2019

Characterizing the cellulose-modifying enzyme BcsG from Escherichia coli

Alexander Anderson
ande5290@mylaurier.ca

Joel T. Weadge
Wilfrid Laurier University

Recommended Citation

Anderson, Alexander and Weadge, Joel T., "Characterizing the cellulose-modifying enzyme BcsG from Escherichia coli" (2019). *Theses and Dissertations (Comprehensive)*. 2193.
https://scholars.wlu.ca/etd/2193

Characterizing the cellulose-modifying enzyme BcsG from *Escherichia coli*

by

Alexander C. Anderson

Bachelor of Science, Wilfrid Laurier University

THESIS

Submitted to the Department of Biology

Faculty of Science

in partial fulfilment of the requirements for the

Master of Science in Integrative Biology

Wilfrid Laurier University

2019

Abstract

Microbial biofilms are communities of microorganisms that exhibit co-operative behaviour, producing a matrix of exopolysaccharide that enmeshes the community. The well-studied human pathogens *Escherichia coli* and *Salmonella enterica* produce a biofilm matrix comprised chiefly of the biopolymer cellulose, along with amyloid protein fibers termed curli. This biofilm matrix confers surface adherence and acts as a protective barrier to disinfectants, antimicrobials, environmental stressors, and host immune responses. Pertaining to this research, the *bcsEFG* operon, conserved in the *Enterobacteriaceae,* encodes an inner membrane-spanning complex responsible for the addition of a phosphoethanolamine (pEtN) modification to microbial cellulose, essential for extracellular matrix assembly and biofilm architecture. Furthermore, *E. coli* deficient in *bcsG* produce a biofilm matrix lacking structural integrity and the self-assembling architecture observed in wild-type *E. coli*. The presence of a pEtN-substituted cellulose matrix was shown to be important in bladder epithelial colonization by uropathogenic *E. coli*, further suggesting its role as a virulence factor in etiological agents of urinary tract infections observed in the clinic. The purpose of this research was to characterize the *Ec*BcsG enzyme, a putative pEtN transferase, by resolving its structure, understanding its role in pEtN substitution of the cellulose matrix in *E. coli*, and by elucidating its catalytic mechanism to enable future efforts in drug discovery. All these research objectives were achieved, shedding light on the biochemical basis of pEtN cellulose in *E. coli*. The *de novo* X-ray crystal structure of the C-terminal catalytic domain of *Ec*BcsG was solved using the single-wavelength anomalous diffraction (SAD) technique phased on L-selenomethionine substituted *Ec*BcsG crystals and revealed *Ec*BcsG folds as a zinc-dependent phosphotransferase belonging to the pEtN transferase family. The *Ec*BcsG active site was mapped using functional complementation, revealing *Ec*BcsG shares a partially conserved active site with other known pEtN transferase family members, including enzymes responsible for resistance to cationic antimicrobial peptides (CAMPs). Using mock cosubstrates *para*-nitrophenyl phosphoethanolamine (*p*-NPPE) and cellooligosaccharides, the specific activity of *Ec*BcsG was measured to be 11.45 +/- 0.51 nmol min^{-1} mg^{-1} in the presence of the cellulo-pentasaccharide. The kinetic parameters were measured as $K_\text{m} = 1.673 \pm 0.382$ mM, $k_\text{cat} = 6.876 \times 10^{-3} \pm 4.14 \times 10^{-4}\ \text{s}^{-1}$, and $k_\text{cat}/K_\text{m} = 4.110 \pm 0.970$ M s^{-1}. The enzymatic product of this reaction was identified and confirmed *in vitro*. Finally, the covalent phospho-enzyme intermediate was isolated, providing evidence for an *Ec*BcsG catalytic mechanism. The structural and functional model of the cellulose modifying complex, resulting from this work, provides new insight into the biochemical basis for the biofilm matrix of *E. coli* and other *Enterobacteriaceae.* The research presented here offers opportunities in structure-based drug discovery and other efforts in inhibiting microbial biofilms and including the possibility of limiting the resilience of biofilm-forming pathogens. Additionally, further understanding of *bcsEFG*-directed phosphoethanolamine cellulose production may enable biosynthetic engineering of new cellulosic materials or confer the advantages of phosphoethanolamine cellulose in new organisms.

Acknowledgements

Above all else, I want to express my most sincere gratitude to my supervisor, Dr. Weadge. You are a natural leader, a supportive mentor, and an incredible source of positive encouragement. I cannot count how many times you have talked me down from worry or doubt and kept me going when I thought I was out of steam. Your unconditional support and guidance have taken me from an untrained volunteer to a confident, skilled scientist. Thank you.

I would also like to thank Dr. Lynette Celeski and Wiriya Thongsomboon at Stanford University for gifting us the AR3110 cell lines that were pivotal in the complementation studies, Denis Spasyuk and Michel Fodje of the CMCF for their expertise and advice in collecting and processing XRD data, Dr. Dyanne Brewer and Dr. Armen Charchoglyan at the University of Guelph AAC for their assistance with mass spectrometry experiments, and Gena Braun at Wilfrid Laurier for her help with the ICP-AES experiments. A very special thank you to Lana Hiscock for her expertise and hard work in the synthesis of *p*-NPPE.

I would also like to express my thanks to Erum Razvi, Dr. Michael Suits, Dr. Geoff Horsman, and Jonah Nechacov for helpful advice and conversations that contributed knowledge to this research. I'd like to thank Thomas Brenner, whose prior work contributed significantly to solving the structure of BcsG. Finally, I would like to thank my labmates and colleagues on the fourth floor of Science Research, far too many to name individually, for their comradery, friendship and assistance. You have all enriched my experience and I wish you all the best in your future endeavours.

Table of Contents

List of Figures and Illustrations

List of Tables

List of Abbreviations

ABC: ATP-binding casette
Bi-bi: Bisubstrate-biproduct
c-di-GMP: bis(3′,5′)-cyclic dimeric guanosine monophosphate
TPR: tetratricopeptide repeat
CAMP: cationic antimicrobial peptide
EDTA: ethylenediaminetetraacetic acid
EPS: extracellular polymeric substances
IEC: ion exchange chromatography
IMAC : immobilized metal ion affinity chromatography
IPTG: Isopropyl β-D-1-thiogalactopyranoside
LB: lysogeny broth
LC-MS: liquid chromatography-mass spectrometry
LC-MS/MS liquid chromatography-tandem mass spectrometry
MBOAT: membrane-bound *O*-acetyltransferase
MCR: mobile colistin resistance (factor)
MIC: minimum inhibitory concentration
N-Boc: *N*-tert-Butyloxycarbonyl
NCBI: National Center for Biotechnology Information
Ni-NTA: nickel-nitrilotriacetic acid
NMR: nuclear magnetic resonance
OD_{600}: optical density, 600nm wavelength
OPG: osmoregulated periplasmic glucan
p-NPPE: *para*-nitrophenyl phosphoethanolamine
PCR: polymerase chain reaction
pEtN: phosphoethanolamine
PNAG: poly-β-(1,6)-*N*-acetyl-D-glucosamine
Q-ToF: quadrupole time-of-flight
RPM: rotations per minute
SAD: single-wavelength anomalous diffraction
SB: super broth
SEC: size exclusion chromatography
THF: tetrahydrofuran
TLC: thin-layer chromatography
TPR: tetratricopeptide repeat (motif)
UDP: uridine diphosphate
UndPP: undecaprenyl pyrophosphate
XRD: X-ray diffraction

1. Introduction

1.1 Microbial biofilms

Microbial biofilms are defined as communities of microorganisms that reside in a matrix of extracellular polymeric substances (EPS), produced by the communities themselves, that adhere to other cells and/or a substratum[1]. Fleming and colleagues have described biofilms as an emergent form of bacterial life that differ substantially from free-living cells[1]. The properties of biofilms cannot be predicted by the study of free-living bacteria alone, and are among the most successful and ubiquitous life strategies on Earth[1]. All higher organisms, including humans, are colonized by at least one organism observed to form biofilms[1]. These communities are often high-cell density, containing over 10^8 cells per gram of wet weight[1]. The composition of these biofilms can be highly heterogeneous, both in the species composition and the gene or protein expression of the constituent cells[1,2]. This heterogeneity is driven by the interplay between communication, co-operation, competition, and the occurrence of ecological microgradients within the biofilm[1–3]. As a result, the biofilm matrix serves as an interface between its constituents and their outside environment, serving many roles to the benefit of the microorganisms that may justify the large resource cost of producing copious quantities of extracellular matrix material[1]. For example, biofilms have been observed to serve in resource capture[1,4–6], accelerated cell growth[7], and the tolerance of stressors or disinfectants, such as shearing forces[1,3,7], dessication[1,8], antimicrobial compounds[1,3,9,10], extreme temperature[1,3,11], or sanitizing agents[1,3,4,9]. Thus, biofilms not only confer emergent properties that have enhanced the success of microorganisms on Earth, they provide constituent cells resilience, persistence, and diversity[1] that has made them a subject of research interest. For example, the formation of biofilms by pathogens pose significant problems to human health, as biofilm formation can occur within the host, on medical devices or implants, or on surfaces of healthcare facilities and equipment[1,3]. In these environments, biofilms would confer enhanced resistance of these pathogens to antimicrobial drugs and disinfectants commonly used to treat infection and contamination, enhancing the survival of biofilms as compared to free-living cells[1,3,12]. Fux *et al.* estimate that biofilms may be implicated in greater than 60% of resistant infections observed by physicians in the clinic[3], making them significant in the study of infectious disease. The formation of biofilms has also been observed in many foodborne microbial pathogens, enhancing their ability to persist through food preservation, packaging, and preparation or cooking[1]. Thus, the persistence of foodborne pathogens or spoilage organisms may also result in devastating economic losses through product recalls and workforce illnesses. The formation of biofilms in water treatment facilities, water distribution pipes and on the surface of underwater infrastructure also pose problems to human health and economics[1,3]. The study of biofilms is not, however, limited to pathogens. Several biofilm-forming species have been applied to the study of new materials[13], wound dressings, polymer science, or water treatment and biocontrol strategies[1,13].

Microbial biofilms are considered a distinct form of life from motile and planktonic cells, and the mechanisms that govern their formation have been reported extensively. Biofilms are formed by planktonic cells through a reversible, then irreversible attachment to a substratum and/or other cells via type IV pili, fimbriae or other cell-surface mediated adhesion mechanisms[11,14,15]. Various cellular signalling events further drive cell division and differentiation that results in the disassembly and downregulation of motility-associated factors, including flagella, and culminate in the secretion of exopolysaccharides and proteinaceous factors that begin to enmesh the growing community[16,17]. Many independent reports conclude that heterogeneity is largely responsible for the complex architecture observed in mature biofilms[16–18], forming highly ordered and differentiated three-dimensional structures that include channels for the exchange of nutrients, solvent, gases, metabolic wastes, and the flow of cells into or out of the biofilm. Metabolic differentiation also confers important emergent properties to biofilms, including altered susceptibility to antimicrobial drugs, the cooperative exchange of resources, and the accumulation of extracellular enzymes that aid in defence, cooperativity and nutrient exchange[1]. Transitioning from a mature biofilm to a planktonic cell or cells is also possible, and participation in a biofilm is a reversible process as entry and exit to the biofilm phenotype is a dynamic and regulated process[1].

Beyond the heterogeneity and the overwhelming diversity observed in biofilms, the single unifying and defining characteristic of biofilms are the extracellular matrices that encase the communities within them. Biofilm matrices themselves are complex, self-assembling structures composed primarily of structurally diverse carbohydrates and proteins, but may also include secreted extracellular DNA, lipids, and other non-structural molecules[1]. In addition to these complex carbohydrate polymers, the existence of large, high molecular-weight amyloid protein fibers has been described in large quantities[1,5,10,11,19]. The identity of the exopolysaccharides that compose an individual biofilm is highly variable, with different organisms producing different exopolysaccharides in their biofilms, both between and within species[1,20]. Isolates of a single organism may produce different exopolysaccharides with vastly different compositions under different conditions, or may produce no EPS at all under others[1,14,20,21], and the presence of multiple genetic elements linked to the production of a variety of EPS components within a single organism is also not uncommon[1,14,22]. For example, the bacterium *Pseudomonas aeruginosa* is capable of producing the exopolysaccharides alginate, Pel, and Psl, although not concomitantly[23]. Despite their apparent importance and near-ubiquity, still little is known about the biochemical basis of EPS biosynthesis, export, and self-assembly into the complex biofilm matrix. However, it is generally understood that EPS biosynthesis is carried out by pathways described as “synthase-dependent”[24]. These pathways are differentiated from others that function in the biosynthesis and export mechanisms of other cell-surface carbohydrates in microorganisms, such as the cell wall, lipopolysaccharide, or capsular polysaccharide[24]. These structural differences are attributed to the nature of the polymer, where capsular or lipopolysaccharides are assembled on a lipid acceptor and are shuttled to the cell surface where they remain attached, whereas, exopolysaccharides are freely secreted at the cell surface and, thus, unique mechanisms for their biosynthesis are necessitated[23,24].

1.2 Synthase-dependent pathways

As described above, the composition of the biofilm matrix is variable between microbial species, and in some cases, may also vary within a single species, as some organisms may possess the means to produce multiple biofilm components under different conditions[7]. In contrast, one unifying feature of these biofilms is that all matrix compositions described at present contain some kind of carbohydrate component[24]. Many of these carbohydrate components can be isolated as high molecular weight, unbranched polysaccharide chains[24]. The biosynthesis of these EPS carbohydrate polymers do not share any physiological mechanisms with other well-studied bacterial metabolic pathways of other cell-surface carbohydrates, such as lipopolysaccharide or capsular polysaccharide. For example, the Wzx-Wzy dependent-pathway is used to assemble the cell-surface carbohydrate O antigen, found as the outer layer of lipopolysaccharide in Gram negative bacteria (Fig. 1)[25]. In this system, glycosyltransferases polymerize the base repeating saccharide chains of O antigen on a lipid acceptor, undecaprenyl pyrophosphate (UndPP) on the cytoplasmic face of the inner membrane[25]. The Wzx flippase then flips the UndPP lipid to the periplasmic face of the membrane via cation-driven antiport[25]. The lipid-linked saccharide unit is transferred to the Wzy polymerase, where it adds a repeating saccharide unit to the reducing end of the growing polymer, one unit at a time[25]. The length of the polymer is then regulated by the Wzz protein, followed by ligation to the lipid A core of lipopolysaccharide[25]. Alternatively, biosynthesis of other cell-surface polysaccharides, such as

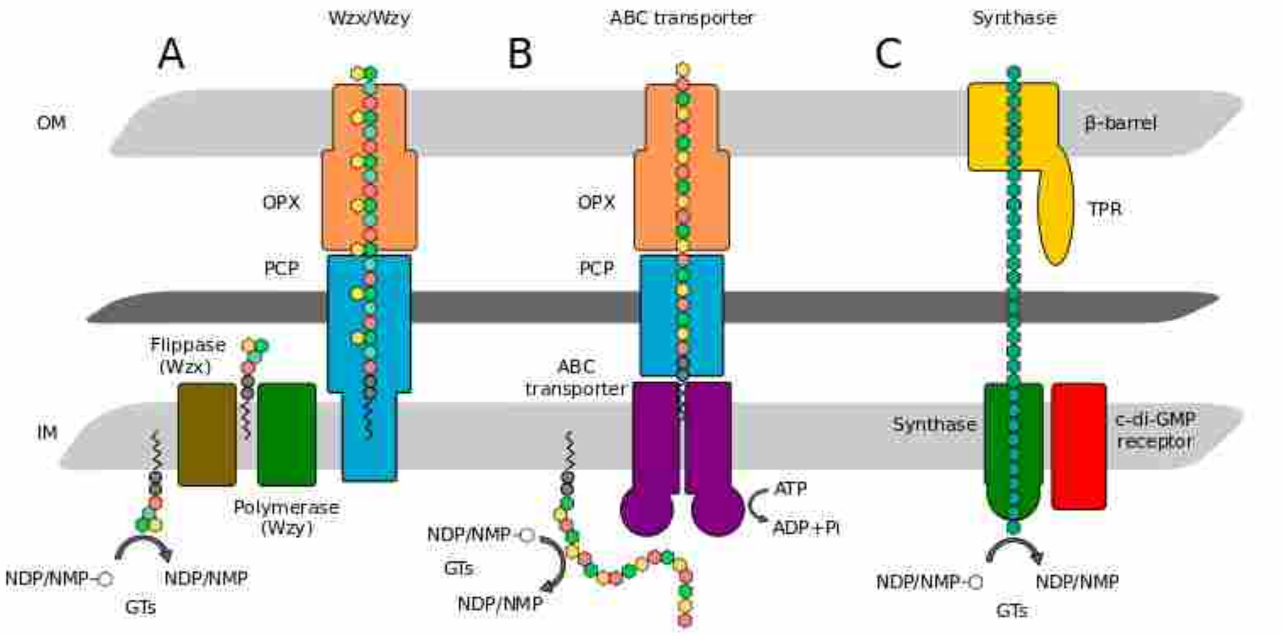


Figure 1: Pathways for the biosynthesis & assembly of cell-surface glycans in Gram-negative bacteria. Lipid-linked glycans, such as lipopolysaccharide or capsular polysaccharide, are assembled on lipid anchors and shuttled to the outer membrane via the Wzx/Wzy dependent pathway (A) or by ATP-binding cassette (ABC) transport (B). Exopolysaccharides biosynthesis and export is instead carried out by synthase complexes (C). PCP: polysaccharide co-polymerase, OPX: outer membrane polysaccharide export, NDP/NMP: nucleotide diphosphate/monophosphate, OM: outer membrane, IM: inner

spore coat or capsular polysaccharide, is carried out in the cytoplasm by cytosolic glycosyltransferases on lipid cores[24], and translocation to the periplasm is carried out by ATP-binding cassette (ABC) transporters (Fig. 1)[24,25]. Export of these lipid-linked polysaccharides is then carried out by the polysaccharide co-polymerase and outer membrane export systems, a series of homologous systems in Gram-negative bacteria that transport a diverse set of lipid-linked glycans to the outer membrane of the cell[26,27].

The free secretion of the EPS components at the cell surface differentiates them from the above mentioned glycans and, thus, necessitates their unique biosynthetic pathways to be independent of lipid acceptors. The biosynthesis of EPS components is instead thought to be carried out by a series of pathways termed “synthase-dependent”, where the EPS is polymerized at the inner membrane and transiently anchored to a synthase complex, allowing for free secretion at the cell surface following polymerization and export[24]. Similar to the other described cell-surface glycan biosynthesis pathways, these synthase-dependent pathways all share common structural characteristics[24,28], although the individual proteins responsible vary based on the EPS they serve to polymerize and export. In Gram-negative bacteria, the focus of this research, these synthase-dependent pathways form a *trans*-envelope complex[24] spanning the inner and outer membranes and traversing the periplasmic space (Fig. 1). In Gram-positive organisms, these synthase complexes reside in the cytoplasmic membrane and facilitate the direct export of the nascent polymer to the extracellular space.

At present, our understanding of the major components of the Gram-negative synthase-dependent pathway include the involvement of a synthase complex[24,29], an outer membrane secretin complex[24,28], and various periplasmic components found to be important for processing of the EPS polymer[5,20,30,31]. The polymerization of the EPS polymer is carried out by the synthase complex, canonically containing a glycosyltransferase domain, responsible for polymerizing the EPS from its metabolic precursor[29], and a regulatory PilZ domain responsible for binding cyclic dimeric guanosine monophosphate (c-di-GMP)[9], an important second messenger molecule in bacteria with a demonstrated physiological role in transcription of community-associated genes and enzymatic activation of their gene products[12,13].

Polymerization of the EPS glycan is linked to translocation through the synthase complex and into the periplasm, where often the nascent polymer may be enzymatically modified by transferases[32], esterases[33–35], epimerases[36], hydrolases[24,30,37], or a combination of these[20]. Finally, export is carried out through an outer-membrane secretin[28], composed of a β-barrel porin[24] in the outer membrane and a tandem-tetratricopeptide repeat (TPR), either as an accessory domain to the porin, or as a separate protein anchored to the outer membrane[28,38]. Together, these biosynthetic genes are operon-encoded, and are upregulated as part of the motile-to-sessile physiological transition in response to cellular signalling events, typically centered around c-di-GMP metabolism[22].

1.3 Microbial cellulose

Among the diverse host of biofilm EPS components, microbial cellulose is one EPS matrix produced by biofilm-forming bacteria of clinical and economic importance[14]. The cellulose biofilm was first described in *Gluconacetobacter xylinus* (formerly *Acetobacter xylinus*), forming a large pellicle composed chiefly of cellulose at the air-liquid interface[39], and the later discovery of cellulose in the well-studied human pathogens *E. coli* and *Salmonella enterica*[14,19] highlight its importance. To date, the production of a biofilm composed of the exopolysaccharide cellulose has shown importance to *Acetobacter*, *Sarcina*, *Rhizobium, Agrobacterium, Cronobacter* and *Pseudomonas* species, among others, to form symbiotic and pathogenic interactions[5,40,41]. Although they share the same name, bacterial cellulose is structurally distinct from plant cellulose, as the former demonstrates malleability and high water-retention as compared to the latter. These properties make bacterial cellulose highly useful for medical applications and wound dressings, as a low cost natural product that can accelerate healing, decrease pain, and greatly lowers the probability of infection for severe burns[42]. Because of its inert and non-immunogenic nature, biocompatibility, and relative resistance to degradation, bacterial cellulose has quickly become a major research topic in the modern biomedical industry[13].

Microbial cellulose is produced as a linear polymer of β-(1,4) linked glucose, although it has been shown to be chemically modified in some cases[43]. In the case of unmodified microbial cellulose, hydrogen bonding occurs between hydroxyl groups, facilitating the formation of dense, fibrous and microcrystalline ultrastructure of the cellulose chains[43]. For this reason, microbial cellulose has high tensile strength and low water retention, making it the subject of study for materials science where it has seen application in wound dressings and stitches[39,43,44]. Moreover, heterogeneity in the structure and assembly of cellulose biofilms have garnered research interest into what shapes these discrete extracellular matrices and the specific properties they confer to constituent bacteria. To date, at least three distinct subgroups of microbial cellulose have been described[5,32,45], which are discussed in further detail below, but all require the essential synthase-dependent components that assemble and export the cellulose exopolysaccharide.

1.4 The bacterial cellulose synthase-dependent exopolysaccharide biosynthetic machine

All cellulose-producing bacteria possess the core *bcsABCZ* operon, or homologues thereof, containing the essential biosynthetic machinery for the production and export of cellulose chains, with the notable absence of outer-membrane localized components in Gram-positive organisms[43]. The BcsA enzyme and its homologues have been demonstrated to be the synthase component in the bacterial inner membrane, possessing processive glycosyltransferase activity[29]. The identification of BcsA/BcsB as the catalytically active glycosyltransferase and the resolution of the bacterial cellulose synthase by X-ray crystallography represented significant milestones in the contemporary understanding of bacterial cellulose biosynthesis[29,46]. The BcsA glycosyltransferase domain catalyzes the addition of a single glucose residue at a time to the growing chain from a uridine 3’-diphosphate (UDP) linked donor molecule, originating from the bacterial cytoplasm[29]. BcsA also possesses a PilZ c-di-GMP binding domain[29], found in other c-di-GMP-responsive proteins[6,22], responsible for control of its glycosyltransferase activity. BcsB forms a heterodimer with BcsA and has been shown to be necessary for BcsA activity and cellulose production both *in vivo*[47] and *in vitro*[29]. Together, the BcsA/B complex also facilitates the translocation of nascent cellulose chains to the periplasmic space via a hydrophilic channel in BcsA and carbohydrate-binding domain in BcsB[29,48].

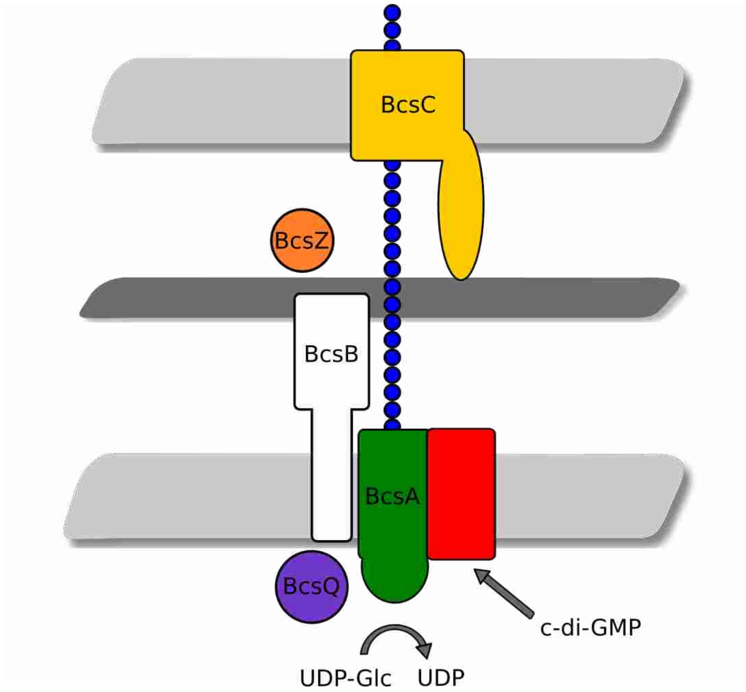


Figure 2: The bacterial cellulose synthase biosynthetic machine. Polymerization of cellulose chains from cytoplasmic UDP-glucose is carried out by the BcsA glycosyltransferase under positive regulation by the second messenger molecule cyclic dimeric guanosine monophosphate (c-di-GMP). Periplasmic translocation is coupled with synthesis and accomplished by the BcsA/BcsB heterodimer. Finally, export is achieved through the BcsC outer membrane secretin. The periplasmic endoglucanase BcsZ is thought to be a negative regulator of the biofilm phenotype and inversely expressed with the synthase components. Model based on Whitney and Howell (2013; Ref. 24).

Secretion of the cellulose polymer across the outer membrane is necessary for Gram-negative organisms, and is thought to be carried out by a class of homologous secretin complexes encoded on synthase-dependent operons[24,38]. The BcsC protein, and its synthase-dependent homologues belong to this novel class of exopolysaccharide secretins[28,38]. A large β-barrel domain in the outer membrane forms a pore for export that matches the chemical nature of the EPS polymer[24]. For example, BcsC is thought to form an uncharged hydrophilic channel of 8Å diameter, suitable for cellulose secretion[38], while AlgE of the alginate EPS system possesses a series of positively-charged periplasmic loops to facilitate the movement of the anionic alginate polymer[49,50]. A TPR region is also present in BcsC and other EPS synthase homologues, and is thought to serve as a scaffold, both for assembling the synthase complex through protein-protein interaction as observed in other TPR-containing proteins, but also by guiding the polymer through the periplasmic space and interacting with the peptidoglycan cell wall in some way[28,38]. Finally, the BcsZ enzyme possesses endo-β(1-4)-glucanase activity, colloquially referred to as a cellulase enzyme, and is responsible for degradation of cellulose polymers from within the chain[30,48]. The activity of BcsZ may be responsible for suppression of the biofilm phenotype and serve to initiate acute infection and virulence[30], enabling rapid transition to motility and host pathogenicity in some organisms. Also identified is the BcsQ protein, thought to be a MinD homologue responsible for the polar localization of cellulose synthase complexes at terminal cell poles and regulating the end-to-end cellular architecture observed in biofilm phenotypes[43].

An understanding of the regulation and assembly of the synthase complex remains elusive, largely due to the complex and transient protein-protein interactions that shape it. At present, it is understood that the assembly of synthase-dependent complexes occurs through at least two distinct mechanisms: first, a molecular checkpoint ensures correct protein folding, membrane insertion and protein-protein interaction of the synthase complex[51], and subsequently, c-di-GMP binding serves to both enzymatically activate the complex and promote its proteolytic stability[52,53]. Furthermore, it appears that loss of c-di-GMP binding or complex ultrastructure is correlated to downregulation of the synthase components and upregulation of the BcsZ glycosyl hydrolase expression[51], although its specific role, perhaps in periplasmic or extracellular clearance of the cellulose exopolysaccharide remains to be demonstrated.

1.5 Microbial crystalline cellulose

The initial discovery of microbial cellulose and the *bcsABCZ* operon in *Acetobacter xylinus* enabled the identification of many more operons in a diverse subset of bacteria that produce cellulose as part of their extracellular matrix[40,54]. Only following the description of modified cellulose exopolysaccharides, and their biofilm phenotypes, did it become clear that unmodified cellulose exopolysaccharides were responsible for a discrete biofilm morphotype of their own. The identification of *acsD* in the proteobacterium *Acetobacter hansenii* and *xcsB* in the cyanobacterium *Thermosynechococcus vulcanus* suggested that a periplasmic adaptor protein may serve as a funnel for the coupling of individual cellulose fibrils into nanocrystalline fibers, shaping the biofilm architecture of the unmodified cellulose exopolysaccharides[45,55]. The apparent structural similarity of the XcsB periplasmic component of the cyanobacterial cellulose

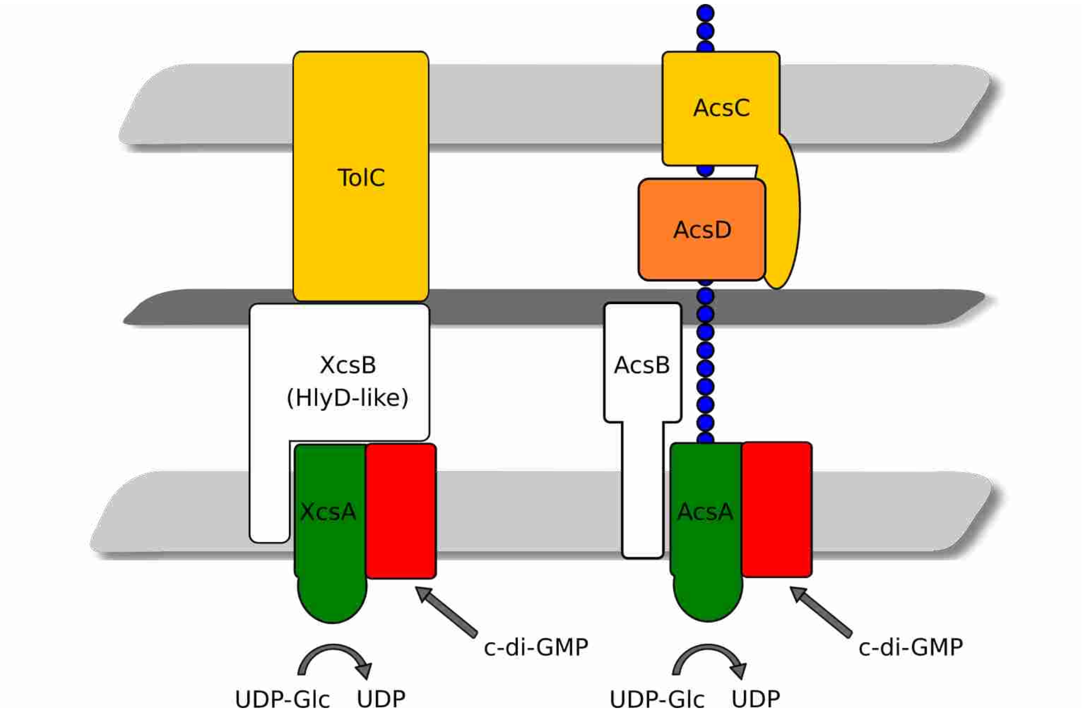


Figure 3: Crystalline cellulose biosynthesis. The proposed bacterial cellulose synthase pathways in the cyanobacterium *Thermosynechococcus vulcanus* (left) and the proteobacterium *Acetobacter hansenii* (left). The presence of a periplasmic adaptor protein in these species may be related to the bundling of cellulose into crystalline fibers in their biofilms. Adapted from Maeda *et al* (2018; Ref 45) and Hu *et al.* (2016; Ref 47).

synthase to the HlyD component of bacterial tripartite efflux pumps suggest that cellulose secretion in cyanobacteria may be mechanistically unique from the Proteobacteria[45,56]. Despite this heterogeneity, numerous reports in the literature demonstrate that the essential components for cellulose production include the synthase complex itself, and the outer membrane secretin[22,29,39,47], and that the lack of accessory periplasmic components is generally correlated with the production of unmodified, crystalline cellulose nanofibrils in the biofilm.

1.6 Microbial acetylcellulose

The first report of microbial acetylcellulose was reported by Spiers and colleagues[5] who demonstrated that an unusual, “wrinkly spreader” biofilm phenotype of *Pseudomonas fluorescens* SBW25 was driven by the production of a modified cellulose exopolysaccharide. In addition to the operon-encoded, core synthase-dependent biosynthetic machinery needed to produce cellulose, this organism also possessed additional genetic components whose deletion resulted in the secretion of an unmodified cellulose exopolysaccharide with different physical characteristics[5]. The acetylation of other carbohydrate polymers in bacteria has been described, including pathways for the O-acetylation of the essential cell wall polymer peptidoglycan[57,58] or the O-acetylation of the exopolysaccharide alginate, first described in *P. aeruginosa*[20,23]*.* The enzymes encoded on this acetylcellulose operon share sequence similarity with both the *bcsABCZ* synthase complex and the *algXFIJ* acetylation pathway that produces the exopolysaccharide alginate, a hemiacetylated polysaccharide consisting of (1-4)-linked β-D-mannuronic acid and α-L-guluronic acid also produced via a synthase-dependent system. Insight into these mechanistic pathways of polymer O-acetylation have been reported extensively in the literature and provide a structural and biochemical basis for acetylation of these polymers.

Generally, an unknown acetyl donor, likely acetyl coenzyme A, provides the acetyl group to a membrane-bound O-acetyltransferase (MBOAT) in the inner membrane[20,23,58]. This acetyl group is then subsequently transferred to periplasmic protein components that serve to covalently attach this acetyl group to the polymer in the periplasmic space[20,58,59].

The essential functions for production and secretion of these acetylcellulose polymers are carried out by the wrinkly spreader structural, or *wssBCDE* and *wssFGHI* operons, named so after the multicellular morphotype their expression confers[5]. The *wssBCDE* operon encodes predicted homologues of the *bcsABCZ* operon, including a synthase complex (*wssB/wssC*), an outer membrane secretin (*wssE*) and a glycosyl hydrolase (*wssD*)[5,20,23]. The role of the *wssFGHI* gene products in cellulose O-acetylation is not well understood at present, although their

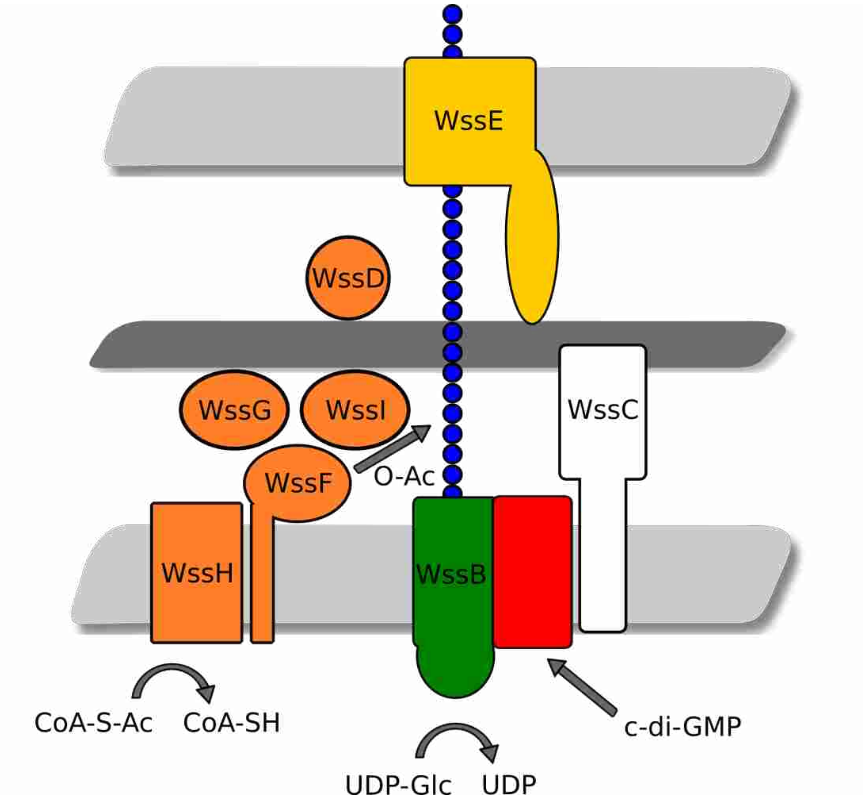


Figure 4: Acetylcellulose biosynthesis. While the molecular mechanisms that govern cellulose O-acetylation have yet to be elucidated, it is generally believed that the *wssFGHI* operon encodes a series of proteins responsible for shuttling an acetyl group from a cytoplasmic donor, likely acetyl coenzyme A, to the exopolysaccharide in the periplasmic space. Adapted from Whitney and Howell (2013; Ref 24).

sequence similarity suggests they play an analogous role to the alginate O-acetylation proteins *algXFIJ*, and to a lesser degree, to the Gram-negative peptidoglycan O-acetyltransferase proteins *patAB*[5,31]. Initial acceptance of an acetyl group from a cytoplasmic donor by WssH, which shares sequence similarities with the membrane-bound O-acetyltransferases PatA and AlgI, whose function are well-characterized[23,58]. The shuttle of these acetyl groups from the membrane-bound O-acetyltransferase to the nascent exopolysaccharide in the periplasmic space is then carried out by WssF, WssI, and possibly WssG, evidenced by the sequence similarity of these proteins to AlgX, AlgJ and AlgF, respectively[5]. These alginate O-acetyltransferase components have been shown to have separate and distinct roles in polymer acetylation, despite having similar sequences and functional roles[31].

