2015

PURIFICATION AND CHARACTERIZATION OF BcsC; AN INTEGRAL COMPONENT OF BACTERIAL CELLULOSE EXPORT

Emily D. Wilson Ms
Wilfrid Laurier University, wils3420@mylaurier.ca

PURIFICATION AND CHARACTERIZATION OF BcsC; AN INTEGRAL

COMPONENT OF BACTERIAL CELLULOSE EXPORT

by

Emily Wilson

Bachelor of Science Honours Biology, Wilfrid Laurier University, 2013

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

2015

Acknowledgements:

Emily Wilson

Characterization of BcsC

Abstract:

Biofilms are a growing concern in the medical field due to their increased resistance to antibiotics. When found in a biofilm, bacteria can have antibiotic resistance 10-1000 times that of their planktonic counterparts. Therefore, it is important to study the formation of biofilms. Cellulose biofilms are formed by *Enterobacteriaceae,* such as many *Escherichia coli* and *Salmonella* spp. strains. Biofilms provide these species with benefits including antimicrobial protection, development of bacterial communities, promotion of DNA exchange, uptake of nutrients, and, in the case of cellulose biofilms, immune system evasion. Cellulose biofilms are controlled by the Bacterial cellulose synthesis (Bcs) complex located at the cell membrane of bacteria able to form cellulose-based biofilms. Proteins, BcsA and BcsB, have been characterized for cellulose synthesis, however, cellulose export has yet to be described. BcsC is believed to play a role in this export process due to its homology to other polysaccharide export proteins in the alginate and poly β-1,6-GlcNAc (PGA) systems. Herein, a series of bioinformatics analysis was performed that supported the hypothesis that BcsC consists of an outer membrane β-barrel connected to a periplasmic tetratricopeptide repeat (TPR) region and that these two regions play different roles in the export process. To begin addressing this hypothesis, the research focused on the structure-function characterization of these regions of BcsC. While practical quantities of the β-barrel region could be purified, this region proved to be recalcitrant to folding into its native state following purification. However, high yields of all TPR constructs were obtained and subjected to further analyses. Circular dichroism studies confirmed our bioinformatics analyses that the secondary structure of the TPR constructs have a predominantly α-helical content. This technique also provided preliminary evidence that there are structural changes upon binding of the TPR to soluble carboxymethyl cellulose (CMC). Intrinsic fluorescence spectroscopy quenching results further demonstrated that the TPR region has a single binding site along with high K_D values (ex. 416 μM for the longest construct) for carboxymethyl cellulose. These results were further confirmed with an Avicel insoluble substrate binding assay which also demonstrated that binding of the TPR to cellulose occurred across a biologically relevant pH range (pH 6-8) and that the majority of the binding may be due to an N-terminal portion of the TPR region (amino acids 24-342). Thus, this collective evidence supports that the TPR region of BcsC plays an integral role in the transport of cellulose polymers across the bacterial cell wall into the external environment where biofilm formation can occur. Future studies regarding BcsC would benefit in investigating potential protein-protein interactions with periplasmic proteins, such as BcsG, as well as profiling the β-barrel domain.

TABLE OF CONTENTS

List of Tables:

List of Figures:

List of Abbreviations:

