

Wilfrid Laurier University

Scholars Commons @ Laurier

Theses and Dissertations (Comprehensive)

2017

Structure-Function Investigation of Proteins Involved in Cellulose Biosynthesis by *Escherichia coli*

Thomas Brenner
bren0690@mylaurier.ca

Follow this and additional works at: <https://scholars.wlu.ca/etd>



Part of the [Amino Acids, Peptides, and Proteins Commons](#), [Bacteria Commons](#), [Bacteriology Commons](#), [Bioinformatics Commons](#), and the [Biotechnology Commons](#)

Recommended Citation

Brenner, Thomas, "Structure-Function Investigation of Proteins Involved in Cellulose Biosynthesis by *Escherichia coli*" (2017). *Theses and Dissertations (Comprehensive)*. 1991.
<https://scholars.wlu.ca/etd/1991>

This Thesis is brought to you for free and open access by Scholars Commons @ Laurier. It has been accepted for inclusion in Theses and Dissertations (Comprehensive) by an authorized administrator of Scholars Commons @ Laurier. For more information, please contact scholarscommons@wlu.ca.

**Structure-Function Investigation of Proteins Involved in Cellulose Biosynthesis by
*Escherichia coli***

by

Thomas Brenner

Bachelor of Science Honours Biology, Wilfrid Laurier University, 2015

THESIS

Submitted to the Department of Biology

Faculty of Science

In partial fulfillment of the requirements for the

Master of Science in Integrative Biology

Wilfrid Laurier University

2017

Thomas Brenner 2017 ©

Acknowledgements:

I would first like to thank Dr. Joel Weadge for catalyzing my passion for scientific research.

Thank you for the opportunity to undertake this research project, for your perpetual support and guidance along the way, and for all of the time and energy that you have devoted to getting me to where I am today.

I would also like to thank my committee members, Dr. Stephanie Dewitte-Orr and Dr. Michael Suits for their invaluable guidance and feedback during this project.

A special thank you goes out to all of my lab mates and everyone from the fourth floor of the Science Research Building for your help, your input, some laughs and of course the occasional delicious baking that helps get us all through the long hours of research!

I would also like to acknowledge NSERC (Natural Sciences and Engineering Research Council of Canada) for their funding and making this project possible.

Abstract:

Bacteria thrive within multicellular communities called biofilms consisting of a self-produced matrix. Biofilm matrices improve bacterial adherence to surfaces while creating a barrier from host immune responses, disinfectants, antibiotics and other environmental factors. Persistent colonization by the widely distributed pathogens, *Escherichia coli* and *Salmonella spp.*, has been linked to production of biofilms composed of the exopolysaccharide cellulose. Cellulose-containing biofilms are also important to *Acetobacter*, *Sarcina*, *Rhizobium* and *Agrobacterium* species to form symbiotic and pathogenic interactions. In Enterobacteriaceae, two operons (*bcsABZC* and *bcsEFG*) are proposed to encode for proteins that form a cellulose biosynthetic complex that spans the bacterial cell wall. Using established recombinant DNA techniques, crystallography and functional assays, the overarching objectives of this research included the investigation of the structures and functions of the BcsE and BcsG proteins to gain insight into how they contribute to the Bcs system. The cytoplasmic protein BcsE has been shown via knockout studies to be required for optimal cellulose biosynthesis and recently the C-terminus was proven to be the second protein domain, after PilZ, dedicated to c-di-GMP binding in Enterobacteriaceae. The N-terminal structure and function of BcsE are still uncharacterized. One hypothesis for the function of BcsG is that it is involved in the labelling of cellulose with phosphoethanolamine (PEA) during export due to its homology to other characterized proteins. For example, the external modification with PEA is a strategy that allows organisms like *Neisseria gonorrhoeae* to evade components of the host immune response. However, the structure, cellular localization and specific mechanism of action of BcsG are yet unknown. To gain insight into the hypothetical properties of BcsE and BcsG, bioinformatics analyses were first conducted. The following research focused on the structure-function characterization of these proteins using recombinant truncated constructs for hypothetical N- and C-terminal domains. While practical quantities of BcsE constructs could be expressed and purified, these constructs proved challenging to isolate in sufficient purity and concentration for structural analyses. High yields of the C-terminal, soluble BcsG construct (amino acids 164 – 559) were ideal for structural and functional analyses. Malachite green-based colorimetric phosphate detection assays supported the bioinformatics analyses that the soluble C-terminus of BcsG has phosphatase activity in the presence of ATP, GTP, CTP and PEA. Metal dependency and pH tests showed that optimal BcsG activity occurs at pH 7.5 with a magnesium cofactor ($2.03 \times 10^{-1} \pm 0.008$ nmol/mg/min) which supports bioinformatics predictions. Using a BcsG¹⁻⁵⁵⁹-GFP hybrid, the localization of the soluble C-terminus of BcsG was shown to reside in the periplasm of *E. coli*. This localization aligns with bioinformatics analyses and would give BcsG a logical vantage point for cellulose modification during export from the cell. Numerous crystallization screens were attempted for BcsE and BcsG constructs. High quality BcsG¹⁶⁴⁻⁵⁵⁹ native protein crystals were achieved with resolutions as sharp as 2.1 Å as measured by X-ray analysis at the Canadian Light Source. Experimental phasing with heavy metal soaking and selenomethionine labeling techniques were attempted in search of missing phase information for BcsG¹⁶⁴⁻⁵⁵⁹. These techniques have shown promise; however, experiments are ongoing. Future studies with BcsG should continue phasing experiments, test more substrates from the PEA metabolism pathway and attempt active site characterization. Future BcsE research should focus on N-terminal functional investigations and structural experiments for the N- and C-termini.

List of Abbreviations

% i.d.: percentage identity (homology at amino acid level between query sequence and template)

ATP: Adenosine triphosphate

C-di-GMP: Bis-(3'-5')-cyclic dimeric guanosine monophosphate

C-terminal: Carboxyl terminal

CTP: Cytidine triphosphate

EPS: Exopolysaccharide

FPLC: Fast protein liquid chromatography

GTP: Guanosine-5'-triphosphate

His-tag: Hexahistidine affinity tag

IMAC: Immobilized metal ion affinity chromatography

IPTG: Isopropyl β -D-1-thiogalactopyranoside

kDa: Kilodalton

LB: Luria Broth

Ni-NTA: Nickel-nitrilotriacetic acid chelate

N-terminal: Amino-terminal

OD₆₀₀: Optical density at 600nm wavelength

PEA: phosphoethanolamine

Phyre2: Protein homology/analogy recognition engine, version 2

RFU: relative fluorescence units

SB: Super broth media

SDS-PAGE: Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

TPR: Tetratricopeptide repeat

Tris: Tris(hydroxymethyl)aminomethane

UDP: Uridine diphosphate

TABLE OF CONTENTS

Acknowledgements:.....	i
Abstract:	ii
List of Abbreviations	iii
List of Tables	vii
List of Figures.....	viii
1. INTRODUCTION	1
1.1 Biofilms.....	1
1.2 The Bacterial Cellulose Synthase (Bcs) Complex	3
1.3 Cyclic Dimeric (3 → 5) Guanosine Monophosphate.....	7
1.4 Alkaline Phosphatase Superfamily.....	9
1.5 Phosphoethanolamine.....	10
2. RESEARCH NEED	13
3. OBJECTIVES	15
4. MATERIALS AND METHODS	16
4.1 Bioinformatics and Cloning	16
4.2 Transformation and Expression	17
4.3 Purification	18
4.3.1 Immobilized Metal Affinity Chromatography (IMAC)	18
4.3.2 SDS-PAGE and Western Blot Analysis	20
4.3.3 Dialysis and Buffer Exchange	21
4.3.4 Ion Exchange and Size Exclusion Chromatography (SEC)	22
4.4 Functional Analysis of BcsG	23

4.4.1 Investigating Phosphatase Activity	23
4.4.2 Localization of BcsG ¹⁻⁵⁵⁹ Using GFP as a Reporter Protein	25
4.5 Structural Analysis of BcsE and BcsG.....	26
4.5.1 Crystallization Screening.....	26
4.5.2 X-Ray Diffraction.....	29
4.5.3 Multiple Isomorphous Replacement.....	29
4.5.4 Selenomethionine Crystallization.....	30
5. RESULTS	31
5.1 Bioinformatics Analyses	31
5.1.1 BcsE Constructs Bioinformatics Analysis.....	32
5.1.2 BcsG Bioinformatics Analyses.....	39
5.1.3 BcsG SDM Constructs Based on Bioinformatics Analysis.....	43
5.2 Protein Expression and Purification	44
5.2.1 Protein Expression.....	44
5.2.2 Immobilized Metal Affinity Chromatography	45
5.2.3 Ion Exchange Chromatography	47
5.2.4 Size Exclusion Chromatography	49
5.3 Functional Analysis of BcsG ¹⁶⁴⁻⁵⁵⁹	50
5.3.1 Investigating Phosphatase Activity	50
5.3.2 Localization of BcsG ¹⁻⁵⁵⁹ Using GFP as a Reporter Protein	55
5.4 Protein Crystallization.....	58
5.4.1 Native Crystallization of BcsG ¹⁶⁴⁻⁵⁵⁹	58
5.4.2 Crystal Conditions Expansion Plates.....	61

5.4.3 Heavy Metal BcsG ¹⁶⁴⁻⁵⁵⁹ Crystal Soaks.....	65
5.4.4 Selenomethionine Expression, Purification and Crystallization Attempts.....	66
6. DISCUSSION.....	70
6.1 Objective 1: Bioinformatics.....	70
6.2 Objective 2: Protein Expression and Purification	70
6.2.1 Protein Expression.....	70
6.2.2 Protein Purification.....	71
6.3 Objective 3: Functional Analysis of BcsE and BcsG¹⁶⁴⁻⁵⁵⁹	72
6.3.1 Investigating Phosphatase Activity of BcsG ¹⁶⁴⁻⁵⁵⁹	72
6.3.2 Localization of BcsG ¹⁻⁵⁵⁹ Using GFP as a Reporter Protein	78
6.4 Objective 4: Protein Crystallization	80
6.4.1 Native Crystallization of BcsG ¹⁶⁴⁻⁵⁵⁹	80
6.4.2 Heavy Metal BcsG ¹⁶⁴⁻⁵⁵⁹ Crystal Soaks.....	84
6.4.3 Selenomethionine Crystallization.....	85
6.5 Summary and Significance	86
6.6 The Integrative Nature of this Research.....	89
7. REFERENCES	91
8. APPENDIX.....	102

List of Tables

Table 1. BcsE and BcsG protein constructs generated from bioinformatics	17
Table 2. Summary flow chart of phosphate assay (from Abcam protocol booklet)	24
Table 3. SeMet Crystal Expansion Plate Layouts.....	30
Table 4. ProtParam protein characteristics for BcsE constructs	33
Table 5. ProtParam protein characteristics for BcsG constructs.....	40
Table 6. Phyre2 top five homology suggestions for BcsG ¹⁻⁵⁵⁹	40
Table 7. TMPred predicted transmembrane regions for BcsG ¹⁻⁵⁵⁹	41
Table 8. Site-directed mutants for BcsG ¹⁶⁴⁻⁵⁵⁹	43
Table 9. Purification summary.....	50
Table 10. Substrate profiling of the metal dependency of BcsG	53
Table 11. Diffraction statistics for native BcsG ¹⁶⁴⁻⁵⁵⁹ protein crystals*	62
Table 12. A rough guide to the resolution of protein structures	83

List of Figures

Figure 1. Components of a biofilm.....	2
Figure 2. Diversity of the bacterial cellulose synthase (bcs) operons.....	3
Figure 3. The Bcs proteins spanning the cell membrane	4
Figure 4. Phosphoethanolamine and phosphocholine metabolisms	10
Figure 5. PEA-decorated and PEA-deficient lipid A molecules.....	12
Figure 6. Workflow from plasmid to overexpressed recombinant protein.....	18
Figure 7. Binding between a histidine-tag and Ni-NTA resin (Qiagen Ltd.)	19
Figure 8. Principles of ion exchange chromatography from Bio-Rad product overview	23
Figure 9. Summary flow chart of phosphate assay (from Abcam protocol booklet).....	24
Figure 10. Colourimetric Phosphatase Assay Principles	25
Figure 11. Crystal Expansion Plate Layouts.....	27
Figure 12. Crystallization zone chart from Seed Bead protocol booklet	28
Figure 13. Crystal seed preparation cartoon from Seed Bead protocol booklet	28
Figure 14. BcsE ¹⁻²¹⁵ N-terminal construct Phyre2 homology results	35
Figure 15. BcsE ²²⁴⁻⁵²³ C-terminal construct Phyre2 homology results.....	36
Figure 16. BcsE ¹⁻⁵²³ Clustal Omega alignment with top five suggested Phyre2 homologs	37
Figure 17. 3DLigand BcsE tertiary structure model and predicted binding	38
Figure 18. BcsG ¹⁶⁴⁻⁵⁵⁹ Clustal Omega alignment with top five suggested Phyre2 homologs	41
Figure 19. 3DLigand BcsG tertiary structure model and predicted binding.....	42
Figure 20. SDS-PAGE showing BcsG ¹⁶⁴⁻⁵⁵⁹ expression for 1 L <i>E. coli</i> BL21 cultures	45
Figure 21. SDS-PAGE IMAC Purification of BcsG ¹⁶⁴⁻⁵⁵⁹	46
Figure 22. SDS-PAGE IMAC Purification of BcsE ¹⁻⁵²³	46

Figure 23. Anion Exchange Purification Profile for BcsE ¹⁻⁵²³	48
Figure 24. SDS-PAGE Anion Exchange Purification of BcsE ¹⁻⁵²³	48
Figure 25. BcsE ¹⁻⁵²³ Size Exclusion Chromatography (SEC) SDS-PAGE gel analysis.....	49
Figure 26. Functional assay of BcsG ¹⁶⁴⁻⁵⁵⁹ with ATP, GTP and CTP substrates	52
Figure 27. Functional assay of BcsG ¹⁶⁴⁻⁵⁵⁹ with CTP substrate and EDTA inhibitor	53
Figure 28. pH profile for BcsG ¹⁶⁴⁻⁵⁵⁹ with CTP in various biological buffers	54
Figure 29. Functional assay of BcsG ¹⁶⁴⁻⁵⁵⁹ with CTP and PEA substrates.....	54
Figure 30. SDS-PAGE gel for GFP and BcsG ¹⁶⁴⁻⁵⁵⁹ -GFP hybrid cleared lysate samples	56
Figure 31. Growth Plates for GFP (+) Control and BcsG-GFP Hybrid	56
Figure 32. Excitation-emission for BcsG ¹⁶⁴⁻⁵⁵⁹ cleared lysate samples.....	57
Figure 33. Tris buffer purified BcsG ¹⁶⁴⁻⁵⁵⁹ crystal hits from the MCSG-4T F7 condition	59
Figure 34. Tris buffer purified BcsG ¹⁶⁴⁻⁵⁵⁹ crystal hits from the MCSG-4T G11 condition	59
Figure 35. Tris buffer purified BcsG ¹⁶⁴⁻⁵⁵⁹ crystal hits from the MCSG-1T D1 condition	60
Figure 36. Tris buffer purified BcsG ¹⁶⁴⁻⁵⁵⁹ crystal hits from the MCSG-3T A5 condition	60
Figure 37. Representative large crystal isoforms from expansion trials.....	63
Figure 38. Representative native BcsG ¹⁶⁴⁻⁵⁵⁹ crystal X-ray diffraction pattern	64
Figure 39. Crystals Soaked in Heavy Atom Solutions	65
Figure 40. SDS-PAGE expression gel for BcsG ¹⁶⁴⁻⁵⁵⁹ grown in SeMet minimal media.....	66
Figure 41. SeMet concentrated IMAC purification samples	67
Figure 42. Native BcsG ¹⁶⁴⁻⁵⁵⁹ crystals used to streak seed through first SeMet conditions	69
Figure 43. SeMet BcsG ¹⁶⁴⁻⁵⁵⁹ microcrystals from second SeMet attempt	69
Figure 44. Potential BcsG substrates from PEA metabolism pathway.....	76

1. INTRODUCTION

1.1 Biofilms

A common survival strategy for many bacteria is through the biosynthesis and export of polysaccharides to form a protective biofilm. A biofilm consists of a cluster of microorganisms fixed within a synthesized matrix of exopolysaccharides (Figure 1). Biofilms permit bacteria to stick to surfaces and provide them with protection from their environment during colonization so that they may thrive in otherwise inhospitable settings (Mah and O'Toole, 2001). Horizontal gene transfer between organisms and expression of genes conferring antimicrobial resistance are also promoted within biofilm communities (Pozo and Patel, 2007).

Bacterial cellulose is manufactured by many persistent organisms, including the widely distributed pathogens *Escherichia coli* and *Salmonella spp.* (Saldaña *et al.*, 2009) to produce their biofilms. Cellulose-containing biofilms are also important to *Acetobacter*, *Sarcina*, *Rhizobium* and *Agrobacterium* species to form symbiotic and pathogenic interactions (Augimeri *et al.*, 2015). The heterogeneous composition of biofilm affects the rates of growth for bacteria due to varying oxygen and waste product concentrations (McDougald *et al.*, 2012). Secondary elements include proteins, signalling molecules, and bacterial waste (Figure 1). Cellulose is produced by the polymerization of UDP-activated-glucose to form long chains of $\beta(1\rightarrow4)$ linked D-glucose units. As one of the most abundant biopolymers on Earth, cellulose is found in vascular plants, algae and bacteria (Zogaj, 2001). Cellulose is the focus of this study because modern ecology, medicine and the economy are largely influenced by the ubiquitous bacterial utilization of cellulose-based biofilms. Means of cellulose biosynthesis remain largely unknown.

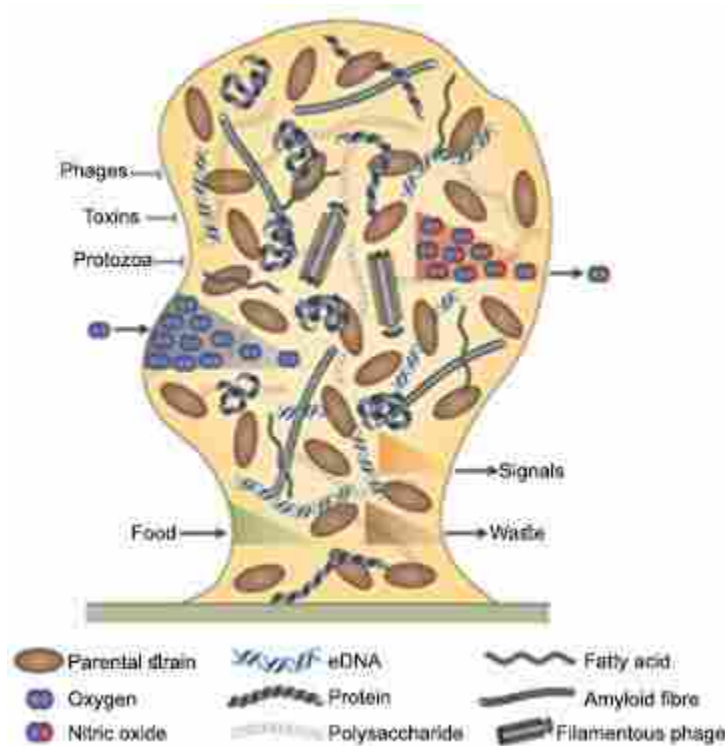


Figure 1. Components of a biofilm

Biofilms consist of cells embedded in a self-produced matrix. Major components of the matrix are polysaccharides (clear chains of circles) that can vary in chemical and physical properties. (McDougald *et al.*, 2012).

Bacterial cellulose is distinct from plant cellulose because it is synthesized as a pure polymer; free of lignins and hemicelluloses (Peterson and Gatenholm, 2011). Bacterial cellulose also has a relatively high strength, moldability and water absorbance compared to plant cellulose (Petersen and Gatenholm, 2011). Properties such as porosity, water content and shape can be manipulated without difficulty making bacterial cellulose useful as a wound dressing for severe burns, while accelerating healing and decreasing pain and probability of infection when used for dressing wounds (Petersen and Gatenholm, 2011). Bacterial cellulose also has the potential to be used as a novel material for long-term implants (Kim *et al.*, 2008). The biocompatibility, immunologically inert nature and resistance to degradation of bacterial cellulose have made it a major research topic in the modern biomedical industry (Kim *et al.*, 2008).

1.2 The Bacterial Cellulose Synthase (Bcs) Complex

In most Gram-negative bacteria, the protein complex that synthesizes and exports cellulose is referred to as the bacterial cellulose synthesis (Bcs) complex. The suspected function of the Bcs protein complex is based on its resemblances to other polysaccharide excretion systems, like the well-understood alginate system from *Pseudomonas* spp. (Whitney and Howell, 2013). In enteric bacteria, two operons (*bcsABZC* and *bcsEFG*) are proposed to encode for proteins that form a cellulose biosynthetic complex that spans the bacterial cell wall (Morgan *et al.*, 2014). Knockout studies have revealed that cellulose production relies on the *bcsABZC* and *bcsEFG* operons (Solano *et al.*, 2002); however, specific mechanisms remain unknown for the proteins encoded by the *bcsEFG* operon. To date, the *bcsABZC* operon has been found in all cellulose-synthesizing bacteria, but the *bcsEFG* operon is restricted to Enterobacteriaceae.

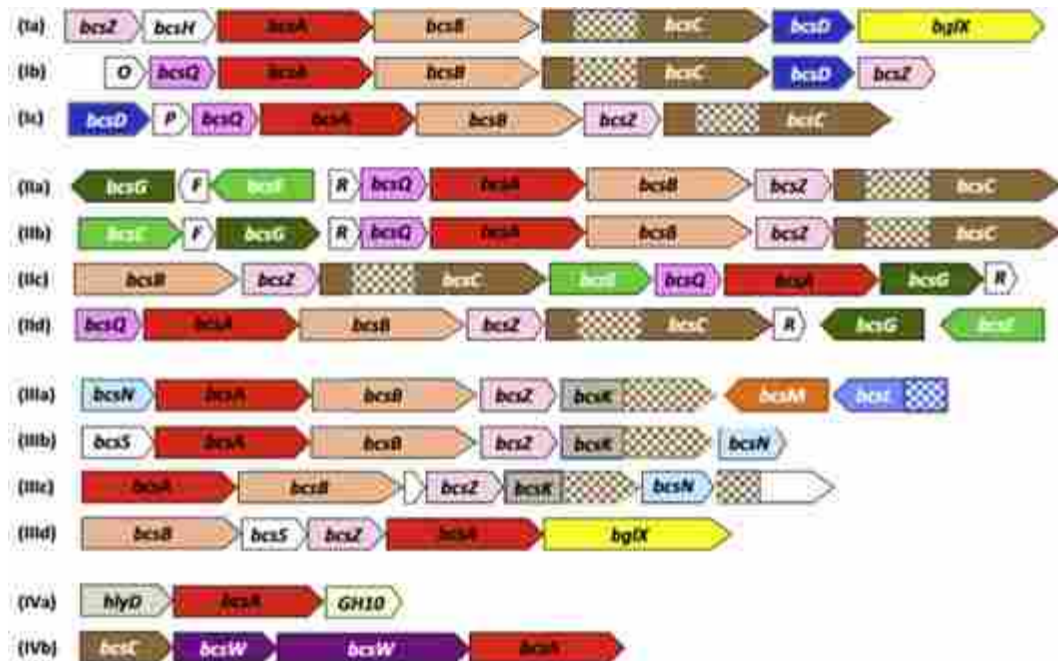


Figure 2. Diversity of the bacterial cellulose synthase (bcs) operons

Variations of Type I, II, III and IV bcs operons. BcsE (light green) and BcsG (dark green) are only present in Type II bcs operons (RömLing, 2015).

The type II, *E. coli*-like, *bcs* operon is prevalent across Enterobacteriaceae (McLaughlin *et al.*, 2017) (See Figure 2; subdivisions IIa through IIc). The distinctive characteristic of this operon is the presence of *bcsE* and *bcsG* genes and the absence of *bcsD* (Fang *et al.*, 2014). In type II *bcs* operons *bcsF* and *yhjR* are often present, as well, but can either be missing or simply unannotated due to being very short genes.

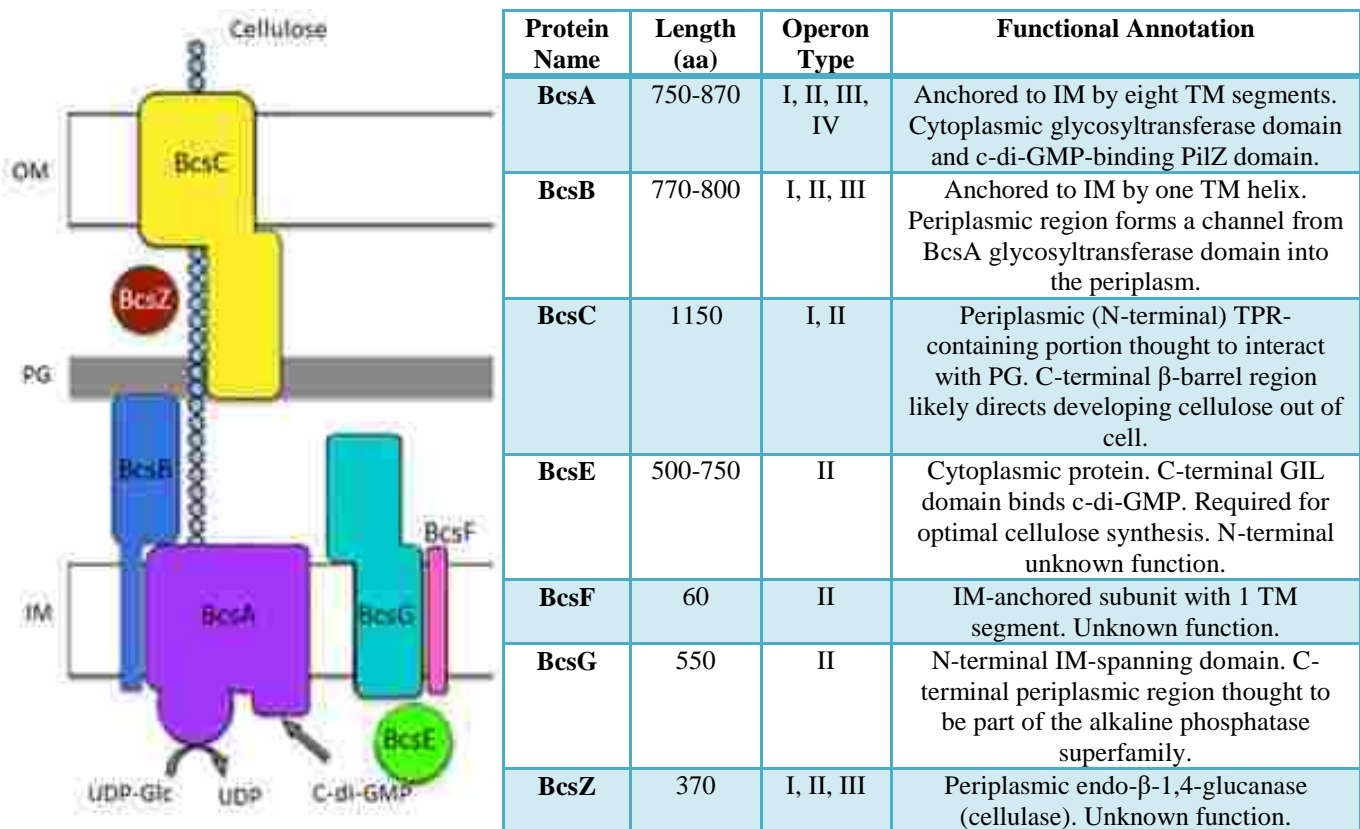


Figure 3. The Bcs proteins spanning the cell membrane

A representation of the Type II bacterial cellulose synthase and export system in Gram negative bacteria. C-di-GMP is seen binding to the cytoplasmic BcsA PilZ domain to activate the BcsA glycosyltransferase domain. BcsB is anchored in the IM and channels the nascent cellulose through the periplasm towards the N-terminal, TPR-containing region of BcsC. Cellulose is exported through the β -barrel domain of BcsC in the OM. BcsZ is likely involved in processing cellulose as it exits the cell. BcsEFG functions are not depicted; hypothetical locations are displayed. Abbreviations: IM: inner membrane, PG: peptidoglycan, OM: outer membrane. Adapted from hypothesized model in Romling and Galperin (2015).

Solano *et al.* (2002) have hypothesized that the combination of the BcsABZC and BcsEFG proteins form a larger membrane-spanning complex in enteric bacteria which would allow UDP-glucose to be polymerized and exported through the cell membrane. Crystal structures of BcsA and BcsB have provided insight into how this polymerization occurs (Omadjela *et al.*, 2013). When the PilZ domain within BcsA binds cyclic-di-GMP the glycosyltransferase catalytic site of BcsA has a conformational change which initiates UDP-glucose polymerization (cellulose production) (Morgan *et al.*, 2014). The cellulose chain is exported through a polysaccharide channel composed of eight transmembrane helices (Morgan *et al.*, 2014). The periplasmic BcsB protein is anchored to BcsA by a C-terminal transmembrane helix (Morgan *et al.*, 2014). BcsB also has two carbohydrate binding domains which may aid in cellulose translocation (Morgan *et al.*, 2013). BcsZ is a non-processive periplasmic protein that lacks carbohydrate binding domains and is responsible for hydrolyzing single cellulose chains (Mazur and Zimmer, 2011).

BcsC is thought to contain distinct outer membrane β -barrel and periplasmic tetratricopeptide repeat (TPR) domains (Figure 3), which also exist in the analogous alginate exopolysaccharide biosynthetic system (Keiski *et al.*, 2010). The TPR region of BcsC is thought to conduct protein-protein and/or protein-carbohydrate interactions that allow cellulose transport across the periplasm while the outer membrane β -barrel is hypothesized to facilitate cellulose export out of the cell (Whitney and Howell, 2013). Similarly, the homologous alginate biosynthetic system in *Pseudomonas aeruginosa* has a β -barrel protein (AlgE) and TPR protein (AlgK) (Keiski *et al.*, 2010). Together, the AlgE and AlgK proteins facilitate the export of alginate during biofilm formation (Keiski *et al.*, 2010).

This project will focus on the BcsE and BcsG proteins encoded by the largely uncharacterized *bcsEFG* operon within the model organism *E. coli* (Figure 3). The *bcsE* and *bcsG* genes are found in the type II *bcs* operons (Figure 2) and have been shown to be required for optimal cellulose biosynthesis (Solano *et al.*, 2002; RomLing and Galperin, 2015). One hypothesis for the function of BcsG is that it is involved in the labelling of cellulose with PEA during export due to its homology to other characterized proteins; likely during export of cellulose from the bacteria. Phyre2 amino acid-level homology searches have suggested highest coverage (82%) for BcsG¹⁻⁵⁵⁹ (residues 14-474) to an integral membrane protein (accession # 5fgn) lipooligosaccharide phosphoethanolamine transferase a (EptA) from *Neisseria meningitidis*. BcsG combines an N-terminal transmembrane region with a C-terminal soluble domain that predicted to reside in the periplasm, however, the location and exact mechanism of possible cellulose labelling by BcsG during biosynthesis remains to be characterized (Lynette *et al.*, personal communication). Phyre2 homology searches have also shown high coverage (77% total covering residues 86-410) for the C-terminus of BcsG with a protein from the alkaline phosphatase superfamily from *Vibrio parahaemolyticus* (accession # 3lxq) which may give some indication of its mode of operation during cellulose modification.

All BcsE amino acid homology searches show adequate homologs for the N-terminus portion, but not for the C-terminus indicating the likelihood of a novel domain. For the N-terminus of BcsE, Phyre2 indicates 67% coverage to a P-loop containing nucleoside triphosphate hydrolase from the RecA protein-like family. Briefly, RecA is an ATP-dependent enzyme responsible for the homologous recombination of broken single-stranded DNA segments by using double-stranded parent molecules as a template for repair (Kelso *et al.*, 2017). Since bioinformatics analyses indicates that BcsE¹⁻²¹⁵ matches RecA and also has the appropriate

conserved residues in these motifs to bind ADP and Mg^{2+} cations as well, BcsE may also be ATP-dependent and may even be allosterically inactivated by ADP.

In a recent study, a Differential Radial Capillary Action of Ligand Assay (DRaCALA) was utilized with fluorescently and radioisotope-labelled cyclic diguanylate (c-di-GMP) to show that a domain in the C-terminus of BcsE is a c-di-GMP receptor (Fang *et al.*, 2014). Fang *et al.* have revealed that c-di-GMP binds in the DUF2819 domain which is now designated as a GGDEF I-site like domain (GIL domain). Fang *et al.* (2014) demonstrated that cellulose production decreased for BcsE mutants with GIL domains incapable of c-di-GMP binding. This indicates that c-di-GMP binding to the GIL domain is fundamental for BcsE activity. Since the glycosyltransferase, BcsA, is also c-di-GMP-dependent (Ross *et al.*, 1987; Ryjenkov *et al.*, 2006) it is possible that Enterobacteriaceae cellulose synthesis has two levels of c-di-GMP-dependent regulation. BcsE is likely activated by c-di-GMP to coordinate its work with the work of BcsA so that they are only active at the same time. There would be no need to turn on BcsE for potential cellulose modification if BcsA is not actively synthesizing cellulose; therefore, they are linked through the second messenger.

1.3 Cyclic Dimeric (3 → 5) Guanosine Monophosphate

Cyclic -di-GMP is a universal second messenger that coordinates the signalling for numerous vital bacterial processes. In bacteria, high c-di-GMP concentrations are linked to features associated with a sessile lifestyle, including: production of pili, protein adhesins and exopolysaccharides important for biofilm formation (RömLing *et al.*, 2013). Cyclic di-GMP strongly affects bacterial survival, stress responses, cell differentiation and virulence factors (RömLing *et al.*, 2013). Production of c-di-GMP is facilitated by the GGDEF protein domains

within diguanylate cyclases (DGCs) and c-di-GMP hydrolysis is caused by either the EAL or HD-GYP domains in phosphodiesterases (PDEs) (Cruz *et al.*, 2012).

C-di-GMP can allosterically inhibit DGCs by binding the I site (amino acids: RxxD) located five amino acids away from the A site (amino acids: GGEEF or GGDEF), which is the active site responsible for binding a GTP molecule (Cruz *et al.*, 2012). If any of these sites are degenerate in an organism, they can sometimes have alternative functions as effector proteins; with or without c-di-GMP (Cruz *et al.*, 2012). Phosphodiesterases are frequent targets for pharmacological inhibition because inhibitors of PDE can prolong or enhance the effects of physiological processes mediated by cyclic AMP or cyclic GMP by inhibition of their degradation by PDE (Jeon *et al.*, 2005).

Since c-di-GMP is responsible for controlling fundamental metabolic processes within bacteria, two levels of c-di-GMP-dependent control could safeguard against inappropriate metabolic shifts to ensure a smooth shift between the motile and sessile states; especially since c-di-GMP levels are known to change rapidly in planktonic bacteria (Russell *et al.*, 2013).