1.7 Microbial phosphoethanolamine cellulose

The identification of phosphoethanolamine (pEtN) cellulose was reported by Thongsomboon *et al.*[32] based on discoveries by McCrate *et al.* that the extracellular matrix of *E. coli* UTI89 harbors a modified form of cellulose[10]. Solid-state nuclear magnetic resonance (NMR) experiments on the isolated extracellular matrix of *E. coli* demonstrated that its biofilm was composed of amyloid curli fibers and pEtN cellulose[32]. Despite the acknowledgement of cellulose synthase operons in the *Enterobactericeae*[11,22,39,60], it was long thought that these organisms produced unmodified cellulose strands as the pEtN evaded detection by traditional methods, often lost by treatments with acid or base that were commonplace in the isolation of the matrix material from the biofilm[32]. Instead, the mature polymer isolated from several strains of pathogenic *E. coli* and *Salmonella enterica* is a cellulose chain of which every-other glucose unit is substituted at the C6 hydoxyl with an O-phosphorylethanolamine group[32].

Similar to the phenomenon observed in acetylcellulose[5], this modification with phosphoethanolamine conferred unique physical properties to the biofilm that were lost with its absence, and these properties appear important in the colonization of hosts of pathogenic *Enterobactericeae*[32,61]

The biochemical basis for the phosphoethanolamine substitution of microbial cellulose has not yet been demonstrated. However, an additional *bcsEFG* operon has been reported adjacent to the *bcsABCZ* operon in many cellulose-producing members of the *Enterobacteriaceae,* termed Type II operons in the literature[22,60]. Until recently, the role of this secondary *bcsEFG* operon remained largely unknown. Romling *et al.* reported that deletion of the operon in *S. enterica* demonstrated reduced, but detectable cellulose production[39]. Fang *et al.*[62] reported that the BcsE protein represented a new class of c-di-GMP receptors[62]. Thongsomboon *et al.*[32] further demonstrated that deletion of *bcsG,* and to a lesser degree, *bcsE* and *bcsF* resulted in a reduction or loss of phsophoethanolamine substitution of cellulose in *E. coli* and *S. enterica*. The proposed model of *bcsEFG* involves assembly into an inner membrane-spanning complex responsible for phosphoethanolamine substitution of microbial cellulose. This complex is suggested to be regulated by c-di-GMP binding, independently of the BcsAB synthase complex[32]. Sequence analysis and mutagenesis suggest that *bcsG* encodes the key phosphoethanolamine transferase enzyme responsible for the covalent phosphoethanolamine modification of the cellulose exopolysaccharides in the *Enterobacteriaceae*[32].

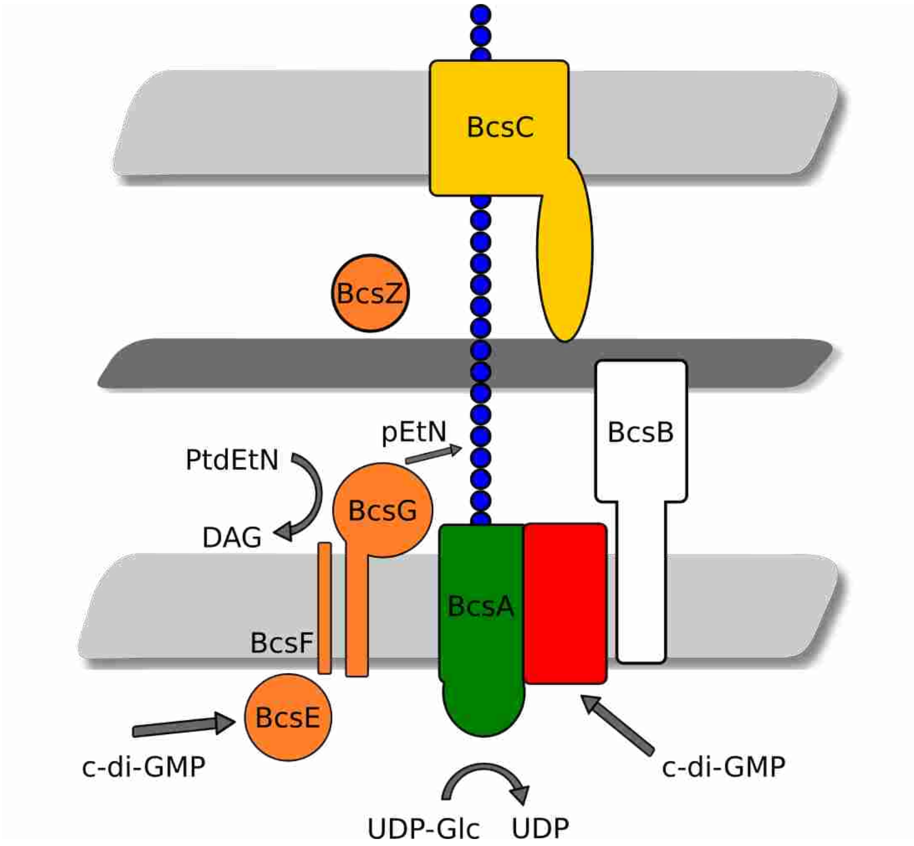


Figure 5: Phosphoethanolamine cellulose biosynthesis. The accessory *bcsEFG* operon in the *Enterobacteriaceae* is proposed to encode a novel c-di-GMP receptor and a phosphoethanolamine transferase responsible for modification of the exopolysaccharide. PtdEtN: phosphatidylethanolamine, DAG: diacyldlycerol, pEtN: phosphoethanolamine. Adapted from Thongsomboon *et al.* (2018; Ref 32).

1.8 Phosphoethanolamine transferases

The substitution of polysaccharides with pEtN outside the context of this research has been described in the literature. Several pEtN transferases that serve to substitute lipopolysaccharide with pEtN have been reported[63–66]. Lipopolysaccharides are high-molecular weight glycolipids composed of a lipid A glycophospholipid, an inner and outer core branched glycan, and an outer unbranched glycan repeating unit. Lipopolysaccharides are also antigenic determinants – the outer repeating glycan units of these molecules referred to as “O antigen”, as the lipopolysaccharide layer of unique bacterial isolates demonstrate significant structural variability both between species or strains and across isolates of the same species or strain[67].

The lipid A component is synthesized by a conserved pathway of nine sequential enzymes[67] that assemble the metabolic precursor, UDP-N-acetylglucosamine into lipid A-Kdo$_2$, which becomes the ultimate acceptor of core sugars[67,68]. Lipid A structure may vary widely both within and between species based on subtle structural changes in the lipid A biosynthetic enzymes and their regulation[67]. The surface O-antigen is assembled separately, and varies based on the activity of many diverse glycosyltransferases, and is often strain specific with over 160 described O-antigen variants found in *E. coli* alone[64,68]. Further variability of the mature LPS structure is attributable to the activity of aforementioned phosphoethanolamine transferases that further decorate these glycans, either at the lipid A glycophospholipid, or the core saccharides, and enhance their variability[64,68–71].

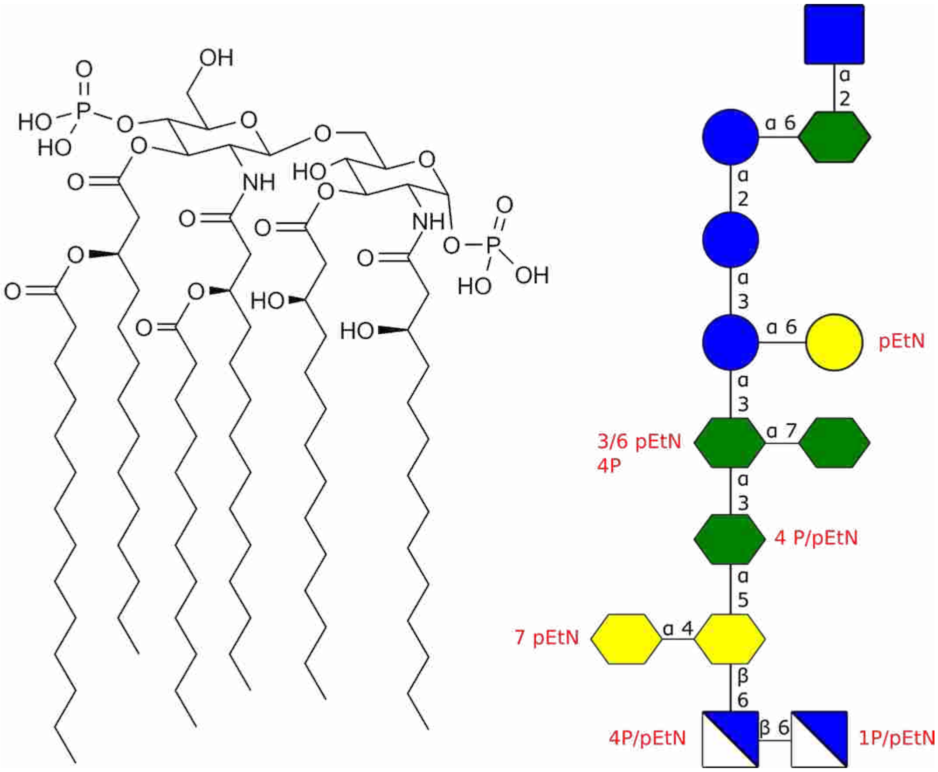


Figure 6: Lipid A and lipopolysaccharide structure. The addition of pEtN to lipid A (left), or to core oligosaccharides (right) of the mature lipopolysaccharide molecule confers resistance to polymyxins, cathelicidins, and alpha- or beta-defensins. A representative lipopolysaccharide structure from *E. coli* K12 is shown based on work by Washazaki *et al.* (2016; Ref. 71), with phosphorylation or pEtN modification sites annotated in red.

The strong negative charge of the lipid A glycophospholipid at the cell surface serves as the binding site for cathelicidins; antimicrobial peptides of the vertebrate innate immune system, defensins; antimicrobial peptides found in vertebrates, invertebrates and some plants, and polymyxins; a class of cyclic, cationic antimicrobial peptides (CAMPs) targeting Gram-negative bacteria[66,72]. Pertaining to this study, polymyxins are naturally produced by various members of the genus *Bacillus* subgenus *Paenibacillius* and have been used as clinical antimicrobial drugs due to their high potency, causing cell death in milligram per liter conentrations[66,73]. The binding of polymyxins to lipopolysaccharides induces destabilization of the cell membrane, leading to increased permeability, cell death, and lysis[74]. The efficacy of these polymyxins is dependent upon the electrostatic interaction between the positively charged polymyxin and the negative charge of the phosphate groups of lipid A on the cell surface[65]. Thus, in bacteria that share an environmental niche with *Paenibacillius,* the emergence of a mechanism to shield the strong negative charge of the lipid A phosphate groups confers resistance to these peptides, and subsequntly, considerable selective advantage in such an environment[75,76]. The action of characterized pEtN transferases on these lipid A serve such a role, where a covalent pEtN modification shields the negatively charged phosphate that serves as the peptide binding site, both sterically with the ethyl linkage, and electrostatically with a positively charged amine group[77].

Today, polymyxins have been repurposed extensively for clinical use as a natural product antimicrobial[73,75,76]. Despite being nephrotoxic, polymyxins are classified as antibiotics of critical importance by the World Health Organization due to their activity against the rising threat of multidrug resistant Gram negative bacteria[73]. However, many species are or have become intrinsically resistant to polymyxins, such as *Neisseria meningitidis, Campylobacter jejuni, Burkholderia cepacian,* and others[75,76]. The increased use of polymyxins in the clinic or for broader industrial purpose has been causally linked to the emergence of these lipid A pEtN transferases encoded on transferable multidrug resistance plasmids, disseminated through human pathogens, such as *E. coli, P. aeruginosa,* or *Klebsiella pneumoniae,* among others[76]. The flagship example, reported on widely in the literature, are the mobile colistin resistance (MCR) factors[72,76]. Based on sequence analysis, these phosphoethanolamine transferases consist of an N-terminal transmembrane domain that anchors them in the inner membrane of Gram-negative bacteria and contain a C-terminal domain belonging to the alkaline phosphatase superfamily, which is localized to the periplasmic face of the inner membrane. Several of these polymyxin resistance factors have been structurally resolved and biochemically characterized. These enzymes modify nascent lipopolysaccharide co-synthetically, using the membrane lipid phosphatidylethanolamine as a donor of the modification[65], first binding to the lipid via the N-terminal transmembrane domain and cleaving the pEtN group, liberating diacylglycerol[78]. These enzymes are then predicted to proceed through a ping-pong double-displacement mechanism, characterized by a covalent pEtN-enzyme intermediate thats transfers the pEtN to its ultimate lipid A acceptor[72], although the evidence for this mechanism remains limited. Furthermore, these enzymes possess a conserved binding site for one or more zinc cations in their active sites, canonical of related phosphotransferases[65]. This model for activity is presumed to be conserved throughout all known pEtN transferases, and by extension, serves as the proposed model for BcsG structure, activity, and mechanism.

2. Research Need and Hypotheses

Antimicrobial-resistant infections are predicted to be the leading cause of death worldwide by 2050, surpassing annual mortality by cancer and cardiovascular disease[79]. The emergence and spread of resistance to clinical antimicrobial drugs by numerous pathogens has become a widespread and worsening public health problem[80]. Such an emergence of common infections resistant to first- and second-line antibiotic treatment options necessitates the use of drugs by clinicians that may be more costly, more toxic to the patient and less efficacious for treating the infection[80]. Furthermore, infections are being presented in the clinic that are becoming increasingly multidrug resistant, further reducing the treatment options available, and cases have been reported of several pathogens that are completely recalcitrant to all known classes of antimicrobial drugs[80].

The discovery of new molecular targets for clinical therapeutics and the development of novel therapeutic strategies to treat multidrug resistant infections is imperative to combat the rising threat of antimicrobial resistance[81]. Biofilms serve as the first line of defence for their constituents, and biofilms have a demonstrated role in the colonization of the host in several pathogens of clinical significance – for example, Fux *et al*. have estimated biofilms to be implicated in greater than 60% of infections observed by physicians in the clinic[3]. Furthermore, the biofilm matrix itself offers both direct and indirect mechanisms to enhance the resistance of its constituents to clinical antimicrobial drugs[1,8,82,83]. These properties make biofilms an attractive potential target for novel therapeutics, where disruption of the biofilm may improve the efficacy of antimicrobial drugs against resistant pathogens or may serve as a novel treatment to attenuate recalcitrant infections. The development of biofilm dispersal strategies may serve to improve treatment outcomes for patients dependent solely upon contemporary treatments centered around bactericidal drugs that are rapidly losing efficacy. Strategies based upon disruption of bacterial biofilms using the glycoside hydrolases encoded by bacteria themselves have been demonstrated in the literature to be correlated with improved treatment outcomes in various infection models and proof-of-concept experiments[84–86].

Pertaining to this research, the discovery of *bcsG* as the essential genetic component for synthesis of pEtN cellulose and, thus, the architecture and assembly of the biofilm matrix makes BcsG an attractive target for such structure-based drug discovery. The role of pEtN cellulose in host colonization by the clinically important pathogens *E. coli* and *S. enterica* further demonstrate the research need for an understanding of the chemical biology of pEtN cellulose biofilms. Small molecule inhibitors of *bcsG*-directed pEtN cellulose biosynthesis may serve to attenuate the host colonization potential of multidrug resistant pathogens, serving important clinical roles in the treatment of antimicrobial-resistant infections that pose a rising global threat to public health.

Furthermore, an understanding of *bcsEFG*-directed pEtN cellulose production may enable biosynthetic engineering of new cellulosic materials or confer the advantages of pEtN cellulose in new organisms. pEtN cellulose evaded detection through decades of microbial cellulose research owing largely to the harsh acid hydrolysis procedures for isolating the matrix material[10]. Only recently have pEtN cellulose and the role of the *bcsEFG* operon been identified, and our understanding of this modification remains in its infancy. The biochemical basis of BcsG activity and the structure of the BcsEFG complex remain unexplored.

We hypothesize that BcsG is functionally homologous to the phosphatidylethanolamine lipooligosaccharide transferases, catalyzing the metal-dependent transfer of lipid head groups to a cellulose exopolysaccharide acceptor. We also hypothesize a structural and biochemical characterization of BcsG will lead to a greater understanding of the mechanism of cellulose pEtN modification and the underlying molecular mechanisms of bacterial cellulose biosynthesis.

3. Objectives

The objective of this thesis is to investigate the chemical biology of pEtN cellulose in bacteria. The research will focus on the gene product of *bcsG* from the model organism and clinically important human pathogen *E. coli.* In order to gain a meaningful understanding of this phenomenon, the thesis addresses three key objectives: a structural characterization of BcsG through crystallization, X-ray diffraction (XRD) analysis and structural modelling; a functional characterization of BcsG through *in vitro* activity assays; and an understanding of the catalytic mechanism of BcsG.

3.1 Structural characterization

At the onset of this thesis, no homologous structure of BcsG was available. During this work, the structure of BcsG from *S. typhimurium* (*St*BcsG*)* was solved independently. The work presented here expands on the knowledge of *St*BcsG in three distinct ways. First, a structural model of the *E. coli* phosphoethanolamine transferase would provide significant insight to the enzyme fold, the catalytic mechanism, and the architecture of the active site pocket in a second model organism of the *Enterobacteriaceae*. The resolution of the *Ec*BcsG catalytic domain would inform further studies aimed at elucidating its function, interpreting its relationship to other phosphotransferase enzymes, and designing small molecule inhibitors of BcsG activity. Because of the significant contributions to the current knowledge mentioned above that such a model would provide, a structural characterization of the BcsG catalytic domain from the model organism *E. coli* is a primary aim of this thesis.

3.2 Functional characterization of BcsG

The *bcsG* gene has been established as the essential genetic component required for phosphoethanolamine production in the extracellular matrix of *E. coli* and *S. enterica.* However, at the time of writing, the enzymatic activity of BcsG had never before been demonstrated. To test the hypothesis that BcsG is a phosphoethanolamine transferase, a key objective of this thesis was to establish that the C-terminal catalytic domain of BcsG is capable of this enzymatic activity.

3.2 Catalytic investigation

An understanding of the catalytic mechanism and the active site architecture is essential to unlocking the secrets of *bcsG*-directed phosphoethanolamine cellulose. At present, the active site and catalytic mechanism of other phosphoethanolamine transferases have been proposed and are supported by experimental evidence. Using this data as a model for the catalytic function and architecture of BcsG, this thesis will address the molecular mechanism of pEtN cellulose.

4. Experimental Methods

4.1 Recombinant protein production

The full-length open reading frame of *bcsG* from *E. coli* K12 (*Ec*BcsG) and a truncated *bcsG* comprised of the nucleotides encoding amino acids Ala164 to Gln559 (*Ec*BcsG$^{\Delta N)}$ were generated via subcloning into the pET28a overexpression vector according to the protocol by Cohen *et al.* (1973)[87] to generate pBcsG and pBcsG$^{\Delta N}$, respectively. Site-directed mutagenesis of these recombinant plasmids at predicted active-site residues were generated using the polymerase chain reaction (PCR) primer method described by Hutchison *et al.* (1978)[88] with the QuikChange II SDM kit supplied by Agilent Technologies and according to the manufacturer's protocol using synthetic oligonucleotide primers supplied by the University of Guelph Laboratory Services facility, Guelph, ON to produce the plasmids listed in Table 1 below. All plasmids were sequence verified at the Centre for Applied Genomics, University of Toronto, Toronto, ON. Summary information for the *Ec*BcsG and *Ec*BcsG$^{\Delta N}$ recombinant constructs are available in Table 2 below. A list of experimental and laboratory strains of *E. coli* used in this study are listed below in Table 3.

Figure 7: Predicted *Ec*BcsG topology, containing a putative N-terminal transmembrane domain and a C-terminal catalytic domain.

Table 1: Recombinant plasmids and oligonucleotide primers used in this study

Plasmid	Relevant description	Oligonucleotide primer sequence(s) (5’-3’)	Source
pBcsG	Recombinant enzyme	F:GGCGCATATGACTCAATTTACGCAAAATACC R:GCGCCTCGAGTTACTGCGGGTAAG	This study
pBcsG C243A	Metal-binding mutant	F:CTACTGGTGATTAACATCGCGTCGCTTTCCTGGTCGGATATAG R:CTATATCCGACCAGGAAAGCGACGCGATGTTAATCACCAGTAG	This study
pBcsG Y277A	Substrate-binding mutant	F:CCACCTCCGCGAGTGGCCCGGCGGCGATC R:GATCGCCGCCGGGCCACTCGCGGAGGTGG	This study
pBcsG Y277F	Substrate-binding mutant	F:GCCACCTCCTTCAGTGGCCCGGCGGCGATCC R:GGATCGCCGCCGGGCCACTGAAGGAGGTGGC	This study
pBcsG S278A	Catalytic mutant	F:CCACCTCCTACGCGGGCCCGGCGGC R:GCCGCCGGGCCCGCGTAGGAGGTGG	This study
pBcsG C290A C306A	Disulfide bond mutant		BioBasic
pBcsG F329A	Substrate-binding mutant	F:GATGGGACATAACGGCCAGGCGGGCGGTTTTTTGAAAG R:CTTTCAAAAAACCGCCCGCCTGGCCGTTATGTCCCATC	This study
pBcsG H396A	Substrate-binding mutant		BioBasic
pBcsG E442A	Metal-binding mutant	F:GGTCGTGGTGCCGGCACACGGCGGCGCG R:CGCGCCGCCGTGTGCCGGCACCACGACC	This study
pBcsG H443A	Metal-binding mutant	F:GTCGTGGTGCCGGAAGCGGGCGGCGCGC R:GCGCGCCGCCCGCTTCCGGCACCACGAC	This study
pBcsG$^{\Delta N}$	Recombinant catalytic domain	F:GGCGCATATGACTCAATTTACGCAAAATACC R:CTCCTTGCATATGGCGGGACAACC	This study
pBcsG$^{\Delta N}$ C243A	Metal-binding mutant	F:CTACTGGTGATTAACATCGCGTCGCTTTCCTGGTCGGATATAG R:CTATATCCGACCAGGAAAGCGACGCGATGTTAATCACCAGTAG	This study
pBcsG$^{\Delta N}$ Y277A	Substrate-binding mutant	F:CCACCTCCGCGAGTGGCCCGGCGGCGATC R:GATCGCCGCCGGGCCACTCGCGGAGGTGG	This study
pBcsG$^{\Delta N}$ S278A	Catalytic mutant	F:CCACCTCCTACGCGGGCCCGGCGGC R:GCCGCCGGGCCCGCGTAGGAGGTGG	This study
pBcsG$^{\Delta N}$ H396A	Substrate-binding mutant	F:CTACAACACGCTTCCACTGGCGGACGGCAACCATTATCC R:GGATAATGGTTGCCGTCCGCCAGTGGAAGCGTGTTGTAG	This study
pBcsG$^{\Delta N}$ E442A	Metal-binding mutant	F:GGTCGTGGTGCCGGCACACGGCGGCGCG R:CGCGCCGCCGTGTGCCGGCACCACGACC	This study
pBcsG$^{\Delta N}$ H443A	Metal-binding mutant	F:GTCGTGGTGCCGGAAGCGGGCGGCGCGC R:GCGCGCCGCCCGCTTCCGGCACCACGAC	This study

F: forward primer, R: reverse primer, restriction sites and mutations denoted by underline

Recombinant plasmids were introduced into chemically competent *E. coli* BL21 (DE3) supplied by Novagen as an expression host (Table 3) by incubating a mixture of cells and plasmid at 42°C for 1 min. These cells were allowed to recover at 4°C for 5 min, followed by incubation at 37°C in 500 µL of Lauria Broth (LB) media (10 g tryptone, 5 g yeast extract, 10 g NaCl per liter of culture). The resulting culture was spread on LB agar plates (LB plus 15g of agar per liter of media) containing 50 µg/mL kanamycin and incubated at 37°C for 12-18 h to select for successful transformants. Pure colonies were picked from this plate and used to inoculate 20 mL of LB containing 50 µg/mL kanamycin per liter of culture. This media was incubated 12-18 h (or until turbid) at 37°C with shaking at 220 rotations per minute (rpm). These cultures were used to inoculate each liter of Super Broth (SB) media (32 g tryptone, 20 g yeast extract, 10 g NaCl per liter of culture) containing kanamycin at a final concentration of 50 µg/mL. These cultures were incubated at 37°C and 220rpm until an optical density at 600nm (OD_{600}) of > 0.6 was reached. Expression was induced by the addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG) of 1mM, as expression of the recombinant gene is under the control of the inducible T7 overexpression system encoded on the pET28a plasmid. Cells were harvested by subjecting the culture to centrifugation at 5000 x *g* for 15 min, decanting the supernatant, and storing the resulting pellet at -20°C until needed.

Table 2: Recombinant proteins used in this study

Construct	*Ec*BcsG$^{\Delta N}$	*Ec*BcsG
Molecular mass (Da)†	45345	62086
Length (residues)	417	581
Predicted pI†	6.11	6.04
Extinction coefficient (M^{-1} cm^{-1}) λ=280 nm†	122965	49390
His-tag	N-terminal	N-terminal

†Values estimated by ProtParam tool[89]

Table 3: Strains used in this study.

Strain	Genotype	Source or reference
E. coli TOP10	F$^-$ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80l*acZ*ΔM15 Δl*acX74 nupG recA1 araD139* Δ(*ara-leu*)7697 *galE15 galK16 rpsL*(Str^R) *endA1* λ$^-$	Invitrogen
E. coli BL21	B F$^-$ *ompT gal dcm lon hsdS*$_B$($r_B^-m_B^-$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K-12}$(λ^S)	Novagen
E. coli AR3110	K-12 W3110 Δ*bcsQ::bcsQ*	32
E. coli AR3110 Δ*bcsG*	AR3110 *bcsG::*scar	32

4.2 Cell lysis and recombinant protein purification

Cell pellets were resuspended to homogeneity in a minimal volume of buffer A (50 mM Tris, 300 mM NaCl, 5 mM MgCl$_2$ pH 8.0), typically 20 mL / L culture. Cell lysis was performed on this suspension using a Constant Systems cell disrupter operating at 17kpsi. The soluble cellular fraction was isolated from the crude lysate by subjecting it to centrifugation at 28000 x *g* for 45 min to remove insoluble debris. Initial His$_6$-tagged protein purification from the soluble cellular fraction was performed by immobilized metal ion affinity chromatography (IMAC). The soluble cellular fraction containing the recombinant EcBcsG$^{\Delta N}$-His$_6$ tagged fusion proteins was incubated in 1-3 mL nickel-nitrilotriacetic acid agarose (Ni-NTA) resin for 1 h at 4°C. The mixture was passed over an open column to separate the Ni-NTA from the unbound cellular fraction. The Ni-NTA was washed with 50 mL of buffer A, followed by washing in 50 mL of wash 1 buffer (buffer A with the addition of 20 mM imidazole) to remove unwanted contaminants. His$_6$-tagged protein was eluted from the column by the addition of 20 mL of elution buffer (buffer A with the addition of 250 mM imidazole) to the resin.

Purification to homogenity was performed using size exclusion chromatography (SEC) or ion-exchange chromatography (IEC) where deemed necessary. Ion exchange chromatography was performed on a GE Akta Pure instrument using a GE Q FF HiTrap anion exchanger. Eluted protein from IMAC purification was buffer exchanged with dialysis 12 h against 50 mM Tris-HCl pH 8.0 and the sample was passed over the column twice to facilitate binding. Protein was eluted with a gradient of 50 mM Tris-Hcl pH 8.0 with increasing concentrations of sodium chloride, up to 1 M. Fractions containing protein, typically eluting in the range of 75-150 mM NaCl, were pooled. Size exclusion chromatography was performed on a GE Akta Pure instrument using a BioRad Enrich SEC 650 column. His_6-*Ec*BcsG from IMAC purification was dialyzed against SEC buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$) for 12 h. The resulting protein solution was concentrated to a volume of 1 mL using a GE Healthcare centrifugal concentrator (30,000 Da molecular weight cut-off) and loaded onto the SEC column. The protein was eluted with SEC buffer at a rate of 1 mL min^{-1} and fractions containing protein were pooled. The purity of protein after each purification step was assessed using SDS-PAGE according to the protocol by Laemmli (1970)[90]. Protein quantification was performed with UV spectrophotometry (λ = 280 nm) using a BioTek Cytation 5 imaging microplate reader and the molar extinction coefficients listed in Table 2.

4.3 Functional bioassay

4.3.1 Synthesis of paranitrophenyl phosphoethanolamine. A solution of 1.1 mL (7.11 mmol) N-(tert-butoxycarbonyl) ethanolamine (N-Boc ethanolamine) was prepared in 20 mL of anhydrous tetrahydrofuran (THF). This mixture was slowly added to a vaccuum-purged and stoppered 50 mL round bottom flask containing a solution of 3.65 g paranitrophenyl phosphodichloridate and

2.30 mL anhydrous pyridine in 35 mL anhydrous THF as solvent. The mixture was placed at 4°C for 1 h, and then placed at 20°C for a minimum of 7 h with agitation. A solution of 2.3 mL pyridine in 10.2 mL of water was added to this reaction mixture for 10 min with agitation. Following this incubation, the solvent was removed by rotary evaporation under vacuum, producing a faintly yellow liquid organic phase with an approximate volume of 10 mL. This organic intermediate was diluted in approximately 50 mL of water and washed in at least three 50 mL volumes of ethyl acetate. The aqueous phase was acidified with two 50 mL volumes of 5% w/v potassium bisulfate until a pH of between 3 and 4 was achieved. The desired organic product was extracted with three 50 mL volumes of methylene chloride. The extracted organic phase was washed with one 50 mL volume of saline water, dried over sodium sulfate, and concentrated by rotary evaporation under vacuum. This reaction intermediate was sampled and dissolved in deuterated chloroform for analysis by H^1, C^{13} and F^{19} nuclear magnetic resonance (NMR) spectroscopy on an Agilent 400 MHz spectrometer. To cleave the N-Boc protecting group, a 72 mg quantity of the dried intermediate was dissolved in a mixture of 1 mL methylene chloride containing 0.1 mL trifluoroacetic acid and stirred at 20°C for at least 1 h. The reaction mixture was concentrated by rotary evaporation under vacuum and the residue was triturated with two volumes of 1 mL of ethyl acetate. The resulting product crystallized as a fine white crystalline solid of paranitrophenyl phosphoethanoamine (*p*-NPPE) trifluoroacetate. The crystals were washed and recrystallized in ethyl acetate. The identity of the organic product was verified by dissolving a small quantity in deuterated DMSO and analyzing by NMR spectroscopy as described above.

4.3.2 Cellulose transferase assay. Purified *Ec*BcsG$^{\Delta N}$ was incubated with *p*-NPPE as the donor of the phosphoethanolamine group and cello-oligosaccharides as the ultimate acceptor for the transferase reaction. Unless otherwise specified, the reaction conditions were assayed with 50 µM *Ec*BcsG$^{\Delta N}$ in a 96-well microtiter plate (Sarstedt) using a total well volume of 100 µL with a buffer solution containing 50 mM HEPES pH 7.5 and 50 mM sodium chloride. The reaction mixtures were incubated at 37°C for 30 min and monitored every 1 min for their absorbance at 414 nm. A standard curve (0 -100 µM) was prepared by dilution of a *para*-nitrophenolate standard solution (10 mM, Sigma) and used to calculate molar equivalents of *para*-nitrophenolate released by enzymatic activity. Negative controls were prepared with the absence of *Ec*BcsG$^{\Delta N}$ from reaction mixtures. Kinetics were performed with variable concentrations of *p*-NPPE (0-10 mM) and cellopentaose (0-2.5 mM) as specified. Kinetic parameters were calculated using GraphPad Prism. The esterase activity of *Ec*BcsG$^{\Delta N}$ was measured in the absence of cello-oligosaccharides in the reaction mixture, whereas transferase activity was measured indirectly in the presence of both donor and cello-acceptor substrates. Dithiothreitol (DTT; 1 mM) and ethylenediaminetetraacetic acid (EDTA; 1 mM) were introduced to *Ec*BcsG$^{\Delta N}$ by buffer exchange for at least 12 h prior to measurement of the enzymatic rate. All measurements were repeated in at least triplicate.

4.3.3 Thin-layer chromatography. Samples from the cellulose transferase assay were prepared for thin-layer chromatography (TLC) by centrifugation at 10,000 x *g* for 5 minutes to remove precipitate. The pellet was discarded, and the supernatant was retained as the analytical sample. Samples were spotted onto silica plates (Sigma) with samples at 1 cm intervals, a minimum distance from the solvent line of 1 cm using glass capillary tubes approximately 4-8 times and allowed to dry 10 min between spots. The mobile phase used to separate the enzymatic products was a solution of ethyl acetate, water, and methanol in an 8:3:4 ratio, respectively. The plate was placed in a 10 L tank, vapor-equilibrated with 50 mL of mobile phase and allowed to run until the solvent front reached 1 cm from the top of the plate, approximately 45 min – 1 h. The plate was dried by heating at 90°C on a hot plate and stained in a solution of 1 g/L α-napthol dissolved in methanol, sulfuric acid and water in a 25:3:2 ratio, respectively. The stain was developed by heating at 90°C for 5 minutes or until spots were visible.

4.3.4 Liquid chromatography-mass spectrometry (LC-MS).

Liquid chromatography–mass spectrometry analyses were performed on an Agilent 1200 HPLC interfaced with an Agilent UHD 6530 Q-ToF (quadrupole time-of-flight) mass spectrometer at the Mass Spectrometry Facility of the Advanced Analysis Centre, University of Guelph. A C18 column (Agilent Extend-C18 50 mm x 2.1 mm 1.8 µm) was used for chromatographic separation with the following solvents: water with 0.1% (v/v) formic acid (A) and acetonitrile with 0.1% (v/v) formic acid (B). The mobile phase gradient was as follows: initial conditions were 10% B hold for 1 min then increasing to 100% B in 29 min followed by column wash at 100% B for 5 min and 20 min re-equilibration. The flow rate was maintained at 0.4 mL/min. The mass spectrometer electrospray capillary voltage was maintained at 4.0 kV and the drying gas temperature at 250° C with a flow rate of 8 L/min. Nebulizer pressure was 30 psi and the fragmentor was set to 160. Nozzle, skimmer and octapole RF voltages were set at 1000 V, 65 V and 750 V, respectively. Nitrogen (purity >99%) was used as both nebulizing, drying and collision gas. The mass-to-charge ratio was scanned across the m/z range of 50-1500 m/z in 4 GHz (extended dynamic range) positive and negative ion modes. Data were collected by data independent MS/MS acquisition with an MS and MS/MS scan rate of 1.41 spectra/sec. The acquisition rate was set at 2 spectra/s. The mass axis was calibrated using the Agilent tuning mix HP0321 (Agilent Technologies) prepared in acetonitrile. Mass spectrometer control, data acquisition and data analysis were performed with MassHunter® Workstation software (B.04.00).