3D……………………………………………………………………three dimensional

BCIP…………………………………………………….5-bromo-5-chloro-3-indolylphosphate

bcs…...…………………………….………………………… bacterial cellulose syntheses

BSA...…………………………………………………………bovine serium albumin

CAZy……………...………………………………………carbohydrate active enzymes

CBD…………………………………………………………carbohydrate binding domain

CBM…………………………………………………………carbohydrate binding module

CD………………………………………………………………circular dichroism

CMC………………………………………………………carboxy-methyl cellulose

c-di-GMP……………………………………………………yclic-di-GMP

E-gate…………………………………………………………extracellular gate

EPS…………………………………………………………… exopolymeric substance

GuHCl…………………………………………………………guanidine hydrochloride

Histidine tag……………………………………………………………His$_6$ tag

IMAC……………………..……………………immobilized metal affinity chromatography

IPTG…..…………………………………………isopropyl β-D-1-thiogalactopyranoside

LB……………………………………………………………………Lauria Bertani

LDAO…………………………………………… N,N-dimethyldodecylamine-N-oxide

NRMSD………………………………………….normalized root mean square displacement

NTB……………………………………………………………nitroblue tetrazolium

OD…………………………………………………..………………optical density

OGT………………………………………………………O-linked GlcNAc transferase

OMP…………………………………………………………outer membrane protein

PEG. …………………………………………………………..polyethylene glycol

P-gate……………………………………………………...………periplasmic gate

PNAG……………………………………………………poly-β-1,6-*N*-acetyl-ᴅ-glucosamine

SB……………………………………………………………….………super broth

SDS……………………………………………………………...sodium dodecyl sulfate

1. INTRODUCTION:

1.1 Bacterial Biofilms

A biofilm is the collection of one or more species of bacteria living in a self-produced protective coating attached to a surface (Banerjee and Joshi, 2013; Prakash *et al.,* 2003; Jonas, 2007). By living in these microecosystems, bacteria are able to utilize byproducts of other bacterial species allowing for communities to consist of a variety of bacterial species within a biofilm. The convenience of living with other bacteria that produce consumable byproducts is a contributing factor to the estimate that greater than 90% of bacteria exist in biofilms (Prakash *et al.,* 2003). The advantages also manifest in that bacteria found in a biofilm can be 10-1000 times more resistant to antibiotics than free-floating (planktonic) bacterial cells (Mah and O'Toole, 2001). It has been estimated that over 60% of infections are due to bacteria forming biofilms, making research on biofilm formation and turnover of great importance in prevention and treatment of disease (Lewis, 2001). In cystic fibrosis patients, the accumulation of thick secretions in the patient's airways allows for bacteria to colonize the respiratory tract of these patients (Girón *et al.,* 2005). One bacterial species that has been an opportunistic pathogen for people with cystic fibrosis is *Pseudomonas aeruginosa*, a biofilm forming bacteria that secretes large amounts of the polysaccharide alginate (Pier *et al.,* 2001; Whitney and Howell, 2013). Figure 1.1 depicts the presence of a *P. aeruginosa* infection (and the profuse amounts of polysaccharide produced) throughout the lungs of a cystic fibrosis patient.

Figure 1.1: A chest X-ray of a cystic fibrosis patient with a *P. aeruginosa* infection. *P. aeruginosa* is an opportunistic pathogen that secretes alginate during biofilm formation, evident by the white masses throughout the lungs (taken from Girón *et al.,* 2005).

Biofilms are also of importance to the food and bioremediation industries. Certain strains of bacteria, such as *Bacillus amyloliquefaciens* C06, can be used in the prevention of brown rot in post-harvest fruits (Liu *et al.,* 2011). The fruit is protected by the formation of a sticky biofilm consisting of γ-polyglutamic acid that prevents other bacteria, like *Monilinia fructicola*, from colonizing post-harvest fruits, such as apples and peaches (Liu *et al.,* 2010; Liu *et al.,* 2011). In bioremediation, microorganisms that are capable of forming biofilms on the surface of hydrocarbons are prime candidates for the breakdown of harmful chemicals. Biofilms are able to support a high biomass density; accelerating the degradation process of these complex chemicals, meanwhile still being able to maintain healthy living conditions (Singh *et al.,* 2006). Due to their prevalence it

is important to understand biofilms on many different levels including their community structure, the differences between sessile and planktonic forms, how biofilms are synthesized and the biological and chemical triggers that cause biofilm formation.

1.2 Biofilm Composition

The matrix of bacterial biofilms are often made up of aggregative proteins called fimbriae and a thick extracellular polysaccharide matrix, among other components. Figure 1.2 (A and B) depicts a network of curli fimbriae aggregative proteins extending from the surface of the bacterial cell. Figure 1.2 (C) also shows a bacterial cell encased in its matrix composed of cellulose chains. As a biofilm matures its components may change since different species may be introduced or removed, as well as shifts in species dominance. Mature biofilms can also consist of other extracellular matrix components synthesized by the encased bacteria, such as extracellular DNA, enzymes and nutrients (Flemming and Wingernader, 2010).
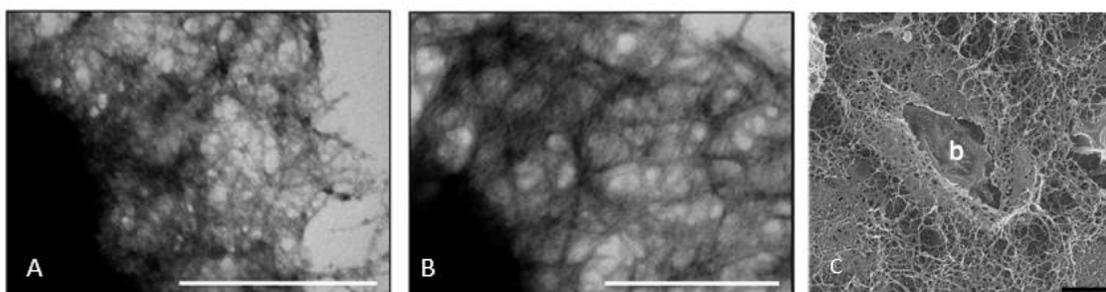
Figure 1.2: Biofilm components. Negative staining TEM of curli fimbriae produced by an (A) O157 strain and (B) non-O157 strain of *E. coli* (Biscola *et al.,* 2011) (C) Polysaccharide matrix composed of cellulose surrounding a bacterial cell (b), visualized by scanning electron microscopy (SEM) (Hung *et al.,* 2013). After the initial attachment of curli fimbriae to the surface the materials that make up the polysaccharide matrix are

exported from the cell, forming this complex network. Scale bars 0.5 μm A and B; 500 nm C.