Biosynthesis poly- β -1,6-*N*-acetylglucosamine (poly-GlcNAc), another common exopolysaccharide produced by Enterobacteriaceae, provides a good example of these two levels of regulation. Steiner *et al.* (2013) demonstrated that c-di-GMP activates biosynthesis of poly-GlcNAc, which is a fundamental constituent within the extracellular matrix of *E. coli* biofilms (Steiner *et al.*, 2013). Activation occurs when c-di-GMP binds directly to both PgaC and PgaD proteins simultaneously (Steiner *et al.*, 2013). When present in sufficient quantities, c-di-GMP stabilizes the PgaD protein allowing for tandem binding of proteins stimulating the glycosyltransferase activity of PgaC and PgaD on the inner membrane (Steiner *et al.*, 2013). However, low c-di-GMP levels cause PgaD protein to rapidly degrade because it fails to interact

with PgaC to form a stable complex (Steiner *et al.*, 2013). Therefore, when c-di-GMP levels drop significantly, PgaD turnover leads to the irreversible inactivation of the Pga machinery, which temporarily precludes the effects of c-di-GMP (Steiner *et al.*, 2013).

1.4 Alkaline Phosphatase Superfamily

Alkaline phosphatase (AP) enzymes occur commonly in many organisms, from bacteria to mammals (McComb *et al.*, 1979). APs are most frequently homodimers with each catalytic site containing three metal ions necessary for enzymatic activity (Millan, 2006). APs catalyze the hydrolysis of monoesters of phosphoric acid phosphate transphosphorylation reaction, as well as transphosphorylation when high concentrations of phosphate acceptors are present (Millan, 2006). The core AP catalytic structures are generally conserved between bacterial and mammalian APs, but the specific activities, K_M values, optimum pH, stability and inhibition mechanisms can differ significantly; likely reflecting diverse *in vivo* niches (Millan, 2006).

The alkaline phosphatase superfamily of enzymes share overlapping substrate specificities and catalytic properties (Millan, 2006). For example, APs show substrate specificity and structural similarity to other nucleoside pyrophosphatase and phosphodiesterases (Millan, 2006). The phosphoethanolamine (PEA) phosphatase, PHOSPHO1, is important to the PEA metabolism pathway (and this thesis in particular) and will be discussed in more detail below (Figure 4). Substrates of the PEA metabolism include: phosphatidylethanolamine, CDP-ethanolamine, phosphoethanolamine and ethanolamine (Figure 4). The phosphocholine pathway is also important as it is vital to numerous human pathogens including *Streptococcus pneumoniae* which causes diseases like sepsis, pneumonia and meningitis that primarily affect the elderly, children and the immunocompromised (Henriques-Normark and Tuomanen, 2013). Choline (Cho), produced by the phosphocholine pathway, is required for the growth of

pneumococci (Rane and Subbarow, 1940) and modification of teichoic and lipoteichoic acids with Cho is important to pneumococcal physiology, colonization and virulence (Fischer *et al.*, 1993; Kharat and Tomasz, 2006).

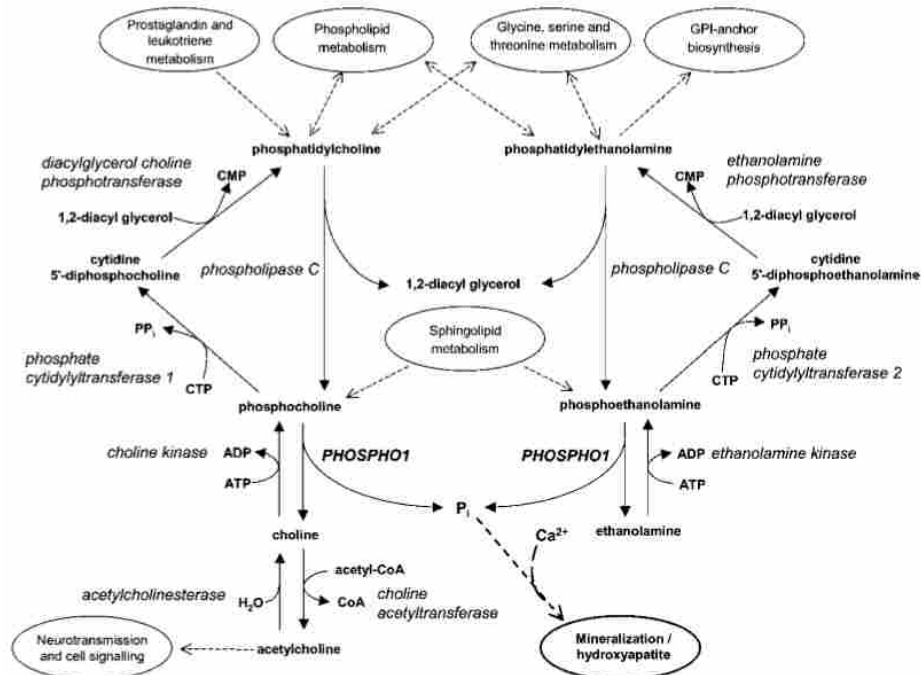


Figure 4. Phosphoethanolamine and phosphocholine metabolisms
 Diagram depicting parallels between human phosphoethanolamine and phosphocholine metabolisms from Scott *et al.*, 2004.

1.5 Phosphoethanolamine

Phosphoethanolamine plays a vitally important role on the lipopolysaccharide (LPS) of numerous Gram-negative and Gram-positive bacteria (Mackinnon *et al.*, 2002). As depicted in Figure 5, the Gram-negative LPS contains a lipid A moiety, a core oligosaccharide, and an O-antigen polysaccharide (Mackinnon *et al.*, 2002). Since lipid A triggers the innate immune response in humans, many bacteria use PEA as a decoration on their LPS to become resistant to cationic antimicrobial peptides by decreasing the overall negative charge of the bacterial surface. The Enterobacteriaceae *Salmonella enterica*, for example, modifies the lipid A region with

phosphoethanolamine in the presence of cationic antimicrobial peptides (CAMPs) (Hankins *et al.*, 2012). The supplemented amine-containing molecule on the lipid A domain prevents CAMPs from binding to the bacterial LPS (Lewis *et al.*, 2009).

Neisseria gonorrhoeae, also adds PEA to the 4' position of the lipid A moiety by the PEA transferase LptA to prevent CAMP binding (Lewis *et al.*, 2009). A 2015 study by Zughair *et al.* showed that PEA-lipid A-producing gonococci significantly decreased autophagy in macrophages leading to heightened gonococcal persistence during infection. Labeling of this lipid A also enhanced *N. gonorrhoeae* resistance to polymorphonuclear leukocytes (PMNs), likely due to a decreased ability of antimicrobial proteins, like cathepsin G, to result in lysis (Handing and Criss, 2015). Additionally, PEA decoration has been shown to decrease complement-mediated killing by C4b binding proteins in the classical pathway (Lewis *et al.*, 2013), delay fusion of azurophilic granules with phagolysosomes (Johnson and Criss, 2013), and down-regulate expression of the CAMP LL-37 in epithelial cells (Bergman *et al.*, 2005).

Gonorrhoeae have recently been listed as an urgent threat pathogen by the CDC due to a spike in antibiotic-resistant strains leading to predictions that it may soon become untreatable (Unemo and Shafer, 2014), which makes the knowledge of its ability to evade the immune response especially significant. Gonorrhoeae is a strictly human (sexually transmitted) disease with an estimated 100 million new cases arising each year as predicted by the World Health Organization (WHO, 2011). Lower genital tract infections by gonococci are often asymptomatic, but if untreated, they can eventually cause pelvic inflammatory disease in men and women, or, ectopic pregnancies, and infertility in women (Vonck *et al.*, 2011).

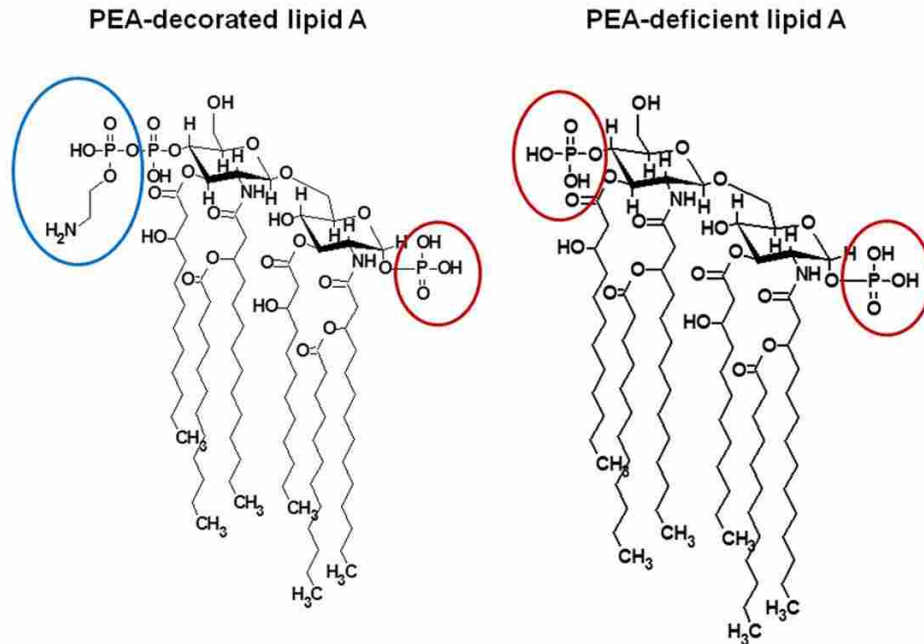


Figure 5. PEA-decorated and PEA-deficient lipid A molecules

From Zughair *et al.* (2015) showing Gc lipid A with and without PEA modification. Gc strains with this modification present have been shown to have significantly enhanced survival during human infection.

As mentioned above, the soluble C-terminal region of BcsG shares strong amino acid-level homology to an integral membrane protein lipooligosaccharide phosphoethanolamine transferase a (EptA) from *N. meningitidis*. Amino acid homology searches have also shown amino acid level homology of BcsG with a protein from the alkaline phosphatase superfamily from *Vibrio parahaemolyticus* which may give some suggestion of its currently unknown mode of operation during cellulose modification. Given these suggested *in silico* comparisons to the PEA transferase EptA and alkaline phosphatase superfamily proteins, we hypothesize that BcsG may have a hydrolyzing effect on phosphate-rich molecules leading to PEA labelling of bacterial cellulose.

2. RESEARCH NEED

The majority of observed enteric bacterial infections are propagated by biofilms, so it is critical to research bacterial exopolysaccharide biosynthesis mechanisms due to their ubiquitous appearance as a bacterial survival strategy (Fux *et al.*, 2005). Studying bacterial cellulose biosynthesis proteins specifically may lead to discoveries fundamental to exopolysaccharide biosynthesis across bacterial species. Furthermore, cellulose is the most abundant and only fully renewable biopolymer on earth (RömLing and Galperin, 2015). This research was conducted to expand upon the currently small foundation of knowledge for the largely uncharacterized BcsE and BcsG proteins which are somehow involved in bacterial cellulose biosynthesis in Enterobacteriaceae.

Bacterial cellulose is already used for tissue regeneration and as a dressing for wounds, but understanding the underlying mechanisms for bacterial cellulose modification could provide a toolkit to give cellulose unique physical and chemical properties. A deeper understanding of bacterial cellulose modification machinery could potentiate synthetic modification of bacterial cellulose for unique biopolymer generation in medicine and bioengineering. Knowledge of proteins involved in bacterial cellulose biosynthesis and modification may one day give clues for efficient energy generation, production of novel materials for a spectrum of situations and for the elimination of persistent bacterial infections.

The BcsE and BcsG proteins may offer targets for circumventing the optimal production of the biofilm, or important biofilm components, thereby aiding in controlling bacterial spread, contamination and disease progression. BcsE has low homology to previously structurally characterized proteins so its characterization could unveil novel structural or functional features that have formerly remained unknown. BcsE has recently been shown to have a c-di-GMP

binding domain on the C-terminus; recently named the GIL domain (Fang *et al.*, 2014). GIL is the second protein domain, after PilZ, shown to be dedicated to c-di-GMP binding (Fang *et al.*, 2014). It has also been shown that, in *S. enterica*, BcsE is not essential for cellulose synthesis but is required for maximal cellulose production, and that c-di-GMP binding is critical for BcsE function (Fang *et al.*, 2014). It appears that cellulose production in Enterobacteriaceae is somehow controlled by a two-tiered c-di-GMP-dependent system involving BcsE and the PilZ domain containing glycosyltransferase BcsA.

There are numerous noteworthy observable scenarios in which bacteria have used molecular modifications to their great advantage. BcsG has a hypothesized association with the PEA labelling of bacterial cellulose on the synthesized biofilm; however, the mechanism by which BcsG functions is currently unknown. PEA is known to facilitate *N. gonorrhoeae* resistance to autophagy by decreasing its susceptibility to cationic antimicrobial peptides, as well as complement-mediated and intraleukocytic killing by neutrophils (Zughaier *et al.*, 2015). *N. gonorrhoeae* is one example of a strictly human pathogen that is already largely antibiotic resistant and also utilizes PEA modification to evade the host response. Another known bacterial modification used by organisms such as pathogenic strains of *Staphylococcus aureus* is acetylation of the exopolysaccharide peptidoglycan which also actively guards against lysis by host immune systems (Scheurwater *et al.*, 2008; Moynihan and Clarke, 2011). Additionally, *Pseudomonas fluorescens* SBW25 (known to grow with a wrinkly spreader phenotype) has been shown to require its synthesized cellulose to be acetylated in order to colonize the air-liquid boundary (Spiers *et al.*, 2003). Therefore, investigating the putative cellulose modifying protein BcsG may lead to discoveries useful for battling many organisms that utilize similar modifications for host-evasion or colonization.

3. OBJECTIVES

The current study investigates the roles of BcsE and BcsG proteins within the larger transmembrane Bcs protein complex in Enterobacteriaceae. Knockout studies indicate that the *bcsEFG* operon is required for optimal bacterial cellulose synthesis and labelling with PEA which is an important virulence factor to many bacteria, including *N. gonorrhoeae* (Handing and Criss, 2015). The structure and mechanism of action for the BcsE and BcsG proteins are currently uncharacterized and the cellular localization of BcsG is also unknown. BcsG may somehow lead to PEA labelling of bacterial cellulose in the biofilm, but the details of how this occurs remains unknown. ***We hypothesize that structure-function analysis of the BcsE and BcsG proteins will help us to better understand their role within the Bcs protein complex.*** The four major experimental objectives for testing this hypothesis were:

1. Utilize bioinformatics tools to generate a foundation of hypothetical properties for both BcsE and BcsG proteins. Amino acid level homology searches will create the informational foundation for these proteins regarding potential activity, essential residues and numerous other parameters that are instrumental for establishing new protein research.
2. Optimize protein expression and purification for BcsE and BcsG constructs. Constructs that are sufficiently isolated will be evaluated by downstream functional and structural tests.
3. Analyze BcsE and BcsG activity with substrates, metals and conditions proposed by bioinformatics research to gain insight into protein functions.
4. Produce high quality BcsE and BcsG protein crystals to be analyzed by X-ray diffraction. High resolution X-ray diffraction patterns are used to determine 3D protein models which are invaluable for observing the form and function of proteins.

4. MATERIALS AND METHODS

4.1 Bioinformatics and Cloning

At the onset of this project, bioinformatics searches for whole BcsE and BcsG proteins were conducted using open source online tools. The Phyre2 server was used to align the amino acid sequence of proteins with previously solved models. Phyre2 results provided hypothetical functions, secondary and tertiary structures. Protein template matches can be useful for hypothesizing potential functions and binding sites based on percent identity with already solved proteins. The 3DLigandSite then used top-ranking protein models to predict potential ligand binding sites. TMPredict was useful for returning hypothetical transmembrane regions.

ProtParam was also utilized in order to acquire theoretical protein instability, isoelectric points, extinction coefficients, transmembrane segments and cellular localizations. These theoretical metrics offered a starting point for researching these proteins which were largely uncharacterized. Pfam and PSIPRED were also used to find protein domains and secondary structure predictions, respectively. Finally, ClustalW is another useful alignment program that suggested specific regions of interest within constructs that may have affected protein folding, stability and/or function. All of these tools used in combination offered valuable information for each construct and gave the new research a starting point. Bioinformatics offer a theoretical foundation for uncharacterized proteins as a starting point for physical testing.

All BcsE and BcsG constructs were created based on above bioinformatics (Table 1) and generated by preceding lab members with *E. coli* BL21 DNA as a template. Previous lab members conducted polymerase chain reaction (PCR) to amplify and ligate target products into the pET28a expression vector.

Table 1. BcsE and BcsG protein constructs generated from bioinformatics

Protein	Residues	6xHis Tag	Truncation
BcsE	1-523	N-terminal	Full protein
	1-215	N-terminal	N-terminal (ATPase-domain)
	224-523	C-terminal	C-terminal (GIL-domain)
BcsG	1-559	N-terminal	Full protein
	164-559	N-terminal	C-terminal soluble region
	1-559:GFP	N-terminal	Full protein:GFP hybrid

4.2 Transformation and Expression

Pipetted 5 μ L of pET28a plasmid containing recombinant construct and antibiotic resistance into CaCl_2 treated BL21 *E. coli* competent cells. After the 30 min incubation at 4°C, the sample was heat shocked for 1 min at 42°C then immediately placed back at 4°C. After 5 min, pre-warmed (37°C) LB broth was aseptically added to the heat shocked samples and the sample was incubated at 37°C for 1 hr. Following incubation, the samples were gently resuspended and distributed with a pipet among warmed LB broth + kanamycin (50 μ g/mL) culture tubes. Samples were incubated at 37°C and 200 rpm shaking. After 12 hr, turbid tubes indicate successful growth of transformed cells.

The starter culture was then added to 1L of super broth growth media (containing 35 g tryptone, 20 g yeast extract, 5 g NaCl and 1 mL of 50 μ g/mL kanamycin) and incubated at 37°C with shaking at 200 rpm. When the optical density at 600nm ($\text{OD}_{600\text{nm}}$) was measured to be ~ 0.6 against a broth blank with a spectrophotometer, protein expression was induced by the addition of 1mL of isopropyl β -D-1-thiogalactopyranoside (IPTG) to the large-scale growth culture. After allowing protein expression to proceed for 12hr at 22°C, the cultures were then transferred from the growth flask to 500 mL centrifuge bottles (Figure 6). Protein pellets were collected by centrifugation at 5,000 x g for 15 min at 4°C. Pellets were stored at -20°C until needed and the supernatant discarded.

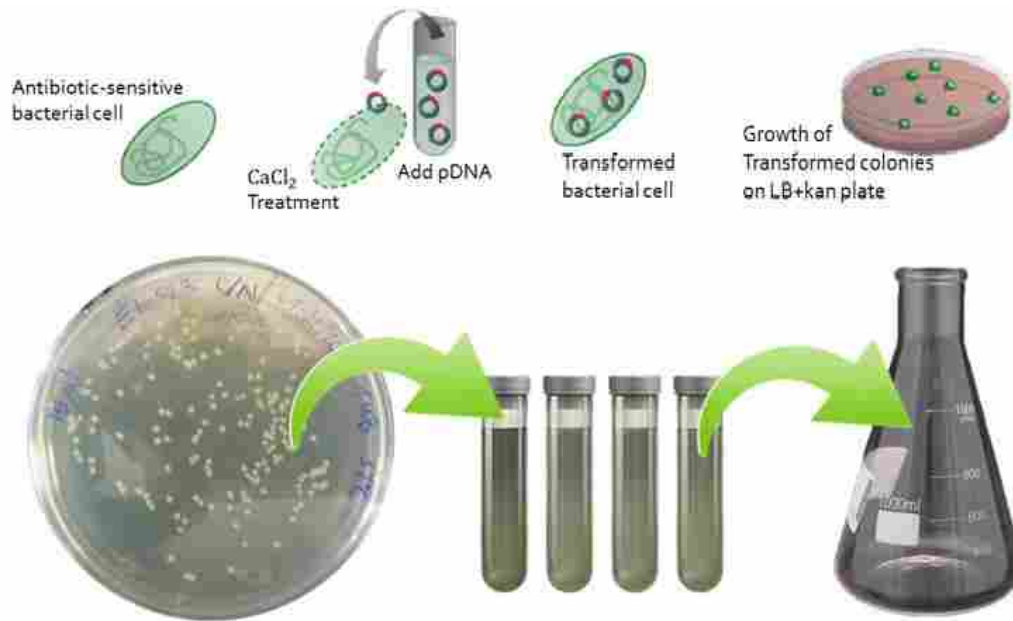


Figure 6. Workflow from plasmid to overexpressed recombinant protein

Plasmid-containing recombinant protein of interest must be transformed into competent BL21 *E. coli* cells. These cells are then selected by antibiotic resistance with kanamycin and grown into high enough quantities to eventually be induced to overexpress recombinant protein in 1 L flasks with IPTG.

4.3 Purification

4.3.1 Immobilized Metal Affinity Chromatography (IMAC)

Cell pellets stored at -20°C were resuspended in $\sim 40\text{mL}$ of Buffer A (optimal contents vary for different protein constructs). Tris-base, NaCl, and imidazole (BioShop in Burlington, ON) were used in the preparation of purification buffers. A Roche protease inhibitor tablet and 1 mL of RNase and DNase were also added to the resuspension. The TS Series Lysing Machine was used to lyse the sample at 17 kpsi. Sample was then centrifuged at $28,000 \times g$ for 45 min at 4°C to separate cellular debris (inclusion bodies and unlysed cells) then the cleared lysate was decanted into a new tube. Nickel nitrilotriacetic acid (Ni-NTA) resin for immobilized metal affinity chromatography (IMAC) (ThermoScientific) was prepared by centrifugation and resuspension in Buffer A. Resin was added to cleared lysate and the sample was allowed to

incubate with gentle mixing at 4°C for ~40 min to maximize recombinant protein binding to the resin.

The supernatant was then loaded into an open column with a capped outlet and the flow-through was collected in a 50 mL Falcon tube. The resin in the column was then washed with 50 mL of Buffer A (containing no imidazole) and the wash 1 fraction was collected in another 50 mL Falcon tube. A second column wash followed with 50 mL of Buffer B (Buffer A containing 25mM imidazole) and collected in a similar fashion. Next, the column was capped at the bottom to stop flow and 15 mL of elution buffer (Buffer A containing 250mM imidazole) was poured onto the resin. The sample was incubated for 5 min in elution buffer to allow saturation of the resin (Figure 7). After the incubation, the column was uncapped and pure recombinant protein was collected 15 x 1 mL elutions in microfuge tubes. A Western blot or SDS-PAGE gel could then be run with the prepared fractions, after heating them briefly at 90°C, to view protein profiles.

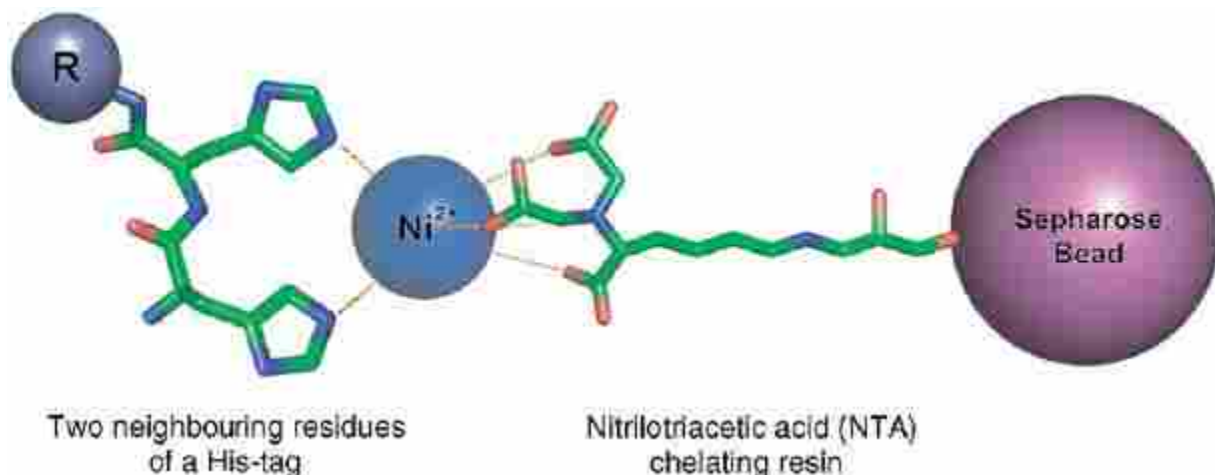


Figure 7. Binding between a histidine-tag and Ni-NTA resin (Qiagen Ltd.)

This diagram from Bolanos-Garcia, V. M., & Davies, O. R. (2006) demonstrates how proteins engineered with histidine tags can be captured with Ni-NTA resin.

4.3.2 SDS-PAGE and Western Blot Analysis

Materials for SDS-PAGE including sodium dodecyl sulfate, ammonium persulfate, and TEMED were purchased from Fisher Scientific, and 30% (w/v) acrylamide/bis-acrylamide was purchased from BioShop. Coomassie stain for dyeing SDS-PAGE gels was prepared with 1% (w/v) Coomassie Brilliant Blue R250, 40% (v/v) methanol, and 10% (v/v) glacial acetic acid from Fisher Scientific. Coomassie destain solution consisted of 10% (v/v) acetic acid and 20% (v/v) methanol in deionized water.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) fractions of flow through, both washes and representative elution fractions were prepared by mixing 20 μ l of each sample with 40 μ l of 5 x SDS with DTT. Gels were setup in the Mini-PROTEAN Tetra cell apparatus (Biorad), filled to correct volume with 1 x Running buffer, wells had samples pipetted into them and apparatus was closed. Gels were run at 200 V for 45 min then either stained with Coomassie R250 or transferred to nitrocellulose paper (Biorad) for a Western blot analysis. If gels were stained, they would then incubate for a minimum of 1 hr in Coomassie dye and then transferred into destain solution until contrast allowed optimal distinction of protein bands.

Western blot transfers were conducted in a Trans-Blot apparatus (Biorad) and run at 4°C in transfer buffer (12 mM Tris, pH 7.5, 96 mM glycine, 20% (v/v) methanol) for 2 hr at 100 V. Transferred blots on nitrocellulose were then blocked with 5% (w/v) skim milk powder in TBS buffer (10mM Tris, pH 7.5, 150 mM NaCl) for ~1 hr. Blots were then washed twice in TTBS (20 mM Tris, pH 7.5, 2 mM NaCl, 0.05% (v/v) Tween) for 7 min each. Primary antibody (mouse anti-His) was diluted 1000 fold in 15 mL blocking buffer and incubated with blots for 45 min. Three subsequent washes were performed in TTBS for 7 min each followed by a 45 min incubation with secondary antibody (alkaline phosphatase conjugated rabbit anti-mouse) diluted

5000 fold in 15 mL blocking buffer. Blots were next washed once in TTBS for 7 min before development. Nitrocellulose was covered in 5-bromo-5-chloro-3-indolyphosphate (BCIP) substrate and incubated in the dark for ~5 min. BCIP cleavage by alkaline phosphatase on the conjugated secondary antibody forms nitroblue tetrazolium (NTB), which leaves a purple colour on His-tagged target protein.

4.3.3 Dialysis and Buffer Exchange

If IMAC was not sufficient for sample purification, dialysis or buffer exchange could be used to gently remove salts, metals or imidazole from the sample prior to further purification steps. First, 4L of dialysis buffer (50mM MgCl₂, 50mM Tris-HCl, pH 7.5) was prepared and cooled to 4°C. Dialysis tubing was cut to fit the total volume of elutions to be dialyzed (2cm of dialysis tube/mL of elution) then soaked in distilled water for 5 min. A knot was tied on one side of the tubing and protein sample was carefully pipetted inside. Once all elutions were transferred into the tubing, all bubbles were removed from the tubing, a second knot tied on the other end and a dialysis clip was clamped against the knot. The dialysis tubing containing the elutions was then placed into a large beaker containing 2L of the previously prepared and cooled dialysis buffer. The beaker was kept at 4°C and allowed to gently mix with a stir bar for 1 hr. The buffer was then cautiously poured out of the beaker (without disturbing the sample within the tubing) and replaced with the remaining 2L of clean dialysis buffer. The sample was then left to sit at 4°C (without stirring) for 16 hr. A buffer exchange could achieve the same results more quickly (but less gently) using a Sartorius (VWR) Vivaspin centrifugal filter device. Vivaspin filter was loaded with the protein sample and concentrated to the desired volume by centrifugation. The addition of a new buffer could simply be added to a concentrated sample afterwards to exchange with the old buffer.

4.3.4 Ion Exchange and Size Exclusion Chromatography (SEC)

Ion exchange (5 mL) HiTrap™ columns were obtained from GE Health Sciences. The column was first equilibrated with 0 M NaCl buffer (25 mM Tris-HCl, pH 7.5) until a consistent profile could be observed on the monitor. The concentrated protein sample was typically loaded onto the column in one injection. Buffers A (no NaCl) and B (1 M NaCl) were used to create a gradient of linearly increasing salt concentration throughout the run based on a protocol created on the FPLC program. As the run progressed, the salt gradient increased at a set rate and a strong independent peak on the ion exchange profile (viewed on the monitor) was used to determine where the desired protein may have eluted (Figure 8). An SDS-PAGE gel and Western Blot were used to verify the elutions that contained the desired protein to be isolated.

Size exclusion chromatography (SEC) is typically used after one or more previous protein purification steps because it works best when there is some separation in size between the protein of interest and remaining sample impurities. SEC columns were obtained from GE Health Sciences. SEC separates molecules according to differences in size as they pass through a SEC medium packed in a column. Proteins do not actually bind to SEC resin (like with IMAC or ion exchange); proteins are separated by size differences. Larger proteins pass through the column more rapidly than smaller proteins that detour through pores in the resin. Like with the ion exchange protocol, the SEC profile can be viewed on the FPLC monitor to determine where the desired protein may have eluted. An SDS-PAGE gel and Western Blot were used to verify the elutions that contained the desired protein to be isolated.

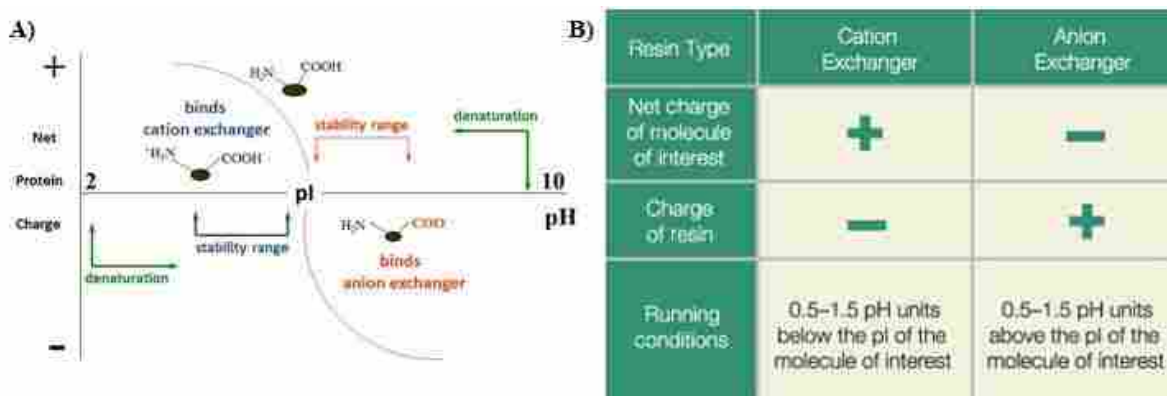


Figure 8. Principles of ion exchange chromatography from Bio-Rad product overview

A) Protein charge above and below pI means that binding to ion exchange media will vary with buffer pH. **B)** Decision table for choosing ion exchange resin.

4.4 Functional Analysis of BcsG

4.4.1 Investigating Phosphatase Activity

A colorimetric phosphate detection kit was purchased from Abcam (Toronto, ON) for the detection of enzymatic hydrolysis of phosphate-rich substrates (Figure 10). All prepared reagents were equilibrated to room temperature prior to use and all standards, controls and samples were done in quadruplicate to ensure accuracy (Figure 9). The ability of BcsG¹⁶⁴⁻⁵⁵⁹ to hydrolyze NTPs was first tested with ATP, GTP and CTP. Reactions were prepared in 96 well plates with varying amounts of enzyme, substrate and buffer to determine optimal conditions for reaction. Subsequent standards and samples were prepared according to the appropriate concentrations of enzyme and buffer found to be acceptable (Table 2). Reactions were analyzed in the Gen5 all-in-one microplate reader at 650 nm. Absorbance measurements for all enzyme reactions were mathematically corrected by subtracting the blanks containing no enzyme. Phosphate concentration of the reaction was extrapolated from the standard curve. Following initial NTP tests, PEA hydrolysis, a wide range of pH buffers, EDTA-treated samples and several metals were tested to obtain more insight into optimal reaction conditions. A BcsG¹⁶⁴⁻⁵⁵⁹ site-directed

mutant (S116A) was also tested with this assay to determine whether the mutated residue was required for activity.

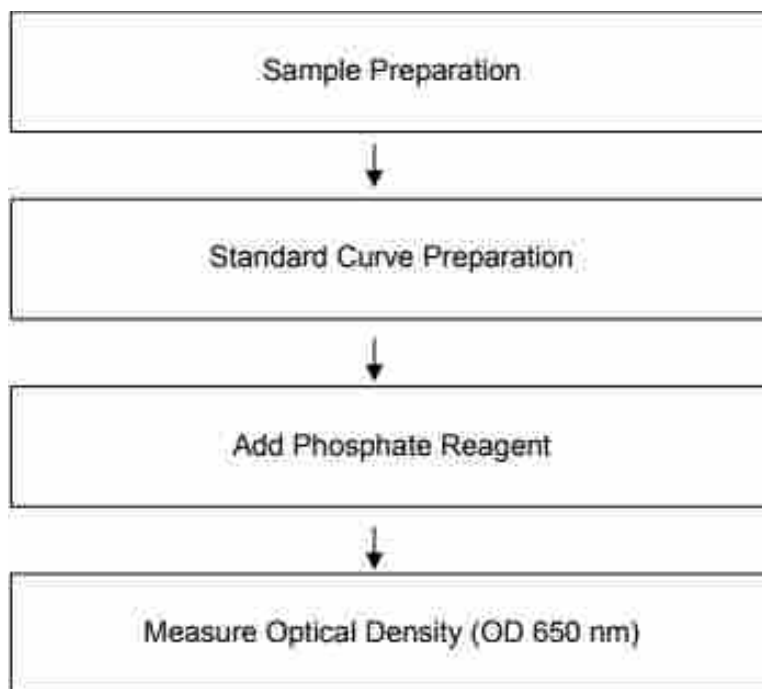


Figure 9. Summary flow chart of phosphate assay (from Abcam protocol booklet)

Table 2. Summary flow chart of phosphate assay (from Abcam protocol booklet)

This table indicates the correct volumes of 100 μM Phosphate Standard and dH_2O to add to standard reaction wells. The 100 μM Phosphate Standard was first prepared by diluting 10 μL of the 10mM Phosphate Standard into 990 μL of dH_2O .

