTT GTT AGG GCG AAT CAC AGG TGC (KpnI)	This study
44. pRSFDuet1*-*s10*	S10: M^{1}GSQ^{2}... ; 12 kDa	Vector MCS1 redesign: as in 13. *s10* cloning in site 1: s: CATATGGGATCCCAGAACCAAAGAATCCGTATCCGCCTGAAAG (BamHI) as: CTATAGGCGGCCGCTTAACCCAGGCTGATCTGCACGTCTAC (NotI)	This study
45. pRSFDuet1*-*s10*$^{(site1)}$-*nusB*$^{(site2)}$	S10: as in 44. NusB: MKLM^{1}... ; 16 kDa	Vector MCS1 redesign: as in 13. *s10* cloning in site 1: as in 44. Vector MCS2 redesign: s: GAATCCCTCGAGTCTGGTAAAGAAACCGCTGCTG (XhoI) as: GGATCCAAGCTTCATATGTATATCTCCTTCTTATAC (HindIII) *nusB* cloning in site 2: s: GGATCCAAGCTTATGAAACCTGCTGCTCGTCGCC (HindIII) as: GGATCCCTCGAGTCACTTTTTGTTAGGGCGAATCACAGGTGC (XhoI)	This study
46. pProExHTB-*nus*^{His}*B*	Nus^{His}B: MSYYHHHHHHDYDIPTTLEVLFQ-✂-GPMGSM^{1}... ; 19 kDa (with tag)	s: CATGTAGGATCCATGAAACCTGCTGCTCGTCGCC (BamHI) as: ACT GTA GGT ACC ATCA CTT TTT GTT AGG GCG AAT CAC AGG TGC (KpnI)	This study
47. pRSFDuet1*-*bcsE*$^{1\text{-}217(site1)}$-*s10*$^{(site2)}$	BcsE$^{1\text{-}217}$: as in 14. S10: M^{1}KLQ^{2}... ; 12 kDa	Vector MCS1 and MCS2 redesign: as in 45. *bcsE*$^{1\text{-}217}$ cloning in site 1: as in 14. *s10* cloning in site 2: s: GGATCCAAGCTTCAGAACCAAAGAATCCGTATCCGCCTGAAAG (HindIII) as: GAATCCCTCGAGTTAACCCAGGCTGATCTGCACGTCTAC (XhoI)	This study
48. pProExHTB-*min*^{His}*DE*	Min^{His}D: MSYYHHHHHHDYDIPTTLEVLFQ-✂-GPMGSM^{1}... MinE: full-length protein; did not co-purify with Min^{His}D from the cytosolic fraction	s: CATATGGGATCCATGGCACGCATTATTGTTGTTACTTCGGGC (BamHI) as: CTAGAAGCGGCCGCTTATTTCAGCTCTTCTGCTTCCGGTAAGGTCAC. (NotI)	This study

Crystallographic data collection and refinement statistics	
	BcsE$^{217-523}$
Crystallized protein	SeMet
Data Collection	
Space group	P 41 21 2
Cell dimensions	
a, b, c (Å)	112.5, 112.5, 106.4
α, β, γ (deg)	90, 90, 90
Wavelength	0.9791
Resolution (Å)	49.7 – 2.2 (2.28 – 2.2)
R-merge	21.1% (257.6%)
R-meas	21.7% (264%)
R-pim	5.2% (63.3%)
Mean I/σ(I)	12.61 (1.1)
Completeness (%)	99.3% (96.8%)
Multiplicity	17.3
$CC_{1/2}$	99.9 (65.2)
Refinement	
Unique reflections	35 074
R-work	20.5%
R-free	23.8%
Number of non-hydrogen atoms	
Proteins	4 091
Ligands	58
B-factors	
Proteins	60.48
Ligands	50.51
R.m.s.d.	
Bond lengths (Å)	0.013
Bond angles (deg)	1.67
Ramachandran plot	
Favored (%)	96.9
Allowed (%)	2.9
Molprobity score	1.51
Crystallization condition ligand	c-di-GMP
Ligand in structure **	c-di-GMP
Protein : c-di-GMP ratio	2 : 1

** Statistics for the highest-resolution shell are shown in parentheses*

4.2 Article 2

4.2.1 Introduction to article 2

The article 2 is a continuation to article 1 in further deciphering the structure and assembly of the assembled Bcs macrocomplex. Moreover, we solved multiple crystallographic snapshots of the regulatory Bcs subcomplexes. Combining structural and functional investigations, we uncover the mechanism of asymmetric Bcs secretion system assembly and periplasmic crown polymerization and reveal unexpected subunit stoichiometry, multisite c-di-GMP recognition, and ATP-dependent regulation.

The work was a close collaboration with Dr. Samira Zouhir, who focused on the cryo-EM studies of the assembled Bcs macrocomplex and the final stages of structural model refinements, whereas I concentrated on the stabilization and crystallization of the BcsRQ and BcsE*RQ multicomponent complexes, the characterization of newly discovered c-di-GMP binding motifs in the BcsE regulator, as well as the validation by cryo-EM of the mechanism of periplasmic crown polymerization via intersubunit structural complementation between neighboring BcsB subunits. Together with Dr. Petya Krasteva, we also conducted all functional assays of cellulose secretion, ligand complexation, ATP-ase activity, and biochemical detection of Bcs subunit expression. This work also benefited from contributions and useful discussions with former colleagues Stéphane Roche and Meryem Caleechurn.

STRUCTURAL BIOLOGY

Architecture and regulation of an enterobacterial cellulose secretion system

Wiem Abidi[1,2,3]*, Samira Zouhir[1,2,3]*, Meryem Caleechurn[1], Stéphane Roche[1], Petya Violinova Krasteva[1,2,3]†

Many free-living and pathogenic enterobacteria secrete biofilm-promoting cellulose using a multicomponent, envelope-embedded Bcs secretion system under the control of intracellular second messenger c-di-GMP. The molecular understanding of system assembly and cellulose secretion has been largely limited to the crystallographic studies of a distantly homologous BcsAB synthase tandem and a low-resolution reconstruction of an assembled macrocomplex that encompasses most of the inner membrane and cytosolic subunits and features an atypical layered architecture. Here, we present cryo-EM structures of the assembled Bcs macrocomplex, as well as multiple crystallographic snapshots of regulatory Bcs subcomplexes. The structural and functional data uncover the mechanism of asymmetric secretion system assembly and periplasmic crown polymerization and reveal unexpected subunit stoichiometry, multisite c-di-GMP recognition, and ATP-dependent regulation.

INTRODUCTION

Bacterial cellulose is a widespread extracellular matrix component that can modulate microbial fitness and virulence in both environmental settings and infected hosts (*1*–*3*). Whereas its biosynthesis typically involves direct c-di-GMP (cyclic diguanosine monophosphate or CDG) control of an inner membrane (IM) BcsAB synthase duo across the bacterial domain of life, in many β- and γ-proteobacteria it is carried out by sophisticated *Escherichia coli*–like Bcs secretion systems, where multiple accessory subunits are essential for secretion or contribute to the optimal production of the polysaccharide in vivo (*1*, *4*).

In particular, the *E. coli* cellulose secretion machinery comprises a total of nine subunits (BcsRQABZCEFG) that span from the cytosol to the extracellular space (Fig. 1A) (*1*). Studies on a BcsAB complex from *Rhodobacter sphaeroides* have shown that processive glucose polymerization is likely carried out by a glycosyl transferase domain of BcsA, whose active site is made accessible by recurrent binding of dimeric c-di-GMP to an adjacent PilZ domain and displacement of a gating loop capping the substrate-binding pocket (*5*–*7*). IM transport is coupled to processive polymerization and is energized by the high-energy phosphoanhydride bonds of the preactivated synthase substrate, uridine diphosphate (UDP)–glucose. The nascent polysaccharide chain is thus extruded, one molecule at a time, through the IM transport domain of BcsA, completed by the C-terminal tail-anchor of partner BcsB (*8*). In the periplasm, the latter features a donut-shaped architecture and is proposed to lead the outcoming cellulose toward the outer membrane secretory component BcsC (*6*). In addition, the periplasmic modules of BcsG and BcsZ can introduce post-synthetic, covalent modifications in the nascent polysaccharide through the addition of phosphoethanolamine residues or limited hydrolysis, respectively (*1*, *9*). Last, *E. coli*–like cellulose secretion in vivo is absolutely dependent on the presence of two small cytosolic proteins, BcsR and BcsQ, and is substantially enhanced by a third cytosolic protein, BcsE, as well as by a short membrane–embedded polypeptide, BcsF (*4*).

We recently uncovered complex interdependence among these regulatory subunits and showed that most of the IM and cytosolic Bcs components (BcsRQABEF) form a megadalton-sized secretory macrocomplex with a multimeric, layered, and asymmetric architecture (Fig. 1A) (*4*, *10*). Nevertheless, the structures, localization, and functional roles of individual building subunits within the secretory assembly remained largely unresolved. Here, we provide near-atomic resolution cryo–electron microscopy (cryo-EM) data that reveal noncanonical synthase tandem assembly and β-sheet complementation–driven BcsB polymerization in the periplasm that can explain both the unusual secretion system asymmetry and its targeting to cell pole in vivo (*4*, *11*). We further present the crystal structures of several multicomponent regulatory subcomplexes, which suggest additional regulatory inputs via BcsRQ-dependent adenosine triphosphate (ATP) complexation, multisite c-di-GMP recognition, and creation of a synthase-proximal c-di-GMP pool for processive enzyme activation.

RESULTS

Structure, polymerization, and asymmetry of the periplasmic crown

To gain further insights into the architecture of the purified Bcs macrocomplex, we resorted to single-particle cryo-EM (Fig. 1, B and C, and fig. S1). The three-dimensional (3D) electron density reconstruction was characterized by substantial gradient of estimated local resolution, with lower values for the transmembrane and cytosolic regions, likely due to the presence of a detergent micelle, partial complex dissociation, and conformational variability among single particles. In contrast, the periplasmic domains were resolved at nearly atomic resolution (average resolution of 2.9 Å without imposed symmetry; fig. S1), with most of the particles featuring up to six BcsB copies in the crown. As the averaged electron density corresponding to the sixth BcsB protomer presented lower intensity and resolution, a pentameric BcsB assembly was refined against the experimental density (table S1). The data confirm that the crown is composed exclusively of multiple BcsB copies and reveal that the protein polymerizes in a

[1]Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, 91198 Gif- sur-Yvette, France. [2]"Structural Biology of Biofilms" Group, European Institute of Chemistry and Biology (IECB), 33600 Pessac, France. [3]CBMN UMR 5248 CNRS, University of Bordeaux, 33600 Pessac, France.
*These authors contributed equally to this work.
†Corresponding author. Email: pv.krasteva@iecb.u-bordeaux.fr

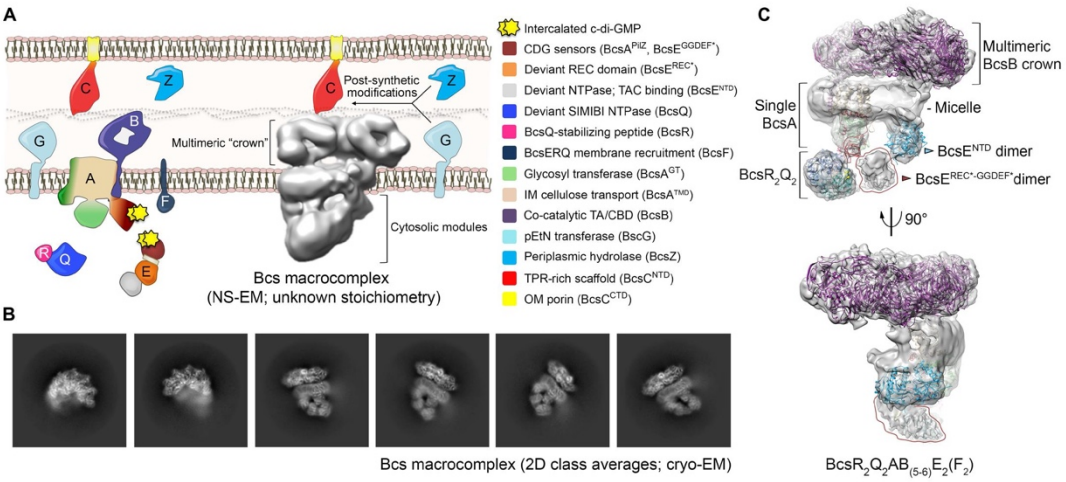


Fig. 1. The *E. coli*–like cellulose secretion system and architecture of the Bcs macrocomplex. (**A**) Left: Topology and functional roles of Bcs subunits (NS-EM, negative-stain electron microscopy; CDG, c-di-GMP; REC, receiver; NTPase, nucleoside triphosphatase; TAC, transcription antitermination complex; NTD, N-terminal domain; SIMIBI, signal recognition particle, MinD, and BioD; IM, inner membrane; TMD, transmembrane domain; TA, tail-anchor; CBD, carbohydrate-binding domains; pEtN, phosphoethanolamine; TPR, tetratricopeptide repeats; OM, outer membrane; CTD, C-terminal domain). Right: Structure of an assembled Bcs macrocomplex encompassing most IM and cytosolic subunits as obtained by single-particle NS-EM (*1*). (**B**) Cryo-EM 2D class averages of detergent-extracted and affinity-purified Bcs macrocomplex. (**C**) Density assignments within the assembled Bcs macrocomplex. A hexameric BcsB crown derived from the locally refined $BcsB^{peri}$ pentamer (see below) was fitted in the periplasmic densities. A model of a CDG-bound BcsA was generated by homology modeling in Robetta and was then fitted and refined against the densities corresponding to the BcsRQAB subcomplex, themselves improved by local refinement. The crystal structure of a sandwich $BcsR_2Q_2$ dimer was fitted in the apical densities, with BcsR adopting an extended four-turn α1 helix, as observed in some of the crystallized states (see below). A Robetta-derived model of dimeric $BcsE^{NTD}$ was fitted in the membrane-proximal densities opposite BcsA. The $BcsE^{NTD}$ copies are fitted in head-to-tail orientation consistent with recently reported interaction data (*10*). Regions corresponding to $BcsE^{Rec*\text{-}GGDEF*}$ did not yield interpretable densities even after local refinement, likely due to conformational heterogeneity (see below). Most BcsB tail-anchors as well as the BcsE-interacting BcsF chains (predicted to fold into a single transmembrane helix each) were not visible in the 3D-reconstructed micelle.

helical fashion to introduce asymmetry to the system rather than axial symmetry that is typical for multicopy secretion system assemblies (fig. S1) (*12*).

Each BcsB subunit adopts a conserved four-domain fold (D1 to D4) in the periplasm ($BcsB^{peri}$) composed of two jellyrolls and two carbohydrate-binding domains, alternating from the N-terminus toward a C-proximal amphipathic helix and an IM tail-anchor (TA) (Fig. 2A and fig. S2). The latter is visible and well refined only in the membrane-proximal BcsB protomer interacting with the BcsA synthase (see below); however, the lack of biochemically detectable BcsB cleavage and the presence of a large detergent micelle underneath the entire crown indicate that all crown BcsB subunits likely feature preserved IM tails. Although the sequence identity with the *R. sphaeroides* homolog is overall low, the four domains adopt similar secondary and tertiary structures with several crucial exceptions (Fig. 2A and fig. S2, A and B). First, D2 of $BcsB^{E.coli}$ features a C-proximal β-strand insertion (*E. coli* residues S^{343}-P^{351}) that extends the central β-sheet of the domain, whereas D4 lacks an amphipathic helix found in the region connecting strands β2 and β3 of the domain's central β-sheet in the $BcsB^{R.sphaeroides}$ homolog (*R. sphaeroides* residues W^{570}-R^{594}). In contrast, surrounding D4 residues in the *E. coli* protein adopt a three-stranded β-sheet fold, which is complemented by the D2 β-strand insertion of the neighboring subunit to form a continuous nine-stranded shared β-sheet and provide a secondary structure-dependent mechanism for periplasmic BcsB polymerization that is likely conserved among enterobacteria (Fig. 2, A and B, and fig. S2). Second, D3 of $BcsB^{E.coli}$ features an extended β-strand hairpin in one of the β-sheet–connecting loops in the jellyroll. Within the macrocomplex, D3 hairpins point toward the center of the BcsB crown, where they are further stabilized by intersubunit β-strand contacts and pack tightly in a helical staircase-like fashion (figs. S1 and S2). In the *R. sphaeroides* BcsAB crystal structures, the newly synthesized cellulose exiting BcsA's IM transport domain positions along the membrane-proximal surface of BcsB's D3 module (Fig. 2A). Structural superposition with the cryo-EM structure presented here indicates that in *E. coli* the secreted cellulose is likely extruded toward the BcsB crown lumen and away from the peripheral D2 and D4 carbohydrate-binding domains. Together, these data suggest that D2 and D4 play a role in interactions with the periplasmic peptidoglycan rather than the secreted polysaccharide, whereas the stacked D3 luminal loops might provide a ratchet-like structural support for the secreted homopolymer on its way toward the outer membrane.

A modeled $BcsB^{peri}$ superhelix features about nine protomers per turn, with the 10th BcsB copy showing a 91-Å vertical displacement relative to the first or almost twice the IM thickness (fig. S1D). Given that BcsB is a tail-anchored protein, its helical polymerization would induce substantial deformation of the IM, which by itself is likely to act as a polymerization-limiting factor. Non-aggregative cross-linking analyses (fig. S1, E and F) (*13*) and the cryo-EM structure of purified,

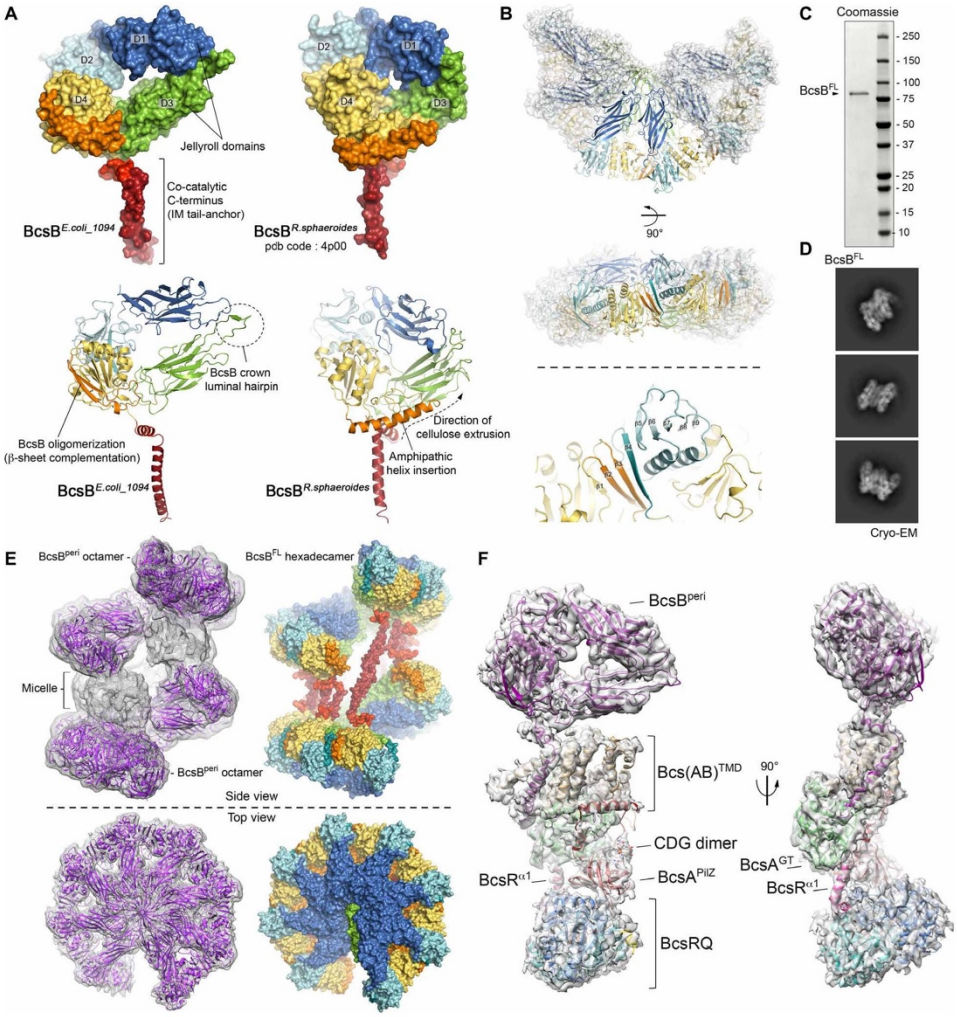


Fig. 2. Structure, polymerization, and asymmetry of the periplasmic crown and noncanonical BcsAB synthase tandem assembly. (**A**) Structural comparison of $BcsB^{FL}$ monomers from *E. coli* (left) and *R. sphaeroides* (right) shown in surface (top) and cartoon (bottom) representation in PyMOL. pdb, Protein Data Bank. (**B**) Top: A model of a hexameric BcsB crown derived from the refined $BcsB^{peri}$ pentamer in top and side views. Bottom: A close-up view of the nine-stranded β-sheet formed by β-sheet complementation between adjacent BcsB subunits. (**C**) Coomassie-stained SDS-PAGE of purified, detergent-extracted, full-length *E. coli* $BcsB^{His}$. (**D**) Cryo-EM 2D class averages of full-length $BcsB^{His}$. (**E**) Left: Cryo-EM density reconstruction of purified full-length $BcsB^{His}$ with density-fitted $BcsB^{peri}$ octamers. Right: A composite atomic model of the $BcsB^{FL}$ hexadecamer with visualized IM tail-anchors as observed in the BcsRQAB assembly. Color coding is as in (A). (**F**) Structure of the BcsRQAB assembly following particle subtraction and local refinement of the cryo–electron density. A single $BcsB^{peri}$ copy was fitted in the periplasmic region, and the amphipathic and transmembrane C-terminal helices were built in the refined density. A CDG-bound BcsA homology model was fitted in the transmembrane and membrane-proximal regions, and the polypeptide backbone was refined against the experimental density. A $BcsR_2Q_2$ crystal structure was fitted in the apical region as in Fig. 1. Separate structural modules and ligand molecules are indicated.

full-length $BcsB^{His}$ (Fig. 2, C to E) show that BcsB polymerization is both self-driven and self-limiting. The 3D reconstruction features two BcsB octamers embedded from opposite sites in a helical belt of detergent micelles (Fig. 2E). Whereas the $BcsB^{peri}$ regions refined to ~4.4-Å resolution following local refinement, the C-terminal BcsB anchors remained largely unresolved in the underlying detergent micelles. Nevertheless, in each of the two BcsB octamers, the periplasmic regions of the first BcsB copy contact the detergent micelle of the respective eight protomers and thus create a steric hindrance for further polymerization (Fig. 2E). Although the isolated $BcsB^{His}$ particles feature different transmembrane region content from the multicomponent Bcs macrocomplex, the intersubunit BcsB contacts remain virtually unchanged between the two assemblies (fig. S1G) and thus reinforce a model of superhelical, asymmetric crown polymerization driven by periplasmic β-sheet complementation. It is important to note that purified Bcs macrocomplex particles typically feature up to six BcsB copies in the crown, likely due to additional restrains on local membrane curvature and BcsB oligomerization potential imposed by interacting IM and cytosolic Bcs partners.

Noncanonical BcsAB synthase tandem assembly

BcsA and BcsB are often encoded by neighboring genes of the same operon and have been shown to engage in 1:1 interactions in both *Gluconacetobacter xylinus* and *R. sphaeroides* (*1*, *6*, *14*). Biochemical data have further shown that the C-terminal BcsB region, composed of an amphipathic and an IM α-helices, is absolutely necessary for glucose polymerization by both the *E. coli* and *R. sphaeroides* synthases in vitro (*8*). Last, structural studies on the $BcsAB^{R.sphaeroides}$ tandem have shown that the BcsB C-terminal anchor tightly associates with and completes the IM transport domain of BcsA, further supporting a structural and functional co-dependency between the subunits (*5*–*7*). On the basis of these data, we proposed earlier a similar 1:1 association between the co-catalytic partners in the assembled Bcs macrocomplex, but we could not distinguish between protein and detergent micelle densities in the negative-stain data (*4*). Unexpectedly, the cryo-EM structure reveals a single BcsA copy in the assembled Bcs macrocomplex (Figs. 1C and 2F), which forms a conserved BcsAB tandem with the membrane-proximal BcsB protomer, as determined by the relative position of the D3 luminal loops in the periplasm. On the other side of the membrane, BcsA folds, as expected, into a glycosyl transferase domain and a PilZ β-barrel module. A homology model could be fitted with minor adjustments into the electron density map [root mean square deviation (RMSD) of 1.6 Å over the Cα atoms between the Robetta (*15*) output and the density-refined model] and reveals extra density consistent with c-di-GMP binding to its canonical binding site onto the C-terminal PilZ module (Figs. 1C and 2F). Although these data suggest overall structural and functional conservation between the *E. coli* and *R. sphaeroides* BcsA homologs, the unexpected BcsA:BcsB stoichiometry, the proximity of a second c-di-GMP sensing protein (BcsE), and the direct interactions of BcsA with the essential for secretion BcsR and BcsQ subunits (see below) suggest that *E. coli* cellulose secretion is dependent on multiple additional regulatory inputs rather than a simple model of BcsAB tandem assembly.