LC-MS/MS analysis of tryptic peptides. LC-MS/MS **(**liquid chromatography tandem mass spectrometry) analyses were performed on an Agilent 1200 HPLC interfaced with an Agilent UHD 6530 Q-TOF mass spectrometer. Separation of the peptides was achieved by reverse-phase HPLC using a C18 column (Agilent AdvanceBio Peptide Map, 100 mm × 2.1 mm 2.7 µm) previously equilibrated with 2% (v/v) acetonitrile in 0.1% (v/v) formic acid at a flow rate of 0.2 mL min^{-1}. Following injection of samples, a linear gradient to 45% (v/v) acetonitrile in 0.1% (v/v) formic acid was applied over 40 min and then to 55% (v/v) acetonitrile over 10 min. The column was washed with 95% (v/v) acetonitrile in 0.1% (v/v) formic acid for 5 min before re-equilibration in starting solvent. The mass-to-charge ratio was scanned across the m/z range of 300−2000 m/z in 4 GHz (extended dynamic range positive-ion auto MS/MS mode). Three precursor ions per cycle were selected for fragmentation. The instrument was calibrated with the ESI TuneMix (Agilent). Raw data files were loaded directly into PEAKS 7 software (Bioinformatics Solutions Inc.) where the data were refined and subjected to deNovo sequencing and database searching. The following modifications were considered within the search parameters: Met oxidation, and phosphorylation or pEtN-ylation of Thr, Ser, Cys, Tyr, and His residues. The data were searched against the *Ec*BcsG sequence. The tolerance values used were 10 ppm for parent ions.

4.4 Functional complementation

Cell lines. *E. coli* AR3110, W3110 (AR3110 *bcsQ::Stop),* AR3110 *ΔbcsG ΔcsgBA* (AR3110 *bcsG::*scar, *csgBA*::*kanR*), AR3110 *ΔbcsG* (AR3110 *bcsG::*scar) and AR3110 *ΔcsgBA* (AR3110 *csgBA*::*kanR*) were received from Thongsomboon and Cegelski (Standford U) as a gift. *E. coli* AR3110 *ΔbcsG* was made chemically competent by first culturing in TYM broth (2% [w/v] tryptone, 0.5% [w/v] Yeast extract, 0.1 M NaCl, 10 mM $MgSO_4$) until an $OD_{600} > 0.6$ was reached. The cells were harvested by centrifugation and washed in TfB I media (10 mM MOPS, 75 mM $CaCl_2$, 10 mM KCl, 15% [v/v] glycerol, pH 7.0). The harvested cells were resuspended in TfB I media and stored at -80°C in 50 µL aliquots. These aliquots were transformed with plasmid according to the protocol outlined above in Section 4.1. Cell lines expressing protein variants of *Ec*BcsG were generated by transforming the pBcsG derivates, whose construction is described in Section 4.1 above, into chemically competent *E. coli* AR3110 *ΔbcsG.*

Pellicle macrocolony phenotype. Biofilm phenotype was assessed in triplicate by inoculating 5 mL of salt-free LB broth containing 50 µg/mL kanamycin and incubating for 5 d at 28°C. The resulting biofilms were removed from the tubes by pouring in glass beakers, discarding the culture media, and staining the biofilms in a solution of 40 µg/mL Congo Red for at least 10 min. The biofilms were then washed in water and imaged on a Bio-Rad VersaDoc 4000 instrument using both a visible white-light filter.

Adherence assay. The adherence potential of biofilms were assessed semi-quantitatively using a crystal violet assay. Eight replicates of each cell line were inoculated into 96-well microtiter plates with each well containing 250 µL of salt-free LB broth containing 50 µg/mL kanamycin and the plates incubated for 5 d at 28°C. The media was removed and the biofilms were washed in three 200µL volumes of sterile phosphate buffered saline (12 mM sodium phosphate, 137 mM sodium chloride, pH 7.4), then stained in 200 µL of 0.1% (v/v) crystal violet solution. The residual crystal violet was washed away with 3 x 200 µL volumes of sterile phosphate buffered saline and allowed to dry for 1 h. The stained biofilm was resolubilized in 200 µL of 30% (v/v) acetic acid and the absorbance was measured at 550 nm in a BioTek Cytation5 imaging plate reader.

Expression of recombinant *Ec*BcsG. Biofilms were grown in 96-well microtiter plates containing 200 µL of salt-free LB with 50 µg/mL kanamycin and 250 µM IPTG at 28°C for 96 h. The biofilms were harvested by mechanical disruption (*i.e.* aspiration by micropipette) and subsequent ultracentrifugation at 21,100 x *g* for 5 min. To prepare the samples for Western blotting analysis, the culture media was removed, and the pellets were resuspended in SDS sample buffer (50 mM Tris-HCl pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 1 mM DTT, 0.02% w/v bromophenol blue) and lysed by heating at 100°C for 1 h. To assess the cellular quantity of *Ec*BcsG supplied *in trans,* these samples were immunoblotted according to the protocol by Burnette (1981)[91] with an antibody against the His-tag conferred by expression from the pET28a vector. Briefly, SDS-PAGE gels were soaked in transfer buffer (12 mM Tris-HCl, 96 mM glycine, 20% (v/v) methanol) and stacked on a nitrocellulose membrane pre-soaked in transfer buffer, loaded into a transfer apparatus (BioRad), and submerged in transfer buffer. Transfer was performed electrophoretically at 100 volts for 1 h 20 min with gentle stirring at 4°C. Transferred membranes were washed in Tris-buffered saline (TBS; 10mM Tris-HCl, 150 mM sodium chloride, pH 7.5) and incubated 1 h in blocking buffer (TBS with 5% (w/v) non-fat milk powder). Buffer was poured off and the blocked membrane was washed in Tween-Tris buffered saline (TTBS; 20 mM Tris-HCl, 2 mM sodium chloride, 0.05% (v/v) Tween-20, pH 7.5), before incubation in primary antibody solution (Mouse anti-His IgG stock, diluted 1000-fold in blocking buffer) for 1 h. Antibody solution was removed, and the blotting membrane was washed twice with 10 mL of TTBS, then incubated in secondary antibody solution (Rabbit-anti mouse IgG-alkaline phosphatase conjugate stock diluted 2000-fold in blocking buffer). Antibody solution was removed, and the blotting membrane was washed once with 15 mL of TTBS and once with 15 mL TBS. To visualize protein bands on the membrane, the blot was washed in BCIP (5-bromo-4-chloro-3-indolyl phosphate) and NBT (nitro- blue tetrazolium chloride) solution for 10 min or until visible indole precipitate formed on the membrane. Blots were imaged on a BioRad VersaDoc 4000 system with an f-stop of 11 and an exposure time of 3 sec.

Minimum inhibitory concentration (MIC) assay. *E. coli* AR3110 and *E. coli* AR3110 Δ*bcsG* were grown for 16 h at 37°C in sterile, cation-adjusted Muller-Hinton media (0.3% (w/v) beef extract, 1.75% (w/v) casein acid hydrolysate, 0.15% (w/v) starch with the addition of 20 mg/L $CaCl_2$ and 10 mg/L $MgSO_4$) and the final OD_{600nm} was adjusted to 0.500 +/- 0.01 with the addition or removal of media. Polymyxin stocks were prepared by dissolving lyophilized peptides (Sigma) in Muller-Hinton media at 64 µg/mL and serially diluting in fresh Muller-Hinton media to achieve the concentrations tested (0-32 µg/mL). Eight replicates were tested per cell line, per peptide in a 96-well microtiter plate (Sarstedt). Microtiter plates containing Muller-Hinton media were inoculated with bacteria at a final OD_{600nm} of 0.050 +/- 0.005 containing 200 µL of liquid media and incubated at 30°C for 24 h. The OD_{600nm} was read on a Cytation5 imaging microplate reader (BioTek).

4.5 X-ray crystallography and structural modelling

Seleno-L-methionine labelling. A minimal media kit for seleno-L-methionine labelling was purchased from Molecular Dimensions. Minimal media was prepared by dissolving 21.6 g of the medium base, containing sodium and potassium phosphate at pH 7.5 in 1 L of ddH_2O and sterilized in an autoclave at 121 °C for 30 min. The media was then supplemented immediately prior to culture by dissolving 5.1 g of nutrient supplement (containing 4g D-glucose, trace vitamins, and all L-amino acids, excluding L-methionine, at a final concentration of 40 µg/mL) into 50mL of ddH_2O and passing through a 0.45 µm syringe filter into sterile medium base. Kanamycin and seleno-L-methionine, at concentrations of 50 µg/mL and 40 µg/mL, respectively, were also supplemented aseptically prior to culture. The medium was inoculated with 50 mL of starter culture from LB media grown 18 h at 37°C with 220 rpm shaking. Prior to inoculation, the cells were separated from the media (*i.e.* containing unlabelled methionine) by centrifugation at 4000 x *g* for 15 min and the media discarded. The cell pellet was washed in 40 mL of sterile ddH_2O and was again separated from the water by centrifugation at 4000 x *g* for 15min. The ddH_2O was discarded and the cell pellet was resuspended in minimal culture media.

Growth of protein crystals. Recombinant *Ec*BcsG$^{\Delta N}$ was expressed and purified to homogeneity as outlined above in sections 4.1 and 4.2 as above for selenomethionine labelled protein. Prepared protein was concentrated in centrifugal filter units (30,000 Da MWCO; Pall Corporation) to a concentration of 1 mM. Screening of crystallization conditions was performed using a Crystal Gryphon instrument (Art Robins Technologies) using 1 µL volumes of protein in a 1:1 and 2:1 ratio with the mother liquor using the sitting drop method. The commercially available MCSG suite (Crystalgen) was screened in its entirety for conditions permissible to crystal growth. After suitable condtions were identified, protein crystals were grown using the hanging drop method in 24-well pregreased crystal plates (Crystalgen). Crystals were grown against 0.1 M HEPES:NaOH, 20% (w/v) polyethylene glycol 4000, 10% (v/v) 2-propanol, pH 7.5 in 1:1, 2:1 and 1:2 ratios of protein solution (1 mM) over 500 µL reservoir volumes, and incubated at 18°C in a Molecular Dimensions crystal growth incubator. Stored plates were periodically observed by microscopy (Olympus SZX16 Stereomicroscope) for precipitation and crystal formation. The *Ec*BcsG$^{\Delta N}$ cosusbtrates *p*-NPPE, cellobiose, cellopentaose, and LysoPE were dissolved in the crystallization condition at a concentration of 5 mM and was used to set up as drops with *Ec*BcsG$^{\Delta N}$ to attempt to crystallize the enzyme-cosubstrate complex.

X-ray diffraction analysis. Crystals were harvested from the crystal drop and subjected to immersion for 5-20s in a cryoprotectant solution of 55% (v/v) 2-propanol, 10% (v/v) poly(ethylene glycol) 4000, and 50 mM HEPES pH 7.5. The cryoprotected crystals were submerged in liquid nitrogen until needed. Crystals were diffracted using synchrotron X-ray light at the Canadian Light Source, Saksastoon, SK, Canada on beamline 08ID-1 and 08BM-1, and at the Advanced Photon Source, Woodridge, IL, USA on beamline 17-ID-B. The presence of Se in crystals was assessed with X-ray fluorescence by scanning an X-ray energy spectrum +/- 0.2 keV from the theoretical Se K edge (12.6 keV). For experimental phasing, single-wavelength anomalous diffraction (SAD) experiments were performed by diffracting crystals at 12.656 keV, an X-ray energy equal to the Se K edge and a complete 360° single-wavelength anomalous diffraction (SAD) dataset was collected for initial structure solution, using 720 images of width 0.5°, an exposure time of 0.2 s per image, and a transmittance of 100%. Unlabelled crystals were diffracted with incident X-ray radiation with a wavelength of 1.0000 Å (12.398 keV) with the same collection parameters.

Data processing and structure solution/refinement. Collected data were processed by indexing, scaling and integrating frames in the XDS and/or Mosflm software suites[92]. Processed SAD data was used to attempt structure solution using experimental phasing in the PHENIX software package[93]. The native structure was solved using molecular replacementwith the program PHASER[93] using the Se-labelled structure as a search model[94]. Structural refinement was performed using iterative rounds of automated refinement using PHENIX[93] followed by manual model building and refinement in Coot[95]. Refinement progress was monitored by following the reduction (and convergence) of the R_{work} and R_{free}. Visualization of the model was done in the PyMOL software package. The biological assembly of *Ec*BcsG$^{\Delta N}$ was assessed using the PDBePISA server[96]. Metal content was assessed with the CheckMyMetal server[97].

Inductively coupled plasma – atomic emission spectroscopy. The metal content of *Ec*BcsG$^{\Delta N}$ was determined using ICP-AES. *Ec*BcsG$^{\Delta N}$ was purified as described in section 4.2 above in the absence of exogenous metals in the purification buffers. The analyzed samples were concentrated to 1 mM and digested in 200µL of HNO_3 by heating at 70°C for 1 h until all *Ec*BcsG$^{\Delta N}$ was dissolved. The sample was then passed through a 0.45 µm centrifugal filter and diluted to 50 µM *Ec*BcsG$^{\Delta N}$ and 2% v/v HNO_3 with 18 megohm-cm^{-1} purity water and analyzed for S, Mg, Co, Cu and Zn at two analytical wavelengths for each element (Mg: 285.213 and 279.077 nm; Zn 206.200 and 213.857 nm; Cu 327.393 and 324.752 nm; Co 228.616 and 238.892 nm) in a Perkin Elmer Optima 8000 ICP-AES. Prepared solutions of copper chloride, magnesium chloride, zinc sulfate, and 2% (v/v) HNO_3 were used as analytical standards and the blank, respectively.

Molecular phylogenetics. The sequence of putative pEtN transferase genes described by Harper *et al.*[63] were downloaded from the National Center for Biotechnology Information (NCBI) protein sequence database. Multiple-sequence alignments were performed using the constraint-based multiple alignment tool COBALT available at NCBI. For phylogenetic analysis, the alignment data were visualized with FigTree 1.4.3 (http://tree.bio.ed.ac.uk/). Structural homologues of *Ec*BcsG$^{\Delta N}$ were queried using the DALI server[98].

5. Results and Discussion

5.1 Expression and purification of *Ec*BcsG$^{\Delta N}$

Expression of *Ec*BcsG$^{\Delta N}$. High yields of soluble *Ec*BcsG$^{\Delta N}$ were observable in the cell lysate of *E. coli* BL21 (DE3) used as an expression host following culture and IPTG (1 mM) induction (Figure 7) for 16 h at 23°C. Protein yield following preliminary IMAC purification was routinely > 50 mg/L of culture at near homogeneity (Figure 7; Lane 3), and did not warrant further optimization as yield and purity was sufficient for downstream applications. Purification to total

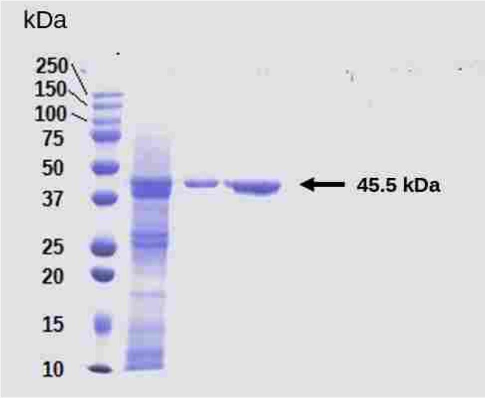


Figure 8: SDS-PAGE of a typical *Ec*BcsG$^{\Delta N}$ purification. Lanes 1: Ladder, Lane 2: cell lysate, Lane 3: IMAC elutions, Lane 4: IEC elutions.

homogeneity at high yield (> 45 mg / L of culture) was possible by secondary purification using either size exclusion chromatography or anion exchange chromatography (Figure 7; Lane 4).

5.2 Crystallization of *Ec*BcsG$^{\Delta N}$

Native *Ec*BcsG$^{\Delta N}$. Primary purification of *Ec*BcsG$^{\Delta N}$ was performed by IMAC and secondary purification was performed by SEC or IEC to total homogeneity and concentrated to 45 mg/mL (1 mM) prior to the preparation of crystallization plates. Initial screening of the MCSG 1-4 Crystallization Suites identified two lead conditions permissible to *Ec*BcsG$^{\Delta N}$ crystal growth: MCSG4-G11 (0.1 M HEPES:NaOH pH 7.5, 20% (w/v) poly-(ethylene glycol) 4000, 10% (v/v) 2-propanol) and MCSG4-F7 (0.2 M magnesium chloride, 0.1 M MES:NaOH pH 6.5, 25% (w/v) PEG 4000). Further expansion screening of these conditions by altering the pH or concentrations of salts and precipitants failed to retain crystal growth and instead resulted in rapid *Ec*BcsG$^{\Delta N}$ precipitation. Crystals of native *Ec*BcsG$^{\Delta N}$ appeared to form optimally in the MCSG4-G11

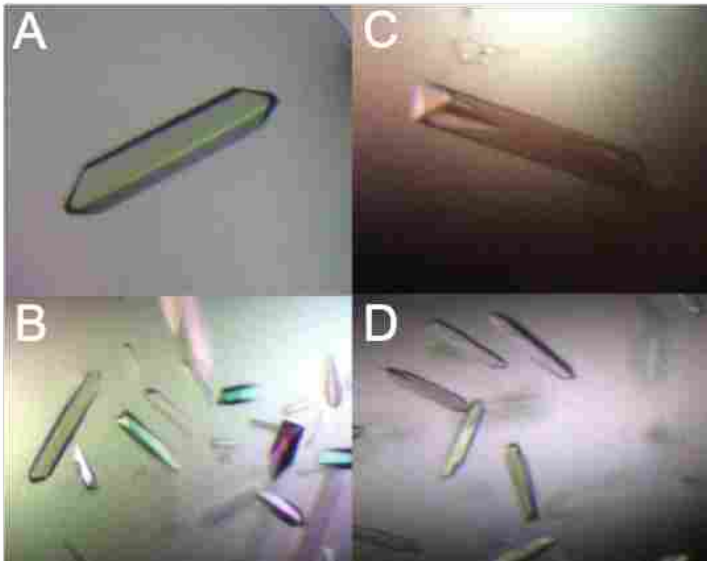


Figure 9: Representative *Ec*BcsG$^{\Delta N}$ crystal size and morphology. Selenomethionine-labelled *Ec*BcsG$^{\Delta N}$ crystals grew with rhomboid morphology, ranging from approximately 150 x 50 µm (A) to 30 x 50 µm (B). Native *Ec*BcsG$^{\Delta N}$ crystals also grew with a rhomboid morphology, ranging from to 50 x 200 µm (C) to 50 x 100 µm (D).

condition (0.1M HEPES:NaOH pH 7.5, 20% (w/v) poly-(ethylene glycol) 4000, 10% (v/v) 2-propanol) when drops were prepared using *Ec*BcsG$^{\Delta N}$ SEC buffer (20 mM Tris, 150 mM NaCl, 5 mM $MgCl_2$, pH 7.5) as the solvent for recombinant protein. The presence of Mg^{2+} was critical for crystal formation, as plates set up in the absence of Mg^{2+} did not display crystal growth following 8 weeks of incubation at 22°C. In the presence of Mg^{2+}, crystal growth could be observed within 5-7 days of incubation as rod shaped crystals, ranging in size from approximately 30 x 50 µm to 50 x 200 µm. Ideal crystal growth was observed after 28 days of elapsed incubation, at which point cracking and damage to the crystalline lattice became evident.

Seleno-L-methionine labelled *Ec*BcsG$^{\Delta N}$. Optimal crystal formation occurred for labelled *Ec*BcsG$^{\Delta N}$ similarly to native *Ec*BcsG$^{\Delta N}$ following purification by IMAC and SEC. Crystals of labelled enzyme grew readily in the MCSG4-G11 condition previously identified by high-throughput screening, and crystal growth also appeared to be dependent upon the presence of Mg^{2+}. Crystals appeared in the conditions tested after 7 days of incubation at 22°C and appeared to cease growth after 24 days of incubation. Large, highly ordered rod-shaped crystals were observed, similarly to crystals grown with native *Ec*BcsG$^{\Delta N}$, ranging in size from approximately 50 x 100 µm to 100 x 250 µm (Figure 8).

5.3 The X-ray crystal structure of *Ec*BcsG$^{\Delta N}$

Model quality. The *de novo* structure of *Ec*BcsG$^{\Delta N}$ was determined using the single-wavelength anomalous diffraction (SAD) method with high-resolution (1.26 Å) anomalous scattering at the selenium K edge (12.656 keV, 0.979647 Å) on crystals prepared with uniformly L-selenomethionine labelled *Ec*BcsG$^{\Delta N}$ (herein *Ec*BcsG$^{\Delta N}$-Se). The diffraction experiment statistics are available in Table 4. The high-resolution diffraction data were reduced and included reflections at a maximum resolution of 1.44 Å to ensure overall data quality as indicated by correlation coefficients (*e.g.* $CC_{1/2} > 0.8$). Model building and refinement improved the R_{work} and R_{free} from 0.212 and 0.229 to 0.197 and 0.212, respectively. Analysis of data suggested crystals grew in space group $P\ 2_1\ 2_1\ 2_1$ with the model containing two copies of *Ec*BcsG$^{\Delta N}$ in the asymmetric unit, consistent with a Matthews coefficient of 2.09 and a solvent content of approximately 40%. However, we observed no evidence that *Ec*BcsG$^{\Delta N}$ may form a dimer in its biological assembly during purification by size exclusion chromatography. The *Ec*BcsG$^{\Delta N}$ model presented excellent stereochemical parameters (Table 4) and covers almost the entirety of the catalytic domain. The structure is complete at the C-terminus and truncates in an unstructured loop immediately adjacent to the His$_6$-tag at the N-terminal region of the soluble domain. One copy of the protein in the asymmetric unit contains weak electron density about residues Ala^{484} and Ala^{382}, and so these residues could not be reliably built into the model. In the second copy, the fit of Ala^{484} and Ala^{382} to the density is satisfactory and the folding of *Ec*BcsG$^{\Delta N}$ at these residues can still be inferred from this model. No Ramachandran outliers are present, and only 16 residues (2.23%) lie outside the ideally preferred region and in the allowed regions. Only one (0.32%) side chain rotamer outlier, Phe^{359}, is present, and only in one copy of the molecule. There are no obvious artefacts of crystallization, although small amounts of unmodelled density about the active site contribute to the residual statistics.

The completed structure of *Ec*BcsG$^{\Delta N}$ labelled with L-selenomethionine was used as a molecular replacement search model to solve a dataset collected on an unlabelled native *Ec*BcsG$^{\Delta N}$ crystal. The data collection and refinement statistics for the unlabelled crystal are also

Table 4: Diffraction data processing, refinement statistics, and model validation

	*Ec*BcsG$^{\Delta N}$-Se	*Ec*BcsG$^{\Delta N}$
Resolution range (Å)	45.97 - 1.44 (1.491 - 1.44)	47.96 - 1.75 (1.813 - 1.75)
Space group	$P\ 2_1\ 2_1\ 2_1$	$P\ 1\ 2_1\ 1$
Cell dimensions		
a, b, c (Å)	78.71, 94.58, 105.19	52.51, 76.92, 95.92
α, β, γ (°)	90, 90, 90	90, 90.578, 90
X-ray energy (eV)	12656.0	12398.4
Total reflections	1832824 (181077)	520255 (43663)
Unique reflections	142127 (14051)	74391 (6595)
Multiplicity	12.9 (12.9)	7.0 (6.6)
Average $I/\sigma(I)$	19.13 (2.64)	28.56 (7.42)
Completeness (%)	99.99 (99.99)	96.64 (85.96)
Anomalous completeness (%)	99.87	N/A
R_{merge}	0.07407 (1.063)	0.0435 (0.2657)
R_{meas}	0.7719 (1.108)	0.04711 (0.2894)
R_{pim}	0.02148 (0.3069)	0.01794 (0.1126)
CC$^{1/2}$	0.999 (0.773)	0.999 (0.969)
CC*	1 (0.934)	1 (0.992)
Refinement statistics		
Reflections used in refinement	142124 (14051)	74372 (6577)
Reflections used for R-free	7107 (703)	3719 (329)
R_{work}	0.197 (0.272)	0.161 (0.214)
R_{free}	0.212 (0.311)	0.198 (0.236)
CC(work)	0.957 (0.828)	0.965 (0.860)
CC(free)	0.953 (0.760)	0.951 (0.789)
Number of non-H atoms		
Total	6362	6742
Macromolecules	5657	5706
Ligands	3	3
Solvent	702	1033
Protein residues	725	732
RMS bond lengths (Å)	0.007	0.007
RMS bond angles (°)	1.24	1.03
Ramachandran favoured (% ; #)	97.77 ; 709	97.39 ; 713
Ramachandran allowed (% ; #)	2.23 ; 16	1.79 ; 13
Ramachandran outliers (% ; #)	0.00 ; 0	0.82 ; 6
Romater outliers (%)	0.32 ; 1	0.32 ; 1
Clashscore	1.62	9.61
Average B-factor		
Total	23.9	23.5
Macromolecules	22.8	21.7
Metal ions	16.3	19.0
Solvent	32.4	33.2

Values in parentheses correspond to the highest resolution shell. *$I/\sigma(I)$:* intensity of a group of reflections divided by the standard deviation of those reflections. $R_{merge} = \sum |I(k) - \langle I \rangle| / \sum I(k)$, where $I(k)$ and $\langle I \rangle$ represent the diffraction intensity values of the individual measurements and the corresponding mean values. The summation is over all unique measurements. $R_{work} = \sum ||F_{obs}| - k|F_{calc}|| / |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. R_{free} is the sum extended over a subset of reflections excluded from all stages of the refinement.

listed in Table 4. A *P* 2_1 space group resolved to 1.75 Å with packing on an alternate face is observed, although two copies are still present in the asymmetric unit. Refinement of the native structure resulted in an improved R_{free} and R_{work} of 0.161 and 0.198, respectively. The model also covers the entire C-terminal catalytic domain, with truncation at the N-terminus near the His_6 tag. Similar stereochemical parameters are observed in the native structure, indicating a high-quality model of *Ec*BcsG$^{\Delta N}$. The same single rotamer outlier, Phe^{359} is present. Six (0.82%) Ramachandran outliers are present, with thirteen (1.79%) residues lying outside the preferred regions. No obvious artefacts of crystallization were observed in this model.

The overall fold of *Ec*BcsG$^{\Delta N}$. The crystal structure of *Ec*BcsG$^{\Delta N}$ reveals a three-dimension al fold bearing 15 α-helical and 13 β-sheet secondary structure elements (Figures 10 and 11). A 7-stand β sheet is present at the core of the protein, flanked by five α helices. Both a large α-helical component and a small 3-strand antiparallel β-sheet are present at the N-terminal region of the fold, presumably oriented towards the inner membrane *in vivo*, based on the predicted orientation of the N-terminal region of *Ec*BcsG$^{\Delta N}$ and its similarity to predicted homologues. The structure truncates in a loop, presumably the linker between the catalytic domain and the N-terminal transmembrane domain.

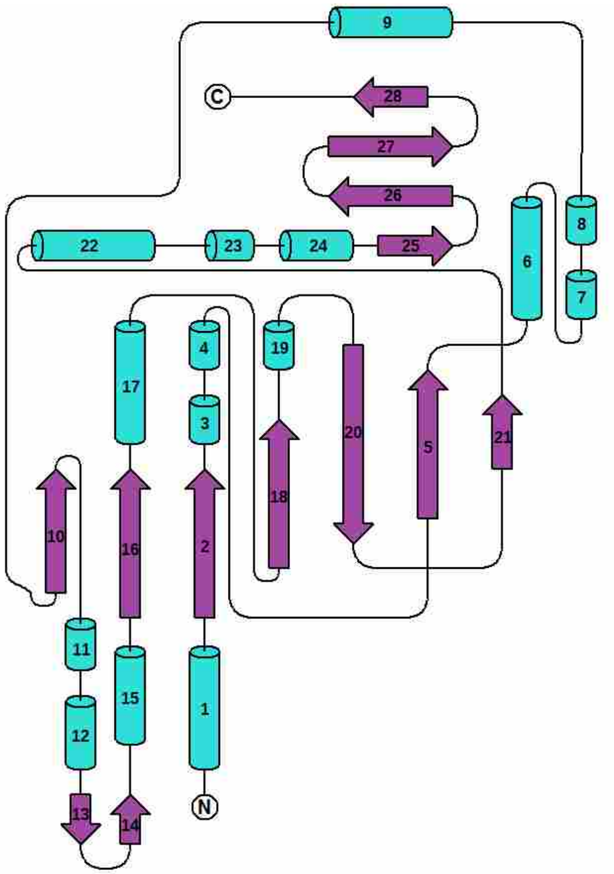


Figure 10: Secondary structure topology cartoon of the overall fold of *Ec*BcsG$^{\Delta N}$. The overall topology shows the $(\beta/\alpha)_7$ fold typical for the alkaline phosphatase superfamily. α-helices are coloured cyan and β-sheets are coloured purple. The N- and C-terminus are denoted as N and C, respectively. Figure generated with TOPDRAW.

5.4 *Ec*BcsG is a member of the pEtN transferase family.

Structural homology. Our results show that the soluble, periplasmic domain of BcsG folds in a globular structure that is typical of the alkaline phosphatase superfamily. A structural homology search on the Dali server[98] showed *Ec*BcsG$^{\Delta N}$ belongs to a family of sulfatases, lipotechoic acid synthetases, and phosphatidyl-ethanolamine transferases. Among the highest scoring results are EptC from *Campylobacter jejuni* (*Cj*EptC, PDB ID 4TN0), EptA from *Neisseria meningitidis* (*Nm*EptA, PDB ID 4KAY), and the mobile colistin resistance factor 1 MCR-1 (PDB ID 5K4P) with z-scores of 20.4, 20.2, 20.4 and 20.2, respectively. A review of the literature reveals all three of these top structural homologues function as phosphoethanolmine (pEtN) transferases, using membrane phosphatidyl-ethanolamine as donor and decorating lipid A-Kdo$_2$ with phosphoethanolamine[64,65,72,77] and belong to a family of related enzymes that have been referred to as the pEtN transferase family[73].

Using PyMOL, a Cα structural superposition of *Ec*BcsG$^{\Delta N}$ with these structures resulted in an average alpha-carbon r.m.s.d. of 2.953 Å across 173 equivalent Cα atoms (Fig. 12). Among the highest regions of conservation in folding architecture are four loops adjacent to the core β-sheet of all structures tested. These loops contribute residues to the active site of the pEtN transferase family and suggest that the active site of *Ec*BcsG may also be comprised of these folds. A subsequent amino acid sequence alignment of *Ec*BcsG$^{\Delta N}$, MCR-1, *Cj*EptC, and *Nm*EptA revealed several blocks of conserved residues among these structures, although the alignment was generally poor due to low sequence identity of *Ec*BcsG and structural features had to be

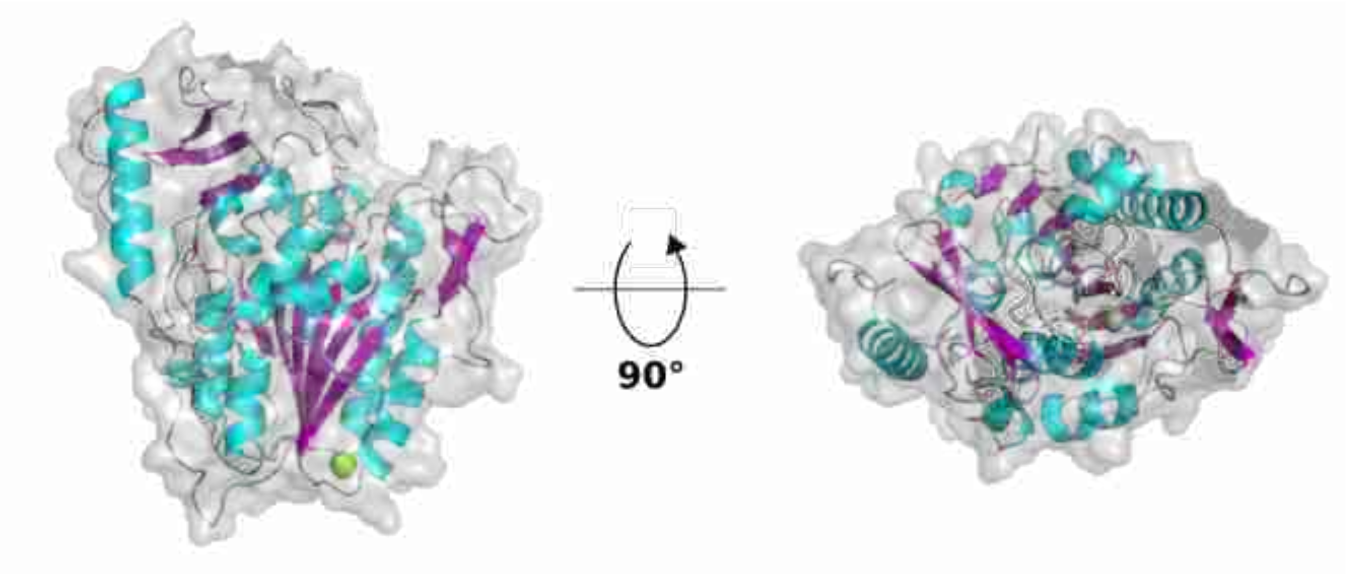


Figure 11: The overall fold of *Ec*BcsG$^{\Delta N}$. The structure is rendered as an α-carbon backbone trace with secondary structure element cartoons. α-helices are coloured cyan, β-sheets are coloured purple, and loops are coloured salmon. The solvent-accessible surface is rendered gray with a solvent radius of 0.5 nm. Rendered using PyMOL.