Standard #	Volume of Phosphate Standard (μL)	ddH_2O (μL)	Final volume standard in well (μL)	Amount standard in well (nmol/well)
1	0	600	200	0
2	30	570	200	1
3	60	540	200	2
4	90	510	200	3
5	120	480	200	4
6	150	450	200	5

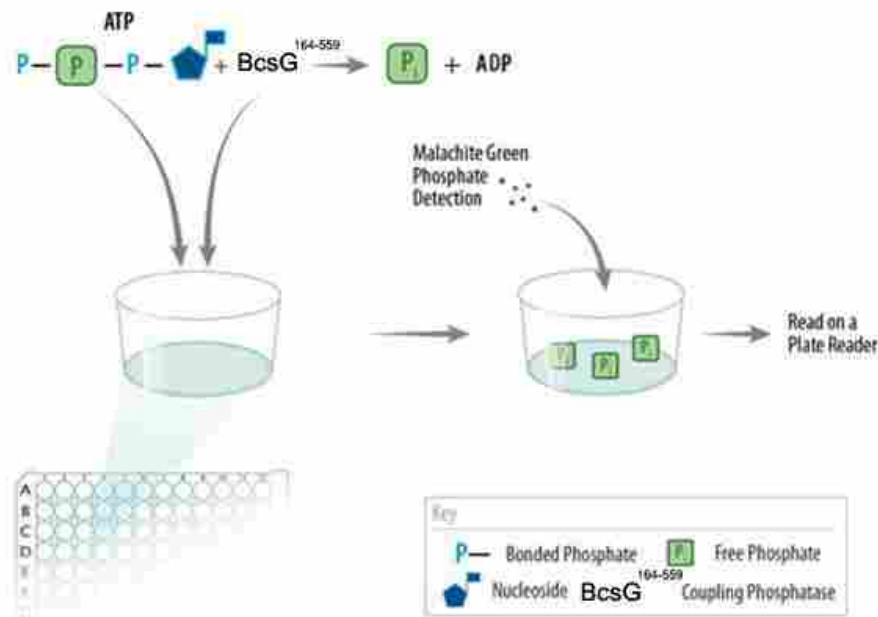


Figure 10. Colourimetric Phosphatase Assay Principles

Assays were conducted in triplicate or quadruplicate at the determined optimal pH of 7.5 and 22°C in Tris-based buffer according to the pH profile results. The reaction was observed at OD_{650nm} which is the optimal wavelength to observe the reaction between released phosphate by-products from NTP substrates and the malachite detection reagent. Assays were run for 40 mins to ensure entire reaction could be observed.

4.4.2 Localization of BcsG¹⁻⁵⁵⁹ Using GFP as a Reporter Protein

To investigate the localization of BcsG in *E. coli*, a BcsG¹⁻⁵⁵⁹:GFP fusion protein was used as a reporter (Feilmeier *et al.*, 2000). A C-terminal GFP fusion protein of the full-length construct was previously created by cloning the BcsG¹⁻⁵⁵⁹ construct into a pET GFP vector by ligation-independent cloning (LIC) using the method described by Addgene (Anonymous Ligation-independent cloning, 2015) to localize BcsG within *E. coli* cells. Subsequent growth conditions were typical of native growth described previously. Cleared lysate of cultures containing hybrid protein was split into two fractions; one was left untreated and the other was treated according to a denaturation-renaturation protocol described by Ward and Bokman (1982) which exposed protein to concentrated HCl followed by 10N NaOH.

Following this, both the treated and untreated samples were allowed to incubate at 4°C for 24 hrs then loaded into a 96 well chimney black microplate (Fluotrack). An excitation, emission reading was taken using the Gen5 all-in-one microplate reader at 365 and 509nm, respectively to determine fluorescence between unhybridized BcsG¹⁻⁵⁵⁹, untreated BcsG¹⁻⁵⁵⁹:GFP, treated BcsG¹⁻⁵⁵⁹:GFP and unhybridized GFP (purchased from Geiner Bio-One International).

4.5 Structural Analysis of BcsE and BcsG

4.5.1 Crystallization Screening

The MCSG suite (Microlytic) and PEG screens were used for crystallization screening for BcsE¹⁻⁵²³ and BcsG¹⁶⁴⁻⁵⁵⁹ constructs. Protein to buffer ratios tested for each construct and condition included 3:1, 2:1, 1:1 and 1:2 to increase the probability of successful crystallization. The Gryphon crystallization robot was pre-programmed for desired drop ratios into sitting drop 96 well plates (Microlytic). Crystalclear tape was used to incubate plates at 18°C until needed. Stored plates were periodically checked by microscopy (Olympus SZX16 Stereomicroscope) for precipitation and crystal formation. Conditions containing newly formed crystals were exposed to UV light or IZIT dye to gain an indication of whether they were salt or protein formations.

Pre-greased hanging drop 24 well crystal plates for crystallographic expansion trials (Crystalgen) were used to expand upon hopeful conditions (Figure 11). Small slides were carefully placed on a clean surface and small drops were pipetted by hand on the slides with varying quantities of buffer and protein sample. Expansion plates were also stored at 18°C until ready for further inspection. Crystal seeding was used on many conditions in an attempt to introduce nucleation sites for crystal growth using the Seed Bead kit protocol (Hampton Research).

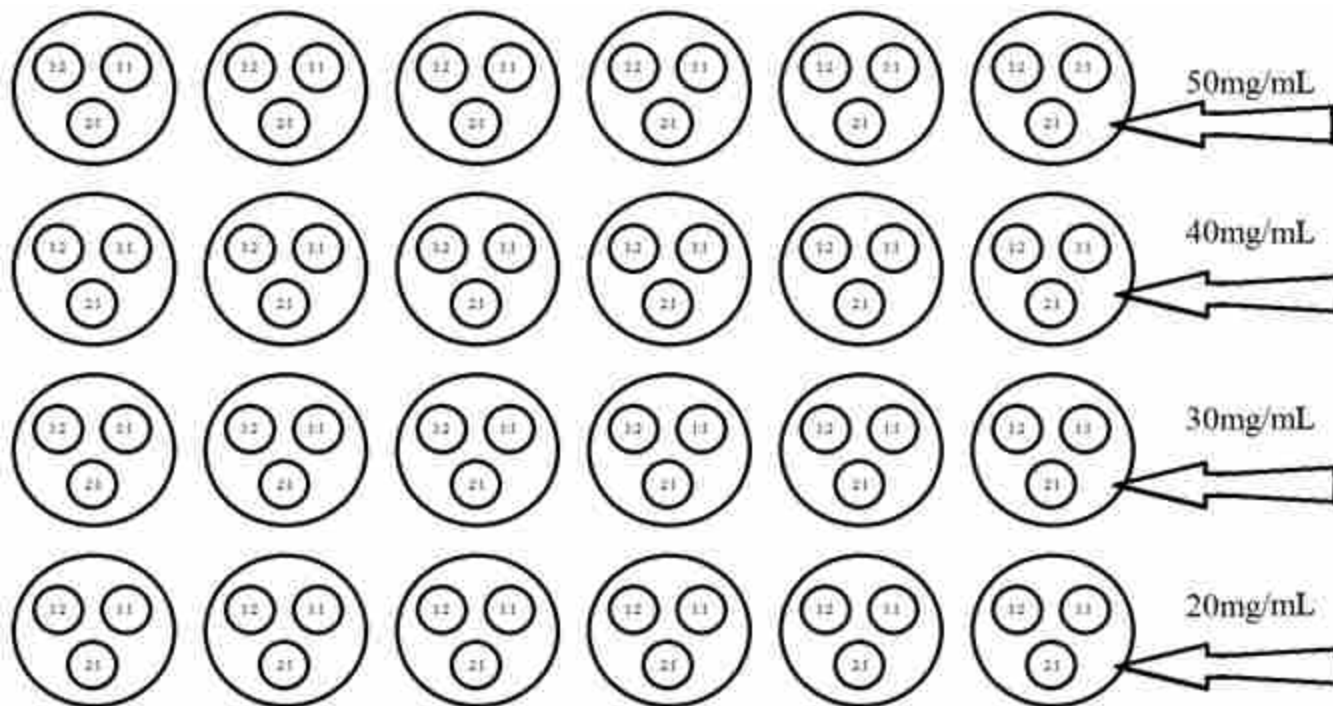


Figure 11. Crystal Expansion Plate Layouts

The hanging drop layout depicted in this figure was done in duplicate: one plate utilized MCSG-4T F7 (0.2 M Magnesium Chloride, 0.1 M MES:NaOH pH 6.5, 25% (w/v) PEG 4000) and the other MCSG-4T G11 (0.1 M HEPES:NaOH pH 7.5 / 20% (w/v) PEG 4000 / 10% (v/v) 2-Propanol). Total drop volumes were 2 or 3 μ L with 2:1, 1:1 and 1:2 (protein:buffer) ratios used to increase the probability of success.

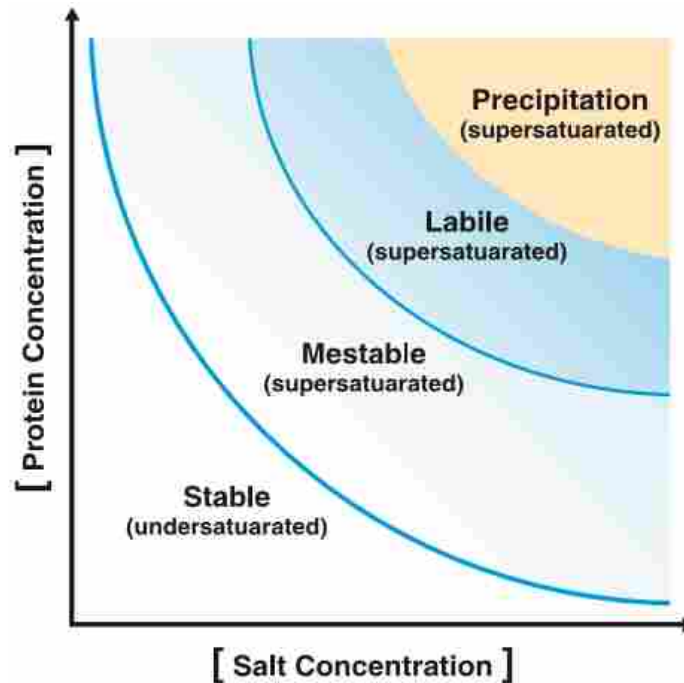


Figure 12. Crystallization zone chart from Seed Bead protocol booklet

The stable zone is undersaturated and, therefore, crystal growth does not occur and the drop remains clear. In the metastable zone nuclei cannot form but the well is supersaturated so crystals can grow if seeds are present. In the labile zone, nuclei can form without seeds and crystals can grow. In the precipitation zone, sample precipitates out of solution and regular crystal growth will not occur.

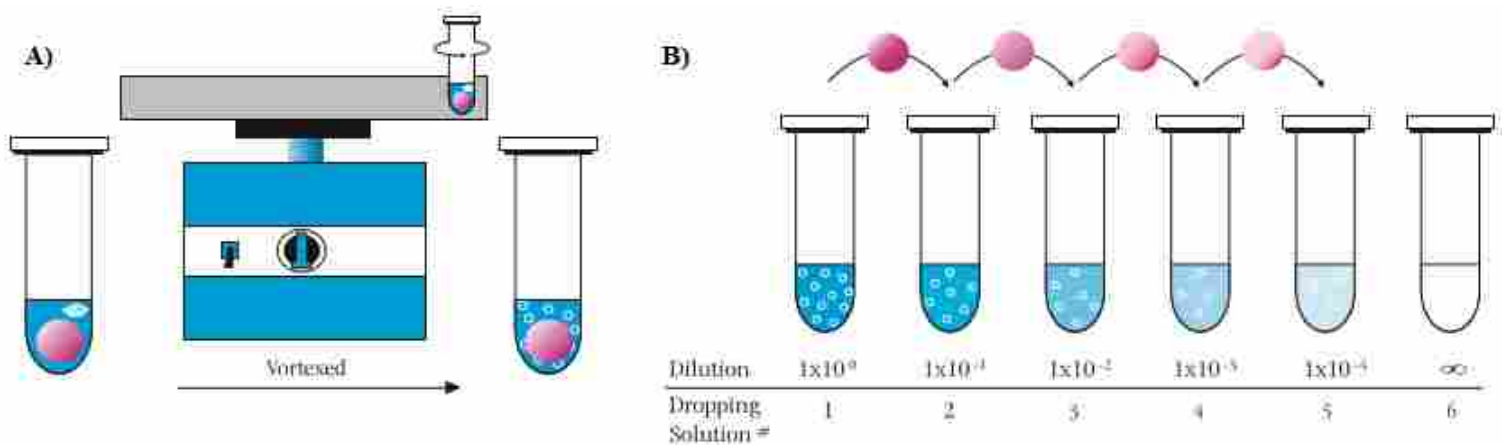


Figure 13. Crystal seed preparation cartoon from Seed Bead protocol booklet

A) Obtained crystal is broken into small crystal seeds by vortexing with seed bead in microfuge tube. B) Dilution series of seed stock is created so that optimal seed concentration can be added to wells with promising crystallization conditions.

4.5.2 X-Ray Diffraction

Crystals selected for analysis were prepared prior to being loaded onto the X-ray crystallography machine at the University of Waterloo or the Canadian Light Source in Saskatchewan. A cryoprotectant, such as glycerol or ethylene glycol, was mixed in a PCR tube with the condition that corresponded to that in which the hopeful crystal resided. Cryoprotectant used for this project was composed of 35% PEG 4000 (v/v) in a 1:1 ratio with the buffer in which the crystal was suspended in. The crystal was carefully pulled out of the well with an appropriately-sized loop and passed through two drops of the cryoprotectant-buffer mixture in succession to saturate the crystal with the protective solution. The needle with the cryoprotected crystal was then stored in a labelled vial within liquid nitrogen until it was ready to be analyzed by X-ray diffraction. To analyze a protein crystal, it was mounted onto the machine so that it was centered on the display monitor for 360° of rotation. The machine was then activated and run for 2 min exposure times and the patterns were saved for analysis.

4.5.3 Multiple Isomorphous Replacement

Crystals selected for analysis were prepared with one alteration from that described above. The cryoprotectant mixed in a PCR tube contained the condition that corresponded to the sample plus 2M KBr or 2M NaBr. The crystal was carefully pulled out of the well with an appropriately-sized loop and passed through two drops of the cryoprotectant-buffer-heavy atom mixture in succession to saturate the crystal with the protective solution and the heavy atom. Soaks for native crystals varied from 5 secs to 22 secs for each exclusive heavy atom mixture in hopes of achieving the right saturation for phase resolution. The loop with the cryoprotected crystal was then stored in a labelled vial within liquid nitrogen until it was ready to be analyzed by X-ray diffraction.

4.5.4 Selenomethionine Crystallization

Selenomethionine ingredients (SelenoMet Base and SelenoMet Nutrient Base) were supplied from Molecular Dimensions and the analytical grade L-(+)-SelenoMethionine was from BioShop. Autoclaved SelenoMet Base dissolved into 1 L of dH₂O and sterile filtered SelenoMet Nutrient Mix were used to create the methionine minus medium. Just prior to use of minimal media, 2 mL of analytical grade L-(+)-SelenoMethionine was added per liter of media for final concentration of 45 mg/L.

BcsG¹⁶⁴⁻⁵⁵⁹ was transformed just as it would be prior to a normal expression; however, expression conditions were distinct. A 100 mL culture of BcsG¹⁶⁴⁻⁵⁵⁹ was grown overnight in SB broth. The cells were pelleted via centrifugation after 16 hrs and rinsed three times in sterile water to remove all traces of SB. Pellets were then resuspended in 1 mL of dH₂O and inoculated into 1 L of pre-warmed minimal media containing L-SeMet. Once the culture reached an OD₆₀₀ expression of target protein was induced with 1mM IPTG. Normal protein capture, storage and purification techniques followed expression.

Table 3. SeMet Crystal Expansion Plate Layouts

Plate layout for SeMet grown BcsG¹⁶⁴⁻⁵⁵⁹ in condition MCSG-4T G11(0.1 M HEPES:NaOH pH 7.5 / 20% (w/v) PEG 4000 / 10% (v/v) 2-Propanol).

		No substrate		2mM CTP Substrate			
		20mg/mL	30mg/mL	20mg/mL	30mg/mL	20mg/mL	30mg/mL
No DTT	10% PEG						
	20% PEG						
2mM DTT	10% PEG						
	20% PEG						

5. RESULTS

5.1 Bioinformatics Analyses

To gain insight into the hypothetical properties of BcsE and BcsG constructs, bioinformatics analyses were conducted. The general trajectory taken for *in silico* investigation of BcsE and BcsG required making homology assessments based on alignments with primary amino acid sequences. Amino acid sequences were submitted into ProtParam to acquire numerous basic parameters for each protein construct such as molecular weights, isoelectric points (pI), percentage of each amino acid, extinction coefficients and absorbance values and instability indexes (Gasteiger *et al.*, 2005). Next, amino acid sequences were entered on PSort to predict cellular localization of proteins (Yu *et al.*, 2010). SignalP was then utilized to postulate if signal peptide cleavage sites, which are stretches of amino acids recognized by signal peptidases for cleavage, were present (Neilson, 2017).

The TMpred program was used to predict if there were any membrane-spanning regions and (if so) what their orientation was using the TMbase database of naturally-occurring transmembrane proteins (Hofmann and Stoffel, 1993). Secondary structures were then predicted by inputting the amino acid sequences of proteins into PSIPRED and Phyre2 (Buchan *et al.*, 2013; Kelley *et al.*, 2015). Phyre2 also proposed hypothetical tertiary structures based on submitted amino acid sequences for each construct. Clustal Omega was used to compare multiple sequence alignments for all protein constructs based on top homology suggestions from Phyre2. Clustal Omega produced multiple sequence alignments and displayed evolutionary relationships for all submitted sequences (Sievers *et al.*, 2011).

Some protein constructs simply turn out not to be crystallizable because of their unique physical properties. If screening optimization yields no fruitful leads, new protein constructs could be generated in order to increase the likelihood of eventual success (Dale *et al.*, 2003).

When deciding on where to cleave proteins, searching bioinformatics resources for less structured strings of residues within a protein is a common strategy. Hypothetical secondary structures give grounds for deciding where to truncate a protein so as most likely not to interfere with folding or stability. This technique was used to decide where to truncate BcsE and BcsG when creating all individual constructs (BcsE¹⁻²¹⁵, BcsE²²⁴⁻⁵²³ and BcsG¹⁶⁴⁻⁵⁵⁹). All protein constructs generated previously in the Weadge lab were created using the hypothetical boundaries identified by PSIPRED and Phyre2 so as not to interfere with active sites, α helices or β strands (refer to Appendix Figures A1 and A2).

5.1.1 BcsE Constructs Bioinformatics Analysis

Table 4 displays a number of important hypothetical characteristics for BcsE¹⁻⁵²³ from ProtParam which is an ExPASy bioinformatics resource used to compute physical and chemical parameters by comparing a query amino acid sequence. These features were taken into consideration later for preparation of appropriate buffers during expression and purification. The theoretical pI gives a good starting point for selecting a buffer pH and for utilizing ion exchange purification which relies on charge. The molecular weights are important to know when viewing SDS-PAGE gels and the extinction coefficients are used when testing protein concentrations. The instability index can indicate whether proteins need to be purified at 4°C or if they may be worked with at room temperature for example. Predicted location of proteins can also be tested if their location remains unproven. Information provided by 3DLigand is valuable when choosing initial substrates and metals when testing protein activity or crystallization. Hypothetical active site residues can also be initial targets for site-directed mutagenesis; with a loss of activity from a mutant indicating a residue's contribution to activity. The heterogens for BcsE¹⁻⁵²³ proposed by the server included: adenosine diphosphate, magnesium and adenosine triphosphate (Figure 17).

Table 4. ProtParam protein characteristics for BcsE constructs

Calculated theoretical values do not include 6xHis tags. All of the listed parameters contribute to the foundation for new protein research (Gasteiger *et al.*, 2005). Abs 0.1% (=1 g/l) values assume all Cys residues are reduced.

Construct	BcsE ¹⁻⁵²³	BcsE ¹⁻²¹⁵	BcsE ²²⁴⁻⁵²³
Number of amino acids	523	215	300
Molecular weight (Da)	59398	24642	33862
Theoretical pI	6.14	5.43	8.37
Extinction coefficients (M-1 cm-1, 280 nm in water)	84950	50210	34740
Abs 0.1% (=1 g/l)	1.430	2.038	1.026
Cellular location on PSort	Unknown	Unknown	Unknown
SignalP	NO	NO	NO
Transmembrane Segments	0	0	0
Instability Index	45.45 (unstable)	40.63 (unstable)	46.23 (unstable)

The Phyre2 results for the top five suggested BcsE¹⁻²¹⁵ homologs are displayed below in Figure 14. High alignment coverage (~67%) is apparent for the N-terminus from residue 8-150 to the template proteins as indicated by the bars on the left side of the panels. The percentage identity (% i.d.) indicates the likely accuracy of the model; for very accurate models this number should be >30%. The % i.d. values are likely acceptably high for models 1, 2 and 4; however, models 3 and 5 are likely too low to be very accurate models. Templates 1, 2 and 5 suggest a potential ATPase domain; template 3 suggests potential circadian clock protein and template 4 suggests possible RecA-like DNA recombination domain.

The Phyre2 results for the top five suggested BcsE²²⁴⁻⁵²³ homologs are displayed below in Figure 15. Low alignment coverage (~30%) is apparent for the C-terminus from residues 26-118. Percent identities for all homologs are too low to be useful for molecular replacement when attempting to model a 3D protein structure. Templates 1, 2 & 5 suggest a potential transport protein homology and templates 3 and 4 suggests potential vitamin B12 and NAD(P) binding

domains, respectively. The results for the N-terminus of BcsE are much stronger and, therefore, likely more reliable than the results for the C-terminal region.

As displayed in Figure 16, Clustal Omega was also used to analyze BcsE¹⁻⁵²³ as a query protein against the top five homologous proteins suggested by Phyre2. Similarly to the Phyre2 output, the greatest amount of alignment can be visibly distinguished on the N-terminus residues between proteins. The 1U9I_A template is the circadian clock protein KaiC with phosphorylation sites and other homologs are RecA protein-like, P-loop containing nucleoside triphosphate hydrolases.

If BcsE¹⁻²¹⁵ has RecA function, it would be an ATP-dependent enzyme responsible for the homologous recombination of broken single-stranded DNA segments by using double-stranded parent molecules as a template for repair (Kelso *et al.*, 2017). P-loop NTPases most commonly hydrolyze the beta-gamma phosphate bond of a bound NTP which causes conformational changes in neighbouring molecules (Leipe *et al.*, 2004). The P-loop NTPase fold, the most prevalent nucleotide-binding conformation, has substrate preference for either ATP or GTP (Leipe *et al.*, 2004). P-loop NTPases have conserved Walker A (P-loop) and Walker B motifs which bind the beta and gamma phosphate moieties of the bound NTP, and a Mg²⁺ cation, respectively (Walker *et al.*, 1982). Bioinformatics analysis indicates that BcsE¹⁻²¹⁵ has the appropriate conserved residues in these motifs to bind ADP and Mg²⁺ cations as well.












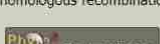


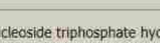
#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	dlubea1 O Q	 Alignment		91.7	16	PDB header: transport protein Fold: P-loop containing nucleoside triphosphate hydrolases Superfamily: P-loop containing nucleoside triphosphate hydrolases Family: RecA protein-like (ATPase-domain) 
2	d1xp8a1 O Q	 Alignment		82.9	17	Fold: P-loop containing nucleoside triphosphate hydrolases Superfamily: P-loop containing nucleoside triphosphate hydrolases Family: RecA protein-like (ATPase-domain) 
3	c1u91A O Q	 Alignment		66.6	10	PDB header: circadian clock protein Chain: A: PDB Molecule: ka1c; PDBTitle: crystal structure of circadian clock protein ka1c with phosphorylation2 sites 
4	c3cmvG O Q	 Alignment		61.5	19	PDB header: recombination Chain: G: PDB Molecule: protein reca; PDBTitle: mechanism of homologous recombination from the recA-ssdna/dsdna2 structures 
5	d1tf7a2 O Q	 Alignment		61.2	8	Fold: P-loop containing nucleoside triphosphate hydrolases Superfamily: P-loop containing nucleoside triphosphate hydrolases Family: RecA protein-like (ATPase-domain) 

Figure 14. BcsE¹⁻²¹⁵ N-terminal construct Phyre2 homology results

Left to right: Template proteins compared to construct query. High alignment coverage (67%) is apparent for the N-terminus from residues 8-150 as indicated by the bars on the left side of the panels. 3D models of template suggesting probable conformation of query protein. Confidence probability that query protein and template are homologous are adequately high for models 1 & 2, but lower than preferred for 3 - 5. %i.d. indicates the likely accuracy of the model; for very accurate models this number should be >30% (however, even at <15% models can be useful if confidence is high. Templates 1, 2 & 5 suggest a potential ATPase domain; template 3 suggests potential circadian clock protein and template 4 suggests possible RecA-like DNA recombination domain. (Accessed August 2017)

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c3c85A 			81.2	12	PDB header: transport protein Chain: A: PDB Molecule: putative glutathione-regulated potassium-efflux system PDBTitle: crystal structure of trka domain of putative glutathione-regulated2 potassium-efflux kefb from vibrio parahaemolyticus Run Investigator
2	c2q1uA 			70.5	12	PDB header: transport protein Chain: A: PDB Molecule: hypothetical protein tm1088a; PDBTitle: crystal structure of a putative transport protein (tm1088a) from2 thermotoga maritima at 1.50 a resolution Run Investigator
3	d7rega2 			66.1	12	Fold: Flavodoxin-like Superfamily: Cobalamin (vitamin B12)-binding domain Family: Cobalamin (vitamin B12)-binding domain Run Investigator
4	d1d1a 			51.8	11	Fold: NAD(P)-binding Rossmann-fold domains Superfamily: NAD(P)-binding Rossmann-fold domains Family: Potassium channel NAD-binding domain Run Investigator
5	c3evvA 			50.3	15	PDB header: transport protein Chain: A: PDB Molecule: c-terminal domain of glutathione-regulated potassium-efflux PDBTitle: crystal structure of the c-terminal domain of e. coli kefc in complex2 with keff Run Investigator

Figure 15. BcsE²²⁴⁻⁵²³ C-terminal construct Phyre2 homology results

Left to right: Template proteins compared to construct query. Low alignment coverage (30%) is apparent for the C-terminus from residues 26-118 as indicated by the bars on the left side of the panels. 3D models of template suggesting probable conformation of query protein. Confidence probability that query protein and template are homologous is fairly high for model 1, but lower than preferred for 2 - 5. Percent identities are too low to be useful considering the corresponding low-level confidence probabilities. Templates 1, 2 & 5 suggest a potential transport protein homology; templates 3 & 4 suggests potential binding domains. (Accessed August 2017)

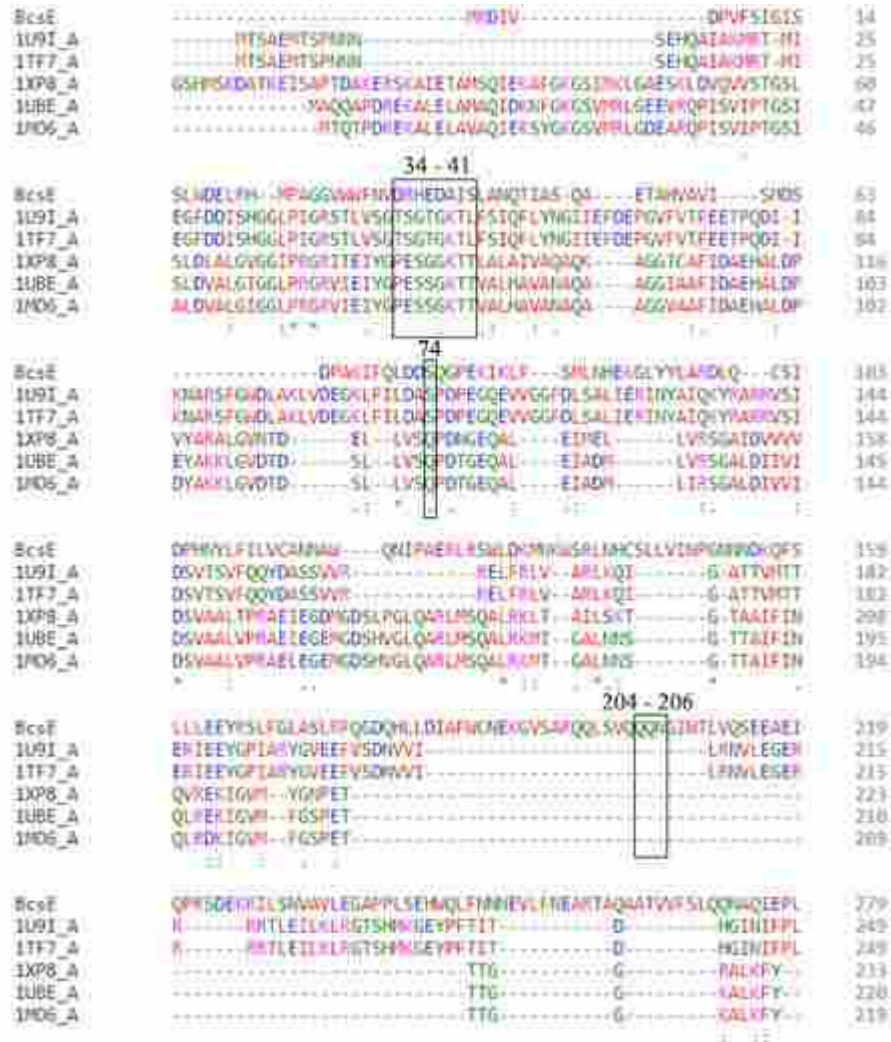


Figure 16. BcsE¹⁻⁵²³ Clustal Omega alignment with top five suggested Phyre2 homologs
 Alignment has been truncated to show greatest homology and 3DLigand predicted binding residues indicated by squares. The top five homologous proteins suggested by Phyre2 were selected for this alignment. 1U9I_A is the circadian clock protein KaiC with phosphorylation sites. Other homologs are RecA protein-like, P-loop containing nucleoside triphosphate hydrolases.

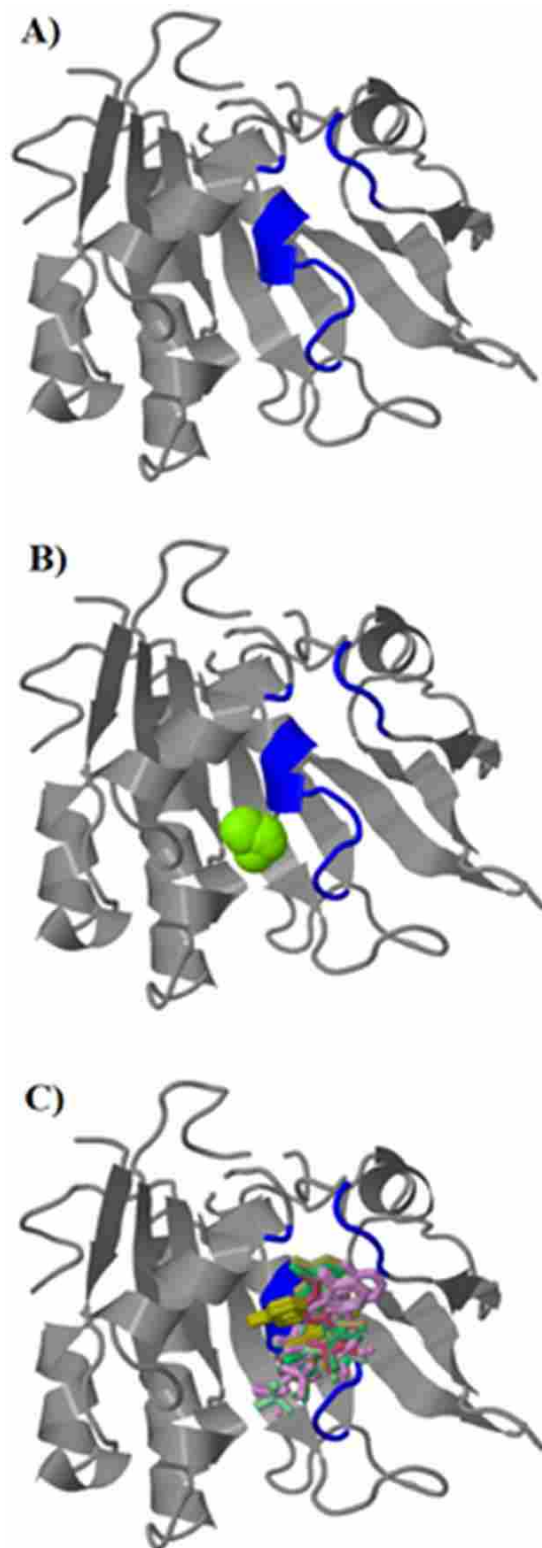


Figure 17. 3DLigand BcsE tertiary structure model and predicted binding

(A) Blue indicates predicted active site. (B) Green indicates magnesium as the predicted metallic heterogen. (C) Various hypothetical orientations within the active site for ADP and ATP which are the predicted non-metallic heterogens.

5.1.2 BcsG Bioinformatics Analyses

For the same reasons covered in Section 5.1.1 for BcsE, Table 5 displays a number of vital characteristics for BcsG constructs proposed by ProtParam. High BcsG¹⁻⁵⁵⁹ alignment coverage was observed for all Phyre2 templates (Table 6) with the highest value observed for template 5 (82% total covering residues 14-474) which is an integral membrane protein lipooligosaccharide PEA transferase (EptA) from *N. meningitidis*. The % i.d. values indicate that all models are fairly accurate (all % i.d. values are ~15%); however, these values are insufficient for molecular replacement 3D protein modeling purposes. Template 1 suggests a potential membrane protein; template 2 suggests potential alkaline phosphatase; templates 3 & 4 suggest possible lipoteichoic acid synthase activity and template 5 suggests possible integral membrane protein with PEA transferase activity. These templates are all intriguing since a hypothesis for BcsG function involved labelling cellulose with PEA during export from the bacteria. The PEA metabolism and its relevance to BcsG are discussed in greater detail in subsequent sections.

As displayed in Figure 18, a Clustal Omega alignment was conducted for BcsG¹⁶⁴⁻⁵⁵⁹ with the same top five homologous proteins as suggested by Phyre2; likewise with high alignment values. BcsG was predicted to contain four N-terminal transmembrane segments by Unitprot and TMPred which matches hypotheses that BcsG is an integral membrane protein. TMPred also suggested that the soluble C-terminus of BcsG would be located outside of the plasma membrane (Table 7). Possible heterogens suggested for BcsG¹⁻⁵⁵⁹ by the 3DLigand server included: AMP, Mg²⁺, Ca²⁺ and Zn²⁺ (Figure 19). Predicted residues are listed in Table 8 and roles of these metals are discussed in greater detail in subsequent sections.