Structural and mechanistic insights into the essential for secretion BcsRQ complex

BcsR is a short, 7-kDa polypeptide with unknown structure and function, whereas BcsQ is a 28-kDa protein predicted to belong to the ancient SIMIBI (signal recognition particle, MinD, and BioD) superfamily of pro- and eukaryotic nucleoside triphosphatases (NTPases). Members of the latter are key to a large variety of cellular processes including divisome positioning (MinD), bacterial flagellum secretion (FlhG and FlhF), and membrane protein sorting in both prokaryotes and higher organisms (SRP54-SR and Get3) (fig. S3) (*16*, *17*). We recently showed that BcsR and BcsQ exhibit chaperone-like function toward each other, where BcsR stabilizes BcsQ to form monodisperse $BcsR_2Q_2$ heterotetramers in solution, while BcsQ itself may play a role in the folding and subsequent stability of BcsR (*10*).

To increase the resolution of the apical BcsRQ subcomplex, we proceeded to protein crystallization. The BcsRQ complex crystallized in a variety of conditions and space groups, and several structures were solved at nearly atomic resolutions (1.6 to 2.1 Å; table S2), with only the ~10 C-terminal BcsQ amino acids remaining unresolved in the electron densities. The refined structural models feature a virtually unchanged conformation for BcsQ (RMSD of 0.19 to 0.42 Å over all atoms), whereas BcsR shows substantial flexibility in its N-terminal domain and the N-proximal end of helix α1.

BcsQ adopts a classical α-β-α SIMIBI NTPase fold with a central seven-stranded β-sheet (a ↑β7•↑β6•↑β1•↑β5•↑β2•↑β4•↓β3 core) sandwiched between flanking α helices (Fig. 3A and fig. S3A) (*18*). Regardless of the purification and crystallization conditions, BcsRQ crystallized as ATP-bound "sandwich" or "tango" dimers of heterodimers. In the latter, an extensive part of the BcsQ homodimerization interface is stabilized by a network of hydrogen bonds and salt bridges among residues from both BcsQ subunits, the "sandwiched" ATP moieties and multiple water molecules resolved in the electron density (Fig. 3, B and C). In particular, the adenine base is stabilized primarily by side-chain interactions with N^{171} in cis, whereas the triphosphate moiety interacts both with residues from the Walker A motif in cis (e.g., T^{16} and T^{17}) and with the side chains of N^{152} and R^{156} in trans (Fig. 3, C and D). Site-directed mutational analysis in cellulo shows that sandwich ATP binding is likely indispensable for cellulose secretion, as observed by the progressive reduction in calcofluor binding between the $BcsQ^{N152D}$, $BcsQ^{N152D•R156E}$, and $BcsQ^{N152A•R156A•N171A}$ mutants (Fig. 3E). Such ATP dependency is especially unexpected given the self-energized nature of cellulose biogenesis, which uses preactivated UDP-glucose as a substrate.

While ATP-dependent tango dimers are an essential part of the catalytic cycle of SIMIBI family members (fig. S3D), capturing them crystallographically has typically required either the introduction of hydrolysis-inactivating mutations or the use of nonhydrolysable ATP homologs and transition state mimics [reviewed in (*16*, *17*)]. In contrast, the dimer-of-heterodimers conformation appears to be the default state of purified BcsRQ (*10*) and the presence of the nucleotide triphosphate (NTP) in all $BcsRQ^{His}$ structures reveals the absence of nucleotide hydrolysis or exchange even after prolonged incubation periods or treatment with chelating agents (table S2). Consistent with this, the $Bcs^{His}RQ^{WT}$ complex exhibited only weak adenosine triphosphatase (ATPase) activity in vitro, which was further inhibited in the context of a $Bcs^{His}RQ^{WT}$-$BcsE^{217–523}$ heterocomplex (fig. S3).

A closer inspection of the BcsQ protein sequence and nucleotide-binding pocket reveals that the protein features a deviant P-loop or Walker A motif as well as an aspartate-to-cysteine (or serine) substitution at the catalytic water-activating residue at position 39 (Fig. 3D and fig. S4). In canonical SIMIBI NTPases, the P-loop typically features two lysine residues—K^{11} and K^{16} in MinD—which

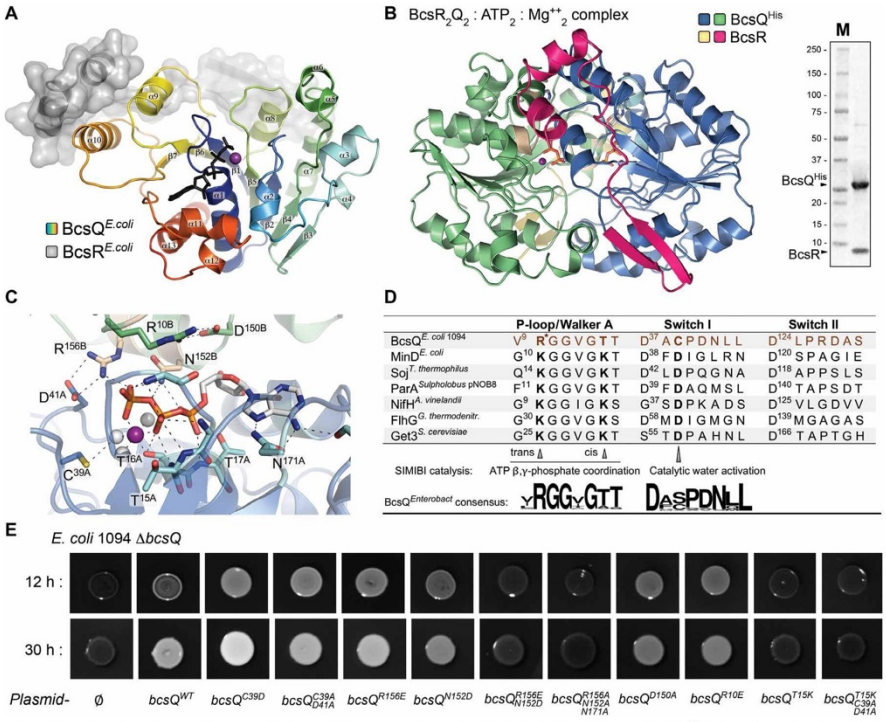


Fig. 3. Structural and functional analyses of the essential for cellulose secretion subunit BcsQ. (**A**) Structure of a BcsRQ^{His} heterodimer showing the conserved SIMIBI fold (in spectrum-colored cartoon), bound ATP (in black sticks), and Mg^{++} ion (as a purple sphere). (**B**) Crystal structure of the ATP-bound BcsRQ^{His} heterocomplex. Right: Coomassie-stained SDS-PAGE of the purified heterocomplex. (**C**) A close-up view of the ATP-binding pocket. BcsQ subunits are shown as cartoon in blue and lime, with ATP-coordinating residues colored in cyan and tan, respectively. Water molecules and the Mg^{++} ion are shown as gray and purple spheres, respectively. ATP and key surrounding residues are shown as sticks. (**D**) Primary sequence comparison of key functional motifs in canonical SIMIBI NTPases and enterobacterial BcsQ. (**E**) Functional complementation assay examining the role of key BcsQ residues in vivo. Colony cellulose secretion is evaluated by fluorescent calcofluor binding. Results are representative of at least six biological replicates. All BcsQ mutants were tested for protein expression and interactions with the BcsR and BcsE partners (E).

coordinate the phosphate moieties in trans and in cis, respectively, and are thought to aid electron density redistribution toward the phosphoanhydride bond oxygen during hydrolysis (Fig. 3D). In particular, the role of the K^{11} residue in trans is akin to that of the arginine finger in small guanosine triphosphatase (GTPase) activating proteins (GAPs), suggesting a common mechanism for intersubunit hydrolysis activation among NTPase superfamilies (*19*). The two lysines are typically key for SIMIBI protein function, as exemplified by *E. coli* MinD^{K11A} and MinD^{K16A} mutants, which are ATPase deficient and unable to interact with the downstream effector MinC (*18*).

In BcsQ, the cis-acting lysine is substituted by a threonine (T^{15}), which points away from the bound ATP and toward the core β-sheet (Fig. 3C). Mutation of this residue to the consensus lysine abolishes cellulose secretion in all tested contexts (BcsQ^{T15K} and BcsQ$^{T15K•C39A•D41A}$), which can be explained with altered geometry within the nucleotide-binding pocket rather than ATP hydrolysis activation. Although the purified Bcs^{His}RQ^{T15K} complex did not exhibit considerable ATPase activity in vitro, it showed ATP complexation, stability, and elution profiles similar to those of the wild-type BcsRQ complex. These data indicate that ATP binding and tango dimer formation alone are not sufficient for cellulose secretion and suggest an additional BcsQ-dependent step, possibly ATP hydrolysis, downstream of BcsERQ complex formation.

Further analysis of the nucleotide-binding pocket reveals that R^{10}, which corresponds to the consensus trans-acting Walker A lysine, is flipped away from the bound ATP and its conformation is stabilized by a salt bridge with the strictly conserved D^{150} from the same BcsQ subunit. Charge-reversal mutation of R^{10} (BcsQ^{R10E}) or, in contrast, a salt bridge–removal mutation of D^{150} that would release the Walker A arginine (BcsQ^{D150A}) lead both to decrease in calcofluor binding in vivo, indicating that the two interacting residues likely play a structural rather than a catalytic role in cellulose biogenesis. Instead of R^{10}, the role of arginine finger in BcsQ is taken by the tandem of conserved trans-acting residues N^{152} and R^{156}. While mutation of both residues leads to loss of cellulose secretion as mentioned above, a charge-reversal mutation of R^{156} (BcsQ^{R156E}) has actually stimulatory effects on cellulose secretion. It is important

to note that while the carboxamide nitrogen of N^{152} forms classical arginine finger interactions with the α-phosphate and the phosphoanhydride bond oxygen of the γ-phosphate, R^{156} coordinates the distal oxygen of the γ-phosphate and also engages into a salt bridge with the D^{41} residue in trans, which itself is in immediate proximity to the canonical catalytic water-activating residue at position 39 (Fig. 3C). The $BcsQ^{R156E}$ mutation could therefore have a twofold stimulatory effect with regard to putative downstream ATP hydrolysis: (i) A negative charge around the distal oxygens of the γ-phosphate would facilitate electron transfer from the attacking hydroxide nucleophile away from the distal oxygens and onto the phosphoanhydride bond oxygen, and (ii) a charge-reversal mutation of R^{156} would release its salt bridge partner D^{41} and increase the negative charge around active site–bound water molecules, therefore favoring water deprotonation and nucleophilic attack on the bound nucleotide. Consistent with this, purified $Bcs^{His}RQ^{R156E}$ exhibited stronger than wild-type ATPase activity in vitro. In addition, the injected protein sample eluted as several distinct species, consistent with BcsRQ complex disassembly, and coeluted nucleotide species did not concentrate together with the protein complex, further indicating nucleotide hydrolysis and release from the $Bcs^{His}RQ^{R156E}$ complex (fig. S3, F to H). Nevertheless, the apparent ATP hydrolytic activity of the $BcsQ^{R156E}$ mutant was inhibited in the context of a ternary $Bcs^{His}RQ^{R156E}$-$BcsE^{217–523}$ complex (fig. S3I), indicating that the BcsE regulator further stabilizes the ATP-loaded, precatalytic sandwich dimer formation.

Although the above results favor a hydrolytically active BcsQ protein, where ATP hydrolysis occurs downstream of the BcsERQ interactions in the context of cellulose biogenesis, it is less evident how the attacking nucleophile is initially activated. As mentioned above, the canonical water-deprotonating aspartate in switch I of the NTPase fold is substituted by a cysteine (C^{39}) or serine residue in enterobacterial BcsQ (Fig. 3D and fig. S4). Although the thiol side chains of cysteines can act as nucleophiles in enzymatic reactions, mutation of both C^{39} and the adjacent D^{41} to alanines ($BcsQ^{C39A•D41A}$) does not negatively affect calcofluor binding in vivo (Fig. 3). Nevertheless, a consensus-mimetic mutation ($BcsQ^{C39D}$) has a stimulatory effect on cellulose secretion, especially in the early stages of biofilm formation and even stronger than that of the $BcsQ^{R156E}$ mutation. The in vitro stability, ATPase activity, and nucleotide-loading states of the $Bcs^{His}RQ^{C39D}$ complex, however, are similar to that of the wild-type protein (fig. S3, F to H). Together, these data favor a model where BcsQ-bound ATP is hydrolyzed downstream of BcsRQ and BcsERQ complex formation, with catalytic water activation being mediated either in trans by a downstream binding partner or involving additional conformational changes within BcsQ to deliver an alternative water-deprotonating residue. In favor of this model, it is important to note that even canonical SIMIBI proteins are actually relatively poor ATPases. Rather, SIMIBI-dependent nucleotide hydrolysis typically requires membrane and/or effector binding and can involve either intrinsic to the NTPase amino acids or catalytic activation in trans by active site residues provided from downstream binding partners [e.g., lipid and MinE-dependent activation of MinD (*20*, *21*), Spo0J-dependent activation of Soj (*22*), and FlhG-dependent activation of FlhF (*23*)].

The crystal structures of the BcsRQ complex also reveal a possible mechanism for BcsR-dependent BcsQ stabilization. As reported recently, neither BcsR nor BcsQ can be purified as stable proteins when expressed separately, whereas BcsRQ coexpression leads to the purification of stable monodisperse $BcsR_2Q_2$ heterocomplexes. The crystal structures show that each BcsR polypeptide interacts stably with both BcsQ subunits (free energy gains of −21.4 and −9.4 kcal/mol) burying 1900 $Å^2$ of combined surface area (Fig. 4A and fig. S3, B and C). In particular, BcsR features a C-terminal V-shaped α-helical domain, $BcsR^{CTD}$, which effectively stitches the interface of the BcsQ homodimer by both polar and hydrophobic interactions, is required for cellulose secretion in vivo, and is sufficient for purification of a stable $Bcs^{ΔNTD}RQ$ heterocomplex (Fig. 4, A to D, and fig. S3, B and C). The interface occupied by $BcsR^{CTD}$ corresponds to the same interface occupied by MinE and Get1 in the MinDE and Get3-Get1 assemblies essential for divisome positioning and tail-anchored membrane protein sorting, respectively, indicating a common regulatory mechanism for cofactor association among SIMIBI family members (Fig. 4A) (*20*, *24*).

The remaining BcsR N-terminus ($BcsR^{NTD}$) features an extended conformation with an apical β-hairpin and adopts variable, partially disordered conformations among the six different crystal structures presented in this study (Fig. 4B). As we discussed recently, such conformational variability could have further stabilizing effects on BcsRQ complex assembly by minimizing the intrinsic aggregation propensity of BcsQ (*10*). Unexpectedly, while it is not requisite for BcsRQ complex formation (Fig. 4C), $BcsR^{NTD}$ appears to be crucial for cellulose secretion and BcsA biochemical detection in cells (Fig. 4, D and E). Fitting of the BcsRQ complex in the cryo-EM structures reveals that the two essential regulators occupy the apical density of the Bcs macrocomplex (Figs. 1C and 2F), with $BcsA^{PilZ}$ buttressing primarily one of the BcsQ subunits, whereas the α1 helix of the BcsA-proximal BcsR copy appears to extend at its N-proximal end to position $BcsR^{NTD}$ onto the catalytic glucosyl transferase domain, $BcsA^{GT}$ (Fig. 2F). Together, these data suggest additional regulatory roles for BcsR and BcsQ in synthase stabilization and activity regulation and can further explain the essential regulatory role of the protein heterocomplex in cellulose biogenesis in vivo.

Together, the cytosolic modules of the BcsARQE subunits enclose a large membrane-proximal vestibule within the Bcs macrocomplex, in which the regulatory $BcsA^{PilZ}$ module occupies a central position (Fig. 1C). Surface conservation analyses of the resolved BcsRQ and BcsRQ-BcsE assemblies (see below) indicate clustering of most conserved residues at the vestibule-facing side of the heterocomplexes (figs. S4 and S5), consistent not only with observed protein-protein interactions but also with the evolution of a specific local environment around the c-di-GMP sensing $BcsA^{PilZ}$ module and the gated catalytic $BcsA^{GT}$ domain. Conserved residues include a series of arginine (R^{174} and R^{201}) and glutamate (E^{203}) residues that adopt alternative conformations to form an electron-dense region at the bottom of the vestibule, which likely preclude nucleotide release and stabilize the tango dimer conformation, as well as composite c-di-GMP sensing sites on the partner BcsE protein, which likely contribute to activating dinucleotide retention for processive glucose polymerization (below).

Structural analyses of the BcsRQ-BcsE interactions and multisite c-di-GMP complexation

We recently showed that BcsE features a tripartite architecture, namely, an N-terminal RecA ATPase-like module ($BcsE^{NTD}$), followed by a phosphotransfer incompetent receiver ($BcsE^{REC*}$) domain and a catalytically inactive GGDEF module ($BcsE^{GGDEF*}$) (*10*). We further showed that while the N-terminal domain oligomerizes in head-to-tail

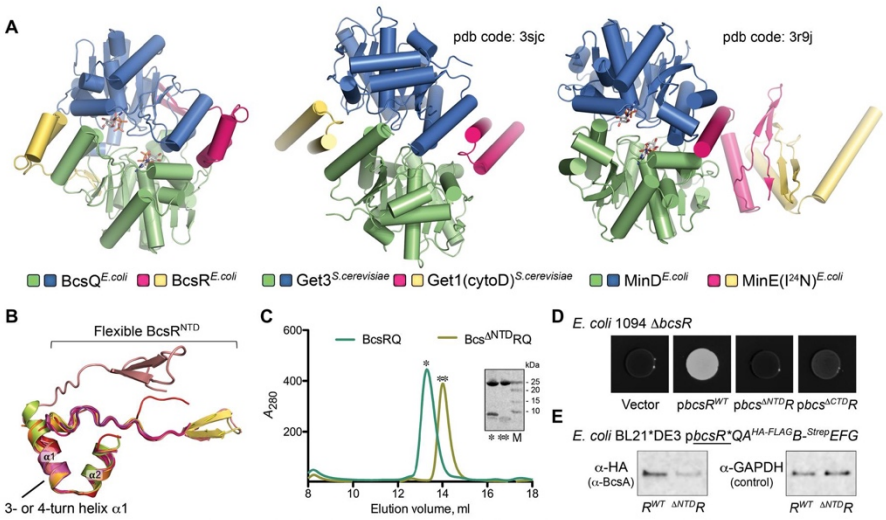


Fig. 4. Structural and functional analyses of the essential for cellulose secretion subunit BcsR. (**A**) Comparison of the BcsRQ complex with two-component complexes of SIMIBI superfamily members. (**B**) Overlay of BcsR subunits found in the asymmetric units of all structures resolved in this study. (**C**) Size exclusion chromatography and SDS-PAGE analysis of purified BcsRQ and Bcs$^{\Delta NTD}$RQ. Molecular weights for the tag-free BcsR and Bcs$^{\Delta NTD}$R are 7.4 and 3 kDa, respectively. (**D**) Functional complementation assay of the role of BcsR domains in cellulose secretion in vivo. Results are representative of at least six biological replicates. (**E**) Western blot detection of cellular BcsA upon Bcs macrocomplex expression in the context of BcsR^{NTD} deletion [GAPDH, glyceraldehyde-3-phosphate dehydrogenase (loading control)]. The results are representative of at least three replicates.

manner and participates in BcsF-driven recruitment of BcsE to the membrane, the C-terminal BcsE^{GGDEF*} moiety is responsible for binding of both dimeric c-di-GMP and BcsQ (*10*). On the basis of these data, we can unambiguously assign the membrane-proximal electron-dense region that lies across from BcsA and underneath the membrane-distal crown repeats to a BcsE^{NTD} dimer (Fig. 2B), but overall, the cryo-EM reconstruction fails to provide further structural insights into the BcsERQ regulatory complex assembly.

To overcome this, we proceeded to purify and crystallize stable BcsRQE* complexes. We obtained crystals for two multicomponent assemblies, BcsRQE$^{349–523}$ and BcsRQ^{R156E}E$^{217–523}$, which diffracted to nearly atomic resolution and yielded structural models refined to 2.5 and 2.9 Å, respectively (table S3). Within the two structures, crystallized in the presence of c-di-GMP and AppCp/Mg^{2+}, the BcsQ dimers are largely unchanged from the those observed in the isolated BcsRQ complexes (RMSD over all atoms < 0.41 Å). The most substantial conformational change from the latter is observed in one of the BcsR copies in the BcsRQ^{R156E}E$^{217–523}$ structure, where the α1 helix features an additional turn at its N-proximal end and the remaining N-terminal region is flipped away from the complex to wrap along a BcsQ protomer from a symmetry-related assembly. This is reminiscent of the apparent conformation of the BcsA-proximal BcsR copy in the cryo-EM reconstruction, where the protein's N-terminal region is flipped away from BcsQ and appears to extend onto the synthase's glycosyl transferase module (Figs. 1C and 2F).

The two crystal structures reveal that BcsE binds BcsQ through a composite interface (1490 Å^2, free energy gain of −9.1 kcal/mol), including a nonlinear binding surface on the GGDEF* module distal from both the c-di-GMP–bound I-site and the N-proximal REC* domain, as well as an extended interface of 1055 Å^2 formed by the ~40 C-terminal residues of BcsE (BcsE^{CT}) (Fig. 5A and fig. S6, A to D). The latter are unresolved in the crystallized BcsE$^{217–523}$ construct alone (*10*), suggesting BcsQ-mediated tertiary structure stabilization. The interface is stabilized by both polar and hydrophobic interactions and particularly by conserved Pro and Leu residues (P^{515} in the enterobacterial consensus/P^{509} in BcsE$^{E.coli}$ and L^{520} in the consensus/L^{514} in BcsE$^{E.coli}$) plugging into two hydrophobic pockets onto the BcsQ surface (fig. S6C). Although deletion of the C-terminal tail or the globular GGDEF* module alone abolishes stable BcsERQ complex purification (fig. S6F), the bipartite interface may allow substantial flexibility in the relative orientations of the BcsQ and BcsE^{GGDEF*} modules in the context of a secretory assembly stabilized by additional protein-protein interactions.

In addition to revealing the structural determinants for BcsQ recruitment, the two BcsRQE structures provide unexpected insights into activating c-di-GMP complexation. Consistent with solution biophysical data (*10*), c-di-GMP is found bound as an intercalated dimer to each GGDEF* domain, partaking in canonical interactions with the conserved R^{415}xxD I-site motif. However, in the BcsRQ^{R156E}E$^{217–523}$ complex, the interstitial α_i helix and the BcsE^{REC*} module of BcsE undergo drastic conformational change [144° rotation and 45 Å displacement relative to the recently reported BcsE$^{217–523}$–c-di-GMP structure (*10*)] to contribute a second conserved RxxD motif (R^{306}ATD), which forms virtually identical I-site–like interactions with the second c-di-GMP molecule (Fig. 5, C and D). We further show that although this second I-site is insufficient for stable c-di-GMP recruitment alone (*10*), it drastically stabilizes the binding of dimeric c-di-GMP in solution as demonstrated by the

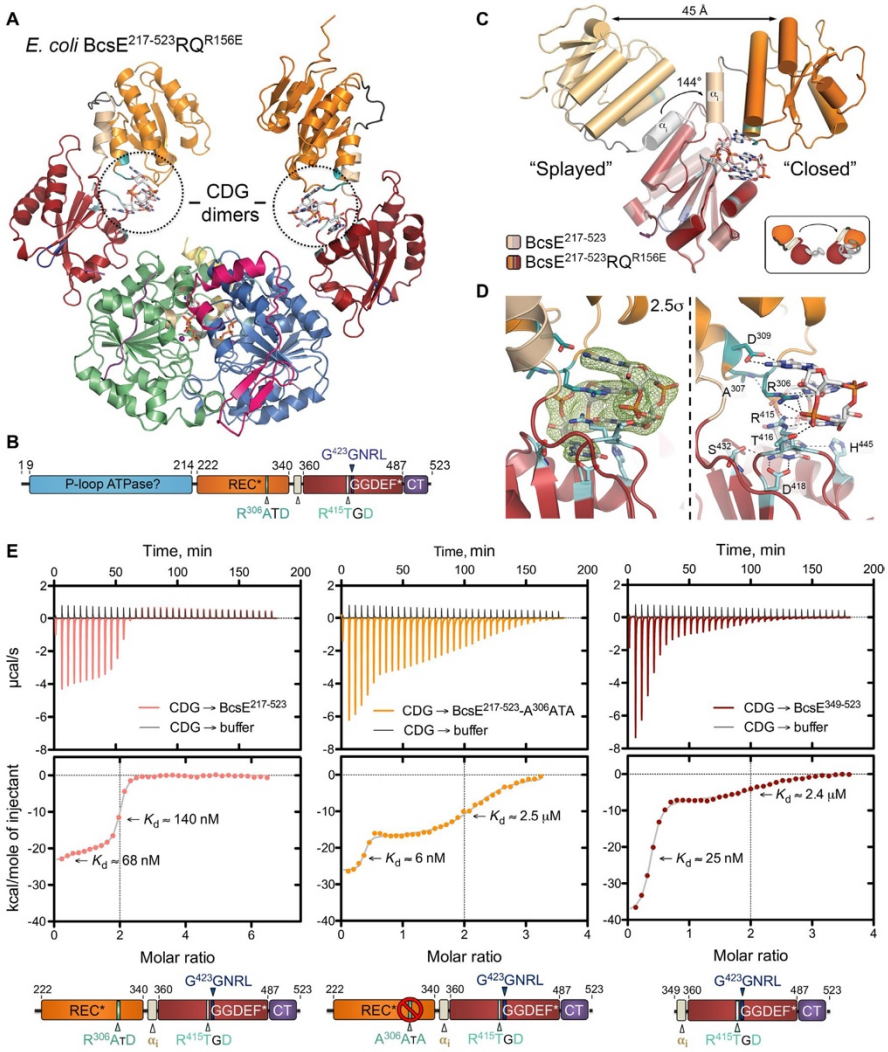


Fig. 5. Multisite c-di-GMP recognition and structure of the BcsRQ^{R156E}-BcsE$^{REC*-GGDEF*}$ complex. (**A**) Crystal structure of untagged BcsRQ^{R156E}-E$^{217–523}$ assembly in complex with ATP/ACP (β,γ-methyleneadenosine triphosphate or AppCp) and c-di-GMP. (**B**) Full-length BcsE domain architecture and functionally validated c-di-GMP–binding motifs (RxxD I-sites I and II) as identified in this study. (**C**) Captured conformational changes in the BcsE$^{REC*-GGDEF*}$ tandem between the BcsE$^{217–523}$ ["Splayed" (*10*)] and BcsRQ^{R156E}E$^{217–523}$ ("Closed") crystal structures. (**D**) Close-up views of c-di-GMP complexation in the three-component complex. Left: An $|F_o|$-$|F_c|$ partial electron density map calculated from a model before inclusion of the dinucleotide and contoured at 2.5 σ. Right: A cartoon-and-stick representation of c-di-GMP coordination. Canonical I-site I residues are colored in cyan; secondary I-site II residues are colored in teal. (**E**) Functional validation of the secondary I-site on the BcsE^{REC*} module (R^{306}ATD) by isothermal titration calorimetry. K_d values were calculated using a two-site binding model.

altered thermodynamic profiles of c-di-GMP binding to truncated (BcsE$^{349–523}$) or point-mutant (BcsE$^{217–523}$•A^{306}ATA) variants (Fig. 5E). As the processive transfer of glucose moieties onto the nascent cellulose chain requires recurrent opening of BcsA's active site by dimeric c-di-GMP binding to the protein's PilZ module, synthase-proximal enrichment of the dinucleotide in already intercalated conformation by the BcsE$^{REC*-GGDEF*}$ tandem can have direct functional implications in the context of an assembled secretion macrocomplex.