The different components of a biofilm serve many purposes for the cells encased within it. The matrix functions both in adhering bacteria to one place, as well as acting as a scaffold that effectively facilitates the translocation of nutrients and numerous biologically active molecules for the bacteria (McDougald *et al.,* 2012). The matrix provides antibiotic resistance for the bacteria living within the biofilm; partially from the thickness of the matrix itself, as well as the potential of the matrix to exclude antimicrobials and impede harmful agents from reaching the encased bacteria (Costerton *et al.,* 1999). In addition to antimicrobial resistance, biofilms provide encased bacteria with other benefits, such as an increase in diversity within the biofilm. As communities form within the matrix, one bacterial by-product may be an energy source for another species (Singh *et al.,* 2006; White, 2007; McDougald *et al.,* 2012; Whitney and Howell, 2013). An increase in biodiversity also increases the opportunity for horizontal gene transfer. Through DNA exchange many species within the biofilm may acquire genes that provide antibiotic resistance, the ability to utilize alternative energy sources, or genes involved in increasing pathogenicity, to name a few. The biofilm matrix may also provide a means for evasion of the immune system within a mammalian host. Exopolysaccharides, such as cellulose, are often immunologically benign. When bacterial cells are encased in exopolysaccharides, antigens on the bacterial surface become hidden

to the host, preventing bacterial detection by the host's immune response system (Kline *et al.,* 2009).
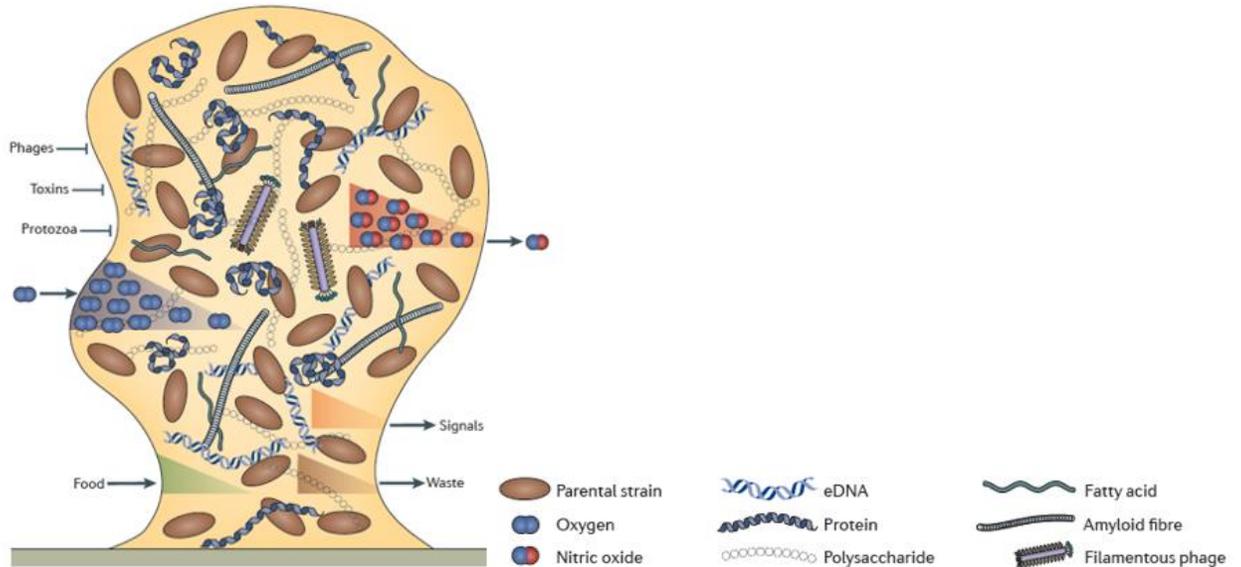
Figure 1.3: Schematic of a bacterial biofilm: a collection of one or more species living in a self-produced polysaccharide matrix. There are many elements to a biofilm, however, the main components are protein and the exopolysaccharide (represented by yellow between cells). The thick exopolysaccharide contributes to the establishment of gradients for oxygen, nutrient and waste diffusion through the biofilm (McDougald *et al.,* 2012).