Table 5. ProtParam protein characteristics for BcsG constructs

All of the listed parameters contribute to the foundation for new protein research. All of the listed parameters contribute to the foundation for new protein research (Gasteiger *et al.*, 2005). Abs 0.1% (=1 g/l) values assume all Cys residues are reduced.

Construct	BcsG ¹⁻⁵⁵⁹	BcsG ¹⁶⁴⁻⁵⁵⁹ +His ₆
Number of amino acids	559	418
Molecular weight (Da)	62086	45496
Theoretical pI	6.11	6.04
Extinction coefficients (M-1 cm-1, 280 nm in H ₂ O)	122965	49390
Abs 0.1% (=1 g/l)	1.981	1.086
Cellular location on PSort	Cytoplasmic membrane	Extracellular
SignalP	NO	NO
Transmembrane segments	4	0

Table 6. Phyre2 top five homology suggestions for BcsG¹⁻⁵⁵⁹

Left to right: Template proteins compared to construct query. High alignment coverage for all templates with highest value observed for template 5 (82%). The confidence probability that query protein and templates are homologous is nearly 100% for all models. The %i.d. indicates the likely accuracy of the model; for very accurate models this number should be >30%; however, all % i.d. values are ~15% which should be sufficient to ensure high accuracy due to the associated high confidence levels. Template 1 suggests a potential membrane protein; template 2 suggests potential alkaline phosphatase; templates 3 & 4 suggest possible lipoteichoic acid synthase activity and template 5 suggests possible an integral membrane protein with PEA transferase activity.

PDB	Coverage	Confidence	% i.d.	PDB Molecule	PDB Title
1. 5i5f	59%	99.0	14	inner membrane protein yejm	salmonella global domain 191
2. 3lxq	57%	99.0	13	uncharacterized protein vp1736	crystal structure of a protein in the alkaline2 phosphatase
3. 4uor	47%	98.8	16	lipoteichoic acid synthase	structure of lipoteichoic acid synthase ltas from listeria2 monocytogenes in complex with glycerol phosphate
4. 2w8d	47%	98.8	15	processed glycerol phosphate lipoteichoic acid synthase	distinct and essential morphogenic functions for wall-and2 lipo-teichoic acids in <i>Bacillus subtilis</i>
5. 5fgn	82%	98.8	14	lipooligosaccharide phosphoethanolamine transferase a	integral membrane protein lipooligosaccharide phosphoethanolamine2 transferase a (EptA) from <i>Neisseria meningitidis</i>

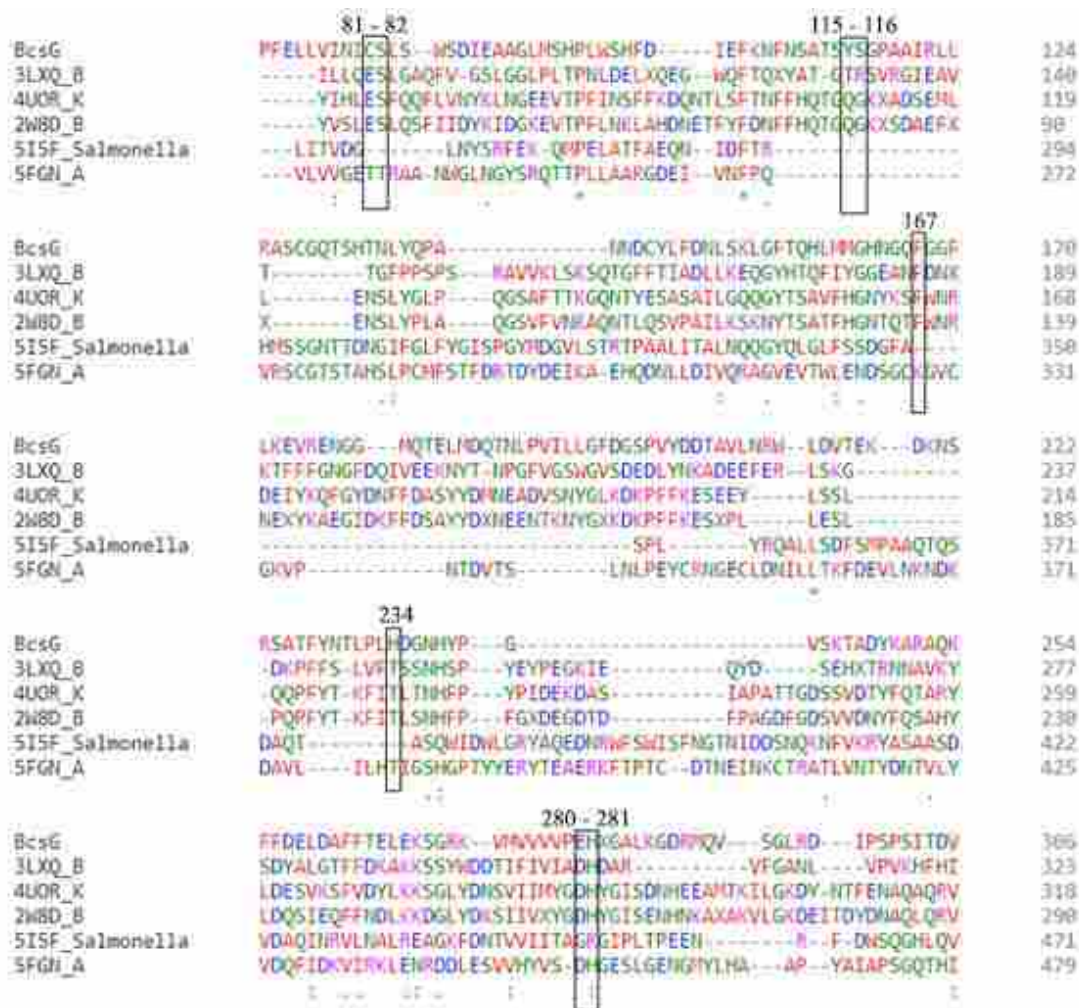


Figure 18. BcsG¹⁶⁴⁻⁵⁵⁹ Clustal Omega alignment with top five suggested Phyre2 homologs
 Alignment has been truncated to show greatest homology and 3DLigand predicted binding residues indicated by squares. The top five homologous proteins suggested by Phyre2 were selected for this alignment. 3LXQ_B is protein vp1736 in the alkaline phosphatase superfamily from *Vibrio parahaemolyticus*. 4UOR_K is the lipoteichoic acid synthase LtaS from *Listeria monocytogenes*. 2W8D_B is a processed glycerol phosphate lipoteichoic acid synthase which conducts distinct and essential morphogenic functions for wall- and lipo-teichoic acids in *Bacillus subtilis*. 5FGN_A is a hydrolase lipooligosaccharide PEA transferase a integral membrane protein lipooligosaccharide PEA transferase a (EptA) from *N. meningitidis*. 515F_D is the inner membrane protein YejM; salmonella global domain 191. Starts at residue #P72.

Table 7. TMPred predicted transmembrane regions for BcsG¹⁻⁵⁵⁹
 Transmembrane segment positions and shape as determined by TMPred.

Segment #	From	To	Score	Orientation
1	44	63	1660	o-i
2	66	84	1002	i-o
3	118	136	2507	o-i
4	138	161	2395	i-o

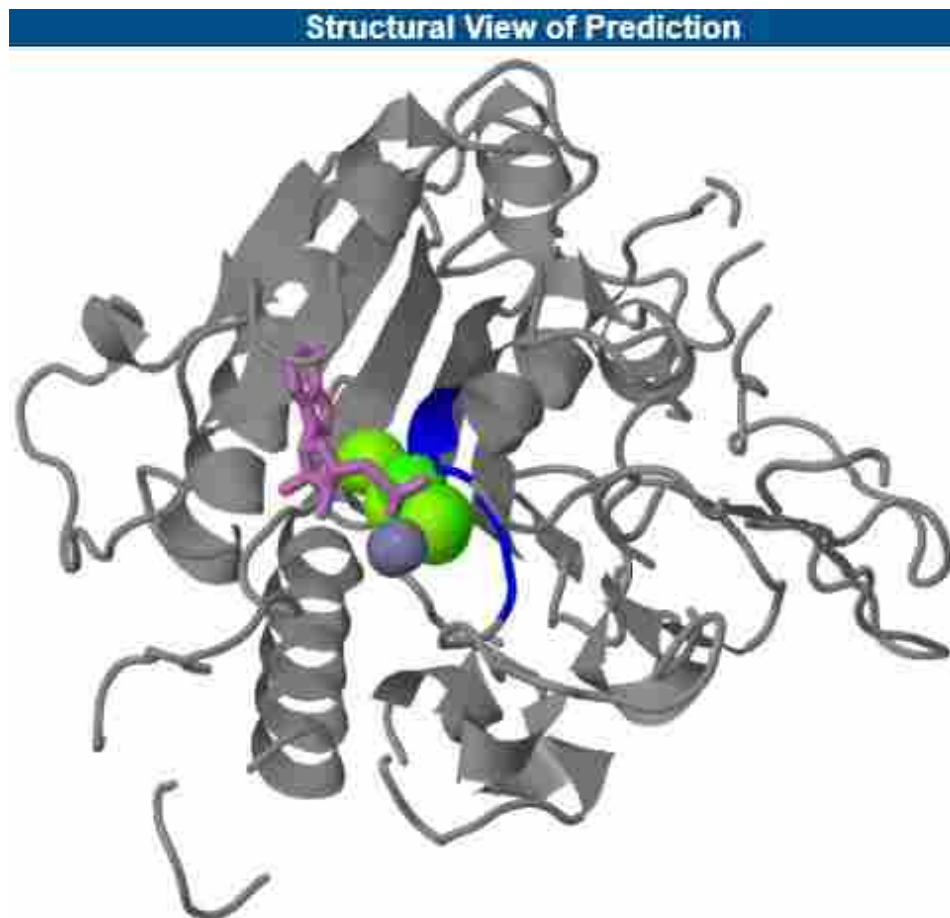


Figure 19. 3DLigand BcsG tertiary structure model and predicted binding

Template used was 3lxqB which is an alkaline phosphatase predicted to be homologous to the soluble portion of BcsG (57% coverage, 99.0% confidence, 13% identity). The blue area represents the predicted binding site. The green portion represents the metallic heterogens which is predicted to be magnesium, calcium or zinc. The purple region depicts the non-metallic heterogen which is predicted to be AMP.

5.1.3 BcsG SDM Constructs Based on Bioinformatics Analysis

Based on the residues identified during bioinformatics to be a part of the active site during homology searches (summary in Figure 18), a number of key mutants were produced (by a previous student in the Weadge lab) by site-directed mutagenesis to test for potential loss of activity (Table 8). The residues of interest were mutated to an alanine first because it is a small non-polar, uncharged residue that should be relatively benign from a structural standpoint. Table 8 below shows all of the BcsG mutants that were sequenced, however for this project, only S116A was tested because it was predicted to be the number one active site residue.

Table 8. Site-directed mutants for BcsG¹⁶⁴⁻⁵⁵⁹

All mutations were satisfactory except for E280A and Y167A. The S116A mutant was tested first because it was predicted to be the number one active site residue from bioinformatics. For all of the alignments below the full BcsG sequence (from *E. coli* K12) was used as the Query even though the constructs are BcsG¹⁶⁴⁻⁵⁵⁹ and the subject is the sequenced plasmids.

Mutation	Hypothetical Role	Sequencing Analysis
E280A	Possible metal binding	Did not cover mutation
H281A	Possible metal binding	Satisfactory
H234A	Possible metal binding	Satisfactory
S116A	This position is modified in many structures – active site residue	Satisfactory
S82A	Possible catalytic nucleophile	Satisfactory
C81A	Possible catalytic nucleophile	Satisfactory
Y115F	predicted to be involved in binding by 3D ligand	Satisfactory
Y167A	predicted to be involved in binding by 3D ligand	Did not cover mutation

5.2 Protein Expression and Purification

5.2.1 Protein Expression

The expression conditions for each protein construct were tested individually to determine the optimal length of expression and temperature requirements. Expression levels did not improve with more than 1 mM IPTG at induction at cell density of $OD_{600} = 0.6$, so 1 mM IPTG was used to conserve resources. Maximal protein expression was observed when allowed to continue for 16 hrs for all constructs. There were no signs of protein degradation after 16 hrs, so samples were routinely collected after this duration. Prior to induction of recombinant protein expression with IPTG, cultures were incubated at 37°C. The two temperatures tested during protein expression were 22°C and 37°C to determine if a lower temperature would produce higher yields of target proteins. Greater protein yields for all constructs were achieved at 22°C, so all subsequent expressions were conducted at 22°C. After expressions were completed, cells were pelleted by centrifugation and stored at -20°C until required.

The highest protein yields were consistently achieved with the BcsG¹⁶⁴⁻⁵⁵⁹ construct with ~40 mg regularly produced per litre of cell culture under optimal expression conditions. Significantly lower yields were observed with BcsE constructs (BcsE¹⁻⁵²³: ~4 mg/L culture, BcsE¹⁻²¹⁵: ~8 mg/L culture, BcsE²²⁴⁻⁵²³: ~4 mg/L culture). Very distinct bands of BcsG¹⁶⁴⁻⁵⁵⁹ can be observed in Figure 20 at ~45kDa which is the expected molecular weight for this construct. These high protein yields for BcsG¹⁶⁴⁻⁵⁵⁹ allowed for numerous downstream structural and functional analyses as outlined later in this document.

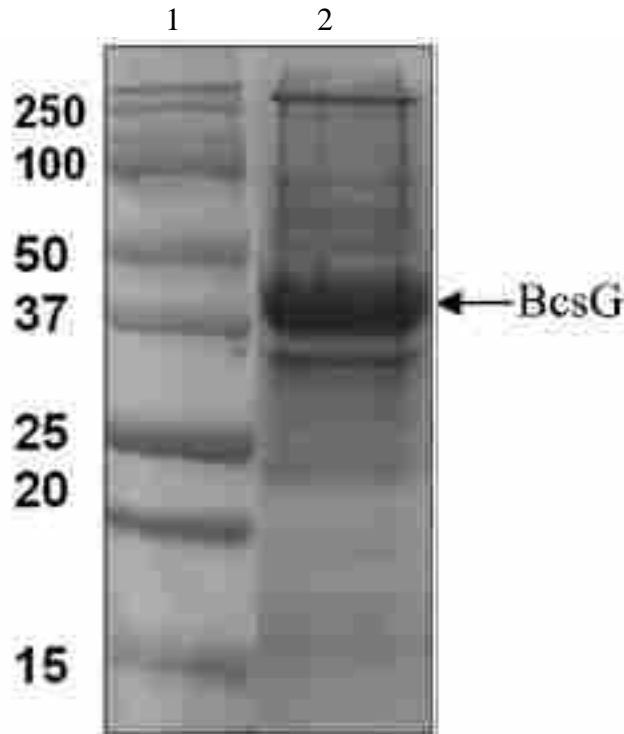


Figure 20. SDS-PAGE showing BcsG¹⁶⁴⁻⁵⁵⁹ expression for 1 L *E. coli* BL21 cultures
SDS-PAGE analysis of BcsG¹⁶⁴⁻⁵⁵⁹ protein expression incubated at 22°C and 200 rpm shaking for 16 h induced with 1 mM IPTG. BcsG¹⁶⁴⁻⁵⁵⁹ indicated with arrow at 45 kDa. Lane 1, molecular-weight markers (kDa); lane 2, SDS-PAGE fraction of expressed BcsG¹⁶⁴⁻⁵⁵⁹ sample.

5.2.2 Immobilized Metal Affinity Chromatography

Following expression, purification by IMAC was used for the isolation of proteins of interest engineered with a histidine tag. If further purification was required, ion exchange or size exclusion chromatography was used to remove persistent contaminants from the sample. IMAC alone was sufficient for the purification of BcsG¹⁶⁴⁻⁵⁵⁹; however, secondary measures were required for BcsE constructs. Figures 21 and 22 show examples of typical IMAC SDS-PAGE profiles for BcsG¹⁶⁴⁻⁵⁵⁹ (~45 kDa) and BcsE¹⁻⁵²³ (~59 kDa), respectively. Some BcsG¹⁶⁴⁻⁵⁵⁹ was lost in wash 2 (W2), but large quantities of pure protein suitable for testing were obtained in all elution fractions (E#) (Figure 21, lanes 4 - 10). Many impurities remain in the BcsE¹⁻⁵²³ elution fractions (Figure 22, lanes 5 - 10) meaning further purification was required.

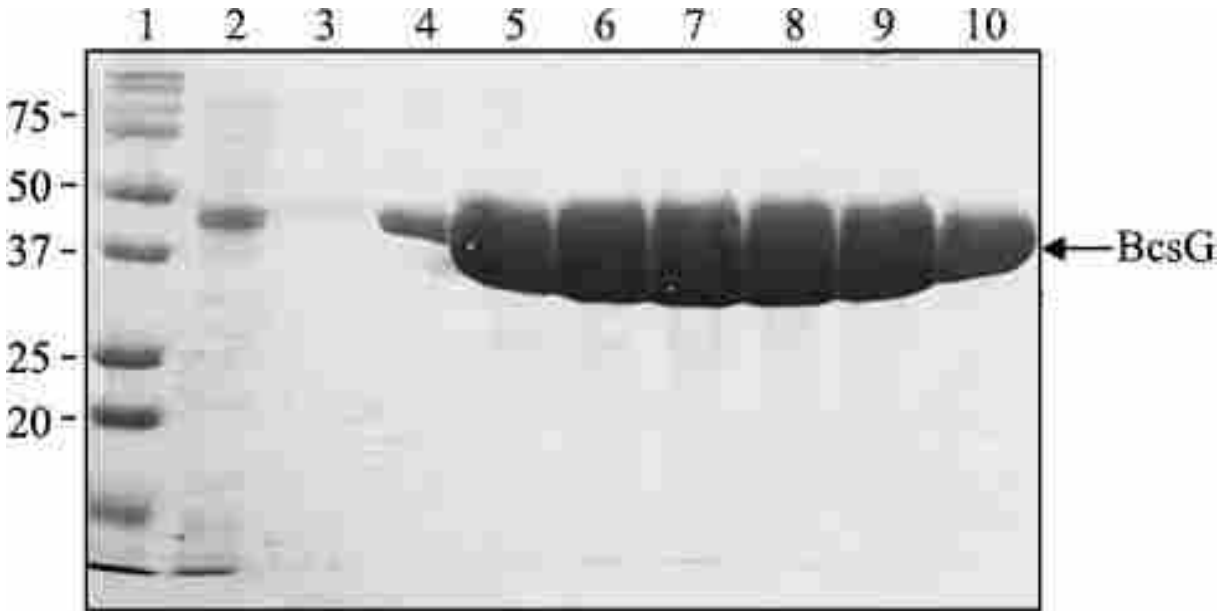


Figure 21. SDS-PAGE IMAC Purification of BcsG¹⁶⁴⁻⁵⁵⁹
SDS-PAGE analysis (12% (v/v)) of BcsG¹⁶⁴⁻⁵⁵⁹ Ni-NTA purification fractions. Lane 1, molecular-weight markers (kDa); lane 2, Ni-NTA column unbound lysate; lane 3, wash with buffer A containing no imidazole; lane 4, wash with buffer A containing 25 mM imidazole; lanes 5-10, purified BcsG¹⁶⁴⁻⁵⁵⁹ elution fractions from the Ni-NTA affinity chromatography column.

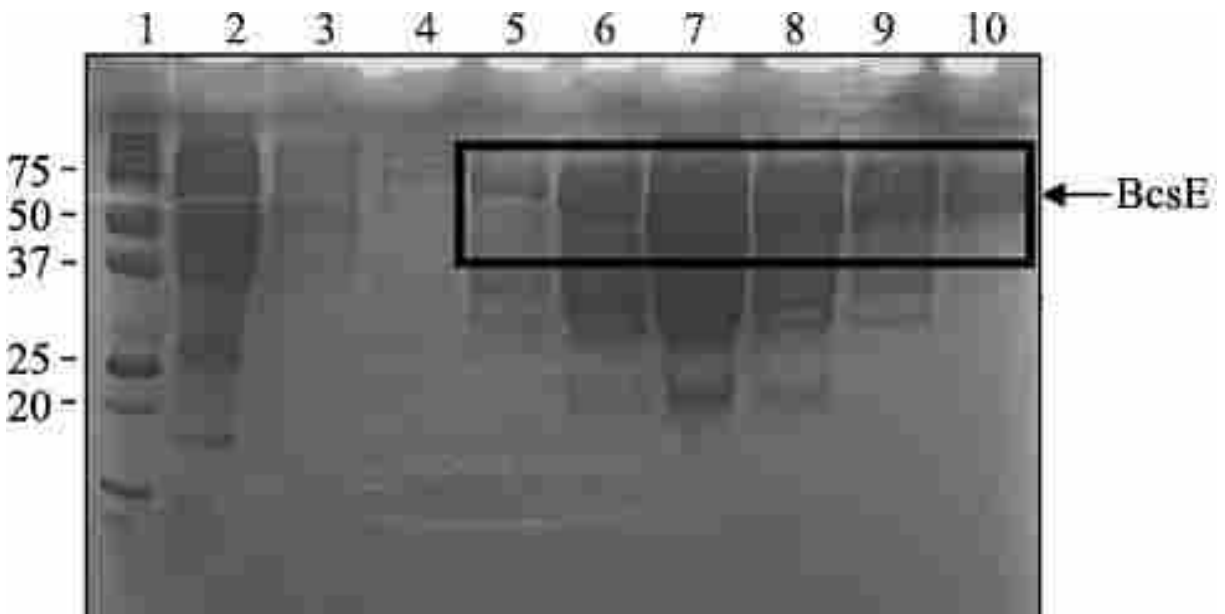


Figure 22. SDS-PAGE IMAC Purification of BcsE¹⁻⁵²³
SDS-PAGE (12% (v/v)) analysis of BcsE¹⁻⁵²³ Ni-NTA purification fractions. Lane 1, molecular-weight markers (kDa); lane 2, Ni-NTA column unbound lysate; lane 3, wash with buffer A containing no imidazole; lane 4, wash with buffer A containing 25 mM imidazole; lanes 5-10, purified BcsE¹⁻⁵²³ elution fractions from the Ni-NTA affinity chromatography column.

5.2.3 Ion Exchange Chromatography

When another purification step was required after IMAC, ion exchange or size exclusion chromatography were used to further polish the sample. IMAC alone was most often sufficient for the purification of BcsG¹⁶⁴⁻⁵⁵⁹; however, secondary steps were taken for BcsE constructs. The purpose of this second chromatographic step was to further purify the protein sample by selecting the target protein based on its charge (Refer to Table 4). Optimal parameters were determined by which conditions lead to the protein of interest being most strongly retained on the column. Buffers A (no NaCl) and B (1 M NaCl) were used to create a gradient of linearly increasing salt concentration throughout the run. An SDS-PAGE gel (12% (v/v)) was used to verify the elutions that contained the desired protein to be isolated.

The strength with which proteins bind to the resin is determined by the location and proportion of charges on the protein. Increasing the salt concentration during the exchange causes molecules with weaker ionic interactions to elute first and molecules with stronger ionic interaction to elute later in the gradient. The pH of the mobile phase (ion exchange buffer) must be between the isoelectric point (pI) of the charged molecule and the pKa of the charged group on resin. For example, a buffer with pH 6.0 may be used during cation exchange if the functional group on the resin has a pKa of 1.2 and the sample protein has a pI of 7.8. Alternatively, a buffer with pH 8.0 may be used for anion exchange when a sample protein has a pI of 6.8 and the pKa of the resin is 10.3. Figures 23 and 24 display an optimal anion exchange profile and SDS-PAGE gel analysis for BcsE¹⁻⁵²³ which shows great improvement from IMAC purification. Several anion exchanges were attempted with this construct before identifying a salt gradient that resulted in a sample of this purity.

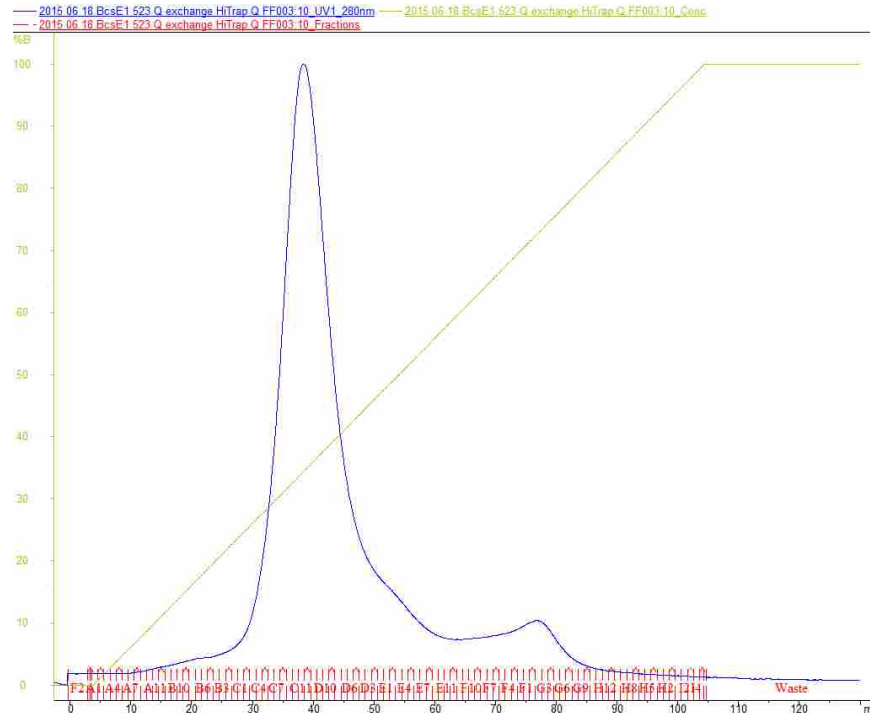


Figure 23. Anion Exchange Purification Profile for BcsE¹⁻⁵²³

The large peak at ~40 mL indicates the release of pure BcsE¹⁻⁵²³ as indicated by the blue line at 280 nm which is the optimal wavelength to observe proteins. The green line indicates increasing salt concentration. BcsE¹⁻⁵²³ begins eluting from the column at approximately 10% (w/v) NaCl.

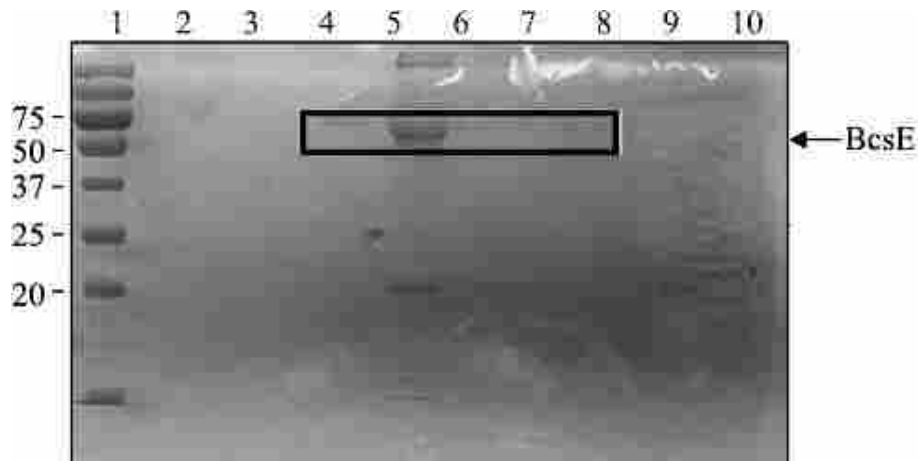


Figure 24. SDS-PAGE Anion Exchange Purification of BcsE¹⁻⁵²³

Elutions from the anion exchange profile above (Figure 23) surrounding the peak reading were analyzed by SDS-PAGE. The band within the square (mainly lane 5) above confirms that the large peak from the above profile was mostly BcsE¹⁻⁵²³ (~59kDa). Although some other contaminants are still observable in this fraction, the sample is notably more pure than after IMAC alone (Figure 22).

5.2.4 Size Exclusion Chromatography

Size exclusion chromatography (SEC) isolates proteins from a sample based on their molecular size. Larger proteins pass through the SEC resin more rapidly than smaller proteins which get detoured through small pores and channels in the resin; causing them to elute from the column more slowly. SEC was used following IMAC or ion exchange when the first purification steps did not remove a sufficient amount of contaminating proteins from the sample. This method worked equally well with all constructs when employed; however, it was most often used on BcsE constructs and not typically required for BcsG¹⁶⁴⁻⁵⁵⁹ purification. Figure 25 depicts a typical SDS-PAGE gel with bands observable at ~59kDa indicating the presence of BcsE¹⁻⁵²³. The purification steps required for each protein construct and subsequent achievable pure protein concentrations are summarized in Table 9 below.

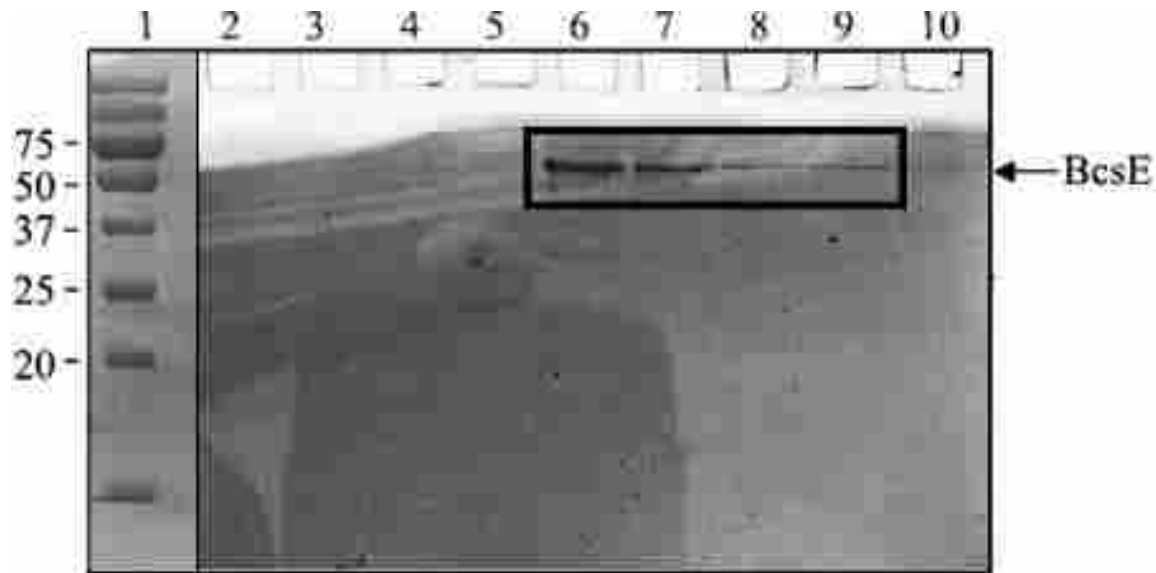


Figure 25. BcsE¹⁻⁵²³ Size Exclusion Chromatography (SEC) SDS-PAGE gel analysis
Elutions from an SEC profile surrounding a peak profile reading were analyzed by SDS-PAGE. The bands within the square above (lanes 6 - 9) represent isolated BcsE¹⁻⁵²³ (~59kDa). Although the protein bands from this profile are not as thick as those found for BcsG¹⁶⁴⁻⁵⁵⁹ IMAC purification, the sample appears to be very pure with almost no observable contaminants.

Table 9. Purification summary

Summary of purification steps required for each construct and subsequent achievable pure protein concentrations.

Construct	Typical purification route	Typical protein yield
BcsG ¹⁶⁴⁻⁵⁵⁹	IMAC	50 mg/mL
BcsE ¹⁻⁵²³	IMAC, then anion exchange and/or SEC	5 mg/mL
BcsE ¹⁻²¹⁵	IMAC, then anion exchange and/or SEC	10 mg/mL
BcsE ²²⁴⁻⁵²³	IMAC, then anion exchange and/or SEC	5 mg/mL

5.3 Functional Analysis of BcsG¹⁶⁴⁻⁵⁵⁹

5.3.1 Investigating Phosphatase Activity

Functional assays were used to test the hypothetical phosphatase activity of BcsG¹⁶⁴⁻⁵⁵⁹ according to bioinformatics homology results. First, a malachite-based colorimetric assay for detecting phosphate by-products was adapted from Mechler and colleagues (2015) and used to determine if the soluble C-terminus of BcsG was able to hydrolyze phosphate-rich NTP molecules (Figure 26). Absorbance changes read at 650 nm indicated that BcsG¹⁶⁴⁻⁵⁵⁹ is in fact capable of hydrolyzing ATP, GTP and CTP. Figure 26 illustrates a set of typical results for this functional analysis of BcsG¹⁶⁴⁻⁵⁵⁹. Over the first 15 mins of this reaction, the specific activity of BcsG¹⁶⁴⁻⁵⁵⁹ was found to be 0.332 nmol/mg/min for ATP hydrolysis, 0.411 nmol/mg/min for GTP hydrolysis and 0.760 nmol/mg/min for CTP hydrolysis.

After observing NTPase activity, metal dependency was explored since bioinformatics analysis (3DLigand) suggested that BcsG may bind magnesium, calcium or zinc; one or more of which could be used as a cofactor for activity. Metal dependency was first tested generally by incubating a BcsG¹⁶⁴⁻⁵⁵⁹ sample with the metal chelating compound EDTA. Figure 27 indicates that after EDTA exposure, BcsG¹⁶⁴⁻⁵⁵⁹ activity was virtually halved (0.366 nmol/min/mg and 0.627 nmol/min/mg of protein for EDTA-treated and wildtype, respectively).

Since BcsG¹⁶⁴⁻⁵⁵⁹ was shown to require a metal cofactor, a metal dependency profile was next generated to determine which metals were ideal as cofactors for hydrolysis of CTP. Samples were incubated with buffers containing 20 mM metal concentrations, without first using a chelating agent on the samples, as conducted by Kennely *et al.* (1993). Table 10 displays all pertinent values for BcsG¹⁶⁴⁻⁵⁵⁹ activity during the metal dependency analysis.

Common biological metal cofactors tested included magnesium, manganese, nickel, calcium and zinc. Specific activity values displayed in Table 10 indicate that the rate of CTP hydrolysis was highest in the presence of magnesium at a rate of $2.03 \times 10^{-1} \pm 0.008$ nmol/mg (of protein) /min. Manganese-reconstituted BcsG¹⁶⁴⁻⁵⁵⁹ had nearly as much activity as magnesium with a specific activity of $1.99 \times 10^{-1} \pm 0.006$ nmol/mg/min suggesting that it is also a suitable cofactor for BcsG¹⁶⁴⁻⁵⁵⁹ activity. For samples containing nickel and calcium, BcsG¹⁶⁴⁻⁵⁵⁹ activity dropped 2- and 4-fold, respectively from optimal reaction rates to $0.093 \times 10^{-1} \pm 0.005$ nmol/mg/min and $0.062 \times 10^{-1} \pm 0.004$ nmol/mg/min, respectively. The BcsG S116A mutant (mutated residue predicted to be an active site residue) also had nearly full activity (0.169 ± 0.013 nmol/mg/min) compared to the wild-type BcsG¹⁶⁴⁻⁵⁵⁹ construct with a magnesium cofactor.