Furthermore, although c-di-GMP binding to receiver domains has been reported earlier, the dinucleotide binding mode in $BcsE^{REC*}$ is markedly different. For example, we first reported c-di-GMP recognition by *Vibrio cholerae* VpsT, where the degenerate receiver domain features a C-terminal α-helical extension (α6) through which it dimerizes to form an intersubunit c-di-GMP binding pocket (*25*). Similarly, some CheY-like response regulators have been reported to bind c-di-GMP through an arginine-rich C-terminal extension of the canonical REC domains, the corresponding peptide alone being sufficient for ligand complexation (*26*). However, to our knowledge, BcsE is the first example where a REC domain fold contributes an intrinsic, I-site–like motif for c-di-GMP coordination.

In addition to the second I-site, the $BcsRQE^{349–523}$ structure reveals a third crystallographic, c-di-GMP–binding motif (fig. S6, F to I). In particular, the dinucleotide dimer bound to the $BcsE^{GGDEF*}$ I-site of a symmetry-related BcsE molecule ($BcsE^{SYM}$) is further coordinated by multiple π-stacking and polar interactions with residues from cognate BcsE ($R^{503}H^{504}$), BcsR ($R^{51}W^{52}$), and BcsQ (R^{219}) protomers (fig. S6G). Although it is not evident that this third binding site has biological relevance in cellulose secretion, its proximity to the $BcsA^{PilZ}$ module raises the possibility of it contributing additional weak interactions for synthase-proximal dinucleotide retention in vivo.

Last, the $BcsRQ^{R156E}E^{217–523}$ structure also provides a structural basis for the poor resolution and likely conformational heterogeneity of the BcsE-corresponding regions in the cryo-EM density reconstruction. Namely, $BcsE^{217–523}$ features substantial changes in the relative orientation of the REC* domain relative to the GGDEF* module when compared not only between the $BcsE^{217–523}$ and $BcsRQ^{R156E}E^{217–523}$ structures (144° rotation, 45-Å displacement) but also between $BcsE^{217–523}$ subunits in the asymmetric unit of the three-component crystal structure alone (20.4° rotation, 6.25-Å displacement). As even local refinement of the cryo-EM data failed to yield electron densities into which we can reliably fit the crystallographic structures, we have refrained from proposing a specific model for the 3D organization of the $BcsE^{REC*\text{-}GGDEF*}$ modules in the context of the assembled Bcs macrocomplex.

DISCUSSION

Apart from some cyanobacteria, all cellulose-producing bacteria feature a BcsAB tandem as the catalytic core of their cellulose biosynthetic machineries (*1*). The latter can be classified in three major types, as found in *G. xylinus* (type I), *E. coli* (type II), and *Agrobacterium tumefaciens* (type III) (*1*). Most of the structural information on the catalytic subunits has come from studies of the BcsAB homologs from *R. sphaeroides*, an organism that features a type I cellulose secretion apparatus (*1*, *6*, *7*). As *R. sphaeroides* secretes amorphous cellulose and no evidence for BcsAB polymerization has been found in vitro or upon analysis of crystallographic interfaces (*4*), it is likely that the *R. sphaeroides* biosynthetic modules consist of separate BcsAB heterodimers in cellulo. In contrast, we showed previously that *E. coli* BcsB likely features polymeric organization in the membrane as seen in detergent-extracted and not reconstituted Bcs macrocomplexes studied by single-particle electron microscopy in negative stain (Fig. 1A) (*4*). Here, we provide nearly atomic resolution data showing that BcsB polymerizes via a β-sheet complementation mechanism among subunits and propose that polymerization would be limited by both the membrane surface tension and the transmembrane domains (TMDs) of interacting Bcs subunits, such as the BcsB membrane anchors, BcsA's transport domain on one side, and, likely, BcsE-interacting BcsF copies on the other side of the crown. Whereas BcsB features a high degree of overall sequence and surface conservation, the D2:D4 β strands and the stacked luminal loops that mediate intersubunit β-strand addition feature relatively low sequence conservation, consistent with the fact that the observed β-strand interactions are backbone-mediated and mostly side-chain independent (figs. S2, B to E, and S5).

Both the D2:D4 intersubunit interactions and the D3 luminal loops appear conserved at the secondary structure level among enterobacteria but are absent in the *R. sphaeroides* BcsB homolog, where D4 features a long amphipathic α-helix insertion in lieu of the interfacial β strand (Fig. 2A and fig. S2) (*6*). This can explain the monomeric conformation of *R. sphaeroides* BcsB and compels the question of whether the absence of a D4 amphipathic helix insertion and an *E. coli*–like predicted secondary structure can be used as a molecular marker for BcsB oligomerization. For example, in *G. xylinus*, which features a type I cellulose secretion machinery and where c-di-GMP and its effect on cellulose biosynthesis were first found (*27*), immunolabeling on freeze-fractured cells has shown that the Bcs complexes arrange in a linear row along the cell surface (*28*). Primary and secondary structure alignments show that $BcsB^{G.xylinus}$ is more akin to the *E. coli* rather than the *R. sphaeroides* BcsB homolog (i.e., presence of putative interfacial β-strand motifs and lack of amphipathic helix insertion), which raises the possibility that the peculiar supramolecular organization of cellulose synthesis in *G. xylinus* could be, at least partly, driven by intermolecular BcsB interactions.

In addition to a conserved BcsB polymerization mechanism, the cryo-EM Bcs macrocomplex reconstruction also revealed a noncanonical BcsA:BcsB stoichiometry that differs substantially from a generally accepted model of equimolar synthase tandem assembly (Fig. 1C). The altered BcsA:BcsB ratio is also particularly interesting, as it demonstrates a biosynthetically expensive stoichiometry for a complex that features a single catalytic subunit. Possible roles for the BcsB polymeric crown might include local perturbations of the IM for facilitated BcsA membrane sorting or function, a ratchet-like organization of the BcsB crown lumen for more efficient cellulose extrusion, a supramolecular assembly for recruitment of regulatory enzymes [e.g., the hydrolase BcsZ, the pEtN transferase BcsG, or BcsB-interacting c-di-GMP metabolizing enzymes PdeK and DgcC (*1*, *9*, *29*)], periplasmic peptidoglycan rearrangements for facilitated guidance toward the outer membrane export channel, or introduction of overall secretion system asymmetry related to the role of the essential SIMIBI subunit BcsQ (see below). Furthermore, the link between BcsB oligomerization and membrane deformation could also explain preferential secretion system assembly at the cell pole (*11*), where the native membrane curvature is inherently highest.

As mentioned above, cellulose secretion in *E. coli* is absolutely dependent on the small cytosolic subunits BcsR and BcsQ [Figs. 3 and 4 and (*4*)]. The structural and functional analyses of the BcsRQ heterocomplex presented here reveal multiple parallels with the structure and mechanism of canonical SIMIBI proteins, which are important multistate regulators in a variety of cellular functions (fig. S3D). An important feature of SIMIBI proteins is the uncoupling of NTP-dependent dimerization and hydrolysis, which provides an additional spatiotemporal control of the proteins' function. The data presented here support a similar model of delayed, dimerization-uncoupled, BcsQ-dependent ATP hydrolysis within the

Bcs secretion system, which occurs downstream of the BcsRQE regulatory complex assembly. Given the substantial effects of BcsQ and BcsR on the detectable levels of IM BcsA [(*4*) and Fig. 4E] and that the BcsERQ complex is recruited to the membrane upon expression of IM BcsF (*10*), it is possible that the BcsRQ tandem plays a direct role in the membrane sorting and/or stability of the BcsA synthase similar to the function of other protein-sorting SIMIBI NTPases [e.g., the SRP54-SR tandem for general membrane protein sorting, Get3 for TA-protein insertion, and FlhF for flagellum assembly initiation (*16*, *17*, *23*)]. As the BcsRQ complex is secretion system specific and involves interactions with a single BcsA copy, it would require no nucleotide or protein recycling mechanisms but rather partake in an energetically economical "one-shot" Bcs macrocomplex assembly.

Although the precatalytic ATP-bound tango dimers of SIMIBI proteins are typically pseudosymmetric, they often participate in asymmetric protein-protein interactions along their catalytic cycles (*16*, *17*). For example, studies on the MinD-MinE complex have shown that the MinE dimer cannot simultaneously affect both catalytic sites of the MinD homodimer and that asymmetric binding of MinE is sufficient to trigger ATP hydrolysis and membrane release of MinD (*20*, *21*). Similarly, in tail-anchored protein sorting homodimeric, guanosine triphosphate (GTP)–loaded Get3 forms an asymmetric complex with the Get4-Get5 partners to secure efficient cargo loading before its delivery to the membrane through GTP hydrolysis, interactions with the Get2-Get1 insertase complex, nucleotide release, and Get3 recycling (*16*, *17*). Similar functional asymmetry is also observed in the assembled Bcs macrocomplex. In particular, the apical BcsRQ subcomplex buttresses a single BcsA copy, where most of the contacts are between the $BcsA^{PilZ}$ domain and a single BcsQ subunit, whereas the N-terminal domain of the proximal BcsR chain appears to extend past the PilZ module and onto the catalytic glycosyl transferase domain. Overall secretion system asymmetry is also secured by the superhelical polymerization of BcsB discussed above, as well as by the substantial conformational flexibility of partner BcsE.

Regarding the role of the third cytosolic regulator BcsE, we recently showed that the protein features an unexpected tripartite architecture composed of degenerate (*) NTPase, REC, and GGDEF domains and that it participates in both secretion system assembly through high-affinity interactions with BcsF and BcsQ and activating cofactor complexation through binding of c-di-GMP to the conserved GGDEF* domain I-site (R^{415}TGD). Here, we demonstrate the contribution of a second I-site (R^{306}ATD), intrinsic to the REC* domain, that stabilizes binding of the dinucleotide in its synthase-activating intercalated dimer form at the REC*-GGDEF* interface (Fig. 5). Both the cryo-EM data and crystallographic structures attest to substantial structural variability between the two modules that could reflect conformation-dependent c-di-GMP binding affinities. In a functional mechanism where dimeric c-di-GMP migrates in and out of the $BcsA^{PilZ}$ domain to allow substrate entry at each cycle of processive glucose polymerization (*5*), BcsE could thus contribute to the retention of a synthase-proximal pool of activating dinucleotide that is made available through minor conformational changes. Last, the tandem $BcsE^{REC*\text{-}GGDEF*}$ I-sites (Fig. 5), the observation of a third crystallographic site for c-di-GMP recognition at the interface of BcsR, BcsQ, and BcsE protomers (fig. S6, F to I), the conspicuous conservation of the vestibule-lining protein surfaces (fig. S5), and the discovery of colocalizing, secretion system–specific, c-di-GMP–metabolizing enzymes (*29*) all point toward *E. coli*–like cellulose secretion evolution toward intersubunit cooperativity and limited dinucleotide diffusion.

MATERIALS AND METHODS

The experiments were not randomized and the investigators were not blinded during experimental design, execution, or outcome assessment. However, most experiments were reproduced independently by different investigators, including crystallographic, biochemical, biophysical, and phenotypic functional assays.

Bacterial strains

Plasmids for recombinant protein expression (see below) were propagated in and isolated from *E. coli* DH5α cells. All recombinant protein expression for structural and in vitro biochemical studies was carried out in BL21(DE3) Star cells, including the expression of selenomethionine-derivatized proteins. *E. coli* 1094 Δ*bcsQ* or Δ*bcsR* strains were used in complementation phenotypic assays with *bcsQ* or *bcsR* variants, respectively, expressed from a low-copy isopropyl-β-D-thiogalactopyranoside (IPTG)–inducible vector (pAM238, below). All bacterial strains and plasmids used in this study are available upon request.

Recombinant DNA techniques

DNA manipulations were carried out using standard protocols for polymerase chain reaction (PCR), molecular cloning, transformation, and DNA analysis. Coding regions for BcsR, BcsQ, BcsRQ, BcsRQAB, BcsE, and BcsEFG variants were amplified using *E. coli* 1094 genomic DNA as a template and a high-fidelity DNA polymerase (Phusion, New England Biolabs) and inserted via digestion/ligation cloning into IPTG-inducible expression vectors with custom-modified multiple cloning sites (MCSs). Point mutations, insertion of stop codons, MCS modifications, and domain deletions were performed using inverse PCR-based protocols and mutation-specific oligonucleotides as primers. All recombinant vectors and introduced mutations were verified by DNA-sequencing and, where applicable, IPTG-inducible protein expression. All BcsQ mutants tested in phenotypic functional assays were verified for expression and copurification with the rest of the $BcsE^{217-523}$RQ complex components.

Protein expression and purification

As BcsR and BcsQ fail to express or purify stably when cloned separately, all variants of the BcsRQ complex were purified following protein coexpression and not complex reconstitution. In particular, the coding region corresponding to the BcsRQ tandem (as found in operon) was amplified and cloned into both the pProExHTB and pET21b expression vectors, adding a cleavable N-terminal or non-cleavable C-terminal hexahistidine tag to BcsR (pProExHTB-^{His}RQ) or BcsQ (pET-RQ^{His}), respectively (*10*). For coexpression purposes, the coding regions corresponding to $BcsE^{217-523}$ and $BcsE^{349-523}$ were cloned into custom-modified pRSFDuet1* expression vector under control of the first T7 promoter, as reported recently (*10*). Overexpression of the Bcs macrocomplexes was performed as previously (*4*) by coexpression of the pCDFDuet1-Bcs^{His}RQ$A^{HA\text{-}FLAG}$B and pRSFDuet1*-^{Strep}EFG constructs after introduction of targeted modifications. Last, for $BcsB^{FL}$ expression, the corresponding coding region was cloned using Nde I/Not I–based restriction/ligation

protocol into a pET21b vector to yield a C-terminally hexahistidine-tagged protein upon recombinant overexpression.

For protein purification, all expression vectors were (co-)transformed into chemically competent *E. coli* BL21(DE3) Star cells. For the expression of native proteins, cells were grown at 37°C under aerobic conditions in terrific broth medium supplemented with appropriate antibiotics [ampicillin (100 μg/ml), kanamycin (40 μg/ml), or a combination of ampicillin (70 μg/ml) + kanamycin (30 μg/ml) for coexpressed vectors]. At a cell optical density corresponding to an absorbance of 0.8 to 1.0 at 600 nm (OD_{600}), the cells were moved to 17°C and overnight protein expression was induced by the addition of IPTG at a final concentration of 0.7 mM. For the expression of selenomethionine-derivatized proteins, 4 liters of cells was initially grown at 37°C in LB medium to an OD_{600} of 0.5 to 0.6. Cells were then pelleted by centrifugation (4000*g*, 15 min, 20°C), gently washed with 200 ml of 1× SelenoMet Medium Base (Molecular Dimensions), collected again, and resuspended in 1 liter of complete SelenoMet Medium (Molecular Dimensions) supplemented with L-selenomethionine (40 mg/liter) and the appropriate antibiotic. Cells were then grown for an additional hour at 37°C, transferred to 17°C, and induced with IPTG as above.

After 16 hours, cells were harvested by centrifugation, resuspended in lysis buffer, and flash-frozen in liquid nitrogen. For BcsRQ purification, the composition of the lysis buffer was 20 mM Hepes (pH 8.0), 120 mM NaCl, 19 mM imidazole (pH 8.0), 2 mM β-mercaptoethanol, and cOmplete protease inhibitors (1 tablet/50 ml) (Roche). For the $BcsRQ^{R156E}$-$BcsE^{217–523}$ and BcsRQ-$BcsE^{349–523}$ complexes, the immobilized metal affinity chromatography (IMAC) buffer was also supplemented with 0.5 μM c-di-GMP (Jena Bioscience or Sigma-Aldrich), 2 μM AppCp (Jena Bioscience), 5 mM $MgCl_2$, and 10% glycerol.

For protein purification, cells were thawed and lysed using an EmulsiFlex-C3 high-pressure homogenizer (Avestin). Cell debris were removed by centrifugation (1 hour at 50,000*g* and 4°C), and the cleared lysates were loaded onto buffer-washed Talon Superflow resin (GE Healthcare) at approximately 0.5 to 1 ml of resin/liter of culture. The resin was subsequently washed with more than 20 volumes of IMAC buffer A (protease inhibitor–free lysis buffer as above), and bound proteins were eluted in a single step with IMAC buffer A supplemented with 200 mM imidazole (pH 8.0) (IMAC buffer B).

For purification of $BcsRQ^{His}$, eluted protein was supplemented with 15 mM EDTA (pH 8.0), concentrated to 2.5 ml using an Amicon Ultra centrifugal filter (30-kDa cutoff; Millipore), and desalted using a disposable PD-10 desalting column (GE Healthcare) and IMAC buffer A. The protein fraction was then incubated with 5 mM $MgCl_2$ and excess AppCp (1 mM), adenosine diphosphate (ADP; 1 mM), or no nucleotide and subjected to size exclusion chromatography on a Superdex 200 Increase 10/300 GL column (GE Healthcare) equilibrated in gel filtration buffer [20 mM Hepes (pH 8.0), 120 mM NaCl, and 2 mM dithiothreitol (DTT)]. Collected protein fractions were analyzed for purity by SDS–polyacrylamide gel electrophoresis (SDS-PAGE), pooled, concentrated, flash-frozen in liquid nitrogen, and stored at −80°C. For purification of untagged $Bcs^{\Delta NTD}RQ$ complex, the coding region covering residues 40 to 62 of BcsR and full-length BcsQ were cloned into a pProExHTB vector and IMAC-purified in parallel to the full-length BcsRQ complex as above. The eluted complexes were desalted for imidazole removal and incubated overnight with His-tagged viral HRV-3C protease for cleavage of the N-terminal hexahistidine tag on BcsR. Following reverse IMAC for protease and tag removal, the flow-through fraction was concentrated and subjected to size exclusion chromatography (SEC). The eluted fractions were analyzed by SDS-PAGE, concentrated, flash-frozen, and stored at −80°C. Molecular weights for the full-length BcsR and $Bcs^{\Delta NTD}R$ following tag cleavage are 7.5 and 3 kDa, respectively.

Complexes $BcsRQ^{R156E}$-$BcsE^{217–523}$ and BcsRQ-$BcsE^{349–523}$ were purified in a similar two-step IMAC procedure. After tag removal, the proteins were subjected to SEC using a Superdex 200 Increase 10/300 GL column and gel filtration buffer with composition 20 mM Hepes (pH 8.0), 120 mM NaCl, 5 mM $MgCl_2$, 0.5 μM c-di-GMP, 2 μM AppCp, 2 mM DTT, and 10% glycerol. Collected protein fractions were analyzed for purity and stoichiometric complex assembly, concentrated, and flash-frozen in liquid nitrogen for storage at −80°C.

Expression of the Bcs macrocomplex (pCDFDuet1-$Bcs^{His}RQA^{HA-FLAG}B$ + pRSFDuet1*-$^{Strep}EFG$) was performed as reported previously (*4*). Briefly, after cotransformation, culture growth, induction and overnight expression, cells were pelleted by centrifugation and resuspended in ice-cold lysis buffer containing 20 mM Hepes (pH 8.0), 120 mM NaCl, 10% glycerol, 5 mM $MgCl_2$, 2 μM AppCp, 2 μM c-di-GMP, 250 μM cellobiose, lysozyme (100 μg/ml), and cOmplete EDTA-free protease inhibitors (1 tablet/50 ml) (Roche). After lysis (EmulsiFlex-C3), cell debris were removed by low-speed centrifugation (12,000*g*, 15 min, 4°C) and the membranes were pelleted by ultracentrifugation using an SW 28 Ti Beckman rotor (26,500 rpm or up to 126,000*g* for 1 hour at 4°C). After removal of the supernatant, the membrane fraction was resuspended in solubilization buffer containing all lysis buffer components but lysozyme and cellobiose, as well as a mix of detergents at the following final concentrations: 0.4% (w/v) digitonin (Sigma-Aldrich), 0.4% (w/v) *n*-dodecyl-β-D-maltopyranoside (anagrade β-DDM, Anatrace), 0.4% (w/v) decyl maltose neopentyl glycol (DM-NPG; Anatrace), and 0.2% lauryl maltose NPG (LM-NPG; Anatrace). After a 60-min incubation at 17°C and under mild agitation, the solubilized membrane fraction was cleared by a second high-speed centrifugation step as above. The supernatant was then incubated with anti-FLAG M2 affinity gel (50 μl of packed resin per liter of induced culture; Sigma-Aldrich) under mild agitation at 4°C for 1 hour. After gravity elution of the nonbound fraction, the resin was washed extensively (>30 column bed volumes) with binding buffer containing all lysis buffer components but lysozyme and cellulase, as well as 0.008% (w/v) LM-NPG. The bound complexes were then eluted using four column bed volumes of elution buffer (affinity buffer supplemented with 3× FLAG peptide at 100 μg/ml) and concentrated on a 100-kDa cutoff Amicon Ultra (Merck Millipore) centrifugal filter. For cryo-EM grid preparation, the Bcs macrocomplex was eluted in a glycerol-free buffer, concentrated to ~0.8 mg/ml, spotted on glow-discharged gold Quantifoil grids, blotted, and plunge-frozen in liquid ethane using a Vitrobot Mark IV device (Thermo Fisher Scientific) at 4°C and 100% humidity.

For BcsA detection in the context of N-terminally truncated BcsR ($Bcs^{\Delta NTD}R$), expression of the corresponding Bcs macrocomplexes was performed as above. To avoid overnight saturation effects, expression was IPTG-induced at 37°C and $BcsA^{HA-FLAG}$ detection was analyzed 120 min after induction. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) detection was used as loading control for the experiments.

Last, $BcsB^{FL}$-overexpressing BL21(DE3)* cells were pelleted and resuspended in buffer containing 20 mM Hepes (pH 8.0), 500 mM NaCl, 19 mM imidazole (pH 8.0), and 10% glycerol and lysed by sonication. Nonlysed cells were removed by a 15-min centrifugation at 12,000*g*, and the supernatant was collected. The sample was then mixed with detergents at final concentrations of 0.4% β-DDM, 0.4% digitonin, 0.4% DM-NPG, 0.2% GDN101, and 0.2% LM-NPG (Anatrace) and incubated at 17°C for 1 hour. The extract was subsequently clarified by a high-speed centrifugation (60,000*g* for 1 hour), and the supernatant was subjected to a single-step IMAC purification. The $BcsB^{FL}$-bound resin was washed extensively with IMAC buffer A [20 mM Hepes (pH 8.0), 500 mM NaCl, 19 mM imidazole, and 0.008% LM-NPG] and eluted with buffer B containing 20 mM Hepes (pH 8.0), 500 mM NaCl, 0.008% LM-NPG, and 200 mM imidazole (pH 8.0). The quality of the eluted sample was verified by SDS-PAGE and used for cryo-EM grid preparation as described above.