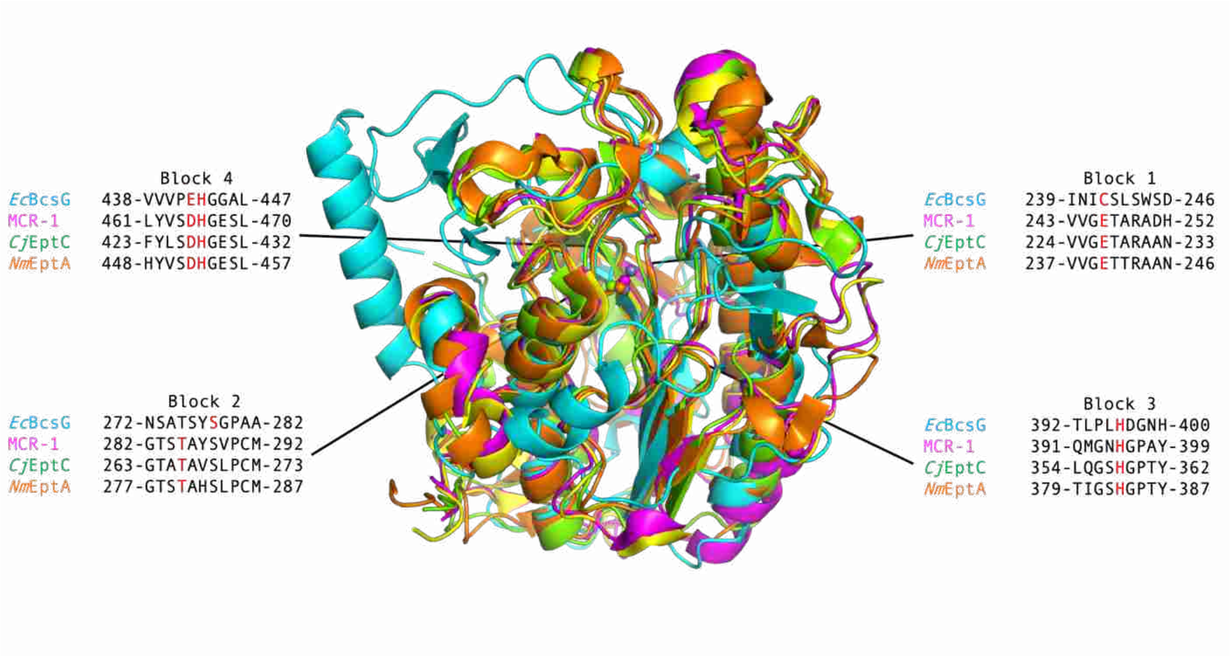


Figure 12: A structural superposition of *Ec*BcsG$^{\Delta N}$ with the homologous phosphoethanolamine transferases MCR-1, *Cj*EptC and *Nm*EptA. Significant folding similarity is observed, particularly across four loops that line the active site pocket of the *Ec*BcsG$^{\Delta N}$ homologues. The amino acid sequence alignment of these loops is shown in blocks 1-4 (inset). Residues highlighted in red are involved in metal binding or catalysis.

manually aligned. The complete sequence alignment is available in Appendix 1 – Figure A8. Not surprisingly, these blocks of conservation are contained within the structurally similar loops of these enzymes (Fig. 12). However, some residues that contribute to the active site of *Nm*EptA, *Cj*EptC, and MCR-1 do not appear conserved in *Ec*BcsG$^{\Delta N}$, and suggest divergent evolution of *Ec*BcsG from these structures, likely owing to their activity on lipid A, a vastly different substrate. Metal coordination in the active site of the MCR-1 like family is achieved by aspartic acid, glutamic acid, histidine, and a threonine residue – observed as phosphothreonine in the structures of MCR-1, *Cj*EptC and *Nm*EptA. Equivalents to the glutamic acid and histidine of other known pEtN transferases are present in *Ec*BcsG$^{\Delta N}$ as residues Glu^{442} and His^{443}. Cys^{243} and Ser^{278} in *Ec*BcsG$^{\Delta N}$ are both positioned near these residues and may instead serve as equivalent to the metal-binding functions of the glutamic acid and threonine of the other pEtN

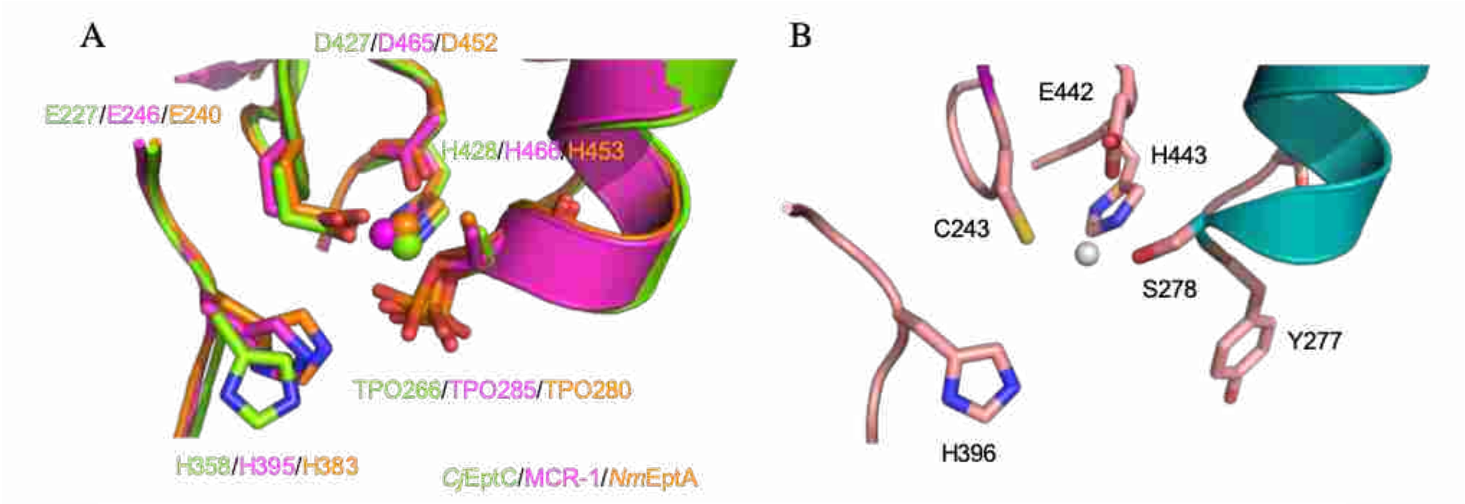


Figure 13: The conserved pEtN transferase active site. The catalytic fold is comprised of a conserved His-Asp-Glu Zn^{2+}-binding triad, a catalytic phospho-threonine nucleophile, and conserved histidine, as demonstrated from representative structures of lipid A pEtN transferases (B). *Ec*BcsG (A) contains an unusual active site for the family, shaped by a Cys in place of the Zn^{2+} binding Glu, and a serine catalytic nucleophile.

transferases, and the hydroxyl of Ser^{278} may also be phosphorylated in a fashion analogous to *Cj*EptC and *Nm*EptA. Additionally, proton abstraction from the catalytic threonine is proposed to be accomplished by the metal-adjacent glutamic acid in the pEtN transferse family[72]. However, despite these amino acid equivalents that may plausibly serve catalytic roles in *Ec*BcsG, the mechanism is unclear from the structure of the holoenzyme alone.

Molecular Phylogenetics. Harper *et al.* reconstructed the molecular phylogeny of known and predicted pEtN transferases and showed the family contains distinct subfamilies, clustered based upon the substrate to which the enzyme transfers the pEtN group[63]. Known substrates for which pEtN transferases have specificity include the 6'OH of the lipopolysaccharide inner core Heptose II[99], the 3' OH of lipopolysaccharide inner core Heptose II[99], various positions of the lipopolysaccharide outer core Gal[63], the 1' phosphate of lipid A[72,77], various periplasmic or extracellular proteins[63,74], and various periplasmic glycans, including the osmoregulated periplasmic glucan (OPG) in *E. coli*[100] and the periplasmic protein N-linked glycan in *C. jejuni*[101].

To date, all available structures of the pEtN transferase family are limited to those belonging to the subfamily specific for the 1’ phosphate of lipid A. Furthermore, Stogios and colleagues[73] constructed a sequence-based phylogenetic analysis of the enzymes belonging to this subfamily specific for the 1’ phosphate of lipid A and found that the subfamily further contains sequence dissimilarity with at least four major clades[73]. Although the membership of *Ec*BcsG to the pEtN transferase family is apparent from its structural homology to the lipid A transferases, its placement within the family is unclear due to the limited number of available structures, of which all belong to the subfamily with substrate specificity for lipid A. To address the relationship of *Ec*BcsG to the other pEtN transferase family members, a sequence-based phylogenetic analysis was performed with a sequence-based alignment using COBALT and visualized as a maximum-likelihood tree using FigTree (Fig. 14). The phylogenetic analysis was based on the known and predicted pEtN transferase family members compiled by Harper *et al.*[63] using sequence data available from repositories and was updated to include *Ec*BcsG. The *Ec*BcsG sequence aligned poorly to other known and predicted pEtN transferases and clustered independently, which was not surprising given the divergence of residues in *Ec*BcsG that are otherwise strictly conserved in the pEtN transferase family. The phylogenetic analysis suggests that BcsG belongs to a novel subfamily of phosphoethanolamine transferases which cluster near to pEtN transferases with specificity for the C6 hydroxyl of lipopolysaccahride Heptose II. This is interesting given that the proposed regioselectivity of *Ec*BcsG is for the C6 hydroxyl of a hexose.

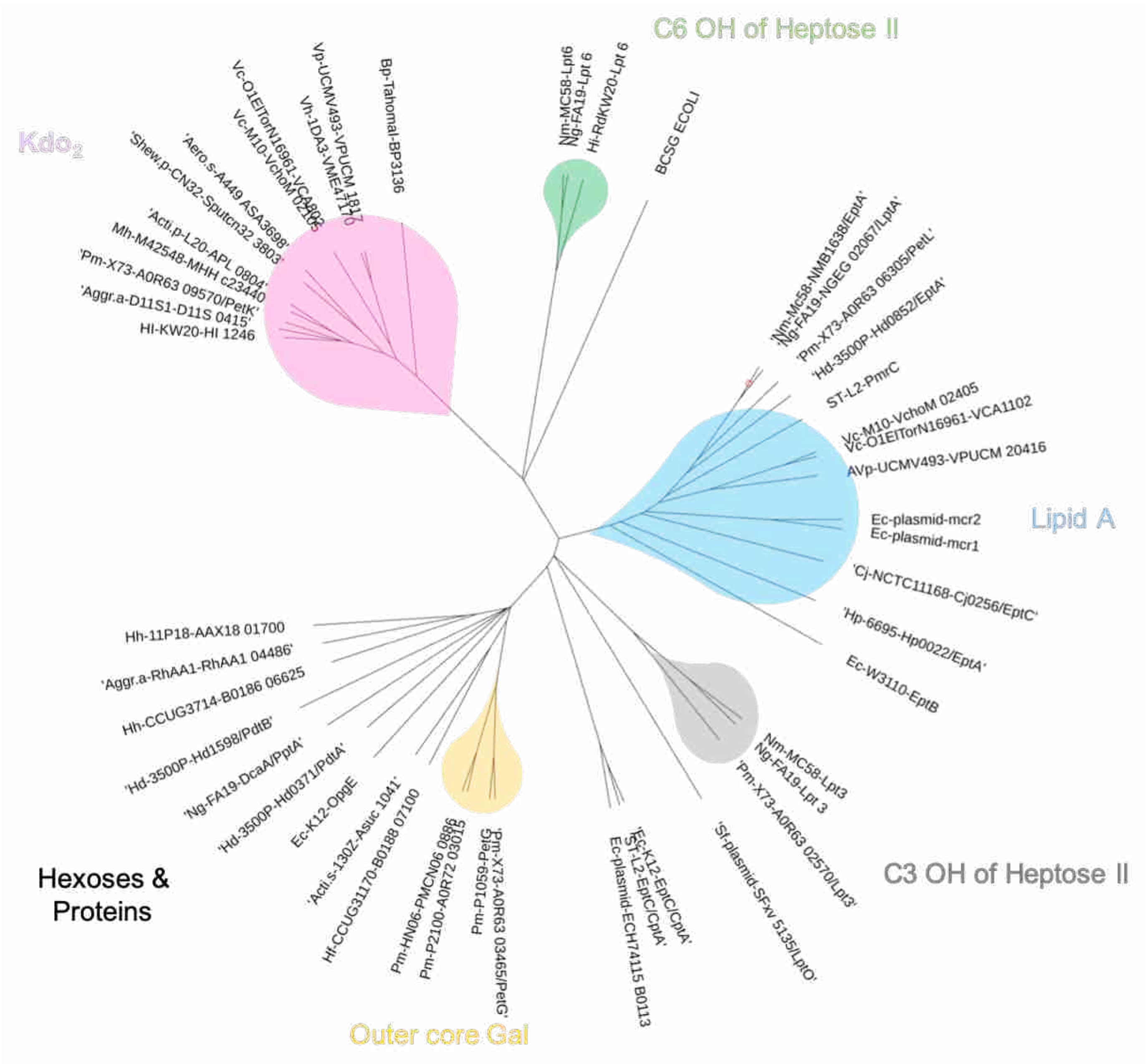


Figure 14: Phylogenetic analysis of representative known and predicted pEtN transferases based on Harper *et al*. Shaded clusters represent subfamilies that share substrate specificity as follows; pink, transferases specific for 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) residues of lipopolysaccharide; green, transferases specific for the C6 OH of Heptose II residues of lipopolysaccharide; gray, transferases specific for the C3 OH of Heptose II residues of lipopolysaccharide; yellow, transferases specific for the outer core Gal residues of lipopolysaccharide; blue, transferases specific for the 1' phosphate of lipid A. Bacterial species names are listed in sequence annotations of the tree as follows; *Acti.p, Actinobacillus pleuropneumoniae; Aero.s, Aeromonas salmonicida; Aggr.a, Aggregatibacter actinomycetemcomitans; Bp, Bordetella pertussis; Cj, Campylobacter jejuni; Ec, Escherichia coli; Hd, Haemophilus ducreyi; Hf, Haemophilus felis; Hh, Haemophilus haemolyticus; Hi, Haemophilus influenzae; Hp, Helicobacter pylori; Mh, Mannheimia haemolytica; Ng, Neisseria gonorrhoeae; Nm, Neisseria meningitidis; Pm, Pasteurella multocida; Sf, Shigella flexneri; Shew.p, Shewanella putrefaciens; ST, Salmonella enterica subsp. enterica serovar Typhimurium; Vc, Vibrio cholerae; Vh, Vibrio harveyi; Vp, Vibrio parahaemolyticus.* Generated with FigTree.

Metal-binding sites. Metal-dependency is typical for the alkaline phosphatase superfamily, although the physiological metal and the coordination environment of superfamily members varies widely[102]. A survey of structurally resolved members of the pEtN transferase family demonstrates dependency on Mg^{2+}, Zn^{2+}, or a combination of these. Because of the overall *Ec*BcsG$^{\Delta N}$ fold typical of metal-dependent members of the alkaline phosphatase superfamily, similarity to the pEtN transferase family, and the apparent stabilizing effect of including Mg^{2+} in purification and crystallization conditions, the structure of *Ec*BcsG$^{\Delta N}$ was surveyed for sites in which Mg^{2+} might bind. In the putative *Ec*BcsG$^{\Delta N}$ active site, the top of helix 6 (residues 278-286) and three loops protruding from the top of β strands 18, 16, and 2 fold into a solvent-accessible pocket. The side chains of Cys^{243}, Ser^{244}, Ser^{278}, His^{396}, Glu^{442}, and His^{443} face in towards this pocket (Fig. 14). A large peak detectable > 7 σ in the F_o-F_C map was calculated in this pocket and was not attributable to the *Ec*BcsG$^{\Delta N}$ macromolecule, observed adjacent the side chains of Cys^{243} Ser^{278}, Glu^{442}, and His^{443} in a tetrahedral geometry, suggesting that a ligand

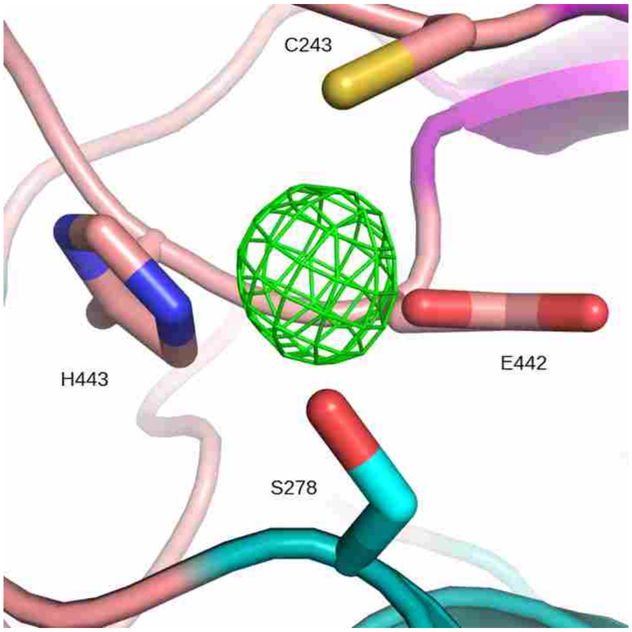


Figure 15: The F_o-F_c map contoured at 7 σ about the putative active site of *Ec*BcsG$^{\Delta N}$. A large positive difference peak in the F_o-F_c map suggests a metal ligand is bound here.

is bound here (Figure 15). Surprisingly, placement of either a Mg^{2+} ion or a water molecule into this electron density refined poorly and suggested a poor fit to the diffraction data. However, a second site at which electron density suggests another ligand-binding site is observed elsewhere in the resolved structure of *Ec*BcsG$^{\Delta N}$. A second large F_o-F_c peak was calculated adjacent to the amide nitrogen of Thr^{478} and Asn^{483}, and the side chain of Asp^{481} (Figure 16). This second ligand site is also coordinated by three highly ordered water molecules in an octahedral arrangement. Occupancy of this site with Mg^{2+} (present from the crystallization conditions) into this density refined optimally and suggested a good fit of this ligand with the X-ray diffraction data. In the second copy of *Ec*BcsG$^{\Delta N}$ present in the asymmetric unit, however, this electron density was not observed and the local folding near this site differs from the ligand-bound state. These observations suggested that Mg^{2+} likely plays a role in *Ec*BcsG$^{\Delta N}$ folding in this local region and accounted for the stabilizing effect of magnesium salts during purification and crystallization of *Ec*BcsG$^{\Delta N}$.

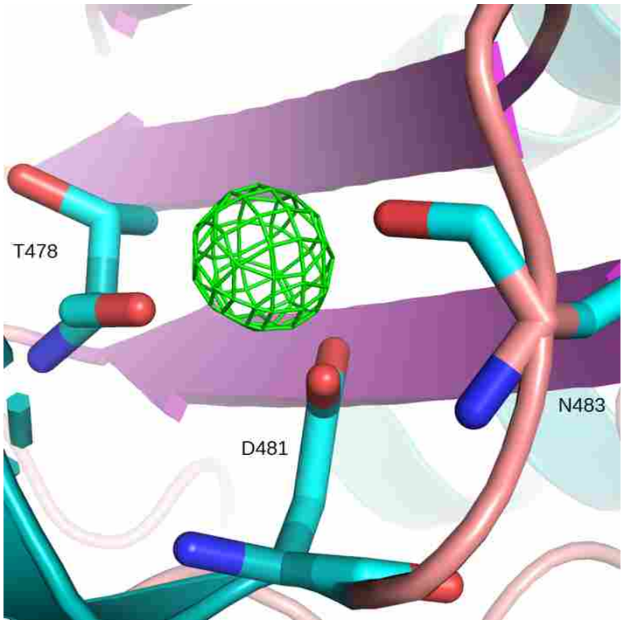
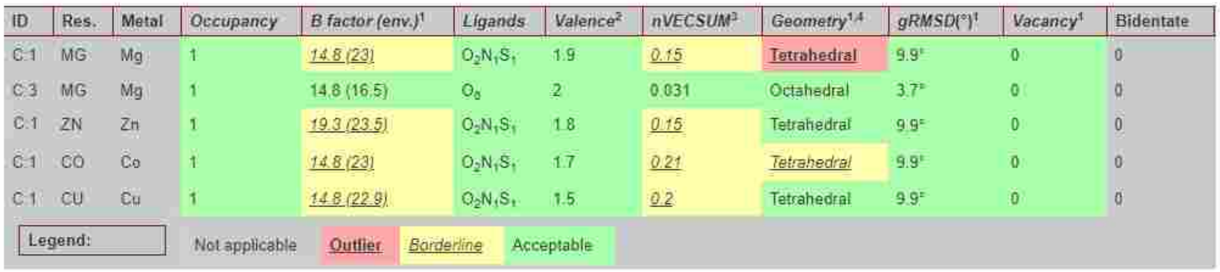


Figure 16: The F_o-F_c map contoured at 7 σ about the secondary metal-binding site. A strong calculated electron density peak is observed bound to D481, N483 and T478, likely a Mg^{2+} ion.

To further investigate the identity of the ligands at these two binding sites, the structure of *Ec*BcsG$^{\Delta N}$ was uploaded to the CheckMyMetal server[97] (https://csgid.org/metal_sites). The results suggest the ligand environment of the secondary ligand site (D481/483/478) is typical of Mg^{2+} binding sites for all parameters tested and provides further evidence that the identity of the secondary site ligand is in fact Mg^{2+}. This analysis also confirmed skepticism that the active site (C243/S278/E442/H443) ligand may also be Mg^{2+}, as tetrahedral geometry is not typical for Mg^{2+} ligands and Mg^{2+} was also a poor fit to the local B factor and bond-valence parameters of the crystal in this region (Fig. 17). Instead, the results suggested a ligand environment in the active site more typical of Co^{2+}, Cu^{2+}, or Zn^{2+}, although no clear preference was observed. Further refinement of the resolved *Ec*BcsG$^{\Delta N}$ structure with all three of these metal ions as ligands in active site resulted in an equally good fit to the experimental data and did not further suggest the identity of the physiological ligand. To unambiguously identify the ligand in the putative active site, *Ec*BcsG$^{\Delta N}$ was purified in the absence of exogenous metal ions using buffers prepared with 18 megaohm cm^{-1} water and was digested in HNO_3 for analysis by ICP-AES. The *Ec*BcsG$^{\Delta N}$ elemental spectra for Zn^{2+}, Cu^{2+}, Co^{2+} and Mg^{2+} unambiguously confirm the identity of the ligands observed by X-ray diffraction (Figure 18). A strong analytical signal, > 40-fold above



ID	Res.	Metal	*Occupancy*	*B factor (env.)*[1]	*Ligands*	*Valence*[2]	*nVECSUM*[3]	*Geometry*[1,4]	*gRMSD(°)*[1]	*Vacancy*[1]	Bidentate
C:1	MG	Mg	1	*14.8 (23)*	$O_2N_1S_1$	1.9	*0.15*	**Tetrahedral**	9.9°	0	0
C:3	MG	Mg	1	14.8 (16.5)	O_6	2	0.031	Octahedral	3.7°	0	0
C:1	ZN	Zn	1	*19.3 (23.5)*	$O_2N_1S_1$	1.8	*0.15*	Tetrahedral	9.9°	0	0
C:1	CO	Co	1	*14.8 (23)*	$O_2N_1S_1$	1.7	*0.21*	*Tetrahedral*	9.9°	0	0
C:1	CU	Cu	1	*14.8 (22.9)*	$O_2N_1S_1$	1.5	*0.2*	Tetrahedral	9.9°	0	0

Legend: Not applicable **Outlier** *Borderline* Acceptable

Figure 17: Output from the CheckMyMetal server. Site ID 1 is the active site and site ID 3 is the secondary site.

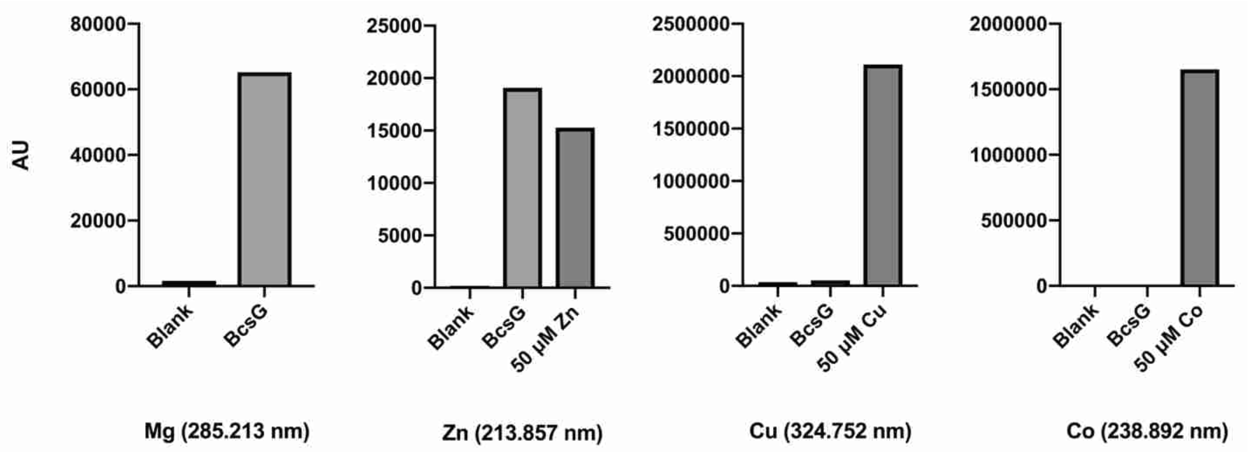


Figure 18: ICP-AES emission spectra at analytical wavelengths for Mg, Zn, Cu and Co. A strong emission singal was observed for Mg and Zn, but not for Cu or Co. Emission of the analytical standards suggest 1 equivalent of Mg and Zn per polypeptide chain.

the blank for Mg^{2+} and > 90-fold for Zn^{2+} was observed, while weak signal intensities for Cu^{2+} and Co^{2+} were observed, approximately 1.5-fold and 4-fold above the blank, respectively. Further analysis of analytical standards of 10 µM Zn^{2+} and 15 µM Cu^{2+} further confirm that the Cu^{2+} and Co^{2+} levels represent negligible quantities, while Zn^{2+} and Mg^{2+} are present in sufficient quantities to conclude that they must comprise the observed ligands in the structure of *Ec*BcsG$^{\Delta N}$. Concurrently, crystals of *Ec*BcsG$^{\Delta N}$ were analyzed by X-ray fluorescence (XRF) spectroscopy near the K edge for Mg (1.3050 keV), Zn (9.6586 keV), Cu (8.9789 keV), and Co (7.7089 keV) and confirmed the presence of Mg and Zn in *Ec*BcsG$^{\Delta N}$ crystals. Final refinement of the structure of *Ec*BcsG$^{\Delta N}$ was completed with these ligand sites occupied by Zn^{2+} and Mg^{2+}.

Internal disulfide bond. While the number of total disulfide bonds identified varies from three (*Cj*EptC and MCR-1) to five (*Nm*EptA) in the pEtN transferase structures examined here, only a single conserved disulfide bond, separated by 10 amino acids, is shared by all three of these structures[65,72,77]. This disulfide bond is shaped by the conserved cysteine pairs Cys262-Cys272 in *Cj*EptC, Cys281-Cys291 in MCR-1, and Cys276-Cys286 in *Nm*EptA (Figure 18) [65,72,77].

The conserved disulfide bond directly fixes the helical secondary structural elements of *Nm*EptA, *Cj*EptC and MCR-1, which contribute the respective catalytic threonine nucleophiles, to the loops immediately N-terminal of each helix [65,72,77]. These conserved disulfides thus orient the catalytic nucleophiles in close proximity to the Zn^{2+} ion present at the respective catalytic centers. Recently, the structure of *Ec*EptC (PDB ID 6A83) was reported to have an overall similar fold and share structural similarity with the pEtN transferase family[66]. However, *Ec*EptC did not confer polymyxin resistance *in vivo* despite sharing the conserved active site architecture and catalytic Zn^{2+} binding discussed here[66]. Instead, Zhao *et al.*[66] proposed that the absence of Cys pairs forming disulfides in *Ec*EptC was responsible for the lack of observed activity, and showed that disruption of Cys pairs in MCR-1 or *Cj*EptC also abrogated polymyxin resistance when expressed in *E. coli* C43 to levels observed with expression of *Ec*EptC[66].

To understand the importance of disulfide bond presence in *Ec*BcsG*,* the structure was examined for the presence of the conserved disulfide bond. The equivalent amino acids to the conserved cysteine pairs from *Cj*EptC, *Nm*EptA and MCR-1 described above are observed as Ala274 and Arg284 in *Ec*BcsG$^{\Delta N}$. Instead, an unambiguous disulfide bond was observed in both the native and Se structures of *Ec*BcsG$^{\Delta N}$ between the cysteine pair Cys290-Cys306, spaced instead by 16 amino acids (Figure 19). This unusual disulfide bond in *Ec*BcsG appears to serve the same function as the conserved pEtN transferase disulfide bond, despite being structurally distinct. Instead of fixing the helix presenting the catalytic nucleophile to a loop directly N-terminal to it, the *Ec*BcsG disulfide instead links the loop immediately C-terminal of the equivalent helix with a neighbouring helical element packed against the central beta-sheet fold. This novel disulfide retains the conserved structural features related to their proposed function; orienting the catalytic *Ec*BcsG nucleophile, Ser[278], in proximity to the catalytic Zn^{2+} of the *Ec*BcsG$^{\Delta N}$ active site in an analogous fashion to *Cj*EptC, *Nm*EptA and MCR-1.

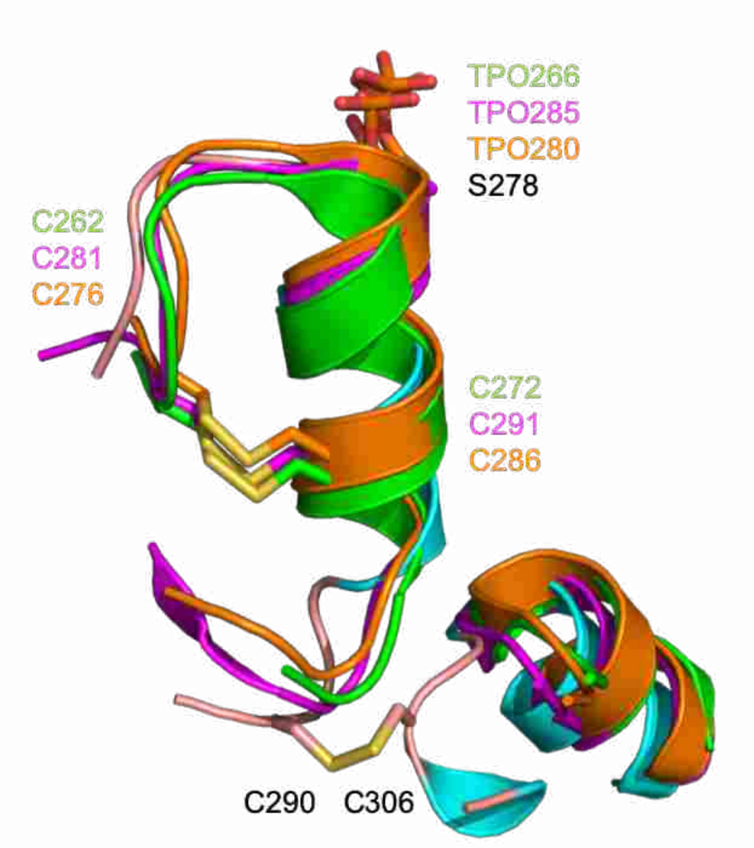


Figure 19*:* The pEtN transferase disulfide bond. A conserved, disulfide bonded cysteine pair in *Cj*EptC (green), MCR-1 (magenta), and *Nm*EptA (orange) fixes the helix presenting the catalytic threonine adjacent to the catalytic Zn^{2+} at the active site and is important for activity. In *Ec*BcsG$^{\Delta N}$ (cyan), a similar disulfide is observed, although not at the conserved location. Figure rendered in PyMOL.

Polymyxin Susceptibility Assay. Several members of the pEtN transferase family have been described to transfer pEtN to both lipid A and other secondary substrates, including the *C. jejuni* periplasmic N-linked glycan[101], flagellins[77], pilins[15,74], and likely other diverse periplasmic targets. To assess if *Ec*BcsG retains activity on lipid A *in vivo*, the susceptibility of *E. coli* AR3110 and *E. coli* AR3110 Δ*bcsG* to the antimicrobial peptides polymyxin E (colistin) and polymyxin B were assessed using a minimum inhibitory concentration (MIC) assay. The observed MIC of colistin was 2 μg/mL in both AR3110 and *E. coli* AR3110 Δ*bcsG*, in agreement with colistin-susceptible *E. coli* isolates that do not possess intrinsic or acquired colistin resistance (Figure 20) [103]. The polymyxin B MIC was observed to be 4 μg/mL for *E. coli* AR3110 and 8 μg/mL for *E. coli* AR3110 Δ*bcsG*. Furthermore, visible quantities of biofilm were present in the microtiter plate at the time of analysis, confirming the conditions of the assay were biofilm-permissive. Our findings did not support the hypothesis that *Ec*BcsG retains activity on lipid A *in vivo* to confer intrinsic polymyxin resistance, since *Ec*BcsG expression was not correlated with colistin MIC.

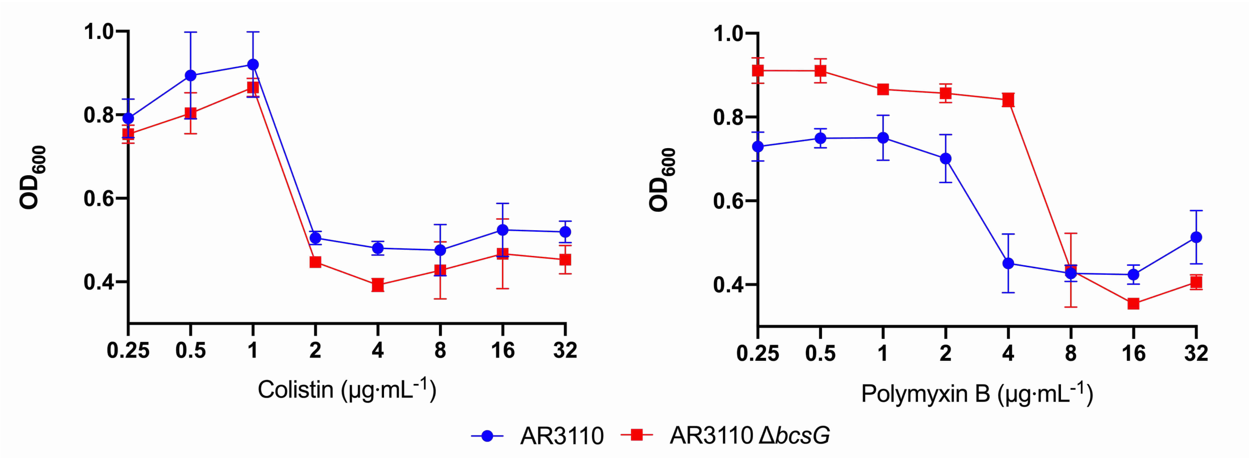


Figure 20: Colistin and polymyxin B minimum inhibitory concentration assay. Growth under biofilm-permissive conditions did not induce intrinsic polymyxin reistance in *E. coli* AR3110, nor did nonpolar deletion of *bcsG* in *E. coli* AR3110 induce polymyxin susceptibility.

5.4 Functional complementation of *Ec*BcsG

Pellicle assay. The colony morphotype of *E. coli* K-12 and *S. enterica* UMR1, and nonpolar *bcsG* chromosomal mutant derivatives thereof, have been described previously as a reporter for BcsG activity when grown on semisolid media containing Congo Red[104]. However, *bcsG* deletion mutants were observed to produce cellulose and, thus, retain a red, dry and rough morphotype using this assay. As a result, the colony morphotype of *E. coli bcsG* deletion strains producing unmodified cellulose was indistinguishable from that of the wild-type *E. coli* strains producing pEtN cellulose. To detect the phenotypic consequences of loss of *bcsG* function, a robust assay was developed to distinguish the functional contribution of *bcsG* and, thus, pEtN cellulose to the *E. coli* K-12 biofilm phenotype. Our results recapitulate those by Thongsomboon *et al.*[32] and demonstrate that the pEtN modification conferred by *Ec*BcsG activity on cellulose is essential for the *E. coli* AR3110 extracellular matrix assembly and architecture. The wild-type cellulose producing strain *E. coli* AR3110 produced a rigid pellicle that formed a cohesive, tissue-like layer when grown in salt-free LB broth that adhered to solid surfaces and was resistant to mechanical stress (Figure 21). The nonpolar chromosomal *bcsG* mutant in the *E. coli* AR3110 background (AR3110 Δ*bcsG*) produced a similar pellicle morphotype, although the biofilm did not display the same cohesive properties and was completely disrupted upon subjection to mechanical stress. Functional complementation of the AR3110 Δ*bcsG* cell line with the pET28a-encoded, full-length recombinant *bcsG* construct restored the rigid, self-assembling properties of the extracellular matrix, while complementation with the C-terminal catalytic domain or the vector alone failed to restore these properties and their biofilms were mechanically disrupted with ease, demonstrating that full-length *Ec*BcsG is both necessary and sufficient for pEtN cellulose and, thus, the rigid pellicle phenotype observed in this experiment.

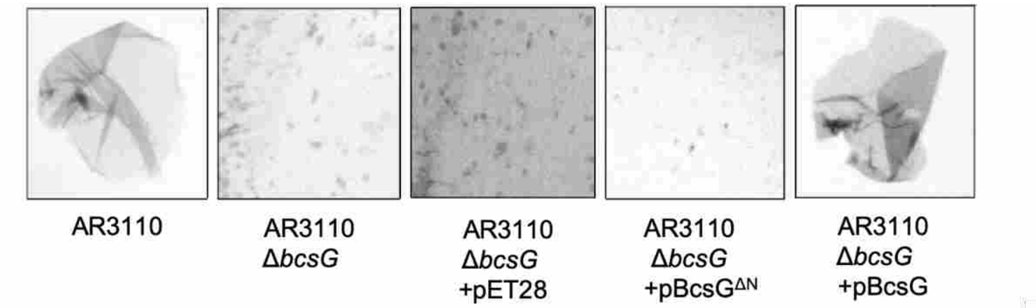


Figure 21: The pEtN cellulose pellicle assay. The rigid pellicle phenotype of *E. coli* AR3110 is lost upon deletion of *bcsG* and is rescued only by complementation with pBcsG, but not pBcsG$^{\Delta N}$ or the pET28a plasmid. Pellicles were stained in 0.01% (w/v) Congo Red following isolation from cultures and visualized with a BioRad VersaDoc 4000.

Crystal Violet Adherence Assay. A semi-quantitative analysis of the adherence potential of these phenotypes was performed using a crystal violet static biofilm assay, and the adherence potential appeared to follow the pellicle phenotype assayed above (Figure 21). Comparison of the AR3110 cell line to its Δ*bcsG* derivative suggested that loss of *bcsG* results in the loss of adherence potential by the biofilm, easily removed from the growth surface by gentle washing in phosphate buffered saline. The AR3110 cell line demonstrated > 4-fold higher crystal violet staining than the nonpolar *bcsG* mutant, restored by functional complementation with *bcsG,* but not *bcsG*$^{\Delta N}$ or the vector alone.