Biofilm-forming bacteria are not limited to organic surfaces. Bacteria can form biofilms on living or dead tissues, medical devices, water pipes, surfaces in aquatic systems (such as rocks), or inert surfaces including plastics, metals and porcelain (Prakash *et al.,* 2003; Jonas, 2007). The ability of bacteria to form biofilms on many different types of surfaces is of importance particularly due to the ability of bacteria at the surface to disassociate from the biofilm (as described in the following section) and re-establish elsewhere, effectively transferring from one surface to another, potentially

causing infections. This is particularly important with surfaces that humans directly or indirectly interact with (such as spoiled food or places food and/or water come in contact with). Studying biofilms at their various levels of organization will help in the prevention of infections; thereby, directly improving public health.

1.3 Stages of Biofilm Development

Bacterial biofilm formation has four main stages of development; adhesion, formation of a monolayer, maturation and dispersal. Figure 1.4 demonstrates a general overview of the lifecycle of a biofilm. In *Salmonella* spp. and *E. coli*, biofilm formation begins with an irreversible attachment of planktonic cells using aggregative proteins that extend from the cell and adhere to the surface where the biofilm will form, these may include proteins such as curli fimbriae, type 1 fimbria or Ag43 (Saldaña *et al.*, 2009; White, 2007; Jonas *et al.,* 2007, Biscola *et al.*, 2011). Figure 1.2 displays curli fimbriae proteins surrounding an *E. coli* cell. In humans and animals the curli fimbriae may attach to host proteins located in the extracellular matrix, like major histocompatibility complex class I proteins on the surface of host cells and laminin of the basal membrane (Saldaña *et al.,* 2009; Prakash *et al.,* 2003). Adherence can also be mediated by other proteins, for example Type 1 fimbriae are responsible for the adherence and invasion of *E. coli* cells in the intestinal epithelial cells of patients with Crohn's disease (Boudeau *et al.,* 2001). Biofilms can also form on plant surfaces by attaching to proteins on roots, leaves, vasculature and in intercellular spaces (Danhorn and Funqua, 2007). The typical surface by which bacterial cells attach to are rough, hydrophobic and coated by conditioning

films made up of nutrients bound to the surface (Prakash *et al.*, 2003). Once a bacterial cell has made an initial attachment to a surface, gene expression changes during the morphological switch from planktonic cells to the sessile lifestyle (White, 2007). The changes from one lifestyle to another includes the up-regulation of genes responsible for biofilm formation, these genes will be discussed later in section 1.7. In the case of motile cells, the down-regulation of mobile associated genes, such as flagella associated genes, also occurs. Motile cells have been found to be more capable of adhering to a surface than cells that do not form flagella or other extensions from the membrane for cell mobility (Prakash *et al.,* 2003; Pratt and Kolter, 1999). Cell mobility is particularly important for biofilms forming where there is liquid flowing over the surface. Allowing the cell to remain in close proximity to the surface (*i.e.*, not washed away by the general flow of the liquid) plays an important role in cell attachment (Prakash *et al.*, 2003). Once the cells are attached to the surface, the need for motile extensions is decreased and at this point the genes for these structures are down-regulated.
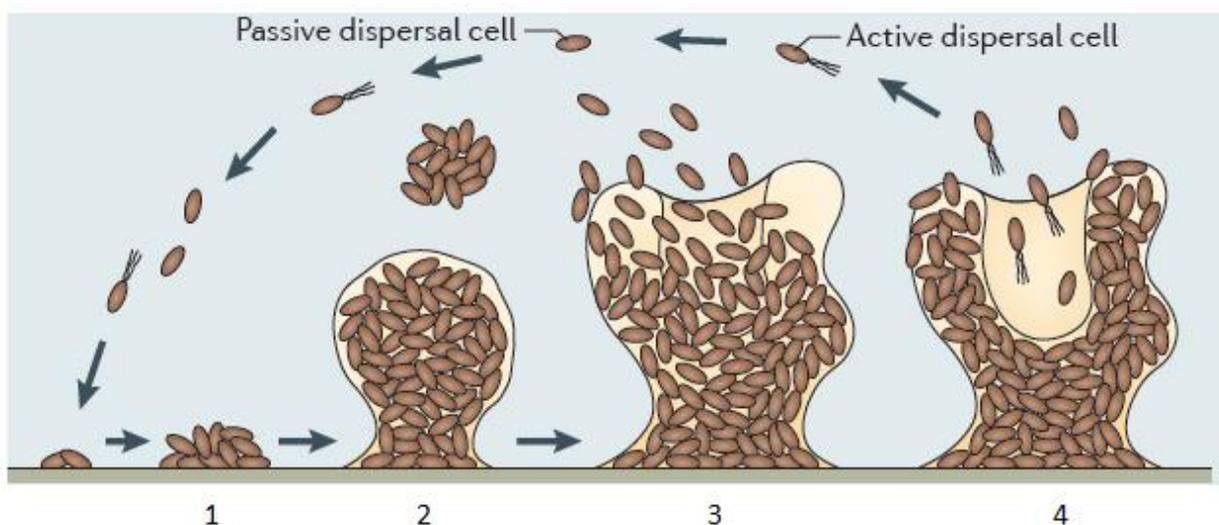
Figure 1.4: Stages of biofilm development. Newly attached cells form the early stages of the biofilm (1). As cells grow and divide they start forming a distinctive structure to the biofilm based on the arrangement of microcolonies within the biofilm (2). Once the biofilm has matured, dispersal cells can leave the biofilm either by the removal of cells due to environmental factors causing cells to slough off (3) or by changing their growth mode to the planktonic form (4) (McDougald *et al.,* 2012).