Next, a variation of the same colorimetric assay was used to determine a pH profile to identify the optimal pH/buffer for BcsG¹⁶⁴⁻⁵⁵⁹ activity with a CTP substrate. Buffers for testing the BcsG¹⁶⁴⁻⁵⁵⁹ pH profile were designed to have overlapping pH values to create a continuous plot of activity across the pH range tested. A broad peak in activity between 6.5 and 8.5 can be observed in Figure 28 below, while there is only about half maximal activity at pH's 5.5 and 9.

Since BcsG may somehow be involved in PEA labelling of cellulose, PEA was selected as the first PEA pathway substrate to test. The reaction of PEA and the PEA phosphatase

PHOSPHO1 (Refer to Figure 4) results in the release of a phosphate; which meant that this reaction could be monitored by the already optimized colorimetric assay described above. As depicted in Figure 29, there was a similar PEA hydrolysis rate (0.136 nmol/mg/min) compared to CTP hydrolysis by BcsG¹⁶⁴⁻⁵⁵⁹ (0.237 nmol/mg/mL).

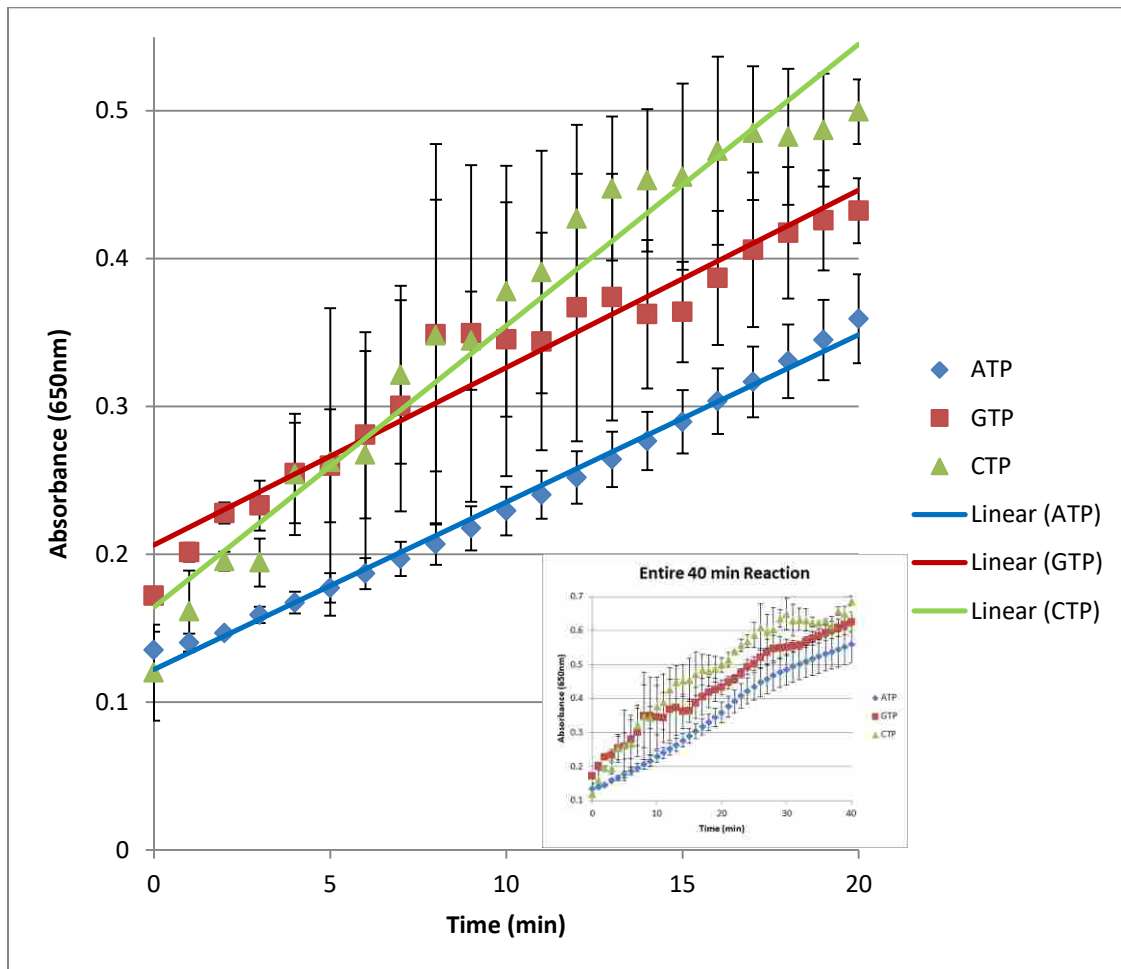


Figure 26. Functional assay of BcsG¹⁶⁴⁻⁵⁵⁹ with ATP, GTP and CTP substrates
Assays were conducted with 50mM buffers, 10 μ M enzyme and with 100 μ M substrate. Hydrolysis appears to be occurring on all NTPs tested with the greatest observed activity for CTP followed by GTP then ATP. These measurements represent an average of four replicates. Large panel shows a close-up of the main reactions which occurred in the first 20 mins. Small panel shows reactions as they slowed over entire 40 min assay.

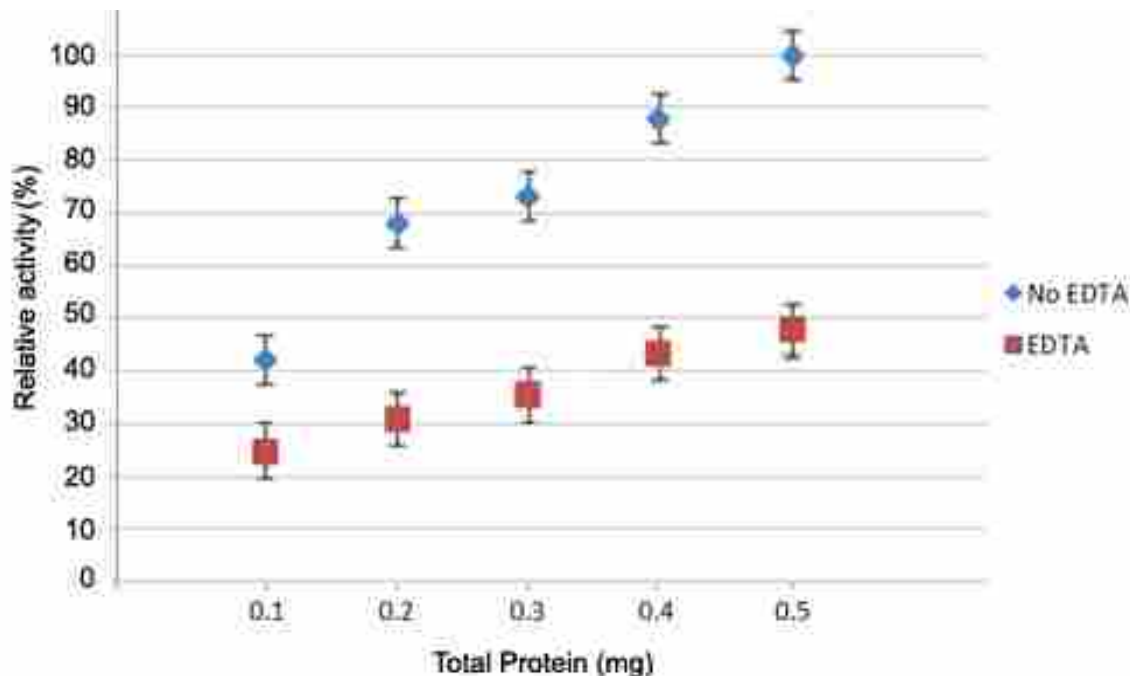


Figure 27. Functional assay of BcsG¹⁶⁴⁻⁵⁵⁹ with CTP substrate and EDTA inhibitor
 Assays were conducted with 50mM buffers, 10 μ M enzyme and with 100 μ M substrate. CTP hydrolysis is impaired by the presence of 2mM EDTA (red) when compared to the wild type sample (blue). These measurements represent an average of four replicates.

Table 10. Substrate profiling of the metal dependency of BcsG

Final metal concentrations assayed at 2mM. Assays were conducted with 50mM buffers, 10 μ M enzyme and with 100 μ M substrate. These measurements represent an average of four replicates.

Metal*	Specific activity (nmol/mg/min)**
Mg	0.213 +/- 0.008
Mn	0.199 +/- 0.006
Ni	0.093 +/- 0.005
Ca	0.062 +/- 0.004
Zn	0.001 +/- 7.31 x 10 ⁻⁵
S116A (Mg)	0.169 +/- 0.013

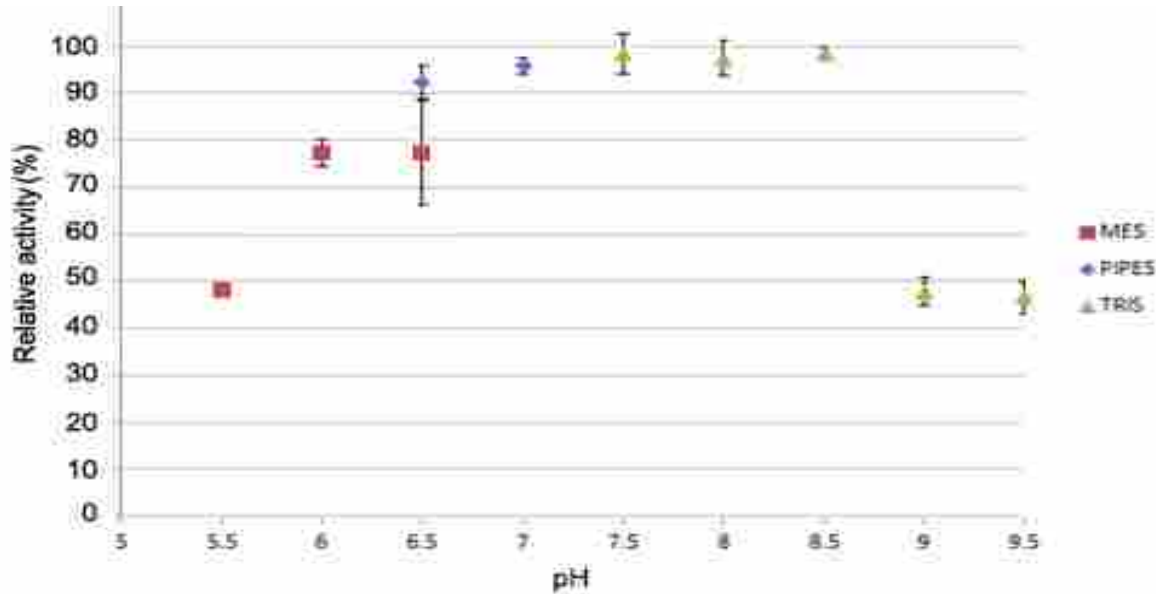


Figure 28. pH profile for BcsG¹⁶⁴⁻⁵⁵⁹ with CTP in various biological buffers

Assays were conducted with 50mM buffers, 10 μ M enzyme and with 100 μ M substrate. Optimal activity appears to be occurring at approximately pH 7.5 with Tris buffer. These measurements represent an average of four replicates.

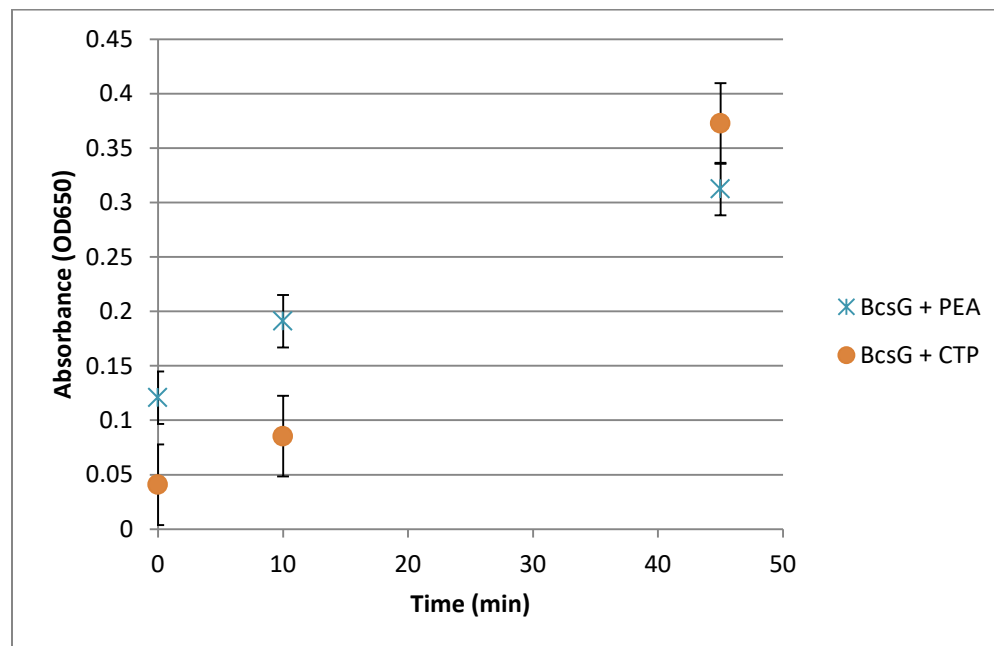


Figure 29. Functional assay of BcsG¹⁶⁴⁻⁵⁵⁹ with CTP and PEA substrates

There appears to be similar PEA hydrolysis occurring compared to CTP hydrolysis by BcsG¹⁶⁴⁻⁵⁵⁹ (0.136 nmol/mg/min and 0.237 nmol/mg/mL, respectively). These measurements represent an average of four replicates.

5.3.2 Localization of BcsG¹⁻⁵⁵⁹ Using GFP as a Reporter Protein

Using a modified protocol from Dinh and Bernhardt (2011), a BcsG¹⁻⁵⁵⁹-GFP hybrid protein was used to determine the localization of the soluble C-terminus of BcsG. GFP, which was engineered onto the C-terminus of BcsG, has several properties that make it a novel protein for periplasmic protein localization. The cleared lysate for the GFP positive control and BcsG¹⁻⁵⁵⁹-GFP hybrids in various buffers can be viewed on an SDS-PAGE gel in Figure 30 according to their molecular weights. Bacteria expressing the BcsG¹⁻⁵⁵⁹-GFP hybrid protein (right) did not fluoresce compared to the GFP control (left) when observed under long-wave UV light (Figure 31) even though qualitatively similar amounts of bacterial growth occurred on both plates. As seen in Figure 32, the untreated BcsG¹⁻⁵⁵⁹-GFP hybrid cleared lysate had a similar fluorescence value to the non-hybrid BcsG¹⁶⁴⁻⁵⁵⁹ construct (14131 and 15199 rfu, respectively). However, when acid-base treated, BcsG¹⁻⁵⁵⁹-GFP hybrid fluorescence was restored markedly (25546 rfu) indicating that the environment was responsible for initial inactivation of the hybrid GFP strongly suggesting that the cellular localization of the soluble portion of the protein was within the periplasm of *E. coli*.

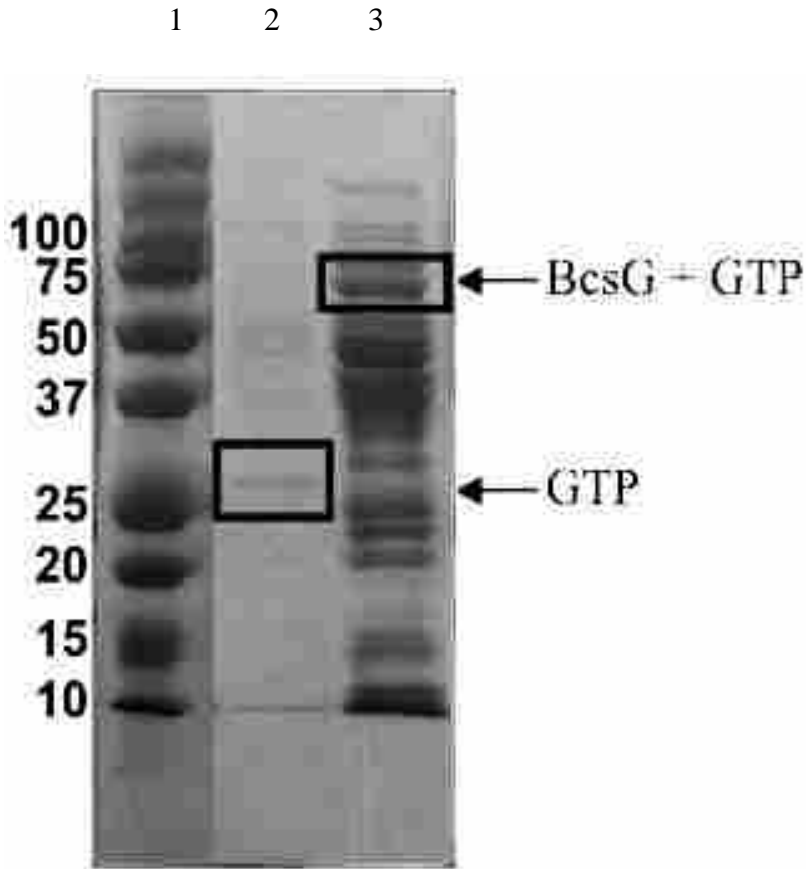


Figure 30. SDS-PAGE gel for GFP and BcsG¹⁻⁵⁵⁹-GFP hybrid cleared lysate samples
The cleared lysates of GFP and BcsG¹⁻⁵⁵⁹ : GFP hybrid protein samples can be seen in the boxes above based on expected molecular weights.



Figure 31. Growth Plates for GFP (+) Control and BcsG-GFP Hybrid
The BcsG¹⁻⁵⁵⁹-GFP hybrid protein (right) does not fluoresce like the GFP control (left) does under UV light. Healthy growth can be seen on both plates.

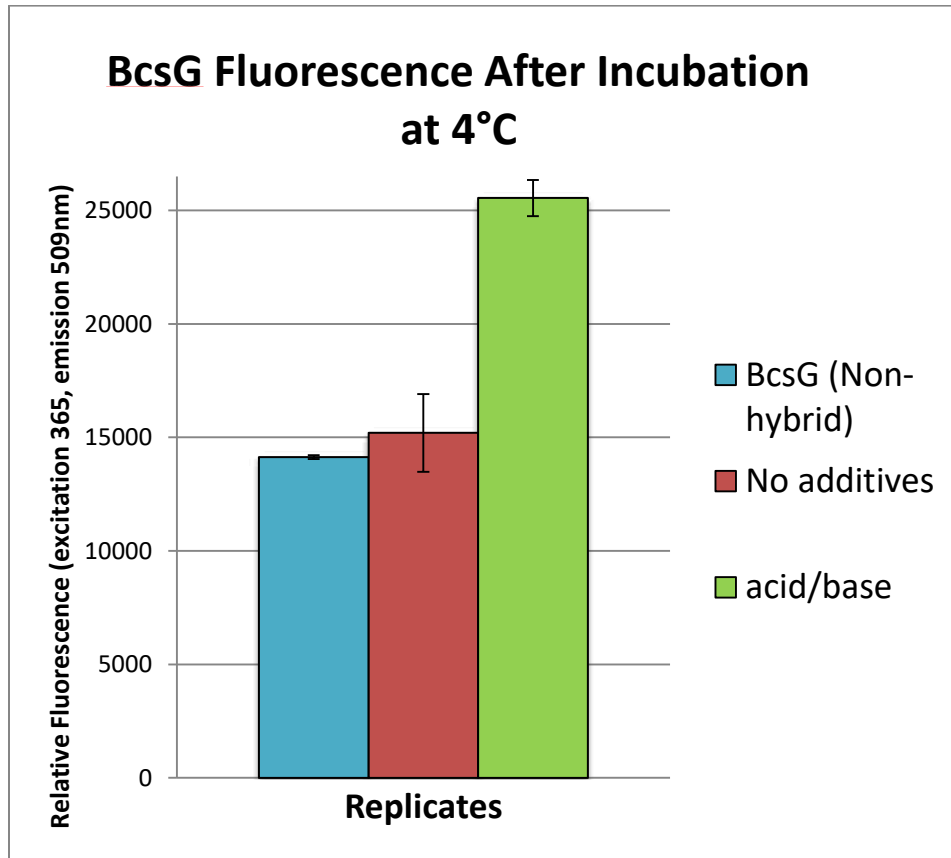


Figure 32. Excitation-emission for BcsG¹⁻⁵⁵⁹ cleared lysate samples

Relative fluorescence units for the cleared lysate of non-hybrid (-) BcsG (left) (14131 rfu) is very similar to the cleared lysate of hybrid BcsG¹⁻⁵⁵⁹-GFP (middle) (15199 rfu) that has not been acid/base treated. Excitation-emission for the cleared lysate of hybrid BcsG¹⁻⁵⁵⁹-GFP that has been acid/base treated (right) (25546 rfu) has the highest reading indicating that GFP fluorescence was restored.

5.4 Protein Crystallization

5.4.1 Native Crystallization of BcsG¹⁶⁴⁻⁵⁵⁹

Proteins are typically large and flexible macromolecules that infrequently self-assemble into the regularly repeating configurations necessary for crystallization. Generally, growing well-diffracting protein crystals requires separation of protein molecules from a solution resulting in self-assembly into a periodic crystal matrix. This can be achieved in the lab by introducing varying ratios of highly concentrated recombinant protein to conditions with assortments of known additives and precipitating reagents to reduce protein solubility. Assuming that nucleation sites are present, whether spontaneously formed or introduced by crystal streaking, protein crystals will grow.

Two crystallization screens were attempted for BcsE¹⁻⁵²³ (in Tris buffer) and BcsG¹⁶⁴⁻⁵⁵⁹ (in phosphate buffer) for each MCSG (1 – 4) screens and custom PEG screen at varying drop ratios and protein concentrations. Prior to loading screens, BcsG¹⁶⁴⁻⁵⁵⁹ was an ideal protein as it purified very quickly (and to a high degree) by a single IMAC, concentrated easily up to 50 mg/mL, and could be stored at 4°C for weeks without precipitating. However, BcsE constructs required a minimum of two (often three) purification steps, concentrated occasionally up to ~5 - 10 mg/mL, and precipitated within days when stored at 4°C. Since BcsE constructs were so much less stable than BcsG¹⁶⁴⁻⁵⁵⁹, attempts to crystallize it were abandoned in order to focus efforts on the more promising BcsG¹⁶⁴⁻⁵⁵⁹ construct.

Six crystal isoforms arose from the initial crystal screens with BcsG¹⁶⁴⁻⁵⁵⁹ purified in phosphate buffer. However, after further investigation of six of these crystals, through a combination of dyes (Iziti) and exposure to X-rays, the crystals grown in phosphate buffer were determined to be salt in composition. Re-optimization of BcsG¹⁶⁴⁻⁵⁵⁹ expression and purification

with Tris buffers did not take long, and then the screens mentioned above were repeated.

Following this, many protein crystal isoforms were amassed (Figures 33 - 36) and the conditions predicted to be best by visual inspection were multiplied upon in hanging drop expansion plates.

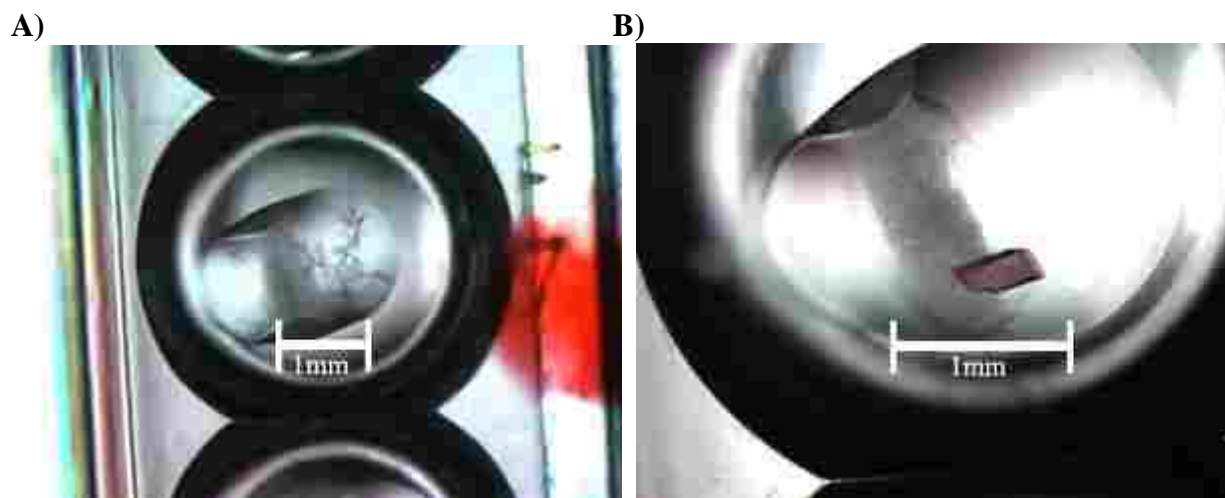


Figure 33. Tris buffer purified BcsG¹⁶⁴⁻⁵⁵⁹ crystal hits from the MCSG-4T F7 condition
Two crystal isoforms were obtained for BcsG¹⁶⁴⁻⁵⁵⁹ (suspended in 50 mM Tris, 50 mM NaCl, 5 mM MgCl₂ buffer) in condition MCSG-4T F7 (0.2 M Magnesium Chloride, 0.1 M MES:NaOH pH 6.5, 25% (w/v) PEG 4000). A) A starburst/urchin isoform grown in sitting drop vapour diffusion plates (96 wells). B) Rhombohedral crystal grown in sitting drop vapour diffusion plates (96 wells).

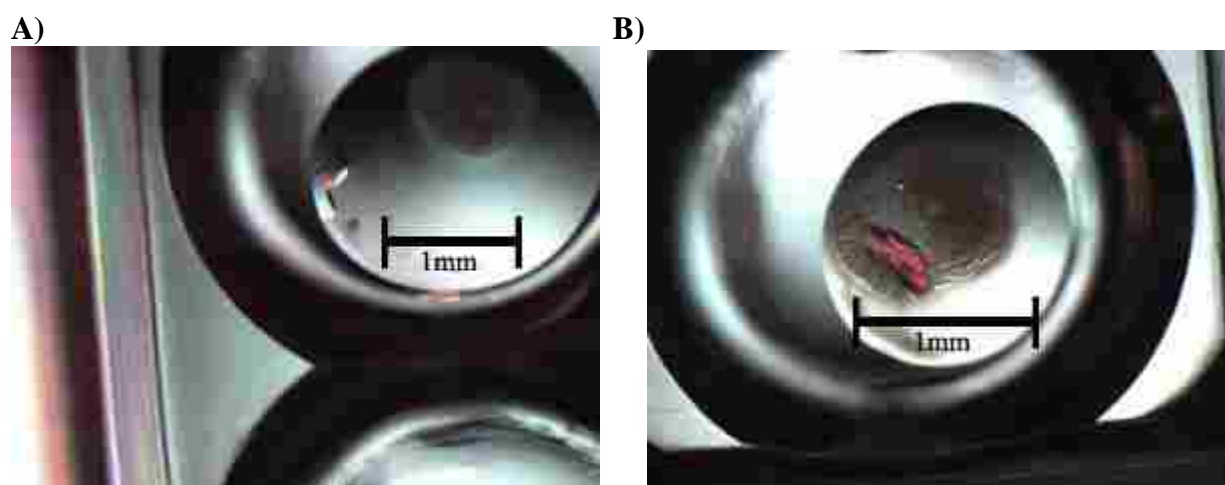


Figure 34. Tris buffer purified BcsG¹⁶⁴⁻⁵⁵⁹ crystal hits from the MCSG-4T G11 condition
Four crystal isoforms were obtained for BcsG¹⁶⁴⁻⁵⁵⁹ (suspended in 50 mM Tris, 50 mM NaCl, 5 mM MgCl₂ buffer) in condition MCSG-4T G11 (0.1 M HEPES:NaOH pH 7.5 / 20% (w/v) PEG 4000 / 10% (v/v) 2-Propanol). A) Large, L-shaped blocky crystal with no precipitation. B) Blocky crystals growing from a central nucleation point out of precipitation in sitting drop vapour diffusion plates (96 wells).

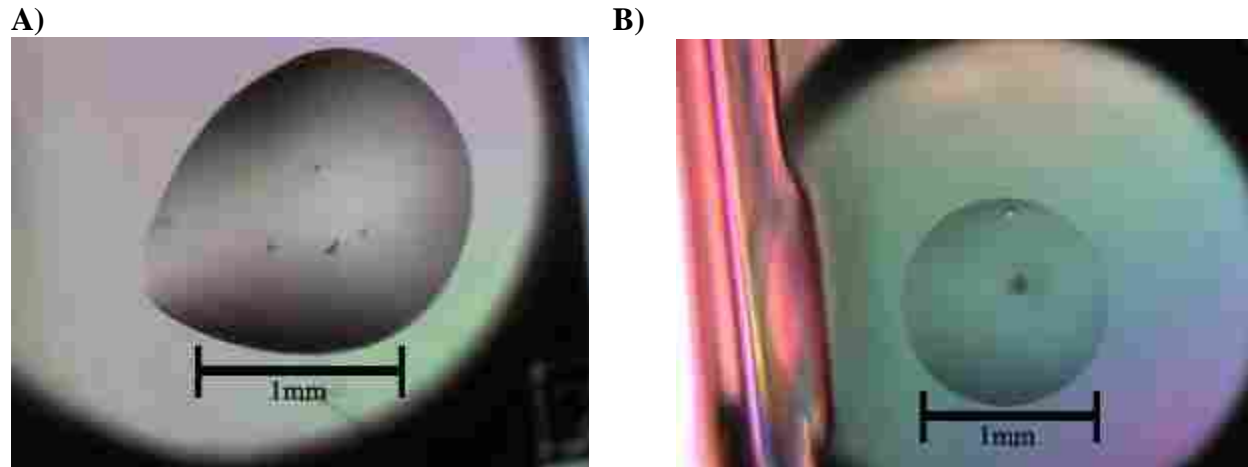


Figure 35. Tris buffer purified BcsG¹⁶⁴⁻⁵⁵⁹ crystal hits from the MCSG-1T D1 condition
Two crystal isoforms were obtained for BcsG¹⁶⁴⁻⁵⁵⁹ (suspended in 50 mM Tris, 50 mM NaCl, 5 mM MgCl₂ buffer) in condition MCSG-1T D1 (0.2M ammonium sulfate, 0.1M bis-tris: HCl pH 6.5, 25% (v/v) PEG 3350). A) Small, irregular prism. B) Small football-shaped crystal in sitting drop vapour diffusion plates (96 wells).

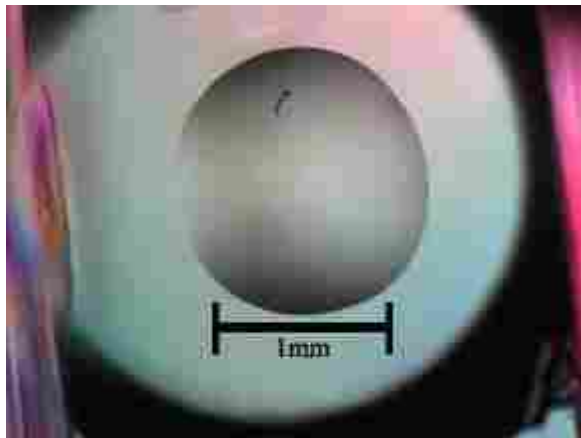


Figure 36. Tris buffer purified BcsG¹⁶⁴⁻⁵⁵⁹ crystal hits from the MCSG-3T A5 condition
One crystal isoform was obtained for BcsG¹⁶⁴⁻⁵⁵⁹ (suspended in 50 mM Tris, 50 mM NaCl, 5 mM MgCl₂ buffer) in condition MCSG-3T A5 (0.1M MES:NaOH pH 6.0, 1.26M ammonium sulfate). Displayed long, sharp prism-shaped crystal in sitting drop vapour diffusion plates (96 wells).

5.4.2 Crystal Conditions Expansion Plates

Expansion plates (layouts depicted in Figure 11) of the most promising conditions with varying protein concentrations yielded numerous diffraction quality crystals (Figure 37). Protein concentrations yielding diffraction quality crystals ranged from 20 mg/mL to 50 mg/mL with a linear trend observed between protein concentration and crystal formation time. Crystals that formed from 50 mg/mL samples took approximately four weeks to reach full size. Crystals resulting from 40 mg/mL took six weeks, 30 mg/mL took about eight weeks and 20 mg/mL took about 10 weeks to reach the same observable size. All of these crystals formed without seeding being required and typically occurred at a 1:1 ratio between protein sample and conditions tested. Large crystals formed around or just below neutral pH. Crystals sent to the Canadian Light Source (CLS) for X-ray diffraction frequently resolved to $\sim 2\text{\AA}$ with the sharpest resolutions reaching as low as 2.1\AA with native protein crystals from conditions MCSG-4T F7 (0.2 M Magnesium Chloride, 0.1 M MES:NaOH, pH 6.5, 25% (w/v) PEG 4000) and G11 (0.1 M HEPES:NaOH, pH 7.5, 20% (w/v) PEG 4000, 10% (v/v) 2-Propanol). Representative crystal isoforms can be seen in Figure 37 (Panels A-D) and a sample diffraction pattern can be observed below in Figure 38.

Table 11. Diffraction statistics for native BcsG¹⁶⁴⁻⁵⁵⁹ protein crystals*

Diffraction source	CMCF beamline 08B1-1
Wavelength (Å)	0.97951
Temperature (K)	100
Detector	MarMosaic mx300
Crystal-detector distance (mm)	350mm
Rotation range per image (°)	1.0
Total rotation range (°)	360
Exposure time per image (s)	12.0
Space group	P2 ₁
a, b, c (Å)	54.4, 79.7, 87.1
α, β, γ (°)	90.0, 99.1, 90.0
Mosaicity (°)	0.271
Resolution range (Å)	44.58 - 2.13 (2.37 - 2.13)**
Total No. of reflections	266810 (11068)**
No. of unique reflections	37856 (7787)**
Completeness (%)	100.0 (70.4)**
Redundancy	7.04 (1.42)**
I/σ(I)	14.7 (5.3)**
R _{meas}	0.093 (4.15)**

* Data collection and processing from the Canadian Light Source in Saskatchewan, Canada.