SEC-coupled on-column cross-linking

To detect $BcsB^{FL}$ oligomerization in solution biochemically, we used SEC-coupled, on-column cross-linking (*13*). First, 200 µl of 0.25% glutaraldehyde was injected to a Superose 6 Increase 10/300 GL column (Cytiva), preequilibrated in buffer containing 20 mM Hepes (pH 8.0), 500 mM NaCl, 0.006% LM-NPG, and 0.006% GDN101. The run was stopped after 5 ml of elution volume, and the sample loop was flush-cleaned with the same buffer. Concentrated IMAC-purified full-length $BcsB^{His}$ was subsequently injected onto the column, and the run was continued. As the higher hydrodynamic radius of the protein sample incurs faster mobility through the chromatography column, protein assemblies passing through the glutaraldehyde bolus would be briefly exposed to the cross-linking agent and any nonspecific aggregates would be simultaneously separated by the gel filtration medium. The mild cross-linking of $BcsB^{FL}$ into higher–molecular weight species was visualized by running individual fractions across the non-aggregative elution peak on 4 to 20% SDS-PAGE gradient gels and Coomassie staining.

SDS-PAGE and Western blot analyses

Protein fractions were analyzed by standard denaturing SDS-PAGE electrophoresis using 4 to 20% gradient mini-gels (Bio-Rad), Expedeon InstantBlue Coomassie stain, and a Li-Cor Odyssey Fc system for Coomassie visualization (700-nm channel). For Western blot analyses, SDS-PAGE–migrated proteins were directly transferred using a standard mini-gel transfer protocol, polyvinylidene difluoride membranes, and a Trans-blot Turbo transfer system (Bio-Rad). Blocking and antibody incubations were performed in the presence of 5% skim milk in TPBS (1X phosphate-buffered saline supplemented with 0.1% Tween™ 20 detergent); all washes between and after antibody incubations were performed with 1× TPBS buffer. Mouse anti-HA (hemagglutinin) (Thermo Fisher Scientific, #26183; dilution 1:1000) and anti-GAPDH (Thermo Fisher Scientific, #MA5-15738; dilution 1:1000) antibodies were used as primary antibodies; Alexa Fluor 680–conjugated donkey anti-mouse antibody (Abcam, ab175774; dilution 1:10,000) was used as secondary antibody. The signal was detected using a Li-Cor Odyssey Fc system in the 700-nm channel.

Crystallization, data collection, and structure determination

All crystals were obtained by sitting or hanging-drop vapor diffusion by mixing equal volumes of protein (1.5 to 6 mg/ml) and reservoir solution followed by incubation at 4°C. Crystallization conditions were as follows: Diffracting $BcsRQ^{His}$ crystals with P 1 21 1 space group appeared in multiple crystallization conditions with typical composition of 15 to 22% polyethylene glycol (PEG) 2000, 100 mM 2-(N-morpholino)ethanesulfonic acid (MES), or bis-tris with pH 5.5 to 6.5 and 0 to 7% glycerol or xylitol; $BcsRQ^{His}$ crystals with P 21 21 21 space group grew in 100 mM MES (pH 6.5) and 12% PEG 20,000; $BcsRQ^{R156E}$-$BcsE^{217–523}$ crystals appeared in 100 mM MES (pH 6.0) and 11% PEG 20,000; BcsRQ-$BcsE^{349–523}$ crystals grew in 100 mM Tris-HCl (pH 9.0), 22 to 25% PEG 6000, and 500 mM LiCl. For crystallization of AppCp- or ADP-pretreated $BcsRQ^{His}$, $MgCl_2$ and either nucleotide were also added to the crystallization condition at 500 and 50 µM final concentrations, respectively. For crystallization of $BcsE^{217–523}$ and the complexes $BcsRQ^{R156E}$-$BcsE^{217–523}$ and BcsRQ-$BcsE^{349–523}$, the crystallization conditions were supplemented with 50 µM c-di-GMP. All $BcsRQ^{His}$ crystals with P 1 21 1 space group, as well as all BcsE-containing crystals, appeared within 3 to 14 days, whereas $BcsRQ^{His}$ crystals with P 21 21 21 space group were detected 6 to 8 months after setting up of the crystallization trials. For cryo-protection, crystals were soaked in reservoir solution supplemented with 25 to 30% glycerol, 1 mM DTT, and, where applicable, 50 µM c-di-GMP. Cryo-preserved crystals were flash-frozen and stored in liquid nitrogen. Data were collected on frozen crystals at 100 K using synchrotron radiation at the European Synchrotron Radiation Facility (ESRF, beamline ID29 MX) and Soleil (beamlines PX1 and PX2) synchrotrons.

Data reduction was carried out with the software package XDS (*30*). Experimental phases were obtained by single-wavelength anomalous diffraction (SAD) experiments on crystals grown from selenomethionine-derivatized proteins and with wavelengths corresponding to the experimentally determined selenium K-edge. Initial models for the BcsRQ complex were built manually in the SAD-derived experimental density using Coot (*31*), and a BcsQ monomer was used as a search model in molecular replacement phasing of x-ray diffraction datasets collected on crystallized native complexes ($BcsRQ^{His}$, BcsRQ-$BcsE^{349–523}$, and $BcsRQ^{R156E}$-$BcsE^{217–523}$; tables S2 and S3). Models for the $BcsE^{GGDEF*}$ and Bcs^{REC*} domains were derived from the recently reported crystal structure of $BcsE^{217–523}$ (*10*) and used as additional search models in the molecular replacement phasing of the BcsRQ-$BcsE^{349–523}$ and $BcsRQ^{R156E}$-$BcsE^{217–523}$ datasets. Reiterative refinements in Phenix and Coot of all structures yielded the final refined models (*31*, *32*). Data collection and refinement statistics are summarized in tables S2 and S3. For illustration purposes, all crystal structures were displayed in the PyMOL Molecular Graphics System (Schrödinger LLC) or UCSF Chimera (*33*). The latter was also used for displaying the 3D reconstructions of the assembled Bcs macrocomplex.

Single-particle cryo-EM

To visualize the self-polymerization propensity of purified His-tagged $BcsB^{FL}$, we collected a dataset of 3830 movies on the Elsa Talos Arctica transmission electron microscope (Thermo Fisher Scientific) at the European Institute of Chemistry and Biology (IECB) Bordeaux operated at 200 kV and equipped with a Gatan K2 Summit direct electron detector. Movies of 67 frames were recorded using SerialEM (*34*) in electron counting mode at a total exposure dose of 49.6 $e^-/Å^2$ and a corrected pixel size of 1.13 $Å^2$. The movies were motion- and contrast transfer function (CTF)–corrected using MotionCor2 (*35*) and Gctf (*36*), respectively, within the cryoSPARC v2 interface. Particles were auto-picked using the latter's "blob picker" function, and after

three rounds of 2D classification in cryoSPARC v2 (*37*), a total of 46,807 particles distributed among 15 classes were retained and used in ab initio model generation and nonuniform refinement. Each $BcsB^{peri}$ octamer was further refined locally after subtraction of the remaining densities, and the resulting reconstructions were combined to yield the locally refined $BcsB^{peri}$ hexadecamer.

We used the same microscope to record a dataset of 6793 movies of the Bcs macrocomplex with similar data collection parameters (64 frames per movie, 48.3 $e^-/Å^2$ total exposure dose, corrected pixel size of 1.13 $Å^2$) and analysis workflow, yielding a final set of 286,771 particles distributed among 64 2D classes, the best of which were kept as templates for particle auto-picking in subsequent analyses (see below). For initial 3D model generation for the assembled Bcs macrocomplex, classes with less than five BcsB copies, as well as classes visually missing densities corresponding to the IM and/or cytosolic regions, were removed during the 2D classification. A 3D model was generated ab initio and subjected to nonuniform refinement; however, the overall resolution of the IM and cytosolic modules was not sufficiently high for reliable density assignment.

To further resolve the 3D architecture of the assembled Bcs macrocomplex, we therefore recorded a total of 9129 movies on the CM01 Titan Krios transmission electron microscope (Thermo Fisher Scientific) at the ESRF Grenoble operated at 300 kV and equipped with a Gatan K2 Summit direct electron detector and a GIF Quantum LS imaging filter (*38*). The images were collected in two separate sessions keeping a constant total electron dose (48 $e^-/Å^2$), defocus range (−0.75 to −2.75 μm), and pixel size (1.05248 $Å^2$). The movies were motion- and CTF-corrected as above, using MotionCor2 and Gctf within the cryoSPARC v2 interface (*35*–*37*). Particles were autopicked using the cryoSPARC v2 template picker and 2D classes generated from the initial Talos Arctica dataset. The resulting particle stacks from each dataset were then subjected to three rounds of 2D classification for removal of "junk" particles. A total of 576,455 particles containing three to six identifiable BcsB copies were then used for ab initio model generation using a single class. Nonuniform refinement of the resultant model yielded a 3D reconstruction corresponding to a BcsB pentamer resolved at ~3.1-Å average resolution accompanied by lower-intensity, lower-resolution regions corresponding to the rest of the Bcs macrocomplex. The map was segmented in UCSF Chimera (*33*) and separate soft masks for the BcsB pentamer, and the rest of the assembly were generated in cryoSPARC v2 (*37*). The low-resolution regions were then subtracted from the refined particles, and the BcsB pentamer was locally refined to 2.9-Å average resolution, with extensive regions featuring local resolution at 2.5 Å. The nonuniformly refined maps before and after local refinement were then autosharpened and combined in Phenix (*32*). Initial BcsB backbone tracing and atomic model building were performed manually for a single BcsB subunit in Coot (*31*). The model was then used for generation of a BcsB pentamer, which was thoroughly refined against the experimental density map through iterative model building and refinement cycles in Coot and Phenix, respectively. Coordinate refinement statistics are summarized in table S1.

To improve the resolution of the BcsRQAB assembly (fig. S7), we first removed a substantial part of the partially assembled Bcs complexes by force-splitting the dataset through 3D classification between two classes: a BcsB crown pentamer and an assembled Bcs macrocomplex as derived from the model generated from the Talos Arctica dataset. The latter class was then subjected to nonuniform refinement in cryoSPARC v2 and two rounds of 3D variability analysis and display in clusters (*37*), where only particles belonging to clusters with resolved IM and cytosolic densities were retained for further analysis. A resulting stack of 173,294 particles were then input for ab initio model generation and nonuniform refinement, and the resulting 3D density reconstruction was segmented in UCSF Chimera (*33*). Separate soft masks were generated for the BcsRQAB assembly (i.e., regions corresponding to the apical $BcsR_2Q_2$ complex; to BcsA's TMD, GT, and PilZ domains; and to the partner BcsB copy) and the rest of the reconstruction. The latter mask was used as input for particle subtraction, and the remaining BcsRQAB assembly was locally refined. A BcsA homology model was generated in Robetta (*15*) based on the $BcsA^{R.sphaeroides}$ crystal structures, and its separate domains were rigid body–fitted and refined against the locally refined BcsRQAB-corresponding cryo-EM density.

Isothermal titration calorimetry

Apparent dissociation constants (K_d) and stoichiometry of interactions (N) were measured by isothermal titration calorimetry (ITC) (*39*) using a Microcal VP-ITC calorimeter from Malvern Panalytical at 20°C. For c-di-GMP binding studies, 0.8 to 1 mM c-di-GMP were used as a ligand in the syringe and purified protein was added to the cuvette at concentrations of 40 to 50 μM. The protein and ligand were purified/diluted in the exact same buffer to minimize nonspecific dilution heat effects. Protein concentrations were determined by a combination of methods including a reducing agent–compatible colorimetric assay (RC DC, Bio-Rad) and 280-nm absorbance measurements under denaturing conditions (A_{280}; 6 M guanidinium chloride) while accounting for potential scattering contributions (A_{330}). All ITC data were analyzed by integrating the injection heat effects, normalized to the amount of ligand and protein present, and curve-fitting based on a single- or two-site binding models using the Origin software package for Microcal. For all titrations, titrations of the ligand into buffer were performed to account for heat dilution effects, and the latter were subtracted during the ligand binding analyses. The apparent K_d and N were derived from the data by using standard procedures, and the graphs were replotted using the GraphPad Prism software.

Calcofluor-binding cellulose secretion assay

To test for the functional effects of BcsQ and BcsR point mutants or truncated variants, we resorted to a functional complementation assay as established previously (*4*). First, chemically competent cells were prepared from *E. coli* 1094 Δ*bcsQ* or Δ*bcsR* deletion strains (*4*). These were then transformed with low-copy pAM-238 plasmids carrying wild-type or mutant *bcsQ* or *bcsR* genes and plated on LB-agar plates (Miller) supplemented with the appropriate antibiotics [streptomycin (60 μg/ml) and chloramphenicol (15 μg/ml)]. Single colonies were inoculated in 5 ml of LB medium with antibiotics and left to grow overnight at 37°C and agitation. On the following morning, 5 μl of the cultures was spotted onto low-salt LB-agar plates [NaCl (1.5 g/liter)] supplemented with the antibiotics, 0.1 mM IPTG, and 0.02% calcofluor (Fluorescent Brightener 28, Sigma-Aldrich). The spots were allowed to air-dry, and the plates were incubated at 30°C. After 24 hours, the plates were photographed under brief illumination with long-wave ultraviolet light (365 nm).

Nucleotide loading state analysis

We used reversed-phase high-performance liquid chromatography to separate nucleotides for the determination of nucleotide loading

states of purified protein complexes. Purified wild-type and mutant BcsRQ complexes were purified by IMAC and gel filtration as described above and concentrated to 20 mg/ml using Microcon Centrifugal Filter units (Merck Millipore) with a 10-kDa cutoff. The concentrated protein samples were denatured on ice using equal volume of chilled isopropanol and immediately resuspended with water to 40 μM final protein concentration. Each sample was subsequently filtered through Microcon Centrifugal Filter units (Millipore) with a 10-kDa cutoff, and nucleotides were separated on a C18 reversed-phase column (Poroshell 120 EC-C18 3 × 150 mm, 2.7 μm, flow rate of 0.5 ml/min) using 100 mM potassium phosphate buffer (buffer A, pH 6.0) as mobile phase (6 min), followed by a methanol-phosphate gradient for column cleanup and reequilibration [1 min 0 to 100% buffer B (30% methanol/70% buffer A), 4 min 100% buffer B, 1 min 100 to 0% buffer B, 2 min buffer A]. Eluted nucleotides were identified by comparison to injected ATP and ADP standards.

ATP hydrolysis assay

ATPase activity assays were conducted using the ATPase/GTPase activity kit (MAK113, Sigma-Aldrich) according to the manufacturer's guidelines. Following IMAC and SEC purification, wild-type and mutant BcsRQ complexes were diluted to final concentrations of 5, 2.5, 1.25, and 0.625 μM in gel filtration buffer [20 mM Hepes (pH 8.0), 120 mM NaCl, and 2 mM DTT]. Ten microliters of each sample was mixed with 20 μl of the provided assay buffer [40 mM tris-HCl (pH 7.5), 80 mM NaCl, 8 mM $MgAc_2$, and 1 mM EDTA] and 10 μl of 4 mM freshly prepared ATP in a microplate. The reactions were incubated for 25 min at room temperature before stopping them by the addition of the malachite green–containing blocking reagent. The samples were left to develop for additional 25 min at room temperature to allow the malachite green to form a stable dark green product with the free phosphate liberated by the ATPase reactions. The colorimetric products, proportional to enzyme activity, were measured in a microplate reader at 620 nm. A previously characterized active ATPase [$FleQ^{T149E}$ (*40*)] was used as a positive control, and negative control wells contained gel filtration buffer in lieu of protein sample.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/5/eabd8049/DC1

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. U. Römling, M. Y. Galperin, Bacterial cellulose biosynthesis: Diversity of operons, subunits, products, and functions. *Trends Microbiol.* **23**, 545–557 (2015).
2. E. Petersen, E. Mills, S. I. Miller, Cyclic-di-GMP regulation promotes survival of a slow-replicating subpopulation of intracellular *Salmonella* Typhimurium. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 6335–6340 (2019).
3. E. Rossi, A. Cimdins, P. Lüthje, A. Brauner, Å. Sjöling, P. Landini, U. Römling, "It's a gut feeling"—*Escherichia coli* biofilm formation in the gastrointestinal tract environment. *Crit. Rev. Microbiol.* **44**, 1–30 (2017).
4. P. V. Krasteva, J. Bernal-Bayard, L. Travier, F. A. Martin, P.-A. Kaminski, G. Karimova, R. Fronzes, J.-M. Ghigo, Insights into the structure and assembly of a bacterial cellulose secretion system. *Nat. Commun.* **8**, 2065 (2017).
5. J. L. W. Morgan, J. T. McNamara, J. Zimmer, Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nat. Struct. Mol. Biol.* **21**, 489–496 (2014).
6. J. L. W. Morgan, J. Strumillo, J. Zimmer, Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* **493**, 181–186 (2013).
7. J. L. W. Morgan, J. T. McNamara, M. Fischer, J. Rich, H.-M. Chen, S. G. Withers, J. Zimmer, Observing cellulose biosynthesis and membrane translocation *in crystallo*. *Nature* **531**, 329–334 (2016).
8. O. Omadjela, A. Narahari, J. Strumillo, H. Mélida, O. Mazur, V. Bulone, J. Zimmer, BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17856–17861 (2013).
9. W. Thongsomboon, D. O. Serra, A. Possling, C. Hadjineophytou, R. Hengge, L. Cegelski, Phosphoethanolamine cellulose: A naturally produced chemically modified cellulose. *Science* **359**, 334–338 (2018).
10. S. Zouhir, W. Abidi, M. Caleechurn, P. V. Krasteva, Structure and multitasking of the c-di-GMP-sensing cellulose secretion regulator BcsE. *MBio* **11**, e01303-20 (2020).
11. B. Le Quéré, J.-M. Ghigo, BcsQ is an essential component of the *Escherichia coli* cellulose biosynthesis apparatus that localizes at the bacterial cell pole. *Mol. Microbiol.* **72**, 724–740 (2009).
12. T. R. D. Costa, C. Felisberto-Rodrigues, A. Meir, M. S. Prevost, A. Redzej, M. Trokter, G. Waksman, Secretion systems in Gram-negative bacteria: Structural and mechanistic insights. *Nat. Rev. Microbiol.* **13**, 343–359 (2015).
13. A. K. Shukla, G. H. Westfield, K. Xiao, R. I. Reis, L.-Y. Huang, P. Tripathi-Shukla, J. Qian, S. Li, A. Blanc, A. N. Oleskie, A. M. Dosey, M. Su, C.-R. Liang, L.-L. Gu, J.-M. Shan, X. Chen, R. Hanna, M. Choi, X. J. Yao, B. U. Klink, A. W. Kahsai, S. S. Sidhu, S. Koide, P. A. Penczek, A. A. Kossiakoff, V. L. Woods Jr., B. K. Kobilka, G. Skiniotis, R. J. Lefkowitz, Visualization of arrestin recruitment by a G-protein-coupled receptor. *Nature* **512**, 218–222 (2014).
14. H. P. Chen, R. M. Brown Jr., Immunochemical studies of the cellulose synthase complex in *Acetobacter xylinum*. *Cellulose* **3**, 63–75 (1996).
15. D. E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32**, W526–W531 (2004).
16. G. Bange, I. Sinning, SIMIBI twins in protein targeting and localization. *Nat. Struct. Mol. Biol.* **20**, 776–780 (2013).
17. S.-O. Shan, ATPase and GTPase tangos drive intracellular protein transport. *Trends Biochem. Sci.* **41**, 1050–1060 (2016).
18. I. Hayashi, T. Oyama, K. Morikawa, Structural and functional studies of MinD ATPase: Implications for the molecular recognition of the bacterial cell division apparatus. *EMBO J.* **20**, 1819–1828 (2001).
19. J. L. Bos, H. Rehmann, A. Wittinghofer, GEFs and GAPs: Critical elements in the control of small G proteins. *Cell* **129**, 865–877 (2007).
20. K. T. Park, W. Wu, K. P. Battaile, S. Lovell, T. Holyoak, J. Lutkenhaus, The Min oscillator uses MinD-dependent conformational changes in MinE to spatially regulate cytokinesis. *Cell* **146**, 396–407 (2011).
21. K.-T. Park, W. Wu, S. Lovell, J. Lutkenhaus, Mechanism of the asymmetric activation of the MinD ATPase by MinE. *Mol. Microbiol.* **85**, 271–281 (2012).
22. T. A. Leonard, P. J. Butler, J. Löwe, Bacterial chromosome segregation: Structure and DNA binding of the Soj dimer—A conserved biological switch. *EMBO J.* **24**, 270–282 (2005).
23. G. Bange, N. Kümmerer, P. Grudnik, R. Lindner, G. Petzold, D. Kressler, E. Hurt, K. Wild, I. Sinning, Structural basis for the molecular evolution of SRP-GTPase activation by protein. *Nat. Struct. Mol. Biol.* **18**, 1376–1380 (2011).
24. S. Stefer, S. Reitz, F. Wang, K. Wild, Y.-Y. Pang, D. Schwarz, J. Bomke, C. Hein, F. Löhr, F. Bernhard, V. Denic, V. Dötsch, I. Sinning, Structural basis for tail-anchored membrane protein biogenesis by the Get3-receptor complex. *Science* **333**, 758–762 (2011).
25. P. V. Krasteva, J. C. N. Fong, N. J. Shikuma, S. Beyhan, M. V. A. S. Navarro, F. H. Yildiz, H. Sondermann, *Vibrio cholerae* VpsT regulates matrix production and motility by directly sensing cyclic di-GMP. *Science* **327**, 866–868 (2010).
26. J. Nesper, I. Hug, S. Kato, C.-S. Hee, J. M. Habazettl, P. Manfredi, S. Grzesiek, T. Schirmer, T. Emonet, U. Jenal, Cyclic di-GMP differentially tunes a bacterial flagellar motor through a novel class of CheY-like regulators. *eLife* **6**, e28842 (2017).
27. P. Ross, H. Weinhouse, Y. Aloni, D. Michaeli, P. Weinberger-Ohana, R. Mayer, S. Braun, E. de Vroom, G. A. van der Marel, J. H. van Boom, M. Benziman, Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* **325**, 279–281 (1987).
28. S. Kimura, H. P. Chen, I. M. Saxena, R. M. Brown Jr., T. Itoh, Localization of c-di-GMP-binding protein with the linear terminal complexes of *Acetobacter xylinum*. *J. Bacteriol.* **183**, 5668–5674 (2001).
29. A. M. Richter, A. Possling, N. Malysheva, K. P. Yousef, S. Herbst, M. Kleist, R. Hengge, Local c-di-GMP signaling in the control of synthesis of the *E. coli* biofilm exopolysaccharide pEtN-cellulose. *J. Mol. Biol.* **432**, 4576–4595 (2020).
30. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
31. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
32. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *PHENIX*: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
33. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

34. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
35. S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
36. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
37. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
38. E. Kandiah, T. Giraud, A. de Maria Antolinos, F. Dobias, G. Effantin, D. Flot, M. Hons, G. Schoehn, J. Susini, O. Svensson, G. A. Leonard, C. Mueller-Dieckmann, CM01: A facility for cryo-electron microscopy at the European Synchrotron. *Acta Crystallogr. D Struct. Biol.* **75**, 528–535 (2019).
39. B. Y. Matsuyama, P. V. Krasteva, M. V. A. S. Navarro, Isothermal titration calorimetry to determine apparent dissociation constants (K_d) and stoichiometry of interaction (n) of C-di-GMP binding proteins. *Methods Mol. Biol.* **1657**, 403–416 (2017).
40. B. Y. Matsuyama, P. V. Krasteva, C. Baraquet, C. S. Harwood, H. Sondermann, M. V. A. S. Navarro, Mechanistic insights into c-di-GMP-dependent control of the biofilm regulator FleQ from *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E209–E218 (2016).

Acknowledgments: We are grateful to all current and former members of the SBB group and especially to A. Thomas-Collignon, M. Poulain, G. Shajepal, and L. Torres-Sanchez for technical assistance and/or work peripheral to the project; to the I2BC Virology department for provided equipment and lab space; to Y. Yamaichi, V. Lioy, J. M. Ghigo, S. Létoffé, J. Bernal-Bayard, and H. Sondermann for providing expression vectors and bacterial strains; to members of the Reyes, Innis, Fronzes, and Hashem laboratories for discussions; and to A. Bezault, E. Kandiah, C. Dian, and the beamline scientists at Soleil for data collection assistance. **Funding:** This project received funding from the ERC Executive Agency under grant agreement 757507—BioMatrix-ERC-2017-StG (to P.V.K.) and was also supported by the I2BC, the IECB, the CNRS, an ATIP-Avenir grant (to P.V.K.), and a Université de Bordeaux IDEX Junior Chair grant (to P.V.K.). The work has also benefited from the I2BC platforms PIM and cryo-EM, which are supported by FRISBI (ANR-10-INBS-05), the IECB cryo-EM platform and Elsa the Talos Arctica, the Soleil synchrotron, and the ERSF. **Author contributions:** P.V.K. conceived the project. P.V.K., W.A., S.Z., and M.C. designed, performed, and optimized the experimental procedures. P.V.K., S.Z., and W.A. analyzed the data. S.R. contributed to the crystallographic data processing and with useful discussions, and P.V.K. wrote the paper with feedback from coauthors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Refined structural models and electron density maps have been deposited in the electron microscopy and protein databanks. Assigned accession codes are as follows: emd-10799 and pdb-6yg8 for the $BcsB^{Peri}$ pentamer; emd-11356 for the $BcsB^{FL}$ hexadecamer; emd-11836 for the BcsRQAB assembly and Bcs macrocomplex; pdb-codes 6yar, 6yay, 6yb3, and 6yb5 for the $BcsRQ^{His}$ complexes; pdb-6ybu for the $BcsRQE^{349-523}$ complex; and pdb-6ybb for the $BcsRQ^{R156E}E^{217-523}$ complex. Additional requests may be addressed to the corresponding author.