The next stage in biofilm formation involves the copious production of an exopolymeric substance (EPS) that aids in the formation of a monolayer of bacterial cells (as depicted in Figure 1.4, panel 2). As mentioned above, EPS consists of extracellular polysaccharide, protein, extracellular DNA and other substances (Felmming and Wingernader, 2010; White, 2007; McDougald *et al.,* 2012). At this stage of biofilm formation the cells become irreversibly attached to a surface and the individual cells become coated in their respective EPS substances (White, 2007). Using type IV pili, bacteria are still able to spread over the surface, thereby increasing the surface area to build the biofilm upon (White, 2007).

The maturation stage of biofilm formation begins with the thickening of the biofilm matrix and multiplication of the cells embedded therein (Figure 1.4, panels 2 and 3). As bacteria produce more exopolysaccharide smaller colonies within the biofilm, termed "microcolonies", begin to form and remain held together by the EPS that the bacteria are producing (White, 2007). Concomitant with the formation of microcolonies, solvent channels are also established throughout the matrix to ensure the bacteria maintain access to water, oxygen and nutrients (White, 2007; Prakash *et al.,* 2003). As the biofilm develops, microcolonies grow larger through cell division of individual bacteria causing a change in the morphology of the microcolonies and of the whole biofilm (White, 2007; Mah and O'Toole, 2001; White, 2007; McDougald *et al.,* 2012). Eventually the biofilm becomes large enough that it begins to segregate into microniches. Bacteria on the bottom of individual colonies of the mature biofilm become buried and have reduced access to nutrients, water and other metabolites (and/or removal of detrimental metabolic waste byproducts) that can lead to cells that are less metabolically active and/or dormant. However, bacteria growing on the surface of a biofilm have continued unhindered access to solvent channels and these cells can maintain relatively high metabolic activity (Prakash *et al.,* 2003).

The final stage of biofilm formation involves the dispersal of cells from the mature biofilm (White, 2007; McDouglad *et al.,* 2012). Cells detach from the EPS through shearing forces and/or by coordinated detachment signals from cells within the biofilm (Figure 1.4, panels 3 and 4). The released cells then become motile and are able

9

to begin the colonization and biofilm formation process elsewhere (White, 2007; McDouglad *et al.,* 2012). Coordinated detachment signals are stimulated by a decrease in resources, such as carbon and nitrogen, and a variety of other factors. After long periods of starvation and repeated detachment events (sloughing off of cells), a biofilm can diminish in size (White, 2007).