**Values for the outer shell are given in parentheses.

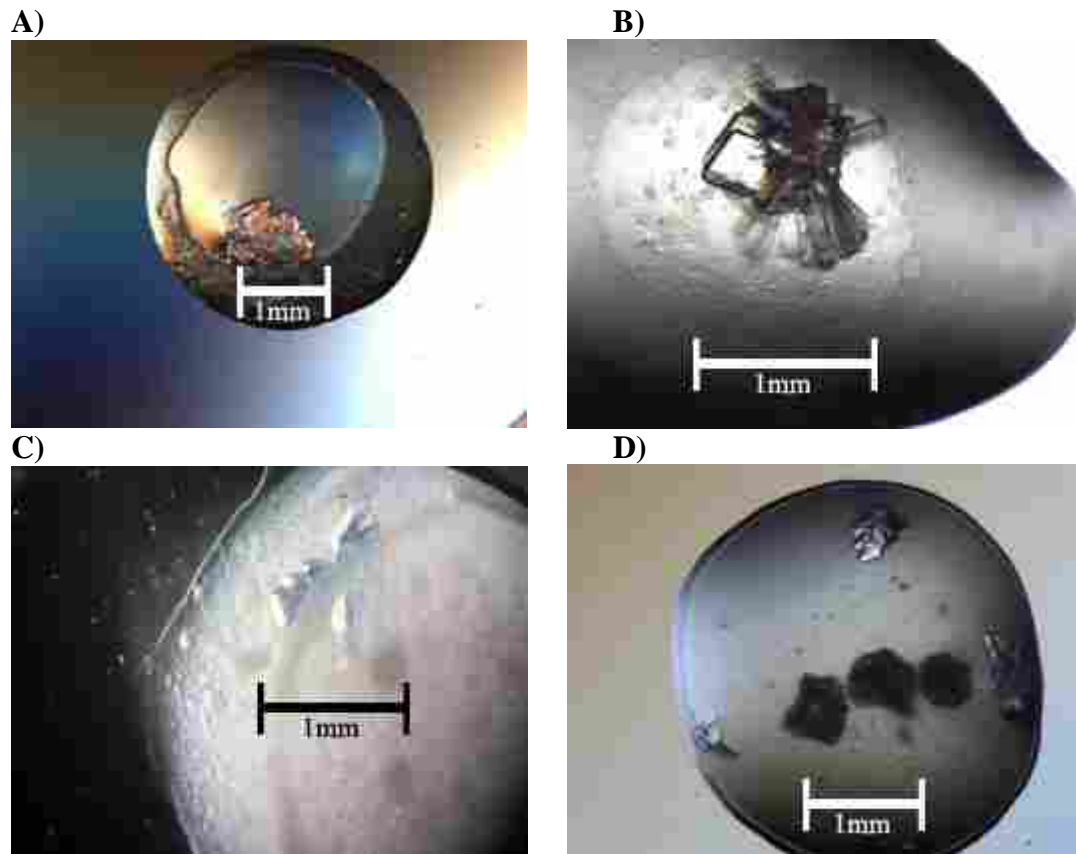


Figure 37. Representative large crystal isoforms from expansion trials

(A) 0.2 M Sodium Chloride, 0.1 M Bis-Tris:HCl, pH 6.5, 25% (w/v) PEG 3350 in a 3:1 ratio of BcsG¹⁶⁴⁻⁵⁵⁹ at 50mg/mL. **(B)** 0.2 M Magnesium Chloride, 0.1 M MES:NaOH, pH 6.5, 25% (w/v) PEG 4000 in a 1:1 ratio of BcsG¹⁶⁴⁻⁵⁵⁹ at 50mg/mL. **(C&D)** 0.1 M HEPES:NaOH, pH 7.5, 20% (w/v) PEG 4000, 10% (v/v) 2-Propanol in a 1:1 ratio of BcsG¹⁶⁴⁻⁵⁵⁹ at 50mg/mL.

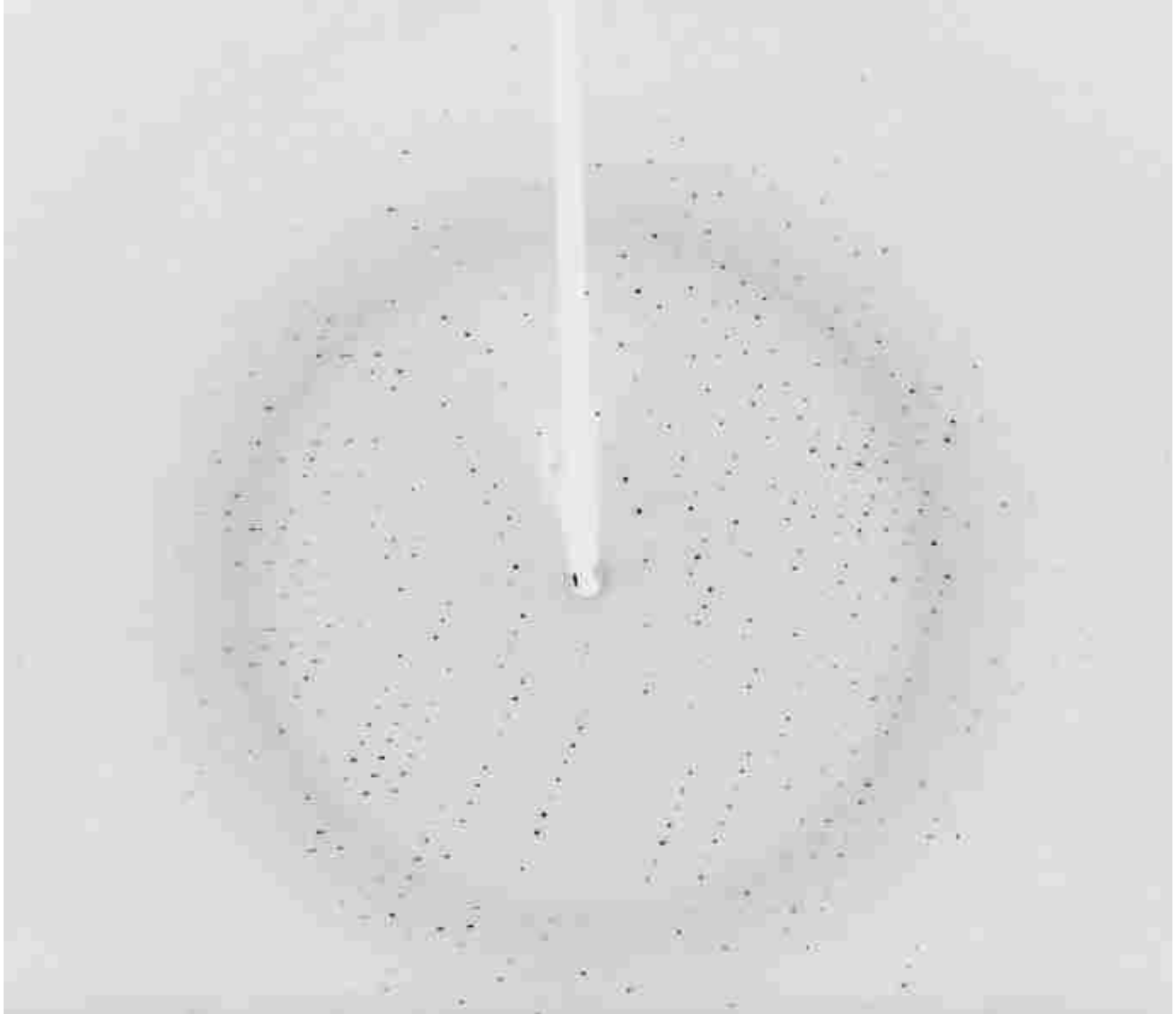


Figure 38. Representative native BcsG¹⁶⁴⁻⁵⁵⁹ crystal X-ray diffraction pattern
This image offers a preview of the quality of the BcsG¹⁶⁴⁻⁵⁵⁹ X-ray diffractions as recorded by the CLS. Diffraction spots are round and sharp indicating high resolution and quality.

5.4.3 Heavy Metal BcsG¹⁶⁴⁻⁵⁵⁹ Crystal Soaks

Phyre2 determined that the nearest structural homologs to BcsG¹⁶⁴⁻⁵⁵⁹ have ~15% identity at the amino acid level to APs. While molecular replacement was attempted with these poor models (by Dr. Weadge), no solutions could be obtained. To circumvent this problem, the use of experimental phasing with heavy metals was explored. Experimental phasing experiments were attempted both by soaking native crystals in heavy atom solutions (NaBr or KBr) and by SeMet replacement in specialized media. Although several NaBr and KBr heavy metal soaks were performed (Refer to Figure 39) and sent for X-ray diffraction analysis at the CLS, the diffraction resolution was low (all were above 3Å) and the spots on the diffraction pattern were smeared, so auto processing could not index the spots properly. As a result, the heavy atom substructure could not be solved to determine the phase for the soaked crystals.

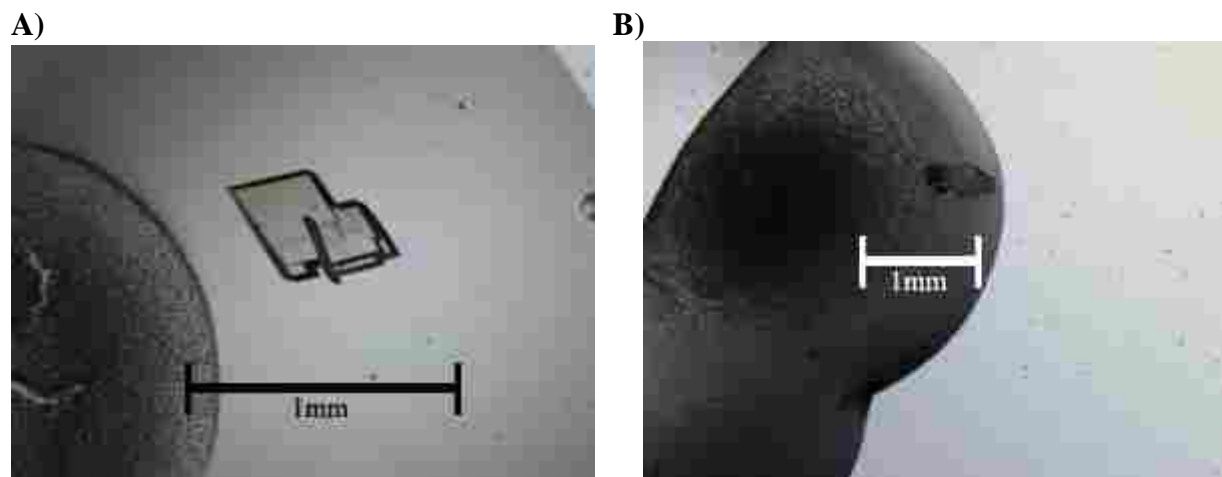


Figure 39. Crystals Soaked in Heavy Atom Solutions

Representative crystal isoforms from expansion plates used for heavy metal soaking in NaBr or KBr. **A)** From MCSG-4T G11 (40mg/mL), 2.5M KBr soaked for 22 secs, Cryoprotectant 35% PEG4000, Stored in liquid nitrogen. **B)** From MCSG-4T G11 (40mg/mL), 2.5M NaBr soaked for 10 secs, Cryoprotectant 35% PEG4000, Stored in liquid nitrogen.

5.4.4 Selenomethionine Expression, Purification and Crystallization Attempts

According to Hendrickson (1991), the introduction of selenium into a recombinant protein can be a superbly effective tool for experimentally solving the phase problem in crystallography. Figure 40 shows that BcsG¹⁶⁴⁻⁵⁵⁹ expression yields were lower with the SelenoMet minimal media compared to expression in super broth. For this reason, all protein pellets were pooled and a single IMAC purification was conducted in order to maximize the final concentration of pure protein for crystallization. After the first IMAC, the wash 1 fraction (containing 25 mM imidazole) was moderately pure and contained the most BcsG¹⁶⁴⁻⁵⁵⁹. Therefore, a second IMAC was conducted on the wash 1 fraction with a longer incubation with Ni-NTA resin in an attempt to allow enhanced binding of the target protein. BcsG¹⁶⁴⁻⁵⁵⁹ eluted in the first wash once again; detected at ~45kDa (as seen in Figure 41). The fraction appeared sufficiently pure, so it was concentrated and prepared for use in a hanging drop crystal plate.

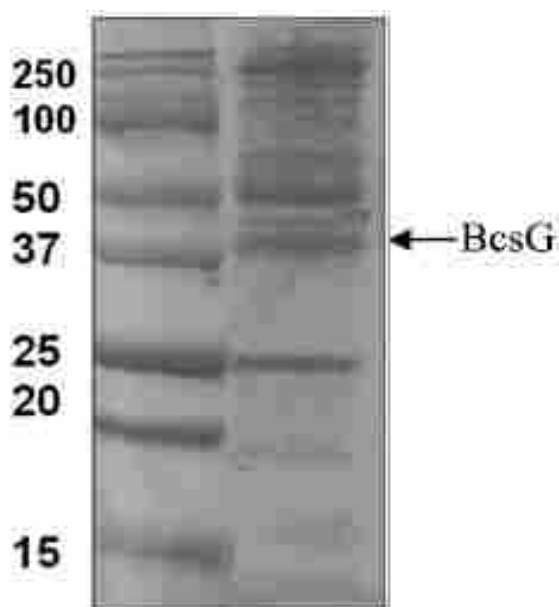


Figure 40. SDS-PAGE expression gel for BcsG¹⁶⁴⁻⁵⁵⁹ grown in SeMet minimal media
SDS-PAGE analysis of BcsG¹⁶⁴⁻⁵⁵⁹ protein expressed in SeMet minimal media; incubated at 22°C and 200 rpm shaking for 16 h induced with 1 mM IPTG. BcsG¹⁶⁴⁻⁵⁵⁹ indicated with arrow at 45 kDa. Lane 1, molecular-weight markers (kDa); lane 2, SDS-PAGE fraction of expressed BcsG¹⁶⁴⁻⁵⁵⁹ sample.

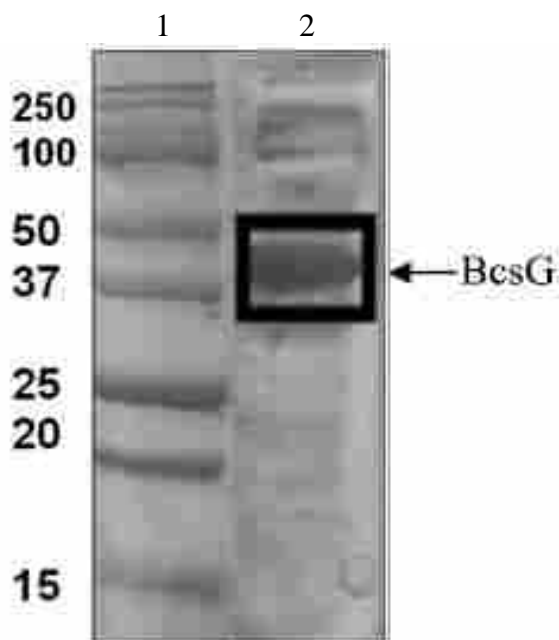


Figure 41. SeMet concentrated IMAC purification samples
SDS-PAGE (12% (v/v)) analysis of BcsG¹⁶⁴⁻⁵⁵⁹ Ni-NTA purification fractions. Lane 1, molecular-weight markers (kDa); lane 2, purified BcsG¹⁶⁴⁻⁵⁵⁹ obtained from wash 1 fraction (~45kDa).

The first expansion plate attempted with BcsG¹⁶⁴⁻⁵⁵⁹ grown in SeMet minimal media was prepared in the same way as with native BcsG¹⁶⁴⁻⁵⁵⁹ at 40mg/mL and 50 mg/mL for conditions MCSG-4T F7 and G11 (Refer to Figure 11). After a month with no crystal hits, the seed bead protocol was followed and freshly prepared undiluted seed stock (from a previous native BcsG¹⁶⁴⁻⁵⁵⁹ crystal) was passed through the hanging drops for both buffers tested in the expansion plate. Crystals used for seeding can be viewed in Figure 42. Next, 1 ul of 2 mM DTT was added to each drop in a separate reservoir for both buffer conditions to prevent intermolecular disulfide bond formation which may have an impact on crystallizability (Wingfield, 2015). After these treatments, three reservoirs were left untreated, two contained added seed stock, and one contained DTT for both expansion plate conditions. The above protocol was repeated two more times with varied protein, PEG and DTT concentrations as outlined in Table 3.

The second expansion setup resulted in microcrystals after 3 weeks with 40 mg/mL BcsG¹⁶⁴⁻⁵⁵⁹ in condition MCSG-4T G11 (0.1 M HEPES:NaOH, pH 7.5, 20% (w/v) PEG 4000, 10% (v/v) 2-Propanol) with drops at 1:1 and 2:1 protein to buffer ratios. Microcrystals from the second SeMet attempt did not grow any further, and nothing has arisen from the first and third plates to date; although all plates are periodically checked as they are still being incubated at 18°C (Figure 43).

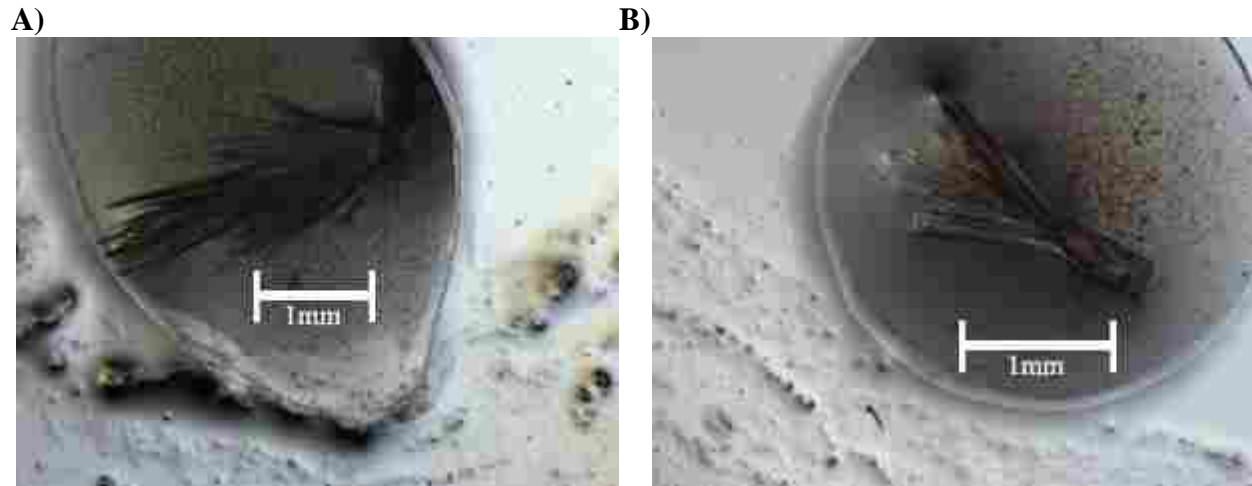


Figure 42. Native BcsG¹⁶⁴⁻⁵⁵⁹ crystals used to streak seed through first SeMet conditions
Representative native BcsG¹⁶⁴⁻⁵⁵⁹ crystal isoforms (40mg/mL) from expansion plates grown in MCSG-4T G11 (0.1 M HEPES:NaOH, pH 7.5, 20% (w/v) PEG 4000, 10% (v/v) 2-Propanol) used for seeding SeMet crystal conditions. **A)** Representative needle-like protein crystal isoform. **B)** Representative rectangular prism-shaped protein crystal isoform.

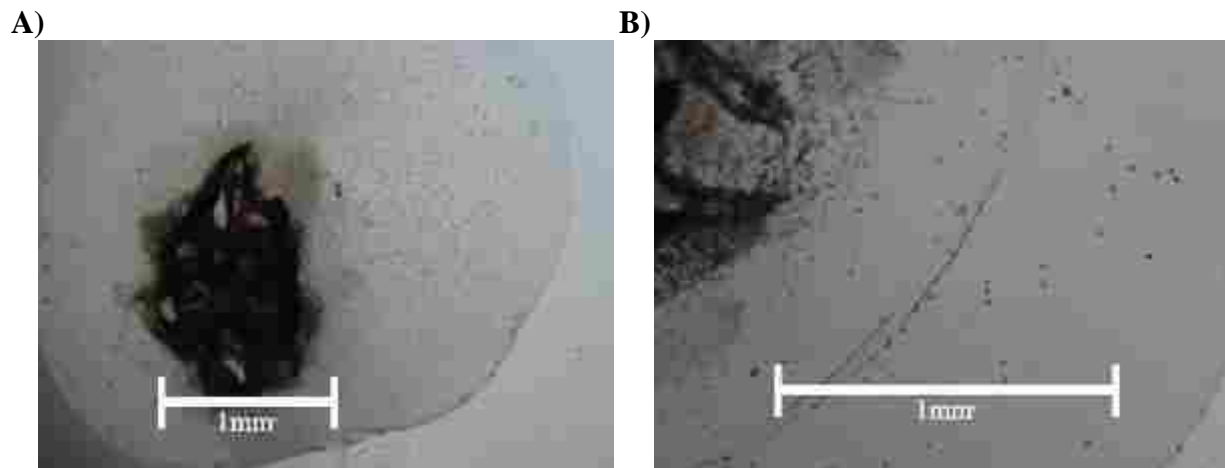


Figure 43. SeMet BcsG¹⁶⁴⁻⁵⁵⁹ microcrystals from second SeMet attempt
SeMet microcrystals observed in precipitation from expansion plate with MCSG-4T G11 (0.1 M HEPES:NaOH, pH 7.5, 20% (w/v) PEG 4000, 10% (v/v) 2-Propanol) in a 1:1 protein to buffer ratio of BcsG¹⁶⁴⁻⁵⁵⁹ at 40mg/mL.

6. DISCUSSION

6.1 Objective 1: Bioinformatics

Concatenated results from open-source bioinformatics tools provided insights into the theoretical properties of BcsE and BcsG constructs prior to experimentation. Amino acid sequence alignments using ClustalW, Pubmed, 3DLigand and Phyre2 identified important residues and hypothetical functions based on protein regions homologous to other proteins. The Phyre2 results for the top five suggested BcsE¹⁻²¹⁵ and BcsE²²⁴⁻⁵²³ homologs are displayed in Figures 14 and 15, respectively. The physical properties and important residues suggested by bioinformatics resources were used when determining mutants, truncated constructs and as starting points for defining expression and purification conditions. Functionality and localization were tested by various assays with BcsG which are discussed in greater detail below. The Phyre2 suggested homologous PEA transferase was noted several times in the results, because BcsG is hypothesized to be involved for PEA labeling of bacterial cellulose. Implications of bioinformatics results are discussed in greater detail throughout subsequent sections.

6.2 Objective 2: Protein Expression and Purification

6.2.1 Protein Expression

The conditions established for optimal expression of constructs were adequate for remaining within physiologically relevant parameters. Final conditions used for expression included: 1 mM IPTG, incubation at 22°C while shaking at 200 rpm for 16 hrs. The BcsG¹⁶⁴⁻⁵⁵⁹ yields achieved (~40 mg/L culture) were ample for downstream functional and structural analyses. However, recombinant protein yields for BcsE constructs (BcsE¹⁻⁵²³: ~4 mg/L culture, BcsE¹⁻²¹⁵: ~8 mg/L culture, BcsE²²⁴⁻⁵²³: ~4 mg/L culture) could have potentially been satisfactory for functional or structural testing if they were made a top priority. Due to tight time restrictions, BcsE constructs were not focused on nearly as much as BcsG during this project since BcsG was

working so well. Fang and colleagues (2014) have also suggested that the full BcsE construct is very difficult to work with compared to other proteins due to its instability; but they were able to prove that it binds c-di-GMP even at comparatively low concentrations (>95% purity at ~1 mg/mL). Future work with BcsE should likely put more time towards optimizing the C- and N-terminal constructs to possibly find conditions that increase stability and allow elucidation of the N-terminal function. Optimization of each termini would also be vital for structural analyses by crystallization as higher protein concentrations are required to crystallize proteins and the BcsE structure is currently unknown.

6.2.2 Protein Purification

Recombinant DNA techniques permit the construction of fusion proteins in which specific affinity tags are added to the protein sequence of interest. Ni-NTA resin was used for IMAC, which was the first purification step in isolating all protein constructs in this project. IMAC was made possible due to all BcsE and BcsG constructs being engineered with six consecutive histidine residues (6xHis tag); which is a common affinity tag to facilitate recombinant protein binding to Ni-NTA.

Protein crystallization has shown frequent success with small 6xHis tags and larger polypeptide tags alike (Segelke *et al.*, 2004). As demonstrated for a wide variety of proteins (including enzymes, transcription factors, and vaccines) by statistical analysis comparisons between tagged and untagged crystal structures, affinity tags have been shown repeatedly not to interfere with the structure or function of purified recombinant proteins; aside from a few residues directly neighboring the terminal tag (Carson *et al.*, 2007). Therefore, we expected that histidine tags used for purification with all protein constructs would not hinder downstream analyses. Successful crystallization of BcsG¹⁶⁴⁻⁵⁵⁹ supports this hypothesis.

If another purification step is required after IMAC, ion exchange or size exclusion chromatography can be used to further polish the sample. IMAC alone was sufficient for the purification of BcsG¹⁶⁴⁻⁵⁵⁹; however, anion exchange was frequently used for BcsE constructs. Anion exchange was conducted on an FPLC by creating a gradient of linearly increasing salt concentration.

Size exclusion chromatography (SEC) worked best after one or more previous purification steps (IMAC and/or ion exchange) because it works optimally when there is some separation in size between the protein of interest and sample impurities. SEC was used following IMAC and/or ion exchange mainly for BcsE constructs since initial purification steps did not sufficiently remove contaminating proteins from the sample. This method worked equally well with BcsG¹⁶⁴⁻⁵⁵⁹ when employed; however, upon optimization of BcsG¹⁶⁴⁻⁵⁵⁹ IMAC protocol, SEC was not required for BcsG¹⁶⁴⁻⁵⁵⁹. Figure 25 depicts a typical chromatogram with a large peak indicating elution of BcsG¹⁶⁴⁻⁵⁵⁹. In summary, BcsE and BcsG were successfully expressed, purified and concentrated to a sufficient degree for meaningful analysis (BcsG¹⁶⁴⁻⁵⁵⁹ = 50 mg/mL, BcsE¹⁻⁵²³ = 5 mg/mL, BcsE¹⁻²¹⁵ = 10 mg/mL, BcsE²²⁴⁻⁵²³ = 5 mg/mL).

6.3 Objective 3: Functional Analysis of BcsE and BcsG¹⁶⁴⁻⁵⁵⁹

6.3.1 Investigating Phosphatase Activity of BcsG¹⁶⁴⁻⁵⁵⁹

As mentioned above, all constructs were successfully expressed and purified. However, the most promising candidate for analysis was BcsG¹⁶⁴⁻⁵⁵⁹; therefore, due to time constraints, functional analyses were exclusively conducted for BcsG¹⁶⁴⁻⁵⁵⁹. Prior to this study mechanistic action of all that was known about BcsG is that it contains an N-terminal transmembrane region and bioinformatics suggested the potential for a periplasmic domain from the alkaline phosphatase superfamily (Galperin and Koonin, 2012). The colorimetric phosphate assay kit

from Abcam was fundamental for observing BcsG¹⁶⁴⁻⁵⁵⁹ activity under a wide range of protein, substrate and additive quantities during this project. Results from numerous assays indicated that BcsG¹⁶⁴⁻⁵⁵⁹ is in fact capable of hydrolyzing ATP at 0.332 nmol/mg/min, GTP at 0.411 nmol/mg/min and CTP at 0.760 nmol/mg/min as seen in Figure 26. Figure 27 shows that when the same assay was run again after EDTA exposure, activity was significantly decreased compared to the wild type enzyme. This result indicates that when metal was sequestered from BcsG¹⁶⁴⁻⁵⁵⁹ by EDTA, its activity decreased meaning that BcsG¹⁶⁴⁻⁵⁵⁹ is in fact dependent on a metal cofactor for activity.

A metal dependency profile was then generated for BcsG¹⁶⁴⁻⁵⁵⁹ by testing activity with a variety of metals associated with phosphatase enzymes (Roberts *et al.*, 2004; Schmidt *et al.*, 2005) in the presence of a CTP substrate. Buffers containing 20 mM metal were introduced to pure protein samples, as conducted by Kennely *et al.* (1993). Metals tested included magnesium, manganese, nickel, calcium and zinc. Table 10 displays all pertinent values for BcsG¹⁶⁴⁻⁵⁵⁹ activity during the metal dependency analysis.

Observed BcsG¹⁶⁴⁻⁵⁵⁹ specific activity values (Table 10) indicate that the rate of CTP hydrolysis was highest with magnesium at a rate of $2.03 \times 10^{-1} \pm 0.008$ nmol/mg/min, closely followed by manganese at $1.99 \times 10^{-1} \pm 0.006$ nmol/mg/min. Bioinformatics analyses with 3DLigand had suggested that magnesium was a metal cofactor for BcsG (Figure 19) and according to this profile that prediction holds true. The phosphoethanolamine phosphatase enzyme PHOSPHO1 also showed maximal activity in the presence of magnesium (Roberts *et al.*, 2004) while the phosphodiesterase YahA has the highest activity in the presence of magnesium and manganese together; which is supportive of the similar activity observed with BcsG¹⁶⁴⁻⁵⁵⁹ with these two metals in particular (Schmidt *et al.*, 2005).

Literature cited activity rates for phosphatases can vary greatly from the low nmol/min/mg range (Ren *et al.*, 2015) to the high $\mu\text{mol}/\text{min}/\text{mg}$ range (Atyaksheva *et al.*, 2008; Sussman *et al.*, 1968). Regardless of the variation between these enzyme activities with their primary substrates, all cited values are higher than those observed for the BcsG¹⁶⁴⁻⁵⁵⁹ hydrolysis of CTP. The comparatively low BcsG¹⁶⁴⁻⁵⁵⁹ values could be an indication that NTPs are not the primary substrate of BcsG. Research by Schmidt *et al.*, (2005) showed that, although the primary substrate for YahA was c-di-GMP, lower secondary activity could be observed with NTPs. Additionally, Roberts *et al.*, (2004) showed that human PHOSPHO1 had 4600 +/- 582 U/mg with PEA as a substrate, but <0.1 U/mg with ATP. Therefore, although abundant metal dependency information was gathered from this research, more substrates should be tested with BcsG¹⁶⁴⁻⁵⁵⁹ to determine if it has greater specificity for another compound.

For BcsG¹⁶⁴⁻⁵⁵⁹ samples tested with nickel and calcium cofactors, enzyme activity dropped 2- and 4-fold, respectively from optimal reaction rates; while in the presence of zinc activity was essentially eliminated (Table 10). Zinc was also shown to nearly abolish phosphatase activity for a protein from *Sulfolobus solfataricus* which was similarly stimulated by magnesium and manganese (Kennely *et al.*, 1993). Since notable differences in activity were consistent across multiple purification fractions, this method was likely sufficient for testing BcsG¹⁶⁴⁻⁵⁵⁹ activity in the presence of different metal cofactors; however, one metal that could also be tested in the future is cobalt, which also stimulated YahA and PHOSPHO1 phosphatase activity to a lesser extent than magnesium (Schmidt *et al.*, 2005; Roberts *et al.*, 2004).

Although BcsG is hypothesized to function in PEA labelling of cellulose, the mechanism of action remains unknown. In order to test a potential substrate from the phosphoethanolamine metabolism pathway (Figure 4) BcsG¹⁶⁴⁻⁵⁵⁹ hydrolase activity was attempted with PEA (Figure

29). Hydrolysis of PEA by BcsG¹⁶⁴⁻⁵⁵⁹ was easily testable with the colorimetric assay previously employed with NTPs, since inorganic phosphate is a by-product of PEA hydrolysis by enzymes like PHOSPHO1 (Figure 4) (Roberts *et al.*, 2004). Figure 29 shows that there appears to have been comparable PEA hydrolysis (0.136 nmol/mg/min) to CTP hydrolysis by BcsG¹⁶⁴⁻⁵⁵⁹ (0.237 nmol/mg/mL). This level of activity with PEA is encouraging since this is the first time that activity has been demonstrated with a substrate present in the periplasm from the PEA metabolism pathway. Future directions for this project will require further investigation with other other biological substrates and acceptors (ie. cellulose) from the PEA metabolism pathway to identify the specific role of BcsG. PEA levels and locations on an acceptor may also vary between organisms (similar to PG acetylation) but this is yet to be observed. A strong potential candidate for future activity testing with BcsG would be phosphatidylethanolamine as it can be readily turned into PEA by enzymes like phospholipase C (Roberts *et al.*, 2004) (Figure 4 and 44, panel B).

A)



B)

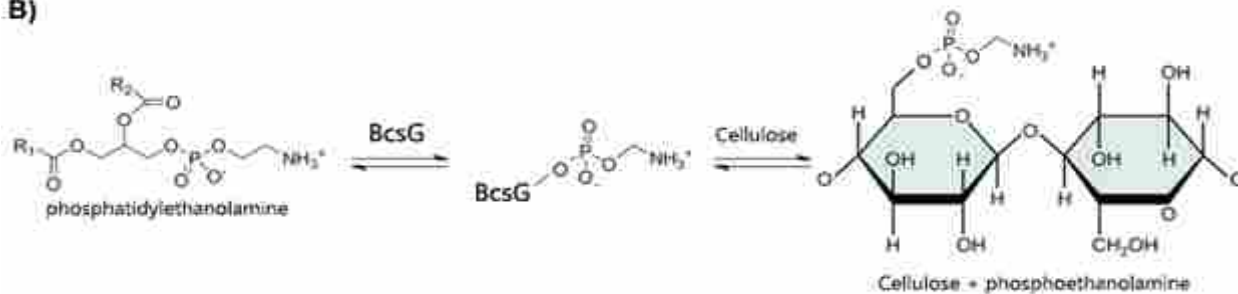


Figure 44. Potential BcsG substrates from PEA metabolism pathway

A) Tested reaction of PEA with BcsG resulting in phosphate liberation from PEA and ethanolamine production. **B)** Hypothetical phosphatidylethanolamine reaction with BcsG directly resulting in PEA labelling of bacterial cellulose.

A pH profile was also created with the phosphate-detecting colorimetric assay with CTP as the substrate for BcsG¹⁶⁴⁻⁵⁵⁹. The 50 mM buffers tested included: MES, PIPES, and Tris-based buffers from pH 5.5 to pH 9.5. A broad peak in activity occurred between pH 6.5 to pH 8.5 with PIPES and Tris-based buffers; the highest activity occurring at pH 7.5 in Tris-based buffer. Activity quickly dropped off outside of this range with low pH MES-based buffers and Tris-based buffers over pH 8.5. Alkaline phosphatases like YahA (Schmidt *et al.*, 2005) or Bovine Alkaline Phosphatase (Fosset *et al.*, 1974) display pH optimum for enzymatic activity at pH 9-10 and pH 8-10, respectively. Even though pH optimum is dependent on the substrate (Fosset *et al.*, 1974), the pH range for BcsG¹⁶⁴⁻⁵⁵⁹ is markedly lower than these other enzymes. However, the phosphatase human PHOSPHO1 involved in the PEA and PCho metabolism pathways displayed a similarly broad peak in activity between pH 6 and pH 7.5 with optimal activity around pH 6.7

(Roberts *et al.*,2004). This aligns with the hypothesis that BcsG¹⁶⁴⁻⁵⁵⁹ is associated with PEA labeling of bacterial cellulose.