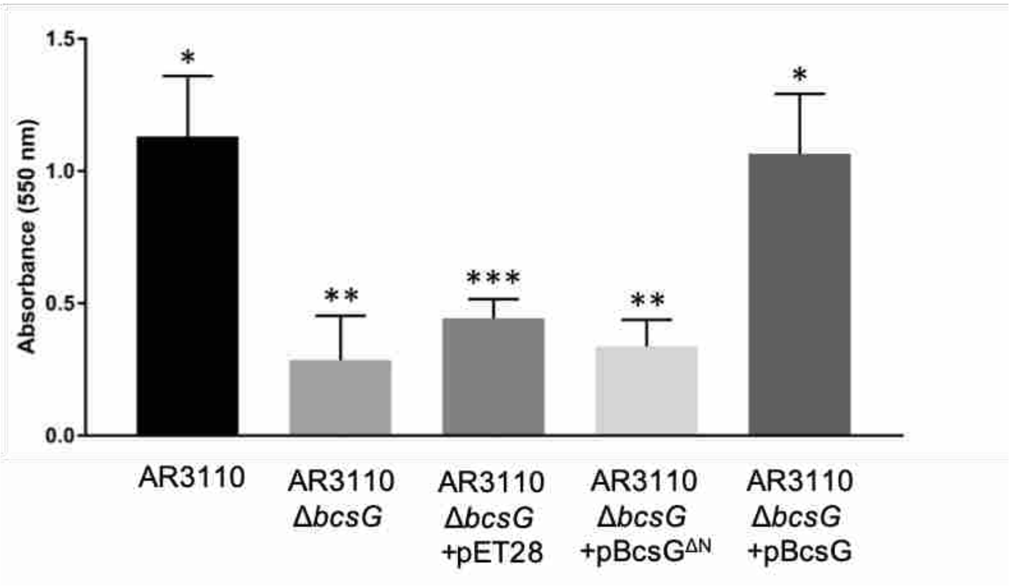


Figure 22: Crystal violet adherence of pellicle biofilms. The adherence potential of *E. coli* AR3110 is lost upon deletion of *bcsG* and was rescued only by complementation with pBcsG, but not pBcsG$^{\Delta N}$ or the pET28a plasmid. Asterisks denote significantly different groups ($p < 0.05$, Welch test)

Site-Directed Mutants. To map the active site of *Ec*BcsG and elucidate the function of key residues identified in its structure, site-directed mutants of the full-length recombinant *Ec*BcsG were generated and their sufficiency to rescue the fragile pellicle phenotype of the *E. coli* AR3110 Δ*bcsG* cell line were assessed using the pellicle (Figure 23) and static crystal violet assays (Figure 2r). Protein variants of *Ec*BcsG, supplemented to *E. coli* AR3110 Δ*bcsG in trans,* containing single alanine substitutions in amino acids Cys^{243}, Ser^{278}, Glu^{442}, His^{443}, His^{396}, and Tyr^{277} failed to functionally complement the chromosomal *bcsG* deletion, and resulted in the growth of fragile pellicle that was sensitive to mechanical disruption (Fig. 22). A Tyr^{277}Phe protein variant was also constructed by mutagenesis of pBcsG to assess if Tyr^{277} is involved in substrate accommodation of cellulose, noted as a possibility based on its position in the crystal structure. To our surprise, the Tyr^{277}Phe mutation also did not rescue the fragile pellicle phenotype of *E. coli* AR3110 Δ*bcsG*, suggesting that the tyrosine hydroxyl may instead be involved in forming polar contacts with substrate, rather than through ring-stacking interactions. Additionally, the Phe^{329}Ala protein variant rescued the fragile pellicle phenotype of *E. coli* AR3110 Δ*bcsG*, also suggesting that ring-stacking interactions by both aromatic residues lining the predicted active-site pocket do not function in substrate-binding or catalysis by *Ec*BcsG.

To further investigate the hypothesis that the observed disulfide bond is important for catalytic activity of *Ec*BcsG, a disulfide-deficient variant, *Ec*BcsG C290A C306A, was constructed and assessed for its activity *in vivo* using the functional complementation experiment. The

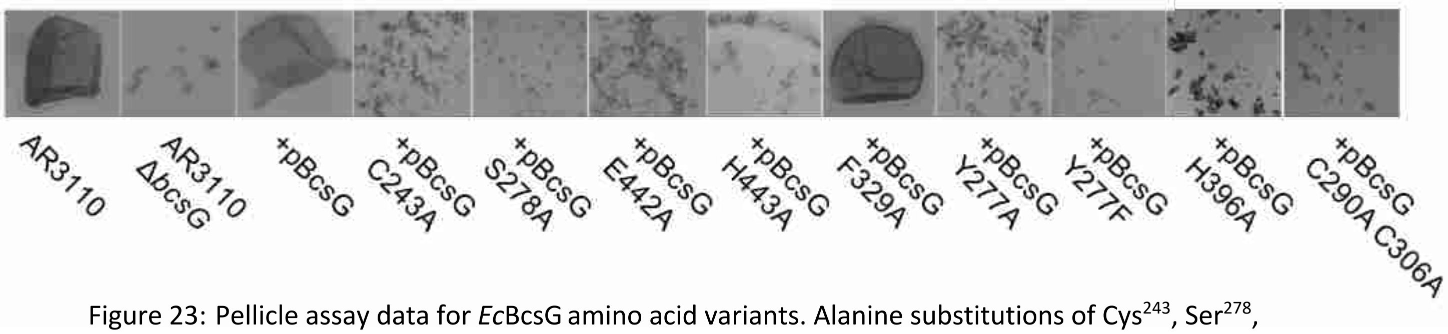


Figure 23: Pellicle assay data for *Ec*BcsG amino acid variants. Alanine substitutions of Cys^{243}, Ser^{278}, Glu^{442}, His^{443}, or Tyr^{277} fail to rescue the fragile pellicle phenotype, while an alanine substitution in Phe^{329} did rescue the fragile pellicle phenotype. A further Tyr^{277}Phe substitution failed to rescue the fragile pellicle phenotype.

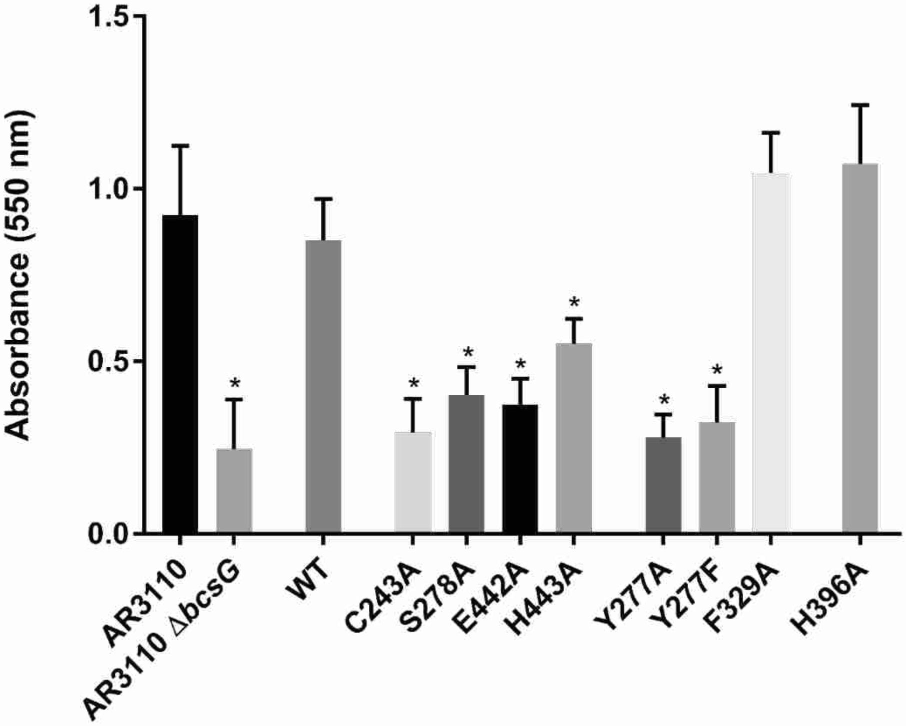


Figure 24: Crystal violet assay data for *Ec*BcsG amino acid variants. Alanine substitutions of Cys^{243}, Ser^{278}, Glu^{442}, His^{443}, or Tyr^{277} fail to rescue the adherence potential of *E. coli* AR3110 Δ*bcsG*, while an alanine substitution in Phe^{329} did rescue the adherence potential. A further Tyr^{277}Phe substitution failed to rescue the adherence potential.

results demonstrated that the *Ec*BcsG C290A C306A did not functionally complement *E. coli* AR3110 Δ*bcsG* and failed to rescue the fragile pellicle phenotype. This data suggests that the disulfide is indispensable for catalytic activity *in vivo*, and disruption of this disulfide results in a phenotype indistinguishable from *Ec*BcsG variants with disruptions in the essential catalytic architecture. This indispensability of the disulfide bond corroborates the role of this disulfide as a mimic of the conserved disulfide in the other pEtN transferases discussed here despite its alternate position in the structure.

Mutant Expression. A Western immunoblot of prepared *E. coli* AR3110 Δ*bcsG* cell lysates (30 µg total protein/well) supplemented with wild-type and variant pBcsG demonstrated invariable expression of *bcsG* and its variants from the pBcsG plasmid *in trans* at low, but detectable, levels indicating that differences in pellicle phenotype of these cell lines are attributable to differences in their amino acid composition and not their differential expression (Figure 25).

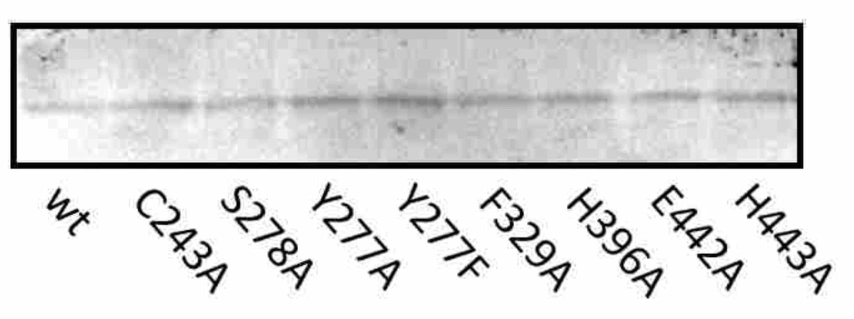


Figure 25: Western immunoblot against the recombinant His-tag of cell lysates from the functional complementation experiments. Uniform expression is observed across the wild-type and amino acid variants.

5.5 Functional bioassay

Synthesis of *p*-NPPE. Repeated screening and analysis of various commercially available substrate analogs failed to detect enzymatic product from *Ec*BcsG$^{\Delta N}$. Substrates tested include O-phosphorylethanolamine (Sigma), 1-myristoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine (14:0 LysoPE, Avanti), deoxyguanosine triphosphate, deoxycytosine triphosphate, deoxyadenosine triphosphate, deoxycytosine triphosphate, and various cellulose analogs, such as Avicel® PH-101, hydroxyethylcellulose, and carboxymethyl cellulose. A review of the literature revealed that at least one other member of the pEtN transferase family, MCR-1, was capable of using the synthetic compound *para*-nitrophenyl phosphoethanolamine (*p*-NPPE), and that its synthesis was possible from commercially available starting materials. We attempted the synthesis of *p*-NPPE as described by Dimitrijevic and colleagues[105] and successfully isolated the synthetic product, *p*-NPPE trifluoroacetate, as a fine white crystalline powder through the work of colleague (Lana Hiscock) in Dr. Maly's laboratory (Wilfrid Laurier University). The organic product was structurally analyzed by ^{1}H, ^{13}C, ^{19}F and ^{31}P NMR spectroscopy and IR spectroscopy (Appendix 1). Quality control of *p*-NPPE was prepared by dissolving the *p*-NPPE trifluoroacetate in 10 M NaOH and performing an absorbance spectral scan and the λ_{max} was found to be 414 nm (Appendix 1). This value was used for all further functional bioassay experiments utilizing *p*-NPPE. A standard curve of base-hydrolyzed *p*-NPPE was performed and compared to an analytical reference standard of paranitrophenolate (Sigma). No significant difference was detected between the reference standard and the *p*-NPPE we prepared (Appendix 1), and so it was deemed suitable for downstream analysis with *Ec*BcsG$^{\Delta N}$.

Activity of *Ec*BcsG$^{\Delta N}$. To test *Ec*BcsG$^{\Delta N}$ activity, the enzyme was overloaded with substrate at levels described by Wanty and colleagues for MCR-1 (*i.e.* 5 mM)[65]. Incubation of *Ec*BcsG$^{\Delta N}$ with the chromogenic substrate *p*-NPPE resulted in the release of the *para*-nitrophenolate ion giving a strong absorbance at 414 nm at a rate significantly higher than an enzyme-free control (t-test, $p < 0.05$, Figure 26). To test if *Ec*BcsG$^{\Delta N}$ was capable of transferase activity *in vitro*, an identical reaction to the one described above was performed with the addition of 1 mM of either D-glucose (Sigma), β-D-cellobiose, β-D-cellotriose, β-D-cellotetraose, β-D-cellopentaose, or β-D-cellohexaose (Megazyme). Under conditions containing 1 mM of β-D-cellotetraose, β-D-cellopentaose, or β-D-cellohexaose, a significant increase in the rate of formation of the *para*-nitrophenolate ion was observed in a length-dependent fashion (Figure 27). To then test the substrate specificity of *Ec*BcsG$^{\Delta N}$ for cellulose, further transferase assays were performed with the addition 1 mM β-D-chitobiose, β-D-chitotriose, β-D-chitotetraose, β-D-chitopentaose, or β-D-chitohexaose. No significant increase in the rate of formation of the *para*-nitrophenolate ion was observed in the presence of any of these chitooligosaccharides as compared to an acceptor-free control reaction (Figure 27).

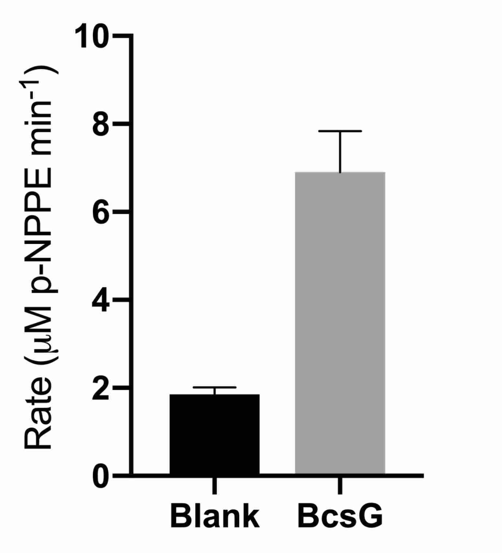


Figure 27: *Ec*BcsG$^{\Delta N}$ liberates *para*-nitrophenolate from *p*-NPPE. A significantly increase in the rate of *p*-NPPE cleavage and *para*-nitrophenolate formation was observed in the presence of 50 µM *Ec*BcsG$^{\Delta N}$.

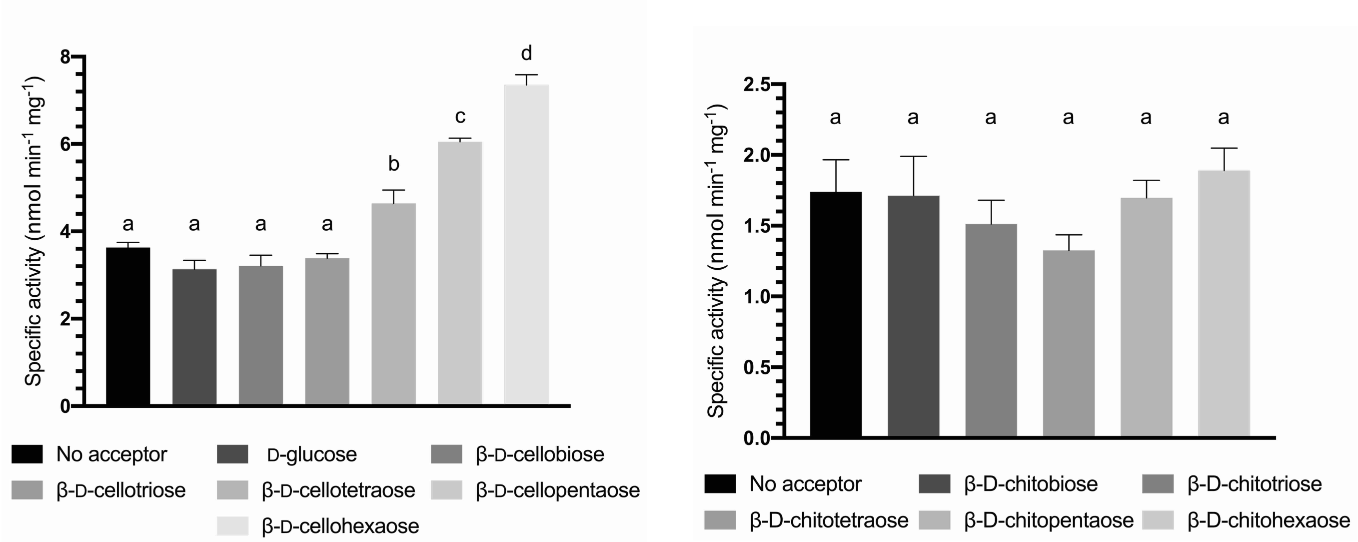


Figure 26: Phosphoethanolamine transferase assay. The rate of *para*-nitrophenolate formation in the presence of *Ec*BcsG$^{\Delta N}$ was increased in the presence of celloligosaccharides 4 residues or longer, and in a length-dependent fashion (left). No such increase was observed with chitooligosaccharides (right). Letters denote statistically significant groupings (ANOVA, $p > 0.0001$).

In an effort to detect the enzymatic product, both the reaction condition containing cellopentaose and an enzyme-free control were repeated for a 24 h duration. The resulting reaction mixtures were spotted on a silica plate for analysis by thin-layer chromatography. Following separation, staining of the silica matrix revealed a shift in the retention of the carbohydrate component of the reaction mixture following treatment with *Ec*BcsG$^{\Delta N}$ (Fig. 28).

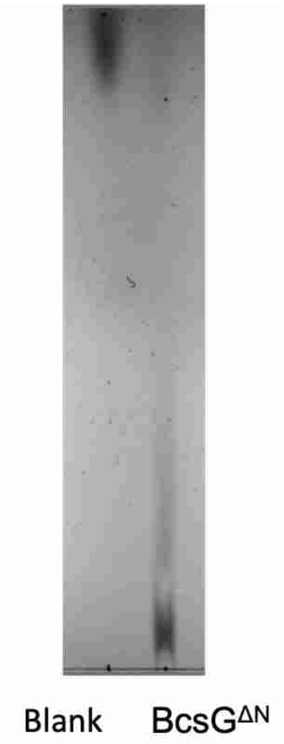


Figure 28: Thin-layer chromatography of the *Ec*BcsG$^{\Delta N}$ product. Cellopentaose (1 mM) in the presence of *p*-NPPE (5 mM) has a high retention time in the nonpolar mobile phase (left) that is reduced by treatment with 50 µM *Ec*BcsG$^{\Delta N}$ (right). TLC plate was stained in a 3% (w/v) α-napthol dissolved in 25:3:2 methanol:water:sulfuric acid solution.

To confirm the increase in the rate of turnover of *p*-NPPE was the result of transferase activity, the reaction products from the sample containing cellopentaose were then separated and analyzed by LC-MS to confirm the transfer of the pEtN group. Along with unreacted *p*-NPPE ($[C_8H_{12}NO_6P]+H^+$, m/z 263.1614) and cellopentaose ($[C_{30}H_{52}O_{26}]+H^+$, m/z 829.2814), there was observed a new species with a novel m/z corresponding to the expected molecular weight of the monosubstituted pentamer ($[C_{32}H_{58}NO_{29}P]+H^+$, m/z 952.2892: Fig. 29). Such a species was absent in an enzyme-free control, which instead contained only the m/z species corresponding to the starting materials.

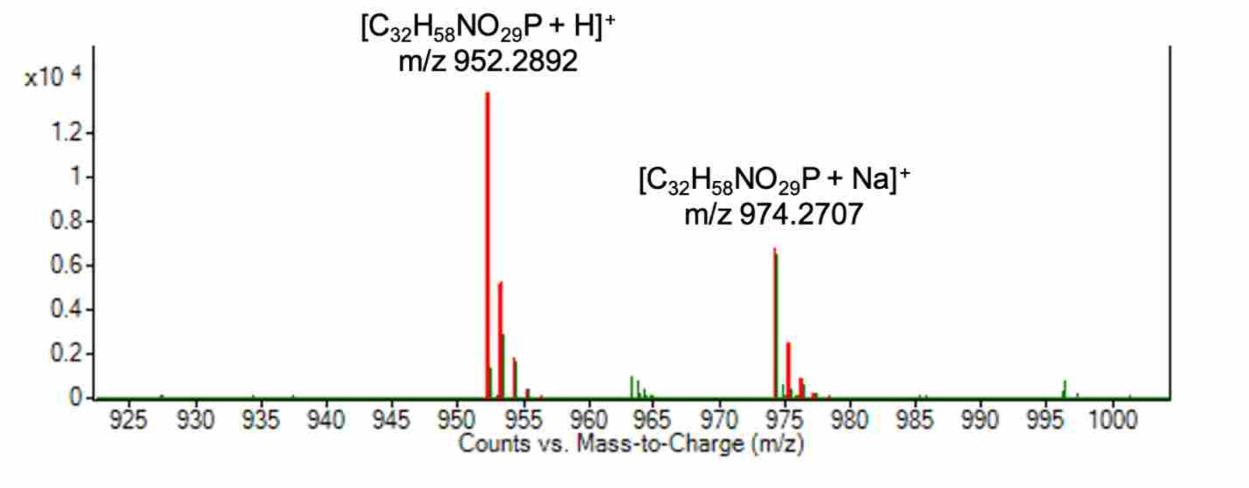


Figure 29: The *Ec*BcsG$^{\Delta N}$ enzymatic product mass spectra. Two nominal m/z species corresponding to mono-pEtN substituted cellopentaose were observed in the presence of *Ec*BcsG$^{\Delta N}$.

Rate kinetics. To measure the apparent kinetic parameters of *Ec*BcsG$^{\Delta N}$ using the *in vitro* assay described here, the apparent initial enzymatic rate was measured at varying concentrations of *p*-NPPE using a fixed concentration of 1 mM cellopentaose, and vice versa. Variation of *p*-NPPE from 0-10 mM and curve fitting (Fig. 30) of the subsequent initial rates as a function of substrate concentration resulted in an apparent K_m of 1.171 ± 0.267 mM, maximal rate (V_{max}) of 5.728 ± 0.345 nmol min^{-1} mg^{-1}, and k_{cat} of $1.91 \times 10^{-6} \pm 1.15 \times 10^{-7}$ s^{-1}. The derived catalytic efficiency k_{cat}/K_m was $1.63 \times 10^{-3} \pm 3.84 \times 10^{-4}$ M^{-1} s^{-1}. The experiment was repeated at a fixed concentration of 5 mM *p*-NPPE (approximately 3x the K_m) and cellopentaose concentration varying from 0 to 2.5 mM cellopentaose in 0.25 mM increments. Curve fitting of the measured initial rates as a function of substrate concentration resulted in an apparent K_m of 1.435 ± 0.632

mM, V_{max} of 12.27 ± 2.67 nmol min^{-1} mg^{-1}, and k_{cat} of $4.09 \times 10^{-6} \pm 8.86 \times 10^{-7}$. The derived catalytic efficiency k_{cat}/K_m was then $2.4 \times 10^{-3} \pm 7.00 \times 10^{-4}$ M^{-1} s^{-1}. However, this model did not account for considerable rate of esterase activity observed of *Ec*BcsG$^{\Delta N}$ (*i.e.* transfer to water rather than cellopentaose). To better model the kinetic behaviour of *Ec*BcsG$^{\Delta N}$ under varied acceptor cosubstrate concentrations, the initial rates were corrected to discount the contribution of esterase activity against *p*-NPPE and the resulting initial rates were fit to a sigmoidal model (Figure 30). The model parameters were in good agreement with the *p*-NPPE kinetic values, with a calculated $K_{1/2}$ of 1.701 ± 0.332, V_{max} of 8.219 ± 1.435 nmol min^{-1} mg^{-1}, and Hill slope of 1.877 ± 0.371. The kinetic parameters are summarized below in Table 5.

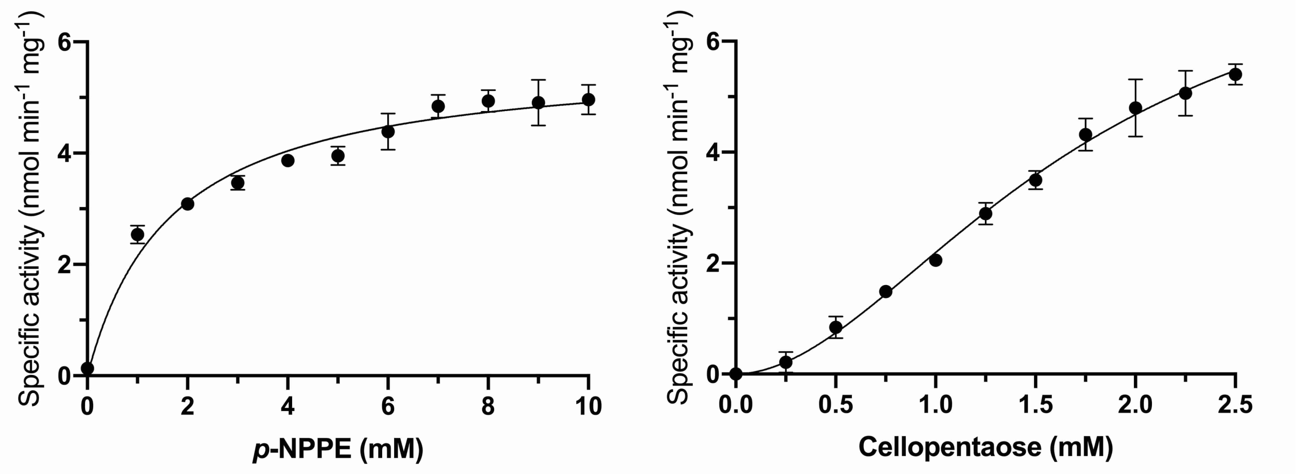


Figure 30: Curve fitting of *Ec*BcsG$^{\Delta N}$ kinetic parameters. The specific *Ec*BcsG$^{\Delta N}$ activity with varying *p*-NPPE (left) or cellopentaose (right) at a fixed concentration of cosubstrate were fit to Michaelis-Menten and sigmoidal models, respectively.

Table 5: Measured kinetic parameters of *Ec*BcsG$^{\Delta N}$

	$K_m/K_{1/2}$ (mM)	V_{max} (nmol min^{-1} mg^{-1})	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})
p-NPPE	1.171 ± 0.267	5.728 ± 0.345	$1.91 \times 10^{-6} \pm 1.15 \times 10^{-7}$	$1.63 \times 10^{-3} \pm 3.84 \times 10^{-4}$
Cellopentaose	1.701 ± 0.332	8.219 ± 1.435	*$4.09 \times 10^{-6} \pm 8.86 \times 10^{-7}$	*$2.4 \times 10^{-3} \pm 7.00 \times 10^{-4}$

* Based on ambiguous fitting to Michaelis-Menten model

Amino Acid Replacements. To assess the catalytic consequences of amino acid substitutions within the *Ec*BcsG active site, a library of amino acid *Ec*BcsG$^{\Delta N}$ variants was constructed with substitutions of active site residues we observed to be essential for catalytic activity *in vivo*. All of the constructs used in this study, C243A, Y277F, S278A, H396A, E442A, and H443A were generated using site-directed mutagenesis of pBcsG$^{\Delta N}$. Each of these recombinant proteins expressed in *E. coli* BL21 transformants with observed yields similar to that of wild-type *Ec*BcsG$^{\Delta N}$. These variant proteins were purified to apparent homogeneity using the protocol described previously, with no modifications (Fig. 31). Care was taken to use new chromatography media for the purification of each respective enzyme to prevent enzyme carry-over and cross-contamination of assay data.

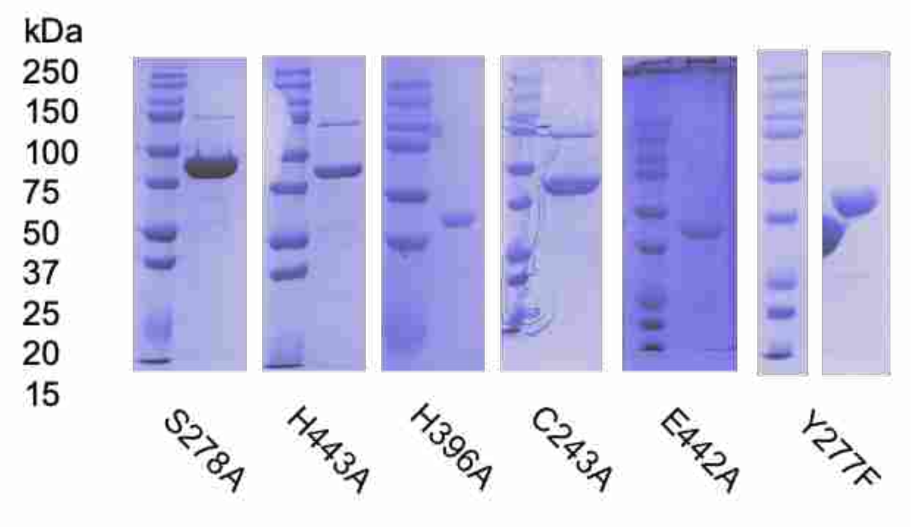


Figure 31: SDS-PAGE of purified *Ec*BcsG$^{\Delta N}$ amino acid variants. All amino acid variants tested expressed and purified to near homogeneity in a fashion similar to wild-type *Ec*BcsG$^{\Delta N}$. Each gel shows a major band representing recombinant and amino-acid substituted *Ec*BcsG$^{\Delta N}$ at 45.5 kDa.

As expected, replacement of the catalytic nucleophile Ser^{278} resulted in a loss of any detectable activity. The observed rates of both esterase and transferase activity of the Ser^{278} Ala replacement were not different from an enzyme free control (*e.g.* 1.96 ± 0.05 nmol min^{-1} mg^{-1}, 1.93 ± 0.13 nmol min^{-1} mg^{-1}, and 1.94 ± 0.06 for the control, esterase, and transferase activity, respectively; Fig. 32). The next greatest observed loss of activity was measured for the replacement of Glu[442] with Ala, also with an esterase activity indistinguishable from an enzyme-free control (*e.g.* 1.96 ± 0.07 nmol min^{-1} mg^{-1}; Figure 31). Only 2.3% residual transferase activity for the Glu[442] variant could be detected in the presence of β-D-cellopentaose.

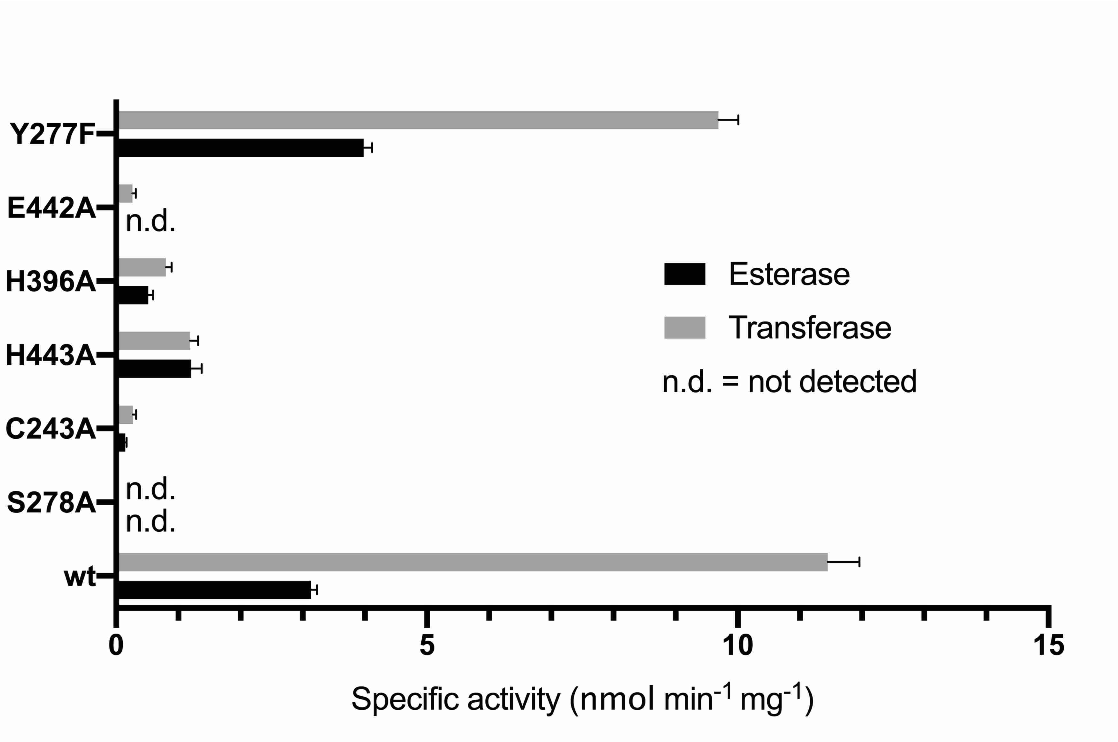


Figure 32: Measured specific activities of *Ec*BcsG$^{\Delta N}$ amino acid variants both in the absence of cellopentaose (esterase) and in the presence (transferase). wt = wild-type *Ec*BcsG$^{\Delta N}$.

Similarly to Glu[442], replacement of Cys[243] with Ala resulted in residual specific activities of 4.7% (esterase) and 2.4% (transferase), consistent with the essential Zn^{2+}-binding roles of these residues (Fig. 32). The His[443] substitution also demonstrated reduced specific activities, although to a lesser degree, representing 38.5% residual esterase and 10.4% residual transferase activity. Furthermore, replacement of His[396] with Ala produced an enzyme with a residual esterase activity of 16.6% and a residual transferase activity of 7.0%, consistent with the observation that this histidine is highly conserved and catalytically important in other pEtN transferases.

Replacement of Tyr^{277} with Phe produced an enzyme variant with greater esterase activity and 84% residual transferase activity compared to wild-type *Ec*$BcsG^{\Delta N}$ (Fig. 32). This observation did not support our earlier conclusion that Tyr^{277} may function in cellulose accommodation in proximity to the Zn^{2+} ion at the active center. The thesis discussed previously that an equivalent tyrosine residue is not observed in available structures of other pEtN transferases, nor was a tyrosine residue observed to participate in contact with the lipid-binding site in the structure of the full-length pEtN transferase EptA from *Neisseria meningitidis* (*Nm*$EptA^{FL}$; PDD ID 5FGN). Instead, the equivalent residue to Tyr^{277} in *Nm*$EptA^{FL}$, Ser^{279}, appears to shape the lipid-binding pocket at the interface of the catalytic and membrane domains. Consequently, we rationalized that this disparity in our data may indicate that the role of the Tyr^{277} hydroxyl may instead be in stabilizing the local folding and interaction of the two *Ec*BcsG domains, and thus, allowing for *Ec*BcsG to sample productive conformations during the catalytic cycle. This hypothesis would account for the importance of Tyr^{277} to *Ec*BcsG activity *in vivo* but not to *Ec*$BcsG^{\Delta N}$ *in vitro*. The remainder of our variant assay data was in good agreement with the proposed active site, demonstrating strict dependence on the catalytic nucleophile/base pair Ser^{278}/His^{396} and coordination of the Zn^{2+} ion at the active center by Cys^{243}, Glu^{442} and, to a lesser extent, His^{443}.

Molecular Determinants. To assess if the loss of activity observed for substitutions of the Zn^{2+}-binding residues was in fact due to the loss of Zn^{2+} in the *Ec*$BcsG^{\Delta N}$ active site or was due to structural changes in the enzyme, the metal chelating agent EDTA was supplied to *Ec*$BcsG^{\Delta N}$ prior to assay. Treatment with 1 mM EDTA was observed to reduce the specific esterase activity of *Ec*$BcsG^{\Delta N}$ to 10.6% of the untreated activity (Fig. 33). This significant loss of esterase activity was in agreement with the large reduction in esterase activity observed for the replacement of Cys^{243}, Glu^{442} and His^{443}. Surprisingly, however, the loss of transferase activity observed for EDTA treatment was markedly less than that observed for single amino acid substitutions of the Zn^{2+}-binding residues*, e.g.* 43% residual activity of wild-type for EDTA treated, compared to 2.4%, 2.3% and 10.4% residual activity for replacements of Cys^{243}, Glu^{442}, and His^{443}, respectively. This data suggests that in addition to Zn^{2+} binding, Cys^{243}, Glu^{442} and His^{443} may also be catalytically important residues, particularly during the latter step of catalysis when pEtN is transferred to cellulose.