1.4 Cellulose

Cellulose is the most abundant biological polymer and varies in its organization and uses among the different organisms that produce it (Morgan *et al.,* 2013; Omadjela *et al.,* 2013). At the structural level, cellulose is an organic polymer of glucose molecules connected by β-1,4 linkages into chains of various sizes (Zogaj *et al.,* 2013; Omadjela *et al.,* 2013). Cellulose can be found in plants, algae, fungi, flagellates and some animals (Ross *et al.,* 1991; Zogaj *et al.,* 2001). Figure 1.5 depicts cellulose produced by a variety of organisms, which can adopt varying macromolecular structures/configurations. In plants, cellulose is a key component of the primary and secondary cell walls and is organized into microfibrils (Czaja *et al.,* 2007; Mizrachi *et al.,* 2012). Plant cellulose is also associated with other materials in the cell wall such as lignins and hemicelluloses (Czaja *et al.,* 2007). The microfibril arrangement of cellulose in plants provides the basis for absorption in plant-based cellulose products such as cotton gauze (Czaja *et al.,* 2007). As an implant, plant cellulose has shown biocompatibility with bone tissue and hepatocytes but displays slow biodegradability preventing its use as permanent biomedical material (Czaja *et al.,* 2007; Märtson *et al.,* 1999).
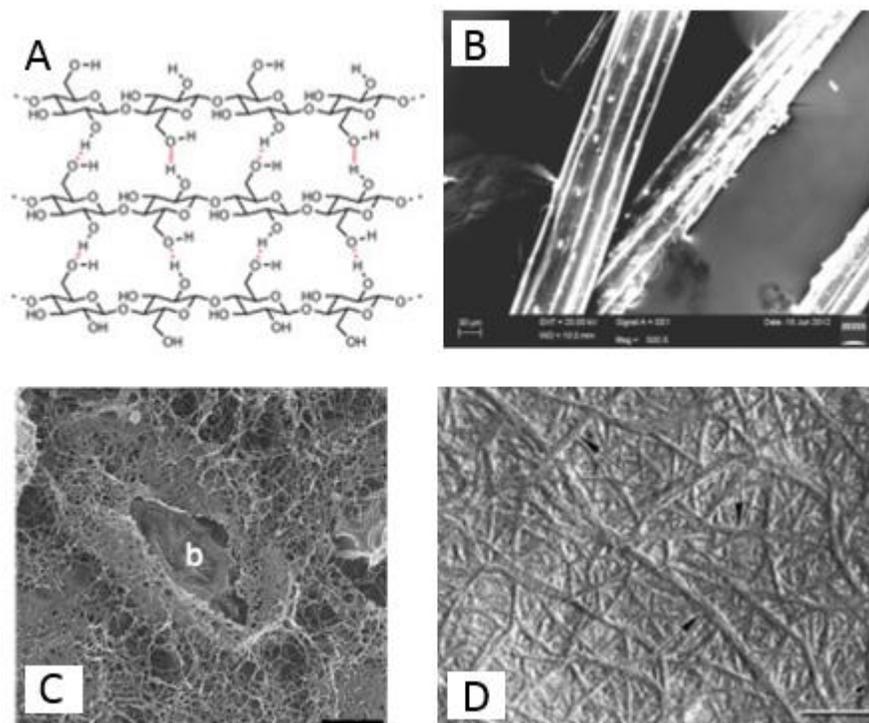
Figure 1.5: Cellulose at different levels of organization. Cellulose is made up of β 1-4 linked glucose molecules that form a network as portrayed in (A) (Modified from Seery, 2013). (B) Depicts an SEM image of fibers from *Hibiscus sabdariffa* made up of cellulose, hemicelluloses and holocelluloses that are located in the cell wall (Sonia and Priya Dasan, 2013). In comparison to the fibers produced in plant species, bacteria produces cellulose in the form of extracellular extensive networks as seen in (C); In the (C) panel, the "b" indicates the bacterium surrounded by the fibrous network (Hung *et al.,*, 2013). In (D) the algae slime mold *Dictyostelium discoideum*, cellulose is produced both in the cell wall and intercellularly during late aggregation and continues through culmination phases. The arrows indicate bunches of microfibrils (Grimson *et al.,* 1996). Scale bars indicate 30 μm (A), 500 nm (C), and 100 nm (D).

In contrast to plant cellulose, bacterial cellulose is extracellular and is organized

into nano- and microfibrils that are present throughout a biofilm matrix (Czaja *et al.,*