The BcsG S116A mutant activity was also tested with magnesium as the metal cofactor and CTP as the substrate. As seen in Table 8, a number of key residues were mutated to alanine residues since it is a small, non-polar, uncharged residue known to be structurally benign. Although all of the BcsG mutants were sequenced, for this project, time only allowed for testing of the S116A mutant which was predicted to be the number one active site residue from previously discussed bioinformatics by 3DLigand. However, the BcsG S116A mutant had ~79% of full activity compared to the non-mutant BcsG¹⁶⁴⁻⁵⁵⁹ construct indicating that this residue is likely not actually vital to BcsG activity as predicted. Low conservation between BcsG homologs at residue S116 observed with a Clustal Omega alignment (Figure 18) indicates that this residue would not be a likely active site residue. Therefore, based on the high functionality of the S116A mutant and the Clustal Omega alignment showing poor conservation between homologs, other mutants should be tested in the future to identify the most important residues for activity. Future activity tests could be attempted with E280A, H281A and H234A next since they are part of the predicted metal binding site in BcsG. C81A and S82A, which are conserved across most sequences from different organisms (Figure 18), may be the catalytic nucleophile instead. The Y115A and Y167A may also be tested as they are predicted to be involved in binding. Loss of function due to mutation would strongly suggest importance of the mutated residue.

This project to date has identified conditions satisfactory for observing differences in hydrolysis rates of phosphate-rich molecules by BcsG¹⁶⁴⁻⁵⁵⁹ as well as optimal pH and metal cofactors required for activity. Since BcsG presence is hypothesized to be required for PEA labeling of bacterial cellulose because of identified homologs during bioinformatics, its substrate

is most likely from the PEA metabolism pathway. Therefore, the highest priority substrate test used next would involve mass spectrometry or evaporative light scattering detector (ELSD) analysis of BcsG¹⁶⁴⁻⁵⁵⁹ and phosphatidylethanolamine in the presence of cellulose to determine if BcsG is capable of directly converting this substrate to PEA and tagging cellulose. Another reaction to test by mass spectrometry or ELSD would include BcsG¹⁶⁴⁻⁵⁵⁹ in the presence of CDP-ethanolamine and cellulose to determine if BcsG produce phosphatidylethanolamine which can later be converted to PEA and added to cellulose. Finally, BcsG¹⁶⁴⁻⁵⁵⁹ in the presence of PEA and CTP could be analyzed by HPLC to determine if BcsG is responsible for the production of CDP-ethanolamine. All of these proposed reactions can be observed in Figure 4.

6.3.2 Localization of BcsG¹⁻⁵⁵⁹ Using GFP as a Reporter Protein

Green fluorescent protein, which was engineered onto the C-terminus of BcsG, has several properties that make it novel for periplasmic protein localization. Firstly, GFP contains an internal fluorophore composed of the Ser-Tyr-Gly residues which fluoresce within the cytoplasm of *E. coli* but not in the periplasm (Feilmeier *et al.*, 2000). The highly reducing periplasmic environment inhibits GFP fluorescence by causing improper GFP folding whereas the stable cytoplasm allows for proper GFP folding, and thus, fluorescence (Feilmeier *et al.*, 2000). Importantly, GFP can also be exported from the cytoplasm as a fusion protein which is necessary for locating a periplasmic protein. Finally, after environmental deactivation of GFP, fluorescence can be restored to the internal fluorophore by effectively unfolding and properly refolding the protein with an acid/base treatment as was achieved during localization of a periplasmic maltose binding protein (Feilmeier *et al.*, 2000).

The fact that the plated BcsG¹⁻⁵⁵⁹-GFP hybrid protein-expressing culture did not fluoresce like the GFP control under long-wave UV light on its own is not enough to prove that the C-

terminus of BcsG is outside of the cytoplasm. The untreated cleared lysate of the BcsG¹⁻⁵⁵⁹-GFP hybrid had very similar excitation, emission readings (365 nm, 509 nm) to the non-hybrid BcsG (14131 and 15199, respectively) which suggests that the hybrid GFP protein is not fluorescing (Figure 32). However, the acid-base treated BcsG¹⁻⁵⁵⁹-GFP hybrid excitation, fluorescence value (25546) is significantly higher, indicates that hybrid GFP fluorescence was restored. This restoration of inactivated GFP fluorescence suggests that the GFP fluorophore was, therefore, inactivated by its environment which means that the soluble region of BcsG is in fact located within the periplasm.

Furthermore, Phyre2 homology findings suggest that the C-terminal domain of BcsG may belong to the alkaline phosphatase superfamily which is comprised of proteins that tend to be in the periplasmic space in Gram-negative bacteria (Horiuchi *et al.*, 1959). Since the periplasmic space is much more subject to environmental variation than the cytoplasmic interior of the cell, bacterial alkaline phosphatases are resistant to inactivation, denaturation, and degradation, and contain a higher rate of activity (Horiuchi *et al.*, 1959). Like APs, BcsG remained soluble for long periods of time (weeks) even at 4°C indicating very high stability. Localization in the periplasm would also align with the comparatively low BcsG¹⁶⁴⁻⁵⁵⁹ specific activities observed with NTPs since these energy-rich molecules are not found in high concentrations within the periplasm. This localization would, therefore, also further suggest that BcsG has a non-NTP primary substrate. This location would also give BcsG a logical vantage point for labelling cellulose with PEA while it is being exported from the cell. Together, these results strongly suggest that the soluble C-terminal region of BcsG is located within the periplasm of bacterial cells with Type II bcs operons.

6.4 Objective 4: Protein Crystallization

6.4.1 Native Crystallization of BcsG¹⁶⁴⁻⁵⁵⁹

A good starting point for protein purity during crystallization screens is a distinct single band on a SDS-PAGE gel indicating that the sample is ~95% pure. Although initial crystallization screens were attempted for all BcsE constructs, each of the constructs was significantly less stable and exceedingly more difficult to work with than BcsG¹⁶⁴⁻⁵⁵⁹. Purifications with BcsE constructs took much longer and resulted in samples of lower purity which are both detrimental to crystallization since protein conformational purity also correlates to crystallizability by light scattering techniques and protein characteristically degrades and/or oligomerizes as time passes (D'arcy, 1994). This rate varies from construct to construct; however, using protein that is as fresh as possible is of great benefit since aggregates are known to severely hinder crystallization even in small quantities. Even after numerous attempts to optimize expression, purification and storage conditions, the highest concentration of pure BcsE¹⁻²¹⁵ (the best BcsE construct to work with) achieved was ~10 mg/mL. Although it is theoretically possible to crystallize a protein at concentrations this low, it is not favorable, and furthermore, BcsE constructs all rapidly precipitated even at concentrations as low as 1 mg/mL within 24 hrs of storage at 4°C.

The average protein concentration used for crystallization according to the Protein Database (PDB) is ~14 mg/mL; there are accounts of some proteins that have crystallized at lower concentrations than this, but usually concentrations above this value are preferable (Krupka *et al.*, 2002). In order to solve the BcsE structure, more optimization for the N- and C-terminal domain constructs (BcsE¹⁻²¹⁵ and BcsE²²⁴⁻⁵²³) will need to be conducted individually for crystallization or new truncations will need to be generated. However, purified BcsG¹⁶⁴⁻⁵⁵⁹ could

consistently be concentrated to as high as 50 mg/mL and stored at 10 mg/mL or less for weeks without precipitating. Therefore, all efforts were shifted solely to crystallizing BcsG¹⁶⁴⁻⁵⁵⁹ in order to increase the chances for achieving diffraction quality crystals in the time available.

Trials were run to identify the proper protein concentration and buffer conditions for high quality crystal formation. A detailed table was created and used to identify the conditions that resulted in the best crystal formation from numerous initial crystal screens. Crystals that fluoresce when exposed to UV light are auspicious, since this often accompanies the presence of aromatic residues in protein crystals. Crystals that turn dark blue after the application of IZIT dye are also likely protein since, unlike salt crystals, they contain channels that can readily take up liquid and become saturated.

The aesthetics of a protein crystal do not automatically indicate diffractibility. Subjectively beautiful crystals may diffract quite poorly while unappealing pieces can have the potential to result in high-quality diffraction patterns. For this reason it is advisable to pursue optimization of numerous isoforms after initial screens. During optimization, commonalities in buffer composition and protein concentration allowed for a small subset of conditions to be tested with varying protein to buffer ratios on hanging drop crystal plates. The subset of conditions included: MCSG-1T D1 (0.2 M Ammonium Sulfate, 0.1 M Bis-Tris:HCl, pH 6.5, 25% (w/v) PEG 3350), MCSG-3T A5 (0.1 M MES:NaOH, pH 6.0, 1.26 M Ammonium Sulfate), MCSG-4T F7 (0.2 M Magnesium Chloride, 0.1 M MES:NaOH, pH 6.5, 25% (w/v) PEG 4000) and MCSG-4T G11 (0.1 M HEPES:NaOH, pH 7.5, 20% (w/v) PEG 4000, 10% (v/v) 2-Propanol). Additionally, buffers used for expression and purification of BcsG¹⁶⁴⁻⁵⁵⁹ contained a low concentration of MgCl₂ which may have aided in protein stability during crystallization since magnesium was shown to be the preferred metal cofactor of BcsG (discussed above).

As depicted in Figure 11, four different protein concentrations were tested (20, 30, 40 and 50 mg/mL) in varying protein : buffer ratios with selected conditions. After approximately one month, numerous large and blocky crystals were observed in many of the 50 mg/mL BcsG¹⁶⁴⁻⁵⁵⁹ wells for conditions MCSG-4T F7 and G11 (Figure 37). Thirteen diffraction quality crystals were shipped to the CLS for X-ray analysis and resulting diffraction readings indicated resolution quality as sharp as 2.1Å with native protein crystals (Figure 38). A lower Angstrom value corresponds to finer electron density resolution. As resolution decreases, the observed electron density becomes fuzzier and more errors will occur during structural interpretation. At a resolution of 2.1Å (as achieved for BcsG¹⁶⁴⁻⁵⁵⁹) few residues would have a wrong rotamer and folds would rarely be incorrect for a generated 3D model. This level of resolution is suitable for solving the structure of a protein if the phase is also known. Therefore, following successful native crystallization of BcsG¹⁶⁴⁻⁵⁵⁹, experimental phasing tests had to be conducted in an attempt to solve the phase problem and, therefore, to be able to propose a structure of the protein. Table 12 displays a rough guide to the resolution of protein structures.

Table 12. A rough guide to the resolution of protein structures

Adapted from Blow, D. (2002). Outline of crystallography for biologists. Oxford University Press on Demand.

Resolution (Å)	Meaning
>4.0	Individual coordinates meaningless. Secondary structure elements can be determined.
3.0 – 4.0	Fold possibly correct, but errors are very likely. Many sidechains placed with wrong rotamer.
2.5 – 3.0	Fold likely correct except that some surface loops might be mismodelled. Several long, thin sidechains (lys, glu, gln, etc.) and small sidechains (ser, val, thr, etc.) likely to have wrong rotamers.
2.0 – 2.5	As 2.5 - 3.0, but number of sidechains in wrong rotamer is considerably less. Many small errors can normally be detected. Fold normally correct and number of errors in surface loops is small. Water molecules and small ligands become visible.
1.5 – 2.0	Few residues have wrong rotamer. Many small errors can normally be detected. Folds are rarely incorrect, even in surface loops.
0.5 – 1.5	In general, structures have almost no errors at this resolution. Individual atoms in a structure can be resolved. Rotamer libraries and geometry studies are made from these structures.

6.4.2 Heavy Metal BcsG¹⁶⁴⁻⁵⁵⁹ Crystal Soaks

Observable diffraction spots (reflections) only indicate the amplitude or magnitude of diffracted X-rays, but not the phase relations between them. In order to reconstruct the electron density from an X-ray diffraction pattern the phases for all reflections are required. When phase information is unknown, reconstruction by Fourier transforms from reflection data is not possible. This is known as the phase problem. Therefore, the missing phases must be obtained by separate phasing experimentation in order to solve a protein structure by X-ray diffraction.

Since a previously solved, structurally similar protein model is currently unavailable for BcsG, experimental phasing was attempted both by soaking native crystals in heavy atom solutions (NaBr and KBr were used) and by SeMet replacement in specialized minimal media. Several NaBr and KBr heavy metal soaks were attempted and sent for X-ray diffraction analysis at the CLS. The resulting diffraction patterns were observed at a resolution of $\sim 3\text{\AA}$; however, the readings were smeared. This smearing caused auto processing to be unable to index the results meaning that the software could not solve the phase for the soaked crystals. This smearing was likely caused by damage to the crystals since the heavy metal solution can cause degradation to occur within the protein's internal channels. Manual data processing may be possible in the future, however, this could inevitably prove impossible and SeMet crystallization is another route that could allow for phase solution.

6.4.3 Selenomethionine Crystallization

As expected, BcsG¹⁶⁴⁻⁵⁵⁹ expression yields were lower with the SeMet minimal media compared to expression in super broth (Figure 40) since minimal media contains lower nutritional content. Pooling of multiple batches allowed for protein concentration similar to non-SEMET conditions upon completion of purification steps. The first attempted expansion plate with BcsG¹⁶⁴⁻⁵⁵⁹ grown in SeMet minimal media was prepared with 40mg/mL and 50 mg/mL BcsG¹⁶⁴⁻⁵⁵⁹ with MCSG-4T F7 and G11 conditions (Figure 11). Crystal seeding was also utilized as it is an excellent technique used to introduce heterogeneous nucleation sites since spontaneous nucleation is known to be quite infrequent at metastable supersaturation levels. Therefore, slow protein crystal growth, which is desirable for generating high-quality diffracting crystals, may be initiated when it is unlikely that spontaneous nucleation will occur (Bergfors, 2003). Two more SeMet batches of BcsG¹⁶⁴⁻⁵⁵⁹ were attempted with varied protein, PEG and DTT concentrations since different purified protein batches often do not exhibit the same crystallizability. One of the plates resulted in microcrystals after 3 weeks with 40 mg/mL in MCSG-4T G11 condition; however, to date no notable SeMet-grown BcsG¹⁶⁴⁻⁵⁵⁹ crystals have been obtained.

There are many potential reasons why the SeMet-grown BcsG¹⁶⁴⁻⁵⁵⁹ crystals have not grown as the native protein crystals did. This could be due to the media it was expressed in or perhaps even the fact that it took two purification steps rather than one to get a pure enough sample for use. Proteins can pick up any number of cofactors, detergents or lipids that fluctuate between batches (Luecke *et al.*, 1999). Batch variation also often prevents crystal optimization after initial screens. The hanging drop plates containing BcsG¹⁶⁴⁻⁵⁵⁹ grown in SeMet media are still currently being incubated and it is also possible that they will simply take longer than the native protein to crystallize. These plates are still periodically analyzed.

6.5 Summary and Significance

The majority of observed enteric bacterial infections utilize biofilms, so it is critical to research bacterial exopolysaccharide biosynthesis mechanisms due to their ubiquitous appearance as a bacterial survival strategy (Fux *et al.*, 2005). Studying bacterial cellulose biosynthesis proteins specifically may lead to discoveries fundamental to exopolysaccharide biosynthesis across bacterial species. Furthermore, cellulose is the most abundant and only fully renewable biopolymer on earth (RömLing and Galperin, 2015). This research was conducted to expand upon the currently small foundation of knowledge for the largely uncharacterized BcsE and BcsG proteins which are somehow involved in bacterial cellulose biosynthesis in members of Enterobacteriaceae.

The BcsE and BcsG proteins may offer targets for circumventing the optimal production of the biofilm, or important biofilm components, thereby, aiding in controlling bacterial spread, contamination and disease progression. BcsE has low homology to previously solved proteins so its characterization could unveil novel structural or functional features that have formerly remained unknown. BcsE was recently shown to have a c-di-GMP binding domain on the C-terminus; recently named the GIL domain (Fang *et al.*, 2014). GIL is the second protein domain, after PilZ, shown to be dedicated to c-di-GMP binding (Fang *et al.*, 2014). It has also been shown that, in *S. enterica*, BcsE is not essential for cellulose synthesis but is required for maximal cellulose production, and that c-di-GMP binding is critical for BcsE function (Fang *et al.*, 2014). It appears that cellulose production in Enterobacteriaceae is somehow controlled by a two-tiered c-di-GMP-dependent system involving BcsE and the PilZ domain containing glycosyltransferase BcsA.

During this project, conditions for successful BcsE expression and purification for meaningful analysis were achieved for all three constructs (BcsE¹⁻⁵²³ = 5 mg/mL, BcsE¹⁻²¹⁵ = 10 mg/mL, BcsE²²⁴⁻⁵²³ = 5 mg/mL). The most promising candidate for structural and functional analyses was BcsG¹⁶⁴⁻⁵⁵⁹; therefore, due to time constraints, limited attempts to crystallize BcsE were attempted and the functional analyses were exclusively conducted for BcsG. However, impending research can use the foundation provided for BcsE constructs by this project to streamline future investigations. Future BcsE research should focus on N-terminal functional investigations and structural experiments for the N- and C-terminal constructs.

There are numerous noteworthy observable scenarios in which bacteria have used molecular modifications to their great advantage. For example, PEA is known to facilitate *N. gonorrhoeae* resistance to autophagy by decreasing its susceptibility to cationic antimicrobial peptides, as well as complement-mediated and intraleukocytic killing by neutrophils (Zughaier *et al.*, 2015). *N. gonorrhoeae* is one of many examples of human pathogens that are largely antibiotic resistant and also utilize PEA modification to evade the host response. Therefore, investigating the putative cellulose modifying protein BcsG may lead to discoveries useful for battling many organisms that utilize similar modifications for host-evasion or colonization.

The research conducted on BcsG during this project confirmed C-terminal phosphatase activity for ATP, GTP, CTP and PEA; the latter of which is from the PEA metabolism pathway. The preferred magnesium and manganese metal cofactors identified aligned with bioinformatics predictions and optimal pH for activity matched with putatively similar enzymes like the PEA phosphatase human PHOSPHO1. The C-terminus of BcsG was also shown to be located in the periplasm with use of a BcsG-GFP hybrid protein. All of these results are very encouraging since they align with the hypothesis that BcsG modifies bacterial cellulose as it exits the cell. More

functional assays will need to be tested in the future for other biological substrates and acceptors (ie. cellulose) in order to identify the specific role of BcsG in the PEA metabolism pathway. A strong potential candidate for future activity testing with BcsG would be phosphatidylethanolamine as it can be readily turned into PEA by enzymes like phospholipase C (Roberts *et al.*, 2004).

During functional studies, numerous large and blocky crystals were observed in many of the 50 mg/mL BcsG¹⁶⁴⁻⁵⁵⁹ wells for conditions MCSG-4T F7 (0.2 M Magnesium Chloride, 0.1 M MES:NaOH, pH 6.5, 25% (w/v) PEG 4000) and MCSG-4T G11 (0.1 M HEPES:NaOH, pH 7.5, 20% (w/v) PEG 4000, 10% (v/v) 2-Propanol). Tens of diffraction quality crystals were shipped to the CLS for X-ray analysis and resulting diffraction readings indicated resolution quality as sharp as 2.1Å with native protein crystals. At this resolution residues would be unlikely to be assigned incorrect rotamers and folds for a generated 3D model. This level of resolution is suitable for solving the structure of a protein if the phase is known. Therefore, following successful native crystallization of BcsG¹⁶⁴⁻⁵⁵⁹, experimental phasing with heavy metal soaking and selenomethionine labeling techniques were attempted in search of missing phases for BcsG¹⁶⁴⁻⁵⁵⁹. These techniques have shown promise; however, experiments are ongoing. Future studies with BcsG should continue optimizing selenomethionine labeling since microcrystals were grown during this project indicating that SeMet-grown BcsG¹⁶⁴⁻⁵⁵⁹ can crystallize.

6.6 The Integrative Nature of this Research

This project integrated many biological disciplines ranging from computational biology to DNA recombination, protein isolation, functional enzymatic investigations and structural analyses by X-ray crystallography techniques. However, it is because of the prominence of Enterobacteriaceae on Earth that makes this research truly integrative. Investigating the sub-microscopic level of exopolysaccharide synthesis impacts how we will approach large-scale issues regarding cellulose-utilizing organisms in nearly every setting where life can be found. The cellulose-producing Enterobacteriaceae are a large family of Gram-negative bacteria made up of numerous symbionts, and pathogens, including: *Escherichia coli*, *Salmonella*, *Shigella*, *Klebsiella*, and *Yersinia pestis*. Many Enterobacteriaceae compose the gut microbiome in humans and other animals, while countless others are found in soil and water. In the laboratory, *E. coli* is also one of the most important model organisms used in microbiology today.

Bacterial cellulose has a multitude of possible applications owing to its numerous unique properties. Bacterial cellulose has already been implemented as a food stabilizer and as a thickener; plus it has other far-reaching applications in commercial industries as well. It is currently used in the production of ultra-strength paper, hi-fidelity loudspeakers and headphones, as an additive in cosmetics, and it is now being tested for the possible development of cellulose-based clothing. Bacterial cellulose may even soon be an effective material for generating electronic paper due to its durability, higher purity and microfibril structure.

Bacterial cellulose is also important in the modern medical industry as a wound dressing for burns because of its ability to hold water to injuries; which is critical for rapid healing and lower instances of scarring. Bacterial cellulose has also been used during bone grafts and other tissue regeneration treatments because it retains all of its useful properties while easily being

reshaped. Furthermore, bacterial cellulose can be modelled into long, hollow tubes as a replacement for components of the cardiovascular system, the digestive or urinary tracts and the trachea without being rejected by existing biological tissues.

Understanding the underlying mechanisms for bacterial cellulose modification could provide a toolkit to give cellulose unique physical and chemical properties. A deeper understanding of bacterial cellulose modification machinery could potentiate synthetic modification of bacterial cellulose for unique biopolymer generation in medicine and bioengineering. Biofilms are already utilized in bioremediation to degrade toxic pollutants, but more biofilm information means identifying more biofilm promoters, and strategies for eliminating harmful environmental compounds safely in the future. Knowledge of proteins involved in bacterial cellulose biosynthesis and modification may soon lead to production of novel materials for a spectrum of situations and for combatting persistent bacterial infections.

Enterobacteriaceae are enormously versatile and, therefore, they are found ubiquitously throughout nature. From the soil to our water, our food and in our gut; these organisms represent the majority. Since these omnipresent and adaptable organisms utilize cellulose as a biofilm exopolysaccharide, we have an imperative to learn about their mechanisms for cellulose biosynthesis so we may control them in settings where it matters most to us: hospitals, industry, and in the environment. This project represents one of many tireless efforts to keep up with the ever-adapting Enterobacteriaceae so that we may utilize their adaptations for our own ends; rather than be ruled by them.

7. REFERENCES

1. Atyaksheva LF, Chukhrai ES & Poltorak OM. 2008. The catalytic properties of alkaline phosphatases under various conditions. *Russian Journal of Physical Chemistry A, Focus on Chemistry*, 82(11), 1947-1951.
2. Augimeri RV, Varley AJ & Strap JL. 2015. Establishing a role for bacterial cellulose in environmental interactions: lessons learned from diverse biofilm-producing Proteobacteria. *Frontiers in microbiology*, 6.
3. Bergfors T. 2003. Seeds to crystals. *Journal of structural biology*, 142(1), 66-76.
4. Bergman P, Johansson L, Asp V, Plant L, Gudmundsson GH, Jonsson AB & Agerberth B. 2005. *Neisseria gonorrhoeae* downregulates expression of the human antimicrobial peptide LL-37. *Cellular microbiology*, 7(7), 1009-1017.
5. Blow D. 2002. *Outline of crystallography for biologists*. Oxford University Press on Demand.
6. Bolanos-Garcia VM & Davies OR. 2006. Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1760(9), 1304-1313.
7. Costerton JW, Stewart PS, and Greenberg EP. 1999. Bacterial biofilms: a common cause of persistent infections. *Science*, 284(5418), 1318-1322.
8. Carson M, Johnson DH, McDonald H, Brouillette C & DeLucas LJ. 2007. His-tag impact on structure. *Acta Crystallographica Section D: Biological Crystallography*, 63(3), 295-301.
9. Crnich CJ and Drinka P. 2012. Medical device-associated infections in the long-term care setting. *Infect. Dis. Clin. North Am.* 26, 143-164.

9. Cruz DP, Huertas MG, Lozano M, Zárata L and Zambrano MM. 2012. Comparative analysis of diguanylate cyclase and phosphodiesterase genes in *Klebsiella pneumoniae*. BMC microbiology, 12(1), 1.
10. Dale GE, Oefner C & D'Arcy A. 2003. The protein as a variable in protein crystallization. Journal of structural biology, 142(1), 88-97.
11. D'Arcy A. 1994. Crystallizing proteins—a rational approach?. Acta Crystallographica Section D: Biological Crystallography, 50(4), 469-471.
12. Dinh T & Bernhardt TG. 2011. Using superfolder green fluorescent protein for periplasmic protein localization studies. Journal of bacteriology, 193(18), 4984-4987.
13. Fang X, Ahmad I, Blanka A, Schotchkowski M, Cimdins A, Galperin MY, and Gomelsky M. 2014. GIL, a new c-di-GMP-binding protein domain involved in regulation of cellulose synthesis in enterobacteria. Molecular microbiology, 93(3), 439-452.
14. Feilmeier BJ, Iseminger G, Schroeder D, Webber H & Phillips GJ. 2000. Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*. Journal of bacteriology, 182(14), 4068-4076.
15. Fischer W, Behr T, Hartmann R, Peter-Katalinic J & Egge H. 1993. Teichoic acid and lipoteichoic acid of *Streptococcus pneumoniae* possess identical chain structures. The FEBS Journal, 215(3), 851-857.
16. Fosset M, Chappelet-Tordo D & Lazdunski M. 1974. Intestinal alkaline phosphatase. Physical properties and quaternary structure. Biochemistry, 13(9), 1783-1788.
17. Fux CA, Costerton JW, Stewart PS, Stoodley P. 2005. Survival Strategies of infectious biofilms. Trends in Microbiology 13:34-40.

18. Galperin MY & Koonin EV. 2012. Divergence and convergence in enzyme evolution. *J. Biol. Chem.* 287, 21-28.
19. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD & Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press pp. 571-607.
20. Handing JW & Criss AK. 2015. The lipooligosaccharide-modifying enzyme LptA enhances gonococcal defence against human neutrophils. *Cellular microbiology*, 17(6), 910-921.
21. Hankins JV, Madsen JA, Giles DK, Brodbelt JS & Trent MS. 2012. Amino acid addition to *Vibrio cholerae* LPS establishes a link between surface remodeling in gram-positive and gram-negative bacteria. *Proceedings of the National Academy of Sciences*, 109(22), 8722-8727.
22. Hendrickson WA. 1991. Determination of Macromolecular Structures from Anomalous Diffraction of. *Science*, 254, 5028.
23. Henriques-Normark B & Tuomanen EI. 2013. The pneumococcus: epidemiology, microbiology, and pathogenesis. *Cold Spring Harbor perspectives in medicine*, 3(7), a010215.
24. Hofmann K & Stoffel W. 1993. TMbase - A database of membrane spanning proteins segments *Biol. Chem. Hoppe-Seyler* 374,166
25. Horiuchi T, Horiuchi S and Mizuno D. 1959. A possible negative feedback phenomenon controlling formation of alkaline phosphomonoesterase in *Escherichia coli*. *Nature* 183, 1529 – 1530.

26. Jeon Y *et al.* 2005. Phosphodiesterase: overview of protein structures, potential therapeutic applications and recent progress in drug development. *Cell Mol Life Sci* 62 (11): 1198–220.
27. Johnson M & Criss AK. 2013. *Neisseria gonorrhoeae* phagosomes delay fusion with primary granules to enhance bacterial survival inside human neutrophils. *Cellular microbiology*, 15(8), 1323-1340.
28. Keiski CL *et al.* 2010. AlgK Is a TPR-Containing Protein and the Periplasmic Component of a Novel Exopolysaccharide Secretin. *Structure* 18:265-273.
29. Kelso AA, Goodson SD, Watts LE, Ledford LL, Waldvogel SM, Diehl JN, ... & Sehorn MG. 2017. The β -isoform of BCCIP promotes ADP release from the RAD51 presynaptic filament and enhances homologous DNA pairing. *Nucleic acids research*, 45(2), 711-725.
30. Kennelly PJ, Oxenrider KA, Leng J, Cantwell JS & Zhao N. 1993. Identification of a serine/threonine-specific protein phosphatase from the archaebacterium *Sulfolobus solfataricus*. *Journal of Biological Chemistry*, 268(9), 6505-6510.
31. Kharat AS & Tomasz A. 2006. Drastic reduction in the virulence of *Streptococcus pneumoniae* expressing type 2 capsular polysaccharide but lacking choline residues in the cell wall. *Molecular microbiology*, 60(1), 93-107.
32. Kim IY *et al.* 2008. Chitosan and its derivatives for tissue engineering applications. *Biotechnology advances*, 26(1), 1-21.
33. Krupka HI, Segelke BW, Ulrich RG, Ringhofer S, Knapp M & Rupp B. 2002. Structural basis for abrogated binding between staphylococcal enterotoxin A superantigen vaccine and MHC-II α . *Protein science*, 11(3), 642-651.

34. Leipe DD, Koonin EV and Aravind L. 2004. STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *Journal of molecular biology*, 343(1), 1-28.
35. Lewis LA, Choudhury B, Balthazar JT, Martin LE, Ram S, Rice PA, ... & Shafer WM. 2009. Phosphoethanolamine substitution of lipid A and resistance of *Neisseria gonorrhoeae* to cationic antimicrobial peptides and complement-mediated killing by normal human serum. *Infection and immunity*, 77(3), 1112-1120.
36. Lewis LA, Shafer WM, Ray TD, Ram S & Rice PA. 2013. Phosphoethanolamine residues on the lipid A moiety of *Neisseria gonorrhoeae* lipooligosaccharide modulate binding of complement inhibitors and resistance to complement killing. *Infection and immunity*, 81(1), 33-42.
37. Luecke H, Schobert B, Richter HT, Cartailler JP & Lanyi JK. 1999. Structural changes in bacteriorhodopsin during ion transport at 2 angstrom resolution. *Science*, 286(5438), 255-260.
38. Mackinnon FG, Cox AD, Plested JS, Tang CM, Makepeace K, Coull PA, ... & Moxon ER. 2002. Identification of a gene (*lpt-3*) required for the addition of phosphoethanolamine to the lipopolysaccharide inner core of *Neisseria meningitidis* and its role in mediating susceptibility to bactericidal killing and opsonophagocytosis. *Molecular microbiology*, 43(4), 931-943.
39. Mah TFC, O'Toole GA. 2001. Mechanisms of biofilm resistance to antimicrobial agents. *Trends in Microbiology* 9: 34-39.

40. Mazur O, and Zimmer J. 2011. Apo- and Cellulose-Bound Structures of the Bacterial Cellulose Synthase Subunit BcsZ. *Journal of biology and chemistry* 1-11.
41. McComb RB, Bowers GN Jr and Posen S. 1979. *Alkaline Phosphatase*. New York: Plenum.
42. McDougald D, Rice SA, Barraud N, Steinberg PD & Kjelleberg S. 2012. Should we stay or should we go: mechanisms and ecological consequences for biofilm dispersal. *Nature reviews. Microbiology*, 10(1), 39.
43. McLaughlin K, Folorunso AO, Deeni YY, Foster D, Gorbatiuk O, Hapca SM, ... & Rogalsky S. 2017. Biofilm formation and cellulose expression by *Bordetella avium* 197N, the causative agent of bordetellosis in birds and an opportunistic respiratory pathogen in humans. *Research in Microbiology*.
44. Mechler L, Herbig A, Paprotka K, Fraunholz M, Nieselt K & Bertram R. 2015. A novel point mutation promotes growth phase-dependent daptomycin tolerance in *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy*, 59(9), 5366-5376.
45. Millán JL. 2006. *Mammalian Alkaline Phosphatases. From Biology to Applications in Medicine and Biotechnology*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co; 1–322.
46. Morgan JL, Strumillo J, and Zimmer J. 2013. Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* 0: 1-5.
47. Morgan JL, McNamara JT, and Zimmer J. 2014. Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nat. Struc. & Mol. Biology*, 21(5), 489-496.

48. Moynihan PJ & Clarke AJ. 2011. O-Acetylated peptidoglycan: controlling the activity of bacterial autolysins and lytic enzymes of innate immune systems. *The international journal of biochemistry & cell biology*, 43(12), 1655-1659.
49. Nielsen H. 2017. Predicting Secretory Proteins with SignalP. *Protein Function Prediction: Methods and Protocols*, 59-73.
50. Omadjela O, Narahari A, Strumillo J, Mélida H, Mazur O, Bulone V & Zimmer J. 2013. BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis. *Proceedings of the National Academy of Sciences*, 110(44), 17856-17861.
51. Petersen N, Gatenholm P. 2011. Bacterial cellulose-base materials and medical devices: Current state and perspectives. *Applied Microbiol Biotechnol* 91: 1277-1286.
52. Pozo JL and Patel R. 2007. The challenge of treating biofilm-associated bacterial infections. *Clinical Pharmacology & Therapeutics*, 82(2), 204-209.
53. Rane L & Subbarow Y. 1940. Choline, pantothenic acid, and nicotinic acid as essential growth factors for pneumococcus. *Journal of Biological Chemistry*, 134, 455-456.
54. Ren Z, Bechkoff G, Mebarek S, Keloglu N, Ahamada S, Meena S, ... & Buchet R. 2015. Direct determination of phosphatase activity from physiological substrates in cells. *PloS one*, 10(3), e0120087.
55. Roberts SJ, Stewart AJ, Sadler PJ & Farquharson C. 2004. Human PHOSPHO1 exhibits high specific phosphoethanolamine and phosphocholine phosphatase activities. *Biochemical Journal*, 382(1), 59-65.
56. RömLing U, Galperin MY and Gomelsky M. 2013. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbio and Mol Bio Reviews*, 77(1), 1-52.

57. RömLing U and Galperin MY. 2015. Bacterial cellulose biosynthesis: diversity of operons, subunits, products, and functions. *Trends in microbiology*, 23(9), 545-557.
58. Ross P *et al.* 1987. Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* 325: 279–281.
59. Ross P, Mayer R, and Benziman M. 1991. Cellulose biosynthesis and function in bacteria. *Microbiological reviews*, 55(1), 35-58.
60. Rupp B & Segelke B. 2001. Questions about the structure of the botulinum neurotoxin B light chain in complex with a target peptide. *Nature Structural & Molecular Biology*, 8(8), 663-664.
61. Russell MH, Bible AN, Fang X, Gooding JR, Campagna SR, Gomelsky M and Alexandre G. 2013. Integration of the second messenger c-di-GMP into the chemotactic signaling pathway. *MBio*, 4(2), 1-13.
62. Ryjenkov DA, Simm R, RömLing U and Gomelsky M. 2006. The PilZ domain is a receptor for the second messenger c-di-GMP: the PilZ domain protein YcgR controls motility in enterobacteria. *J Biol Chem* 281: 30310–30314.
63. Saldaña, Z. *et al.* 2009. 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. *Environmental microbiology*, 11(4), 992-1006.
64. Scheurwater E, Reid CW & Clarke AJ. 2008. Lytic transglycosylases: bacterial space-making autolysins. *The international journal of biochemistry & cell biology*, 40(4), 586-591.