Subsequently, the reducing agent DTT was introduced to *Ec*BcsG$^{\Delta N}$ to assess the importance of disulfide bond formation to enzymatic activity. An observed decrease in

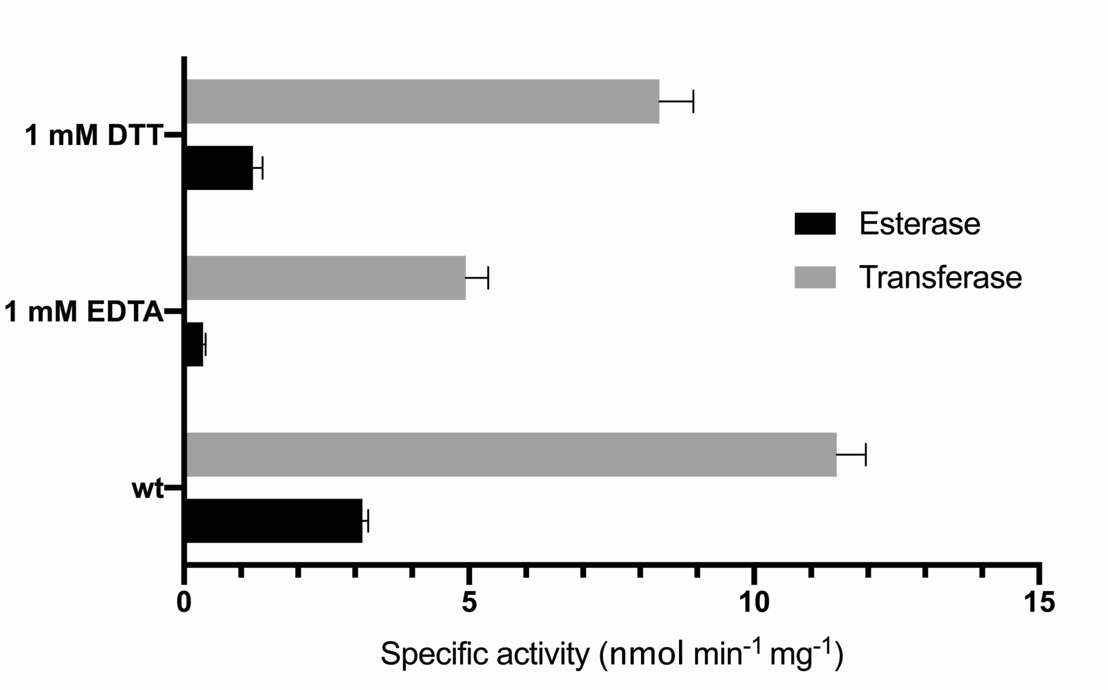


Figure 33: Measured specific activities of *Ec*BcsG$^{\Delta N}$ treated with the metal chelating agent EDTA and the reducing agent DTT, both in the absence of cellopentaose (esterase) and in the presence (transferase). wt = wild-type *Ec*BcsG$^{\Delta N}$.

enzymatic activity, representing 31.9% of residual esterase activity was observed (Fig. 33). In contrast, the measured residual transferase activity was 72.8% of wild-type. Large reductions in the enzyme rate would be expected to agree with the functional complementation data, and result in an inadequate level of pEtN, and thus, the fragile pellicle phenotype. However, the limited reduction of the enzymatic rate in the presence of the reducing agent DTT is in apparent disagreement with the fragile pellicle phenotype of the AR3110 Δ*bcsG* pBcsG$^{\Delta N}$ C290A C306A, which displayed a pellicle phenotype not different from AR3110 Δ*bcsG*. This data is instead rationalized by the highly specific requirements of exopolysaccharide modification for biofilm formation. For example, the biofilm polysaccharide poly-β-(1,6)-*N*-acetyl-D-glucosamine, sometimes abbreviated PNAG, requires only partial de-acetylation by the enzyme PgaB or its orthologues for successful biofilm formation in various bacteria, including *E. coli*, *S. aureus, and Bortadella bronchiseptica*[21,34,106]. PNAG found in biofilms has been observed to have approximately 15-20% of the *N*-acetyl-D-glucosamine saccharide units to be deacetylated, with either more or less extensive deacetylation causing disruption in biofilm formation, probably due to the loss of a distributed cationic charge on the polymer[33–35]. Although a loss in *Ec*BcsG$^{\Delta N}$ activity of < 30% was measured in the presence of DTT, which could be considered trivial for the biological function of some enzymes, the extent of pEtN modification to 50% of D-glucose saccharide units has been observed of pEtN cellulose biofilms isolated from *E. coli* or *S. enterica*[32]. As a result, even subtle reductions in the enzymatic rate, such as those observed here, could still plausibly alter the degree of pEtN substitution so that is it not matched to the rate of cellulose synthesis, disrupting the otherwise uniform charge and chemistry required for biofilm formation.

5.6 Evidence for the *Ec*BcsG catalytic mechanism

Identification of the covalent pEtN enzyme intermediate. A comprehensive catalytic mechanism for any pEtN transferase family member has yet to be experimentally elucidated[72]. Characterized enzymes belonging to the alkaline phosphatase superfamily tend to vary greatly in the nature and number of metal ions in the active site, based on available structures of these enzymes[102]. While the number of available structures in the pEtN transferase family remain limited at present, it appears that at least one catalytic zinc ion is strictly conserved and is indispensable for activity[72,77]. Several structures of MCR-1 are available, with one to four zinc ions observed in the active site, although Hinchcliffe *et al.* demonstrated that the single strictly conserved zinc ion is likely sufficient for at least part of the catalytic cycle[72]. Furthermore, pEtN transferases are proposed to proceed by a ping-pong bisubstrate-biproduct (bi-bi) reaction mechanism (Fig. 34), whereby the first step is pEtN transfer from a phospholipid to the enzyme catalytic nucleophile, and the second step is nucleophilic attack of the covalent pEtN intermediate by the acceptor substrate and turnover of the enzyme, although evidence for the second step and its structural feasibility remain to be seen[72]. The structure of *Ec*BcsG$^{\Delta N}$ presented here demonstrates the presence of the single conserved catalytic zinc ion, and an equivalent serine in place of the family consensus catalytic threonine, satisfying all structural

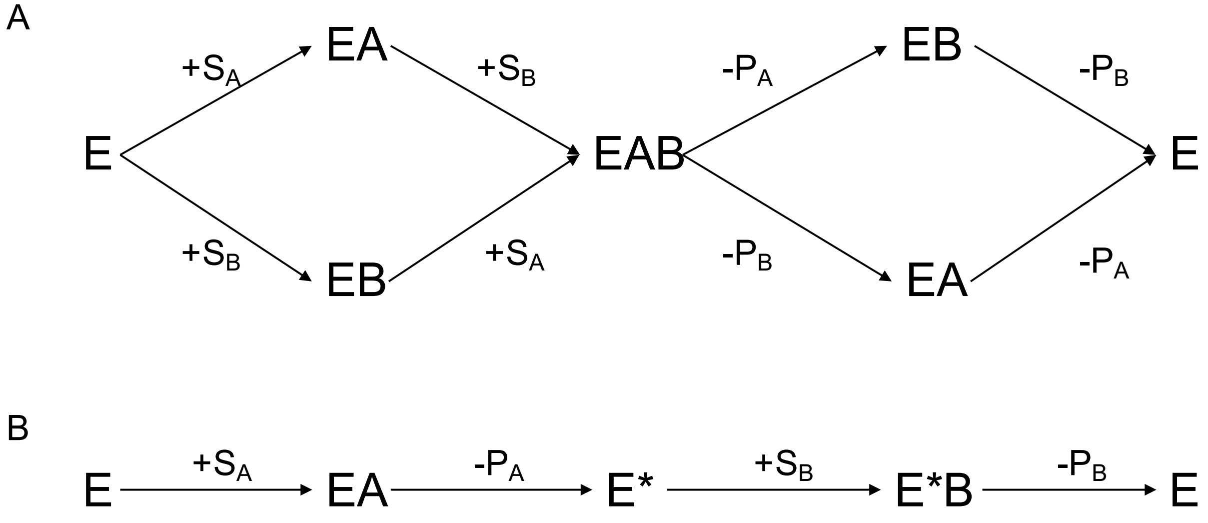


Figure 34: Bisubstrate-biproduct reaction pathways. (A) The sequential bi-bi mechanism, in which both substrates bind before the release of product. (B) The ping-pong bi-bi pathway, in which the release of the first product and the formation of a covalent enzyme intermediate must occur before binding of the second substrate.

criteria presented thus far for catalysis through the probable pEtN transferase ping pong bi-bi mechanism. To further investigate the hypothesis that *Ec*BcsG proceeds through such a ping pong bi-bi reaction pathway, intact *Ec*BcsG$^{\Delta N}$ was analyzed by LC- Q-TOF MS both before and after substrate exposure to identify the covalent enzyme intermediate, if such an intermediate exists (Fig. 35). The intact *Ec*BcsG$^{\Delta N}$ mass spectra was composed of two predominant m/z species corresponding to molecular ions of mass 45238.57 Da and 45318.54 Da. The 45238.57 Da species is not in agreement with the *Ec*BcsG$^{\Delta N}$ sequence, with a theoretical molecular mass of 45344.98 Da. Instead, the hypothetical molecular mass of a post-translationally modified and mature *Ec*BcsG$^{\Delta N}$ polypeptide, which accounts for the presence of a disulfide bond, formylation of the N-terminal methionine residue, and loss of the C-terminal glutamic acid residue, is in good agreement with the observed species, having a predicted molecular mass of 45242.84 Da. The second species, with a molecular mass of 45318.54 Da, corresponds to a mass shift of

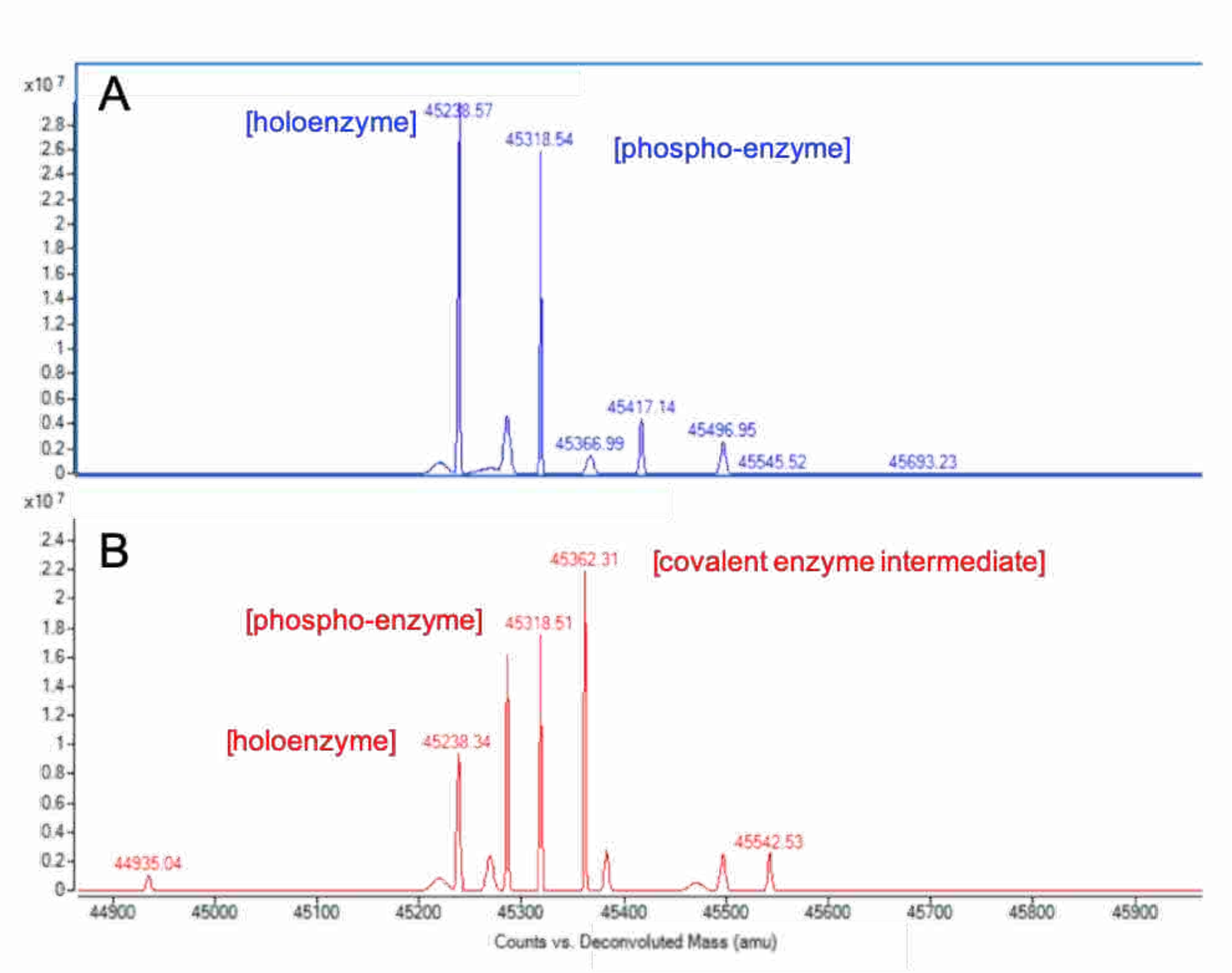


Figure 35: LC- Q-ToF MS of intact *Ec*BcsG$^{\Delta N}$. The mass spectra of intact *Ec*BcsG$^{\Delta N}$ (A) and of *Ec*BcsG$^{\Delta N}$ following substrate exposure (B) demonstrate that BcsG$^{\Delta N}$ proceeds through a covalent pEtN-enzyme intermediate, and that recombinant *Ec*BcsG$^{\Delta N}$ exists in solution as a mixed population of holo-*Ec*BcsG$^{\Delta N}$ and as a phosphorylated form.

+79.97 Da that is indicative of protein phosphorylation (Fig. 35; panel A). The observation of a phosphorylated *Ec*BcsG$^{\Delta N}$ species in solution is consistent with other pEtN transferase family members, for which a phosphorylated species was observed *in crystallo*[65,72,77]. However, the presence of a phosphorylated *Ec*BcsG$^{\Delta N}$, particularly with a comparatively high relative abundance, was surprising as phosphorylated cellooligosaccharides were never detected as enzymatic products by thin-layer chromatography or mass spectrometry, and suggested the phosphorylated species is not catalytically relevant. In addition, *Ec*BcsG$^{\Delta N}$ was purified in the absence of phosphate or inorganic phosphorus, suggesting that the phosphorylation event must have occurred in the *E. coli* overexpression host, and may be an artefact of cytoplasmic expression of the recombinant catalytic domain.

In an attempt to identify the site of phosphorylation, the structure of *Ec*BcsG$^{\Delta N}$ was refined *de novo* using an unbiased calculated electron density map obtained from initial experimental phasing, with a spherical omit map of 10 Å centered around the catalytic zinc ion to remove model bias that may have been introduced during initial refinement, and before the detection of a phosphorylated *Ec*BcsG$^{\Delta N}$ polypeptide in solution state. Surprisingly, no electron density was calculated within 10 Å of the catalytic zinc that could be attributed to a phosphorylated residue. The absence of the phosphorylated *Ec*BcsG$^{\Delta N}$ species in the structure presented here is reconciled by the spatial geometry of crystal packing, whereby the face of *Ec*BcsG$^{\Delta N}$ presenting the active site makes crystal contacts with a molecule in a neighbouring asymmetric unit in both the $P\ 2_12_12_1$ crystal form of *Ec*BcsG$^{\Delta N}$-Se and the $P\ 2_1$ crystal form of *Ec*BcsG$^{\Delta N}$. Thus, phosphorylation of a residue in the active site would plausibly alter the surface electrostatic potential of *Ec*BcsG$^{\Delta N}$, and may impede productive crystal packing, rationalizing the incorporation of only holo- *Ec*BcsG$^{\Delta N}$ into our structural model, despite the high relative abundance of the phosphorylated form observed from intact protein LC- Q-TOF MS.

To then identify the presence of a covalent enzyme intermediate, if such a species exists, LC- Q-TOF mass spectra of *Ec*BcsG$^{\Delta N}$ were collected immediately following exposure to the substrate analogue *p*-NPPE in large (300-fold) stoichiometric excess. In addition to the holo-*Ec*BcsG$^{\Delta N}$ (45238.57 Da) and phospho-*Ec*BcsG$^{\Delta N}$ (45318.54 Da) species, the predominant molecular ion had an observed molecular mass of 45362.31 Da, a mass shift of +123.74 Da corresponding to the covalent addition of pEtN and representing the catalytically relevant covalent enzyme intermediate (Figure 35; panel B). Moreover, the detection of the covalent enzyme intermediate was paralleled by a reduction in the relative intensity of the species corresponding to the mass of holo-*Ec*BcsG$^{\Delta N}$, with no apparent change in the relative abundance of the phosphorylated species. This observation was consistent with the hypothesis that the phosphorylated species may not be catalytically relevant and provides evidence that *Ec*BcsG$^{\Delta N}$ proceeds through a covalent pEtN-enzyme intermediate, a requirement for the first step of catalysis through a ping pong bi-bi reaction mechanism.

Determination of the *Ec*BcsG$^{\Delta N}$ reaction order. Catalysis through a ping pong bi-bi reaction mechanism mandates the interaction of cosubstrates to exclusive forms of the enzyme, that is, the cosubstrate for the first step of the reaction mechanism, which generates the covalent enzyme intermediate from the free enzyme, will not subsequently interact with the covalent enzyme intermediate required for the second step regenerating the free enzyme, and vice versa[107]. As a consequence, the rate of formation of the covalent enzyme intermediate will still affect the net enzymatic rate, even when the concentration of the cosubstrate that regenerates the free enzyme is saturating[107]. Conversely, because the rate of such a mechanism depends on the concentration of both cosubstrates, increasing the concentration of one cosubstrate at a fixed concentration of the other will result in an increase in the initial rate until the concentration of the fixed cosubstrate becomes constraining[107]. These phenomena were first described by Cleland, often termed Cleland's rules[107]. The mechanism can thus be interpreted from a Lineweaver-Burk reciprocal plot, where for enzymes following a ping-pong bi-bi mechanism, plotting the double reciprocal of initial rate against varying concentration of one cosubstrate at a fixed concentration of the other will result in linear relationships with unique intercepts but equal slopes, demonstrating the interaction of the two cosubstrates with exclusive forms of the enzyme as is required for such a mechanism[107]. Alternately, a sequential bi-bi reaction pathway, in which the cosubstrates interact with the same form of the enzyme at some point in the catalytic cycle, would produce a double reciprocal plot of linear equations with equal intercepts but unique slopes[107]. This is observed to be true in enzymes which interact with both cosubstrates simultaneously, as the variation of one cosubstrate would fail to influence the initial rate of the enzyme when the other cosusbtrate is saturating at any non-negligible concentration of the other[107].

To investigate the reaction order, and thus test the hypothesis that *Ec*BcsG$^{\Delta N}$ follows a ping pong bi-bi reaction pathway, the initial velocity was measured at cosubstrate concentrations bracketing the K_m for each (0.7, 1.4, 2.1 and 2.8 mM for *p*-NPPE and 0.5, 1, 2 and 2.5 mM for cellopentaose). Double reciprocal plots of initial velocity against (i) *p*-NPPE concentration at a fixed cellopentaose concentration, and (ii) cellopentaose concentration at a fixed *p*-NPPE concentration, were prepared and fit to linear models (Figure 36).

The resulting linear model data was analyzed with a regular one-way ANOVA, determining that the slopes were not significantly different from each other (i: $F(3,12) = 1.275$, $p = 0.3273$; ii: $F(3,12) = 1.086$, $p = 0.3661$). A Tukey's test was conducted postanalysis for pairwise comparison of each slope, and none were found to be significantly different from any other ($p > 0.3$ for all cases). A subsequent ANCOVA was performed to test the null hypothesis that the intercepts are equal, and the intercepts were found to be significantly different from each other ($F(3,11) = 188.1$, $p < 0.0001$).

The observation of double reciprocal initial velocity versus cosubstrate concentration plots bearing parallel slopes with unique intercepts suggests that each cosubstrate used in the *Ec*BcsG$^{\Delta N}$ activity assay interacts with a different form of the enzyme. Hence, according to Cleland's rules, such a finding supports the hypothesis that *Ec*BcsG$^{\Delta N}$ must proceed through a ping-pong bi-bi reaction mechanism.

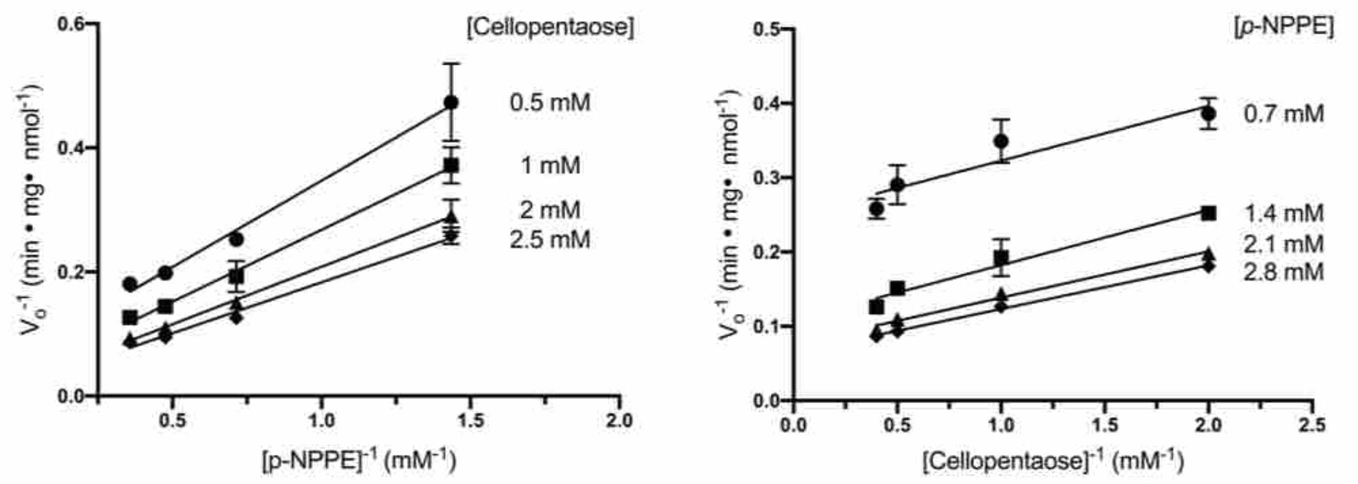


Figure 36. The reaction order of *Ec*BcsG$^{\Delta N}$ by kinetic analysis. Lineweaver−Burk plots of initial rates of reaction of 50 µM *Ec*BcsG$^{\Delta N}$ in 50 mM NaCl, 50 mM HEPES pH 7.5, using varied concentrations of *p*-NPPE (left) as the donor or cellopentaose (right) as the acceptor. Each data point represents 3 replicates.

Identification of Ser^{278} as the catalytic nucleophile. A conserved catalytic threonine is proposed to serve the role of the catalytic nucleophile in the lipid A pEtN transferases, evidenced by the observation of this conserved residue as phosphothreonine in the crystal structures of several lipid A pEtN transferases[65,72,77], and for its indispensability for catalytic activity *in vivo*[72,77]. However, the structure of *Ec*BcsG$^{\Delta N}$ presented here instead possessed the equivalent Ser^{278} in place of the conserved threonine, and was never observed as phosphoserine *in crystallo*, calling into question the identity of the *Ec*BcsG$^{\Delta N}$ catalytic nucleophile. To unambiguously identify the residue serving as the catalytic nucleophile, *Ec*BcsG$^{\Delta N}$ was briefly exposed to the substrate analogue *p*-NPPE (at 15 mM) to facilitate the formation of the covalent enzyme intermediate and was subsequently digested in sequencing-grade trypsin. The resulting tryptic peptide library was separated and analyzed by LC- Q-TOF MS to identify covalently pEtN modified peptides. Approximately 91% coverage was achieved in the tryptic peptide library in complete agreement with the hypothetical *Ec*BcsG$^{\Delta N}$ sequence. Peptides containing the predicted catalytic nucleophile Ser^{278} were observed, along with similar peptides containing a mass shift of +79.97 Da with high confidence ($p < 0.0003$), consistent with the phosphorylated *Ec*BcsG$^{\Delta N}$ species observed in the intact *Ec*BcsG$^{\Delta N}$ mass spectra. However, these peptides also contained other possible phosphorylation sites, including Ser^{273}, Thr^{275}, Ser^{276}, and Tyr^{277}. To then localize the site of phosphorylation, peptides identified likely to be phosphorylated were separated and analyzed by LC- Q-TOF MS/MS. The peptide fragment spectra reveal that the +79.97 Da mass shift, corresponding to phosphorylation, was localized to Ser^{278} for all phosphorylated *Ec*BcsG$^{\Delta N}$ peptides from the trypsin digest with high confidence (error ≤ 0.05 Da, FDR ≤ 1%; Figure 37 & Table 6).

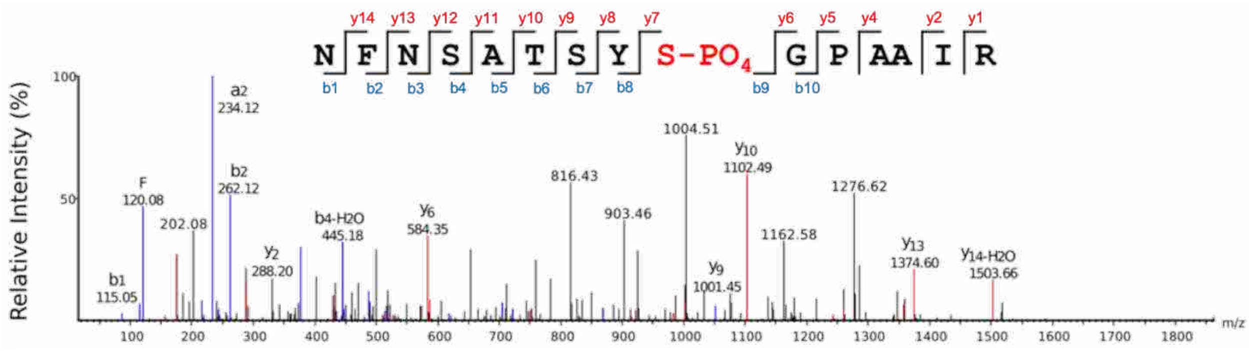


Figure 37: Identification of a phosphorylated *Ec*BcsG$^{\Delta N}$ intermediate by LC- Q-ToF MS/MS. Phosphorylated peptides were observed in the trypic peptide library of digested *Ec*BcsG$^{\Delta N}$, and subsequent MS/MS localized this phosphorylation to the putative catalytic nucleophile Ser^{278}.

Table 6: Identification of phosphorylated peptides by LC- Q-ToF MS/MS[1]

#	Immonium	b	b-H_2O	b-NH_3	Seq	y	y-H_2O	y-NH_3	#
1	87.06	115.05	97.04	98.02	N				15
2	120.08	262.12	244.11	245.09	F	1521.67	1503.64	1504.65	14
3	87.06	376.16	358.15	359.14	N	1374.59	1356.60	1357.61	13
4	60.04	463.19	445.18	446.17	S	1260.58	1242.58	1243.54	12
5	44.04	534.21	516.23	517.20	A	1173.53	1155.50	1156.50	11
6	74.06	635.28	617.27	618.25	T	1102.49	1084.48	1085.47	10
7	60.04	722.30	704.30	705.31	S	1001.49	983.45	984.42	9
8	136.08	885.37	867.36	868.34	Y	914.40	896.40	897.39	8
9	140.01	1052.41	1034.36	1035.35	S (+79.97)	751.35	733.34	734.32	7
10	30.03	1109.41	1091.38	1092.39	G	584.35	566.34	567.32	6
11	70.07	1206.45	1188.44	1189.42	P	527.32	509.32	510.30	5
12	44.05	1277.48	1259.47	1260.46	A	430.27	412.27	413.25	4
13	44.05	1348.52	1330.51	1331.49	A	359.24	341.23	342.21	3
14	86.10	1461.61	1443.59	1444.58	I	288.20	270.19	271.18	2
15	129.11				R	175.12	157.06	158.09	1

[1] All b and y ions corresponding to Figure 31 are coloured in blue and red, respectively. All ions were detected with ≤ 0.05 Da error.

However, equivalent peptides bearing a mass shift of +123.74 Da, which would correspond to the covalent pEtN intermediate, were not detected. The absence of pEtN-linked peptides in the *Ec*BcsG$^{\Delta N}$ tryptic peptide library was not surprising, as the covalent enzyme intermediate would likely be short lived and unstable in the context of active enzyme in an aqueous solvent, an outcome supported by the high apparent rate of esterase activity by *Ec*BcsG$^{\Delta N}$ described here (8.26 µM min^{-1}). As such, the detection of the covalent enzyme intermediate through tryptic digestion in solution would require the stability of the covalent pEtN enzyme intermediate on the order of hours, which would likely not be possible under the constraints of this experiment.

6. Summary Conclusions

6.1 BcsG as a member of the phosphoethanolamine transferase family

The first objective in this thesis was a structural characterization of *Ec*BcsG. The experimental approach was appropriate and successful, revealing the three-dimensional structure of the catalytically important C-terminal domain of *Ec*BcsG to near-atomic resolution. The structural model presented in this thesis respresents a significant contribution to the field of microbial biofilms and the study of phosphoethanolamine transferases in several distinct ways.

First, the data presented here demonstrates that *Ec*BcsG, and presumably, orthologous BcsG proteins, are both structurally and mechanistically similar to the polymyxin resistance factors *Cj*EptC, *Nm*EptA, and MCR-1. This high degree of structural similarity of *Ec*BcsG to the polymyxin resistance factors is in stark contrast to its low amino acid sequence identity to these proteins (22.4% *Cj*EptC, 29.4% *Nm*EptA, 25.0% MCR-1). Additionally, *Ec*BcsG catalyzes the transfer of pEtN to cellulose, while all three of the structural homologues discussed here catalyze the transfer of pEtN to lipid A, a vastly different substrate. However, sequence similarity and other *in silico* predictive tools estimate there are < 60 known or predicted pEtN transferase family members identifiable from bacterial reference genomes, as reported by Harper *et al.*[63]. Furthermore, phylogenetic analysis of these predicted pEtN transferases, first reported by Harper *et al.* and updated here (Figure 13), suggest the grouping of these enzymes into distinct subfamilies, which appear to cluster based on substrate specificity, supported by biochemical characterization of a limited number of the analyzed sequences available in the literature. The low sequence identity of *Ec*BcsG to all other known pEtN transferases not only demonstrates the limited predictive ability of sequence analysis for this family, but also indicates that the pEtN transferase family is likely far more expansive than predicted by Harper *et al.* Furthermore, *Ec*BcsG does not appear to cluster with any pEtN transferase family member in the phylogenetic analysis, which may be indicative that *Ec*BcsG, and related orthologues could represent a novel subfamily of pEtN transferases for which no other members have been described.

At present, the availability of structures of pEtN transferases is limited only to the polymyxin resistance factors described here. While these enzymes are functionally distinct from *Ec*BcsG, it appears *Ec*BcsG shares a highly conserved catalytic fold and mechanism, despite an apparent sequence dissimilarity and phylogenetic distance. Although an understanding of the various roles of pEtN transferases in bacteria are in their infancy, the discovery of BcsG as a pEtN transferase suggests that the broader pEtN transferase family may share a highly conserved overall fold and mechanism, despite the apparent variety of distinct reactions catalyzed by them. Additionally, a link between the enzymes that confer resistance to antimicrobial compounds and the enzymes that assemble the extracellular matrix demonstrates a clear and unifying emergent theme for this enzyme family, serving to enhance bacterial persistence and being shaped by the selective pressure of harsh, inhospitable, or bactericidal environments.

Although it was not used in the structure solution or analysis of *Ec*BcsG discussed here, the structure and function of BcsG from *Salmonella typhimirium* (*St*BcsG; PDB id 5OLT) was published during the timeline of this thesis by Sun *et al.* (2018)[51]. The structure of *Ec*BcsG discussed here and of *St*BcsG, aligned using PyMOL with a r.m.s.d. of 0.531 Å across 2278 atoms, indicating a global similarity and suggesting these enzymes are truly structural orthologues. Further site-directed mutagenesis studies of a limited set of residues reflected particular phenotypes pointing to the importance of *St*BcsG to the pEtN cellulose biofilm[51], in support of the model proposed here. In conjunction, the role of the *St*BcsG N-terminal domain in assembly of the mature bacterial cellulose synthase was confirmed[51], which may be further investigated in *Ec*BcsG.

6.2 BcsG is a phosphoethanolamine transferase essential for extracellular matrix architecture and assembly

The second objective of this thesis was a functional characterization of *Ec*BcsG. The data presented here provides, for the first time, an unambiguous demonstration of the true role of BcsG in *E. coli*, both at the molecular level, and at the level of the microbial biofilm. Furthermore, this thesis not only presented a model for the molecular and biological roles of BcsG and phosphoethanolamine cellulose, but also demonstrated the molecular determinants for BcsG activity.

The discovery of phosphoethanolamine cellulose by Thongsomboon *et al.*[32] and its importance in the role of infection by uropathogenic *E. coli*, reported by Hollenbeck *et al.*[61] demonstrated that the role of *bcsG* in microbial cellulose biosynthesis and in microbial pathogenesis and persistence was not trivial. This thesis demonstrates, for the first time, that *Ec*BcsG-directed pEtN cellulose is required for the structural integrity of the extracellular matrix of *E. coli*. Furthermore, disruptions in the catalytic features of *Ec*BcsG, first identified in this research, result in a biofilm phenotype that is indistinguishable from a chromosomal deletion of *bcsG*. This finding was coupled to an *in vitro* investigation, which demonstrated that the C-terminal catalytic domain of *Ec*BcsG alone, in the absence of any other Bcs complex partner, was sufficient to recapitulate the transfer of pEtN to various cellooligosaccharides, and that disruptions in the catalytic features recapitulated the lack of activity seen *in vivo*.

However, several questions about BcsG still remain. In particular, it still remains to be seen how the activity of BcsG is regulated to produce a polymer that is homogenously pEtN-substituted with a precise stoichiometry of one pEtN to every-other saccharide as noted by Thongsomboon *et al.* (2018)[32]. Furthermore, the interaction of BcsG with other Bcs complex members is not well characterized, and the role of c-di-GMP and/or protein-protein interactions in phosphoethanolamine cellulose biosynthesis are not yet clear. The specific roles of *bcsE* and *bcsF*, encoded on a shared operon with *bcsG*, remain to be seen. Although identified as a novel class of c-di-GMP receptor, BcsE was not necessary for the activity of *Ec*BcsG$^{\Delta N}$ *in vitro*. A bacterial two-hybrid study by Thongsomboon *et al.* showed that BcsE, BcsF and BcsG appear to form stable interactions *in vivo*, and that BcsG appears to interact with the BcsA/B synthase complex[32]. Sun *et al.* also demonstrated that the proper folding and membrane insertion of the BcsG N-terminal domain is required for proteolytic stability of BcsA[51]. However, no clear model of Bcs complex assembly is emergent from these observations.

At present, the role of c-di-GMP signalling in the control of BcsG-directed pEtN cellulose is unknown. Allosteric regulation of cellulose biosynthesis at the level of BcsA enzymatic activity is known to be dependent upon c-di-GMP binding to the PilZ domain of BcsA[53], and so it is tempting to speculate that BcsE might serve an analogous role in the allosteric regulation of

BcsG activity. One other distinct mechanism of exopolysaccharide synthase complex regulation has been described in support of this theory[108]. The biosynthesis of poly-β-1,6-*N*-acetylglucoasmine (PNAG), an exopolysaccharide produced by *E. coli* and other bacteria, is carried out by both PgaC and PgaD, the two inner membrane components of the PNAG synthesis machinery responsible for glycosyltransferase activity[109]. Binding of c-di-GMP to both PgaC and PgaD stabilizes their interaction and promotes enzyme activity resulting in PNAG synthesis, and proteolytic stability of PgaD *in vivo* is c-di-GMP dependent[108].