2007) (Figure 1.5, panel C). The combination of the hydrophilic nature of cellulose and its network of microfibrils creates the ability to chelate a large amount of water, resulting in a hydrogel (Czaja *et al.,* 2007; Peterson and Gatenholm, 2011). Bacterial cellulose is also able to conform/mold to different surfaces and is mechanistically sound due to the presence of hydrogen bonds between the fibrillar units (Czaja *et al.,* 2007). There have been many different proposed uses for bacterial cellulose, such as applications in assisting wound healing and blood vessel replacements, as well as possible roles in forming scaffolding for guided tissue regeneration and tissue engineering (Czaja *et al.,* 2007; Petersen and Gatenholm, 2011, Bodin *et al.,* 2007). An additional example of cellulose commonly used in the biomedical industry is bacterial cellulose membranes, which are commercially available and have been used for wound dressings on victims of severe burns and other skin conditions (Bordin *et al.,* 2007; Czaja *et al.,* 2007; Peterson and Gatenholm, 2011). Bacterial cellulose membranes, depicted in Figure 1.6, have been shown to reduce pain, infection, scarring and exudate buildup, all while keeping the area moist for healing (Czaja *et al.,* 2007; Peterson and Gatenholm, 2011). Bacterial cellulose is easily purified and has not been shown to degrade in the body which makes cellulose a prime candidate for biomedical materials (Klemm *et al.,* 2001; Czaja *et al.,* 2007).
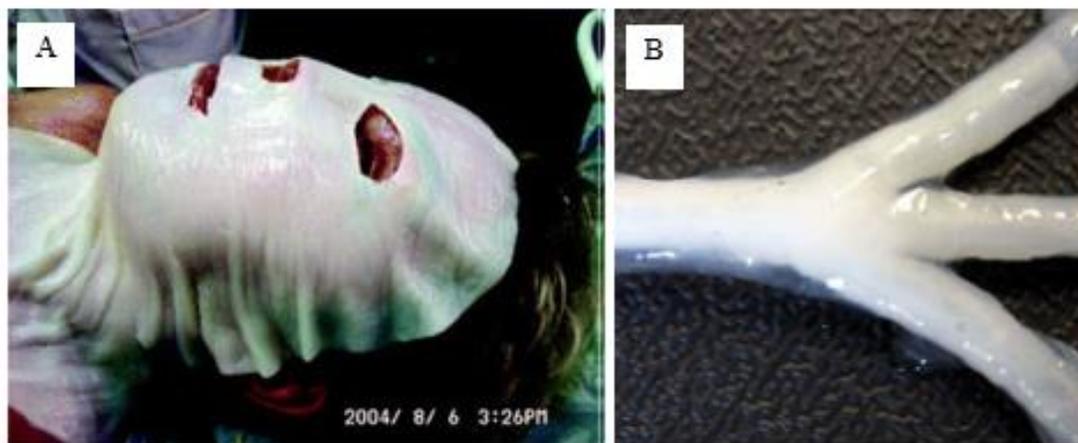
Figure 1.6: Cellulose membranes for biomedical purposes. Cellulose can be grown to adopt many different shapes. A) Cellulose sheets have been used as a topical application for wound healing in burn victims. Benefits to using cellulose include the reduction of wound pain, faster healing, decreasing the likeliness of developing infection in wounds, inhibition of exudate, and allows caretakers more ease in wound inspection (Petersen and Gatenholm, 2011). B) Bacterial cellulose grown on silicon tubing. Using different growth conditions, a variety of cellulose tubing sizes can be created with the intension for possible vessel replacement in procedures such as bypass surgeries (Bordin *et al.,* 2007).

1.5 Disadvantages and Advantages to Bacterial Biofilms

Biofilms play an important role in many environmental applications and often lead to beneficial effects in bioremediation. Many Enterobacteriaceae, such as certain strains of *E. coli* and *Salmonella* spp., are capable of forming biofilms using cellulose as a predominant exopolysaccharide within the biofilm matrix (Saldaña *et al.,* 2009). When pathogenic strains of *E. coli* or *Salmonella* form cellulose biofilms, they pose a larger threat of persistence within the host because the biofilm aids in protecting the bacteria from the immune system (Saldaña *et al.,* 2009; Kline *et al.,* 2009). Research on cellulose biofilm formation and disassembly is, therefore, imperative for developing treatments

against pathogenic biofilm-forming organisms. However, not all biofilm-forming bacteria are harmful and many play an important role in various industries. For example, biofilm-forming bacteria have been shown to be highly effective in treating compounds and biological contents that are otherwise recalcitrant to degradation, such as chlorophenols, azo dyes and herbicides (Singh *et al.,* 2006). Bacterial communities within a biofilm that can degrade pollutants will often "feed off" one another's by-products, providing the full breakdown of these harmful substances. To optimize the efficiency of remediation processes many researchers have used genetic engineering to improve bacterial strains involved in bioremediation using biofilms (Singh, 2006). As we continue to understand more about biofilms, more applications continue to be found. It is, therefore, important to understand the many roles biofilms play, the structural differences of various biofilms, as well as the synthesis and export of exopolysaccharides for biofilm formation. The machinery responsible for generating different biofilm EPS varies between the different systems, all of which have substantial homologous features to one another. By studying the synthesis and export machinery of one polysaccharide, the information gained may be transferrable in studying unclassified proteins of other biofilm-forming systems.