65. Segelke BW, Schafer J, Coleman MA, Lekin TP, Toppani D, Skowronek KJ, ... & Rupp B. 2004. Laboratory scale structural genomics. *Journal of structural and functional genomics*, 5(1-2), 147-157.
66. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, ... & Thompson JD. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), 539.
67. Solano S, *et al.* 2002. Genetic analysis of *Salmonella enteritidis* biofilm formation: critical role of cellulose. *Molecular Microbiology* 43: 793-808.
68. Spiers AJ, Bohannon J, Gehrig SM & Rainey PB. 2003. Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose. *Molecular microbiology*, 50(1), 15-27.
69. Steiner S, Lori C, Boehm A & Jenal U. 2013. Allosteric activation of exopolysaccharide synthesis through cyclic di-GMP-stimulated protein-protein interaction. *The EMBO journal*, 32(3), 354-368.
70. Stewart PS. 2002. Mechanisms of antibiotic resistance in bacterial biofilms. *Int. J. Med. Microbiol.* 292, 107-113.
71. Sussman HH, Small PA & Cotlove E. 1968. Human alkaline phosphatase immunochemical identification of organ-specific isoenzymes. *Journal of Biological Chemistry*, 243(1), 160-166.
72. Unemo M & Shafer WM. 2014. Antimicrobial resistance in *Neisseria gonorrhoeae* in the 21st century: past, evolution, and future. *Clinical microbiology reviews*, 27(3), 587-613.

73. Vonck RA, Darville T, O'Connell CM & Jerse AE. 2011. Chlamydial infection increases gonococcal colonization in a novel murine coinfection model. *Infection and immunity*, 79(4), 1566-1577.
74. Walker JE, Saraste M, Runswick M J and Gay NJ. 1982. Distantly related sequences in the alpha-and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO journal*, 1(8), 945. Whitney JC, *et al.* 2011. Structural basis for alginate secretion across the bacterial outer membrane. *PNAS* 108(32): 13083-13088.
75. Ward WW & Bokman SH. 1982. Reversible denaturation of *Aequorea* green-fluorescent protein: physical separation and characterization of the renatured protein. *Biochemistry*, 21(19), 4535-4540.
76. Whitney JC, and Howell PL. 2013. Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. *Trends in microbiology*, 21(2), 63-72.
77. Wingfield PT. 2015. Overview of the purification of recombinant proteins. *Current protocols in protein science*, 6-1.
78. World Health Organization. Accession date, 24 June 2017. Gonorrhoea— incidence rate (per 100,000 population). Centralized Information System for Infectious Diseases (CISID) Website: Sexually transmitted infections (STI). WHO Regional Office for Europe, Copenhagen, Denmark. <http://data.euro.who.int/cisid>.
79. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ & Brinkman FSL. 2010. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes, *Bioinformatics* 26(13):1608-1615

80. Zogaj X, Nimtz M, Rohde M, Bokranz W, and RomLing U. 2001. The multicellular morphotypes of *Salmonella typhimurium* and *Escherichia coli* produce cellulose as the second component of the extracellular matrix. *Molecular Microbiology* 39:1452-1463.
81. Zughair SM, Kandler JL, Balthazar JT & Shafer WM. 2015. Phosphoethanolamine Modification of *Neisseria gonorrhoeae* Lipid A Reduces Autophagy Flux in Macrophages. *PloS one*, 10(12), e0144347.

8. APPENDIX

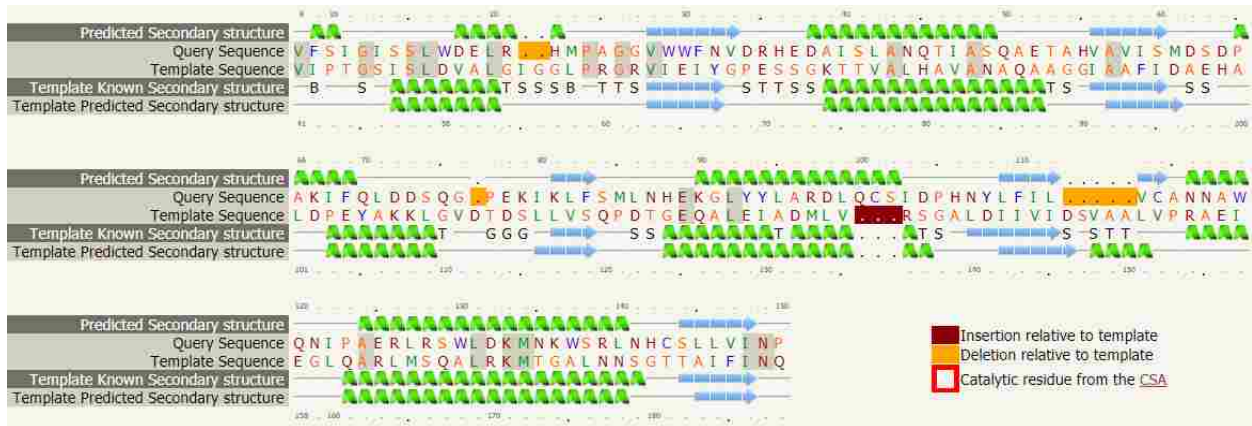


Figure A1. Representative BcsE Phyre2 secondary structure alignment to 1ubea
 The alignment only covers the N-terminus of the protein from residue 8-150 as indicated by the red bar on the left side. This is also true for other listed BcsE homology matches on Phyre2.

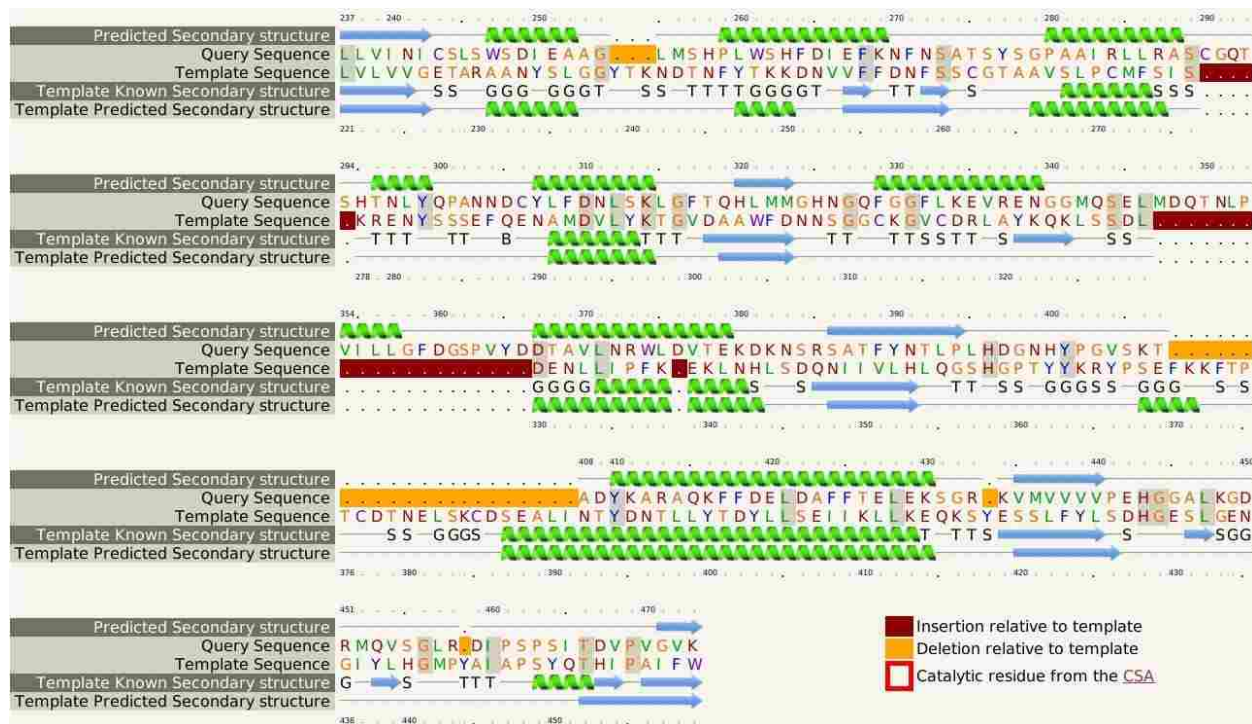


Figure A2. Representative BcsG Phyre2 secondary structure alignment to c4tn0C
 The alignment only covers the C-terminus of the protein from residue 237-450 as indicated by the red bar on the left side.

Topology

Feature key	Position(s)	Length	Description	Graphical view
Transmembrane ¹	34 – 54	21	Helical Sequence analysis	
Transmembrane ¹	68 – 88	21	Helical Sequence analysis	
Transmembrane ¹	113 – 133	21	Helical Sequence analysis	
Transmembrane ¹	139 – 159	21	Helical Sequence analysis	

Figure A3. Unitprot predicted transmembrane regions for BcsG¹⁻⁵⁵⁹.
 Transmembrane segment positions and shape as determined by Unitprot.

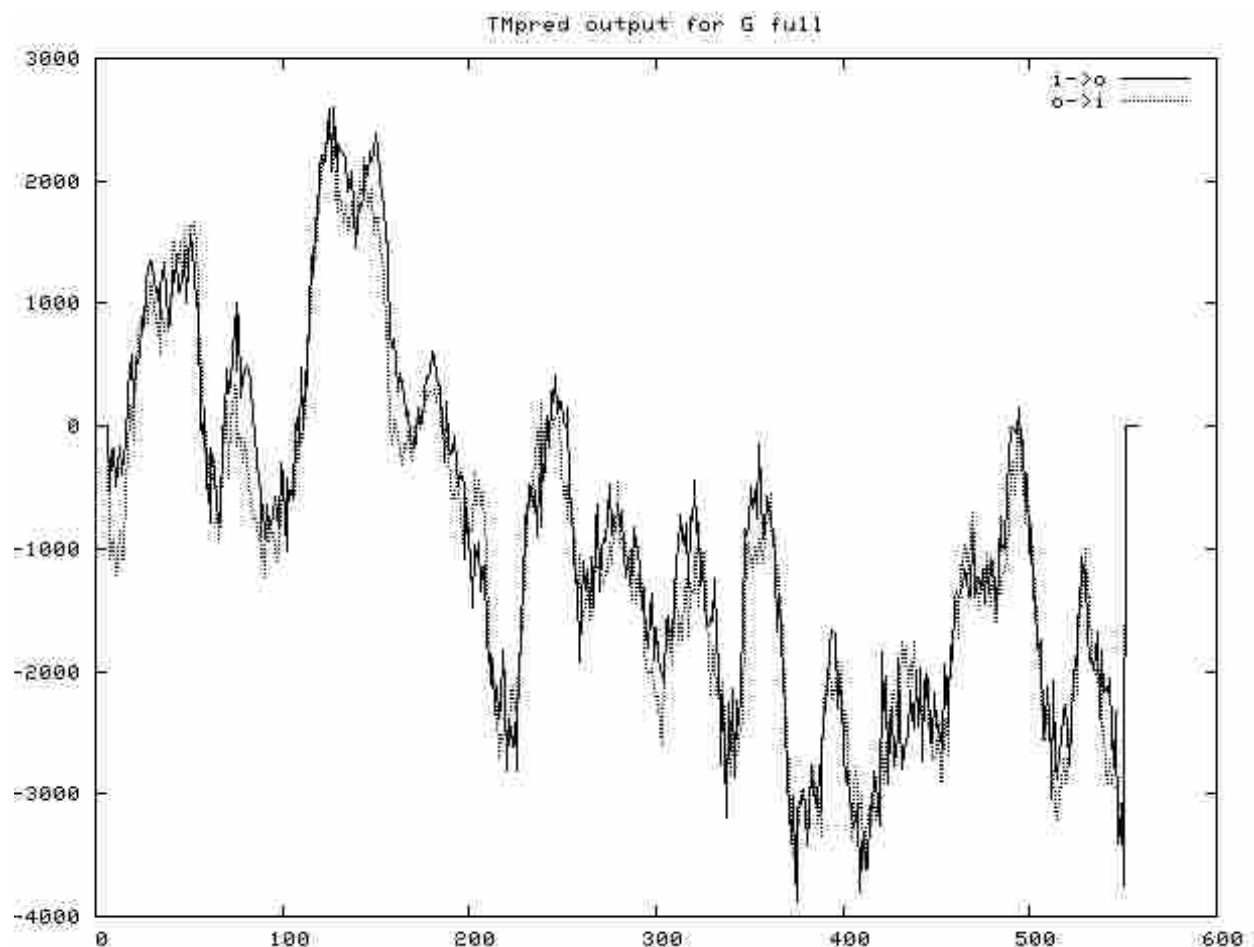


Figure A4. TMPred predicted transmembrane regions for BcsG¹⁻⁵⁵⁹.
 Transmembrane segment positions and shape as determined by TMPred.

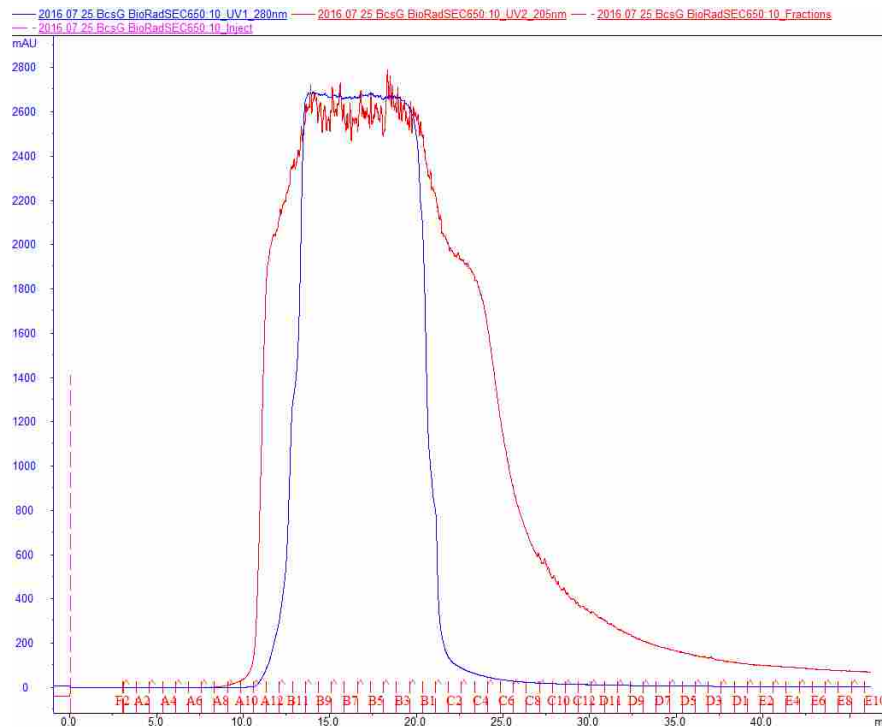


Figure A5. Size Exclusion Purification Profile for BcsG¹⁶⁴⁻⁵⁵⁹

The large peak at ~15 mL indicates the release of pure BcsG¹⁶⁴⁻⁵⁵⁹ as indicated by the blue line at 280 nm, and the red line at 205 nm, which are both optimal wavelengths to observe proteins.

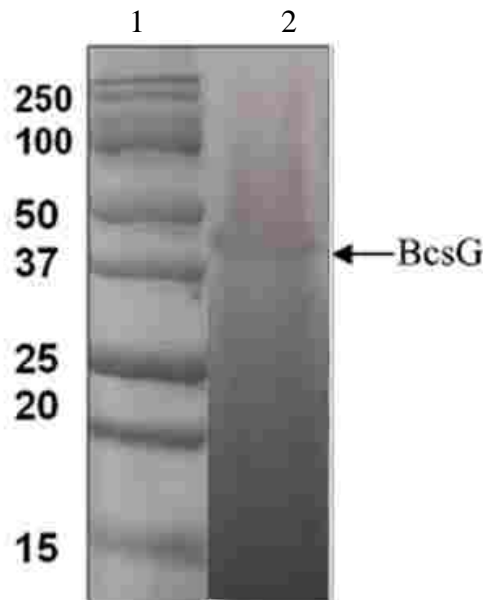


Figure A6. Western blot of BcsG¹⁶⁴⁻⁵⁵⁹ after IMAC purification

Western blot analysis of BcsG¹⁶⁴⁻⁵⁵⁹ after IMAC purification for the elution #1 fraction. BcsG¹⁶⁴⁻⁵⁵⁹ indicated with arrow at 45 kDa. Lane 1, molecular-weight markers (kDa); lane 2, Elution #1 fraction.

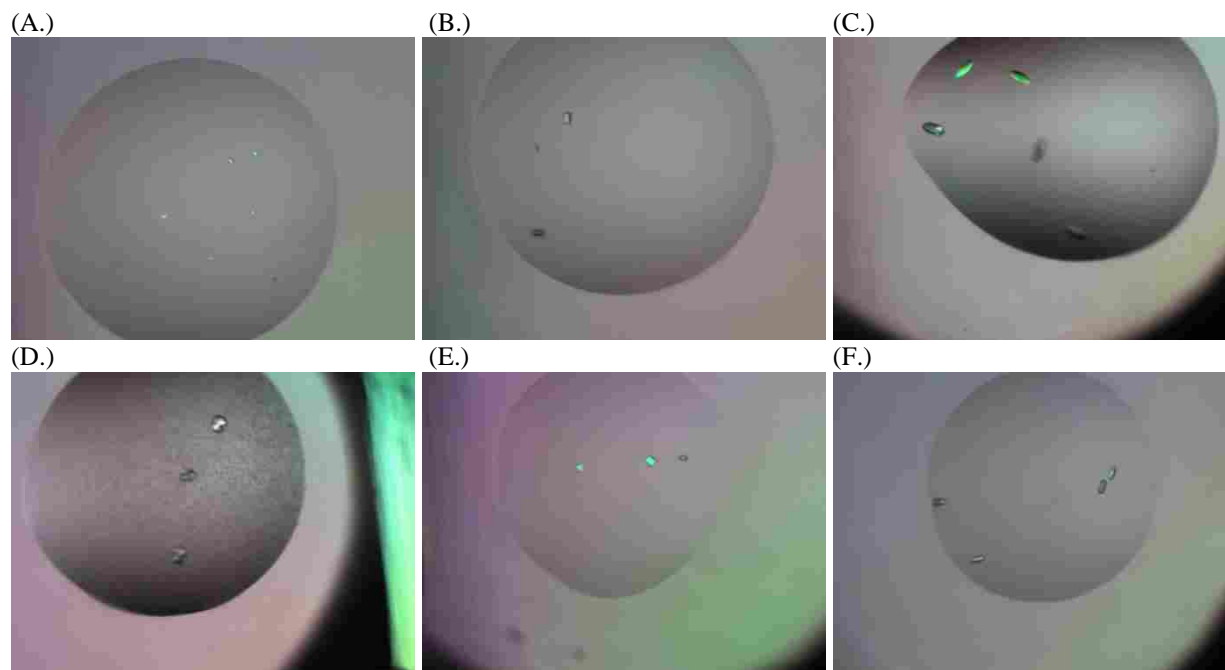


Figure A6. Initial BcsG¹⁶⁴⁻⁵⁵⁹ phosphate buffer attempts resulted only in salt crystals. Numerous crystal isoforms were obtained for BcsG¹⁶⁴⁻⁵⁵⁹ with a phosphate buffer. Analysis with the University of Waterloo X-ray diffractometer indicated that all isoforms were salt. Future crystallization attempts would utilize Tris buffers for purification and protein storage.

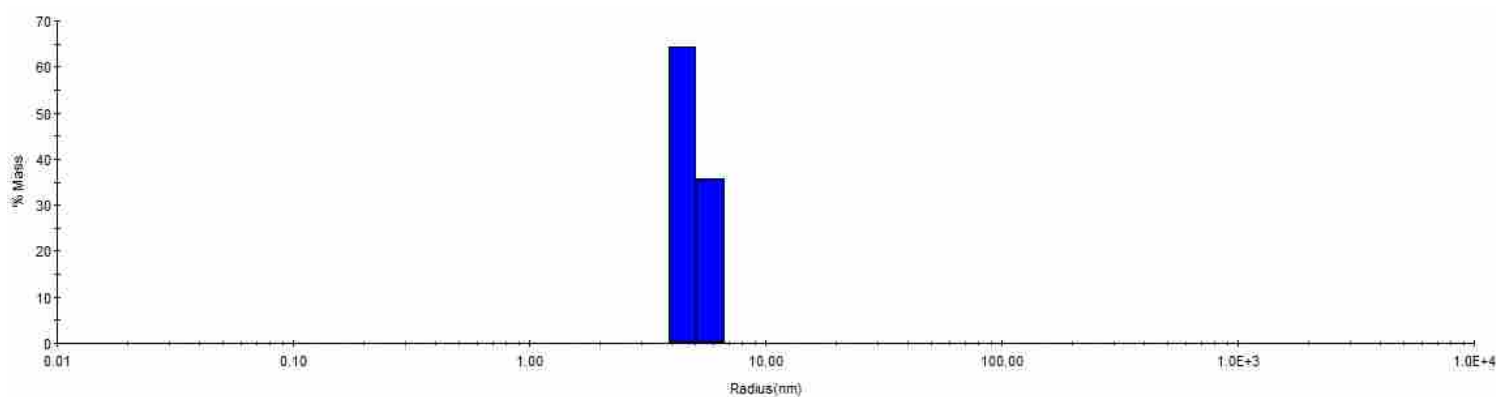


Figure A7. Dynamic Light Scatter (DLS) plot for BcsG¹⁶⁴⁻⁵⁵⁹ This DLS reading for BcsG¹⁶⁴⁻⁵⁵⁹ (0.5mg/mL in 50 mM Tris buffer) indicates that the construct is quite monodisperse under defined optimal conditions which correlates to protein crystallizability.

Table A1. Potential protein crystal isoforms from numerous MCSG-1T-4T crystal screens
After optimization of expression and purification of BcsG¹⁶⁴⁻⁵⁵⁹ with 50 mM Tris buffer, numerous crystal isoforms were observed in a multitude of conditions. Isoforms were tracked and compared so that most promising hits could be determined and further tested.

MCSG-1T				
<u>Well</u>	<u>Salt</u>	<u>Buffer</u>	<u>Precipitant</u>	<u>Isoform</u>
B4	0.2 M Magnesium Chloride	0.1 M Bis-Tris:HCl pH 6.5	25% (w/v) PEG 3350	starburst/urchin
B5	0.2 M Magnesium Chloride	0.1 M Tris:HCl pH 8.5	25% (w/v) PEG 3350	urchin
C5	0.2 M Magnesium Acetate		20% (w/v) PEG 3350	thin plates
C9	0.8 M Lithium Chloride	0.1 M Tris:HCl pH 8.5	32% (w/v) PEG 4000	hexagons (some thick/some thin)
D1	0.2 M Ammonium Sulfate	0.1 M Bis-Tris:HCl pH 6.5	25% (w/v) PEG 3350	football-shaped
D3	0.2 M Magnesium Chloride	0.1 M Tris:HCl pH 8.5	30% (v/v) PEG 400	thin/irregular plates
D8	0.1 M Sodium Chloride	0.1 M Bis-Tris:HCl pH 6.5	1.5 M Ammonium Sulfate	small squares
E1		0.1 M HEPES:NaOH pH 7.5	2.0 M Ammonium Sulfate	small squares
E8		0.1 M Tris:HCl pH 8.5	1.8 M Magnesium Sulfate	3D/irregular/jagged
E10	0.2 M Ammonium Tartrate Dibasic		20% (w/v) PEG 3350	geometric/3D/bright/sharp
E12			2.4 M Sodium Malonate pH7.0	thin/irregular sheet
F2	0.2 M Ammonium Acetate	0.1 M Bis-Tris:HCl pH 6.5	25% (w/v) PEG 3350	small squares
F5	0.2 M Sodium Acetate		20% (w/v) PEG 3350	3D/irregular/jagged
F9	0.05 M Ammonium Sulfate	0.05 M Bis-Tris:HCl pH 6.5	30% (v/v) Pentaerythritol Ethoxylate (15/4 EO/OH)	3D/irregular/jagged
F11	0.2 M Ammonium Sulfate	0.1 M HEPES:NaOH pH 7.5	25% (w/v) PEG 3350	thin rod
G2	0.04 M Potassium Phosphate		16% (w/v) PEG 8000, 20% (v/v)	3D/irregular/jagged

	monobasic		Glycerol	
G5		0.1 M Bis-Tris:HCl pH 6.5	2.0 M Ammonium Sulfate	small squares
G7		0.1 M Tris:HCl pH 8.5	25% (w/v) PEG 3350	thin/irregular sheet
H7		0.1 M Bis-Tris:HCl pH 5.5	2.0 M Ammonium Sulfate	small squares
H8	1.0 M Succinic Acid pH 7.0	0.1 M HEPES:NaOH pH 7.0	1% (w/v) PEG MME 2000 93	hexagonal
MCSG-2T				
<u>Well</u>	<u>Salt</u>	<u>Buffer</u>	<u>Precipitant</u>	<u>Isoform</u>
E6		0.1 M Sodium Acetate:Acetic Acid pH 4.5	1 M Ammonium Phosphate Dibasic	small rods in pptt
G6	0.2 M Sodium Chloride	0.1 M Sodium Citrate:Citric Acid pH 5.5	1.0 M Ammonium Phosphate Dibasic	small rods
G9	0.2 M Sodium Chloride	0.1 M Na₂HPO₄:KH₂PO₄ pH 6.2	20% (w/v) PEG 1000	hexagonal
MCSG-3T				
<u>Well</u>	<u>Salt</u>	<u>Buffer</u>	<u>Precipitant</u>	<u>Isoform</u>
A4		0.1 M Imidazole:HCl pH 8.0	1.0 M Ammonium Phosphate Dibasic	small rods
C4		0.1 M Na₂HPO₄/KH₂PO₄ pH 6.2	25% (v/v) 1,2-Propandiol, 10% (v/v) Glycerol	medium/sharp/rod (single)
E1	0.2 M Sodium Chloride	0.1 M NaH₂PO₄/K₂HPO₄ pH 6.2	50% (v/v) PEG 200	medium/sharp/rod (numerous)
E6		0.1 M Tris:HCl pH 8	20% (v/v) MPD	medium/sharp/rod (in pptt)
G3	0.2 M Magnesium Chloride	0.1 M Sodium Cacodylate:HCl pH 6.5	50% (v/v) PEG 200	small rods
H9		0.1 M Tris:HCl pH	1.0 M	small rods (in pptt)

		8.5	Ammonium Phosphate Dibasic	
A3	0.2 M Lithium Sulfate	0.1 M Tris:HCl pH 7.0	1.0 M Potassium Sodium Tartrate	rod (one/sharp)
A4		0.1 M Imidazole:HCl pH 8.0	1.0 M Ammonium Phosphate Dibasic	rod (numerous)
A5a		0.1 M MES:NaOH pH 6.0	1.26 M Ammonium Sulfate	large/sharp rod (intriguing)
A5b		0.1 M MES:NaOH pH 6.0	1.26 M Ammonium Sulfate	small squares (numerous)
A12		0.1 M HEPES:NaOH pH 7.5	0.3 M Magnesium Formate	hexagonal
B2			1.8 M Ammonium Citrate Tribasic pH 7.0	small squares
B3	0.8 M Potassium Sodium Tartrate	0.1 M Tris:HCl pH 8.5	0.5% (w/v) PEG MME 5000	small squares
B5	0.2 M Ammonium Acetate	0.1 M HEPES:NaOH pH 7.5	45% (v/v) MPD	small squares (heavy pptt)
B8	0.2 M Sodium Chloride	0.1 M Sodium Cacodylate:HCl pH 6.5	2 M Ammonium sulfate	small squares
B9	0.2 M Calcium Acetate	0.1 M Imidazole:HCl pH 8.0	20% (w/v) PEG 1000	hexagonal (very small/single)
B10 a	0.2 M Magnesium Chloride	0.1 M Imidazole:HCl pH 8.0	15% (v/v) Ethanol	thin/irregular sheet
B10 b	0.2 M Magnesium Chloride	0.1 M Imidazole:HCl pH 8.0	15% (v/v) Ethanol	hexagonal (sharp/single/3D)
B12		0.1 M HEPES:NaOH pH 7.5	1.26 M Ammonium	hexagonal

			Sulfate	
C1		0.1 M Imidazole:HCl pH 8.0	2.5 M Sodium Chloride	small squares
C3		0.1 M Sodium Cacodylate:HCl pH 6.5	40% (v/v) MPD, 5% (w/v) PEG 8000	hexagonal (in heavy pptt)
C4		0.1 M Na₂HPO₄/KH₂PO₄ pH 6.2	25% (v/v) 1,2- Propandiol, 10% (v/v) Glycerol	large/3D rectangle (numerous)
C7		0.1 M Bis-Tris Propane:HCl pH 7.0	1.0 M Ammonium Citrate Tribasic pH 7.0	small squares
C8		0.1 M Tris:HCl pH 8.5	3.2 M Sodium Chloride	small squares
C9		0.1 M Bis-Tris Propane:HCl pH 7.0	2.0 M Sodium Formate	thin rod
C10	0.2 M Ammoniu m Acetate	0.1 M Bis-Tris:HCl pH 6.5	45% (v/v) MPD	small squares (heavy pptt)
C11	0.2 M Ammoniu m Acetate	0.1 M Tris:HCl pH 8.5	45% (v/v) MPD	small squares (heavy pptt)
D3	0.2 M Lithium Sulfate	0.1 M Tris:HCl pH 7.0	2.0 M Ammonium Sulfate	rod
D4		0.1 M Imidazole:HCl pH 8.0	1.0 M Sodium Citrate Tribasic	hexagonal (small)
D5	0.2 M Sodium Chloride	0.1 M Imidazole:HCl pH 8.0	1.0 M Potassium Sodium Tartrate	thin rod
D7	0.2 M Lithium Sulfate	0.1 M CHES:NaOH pH 9.5	1.0 M Potassium Sodium Tartrate	small squares

D9			1.6 M Sodium Citrate Tribasic	needles
E1	0.2 M Sodium Chloride	0.1 M NaH₂PO₄/K₂HPO₄ pH 6.2	50% (v/v) PEG 200	large/3D rectangle
E4	0.05 M Cesium Chloride	0.1 M MES:NaOH pH 6.5	30% (w/v) Jeffamine® M-600	small squares
E5		0.1 M Sodium Citrate:HCl pH 5	3.15 M Ammonium Sulfate	small squares
E6		0.1 M Tris:HCl pH 8	20% (v/v) MPD	small squares (various isoforms)
E9		0.09 M HEPES:NaOH pH 7.5	1.26 M Sodium Citrate, 10% (v/v) Glycerol	needles
E10		0.1 M Bis-Tris Propane:HCl pH 7.0	2.5 M Ammonium Sulfate	needles
F3	0.2 M Magnesium Chloride	0.1 M Tris:HCl pH 8.5	50% (v/v) Ethylene Glycol	needles
F4			2.1 M DL-Malic Acid pH 7.0	needles
F5	1.0 M Ammonium Sulfate	0.1 M HEPES:NaOH pH 7.0	0.5% (w/v) PEG 8000	irregular/3D
F8		0.1 M HEPES:NaOH pH 7.0	10% (w/v) PEG MME 5000	irregular/3D
F11		0.1 M Tris:HCl pH 8.5	20% (w/v) PEG 1000	irregular/3D
G1	0.2 M Magnesium Chloride	0.1 M Imidazole:HCl pH 8.0	35% (v/v) MPD	hexagonal (in heavy pptt)
G3	0.2 M Magnesium Chloride	0.1 M Sodium Cacodylate:HCl pH 6.5	50% (v/v) PEG 200	irregular/3D (plus urchins)
G5		0.1 M HEPES:NaOH pH 7.5	70% (v/v) MPD	small squares (heavy pptt)

G10		0.1 M Bis-Tris Propane:HCl pH 7.0	1.5 M Ammonium Chloride	small squares
G11		0.1 M Tris:HCl pH 8.5	0.4 M Magnesium Formate	irregular/3D
H1	0.2 M Proline	0.1 M HEPES:NaOH pH 7.5	10% (w/v) PEG 3350	rod
H7	0.2 M Sodium Chloride	0.1 M HEPES:NaOH pH 7.5	20% (v/v) 1,4- Butanediol	thin/irregular sheet
H8		0.1 M CHES:NaOH pH 9.5	30% (v/v) PEG 400	thin/irregular sheet
H9		0.1 M Tris:HCl pH 8.5	1.0 M Ammonium Phosphate Dibasic	rod
H11		0.1 M Bis-Tris Propane:HCl pH 7.0	0.7 M Sodium Citrate Tribasic	thin/irregular sheet
MCSG-4T				
<u>Well</u>	<u>Salt</u>	<u>Buffer</u>	<u>Precipitant</u>	<u>Isoform</u>
F7	0.2 M Magnesium Chloride	0.1 M MES:NaOH pH 6.5	25% (w/v) PEG 4000	starburst/urchin
G11		0.1 M HEPES:NaOH pH 7.5	20% (w/v) PEG 4000, 10% (v/v) 2- Propanol	large/blocky crystals (with ppt)
A5		0.1 M Na₂HPO₄/KH₂PO₄ pH 6.2	10% (w/v) PEG 3000	large rectangle
B1		0.1 M Bis-Tris Propane:HCl pH 7	2.4 M DL- Malic Acid pH 7.0	rods (tiny)
B2		0.1 M Bis-Tris Propane:HCl pH 7	1.2 M Potassium Sodium Tartrate	irregular sheets (small)
B6		0.1 M Tris:HCl pH 8.5	0.7 M Sodium Citrate Tribasic	irregular sheets
B7		0.1 M Tris:HCl pH	6 M	small squares

		8.5	Ammonium Nitrate	
C2		0.1 M Tris:HCl pH 8.5	1.2 M Sodium Citrate Tribasic	rods
C3		0.1 M Tris:HCl pH 8.5	3.5 M Sodium Formate	needles
C8		0.1 M Bis-Tris Propane:HCl pH 7	2.2 M DL-Malic Acid pH 7.0	needles
D11	0.2 M Lithium Sulfate	0.1 M MES:NaOH pH 6	35% (v/v) MPD	irregular sheets
F4	0.2 M Magnesium Chloride	0.1 M Tris:HCl pH 8.5	16% (w/v) PEG 4000	irregular sheets
F5	0.2 M Lithium Sulfate	0.1 M Tris:HCl pH 8.5	20% (w/v) PEG 4000	small squares
H3	0.2 M Ammonium Sulfate	0.1 M MES:NaOH pH 6.5	30% (w/v) PEG 5000 MME	irregular sheets