Submitted 13 July 2020
Accepted 8 December 2020
Published 27 January 2021
10.1126/sciadv.abd8049

Citation: W. Abidi, S. Zouhir, M. Caleechurn, S. Roche, P. V. Krasteva, Architecture and regulation of an enterobacterial cellulose secretion system. *Sci. Adv.* **7**, eabd8049 (2021).

Figure S1. Quaternary structure of BcsB$^{E.coli}$. (**A**) Views of a refined BcsB$^{E.coli}$ pentamer against the experimental cryo-EM density following local refinement within the assembled Bcs macrocomplex. (**B**) Surface map coloring according to FSC-based local resolution estimations, color key in angstroms. (**C**) FSC plots and average resolution estimations for the locally refined BcsB pentamer at the gold-standard FSC cutoff of 0.143. Cryo-EM data analysis was carried out in cryoSPARC v2, atomic model building and refinement in Phenix, map visualization in UCSF Chimera. (**D**) A modeled decameric BcsB$^{E.coli}$ superhelix based on the experimentally observed inter-subunit contacts presented in (**A**). The stacked luminal loops are shown in surface representation, the rest of each BcsB subunit is shown as cartoon. Calculations of the rotational and translational displacements between the first and last protomer, as well as structure visualization were carried out in PyMol. (**E**) Elution profile of purified full-length BcsB^{His} subjected to size-exclusion chromatography (SEC) coupled with non-aggregative, on-column crosslinking. SEC column: Superose® 6 Increase 10/300 GL (Cytiva). (**F**) SDS-PAGE analysis of the IMAC elution fraction of purified full-length BcsB^{His} (lane 2) and the SEC-elution fractions of the same sample following mild on-column crosslinking and gel filtration, as in (**E**). (**G**) Comparison of BcsB^{peri} packing observed in the structure of IMAC-purified full-length BcsB^{His} (rainbow, one of two experimentally observed octamers presented) vs. BcsB^{peri} packing as observed in the Bcs macrocomplex crown (grey, an octamer model based on inter-subunit contacts presented in (**A**)). The two octamers were aligned onto the first BcsB copy, calculations of the rotational and translational displacements between the last protomers of each assembly were carried out in PyMol.

Figure S2. Sequence and structure comparison of BcsB$^{E.coli}$ and BcsB$^{R.sphaeroides}$. (**A**) Fold comparison of the four periplasmic BcsB domains with emphasis on key motifs participating in inter-subunit β-strand-mediated backbone interactions. (**B**) Sequence and secondary structure alignment of the two homologs based on a published crystal structure of *R. sphaeroides* (pdb code: 4p00) and the cryo-EM data presented here. (**C**) A close-up view of the sequence-independent inter-subunit β-sheet complementation. (**D**) Sequence alignment of the key interface β-strand regions among enterobacteria. (**E**) Corresponding consensus logos and secondary structure predictions. Fold comparison was visualized in PyMol. Sequence alignments were performed in ClustalΩ and visualized in Jalview. Secondary structure prediction was carried out on the full-length enterobacterial BcsB consensus sequence in JPred. Consensus logos were generated in WebLogo.

Figure S3. Structure-function analyses of the BcsRQ complex. (**A**) Secondary structure topology of the BcsR (grey gradient) and BcsQ (rainbow) subunits. (**B**) Interface analysis of the BcsRQ complex. Surface areas and free energy gains were calculated in PDBePISA. (**C**) Close-up views of the BcsR:BcsQ interface. (**D**) Pre-catalytic nucleotide-bound 'tango' dimers of diverse SIMIBI proteins and their cellular functions. (**E**) SDS-PAGE analysis of IMAC elution fractions of Bcs^{His}RQ*-BcsE$^{217-523}$ complexes with BcsQ mutants presented in this study. (**F**) SEC elution

profiles and SDS-PAGE analysis of the elution peaks for BcsRQ* complexes containing a subset of BcsQ variants showing significant phenotypic effects on cellulose secretion. The $BcsQ^{R156E-N152D}$ mutant was unstable following IMAC elution and did not yield stable BcsRQ complex upon gel filtration. (**G**) ATPase activity of the purified BcsRQ* complexes visualized by a colorimetric malachite green-based assay (MAK113, Sigma-Aldrich). The active ATPase $FleQ^{T149E}$ was used as a positive control. (**H**) Nucleotide loading states of purified BcsRQ* complexes as detected by protein complex concentration, protein precipitation and reversed-phase HPLC analysis of the liberated nucleotides. Elution times of stably bound nucleotides concentrated together with the protein complexes are compared to those of individually injected ADP and ATP standards. (**I**) SEC-elution and ATPase activity profiles of purified $Bcs^{His}RQ^{WT}$-$BcsE^{217-523}$ and $Bcs^{His}RQ^{R156E}$-$BcsE^{217-523}$ heterocomplexes. Left: representative gel filtration profiles on a Superdex® 200 Increase 10/300 GL column (Cytiva). SEC buffer: 120 mM NaCl, 20 mM HEPES pH 8.0 and 2 mM DTT. Right: ATPase activity visualized as in (**G**).

Figure S4. Sequence alignments of enterobacterial cellulose secretion regulators. Alignments of representative BcsQ (**A**), BcsR (**B**) and BcsE (**C**) homologs with key residues, sequence motifs, secondary structure elements and domain organization as identified. Blue color intensity/darkness correlates with sequence conservation. The sequence alignment was generated in ClustalΩ and visualized with Jalview.

Figure S5. Surface conservation mapping of Bcs subunits and assemblies. Left, surface representation (PyMol) with domain color-coding as in the rest of the manuscript. Right, surface conservation presented as cyan (0%) – maroon (100%) gradients (UCSF Chimera).

Figure S6. BcsE-BcsQ interactions and crystal structure of the BcsRQ-$BcsE^{GGDEF*}$ complex. (**A**) Crystal structure of the BcsRQ-$BcsE^{349-523}$ complex. BcsQ is colored in green and blue, BcsR in pink and yellow, the GGDEF* domain in dark red, the I-site in cyan, the interstitial helix in tan, and C-terminal BcsE tail ($BcsE^{CT}$) in purple. ATP and c-di-GMP are shown as sticks, Mg^{++} ions as spheres. (**B**) Close-up views of the interface between the $BcsE^{GGDEF*}$ domain and the C-proximal BcsQ region. (**C**) A close-up view of the $BcsE^{CT}$: BcsQ interface. Interface BcsE residues are indicated on the left, BcsQ residues on the right. (**D**) Structural motifs in the $BcsE^{349-523}$ construct and interface analysis of the binary BcsE : BcsQ interaction as calculated in PDBePISA. (**E**) Dependence of BcsE-BcsQ complex formation on both the $BcsE^{GGDEF*}$ and $BcsE^{CT}$ components. (**F**) Head-to-head crystallographic packing of the BcsRQ-$BcsE^{349-523}$ complex. (**G**) C-di-GMP coordination in the $BcsRQE^{349-523}$: c-di-GMP crystals. Left, an (|Fo|-|Fc|) partial electron density map calculated from a model prior to inclusion of the dinucleotide and contoured at 3σ. Right, I-site bound c-di-GMP coordination by BcsE, BcsR, and BcsQ residues from a symmetry related complex. (**H-I**) Summary of c-di-GMP coordinating motifs observed in this study as mapped onto the BcsE domain architecture (**H**) and a composite $BcsRQE^{REC*-GGDEF*}$ assembly featuring a 'splayed' $BcsE^{217-523}$ module (**I**).

Figure S7. Cryo-EM data analysis workflow for the BcsRQAB complex. (**A**) Particle curation strategy for BcsRQAB focused refinement within the assembled Bcs macrocomplex. (**B**) Surface coloring of the unsharpened BcsRQAB reconstruction according to FSC-based local resolution estimations, color key in angstroms. (**C**) FSC plots and average resolution estimations for the locally refined BcsRQAB assembly at the gold-standard FSC cutoff of 0.143. (**D**) A snapshot of the modeled BcsA and BcsB transmembrane regions as fitted in the experimental electron density.

Table S1. Cryo-EM data collection and refinement statistics for the BcsB pentamer

Table S2. Crystallographic data collection and refinement statistics for the BcsRQ complexes

Table S3. Crystallographic data collection and refinement statistics for the BcsRQE* complexes

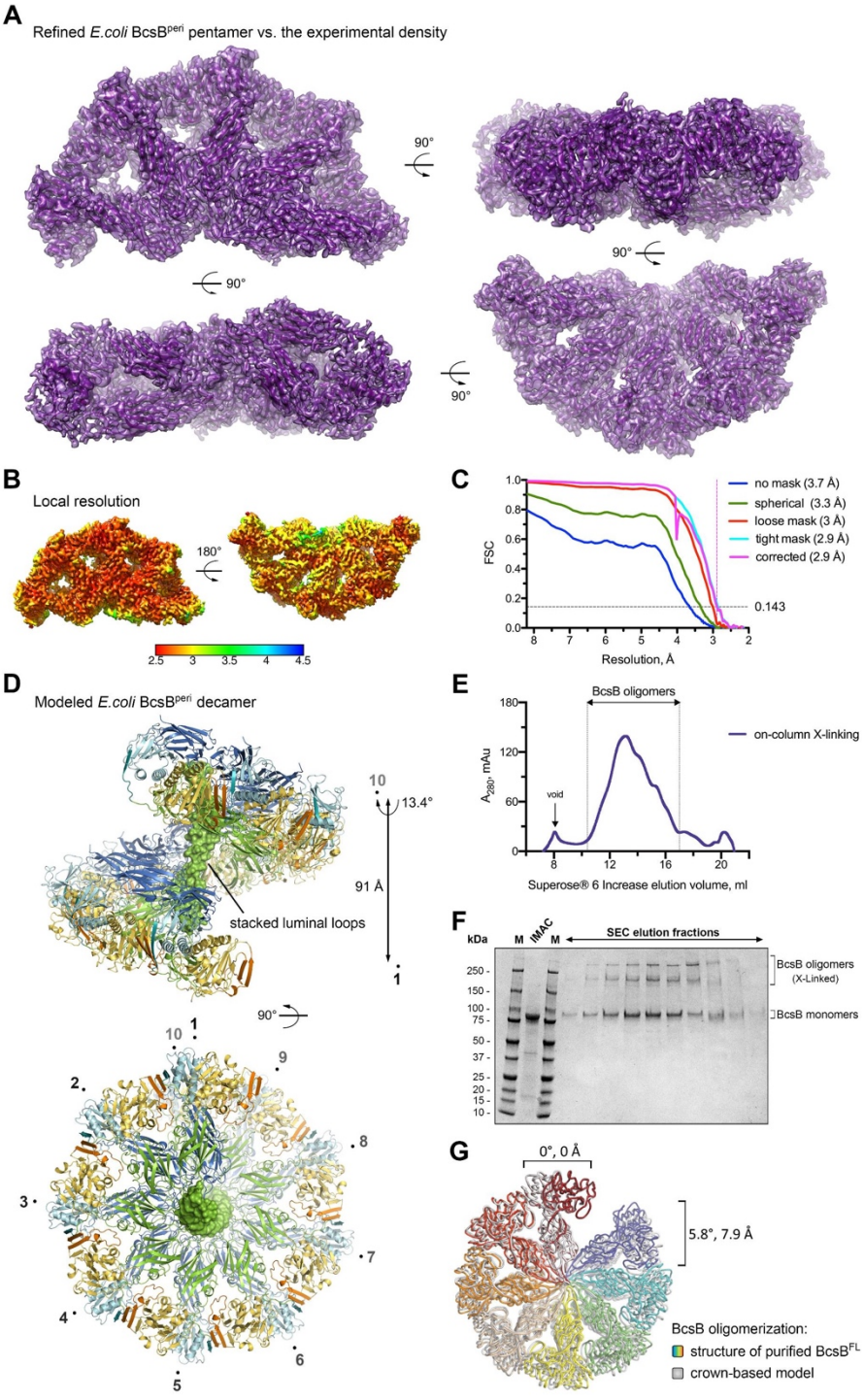

A
Refined E.coli BcsB^peri pentamer vs. the experimental density
90°
90°
90°
90°
B
Local resolution
180°
2.5
3
3.5
4
4.5
C
FSC
1.0
0.8
0.6
0.4
0.2
0.0
8
7
6
5
4
3
2
Resolution, Å
no mask (3.7 Å)
spherical (3.3 Å)
loose mask (3 Å)
tight mask (2.9 Å)
corrected (2.9 Å)
0.143
D
Modeled E.coli BcsB^peri decamer
10
13.4°
91 Å
1
stacked luminal loops
90°
10
1
9
2
8
3
7
4
6
5
E
BcsB oligomers
A280, mAu
180
120
60
0
void
8
12
16
20
Superose® 6 Increase elution volume, ml
on-column X-linking
F
kDa
M
IMAC
M
SEC elution fractions
250 -
150 -
100 -
75 -
50 -
37 -
25 -
20 -
15 -
10 -
BcsB oligomers
(X-Linked)
BcsB monomers
G
0°, 0 Å
5.8°, 7.9 Å
BcsB oligomerization:
structure of purified BcsB^FL
crown-based model

Figure S1

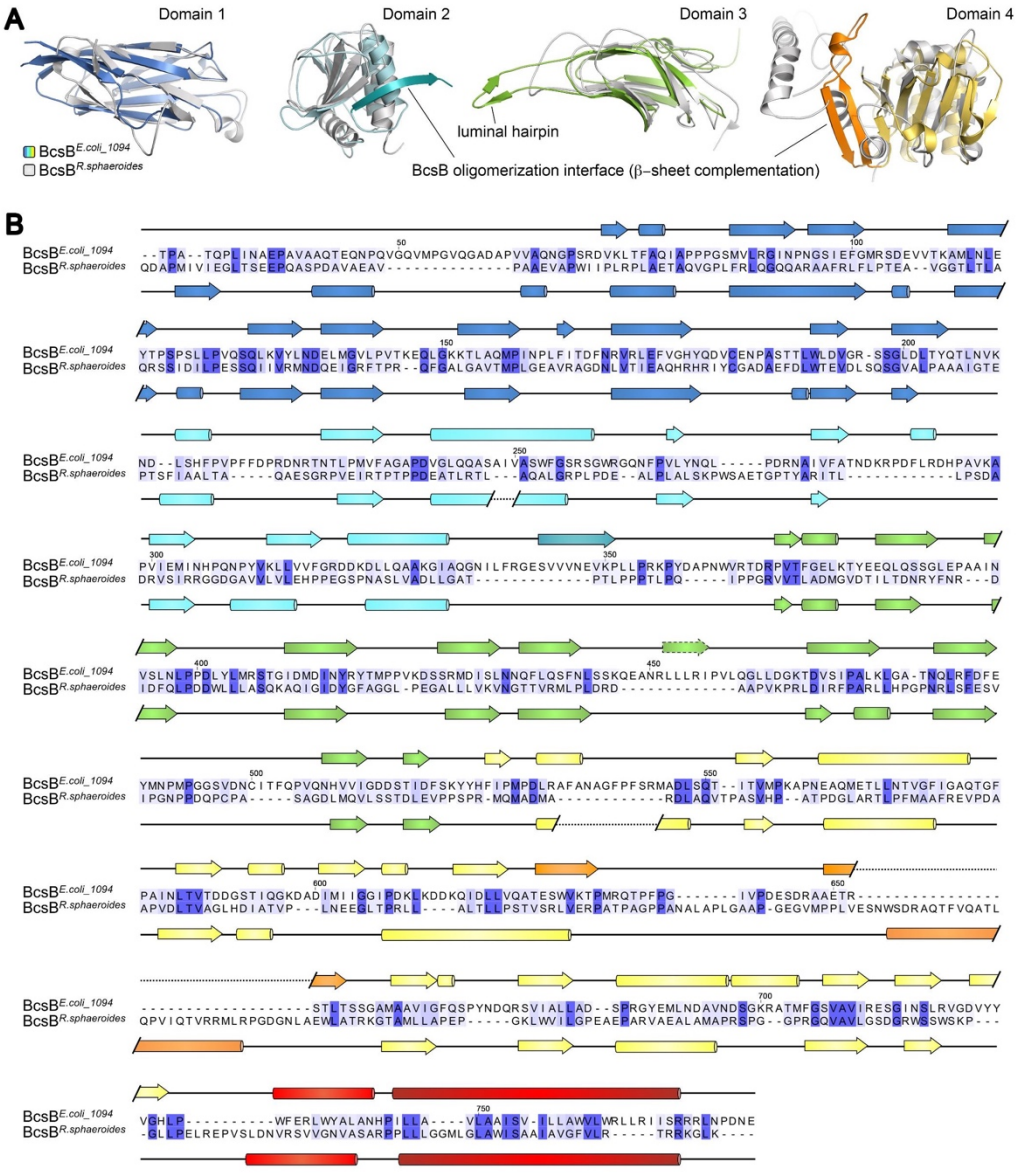

A
Domain 1
Domain 2
Domain 3
Domain 4
BcsB E.coli_1094
BcsB R.sphaeroides
luminal hairpin
BcsB oligomerization interface (β–sheet complementation)
B
C
D
E

Figure S2

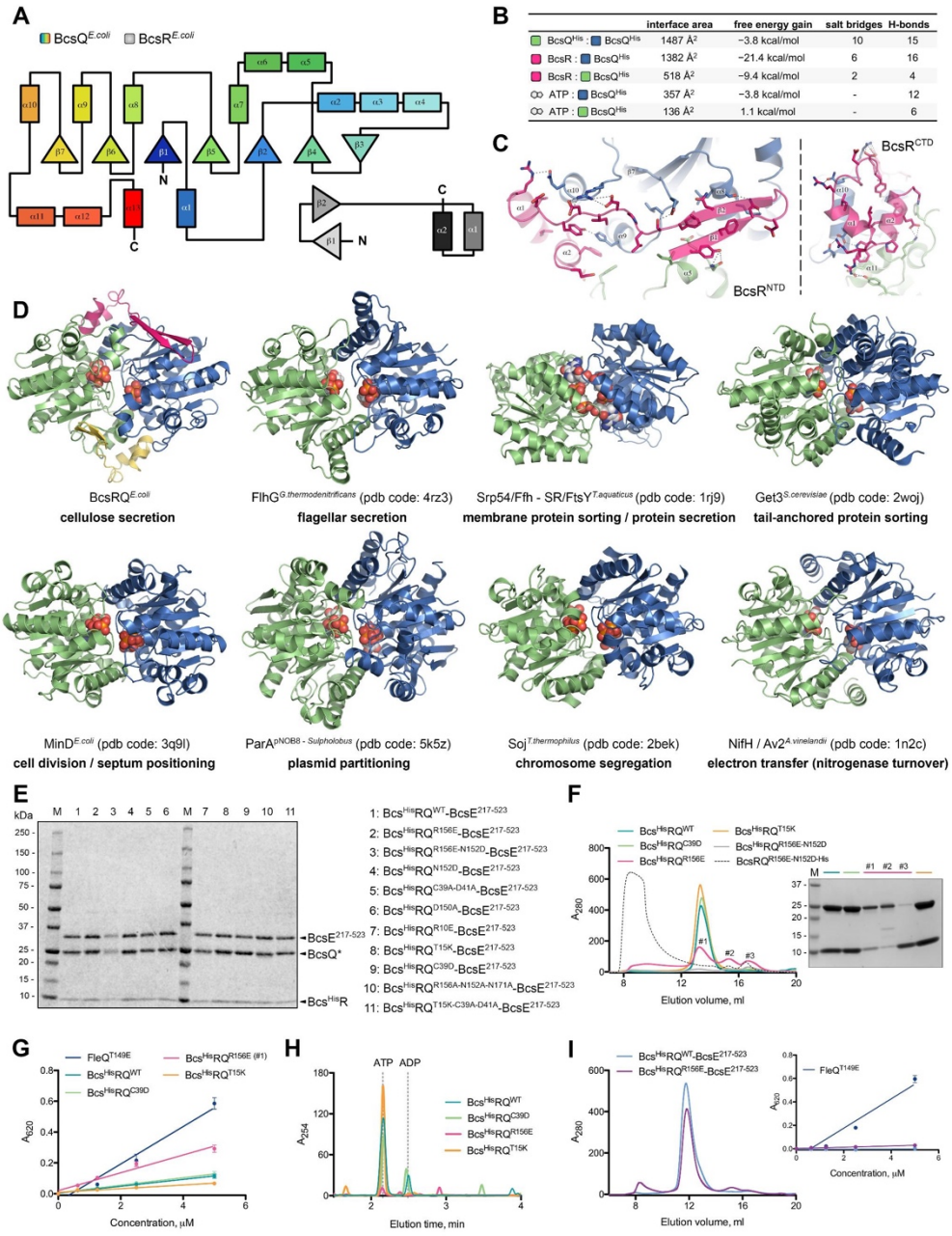

A
BcsQ^E.coli
BcsR^E.coli
B
interface area
free energy gain
salt bridges
H-bonds
BcsQ^His : BcsQ^His 1487 Å² −3.8 kcal/mol 10 15
BcsR : BcsQ^His 1382 Å² −21.4 kcal/mol 6 16
BcsR : BcsQ^His 518 Å² −9.4 kcal/mol 2 4
ATP : BcsQ^His 357 Å² −3.8 kcal/mol - 12
ATP : BcsQ^His 136 Å² 1.1 kcal/mol - 6
C
BcsR^CTD
BcsR^NTD
D
BcsRQ^E.coli
cellulose secretion
FlhG^G.thermodenitrificans (pdb code: 4rz3)
flagellar secretion
Srp54/Ffh - SR/FtsY^T.aquaticus (pdb code: 1rj9)
membrane protein sorting / protein secretion
Get3^S.cerevisiae (pdb code: 2woj)
tail-anchored protein sorting
MinD^E.coli (pdb code: 3q9l)
cell division / septum positioning
ParA^pNOB8 - Sulpholobus (pdb code: 5k5z)
plasmid partitioning
Soj^T.thermophilus (pdb code: 2bek)
chromosome segregation
NifH / Av2^A.vinelandii (pdb code: 1n2c)
electron transfer (nitrogenase turnover)
E
kDa
1: Bcs^His RQ^WT-BcsE^217-523
2: Bcs^His RQ^R156E-BcsE^217-523
3: Bcs^His RQ^R156E-N152D-BcsE^217-523
4: Bcs^His RQ^N152D-BcsE^217-523
5: Bcs^His RQ^C39A-D41A-BcsE^217-523
6: Bcs^His RQ^D150A-BcsE^217-523
7: Bcs^His RQ^R10E-BcsE^217-523
8: Bcs^His RQ^T15K-BcsE^217-523
9: Bcs^His RQ^C39D-BcsE^217-523
10: Bcs^His RQ^R156A-N152A-N171A-BcsE^217-523
11: Bcs^His RQ^T15K-C39A-D41A-BcsE^217-523
BcsE^217-523
BcsQ*
Bcs^His R
F
Elution volume, ml
G
Concentration, μM
H
ATP ADP
Elution time, min
I
Elution volume, ml
Concentration, μM