Such a two-level mechanism of regulation of the pEtN cellulose synthase is certainly possible under the current model for its function and protein-protein interaction. Already, the proteolytic stability of BcsA appears to be dependent upon interaction with the BcsG N-terminal domain[51], and its activity further depends on the binding of c-di-GMP to the BcsA PilZ domain[53]. Whether the proteolytic stability or activity of BcsA and/or BcsG depends upon ligand binding to the BcsE GIL domain remains to be seen, although such a mechanism seems plausible given the apparent structural and functional homology among Gram-negative synthase-dependent exopolysaccharide biosynthetic machines[23].

6.3 Mechanism of action of pEtN transferases

The third objective of this thesis was an investigation of the catalytic mechanism of *Ec*BcsG. The experimental methods and data described here were successful in demonstrating a limited set of the requirements for catalysis through a ping-pong bi-bi reaction mechanism. The results presented here support the hypothesis that pEtN transferases proceed by this conserved mechanism, although the molecular details of the complete catalytic cycle still have yet to be demonstrated.

While the individual substrates for which known and predicted pEtN transferases catalyze transfer of pEtN vary greatly, it is generally believed that the physiological donor of pEtN in all cases is the membrane phospholipid phosphatidylethanolamine, enriched in the inner membrane of Gram-negative bacteria, and supported by experimental evidence from Thoingsomboon *et al.*[32]. Such a model accounts for the presence of an N-terminal transmembrane domain known or predicted to localize to the inner membrane for all pEtN transferases[63]. This model then supports catalysis through a ping-pong bisubstrate-biproduct reaction pathway, in which the catalytic domain cleaves the pEtN group with a catalytic nucleophile and liberating diacylglyerol as the first enzymatic product, and subsequently presents the pEtN to the acceptor cosubstrate, at which point it is cleaved and covalently bonded to the second enzymatic product[72]. A ping-pong mechanism of action would then mandate this catalytic cycle be completed in two discrete steps, where binding of the first and second cosubstrates occur only to the holoenzyme and the covalent enzyme intermediate, respectively. However, this mechanism remains highly speculative, and the only experimental evidence available to support such a mechanism was reported by Hinchliffe *et al.*[72] that density functional theory calculations on available structures of MCR-1 indicate that at least the first step of this mechanism is energetically feasible.

This thesis then provides, for the first time, direct and unambiguous experimental evidence for catalysis through a ping-pong bi-bi mechanism of action. The existence of a covalent pEtN-BcsG enzyme intermediate was identified and is the first enzyme in the pEtN transferase family known to possess such a covalent intermediate. Furthermore, Cleland's rules[107] were applied, and demonstrated that the apparent kinetic behaviour of BcsG support a ping-pong bi-bi mechanism. Based on the existence of a covalent enzyme-intermediate, and the exclusion of a sequential bi-bi mechanism based on kinetic behaviour, the data presented here confirm that BcsG, and likely the known members of the pEtN transferase family described here, proceed through a ping-pong bi-bi reaction mechanism.

However, several mechanistic details remain to be seen, and there is limited evidence for the role of each essential amino acid in the catalytic cycle. For example, the results demonstrate that H396 and Y277 are essential for *Ec*BcsG activity both *in vivo* and *in vitro*. However, the role of these amino acids cannot be inferred from the structural model of *Ec*BcsG. Further efforts to experimentally resolve a BcsG-substrate complex would be required for a complete understanding of how these residues participate in the catalytic cycle[78]. Furthermore, the structure and *in vitro* activity were demonstrated using a recombinant construct of BcsG containing a defined truncation of the N-terminal transmembrane domain. The structure of the full-length *Nm*EptA pEtN transferase, including the N-terminal transmembrane domain, was also reported by Anandan *et al* (PDB id 5FGN), and suggested how conformational change may influence lipid-binding, and play an important catalytic function throughout the catalytic cycle. Instead, our experimental approach focused on an artificial pEtN donor not linked to a lipid. Accordingly, the catalytic importance of a lipid-binding site in the N-terminal domain and its interactions with the membrane and/or other membrane proteins was not investigated. Finally, while the results here establish a plausible, experimentally-supported mechanism for the interaction of *Ec*BcsG with the pEtN donor during the first step of the proposed catalytic cycle, the mechanistic basis for the binding of cellulose and subsequent transfer of pEtN during the second step of the catalytic cycle remain completely unknown. A complete understanding of the molecular determinants of binding to both cosubstrates is required before a complete catalytic mechanism can be confidently proposed.

6.4 BcsG as a therapeutic target for microbial infections

The success of this thesis in addressing all three key objectives outlined within has provided significant advancement of the contemporary understanding of microbial phosphoethanolamine cellulose biosynthesis. Beyond this, it is critically important to contextualize the research findings of this thesis and highlight the potential impacts and significance that this new knowledge will provide to collective humankind.

Biofilms have been estimated to be implicated in >60% of infections observed in the clinic[3], making them an important factor in public health and infectious bacterial disease. Additionally, biofilm-associated genes in pathogens are considered virulence factors[12], contributing to antimicrobial resistance. The importance of pEtN cellulose to uropathogenic *E. coli* during the course of urogenital tract infection further highlights the link between biofilm formation and infection[20,61]. Antimicrobial resistant infections are predicted to be the leading cause of death worldwide by 2050, surpassing annual mortality by cancer and cardiovascular disease[79]. Beyond infection, the presence of biofilms on fresh produce or other food items can also cause significant economic losses in addition to health concerns[14]. For these reasons, the development of new therapeutic strategies, biocontrol agents, or new clinical drugs is essential to combat the rising threat of recalcitrant bacterial infection and persistence.

The structure, activity, and mechanism of BcsG that is explored in this thesis all provide promising evidence that BcsG would be a suitable therapeutic target to attenuate the virulence and persistence of bacteria who produce pEtN cellulose. As discussed above, the work presented here demonstrates that BcsG alone is essential for the architecture and assembly of the biofilm matrix in *E. coli*. Accordingly, the development of a small-molecule inhibitor, or other method, to abrogate BcsG activity would represent an effective strategy for the treatment of drug-resistant or otherwise persistent colonization by bacteria. Furthermore, the results in this thesis show that obstruction of pEtN cellulose is not cytotoxic to cells, and such a strategy to inhibit the biofilm of bacteria likely would not be met with strong selective pressure for resistance. This is in contrast to the emerging resistance of bacteria to conventional antimicrobial drugs, which has rapidly become a global threat to public health[79]. However, pEtN and phosphatidylethanolamine also contribute to eukaryotic cell membranes, and so the genes responsible for the biosynthesis and turnover of pEtN and pEtN-associated lipids may be conditionally essential to humans, making pEtN transferases a challenging molecular target. For example, the *PHOSPHO1* gene, encoding a phosphocholine/phosphoethanolamine phosphatase is important for liberating inorganic phosphate in mineralizing cells[110], and so disruption of the cellular pEtN pool may have deleterious effects on bone growth and development in humans and animals. Furthermore, because of the apparently conserved structure and mechanism of pEtN transferases explored here, such a strategy to abrogate BcsG-directed pEtN cellulose may also inhibit the activity of other antimicrobial drug resistance enzymes, limiting the potential for resistance for polymyxins - at least by the mechanisms described here.

The requirements for pEtN transferase activity appear to be conserved, as the results herein recapitulate those of Hinchcliffe *et al.* for MCR-1[72]. Zinc-binding, disulfide bond formation, and the correct orientation and chemistry of amino acids involved in contacts with either cosubstrate are indispensable for the activity of BcsG both *in vivo* and *in vitro*. This knowledge of the essential catalytic features of BcsG, and the resolution of its overall structure, represent significant advances in contemporary knowledge of bacterial cellulose biosynthesis and enable future efforts for structurally informed inhibitor design.

Although only just recently discovered in *E. coli* and *S. enterica* at the time of writing, pEtN cellulose represents a naturally produced, chemically modified form of cellulose which may prove useful as a biomaterial. The research presented in this thesis, in addition to the findings contributed by Hollenbeck *et al.*[61] and Thongsomboon *et al.*[32], demonstrate that pEtN cellulose has unique properties from unmodified microbial cellulose that contribute to the rigidity of the extracellular matrix. Although the mechanism for this apparent rigidity remains to be seen, Hollenbeck *et al.* demonstrated that pEtN cellulose appears to serve in maintaining the association of curli fimbriae to the bacterial cell surface[61]

7. Integrative Nature of This Research

Biofilms are one of the most successful and widely distributed forms of life on Earth and are the predominant form of life for bacteria and archaea[1,111]. Biofilms are complex, self-assembling extracellular matrices with large diversity in the biosynthesis of the materials that constitute them, the properties they confer to their constituents, the microorganisms that produce them, and the habitats they colonize. Accordingly, a complete understanding of biofilms requires a comprehensive model that accounts for the biological processes that govern them, ranging from the molecular level to the level of microbial ecology. The research described in this thesis concerns, primarily, the molecular mechanisms underlying the biosynthesis of the exopolysaccharide from the model organism *E. coli*. In order to address the complex phenomenon of biofilms that span many levels of the hierarchy of biological science, the experimental approach of this thesis included a corresponding investigation of the biofilm at many different levels. Beginning at the level of small molecules, this research describes the enzymatic product of BcsG, pEtN cellulose, as well as the individual molecular determinants for BcsG catalysis. The thesis also addresses this phenomenon at the enzyme level, demonstrating the overall fold and structure of the BcsG enzyme. Furthermore, this thesis addresses the consequences of BcsG-mediated pEtN cellulose on the properties of the biofilm at the cellular and the population level using the pellicle phenotype experiment. This experimental approach that encompasses many levels of biology is exemplary of integrative research.

Beyond the study of biological processes, this thesis also describes multidisciplinary research spanning the disciplines of the life sciences. For example, synthetic and analytical chemistry were paramount to the experimental approaches used here to demonstrate the activity of *Ec*BcsG *in vitro*. The near-atomic resolution structure of *Ec*BcsG described here was accomplished by X-ray crystallography, a technique centered around modern innovations in the fields of physics, chemistry and mathematics. Finally, computational biology and sequence analysis were important experimental techniques necessary to contextualize the structure of *Ec*BcsG and understand its evolutionary relationship to other known sequences and structures. The research described in this thesis involved many multidisciplinary approaches to answer important and meaningful questions about biological processes that could not be addressed with the conventional tools for the study of biology alone.

Finally, the results presented here are the result of a largely collaborative effort, both within the institution and outside. The strains used in this study were the result of a collaboration with Thongsomboon *et al.*, the chemical synthesis of *p*-NPPE was possible through a collaboration with Dr. Ken Maly in the Department of Chemistry & Biochemistry at Wilfrid Laurier University, and the mass spectrometry experiments were performed with the expert supervision of Dr. Armen Charchoglyan and Dr. Dyanne Brewer of the Advanced Analysis Facility, University of Guelph.

Through the study of biology at many levels, through many diverse and multidisciplinary techniques, and through many collaborations, the research presented here was made possible. For these reasons, the very nature of this thesis is Integrative Biology.

8. Summary and significance

Biofilms have been estimated to be implicated in >60% of infections observed in the clinic[3], making them an important factor in public health and infectious bacterial disease. Additionally, biofilm-associated genes in pathogens are considered virulence factors[12], contributing to antimicrobial resistance. Antimicrobial resistant infections are predicted to be the leading cause of death worldwide by 2050, surpassing annual mortality by cancer and cardiovascular disease[79]. The presence of biofilms on fresh produce or on other foodstuff can also cause significant economic losses in addition to health concerns[14]. The structural and functional model of the cellulose modifying complex offers opportunities in structure-based drug discovery and other efforts in inhibiting microbial biofilms and limiting the resilience of biofilm-forming pathogens. Additionally, the new understanding of BcsG-directed phosphoethanolamine cellulose production provided by this thesis may enable biosynthetic engineering of new cellulosic materials or confer the advantages of phosphoethanolamine cellulose in new organisms.

9. References Cited

1. Flemming, H.-C. *et al.* Biofilms: an emergent form of bacterial life. *Nat. Rev. Microbiol.* **14**, 563–575 (2016).
2. Papenfort, K. & Bassler, B. L. Quorum sensing signal–response systems in Gram-negative bacteria. *Nat. Rev. Microbiol.* **14**, 576–588 (2016).
3. Fux, C. A., Costerton, J. W., Stewart, P. S. & Stoodley, P. Survival strategies of infectious biofilms. *Trends Microbiol.* **13**, 34–40 (2005).
4. Marsh, P. Dental plaque as a microbial biofilm dental plaque – Existing perspective. *Caries Res* **38**, 204–211 (2004).
5. Spiers, A. J., Bohannon, J., Gehrig, S. M. & Rainey, P. B. Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose. *Mol. Microbiol.* **50**, 15–27 (2003).
6. Valentini, M. & Filloux, A. Biofilms and cyclic di-GMP (c-di-GMP) signaling: Lessons from *Pseudomonas aeruginosa* and other bacteria. *J. Biol. Chem.* **291**, 12547–55 (2016).
7. Flickinger, S. T. *et al.* Quorum sensing between *Pseudomonas aeruginosa* biofilms accelerates cell growth. *J. Am. Chem. Soc.* **133**, 5966–5975 (2011).
8. Stewart, P. S. & William Costerton, J. Antibiotic resistance of bacteria in biofilms. *Lancet* **358**, 135–138 (2001).
9. Marsh, P. D. Microbial ecology of dental plaque and its significance in health and disease. *Adv. Dent. Res.* **8**, 263–71 (1994).
10. McCrate, O. A., Zhou, X., Reichhardt, C. & Cegelski, L. Sum of the parts: Composition and architecture of the bacterial extracellular matrix. *J. Mol. Biol.* **425**, 4286–4294 (2013).
11. Römling, U. *et al.* Occurrence and regulation of the multicellular morphotype in *Salmonella* serovars important in human disease. *Int. J. Med. Microbiol.* **293**, 273–285 (2003).

12. Ciofu, O., Tolker-Nielsen, T., Jensen, P. Ø., Wang, H. & Høiby, N. Antimicrobial resistance, respiratory tract infections and role of biofilms in lung infections in cystic fibrosis patients. *Adv. Drug Deliv. Rev.* **85**, 7–23 (2015).

13. Rajwade, J. M., Paknikar, K. M. & Kumbhar, J. V. Applications of bacterial cellulose and its composites in biomedicine. *Appl. Microbiol. Biotechnol.* **99**, 2491–2511 (2015).

14. Beloin, C., Roux, A. & Ghigo, J. M. *Escherichia coli* biofilms. *Curr. Top. Microbiol. Immunol.* **322**, 249–89 (2008).

15. Aas, F. E. *et al. Neisseria gonorrhoeae* type IV pili undergo multisite, hierarchical modifications with phosphoethanolamine and phosphocholine requiring an enzyme structurally related to lipopolysaccharide phosphoethanolamine transferases. *J. Biol. Chem.* **281**, 27712–27723 (2006).

16. Waldner, L., MacKenzie, K., Köster, W. & White, A. From exit to entry: Long-term survival and transmission of *Salmonella*. *Pathogens* **1**, 128–155 (2012).

17. MacKenzie, K. D., Palmer, M. B., Köster, W. L. & White, A. P. Examining the link between biofilm formation and the ability of pathogenic *Salmonella* strains to colonize multiple host species. *Front. Vet. Sci.* **4**, 1–19 (2017).

18. Beebout, C. J. *et al.* Respiratory heterogeneity shapes biofilm formation and host colonization in uropathogenic *Escherichia coli*. *MBio* **10**, 1–16 (2019).

19. Saldaña, Z. *et al.* Synergistic role of curli and cellulose in cell adherence and biofilm formation of attaching and effacing *Escherichia coli* and identification of Fis as a negative regulator of curli. *Environ. Microbiol.* **11**, 992–1006 (2009).

20. Whitfield, G. B., Marmont, L. S. & Howell, P. L. Enzymatic modifications of exopolysaccharides enhance bacterial persistence. *Front. Microbiol.* **6**, 1–21 (2015).

21. Arciola, C. R., Campoccia, D., Ravaioli, S. & Montanaro, L. Polysaccharide intercellular adhesin in biofilm: structural and regulatory aspects. *Front. Celluar Infect. Microbiol.* **5**, 1–10 (2015).

22. Römling, U. Molecular biology of cellulose production in bacteria. *Res. Microbiol.* **153**, 205–212 (2002).

23. Low, K. E. & Howell, P. L. Gram-negative synthase-dependent exopolysaccharide biosynthetic machines. *Curr. Opin. Struct. Biol.* **53**, 32–44 (2018).

24. Whitney, J. C. & Howell, P. L. Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. *Trends Microbiol.* **21**, 63–72 (2013).

25. Islam, S. T. & Lam, J. S. Synthesis of bacterial polysaccharides via the Wzx/Wxy-dependent pathway. *Can. J. Microbiol.* **716**, 697–716 (2014).

26. Kalynych, S., Valvano, M. A. & Cygler, M. Polysaccharide co-polymerases: the enigmatic conductors of the O-antigen assembly orchestra. *Protein Eng. Des. Sel.* **25**, 797–802 (2012).

27. Cuthbertson, L., Mainprize, I. L., Naismith, J. H. & Whitfield, C. Pivotal Roles of the outer membrane polysaccharide export and polysaccharide copolymerase protein families in export of extracellular polysaccharides in Gram-Negative bacteria. *Microbiol. Mol. Biol. Rev.* **73**, 155–177 (2009).

28. Keiski, C.-L. *et al.* AlgK Is a TPR-containing protein and the periplasmic component of a novel exopolysaccharide secretin. *Structure* **18**, 265–273 (2010).

29. Omadjela, O. *et al.* 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**, 17856–61 (2013).

30. Ahmad, I. *et al.* BcsZ inhibits biofilm phenotypes and promotes virulence by blocking cellulose production in *Salmonella enterica* serovar Typhimurium. *Microb. Cell Fact.* **15**, 177 (2016).

31. Baker, P. *et al. P. aeruginosa* SGNH hydrolase-like proteins AlgJ and AlgX have similar topology but separate and distinct roles in alginate acetylation. *PLoS Pathog.* **10**, (2014).

32. Thongsomboon, W. *et al.* Phosphoethanolamine cellulose: A naturally produced chemically modified cellulose. *Science (80-.).* **359**, 334–338 (2018).

33. Little, D. J. *et al.* Modification and periplasmic translocation of the biofilm exopolysaccharide poly-β-1,6-N-acetyl-D-glucosamine. *Proc. Natl. Acad. Sci.* **111**, 11013–11018 (2014).

34. Little, D. J. *et al.* The protein BpsB is a poly-β-1,6-N-acetyl-D-glucosamine deacetylase required for biofilm formation in *Bordetella bronchiseptica*. *J. Biol. Chem.* **290**, 22827–22840 (2015).

35. Pokrovskaya, V. *et al.* Functional characterization of *Staphylococcus epidermidis* IcaB, a de-N-acetylase important for biofilm formation. *Biochemistry* **52**, 5463–5471 (2013).

36. Jerga, A., Raychaudhuri, A. & Tipton, P. A. *Pseudomonas aeruginosa* C5-mannuronan epimerase: Steady-state kinetics and characterization of the product. *Biochemistry* **45**, 552–560 (2006).

37. Little, D. J. *et al.* PgaB orthologues contain a glycoside hydrolase domain that cleaves deacetylated poly-β-(1,6)-N-acetylglucosamine and can disrupt bacterial biofilms. *PLoS Pathog.* **14**, 1–29 (2018).

38. Whitfield, C. & Mainprize, I. L. TPR motifs: Hallmarks of a new polysaccharide export scaffold. *Structure* **18**, 151–153 (2010).

39. Römling, U. & Gaperin, M. Y. Bacterial cellulose biosynthesis: diversity of operons, subunits, products, and functions. *Trends Microbiol.* **23**, 545–557 (2015).

40. Ross, P., Mayer, R. & Benziman, M. Cellulose biosynthesis and function in bacteria. *Microbiol. Rev.* **55**, 35–58 (1991).

41. Singh, N., Patil, A., Prabhune, A. A., Raghav, M. & Goel, G. Diverse profiles of N-acyl-homoserine lactones in biofilm forming strains of *Cronobacter sakazakii*. *Virulence* **8**, 275–281 (2017).

42. Petersen, N. & Gatenholm, P. Bacterial cellulose-based materials and medical devices: Current state and perspectives. *Appl. Microbiol. Biotechnol.* **91**, 1277–1286 (2011).

43. McNamara, J. T., Morgan, J. L. W. & Zimmer, J. A molecular description of cellulose biosynthesis. *Annu. Rev. Biochem.* **84**, 895–921 (2015).

44. Knott, B. C., Crowley, M. F., Himmel, M. E., Zimmer, J. & Beckham, G. T. Simulations of cellulose translocation in the bacterial cellulose synthase suggest a regulatory mechanism for the dimeric structure of cellulose. *Chem. Sci.* **7**, 3108–3116 (2016).

45. Maeda, K. *et al.* Genetic identification of factors for extracellular cellulose accumulation in the thermophilic cyanobacterium *Thermosynechococcus vulcanus*: proposal of a novel tripartite secretion system. *Mol. Microbiol.* **109**, 121–134 (2018).

46. Morgan, J. L. W. *et al.* Observing cellulose biosynthesis and membrane translocation *in crystallo*. *Nature* **25**, 289–313 (2016).

47. Hu, L. *et al.* Analysis of the cellulose synthase operon genes, *bcsA, bcsB*, and *bcsC* in *Cronobacter* species: Prevalence among species and their roles in biofilm formation and cell-cell aggregation. *Food Microbiol.* **52**, 97–105 (2015).

48. Mazur, O. & Zimmer, J. Apo- and cellopentaose-bound structures of the bacterial cellulose synthase subunit BcsZ. *J. Biol. Chem.* **286**, 17601–6 (2011).

49. Hay, I. D., Rehman, Z. U. & Rehm, B. H. A. Membrane topology of outer membrane protein AlgE, which is required for alginate production in *Pseudomonas aeruginosa*. *Appl. Environ. Microbiol.* **76**, 1806–12 (2010).

50. Whitney, J. C. *et al.* Structural basis for alginate secretion across the bacterial outer membrane. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13083–8 (2011).

51. Sun, L. *et al.* Structural and functional characterization of the BcsG subunit of the cellulose synthase in *Salmonella typhimurium*. *J. Mol. Biol.* **430**, 3170–3189 (2018).

52. Marmont, L. S. *et al.* PelA and PelB form a modification and secretion complex essential for Pel polysaccharide-dependent biofilm formation in *Pseudomonas aeruginosa*. *J. Biol. Chem.* jbc.M117.812842 (2017). doi:10.1074/jbc.M117.812842

53. Morgan, J. L. W., McNamara, J. T. & Zimmer, J. Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nat. Struct. Mol. Biol.* **21**, 489–496 (2014).

54. Ross, P. *et al.* Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* **325**, 279–281 (2003).

55. Du, J., Vepachedu, V., Cho, S. H., Kumar, M. & Nixon, B. T. Structure of the cellulose synthase complex of *Gluconacetobacter hansenii* at 23.4 Å resolution. *PLoS One* **11**, 1–24 (2016).

56. Hinchliffe, P., Symmons, M. F., Hughes, C. & Koronakis, V. Structure and operation of bacterial tripartite pumps. *Annu. Rev. Microbiol.* **67**, 221–242 (2013).

57. Bera, A., Biswas, R., Herbert, S. & Gotz, F. The Presence of peptidoglycan O-acetyltransferase in various Staphylococcal species correlates with lysozyme resistance and pathogenicity. *Infect. Immun.* **74**, 4598–4604 (2006).

58. Moynihan, P. J. & Clarke, A. J. *O*-Acetylation of peptidoglycan in Gram-negative bacteria. *J. Biol. Chem.* **285**, 13264–13273 (2010).

59. Moynihan, P. J. & Clarke, A. J. Mechanism of action of peptidoglycan O-acetyltransferase B involves a Ser-His-Asp catalytic triad. *Biochemistry* **53**, 6243–6251 (2014).

60. Solano, C. *et al.* Genetic analysis of *Salmonella enteritidis* biofilm formation: Critical role of cellulose. *Mol. Microbiol.* **43**, 793–808 (2002).

61. Hollenbeck, E. C., Antonoplis, A., Chai, C., Thongsomboon, W. & Fuller, G. G. Phosphoethanolamine cellulose enhances curli-mediated adhesion of uropathogenic *Escherichia coli* to bladder epithelial cells. *Proc. Natl. Acad. Sci.* **115**, 10106–10111 (2018).

62. Fang, X. *et al.* GIL, a new c-di-GMP-binding protein domain involved in regulation of cellulose synthesis in enterobacteria. *Mol. Microbiol.* **93**, 439–452 (2014).

63. Harper, M. *et al.* Characterization of two novel lipopolysaccharide phosphoethanolamine transferases in *Pasteurella maltocida* and their role in resistance to cathelicidin-2. *Infect. Immun.* **85**, 1–16 (2017).

64. Salazar, J., Alarcón, M., Huerta, J., Navarro, B. & Aguayo, D. Phosphoethanolamine addition to the Heptose I of the lipopolysaccharide modifies the inner core structure and has an impact on the binding of Polymyxin B to the *Escherichia coli* outer membrane. *Arch. Biochem. Biophys.* **620**, 28–34 (2017).

65. Wanty, C. *et al.* The structure of the neisserial lipooligosaccharide phosphoethanolamine transferase A (LptA) required for resistance to polymyxin. *J. Mol. Biol.* **425**, 3389–3402 (2013).

66. Zhao, Y. *et al.* Structural and mechanistic insights into polymyxin resistance mediated by EptC originating from *Escherichia coli. FEBS J.* **286**, 750–764 (2018).

67. Maldonado, R. F., Sá-Correia, I. & Valvano, M. A. Lipopolysaccharide modification in gram-negative bacteria during chronic infection. *FEMS Microbiol. Rev.* **40**, 480–493 (2016).

68. Steimle, A., Autenrieth, I. B. & Frick, J. S. Structure and function: Lipid A modifications in commensals and pathogens. *Int. J. Med. Microbiol.* **306**, 290–301 (2016).

69. Hashii, N., Isshiki, Y., Iguchi, T. & Kondo, S. Structural characterization of the carbohydrate backbone of the lipopolysaccharide of *Vibrio parahaemolyticus* O-untypeable strain KX-V212 isolated from a patient. *Carbohydr. Res.* **338**, 2711–2719 (2003).

70. Zughaier, S. M., Kandler, J. L., Balthazar, J. T. & Shafer, W. M. Phosphoethanolamine modification of N*eisseria gonorrhoeae* lipid A reduces autophagy flux in macrophages. *PLoS One* **10**, 1–17 (2015).

71. Washizaki, A., Yonesaki, T. & Otsuka, Y. Characterization of the interactions between *Escherichia coli* receptors, LPS and OmpC, and bacteriophage T4 long tail fibers. *Microbiologyopen* **5**, 1003–1015 (2016).

72. Hinchliffe, P. *et al.* Insights into the mechanistic basis of plasmid-mediated colistin resistance from crystal structures of the catalytic domain of MCR-1. *Sci. Rep.* 1–10 (2017). doi:10.1038/srep39392

73. Stogios, P. J. *et al.* Substrate recognition by a colistin resistance enzyme from *Moraxella catarrhalis*. *ACS Chem. Biol.* **13**, 1–10 (2018).

74. Cullen, T. W. & Trent, M. S. A link between the assembly of flagella and lipooligosaccharide of the Gram-negative bacterium *Campylobacter jejuni. Proc. Natl. Acad. Sci.* **107**, 5160–5165 (2010).

75. Storm, D. R., Rosenthal, K. S. & Swanson, P. E. Polymyxin and related peptide antibiotics. *Annu. Rev. Biochem.* **46**, 723–763 (2003).

76. Olaitan, A. O., Morand, S. & Rolain, J. M. Mechanisms of polymyxin resistance: Acquired and intrinsic resistance in bacteria. *Front. Microbiol.* **5**, 1–19 (2014).

77. Fage, C. D., Brown, D. B., Boll, J. M., Keatinge-Clay, A. T. & Trent, M. S. Crystallographic study of the phosphoethanolamine transferase EptC required for polymyxin resistance and motility in *Campylobacter jejuni*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 2730–2739 (2014).

78. Anandan, A. *et al.* Structure of a lipid A phosphoethanolamine transferase suggests how conformational changes govern substrate binding. *Proc. Natl. Acad. Sci.* **114**, 2218–2223 (2017).

79. The Review on Antimicrobial Resistance. (2014). doi:10.1038/510015a

80. U.S. Centers for Disease Control. Antibiotic Reisstance Threats (Report). (2013).

81. Farha, M. A. & Brown, E. D. Strategies for target identification of antimicrobial natural products. *Nat. Prod. Rep.* **33**, 668–680 (2016).

82. Mah, T. F. & O'Toole, G. A. Mechanisms of biofilm resistance to antimicrobial agents. *Trends Microbiol.* **9**, 34–39 (2001).

83. Del Pozo, J. L. & Patel, R. The challenge of treating biofilm-associated bacterial infections. *Clin. Pharmacol. Ther.* **82**, 204–209 (2007).

84. Snarr, B. D., Howell, P. L. & Sheppard, D. C. Hoisted by their own petard: do microbial enzymes hold the solution to treating and preventing biofilm infections? *Future Microbiol.* **13**, 395–398 (2018).

85. Asker, D., Awad, T. S., Baker, P., Howell, P. L. & Hatton, B. D. Non-eluting, surface-bound enzymes disrupt surface attachment of bacteria by continuous biofilm polysaccharide degradation. *Biomaterials* **167**, 168–176 (2018).

86. Baker, P. *et al.* Exopolysaccharide biosynthetic glycoside hydrolases can be utilized to disrupt and prevent *Pseudomonas aeruginosa* biofilms . *Sci. Adv.* **2**, e1501632 (2016).

87. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 3240–3244 (1973).

88. Hutchison, C. A. *et al.* Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.* **253**, 6551–6560 (1978).

89. Gasteiger, E. *et al.* Protein identification and analysis tools on the ExPASy server. (Humana Press, 2005).

90. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).

91. Burnette, W. N. ‘Western Blotting’: Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal. Biochem.* **112**, 195–203 (1981).

92. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).

93. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).

94. Evans, P. & McCoy, A. An introduction to molecular replacement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 1–10 (2007).

95. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

96. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

97. Heping Zheng, W. M. Validating metal binding sites in macromolecule structures using the CheckMyMetal web server. *Nat. Protoc.* **9**, 156–170 (2014).

98. Holm, L., Kääriäinen, S., Wilton, C. & Plewczynski, D. Using Dali for structural comparison of proteins. *Curr. Protoc. Bioinforma.* 1–24 (2006). doi:10.1002/0471250953.bi0505s14

99. Wenzel, C. Q. *et al.* Functional characterization of Lpt3 and Lpt6, the inner-core lipooligosaccharide phosphoethanolamine transferases from *Neisseria meningitidis*. *J. Bacteriol.* **192**, 208–216 (2010).

100. Bontemps-Gallo, S. *et al.* Biosynthesis of osmoregulated periplasmic glucans in *Escherichia coli*: The phosphoethanolamine transferase is encoded by *opgE*. *Biomed Res. Int.* **2013**, 1–8 (2013).

101. Scott, N. E. *et al.* Modification of the *Campylobacter jejuni* N-linked glycan by EptC protein-mediated addition of phosphoethanolamine. *J. Biol. Chem.* **287**, 29384–29396 (2012).

102. Duarte, F., Amrein, B. A. & Kamerlin, S. C. L. Modeling catalytic promiscuity in the alkaline phosphatase superfamily. *Phys. Chem. Chem. Phys.* **15**, 11160–11177 (2013).

103. Chew, K. L., La, M.-V., Lin, R. T. P. & Teo, J. W. P. Colistin and Polymyxin B susecptibility testing for carbapenem-reistant and MCR-positive Enterobacteriaceae: Comparison of Sensititre, MicroScan, Vitek 2, and Etest with broth microdilution. *J. Clin. Microbiol.* **55**, 2609–2616 (2017).

104. Reichhardt, C. *et al.* Influence of the amyloid dye Congo Red on curli, cellulose, and the extracellular matrix in *E. coli* during growth and matrix purification. *Anal. Bioanal. Chem.* **408**, 7709–7717 (2016).

105. Dimitrijevic, M., Grujic-injac, B. & Lajsic, S. The synthesis of a new phospholipid from the Koilin glandular layer of chicken gizzard. *Hoppe Seylers Z Physiol Chem.* **360**, 477–480 (1979).

106. Little, D. J. *et al.* The structure- and metal-dependent activity of *Escherichia coli* PgaB provides insight into the partial de-N-acetylation of poly-β-1,6-N-acetyl-D-glucosamine. *J. Biol. Chem.* **287**, 31126–31137 (2012).

107. Cleland, W. W. The kinetics of enzyme-catalyzed reactions with two or more substrates or products. III. Prediction of initial velocity and inhibition patterns by inspection. *Biochim. Biophys. Acta* **67**, 188–96 (1963).

108. Steiner, S., Lori, C., Boehm, A. & Jenal, U. Allosteric activation of exopolysaccharide synthesis through cyclic di-GMP-stimulated protein–protein interaction. *EMBO J.* **32**, 354–368 (2012).

109. Wang, X., Iii, J. F. P. & Romeo, T. The *pgaABCD* locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required for biofilm formation. *J. Bacteriol.* **186**, 2724–2734 (2004).

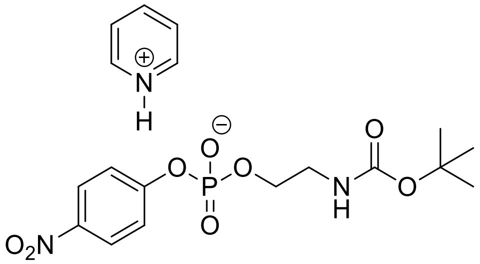
110. Roberts, S. J., Stewart, A. J., Sadler, P. J. & Farquharson, C. Human PHOSPHO1 exhibits high specific phosphoethanolamine and phosphocholine phosphatase activities. *Biochem. J.* **382**, 59–65 (2004).

111. Flemming, H. C. & Wuertz, S. Bacteria and archaea on Earth and their abundance in biofilms. *Nat. Rev. Microbiol.* **17**, 247–260 (2019).

Appendix 1

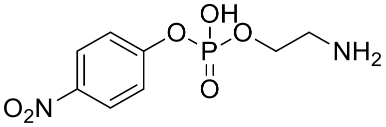
Synthesis of *p*-nitrophenylphosphorylethanolamine. Reagents and starting materials were purchased from Millipore-Sigma and used as received. Anhydrous and oxygen-free solvents were dispensed from a custom-built solvent purification system which used purification columns packed with activated alumina and supported copper catalyst (Glasscontour, Irvine, CA) and maintained under argon. Melting points were determined on a Mel-Temp® Electrothermal melting point apparatus and are uncorrected. High resolution mass spectra were recorded at the Centre Régional de Spectrométrie de Masse à l'Université de Montréal using an Agilent LC-MSD TOF spectrometer in the ionisation mode indicated. Elemental analysis was conducted by Midwest Microanalysis. NMR spectra were recorded on an Agilent Technologies 400 MHz (^{1}H, ^{13}C, ^{19}F) or 300 MHz (^{31}P) spectrometer using deuterated solvents purchased from CIL Int. Chemical or Millipore-Sigma; shifts are reported (δ-scale) using the residual solvent peak as reference. In the case of ^{19}F and ^{32}P NMR no reference was used.