Figure S3

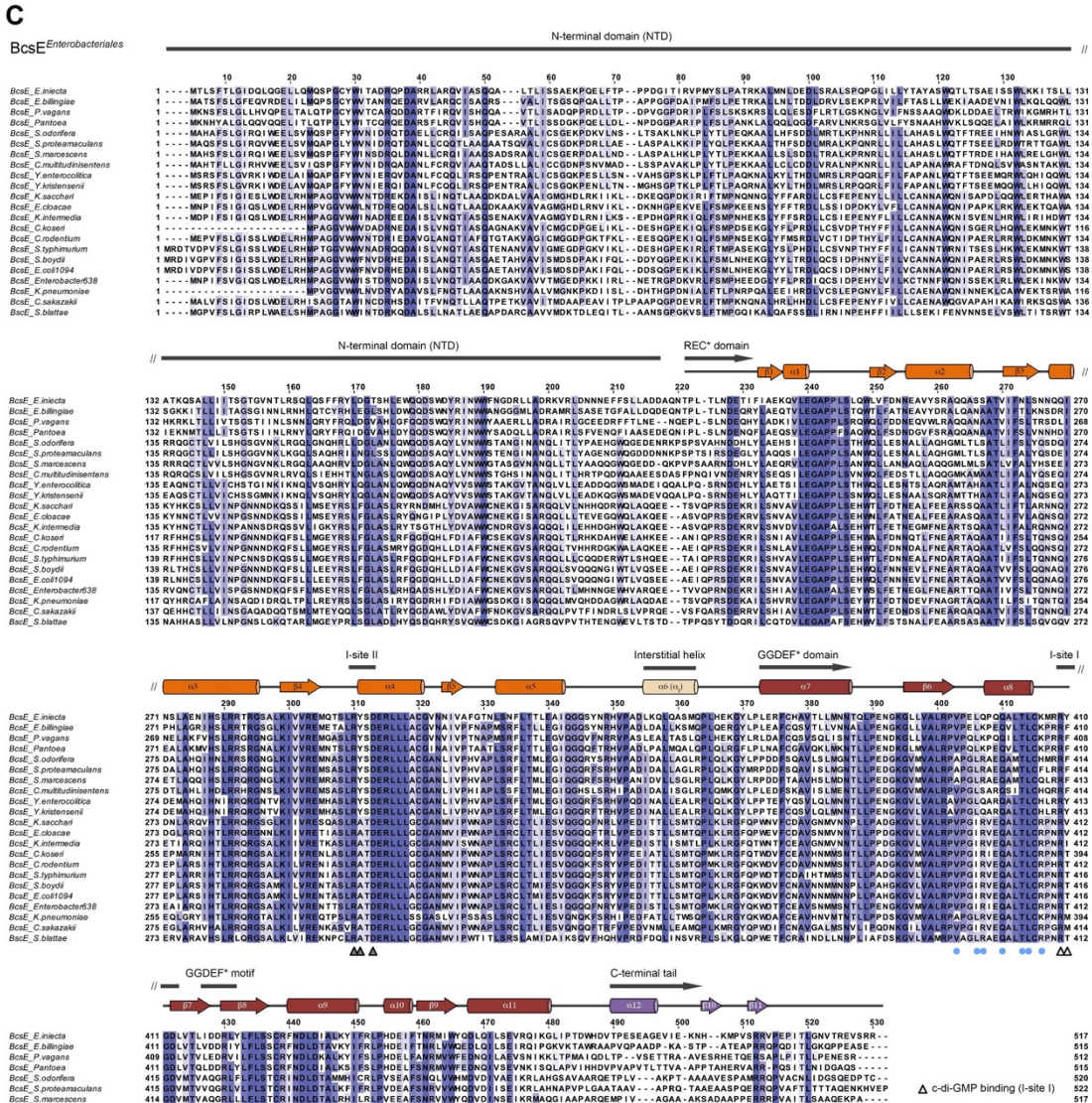

C
BcsE Enterobacteriales
N-terminal domain (NTD)
REC* domain
I-site II
Interstitial helix
GGDEF* domain
I-site I
GGDEF* motif
C-terminal tail
c-di-GMP binding (I-site I)
c-di-GMP binding (I-site II)
c-di-GMP binding (X-tal)
BcsQ binding
BcsR binding

Figure S4

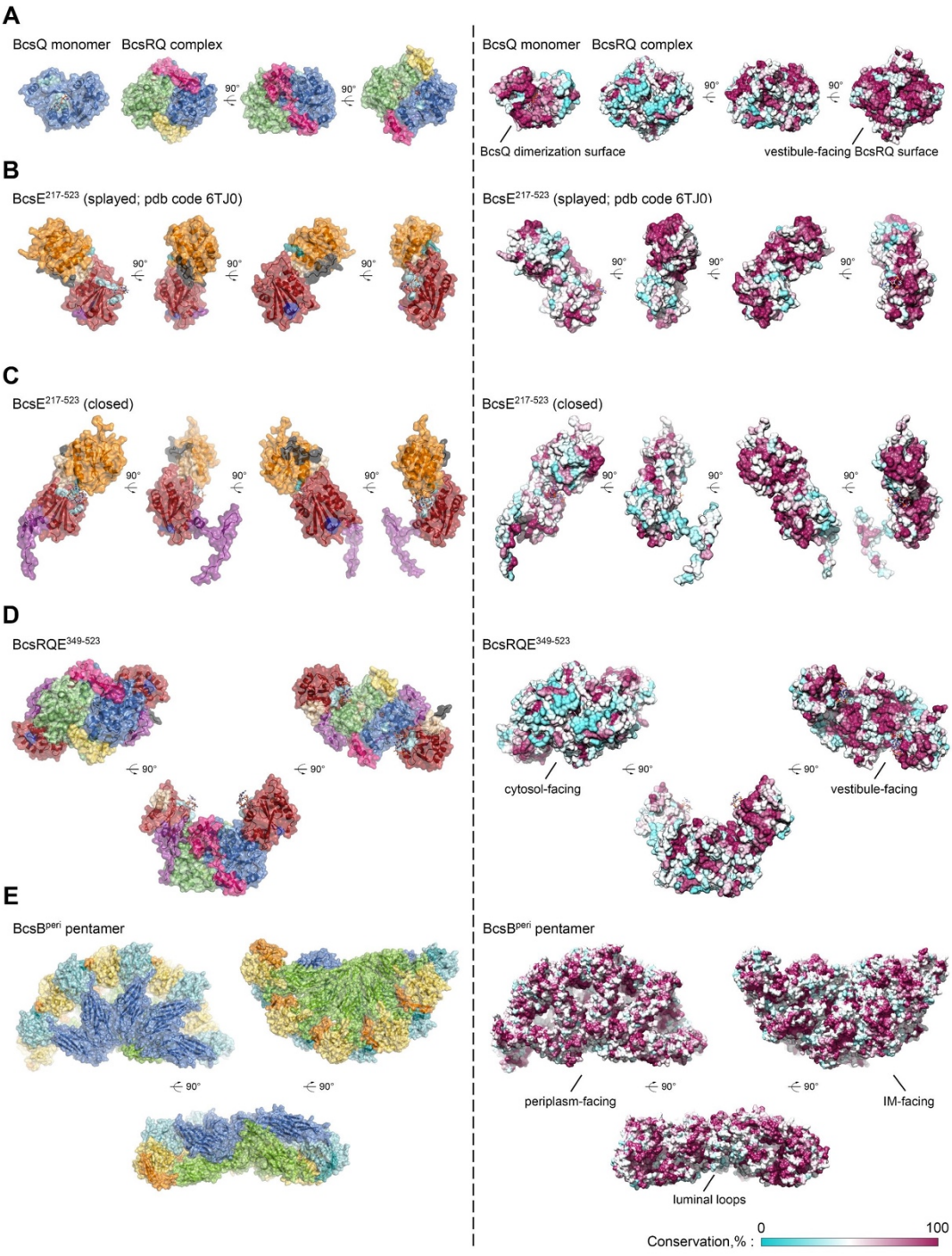

A
BcsQ monomer
BcsRQ complex
BcsQ monomer
BcsRQ complex
BcsQ dimerization surface
vestibule-facing BcsRQ surface
B
BcsE217-523 (splayed; pdb code 6TJ0)
BcsE217-523 (splayed; pdb code 6TJ0)
C
BcsE217-523 (closed)
BcsE217-523 (closed)
D
BcsRQE349-523
BcsRQE349-523
cytosol-facing
vestibule-facing
E
BcsBperi pentamer
BcsBperi pentamer
periplasm-facing
IM-facing
luminal loops
Conservation,% :
0
100
90°

Figure S5

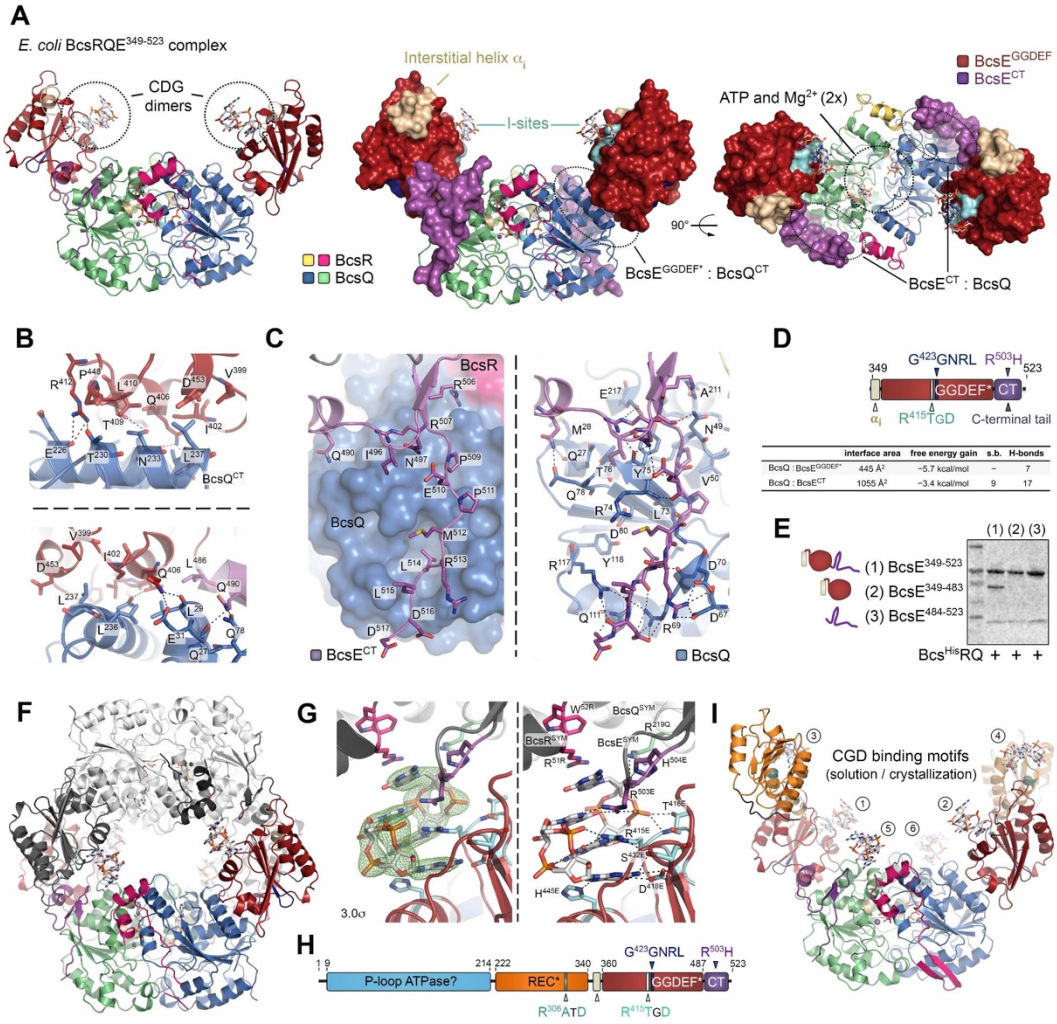

A
E. coli BcsRQE349-523 complex
CDG dimers
Interstitial helix α1
I-sites
BcsR
BcsQ
BcsEGGDEF* : BcsQCT
90°
ATP and Mg2+ (2x)
BcsEGGDEF
BcsECT
BcsECT : BcsQ
B
BcsQCT
C
BcsR
BcsQ
BcsECT
BcsQ
D
G423GNRL R503H
349
523
GGDEF*
CT
α1
R415TGD
C-terminal tail
interface area
free energy gain
s.b.
H-bonds
BcsQ : BcsEGGDEF*
445 Å2
~5.7 kcal/mol
~
7
BcsQ : BcsECT
1055 Å2
~3.4 kcal/mol
9
17
E
(1) (2) (3)
(1) BcsE349-523
(2) BcsE349-483
(3) BcsE484-523
BcsHisRQ + + +
F
G
3.0σ
I
CGD binding motifs
(solution / crystallization)
H
1
9
214
222
340
360
487
523
P-loop ATPase?
REC*
GGDEF*
CT
G423GNRL
R503H
R306ATD
R415TGD

Figure S6

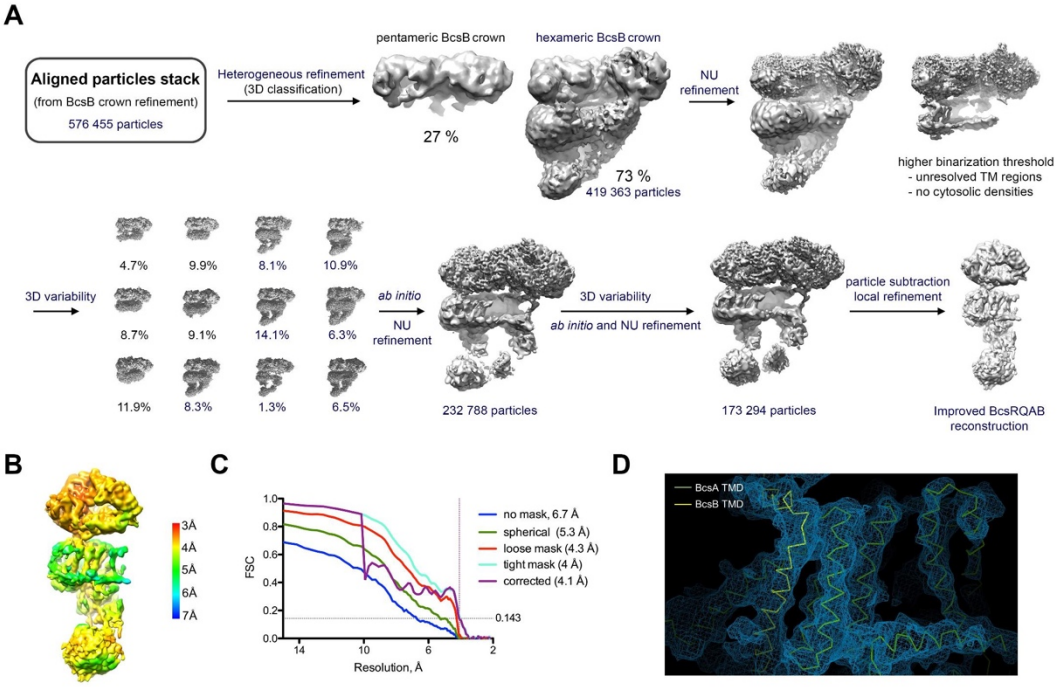

A
Aligned particles stack
(from BcsB crown refinement)
576 455 particles
Heterogeneous refinement
(3D classification)
pentameric BcsB crown
27 %
hexameric BcsB crown
73 %
419 363 particles
NU
refinement
higher binarization threshold
- unresolved TM regions
- no cytosolic densities
3D variability
4.7%
9.9%
8.1%
10.9%
8.7%
9.1%
14.1%
6.3%
11.9%
8.3%
1.3%
6.5%
ab initio
NU
refinement
232 788 particles
3D variability
ab initio and NU refinement
173 294 particles
particle subtraction
local refinement
Improved BcsRQAB
reconstruction
B
3Å
4Å
5Å
6Å
7Å
C
FSC
1.0
0.8
0.6
0.4
0.2
0.0
0.143
14
10
6
2
Resolution, Å
no mask, 6.7 Å
spherical (5.3 Å)
loose mask (4.3 Å)
tight mask (4 Å)
corrected (4.1 Å)
D
BcsA TMD
BcsB TMD

Figure S7

Table S1 | Cryo-EM data collection and refinement statistics for BcsB*	
Protein sample	Bcs macrocomplex ($Bcs^{His}RQA^{HA\text{-}FLAG}B^{Strep}EF(G)$)
Data collection	
Microscope	Titan Krios (Thermo Fisher Scientific) CM01 (ESRF)
Voltage (kV)	300
Camera	Gatan K2 Summit
Energy filter	GIF Quantum LS
Pixel size ($Å^2$)	1.05258
Collection mode	electron counting
Total electron dose ($e^-/Å^2$)	48
Movies	9 129
Frames/movie	40 (dataset I), 50 (dataset II)
Defocus range (μm)	–0.75 to –2.75
Single particles	576 455
Average map resolution	2.9 Å
Symmetry	none (C1)
Sharpening B-factor	84.4
Atomic model refinement	
* Refined macromolecular assembly	$BcsB^{peri}$ pentamer
Number of chains / residues	5 / 3132
Number of non-hydrogen atoms	
Proteins	24 330
B-factors	
Proteins (min/max/mean)	38.34/131.30/71.56
R.m.s.d.	
Bond lengths (Å)	0.005
Bond angles (deg)	0.602
Ramachandran plot	
Favored (%)	95.29
Allowed (%)	4.71
Rotamer outliers	0
Cβ outliers	0
Peptide plane (%)	
Cis proline/general	3.6 / 0
Twisted proline/general	0 / 0
CaBLAM outliers (%)	2.38
All-atom clashscore	7.02
Molprobity score	1.71

Table S2 | Crystallographic data collection and refinement statistics for the BcsRQ complex *

	$BcsRQ^{His}$	$BcsRQ^{His}$	$BcsRQ^{His}$	$BcsRQ^{His}$
Crystallized protein	SeMet	SeMet	Native	Native
Data Collection				
Space group	P 1 21 1	P 1 21 1	P 21 21 21	P 1 21 1
Cell dimensions				
a, b, c (Å)	56.1, 78.4, 65.2	57.1, 78.8, 66.6	59.1, 73.2, 140.5	57, 79, 66.8
α, β, γ (deg)	90, 97.41, 90	90, 99.28, 90	90, 90, 90	90, 99.26, 90
Wavelength	0.9793	0.9786	0.9800	0.9793
Resolution (Å)	45.2-1.89 (1.97-1.9)	46.6-2.09 (2.16-2.09)	45.2-1.59 (1.65-1.59)	46.7-1.59 (1.65-1.59)
R-merge	9.5% (116.9%)	9.5% (135%)	9.4% (192.6%)	8.5% (127%)
R-meas	10% (126.6%)	10.15% (146.7%)	9.9% (201.9%)	9.7% (156.7%)
R-pim	3.1% (47.6 %)	3.6% (56.2%)	2.7% (59.3%)	4.7% (90%)
Mean I/σ(I)	14.4 (1.2)	13.4 (1.35)	17.2 (1.05)	11.4 (0.7)
Completeness (%)	94% (69.66%)	98.7% (92.7%)	99% (91%)	95.7% (74.5%)
Multiplicity	10.1 (6.7)	7.6 (6.4)	13.2 (11.1)	3.8 (2.5)
$CC_{1/2}$	99.9 (72.7)	99.8 (68.2)	99.9 (45.9)	99.8 (41.5)
Refinement				
Unique reflections	41 607	34 175	81 658	74 521
R-work	17.8%	19.2%	16.0%	18.5%
R-free	21.8%	22.3%	18.4%	20.9%
Number of non-hydrogen atoms				
Proteins	4 328	4 367	4 592	4 596
Ligands	64	64	64	64
B-factors				
Proteins	52.07	67.37	28.93	45.65
Ligands	34.71	48.03	21.88	29.36
R.m.s.d.				
Bond lengths (Å)	0.006	0.002	0.006	0.006
Bond angles (deg)	0.79	0.51	0.78	0.81
Ramachandran plot				
Favored (%)	98.9	98.5	98.5	98.9
Allowed (%)	1.2	1.5	1.5	1.1
Molprobity score	1.01	1.02	0.94	1.08
Crystallization condition ligand	AppCp (ACP)	ADP	ADP	none
Ligand in structure	ATP / ACP **	ATP	ATP	ATP

** Statistics for the highest-resolution shell are shown in parentheses.*

*** ACP binding cannot be ruled out as the compound was present in the crystallization condition and the electron density is indistinguishable from that of ATP*

**** Ligands shown in the |Fo| - |Fc| maps calculated from models prior to inclusion of the ligands and contoured at 2σ*

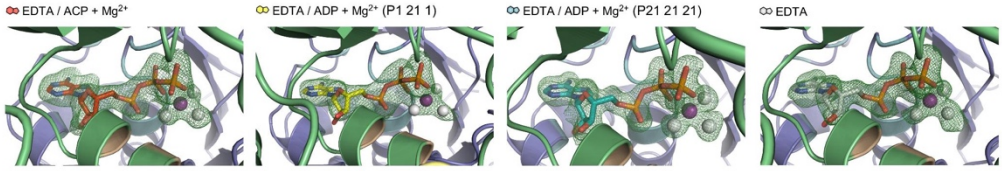


Table S3 | Crystallographic data collection and refinement statistics for BcsRQE* heterocomplexes		
	$BcsRQ^{R156E}$-$BcsE^{217-523}$	**BcsRQ-$BcsE^{349-523}$**
Crystallized protein	Native	Native
Data Collection		
Space group	P 21 21 21	P 1 21 1
Cell dimensions		
a, b, c (Å)	61.4, 169.6, 177.6	108, 72.6, 161.3
α, β, γ (deg)	90, 90, 90	90, 98.37, 90
Wavelength	0.9801	0.9801
Resolution (Å)	48-2.9 (3.004-2.9)	49-2.49 (2.58-2.49)
R-merge	19.18% (307%)	20.15% (156.5%)
R-meas	19.94% (319%)	21.8% (169.6%)
R-pim	5.8% (85.3%)	8.3% (64.61%)
Mean I/σ(I)	14 (0.78)	11.45 (1.1)
Completeness (%)	99.7% (99.4%)	99% (91.6%)
Multiplicity	13.6	6.9
$CC_{1/2}$	99.9 (52.5)	99.5 (46.9)
Refinement		
Unique reflections	41 967	86 089
R-work	19.96%	18.08%
R-free	22.9%	22.5%
Number of non-hydrogen atoms		
Proteins	8 851	13 823
Ligands	260	496
B-factors		
Proteins	104.58	62.38
Ligands	95.4	53.15
R.m.s.d.		
Bond lengths (Å)	0.004	0.004
Bond angles (deg)	0.68	0.7
Ramachandran plot		
Favored (%)	98.46	98.43
Allowed (%)	1.54	1.57
Molprobity score	1.24	1.09
Crystallization condition ligand	ACP, c-di-GMP	ACP, c-di-GMP
Ligand in structure **	ATP/ACP, c-di-GMP	ATP/ACP, c-di-GMP
Protein : c-di-GMP ratio	1 : 2	1 : 2

** Statistics for the highest-resolution shell are shown in parentheses*

*** ACP binding cannot be ruled out as the compound was present in the crystallization condition and the electron density is indistinguishable from that of ATP*

5 General discussion

Carbohydrates are arguably the most important energy carrier and the most abundant biological components in all kingdoms of life. Evolutionary speaking, many bacteria dedicate a part of their metabolism to produce a plethora of carbohydrates, depending on both the disposable building blocks and the surrounding environment. Gram-negative and Gram-positive bacteria use their cell factories to secrete various classes of biopolymers and polysaccharides are oftentimes the main class of secreted extracellular substances [1], [2], [6], [7]. The polysaccharides either remain linked to the cell surface a capsular coat or are freely secreted into the extracellular milieu to contribute to the scaffolding of the biofilm matrix. Along the previous chapters, I spot the light on the cellulose as a main biofilm polysaccharide secreted by many bacteria, due to its unique characteristics that contribute to maintain the cell-cell and cell-surface adhesion and establish a highly protective and biocompatible environment for the multicellular communities. Through the present studies, I focused on the multicomponent cellulose secretion machinery in Gram-negative bacteria, particularly the *E. coli*-like Type II cellulose secretion system, which comprises a total of nine subunits (BcsRQABZCEFG) that span from the cytosol to the extracellular space. Along the results chapter, we marked a milestone in the mechanistic understandings for cellulose secretion system assembly and nucleotide-dependent regulation. Before our discoveries, several studies had shown that processive glucose polymerization is performed by the glycosyl transferase domain of BcsA, whose active site is made accessible by binding of dimeric c-di-GMP to an adjacent PilZ domain. The transport through the IM is coupled to the polymerization and is energized by the high-energy phosphoanhydride bonds of the preactivated synthase substrate, UDP-glucose. The nascent polysaccharide chain is subsequently extruded, one molecule at a time, through the IM transport domain of BcsA assisted by the C-terminal tail-anchor of BcsB. The latter partner adopts a donut-shaped periplasmic architecture and is proposed to guide the polysaccharide on its way to the outer membrane secretory component BcsC. Additional studies revealed that the cellulose secreted by *E. coli*-like systems is covalently decorated by pEtN residues in a post synthetic way by the periplasmic subunit BcsG. Another component is the presumably periplasmic endonuclease BcsZ, which is suggested to be required for degradation of cellulose in case it remains accumulated in the periplasm, for cleavage and restructuring of nascent glucan chains to allow microfibril formation to take place outside the cell, or for release of the polysaccharide from the cell surface. On the intracellular side of the bacterial membrane, the regulatory subunits BcsR, BcsQ and BcsE remained enigmatic. When we started our investigations, we had little information available about these components. We knew that *E. coli*–like cellulose secretion *in vivo* is absolutely dependent on the presence of the two small

cytosolic proteins BcsR and BcsQ, and is boosted the third cytosolic protein BcsE, as well as by the short membrane-embedded polypeptide BcsF. A study by Krasteva and colleagues revealed that most of the IM and cytosolic Bcs components (BcsRQABEF) form a megadalton-sized secretory macrocomplex with a multimeric, layered, and asymmetric architecture. However, the structures, localization, and functional roles of the individual subunits remained largely unresolved. Through our two related studies (article 1 and article 2), we provided structural and functional insights into the individual subunits, together with the integral complex assembly. Based on crystallographic and functional data, we showed that BcsE actually features a tripartite architecture. In it, an N-terminal catalytically incompetent ATPase-like domain aids BcsE dimerization, participates in BcsF-mediated membrane recruitment of the essential BcsR and BcsQ subunits and interacts with conserved TAC (transcription antitermination complex) components suggesting additional regulatory roles at the gene expression level. Moreover, we showed that the previously hypothesized GIL domain providing a second c-di-GMP sensor to the system is in fact a degenerate receiver–GGDEF domain tande, ($BcsE^{REC*\text{-}GGDEF*}$), where the divergent diguanylate cyclase module binds both c-di-GMP and BcsQ through mutually independent interfaces. Contrasting degrees of sequence conservation between the N-terminal module and the REC-GGDEF tandem, as well as the identification of organisms where the corresponding BcsE parts are encoded by separate genes, point toward multidomain BcsE evolution and function integration via separate gene fusion events. The c-di-GMP–bound $BcsE^{REC*\text{-}GGDEF*}$–$BcsRQ^{R156E}$ complex brought up more surprises, as the BcsE variant in the multicomponent complex adopts a strikingly different conformation from the structure of c-di-GMP–bound $BcsE^{REC*\text{-}GGDEF*}$. While the canonical I-site RXXD motif on the catalytically incompetent diguanylate cyclase module coordinates a c-di-GMP moiety, the degenerate receiver domain and a so-called ‘interstitial helix’, linking it to the GGDEF module, undergo a 144° rotation and 45Å displacement to contribute a distinct conserved RXXD motif (R^{306}ATD) and coordinate a second intercalated dinucleotide molecule via virtually identical arginine/aspartate-dependent interactions. We further revealed the importance of the REC domain I-site bound to the complexed c-di-GMP, using solution-based isothermal titration calorimetry experiments, where truncated ($BcsE^{GGDEF*}$) or point-mutated BcsE ($BcsE^{REC*\text{-}GGDEF*}$•$A^{306}$ATA) exhibited drastically altered thermodynamic profiles of the ligand binding reactions. Finally, the crystal structure of the $BcsE^{GGDEF*}$–BcsRQ complex showed an additional, crystallographic c-di-GMP–binding interface involving multiple π-stacking and polar interactions with conserved residues from BcsE ($Arg^{503}His^{504}$), BcsR ($Arg^{51}Trp^{52}$) and BcsQ (Arg^{219}). Even though the biological relevance of this third dinucleotide binding site

remains unverified, it is possible that it contributes additional weak interactions within the assembled cellulose secretion machinery, where the BcsERQ complex forms a swaddle-like vestibule around the BcsA PilZ domain and through multi-site c-di-GMP complexation can provide a synthase-proximal pool of circulating dinucleotide for enzyme activation. Regarding the assembled cellulose secretion machinery, we provided nearly atomic resolution data showing that BcsB polymerizes via a ß-sheet complementation mechanism among neighboring subunits and propose that polymerization would be simultaneously limited by the polymerization-induced membrane curvature due to BcsB's superhelicity and C-terminal anchors, the intrinsic membrane surface tension and the protein-protein interactions with the rest of the secretion system components. The integral assembled Bcs secretion macrocomplex thus contains a periplasmic crown of up to six BcsB copies, a single BcsA synthase, a likely dimer of BcsF inner membrane peptides and the $BcsR_2Q_2E_2$ cytosolic regulatory complex, which supports the synthase through BcsRQ-$BcsA^{PilZ}$ interactions on one side and is anchored to the inner membrane through $BcsE^{NTD}$-BcsF interactions on the other. Placing the cellulose machinery in a wider picture, the Bcs system belongs to the larger family of synthase-dependent exopolysaccharide secretion systems. Members of the latter include the alginate, the PNAG (poly-N-acetylglucosamine) and the Pel exopolysaccharide biosynthesis systems, many of which are found in both Gram-positive and Gram-negative bacteria. The common trait among the synthase-dependent family is the processive coupling of the polymerization with the export of the glycan chain. Opposingly to the Bcs system, the glycosyl-transferase and inner membrane translocation can be performed by separate components, as for example the PelF and PelG subunits of the Pel systems. In line with the Bcs system, in which the glycosyl transfer and the extrusion through the IM are carried out by the single subunit BcsA, alginate and PNAG systems employ Alg8 and PgaC respectively (Figure 15). Another common feature imposes itself among the members of the synthase-dependent exopolysaccharide family is the c-di-GMP dependent activation engaging one or more dinucleotide-sensing protein modules such as $Alg44^{PilZ}$, $PelD^{GGDEF}$ domains and a composite binding site at the PgaC-PgaD interface for the alginate, Pel and PNAG systems, respectively. Additionally, AlgK, PelE, PelB and PgaA, on one hand, and AlgL, PelA, PgaB on the other, substitute the TPR-rich BcsC modules and the hydrolase BcsZ, respectively. Within the last years, numerous studies uncovered the structures of different modules of the latter synthase-dependent exopolysaccharides systems, whereas genomic studies allowed to establish comparisons-based parallels in a wide range of bacteria. Nevertheless, the intricate mechanisms for secretion systems assembly and biofilm matrix

secretion require further investigations, especially considering a rising biotechnological interest and many already enabled applications of these intriguing biological polymers.

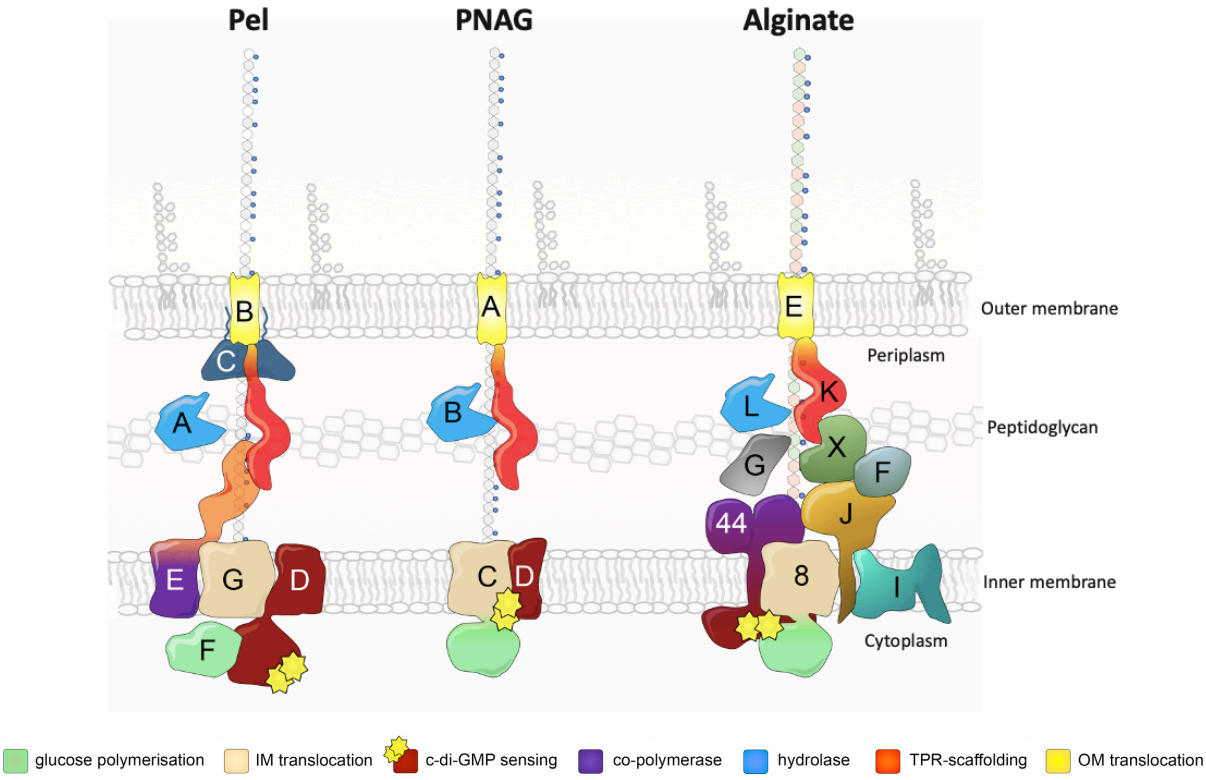


Figure 15. Examples of bacterial exopolysaccharide secretion systems

Examples of three synthase dependent systems. Left, Pel system; middle PNAG (poly-N-acetylglucosamine) system; right, Alginate system. Shown are the subunits and their proposed topologies in Gram-negative bacteria [47].

6 Final conclusion and perspectives

My dissertation work is focused on the study of the bacterial cellulose secretion machineries, with a special attention to the *E. coli*-like Type II Bcs system for pEtN-cellulose secretion. I dedicated my work on bacterial cellulose for two major reasons. The first reason is the contribution of cellulose and exopolysaccharides in general to biofilm formation in many and diverse bacterial species including clinically important infectious agents. Indeed, bacterial biofilms are usually linked to pathogen persistence and antibiotics resistance development in chronic diseases, as in nosocomial infections or in patients with cystic fibrosis, urinary tract infections or medical implants. Indeed, bacterial cellulose secretion provides an excellent model for deciphering the roles of c-di-GMP signaling and synthase-dependent polysaccharide production, which are also the basis for biofilm formation via alternative secretion systems in many medically and economically important human or plant pathogens. The second reason is the real economic and ecological need for the use of the bacterial cellulose in a variety of biomedical, materials science and agricultural biocontrol applications [47].

We know that even though there had been a tremendous amount of work in deciphering bacterial cellulose biosynthesis spanning multiple decades, relevant pieces were still missing from big picture of this important biosynthetic process. Through the present work, I significantly contributed to the understanding of the complex mechanisms governing cellulose secretion. Following a strategy where I combined structural and functional studies, I solved the structures of multiple complexes of the so-far enigmatic regulatory subunits BcsR, BcsQ and BcsE. These revealed an unexpected role for ATP, via sandwiched complexation the essential for secretion BcsRQ complex. Viewed from an energetic perspective, use of UDP-complexed sugars should be sufficient for powering polymerization and nascent cellulose extrusion. Nevertheless, our work demonstrated that both ATP binding and hydrolysis after recruitment to the membrane are necessary for secretion likely through effects on BcsA stability or membrane sorting. Additionally, together with my colleagues I showed that the BcsRQ complex interacts with the second c-di-GMP sensor BcsE to form a heterohexameric complex. I captured multiple crystallized states of the BcsERQ regulatory complex that support multi-site mechanism of c-di-GMP sensing. Put in the context of the assembled Bcs secretion system, where the BcsERQ subunits form a BcsA-surrounding cytosolic vestibule, this c-di-GMP complexation can provide a synthase-proximal pool for processive enzyme activation. Additionally, I showed that the intriguing stoichiometry of the so-called periplasmic crown of the Bcs system is formed via self-driven BcsB polymerization secured by intersubunit ß-sheet complementation and in contrast to a previously existing dogma for equimolar BcsA:BcsB

stoichiometries, we demonstrated a non-canonical $BcsA_1BcsB_6$ complexation likely optimized forBcsA sorting, c-di-GMP vestibule assembly and recruitment of regulatory components such as the pEtN transferase BcsG and the c-di-GMP-metabolizing DgcC/DgcQ and PdeK.

Nevertheless, multiple functional aspects remain to be further investigated. These include the likely activation of ATP hydrolysis at the membrane level, as well as its roles in secretion system stability or assembly; the role of the uncovered by us BcsE interactions with the NusB and NusE factors from the conserved transcription antitermination machinery; the recruitment of the periplasmic cellulose-modifying enzymes BcsZ and BcsG or the functional interactions with c-di-GMP-metabolizing enzymes providing the on/off switch for the system; the mechanisms of polysaccharide crossing through the peptidoglycan layer and recruitment of the outer-membrane secretory component BcsC, as well as the tethering or release of secreted cellulose to or from the cell surface. This said, I hope my work will help guide future studies into this complex and widespread bacterial biosynthetic process.

7 References

[1] H.-C. Flemming, J. Wingender, U. Szewzyk, P. Steinberg, S. A. Rice, and S. Kjelleberg, "Biofilms: an emergent form of bacterial life," *Nat. Rev. Microbiol.*, vol. 14, no. 9, pp. 563–575, Sep. 2016, doi: 10.1038/nrmicro.2016.94.

[2] H.-C. Flemming and J. Wingender, "The biofilm matrix," *Nat. Rev. Microbiol.*, vol. 8, no. 9, pp. 623–633, Sep. 2010, doi: 10.1038/nrmicro2415.

[3] L. Hall-Stoodley, J. W. Costerton, and P. Stoodley, "Bacterial biofilms: from the Natural environment to infectious diseases," *Nat. Rev. Microbiol.*, vol. 2, no. 2, pp. 95–108, Feb. 2004, doi: 10.1038/nrmicro821.

[4] G. O'Toole, H. B. Kaplan, and R. Kolter, "Biofilm Formation as Microbial Development," *Annu. Rev. Microbiol.*, vol. 54, no. 1, pp. 49–79, Oct. 2000, doi: 10.1146/annurev.micro.54.1.49.

[5] M. F. Moradali and B. H. A. Rehm, "Bacterial biopolymers: from pathogenesis to advanced materials," *Nat. Rev. Microbiol.*, vol. 18, no. 4, pp. 195–210, Apr. 2020, doi: 10.1038/s41579-019-0313-3.

[6] D. O. Serra, A. M. Richter, and R. Hengge, "Cellulose as an Architectural Element in Spatially Structured Escherichia coli Biofilms," *J. Bacteriol.*, vol. 195, no. 24, pp. 5540–5554, Dec. 2013, doi: 10.1128/JB.00946-13.

[7] X. Zogaj, M. Nimtz, M. Rohde, W. Bokranz, and U. Römling, "The multicellular morphotypes of Salmonella typhimurium and Escherichia coli produce cellulose as the second component of the extracellular matrix," *Mol. Microbiol.*, vol. 39, no. 6, pp. 1452–1463, Mar. 2001, doi: 10.1046/j.1365-2958.2001.02337.x.

[8] U. Römling and M. Y. Galperin, "Bacterial cellulose biosynthesis: diversity of operons, subunits, products, and functions," *Trends Microbiol.*, vol. 23, no. 9, pp. 545–557, Sep. 2015, doi: 10.1016/j.tim.2015.05.005.

[9] P. V. Krasteva *et al.*, "Insights into the structure and assembly of a bacterial cellulose secretion system," *Nat. Commun.*, vol. 8, no. 1, Art. no. 1, Dec. 2017, doi: 10.1038/s41467-017-01523-2.

[10] J. L. W. Morgan, J. Strumillo, and J. Zimmer, "Crystallographic snapshot of cellulose synthesis and membrane translocation," *Nature*, vol. 493, no. 7431, pp. 181–186, Jan. 2013, doi: 10.1038/nature11744.

[11] J. L. W. Morgan, J. T. McNamara, and J. Zimmer, "Mechanism of activation of bacterial cellulose synthase by cyclic-di-GMP," *Nat. Struct. Mol. Biol.*, vol. 21, no. 5, pp. 489–496, May 2014, doi: 10.1038/nsmb.2803.

[12] J. L. W. Morgan *et al.*, “Observing cellulose biosynthesis and membrane translocation in crystallo,” *Nature*, vol. 531, no. 7594, pp. 329–334, Mar. 2016, doi: 10.1038/nature16966.

[13] J. T. McNamara, J. L. W. Morgan, and J. Zimmer, “A Molecular Description of Cellulose Biosynthesis,” *Annu. Rev. Biochem.*, vol. 84, pp. 895–921, 2015, doi: 10.1146/annurev-biochem-060614-033930.

[14] B. L. Quéré and J.-M. Ghigo, “BcsQ is an essential component of the Escherichia coli cellulose biosynthesis apparatus that localizes at the bacterial cell pole,” *Mol. Microbiol.*, vol. 72, no. 3, pp. 724–740, 2009, doi: 10.1111/j.1365-2958.2009.06678.x.

[15] X. Fang *et al.*, “GIL, a new c-di-GMP-binding protein domain involved in regulation of cellulose synthesis in enterobacteria,” *Mol. Microbiol.*, vol. 93, no. 3, pp. 439–452, 2014, doi: 10.1111/mmi.12672.

[16] O. E. Petrova and K. Sauer, “A Novel Signaling Network Essential for Regulating Pseudomonas aeruginosa Biofilm Development,” *PLOS Pathog.*, vol. 5, no. 11, p. e1000668, Nov. 2009, doi: 10.1371/journal.ppat.1000668.

[17] R. Hengge, “Principles of c-di-GMP signalling in bacteria,” *Nat. Rev. Microbiol.*, vol. 7, no. 4, pp. 263–273, Apr. 2009, doi: 10.1038/nrmicro2109.

[18] C. Chan *et al.*, “Structural basis of activity and allosteric control of diguanylate cyclase,” *Proc. Natl. Acad. Sci.*, vol. 101, no. 49, pp. 17084–17089, Dec. 2004, doi: 10.1073/pnas.0406134101.

[19] R. Paul *et al.*, “Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain,” *Genes Dev.*, vol. 18, no. 6, pp. 715–727, Mar. 2004, doi: 10.1101/gad.289504.

[20] A. J. Schmidt, D. A. Ryjenkov, and M. Gomelsky, “The Ubiquitous Protein Domain EAL Is a Cyclic Diguanylate-Specific Phosphodiesterase: Enzymatically Active and Inactive EAL Domains,” *J. Bacteriol.*, vol. 187, no. 14, pp. 4774–4781, Jul. 2005, doi: 10.1128/JB.187.14.4774-4781.2005.

[21] U. Römling, M. Y. Galperin, and M. Gomelsky, “Cyclic di-GMP: the first 25 years of a universal bacterial second messenger,” *Microbiol. Mol. Biol. Rev. MMBR*, vol. 77, no. 1, pp. 1–52, Mar. 2013, doi: 10.1128/MMBR.00043-12.

[22] P. V. Krasteva, K. M. Giglio, and H. Sondermann, “Sensing the messenger: The diverse ways that bacteria signal through c-di-GMP,” *Protein Sci.*, vol. 21, no. 7, pp. 929–948, 2012, doi: 10.1002/pro.2093.

[23] P. V. Krasteva and H. Sondermann, “Versatile modes of cellular regulation via cyclic dinucleotides,” *Nat. Chem. Biol.*, vol. 13, no. 4, pp. 350–359, Mar. 2017, doi: 10.1038/nchembio.2337.

[24] K. Sauer, A. K. Camper, G. D. Ehrlich, J. W. Costerton, and D. G. Davies, “Pseudomonas aeruginosa displays multiple phenotypes during development as a biofilm,” *J. Bacteriol.*, vol. 184, no. 4, pp. 1140–1154, Feb. 2002, doi: 10.1128/jb.184.4.1140-1154.2002.

[25] K. P. Rumbaugh and K. Sauer, “Biofilm dispersion,” *Nat. Rev. Microbiol.*, vol. 18, no. 10, pp. 571–586, Oct. 2020, doi: 10.1038/s41579-020-0385-0.

[26] A. Little *et al.*, “Revised Phylogeny of the Cellulose Synthase Gene Superfamily: Insights into Cell Wall Evolution,” *Plant Physiol.*, vol. 177, no. 3, pp. 1124–1141, Jul. 2018, doi: 10.1104/pp.17.01718.

[27] D. R. Nobles, D. K. Romanovicz, and R. M. Brown, “Cellulose in Cyanobacteria. Origin of Vascular Plant Cellulose Synthase,” *Plant Physiol.*, vol. 127, no. 2, pp. 529–542, Oct. 2001, doi: 10.1104/pp.010557.

[28] Y. Nishiyama, P. Langan, and H. Chanzy, “Crystal Structure and Hydrogen-Bonding System in Cellulose Iβ from Synchrotron X-ray and Neutron Fiber Diffraction,” *J. Am. Chem. Soc.*, vol. 124, no. 31, pp. 9074–9082, Aug. 2002, doi: 10.1021/ja0257319.

[29] K. H. Gardner and J. Blackwell, “The structure of native cellulose,” *Biopolymers*, vol. 13, no. 10, pp. 1975–2001, 1974, doi: 10.1002/bip.1974.360131005.

[30] W. Thongsomboon, D. O. Serra, A. Possling, C. Hadjineophytou, R. Hengge, and L. Cegelski, “Phosphoethanolamine cellulose: A naturally produced chemically modified cellulose,” *Science*, vol. 359, no. 6373, pp. 334–338, Jan. 2018, doi: 10.1126/science.aao4096.

[31] “4565.full.pdf.” Accessed: Aug. 20, 2021. [Online]. Available: https://www.pnas.org/content/pnas/73/12/4565.full.pdf

[32] C. Haigler, R. Brown, and M. Benziman, “Calcofluor white ST Alters the in vivo assembly of cellulose microfibrils,” *Science*, vol. 210, no. 4472, pp. 903–906, Nov. 1980, doi: 10.1126/science.7434003.

[33] M. Benziman, C. H. Haigler, R. M. Brown, A. R. White, and K. M. Cooper, “Cellulose biogenesis: Polymerization and crystallization are coupled processes in Acetobacter xylinum,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 77, no. 11, pp. 6678–6682, Nov. 1980.

[34] J. F. Acheson, Z. Derewenda, and J. Zimmer, “Architecture of the Cellulose Synthase Outer Membrane Channel and its association with the Periplasmic TPR Domain,” *Struct.*

Lond. Engl. 1993, vol. 27, no. 12, pp. 1855-1861.e3, Dec. 2019, doi: 10.1016/j.str.2019.09.008.

[35] N. Sunagawa *et al.*, "Cellulose complementing factor (Ccp) is a new member of the cellulose synthase complex (terminal complex) in Acetobacter xylinum," *J. Biosci. Bioeng.*, vol. 115, no. 6, pp. 607–612, Jun. 2013, doi: 10.1016/j.jbiosc.2012.12.021.

[36] S.-Q. Hu *et al.*, "Structure of bacterial cellulose synthase subunit D octamer with four inner passageways," *Proc. Natl. Acad. Sci.*, vol. 107, no. 42, pp. 17957–17961, Oct. 2010, doi: 10.1073/pnas.1000601107.

[37] W. J. Nicolas, D. Ghosal, E. I. Tocheva, E. M. Meyerowitz, and G. J. Jensen, "Structure of the Bacterial Cellulose Ribbon and Its Assembly-Guiding Cytoskeleton by Electron Cryotomography," *J. Bacteriol.*, vol. 203, no. 3, Jan. 2021, doi: 10.1128/JB.00371-20.

[38] A. J. Brown, "XLIII.—On an acetic ferment which forms cellulose," *J Chem Soc Trans*, vol. 49, no. 0, pp. 432–439, 1886, doi: 10.1039/CT8864900432.

[39] F. C. Lin, R. M. Brown, R. R. Drake, and B. E. Haley, "Identification of the uridine 5'-diphosphoglucose (UDP-Glc) binding subunit of cellulose synthase in Acetobacter xylinum using the photoaffinity probe 5-azido-UDP-Glc.," *J. Biol. Chem.*, vol. 265, no. 9, pp. 4782–4784, Mar. 1990, doi: 10.1016/S0021-9258(19)34039-6.

[40] W. Abidi, S. Zouhir, M. Caleechurn, S. Roche, and P. V. Krasteva, "Architecture and regulation of an enterobacterial cellulose secretion system," *Sci. Adv.*, vol. 7, no. 5, p. eabd8049, Jan. 2021, doi: 10.1126/sciadv.abd8049.

[41] J. B. McManus, Y. Deng, N. Nagachar, T. Kao, and M. Tien, "AcsA–AcsB: The core of the cellulose synthase complex from Gluconacetobacter hansenii ATCC23769," *Enzyme Microb. Technol.*, vol. 82, pp. 58–65, Jan. 2016, doi: 10.1016/j.enzmictec.2015.08.015.

[42] I. M. Saxena, K. Kudlicka, K. Okuda, and R. M. Brown, "Characterization of genes in the cellulose-synthesizing operon (acs operon) of Acetobacter xylinum: implications for cellulose crystallization.," *J. Bacteriol.*, vol. 176, no. 18, pp. 5735–5752, Sep. 1994.

[43] K. Maeda, J. Tamura, Y. Okuda, R. Narikawa, T. Midorikawa, and M. Ikeuchi, "Genetic identification of factors for extracellular cellulose accumulation in the thermophilic cyanobacterium Thermosynechococcus vulcanus: proposal of a novel tripartite secretion system," *Mol. Microbiol.*, vol. 109, no. 1, pp. 121–134, 2018, doi: 10.1111/mmi.13977.

[44] H. Xu, K. F. Chater, Z. Deng, and M. Tao, "A Cellulose Synthase-Like Protein Involved in Hyphal Tip Growth and Morphological Differentiation in Streptomyces," *J. Bacteriol.*, vol. 190, no. 14, pp. 4971–4978, Jul. 2008, doi: 10.1128/JB.01849-07.

[45] R. Liman, P. D. Facey, G. van Keulen, P. J. Dyson, and R. Del Sol, "A laterally acquired galactose oxidase-like gene is required for aerial development during osmotic stress in Streptomyces coelicolor," *PloS One*, vol. 8, no. 1, p. e54112, 2013, doi: 10.1371/journal.pone.0054112.

[46] T. J. Silhavy, D. Kahne, and S. Walker, "The Bacterial Cell Envelope," *Cold Spring Harb. Perspect. Biol.*, vol. 2, no. 5, p. a000414, May 2010, doi: 10.1101/cshperspect.a000414.

[47] W. Abidi, L. Torres-Sánchez, A. Siroy, and P. V. Krasteva, "Weaving of bacterial cellulose by the Bcs secretion systems," *FEMS Microbiol. Rev.*, no. fuab051, Oct. 2021, doi: 10.1093/femsre/fuab051.

[48] O. Omadjela *et al.*, "BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis," *Proc. Natl. Acad. Sci.*, vol. 110, no. 44, pp. 17856–17861, Oct. 2013, doi: 10.1073/pnas.1314063110.