(i) pyridin-1-ium 2-((tert-butoxycarbonyl)amino)ethyl (4-nitrophenyl) phosphate

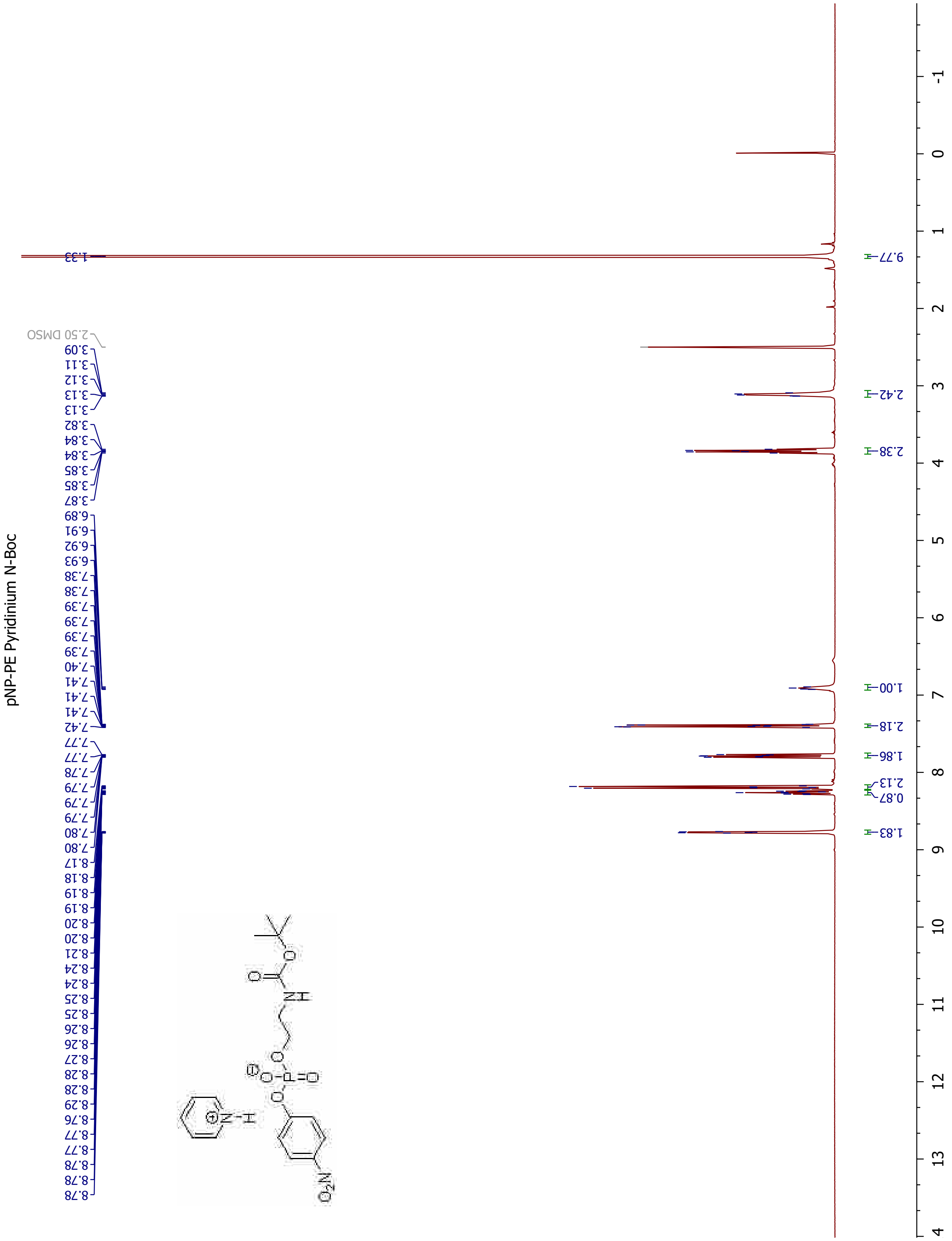


To a 100 mL 2-neck round-bottomed flask, equipped with N_2 inlet and magnetic stir bar, was added *p*-nitrophenyl phosphodichloridate (3.64 g, 14.2 mmol, 2 eq), dry THF (35 mL), and pyridine (2.3 mL, 28.4 mmol, 4 eq), and the resulting clear colourless solution was cooled to 0 °C (ice-water bath). A solution of N-Boc-ethanolamine (1.10 mL, 7.11 mmol, 1 eq) in dry THF (20 mL) was added dropwise *via* syringe to the cold solution with formation of a white precipitate, and the mixture stirred at room temperature overnight. A solution of pyridine (2.3 mL, 28.4 mmol, 4 eq) in H_2O (10 mL) was then added and the clear solution stirred 10 minutes further. The volatiles were removed *via* rotary evaporation yielding a yellow oil which was dissolved in H_2O (50 mL). The aqueous layer was washed with EtOAc (3 × 50 mL), extracted with CH_2Cl_2 (3 × 50 mL), and the CH_2Cl_2 layer washed with H_2O (1 × 25 mL), dried over Na_2SO_4, and concentrated *via* rotary evaporation yielding a clear oil which solidified upon standing to give a white solid (1.43 g, 3.24 mmol, 46%). mp (EtOAc) = 98.5−99.5 °C; ^{1}H NMR (400 MHz, DMSO-d$_6$) δ: 8.77 (dt, J = 4.9, 1.6 Hz, 2H), 8.26 (tt, J = 7.8, 1.7 Hz, 1H), 8.19 (d, J = 9.2 Hz, 2H), 7.79 (dd, J = 7.8, 6.2 Hz, 2H), 7.40 (d, J = 9.2 Hz, 2H), 6.91 (br s, 1H), 3.85 (dt, J = 7.5, 6.0 Hz, 2H), 3.11 (q, J = 5.9 Hz, 2H), 1.33 (s, 9H); ^{13}C NMR (100 MHz, DMSO-*d$_6$*) δ: 158.47 (d, J = 6.1 Hz), 155.91, 145. 67, 142.97, 142.38, 126.27, 125.83, 120.78 (d, J = 5.4 Hz), 78.07, 65.00 (d, J = 6.3 Hz), 40.95 (d, J = 6.3 Hz), 28.57; ^{32}P NMR (162 MHz, DMSO-*d$_6$*) δ: −6.92 (s); HRMS (ESI−): Calc. for $C_{13}H_{18}N_2O_8P$ [M$^-$]: 361.08063 Found: 361.08074.

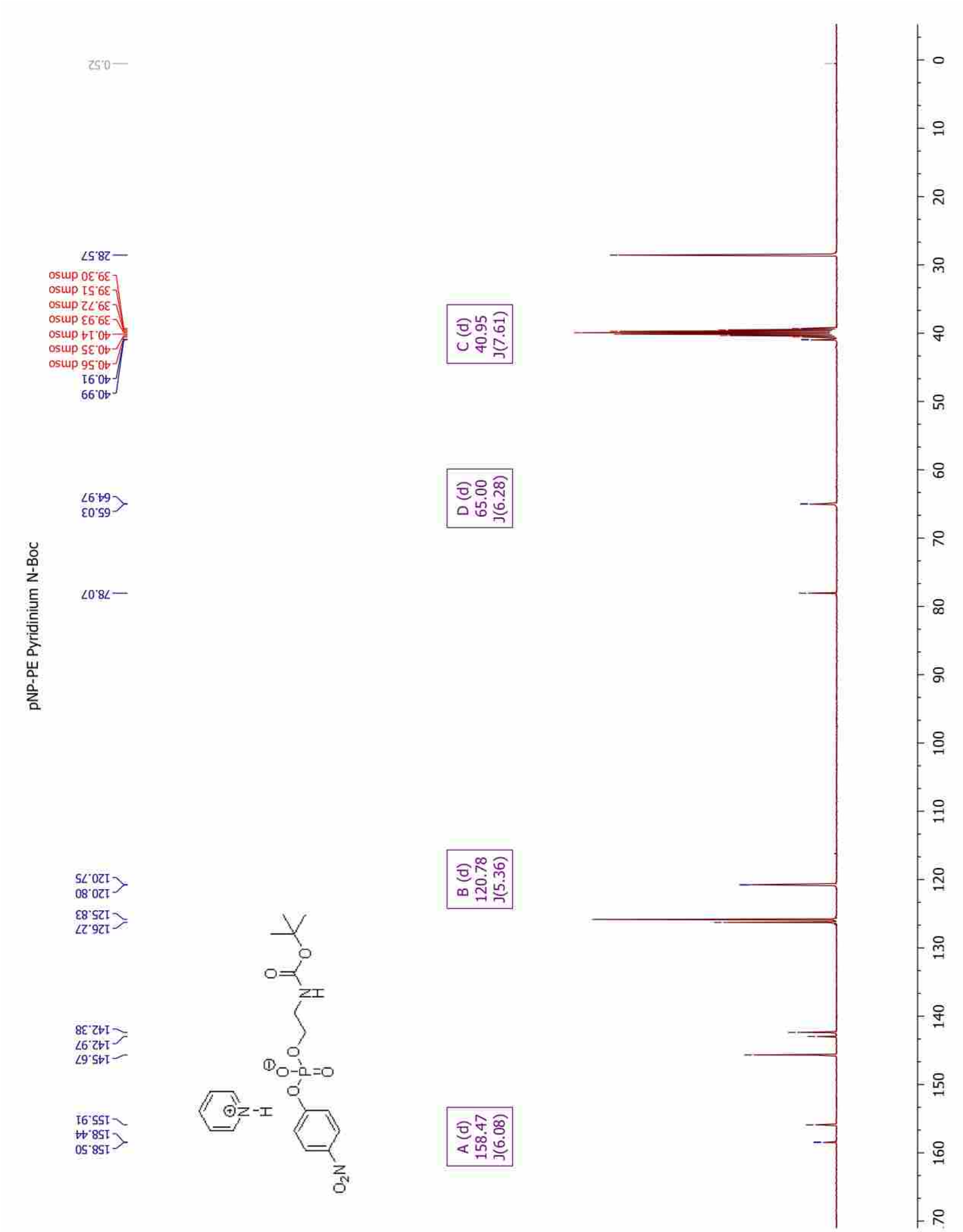
(ii) 2-aminoethyl (4-nitrophenyl) hydrogen phosphate



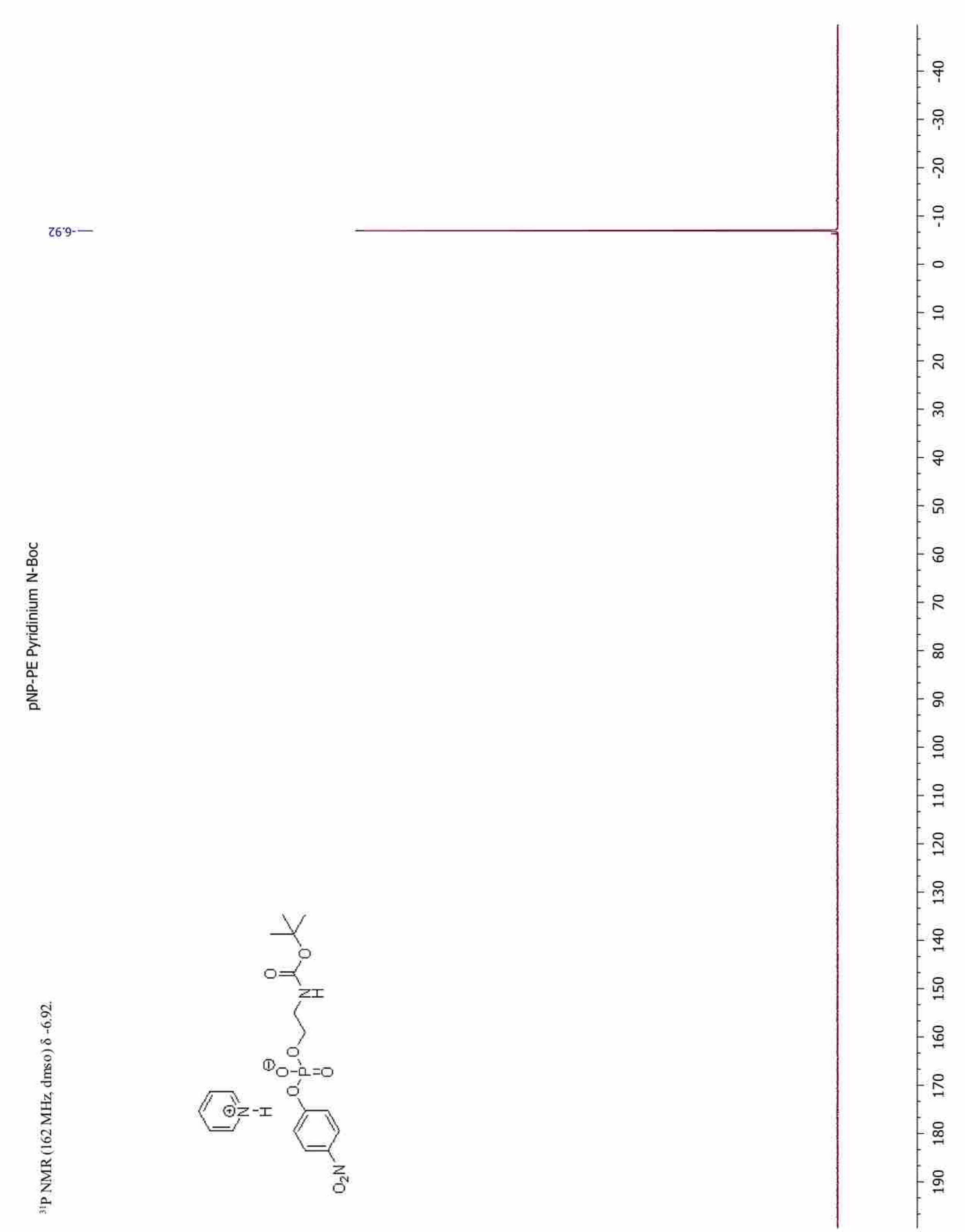
To a solution of N-BOC-ethanolamine pyridinium salt (850 mg, 1.92 mmol) in CH_2Cl_2 (10 mL) was added trifluoroacetic acid (0.8 mL) and allowed to stand 1.5 h. After this time the volatiles were removed *via* rotary evaporation (bath temperature ≈ 50 °C) yielding a clear, colorless oil. White crystals of the trifluoroacetate salt were obtained from EtOAc (7 mL) upon cooling to −15 °C overnight and collected *via* suction filtration, washing with ether (400 mg, 1.53 mmol, 79%). ^1H NMR (400 MHz, D_2O) δ: 8.21 (d, J = 8.8 Hz, 2H), 7.29 (d, J = 9.0 Hz, 2H), 4.19 (q, J = 5.7 Hz, 2H), 3.25 (t, J = 4.5 Hz, 2H); ^{13}C NMR (100 MHz, D_2O) δ: 157.02 (d, J_{C-P} = 6.3 Hz), 143.56, 125.81, 120.31 (d, J = 5.0 Hz), 62.54 (d, J = 5.4 Hz), 39.79 (d, J = 8.2 Hz); ^{32}P NMR (121 MHz, DMSO-d_6) δ: −6.71; Found: C, 36.48; H, 4.22; N, 10.56; P, 11.61. $C_8H_{11}N_2O_6P$ requires C, 36.65; H, 4.23; N, 10.69; P, 11.81.



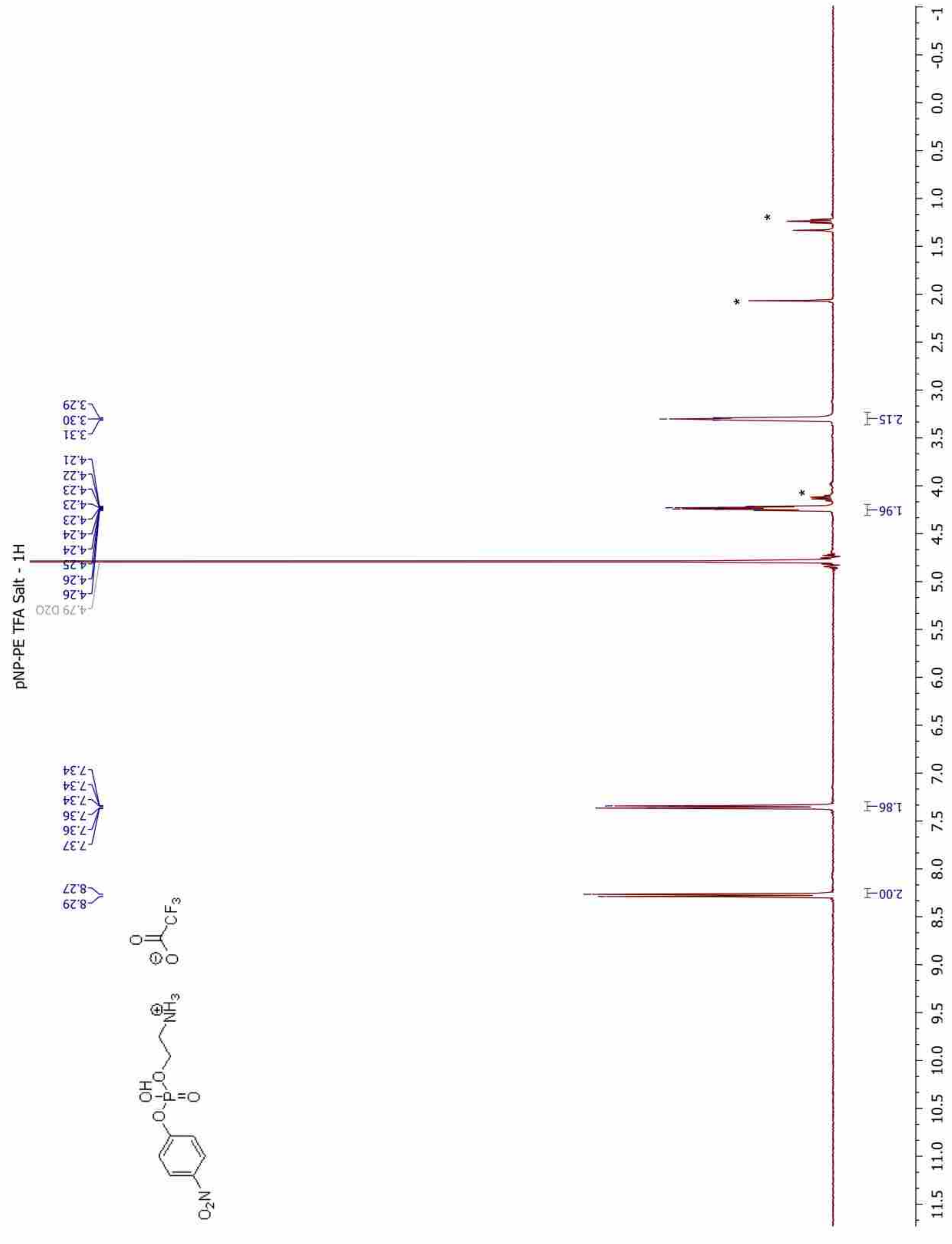
Appendix Figure A1: Annotated ^{1}H NMR spectra for the *para*-nitrophenyl-N-BOC-phosphorylethanolamine pyridinium salt synthetic intermediate.



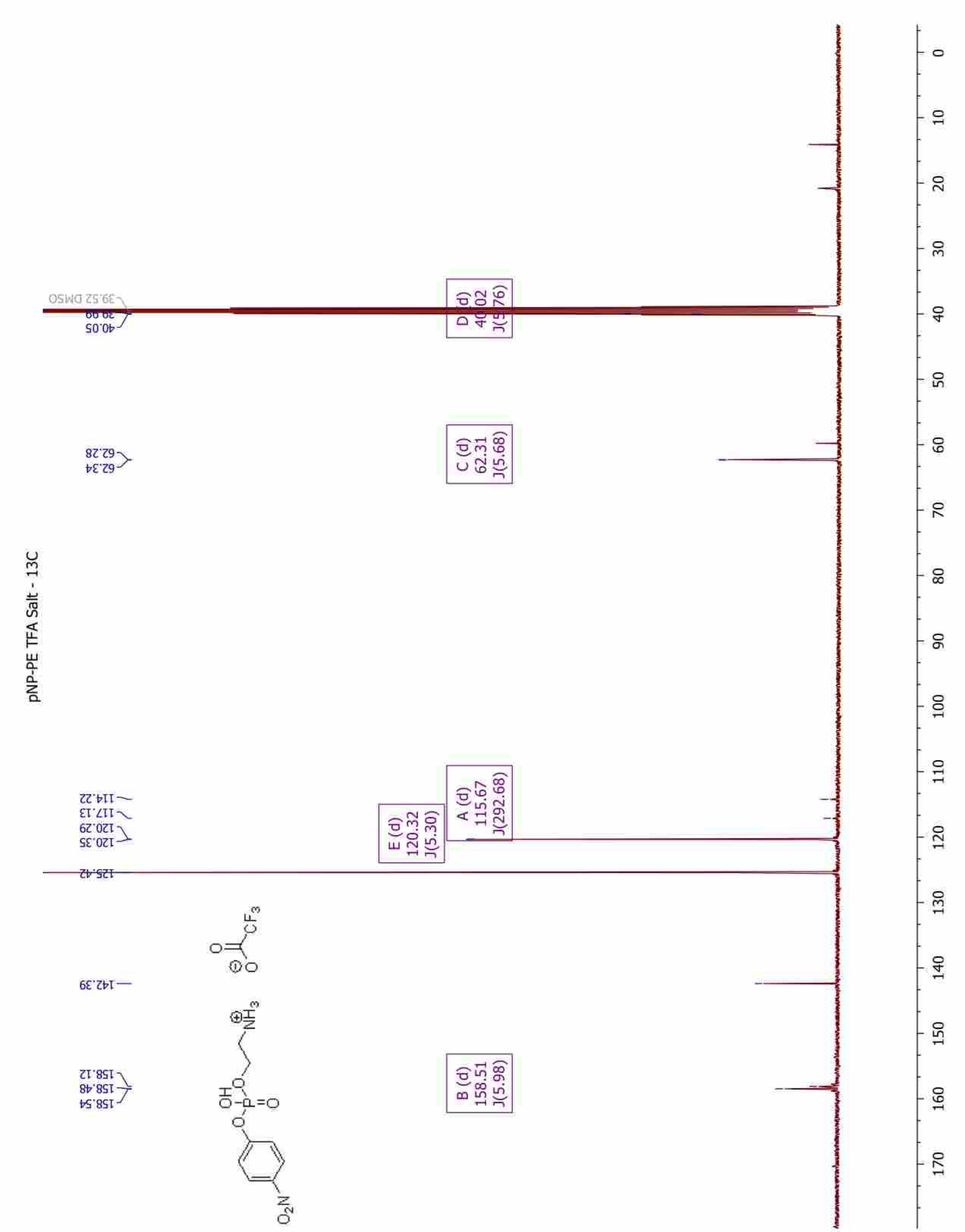
Appendix Figure A2: Annotated ^{13}C NMR spectra for the *para*-nitrophenyl-N-BOC-phosphorylethanolamine pyridinium salt synthetic intermediate.



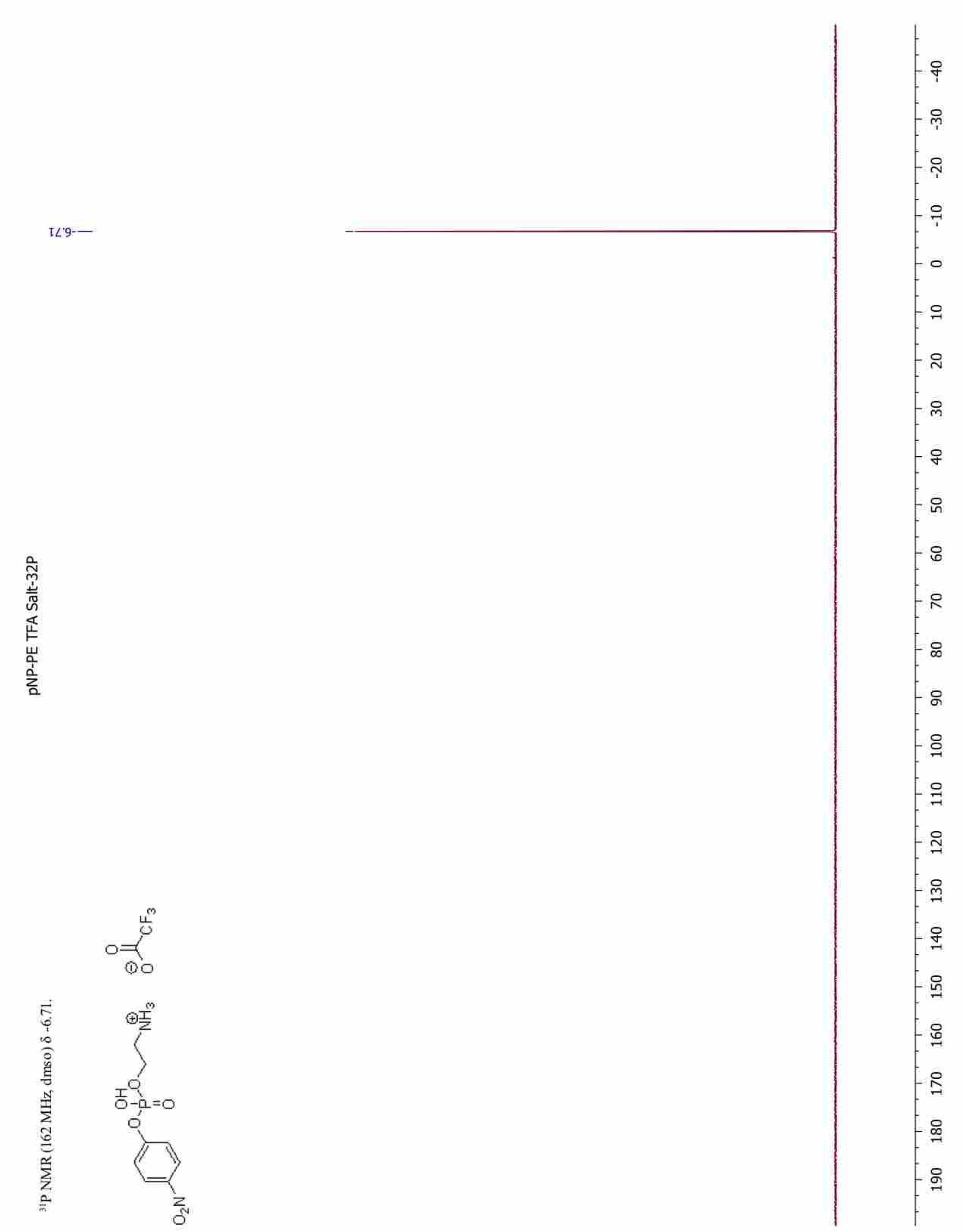
Appendix Figure A3: Annotated ^{31}P NMR spectra for the *para*-nitrophenyl-N-BOC-phosphorylethanolamine pyridinium salt synthetic intermediate.



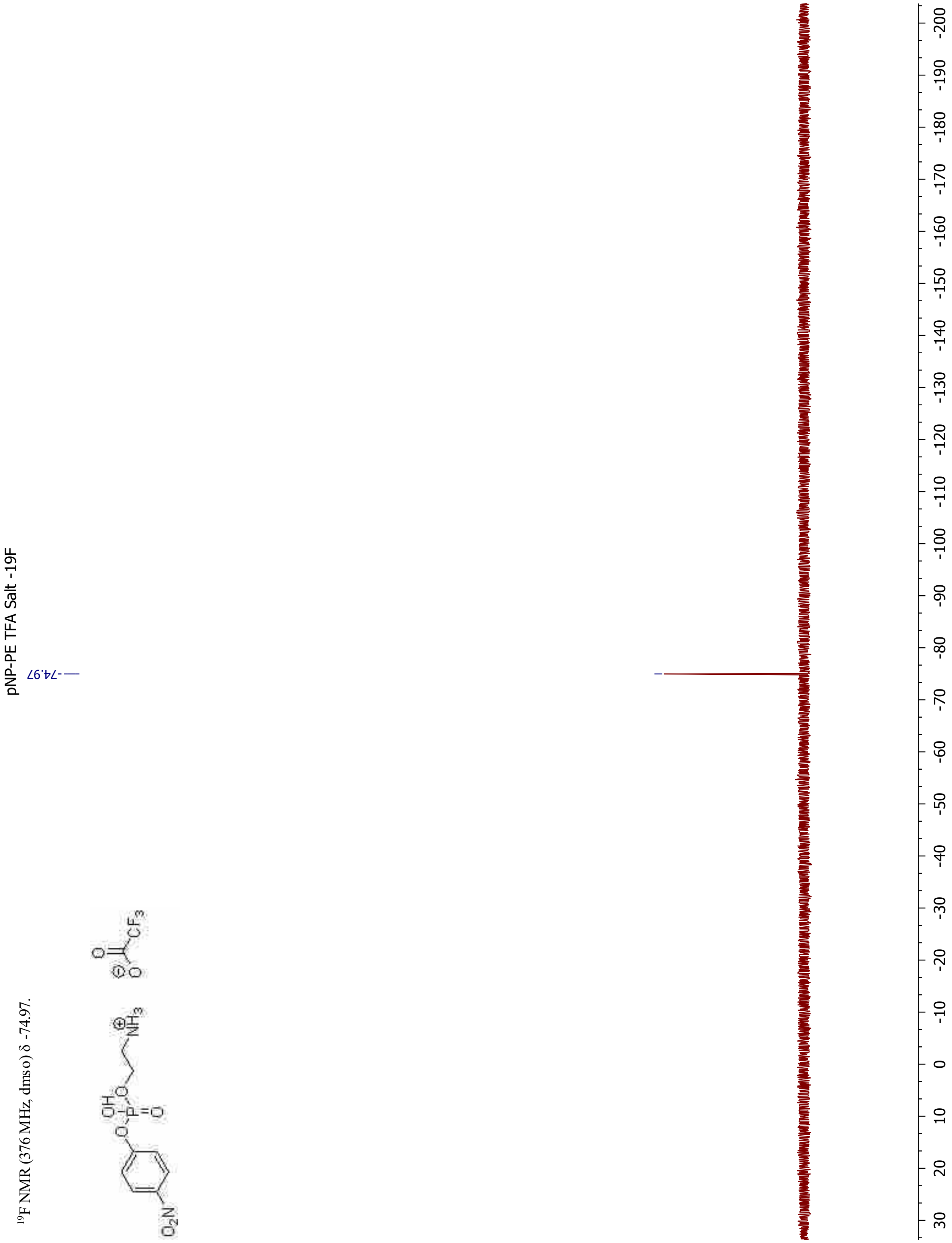
Appendix Figure A4: Annotated ^{1}H NMR spectra for the *para*-nitrophenyl phosphorylethanolamine trifluoroacetate salt. Asterisks denote resident ethyl acetate sovlent detected that could not be assigned to the synthetic product.



Appendix Figure A5: Annotated ^{13}C NMR spectra for the *para*-nitrophenyl phosphorylethanolamine trifluoroacetate salt.



Appendix Figure A6: Annotated ^{31}P NMR spectra for the *para*-nitrophenyl phosphorylethanolamine trifluoroacetate salt.



Appendix Figure A7: Annotated ^{19}F NMR spectra for the *para*-nitrophenyl phosphorylethanolamine trifluoroacetate salt.

```
EcBcsG    1 MTQ----FTQ----NTAMPSSLWQYWRGLSGWNFYFLVKFGLLWAGYL------NFHPL-----LNLVFAAFLLMP  57
CjEptC    1 --MLRLTWFQFTFFNSLM--------IVLLNFNLFYFVYEK-----N-TQNWLITFVFIVAYFALVHVICSLLFI-  59
NmEptA    1 -MIKPN--LRPKLGSSALIAFLSLYSSLVLNYAFFAKVVELHPFNGTGADIFLYT-MPVVLFFLSNFVF-HVIAL-  70
MCR-1     1 MMQHTSVWYRRSVSPFVLVASVAVFLTATANLTFFDKISQTYPIADNL--GFVLT-IAVVLFGAMLLITTLLSSY-  72

EcBcsG   58 LPRYSLHRLRHWIALPIGFALF-WHDTWLPGPESIMSQ-GSQVAGFSTDYLIDLVTRFINWQMI-GAIFVLLVAWL 130
CjEptC   60 --KFFTKFF-SILFIISSFLSVYFISFYGVLIDSDMMQNVVQTDIKEVKDL--LNLKLILFVVL-ALLLVFYVVKV 129
NmEptA   71 --PFVHKVL-IPLILVISAAVSYQEIFFNIYFNKSMLNNVLQTTAAESARL--ITPGYVLWIVCLGVLPALAYIAV 141
MCR-1    73 --RYVLKPV-LILLLIMGAVTSYFTDTYGTVYDTTMLQNALQTDQAETKDL--LNAAFIMRIIGLGVLPSLLVAFV 143

EcBcsG  131 F--LSQW------IRITVFVVAILLWLNVL------------------TLAGPSFSLWPAGQPT-----------T 169
CjEptC  130 KIDYYGSFKSHIKIKIINIISGLIVVCAVLIPLSKTFLPFFRNYNEIRMYNTPFYQIYAVVRY-YVRFVKAKPEFK 204
NmEptA  142 KVKYRVWYKELLT-RLVLAAVSFLCALGIAMLQYQDYASFFRNNKSVTHLIVPSNFI-GAGVSKYKDWKRSNIPYT 215
MCR-1   144 KVDYPTWGKGLMR-RLGLIVASLALILLPVVAFSSHYASFFRVHKPLRSYVNPIMPIYSVGKLASIEYKKASAPKD 218

EcBcsG  170 TVTTTGGNAAATVAATGGAPVVGDMPAQTAPPTTANLNAWLNNFYNAEAKRKSTFPSSLPADAQPFELLVINICSL 245
CjEptC  205 T-----IANDAYRENNHTKKLLVLVVGETARAANYSLGGYTKNDTNFYTK-KDN---------------------- 252
NmEptA  216 ---QLDMAVVQNRPAGSLRRFVVLVVGETTRAANWGLNGYSRQTTPLLAARGDE---------------------- 266
MCR-1   219 TIYHAKDAVQATKPDMRKPRLVVFVVGETARADHVSFNGYERDTFPQLAK-IDG---------------------- 271

EcBcsG  246 SWSDIEAAGLMSHPLWSHFDIEFKNF---NSATSYSGPAAIRLLRASCGQTSHTNLYQPANNDCYLFDNLSKLGFT 318
CjEptC  253 -------------------VVFFDNFSSCGTATAVSLPCMFSISKRENYS--------SSEFQENAMDVLYKTGVD 301
NmEptA  267 -------------------IVNFPQVRSCGTSTAHSLPCMFSTFDRTDYDE------IKAEHQDNLLDIVQRAGVE 317
MCR-1   272 -------------------VTNFSNVTSCGTSTAYSVPCMFSYLGADEYDV------DTAKYQENVLDTLDRLGVS 322

EcBcsG  319 QHLMMGHNGQFGGFLKEVRENGGMQSE--------------------------------LMDQTNLPVILLGFDGS 362
CjEptC  302 AVWFDNNSGGCKGVCDRLAYKQKLSSD----------------LDENLLAPFKEK-LNHLSDQNIIVLHLQGSHGP 360
NmEptA  318 VTWLENDS-GCKGVCGKVPNTDVTSLNLP---EYC---RNGECLDNILLTKFDEV-LNKNDKDAVLILHTIGSHGP 385
MCR-1   323 ILWRDNNS-DSKGVMDKLPKAQFADYKSATNNAICNTNPYNECRDVGMLVGLDDFVAANNGKDMLIMLHQMGNHGP 397

EcBcsG  363 PVYDDTAVLNRWLDVTEKDKNSRSATFYNTLPLHDGNHYPGVSKTAD---YKA---RAQKFFDELDAFFTELEKSG 432
CjEptC  361 TYYKRYP-----------------SEFKKFTPTCDTNELSKCDSEALINTYDNTLLYTDYLLSEIIKLL-KEQK-S 417
NmEptA  386 TYYERYT-----------------EAERKFTPTCDTNEINKCTRATLVNTYDNTVLYVDQFIDKVIRKL-ENRD-D 442
MCR-1   398 AYFKRYD-----------------EKFAKFTPVCEGNELAKCEHQSLINAYDNALLATDDFIAQSIQWL-QTHSNA 455

EcBcsG  433 RKVMVVVVPEHGGALKGDRMQVSGLRDIPSPSITDVPVGVKFFGMKAPHQGAPIVIEQPSSFLAISDLVVRVLDGK 508
CjEptC  418 YESSLFYLSDHGESLGENGIYLHGMPYAIAPSYQTHIPAIFWSNDKNLMNL-AKE------Y-----------KGL 475
NmEptA  443 LESVVHYVSDHGESLGENGMYLHAAPYAIAPSGQTHIPMVMWFSKAFRQHG-GID------FQCLK----QKAAEN 507
MCR-1   456 YDVSMLYVSDHGESLGENGVYLHGMPNAFAPKEQRSVPAFFWTDKQTGITP--------------------MATDT 511

EcBcsG  509 IFTEDNVDWKKLTSGLPQTAPVSENSNAVVIQYQDKPYYVRLNGGDWVPYPQ 559
CjEptC  476 KLSQDNL-----FSTLLGYFDVKTS---V---YE-PEY-DLLNPKLKANP- 512
NmEptA  508 EYSHDHY-----FSTVLGLMDISNS---Q--TYR-KEM-DILAACRRPR-- 544
MCR-1   512 VLTHDAI-----TPTLLKLFDVTAD---K--VKD-RTA-FIR--------- 541
```

Appendix Figure A8: Complete sequence alignment of *Ec*BcsG, *Cj*EptC, *Nm*EptA, and MCR-1 using ClustalOmega. Important active site residues are highlighted in red and alignment gaps are indicated with a hyphen.