[49] U. Römling, "Molecular biology of cellulose production in bacteria," *Res. Microbiol.*, vol. 153, no. 4, pp. 205–212, May 2002, doi: 10.1016/S0923-2508(02)01316-5.

[50] S. Valla *et al.*, "Cloning of a gene involved in cellulose biosynthesis in Acetobacter xylinum: complementation of cellulose-negative mutants by the UDPG pyrophosphorylase structural gene," *Mol. Gen. Genet. MGG*, vol. 217, no. 1, pp. 26–30, May 1989, doi: 10.1007/BF00330938.

[51] L. L. Lairson, B. Henrissat, G. J. Davies, and S. G. Withers, "Glycosyltransferases: Structures, Functions, and Mechanisms," *Annu. Rev. Biochem.*, vol. 77, no. 1, pp. 521–555, 2008, doi: 10.1146/annurev.biochem.76.061005.092322.

[52] C. Brown, F. Leijon, and V. Bulone, "Radiometric and spectrophotometric in vitro assays of glycosyltransferases involved in plant cell wall carbohydrate biosynthesis," *Nat. Protoc.*, vol. 7, no. 9, pp. 1634–1650, Sep. 2012, doi: 10.1038/nprot.2012.089.

[53] B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat, "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics," *Nucleic Acids Res.*, vol. 37, no. Database issue, pp. D233–D238, Jan. 2009, doi: 10.1093/nar/gkn663.

[54] K. E. Low and P. L. Howell, "Gram-negative synthase-dependent exopolysaccharide biosynthetic machines," *Curr. Opin. Struct. Biol.*, vol. 53, pp. 32–44, Dec. 2018, doi: 10.1016/j.sbi.2018.05.001.

[55] S. Nojima *et al.*, “Crystal structure of the flexible tandem repeat domain of bacterial cellulose synthesis subunit C,” *Sci. Rep.*, vol. 7, no. 1, p. 13018, Dec. 2017, doi: 10.1038/s41598-017-12530-0.

[56] M. Y. Galperin and E. V. Koonin, “Divergence and Convergence in Enzyme Evolution,” *J. Biol. Chem.*, vol. 287, no. 1, pp. 21–28, Jan. 2012, doi: 10.1074/jbc.R111.241976.

[57] E. C. Hollenbeck, A. Antonoplis, C. Chai, W. Thongsomboon, G. G. Fuller, and L. Cegelski, “Phosphoethanolamine cellulose enhances curli-mediated adhesion of uropathogenic *Escherichia coli* to bladder epithelial cells,” *Proc. Natl. Acad. Sci.*, vol. 115, no. 40, pp. 10106–10111, Oct. 2018, doi: 10.1073/pnas.1801564115.

[58] L. Sun *et al.*, “Structural and Functional Characterization of the BcsG Subunit of the Cellulose Synthase in Salmonella typhimurium,” *J. Mol. Biol.*, vol. 430, no. 18, pp. 3170–3189, Sep. 2018, doi: 10.1016/j.jmb.2018.07.008.

[59] A. C. Anderson, A. J. N. Burnett, L. Hiscock, K. E. Maly, and J. T. Weadge, “The Escherichia coli cellulose synthase subunit G (BcsG) is a Zn2+-dependent phosphoethanolamine transferase,” *J. Biol. Chem.*, vol. 295, no. 18, pp. 6225–6235, May 2020, doi: 10.1074/jbc.RA119.011668.

[60] F. Nicol, I. His, A. Jauneau, S. Vernhettes, H. Canut, and H. Höfte, “A plasma membrane-bound putative endo-1,4-beta-D-glucanase is required for normal wall assembly and cell elongation in Arabidopsis.,” *EMBO J.*, vol. 17, no. 19, pp. 5563–5576, Oct. 1998, doi: 10.1093/emboj/17.19.5563.

[61] I. Ahmad *et al.*, “BcsZ inhibits biofilm phenotypes and promotes virulence by blocking cellulose production in Salmonella enterica serovar Typhimurium,” *Microb. Cell Factories*, vol. 15, no. 1, p. 177, Dec. 2016, doi: 10.1186/s12934-016-0576-6.

[62] Baker, P., Hill, P. J., Snarr, B. D., Alnabelseya, N., Pestrak, M. J., Lee, M. J., ... & Howell, P. L. (2016). Exopolysaccharide biosynthetic glycoside hydrolases can be utilized to disrupt and prevent Pseudomonas aeruginosa biofilms. Science advances, 2(5), e1501632.

[63] T. Nakai *et al.*, “Formation of Highly Twisted Ribbons in a Carboxymethylcellulase Gene-Disrupted Strain of a Cellulose-Producing Bacterium,” *J. Bacteriol.*, vol. 195, no. 5, pp. 958–964, Mar. 2013, doi: 10.1128/JB.01473-12.

[64] Y. Yasutake *et al.*, “Structural characterization of the Acetobacter xylinum endo-β-1,4-glucanase CMCax required for cellulose biosynthesis,” *Proteins Struct. Funct. Bioinforma.*, vol. 64, no. 4, pp. 1069–1077, 2006, doi: 10.1002/prot.21052.

[65] O. Mazur and J. Zimmer, “Apo- and Cellopentaose-bound Structures of the Bacterial Cellulose Synthase Subunit BcsZ,” *J. Biol. Chem.*, vol. 286, no. 20, pp. 17601–17606, May 2011, doi: 10.1074/jbc.M111.227660.

[66] S. Kawano, K. Tajima, H. Kono, T. Erata, M. Munekata, and M. Takai, “Effects of endogenous endo-β-1,4-glucanase on cellulose biosynthesis in Acetobacter xylinum ATCC23769,” *J. Biosci. Bioeng.*, vol. 94, no. 3, pp. 275–281, Sep. 2002, doi: 10.1016/S1389-1723(02)80162-1.

[67] W. Scott, B. Lowrance, A. C. Anderson, and J. T. Weadge, “Identification of the Clostridial cellulose synthase and characterization of the cognate glycosyl hydrolase, CcsZ,” *PLoS ONE*, vol. 15, no. 12, p. e0242686, Dec. 2020, doi: 10.1371/journal.pone.0242686.

[68] H. M. Koo, S. H. Song, Y. R. Pyun, and Y. S. Kim, “Evidence that a beta-1,4-endoglucanase secreted by Acetobacter xylinum plays an essential role for the formation of cellulose fiber,” *Biosci. Biotechnol. Biochem.*, vol. 62, no. 11, pp. 2257–2259, Nov. 1998, doi: 10.1271/bbb.62.2257.

[69] S. Shan, “ATPase and GTPase Tangos Drive Intracellular Protein Transport,” *Trends Biochem. Sci.*, vol. 41, no. 12, pp. 1050–1060, Dec. 2016, doi: 10.1016/j.tibs.2016.08.012.

[70] G. Bange and I. Sinning, “SIMIBI twins in protein targeting and localization,” *Nat. Struct. Mol. Biol.*, vol. 20, no. 7, pp. 776–780, Jul. 2013, doi: 10.1038/nsmb.2605.

[71] A. J. Spiers, J. Bohannon, S. M. Gehrig, and P. B. Rainey, “Biofilm formation at the air–liquid interface by the Pseudomonas fluorescens SBW25 wrinkly spreader requires an acetylated form of cellulose,” *Mol. Microbiol.*, vol. 50, no. 1, pp. 15–27, 2003, doi: 10.1046/j.1365-2958.2003.03670.x.

[72] P. Ross *et al.*, “Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid,” *Nature*, vol. 325, no. 6101, pp. 279–281, Jan. 1987, doi: 10.1038/325279a0.

[73] D. Amikam and M. Y. Galperin, “PilZ domain is part of the bacterial c-di-GMP binding protein,” *Bioinforma. Oxf. Engl.*, vol. 22, no. 1, pp. 3–6, Jan. 2006, doi: 10.1093/bioinformatics/bti739.

[74] Z. Chen and P. Schaap, “The prokaryote messenger c-di-GMP triggers stalk cell differentiation in Dictyostelium,” *Nature*, vol. 488, no. 7413, pp. 680–683, Aug. 2012, doi: 10.1038/nature11313.

[75] Y. Aloni, D. P. Delmer, and M. BENZIMANt, “Achievement of high rates of in vitro synthesis of 1,4-13-D-glucan: Activation by cooperative interaction of the Acetobacter

xylinum enzyme system with GTP, polyethylene glycol, and a protein factor," *Proc Natl Acad Sci USA*, p. 5, 1982.

[76] Y. Aloni, R. Cohen, M. Benziman, and D. Delmer, "Solubilization of the UDP-glucose:1,4-beta-D-glucan 4-beta-D-glucosyltransferase (cellulose synthase) from Acetobacter xylinum. A comparison of regulatory properties with those of the membrane-bound form of the enzyme.," *J. Biol. Chem.*, vol. 258, no. 7, pp. 4419–4423, Apr. 1983, doi: 10.1016/S0021-9258(18)32639-5.

[77] U. Jenal, A. Reinders, and C. Lori, "Cyclic di-GMP: second messenger extraordinaire," *Nat. Rev. Microbiol.*, vol. 15, no. 5, pp. 271–284, May 2017, doi: 10.1038/nrmicro.2016.190.

[78] R. Hengge *et al.*, "Systematic Nomenclature for GGDEF and EAL Domain-Containing Cyclic Di-GMP Turnover Proteins of Escherichia coli," *J. Bacteriol.*, vol. 198, no. 1, pp. 7–11, doi: 10.1128/JB.00424-15.

[79] A. M. Richter *et al.*, "Local c-di-GMP Signaling in the Control of Synthesis of the E. coli Biofilm Exopolysaccharide pEtN-Cellulose," *J. Mol. Biol.*, vol. 432, no. 16, pp. 4576–4595, Jul. 2020, doi: 10.1016/j.jmb.2020.06.006.

[80] S. Da Re and J.-M. Ghigo, "A CsgD-Independent Pathway for Cellulose Production and Biofilm Formation in Escherichia coli," *J. Bacteriol.*, vol. 188, no. 8, pp. 3073–3087, Apr. 2006, doi: 10.1128/JB.188.8.3073-3087.2006.

[81] W. Abidi, S. Zouhir, M. Caleechurn, S. Roche, and P. V. Krasteva, "Architecture and regulation of an enterobacterial cellulose secretion system," *Sci. Adv.*, vol. 7, no. 5, p. eabd8049, Jan. 2021, doi: 10.1126/sciadv.abd8049.

[82] O. Sarenko *et al.*, "More than Enzymes That Make or Break Cyclic Di-GMP—Local Signaling in the Interactome of GGDEF/EAL Domain Proteins of Escherichia coli," *mBio*, vol. 8, no. 5, pp. e01639-17, doi: 10.1128/mBio.01639-17.

[83] S. Lindenberg, G. Klauck, C. Pesavento, E. Klauck, and R. Hengge, "The EAL domain protein YciR acts as a trigger enzyme in a c-di-GMP signalling cascade in E. coli biofilm control," *EMBO J.*, vol. 32, no. 14, pp. 2001–2014, Jul. 2013, doi: 10.1038/emboj.2013.120.

[84] V. Pfiffer, O. Sarenko, A. Possling, and R. Hengge, "Genetic dissection of Escherichia coli's master diguanylate cyclase DgcE: Role of the N-terminal MASE1 domain and direct signal input from a GTPase partner system," *PLOS Genet.*, vol. 15, no. 4, p. e1008059, avr 2019, doi: 10.1371/journal.pgen.1008059.

[85] U. Römling, W. D. Sierralta, K. Eriksson, and S. Normark, “Multicellular and aggregative behaviour of Salmonella typhimurium strains is controlled by mutations in the agfD promoter,” *Mol. Microbiol.*, vol. 28, no. 2, pp. 249–264, 1998, doi: 10.1046/j.1365-2958.1998.00791.x.

[86] U. Römling, M. Rohde, A. Olsén, S. Normark, and J. Reinköster, “AgfD, the checkpoint of multicellular and aggregative behaviour in Salmonella typhimurium regulates at least two independent pathways,” *Mol. Microbiol.*, vol. 36, no. 1, pp. 10–23, 2000, doi: 10.1046/j.1365-2958.2000.01822.x.

[87] “Structure and Multitasking of the c-di-GMP-Sensing Cellulose Secretion Regulator BcsE.” https://journals.asm.org/doi/epub/10.1128/mBio.01303-20 (accessed Aug. 26, 2021).

[88] H. Weinhouse *et al.*, “c-di-GMP-binding protein, a new factor regulating cellulose synthesis in Acetobacter xylinum,” *FEBS Lett.*, vol. 416, no. 2, pp. 207–211, Oct. 1997, doi: 10.1016/s0014-5793(97)01202-7.

[89] D. A. Ryjenkov, R. Simm, U. Römling, and M. Gomelsky, “The PilZ Domain Is a Receptor for the Second Messenger c-di-GMP: THE PilZ DOMAIN PROTEIN YcgR CONTROLS MOTILITY IN ENTEROBACTERIA*,” *J. Biol. Chem.*, vol. 281, no. 41, pp. 30310–30314, Oct. 2006, doi: 10.1074/jbc.C600179200.

[90] B. T. Nixon *et al.*, “Comparative Structural and Computational Analysis Supports Eighteen Cellulose Synthases in the Plant Cellulose Synthesis Complex,” *Sci. Rep.*, vol. 6, no. 1, p. 28696, Jun. 2016, doi: 10.1038/srep28696.

[91] P. Paximada, A. A. Koutinas, E. Scholten, and I. G. Mandala, “Effect of bacterial cellulose addition on physical properties of WPI emulsions. Comparison with common thickeners,” *Food Hydrocoll.*, vol. 54, pp. 245–254, Mar. 2016, doi: 10.1016/j.foodhyd.2015.10.014.

[92] Z. Shi, Y. Zhang, G. O. Phillips, and G. Yang, “Utilization of bacterial cellulose in food,” *Food Hydrocoll.*, vol. 35, pp. 539–545, Mar. 2014, doi: 10.1016/j.foodhyd.2013.07.012.

[93] K. Fijałkowski, D. Peitler, R. Rakoczy, and A. Żywicka, “Survival of probiotic lactic acid bacteria immobilized in different forms of bacterial cellulose in simulated gastric juices and bile salt solution,” *LWT - Food Sci. Technol.*, vol. 68, pp. 322–328, May 2016, doi: 10.1016/j.lwt.2015.12.038.

[94] K. Ludwicka, M. Kaczmarek, and A. Białkowska, “Bacterial Nanocellulose—A Biobased Polymer for Active and Intelligent Food Packaging Applications: Recent

Advances and Developments," *Polymers*, vol. 12, no. 10, Art. no. 10, Oct. 2020, doi: 10.3390/polym12102209.

[95] A. Fillat *et al.*, "Bacterial cellulose for increasing barrier properties of paper products," *Cellulose*, vol. 25, no. 10, pp. 6093–6105, Oct. 2018, doi: 10.1007/s10570-018-1967-0.

[96] A. H. Basta and H. El-Saied, "Performance of improved bacterial cellulose application in the production of functional paper," *J. Appl. Microbiol.*, vol. 107, no. 6, pp. 2098–2107, Dec. 2009, doi: 10.1111/j.1365-2672.2009.04467.x.

[97] H. S. Barud *et al.*, "Optically transparent membrane based on bacterial cellulose/polycaprolactone," *Polímeros*, vol. 23, pp. 135–142, 2013, doi: 10.1590/S0104-14282013005000018.

[98] J. Shah and R. Malcolm Brown, "Towards electronic paper displays made from microbial cellulose," *Appl. Microbiol. Biotechnol.*, vol. 66, no. 4, pp. 352–355, Jan. 2005, doi: 10.1007/s00253-004-1756-6.

[99] E. R. P. Pinto *et al.*, "Transparent composites prepared from bacterial cellulose and castor oil based polyurethane as substrates for flexible OLEDs," *J. Mater. Chem. C*, vol. 3, no. 44, pp. 11581–11588, Nov. 2015, doi: 10.1039/C5TC02359A.

[100] R. Augustine, R. Rajendran, U. Cvelbar, M. Mozetič, and A. George, "Biopolymers for Health, Food, and Cosmetic Applications," in *Handbook of Biopolymer-Based Materials*, John Wiley & Sons, Ltd, pp. 801–849. doi: 10.1002/9783527652457.ch27.

[101] T. Almeida, A. J. D. Silvestre, C. Vilela, and C. S. R. Freire, "Bacterial Nanocellulose toward Green Cosmetics: Recent Progresses and Challenges," *Int. J. Mol. Sci.*, vol. 22, no. 6, Art. no. 6, Jan. 2021, doi: 10.3390/ijms22062836.

[102] H. Wu *et al.*, "Regenerated chitin fibers reinforced with bacterial cellulose nanocrystals as suture biomaterials," *Carbohydr. Polym.*, vol. 180, pp. 304–313, Jan. 2018, doi: 10.1016/j.carbpol.2017.10.022.

[103] L. Zheng, S. Li, J. Luo, and X. Wang, "Latest Advances on Bacterial Cellulose-Based Antibacterial Materials as Wound Dressings," *Front. Bioeng. Biotechnol.*, vol. 8, p. 593768, Nov. 2020, doi: 10.3389/fbioe.2020.593768.

[104] W. Czaja, A. Krystynowicz, S. Bielecki, and R. M. Brown, "Microbial cellulose--the natural power to heal wounds," *Biomaterials*, vol. 27, no. 2, pp. 145–151, Jan. 2006, doi: 10.1016/j.biomaterials.2005.07.035.

[105] M. de Lima Fontes *et al.*, "Effect of in situ modification of bacterial cellulose with carboxymethylcellulose on its nano/microstructure and methotrexate release properties," *Carbohydr. Polym.*, vol. 179, pp. 126–134, Jan. 2018, doi: 10.1016/j.carbpol.2017.09.061.

[106] X. Lv *et al.*, “Bacterial Cellulose-Based Biomimetic Nanofibrous Scaffold with Muscle Cells for Hollow Organ Tissue Engineering,” *ACS Biomater. Sci. Eng.*, vol. 2, no. 1, pp. 19–29, Jan. 2016, doi: 10.1021/acsbiomaterials.5b00259.

[107] J. Wang, C. Gao, Y. Zhang, and Y. Wan, “Preparation and in vitro characterization of BC/PVA hydrogel composite for its potential use as artificial cornea biomaterial,” *Mater. Sci. Eng. C*, vol. 30, no. 1, pp. 214–218, Jan. 2010, doi: 10.1016/j.msec.2009.10.006.

[108] S. Gonçalves *et al.*, “Acetylated bacterial cellulose coated with urinary bladder matrix as a substrate for retinal pigment epithelium,” *Colloids Surf. B Biointerfaces*, vol. 139, pp. 1–9, Mar. 2016, doi: 10.1016/j.colsurfb.2015.11.051.

[109] M. Scherner *et al.*, “In vivo application of tissue-engineered blood vessels of bacterial cellulose as small arterial substitutes: proof of concept?,” *J. Surg. Res.*, vol. 189, no. 2, pp. 340–347, Jun. 2014, doi: 10.1016/j.jss.2014.02.011.

[110] G. Baniulyte *et al.*, “Identification of regulatory targets for the bacterial Nus factor complex,” *Nat. Commun.*, vol. 8, no. 1, p. 2027, Dec. 2017, doi: 10.1038/s41467-017-02124-9.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Innovation thérapeutique
du fondamental à l'appliqué
(ITFA)

Titre : Études structurales et fonctionnelles de sécrétion de cellulose bactérienne

Mots clés : Biofilms bactériens, cellulose, microbiologie, biologie structurale

Résumé : La cellulose, le composé le plus abondant sur la planète Terre, est un composant important des parois cellulaires des plantes. La biogenèse de la cellulose, cependant, n'est pas limitée aux règnes supérieurs, car les bactéries sécrètent également de la cellulose en tant que composant extracellulaire de leur matrice de biofilm. Alors que la découverte de la cellulose bactérienne remonte au 19ème siècle, ce n'est qu'un siècle plus tard qu'une succession de travaux structuraux a fourni des détails sur la glucosyltransférase BcsA (Bacterial cellulose synthesis A) largement conservée et activée par le second messager intracellulaire c-di-GMP (di-guanosine monophosphate cyclique). La sous-unité catalytique BcsA avec son partenaire BcsB, permet la polymérisation et le passage du polysaccharide à travers la membrane interne des bactéries. Néanmoins, d'autres sous-unités accessoires contribuent ou sont essentielles à l'assemblage et à la stabilité du complexe, ainsi qu'aux modifications post-synthétiques et à la sécrétion de la cellulose naissante à travers l'enveloppe cellulaire bactérienne complexe. La machinerie de sécrétion de cellulose d'*E. coli* (*Escherichia coli*) par exemple, est caractérisée par neuf sous-unités (BcsRQABZCEFG) qui s'étendent du cytosol à l'espace extracellulaire. Une étude récente a révélé que la plupart des sous-unités Bcs chez *E. coli* interagissent pour former une nanomachine sécrétoire. Cependant, sa reconstruction initiale à basse résolution n'a pas permis de distinguer les sous-unités régulatrices ni leurs mécanismes fonctionnels exacts. Dans ce contexte, j'ai étudié le système Bcs de type *E. coli*, structurellement et biochimiquement. J'ai montré que les sous-unités cytoplasmiques BcsR, BcsQ et BcsE interagissent pour former des sous-complexes régulateurs. J'ai également fourni plusieurs structures cristallographiques de ces sous-complexes régulateurs, qui ont révélé une stœchiométrie inattendue de ces sous unités, une interaction à différents motifs du c-di-GMP et une régulation dépendante de l'ATP (adénosine triphosphate). Pour obtenir plus de détails mécanistiques du macrocomplexe dans son intégralité, j'ai eu recours à la cryo-EM (cryo-microscopie électronique) par la technique dite de « particules isolées ». L'étude structurale du macrocomplexe a révélé un macrocomplexe asymétrique d'environ 1 MDa (Megadalton). Je montre en outre que l'asymétrie du macrocomplexe Bcs est due à l'auto-oligomérisation de BcsB par complémentation des feuillets ß dans le périplasme. Ces découvertes permettent de placer les pièces clés de l'assemblage et du fonctionnement du système de sécrétion de cellulose dans un plus large contexte de signalisation bactérienne et de formation de biofilm.

Maison du doctorat de l'Université Paris-Saclay
2ème étage aile ouest, Ecole normale supérieure Paris-Saclay
4 avenue des Sciences,
91190 Gif sur Yvette, France

université PARIS-SACLAY

ÉCOLE DOCTORALE
Innovation thérapeutique
du fondamental à l'appliqué
(ITFA)

Title: Structural and functional studies of bacterial cellulose secretion

Keywords: Bacterial biofilms, cellulose, microbiology, structural biology

Abstract: Cellulose, the most abundant compound on the planet Earth, is an important component of plants' cell walls. The cellulose biosynthesis, however, is not limited to higher kingdoms, as bacteria also secrete cellulose as an extracellular component of their biofilm matrix. While the discovery of bacterial cellulose goes back to the 19th century, it is only a century later that a succession of structural works provided details of the widely conserved glucosyltransferase BcsA (Bacterial cellulose synthesis A), which is allosterically activated by the intracellular second messenger c-di-GMP (Bis-(3′-5′)-cyclic dimeric guanosine monophosphate). The catalytic subunit BcsA together with its co-synthetic partner BcsB, permits the polymerization and extrusion of the polysaccharide through the inner bacterial membrane. Nevertheless, various accessory subunits contribute to or are essential for synthase complex assembly and stability, as well as post-synthetic modifications and secretion of the nascent cellulose through the complex bacterial cell envelope. The *E. coli* (*Escherichia coli*) cellulose secretion machinery for instance, is characterized by nine subunits (BcsRQABZCEFG) that span from the cytosol to the extracellular space. A recent study provided mechanistic understanding of the cellulose secretion system assembly in *E. coli* and revealed that most of the Bcs subunits interact to form a secretory nanomachine. However, its initial low-resolution reconstruction didn't allow to distinguish the regulatory subunits or their exact functional mechanisms, even though the interest in bacterial cellulose in the biotechnological and other fields continues to spark. To address these inadequacies, I studied the conserved *E. coli*-like Bcs system, structurally and biochemically. I showed that the cytosolic subunits BcsR, BcsQ, and BcsE interact to form regulatory subcomplexes. I also provided multiple crystallographic snapshots of these accessory subcomplexes, which revealed unexpected subunit stoichiometry, multisite c-di-GMP recognition, and an unexpected ATP (adenosine triphosphate)-dependent regulation. To gain further mechanistic and structural details on the assembled macrocomplex, I resorted to single particle cryo-EM (cryo-Electron Microscopy). The structural study of the macrocomplex revealed a roughly 1 MDa (Megadalton)-sized asymmetric secretory assembly. I further show that the asymmetry of the Bcs macrocomplex is due to self-driven and membrane-curvature-dependent BcsB homo-oligomerization through ß-sheet complementation in the periplasm. These findings allow to place key pieces of the puzzle of cellulose secretion system assembly and function into the bigger picture of bacterial signaling and biofilm formation.

Maison du doctorat de l'Université Paris-Saclay
2ème étage aile ouest, Ecole normale supérieure Paris-Saclay
4 avenue des Sciences,
91190 Gif sur Yvette, France