1.6 *Escherichia coli* as a Model Organism

With respect to the present research, *E. coli* is a great model organism for a number of reasons. Firstly, there is a plethora of genetic/molecular tools available for the study of particular genes and proteins in this organism. Secondly, there are many strains of commensal, non-pathogenic and pathogenic *E. coli* that are capable of producing

cellulose-containing biofilms. In certain pathogenic strains, the ability to form cellulose-containing biofilms allows the bacteria to evade the immune system, thus contributing to their ability to persist in a host and potentially express additional pathogenic traits. For example, *E. coli* O157:H7 EDL933 is a biofilm-forming enterohemorrhagic pathogen capable of forming cellulose biofilms (Henrissat *et al.,* 2002). *E. coli* O157:H7 EDL933 was responsible for making 2300 people severely ill and contributed to seven deaths during the Walkerton, Ontario *E. coli* outbreak in May of 2000 (Barnett Foster, 2013; Hrudey *et al.,* 2003). The *E. coli* outbreak in Walkerton is merely one local example of many waterborne outbreaks. Previous to Walkerton, there had been 15 cases of *E. coli* waterborne disease outbreaks in Canada and the United States since the mid 1970's. Each of these outbreaks affected between 47 and 400 000 people and totaled 11 deaths (Hrudey *et al.,* 2003). Different strains of *E. coli*, both pathogenic and non-pathogenic, have also been found to form biofilms on food products, particularly plant foods, such as spinach, sprouts and lettuce (Macarisin *et al.,* 2013; Taormina *et al.,* 1999 Ackers *et al.,* 1998). Further studies are required to determine the different roles of cellulose biofilms in the many different strains of pathogenic and commensal strains of *E. coli* and the differences between synthetic machinery.

1.7 Cellulose Gene Expression and the bcs Complex in Enterobacteriaceae

During the switch from planktonic to sessile growth modes (*i.e.*, biofilm formation), there is a difference of approximately 40% in the regulation of a subset of genes responsible for cellular growth (Potera, 1999; Davies and Geesey, 1993; Prakash *et*

al., 2003). This phenomena includes the down-regulation of genes involved in flagellin synthesis and the up-regulation of genes involved in the production of bacterial cellulose (Prakash *et al.,* 2003). When environmental conditions change to benefit bacteria in a sessile lifestyle, cellular levels of the second messenger cyclic-di-GMP (c-di-GMP) increase, causing the up-regulation of biofilm formation associated genes (Morgan *et al*., 2013). For example, c-di-GMP is involved in controlling cellulose biosynthesis at both the gene expression (Merighi *et al.,* 2007), and protein synthesis/export levels (Morgan *et al.,* 2013).

Cellulose production for biofilm formation is under the control of two bacterial cellulose synthesis (*bcs*) operons outlined in Figure 1.7. The first operon, *bcsABZC*, contains genes that encode for the expression of proteins in the bacterial cellulose synthesis complex BcsA, BcsB, BcsZ and BcsC, respectively, and will be discussed in more detail below (Zogaj *et al.,* 2001; Whitney and Howel, 2013; Bokranz *et al.,* 2005). A second cellulose operon, *bcsEFG,* lies both adjacent and divergent to the first operon, and contributes to cellulose synthesis in certain bacteria (primarily thought to be members of the Enterobacteriaceae), but is not dispersed through all bacteria that possess a cellulose complex. In a study performed by Solano and colleagues (2002), the mutation of *bcsE* resulted in *Salmonella enteritidis*, a food borne pathogenic bacteria, losing the ability to form a biofilm. However, more recently it has been found that the deletion of BcsE merely prevents maximal cellulose production, provided the downstream *BcsFG* genes were still expressed (Fang *et al*., 2014). Furthermore, BcsE has been discovered to

contain a C-terminal GIL domain responsible for c-di-GMP binding, leading to the hypothesis that BcsE may play a role in a two-tiered control mechanism (with BcsA) in the transition from the motile lifestyle to a sessile state (Fang *et al*., 2014). Research continues to characterize the N-terminal domain of BcsE, using homology searches it has been hypothesized that the N-terminal domain of BcsE may play a role as a protease (Solao, 2002). However, recent research in the Weadge lab has found no protease activity to date (Brenner and Weadge Unpublished, 2015).

The remaining proteins of the *bcsEFG* operon, BcsG and BcsF, have remained elusive. BcsG has shown homology to endoglucanase-encoding genes leading to the hypothesis that BcsG may be involved in cleaving cellulose in the periplasm (Solano, 2002; Nakamura *et al.,* 1986). However, BcsE demonstrated no endoglucanase activity by standard assays; thereby suggesting that this protein may have an as yet undefined role (Razvi and Weadge Unpublished, 2014). Currently, no close homologs of BcsF have been found/characterized to provide further insight into its function (Le Quéré and Ghigo, 2009). Due to the lack of information available on the proteins produced by genes in the *bcsEFG* operon, this document will focus on the current studies with respect to the *bcsABZC* operon.
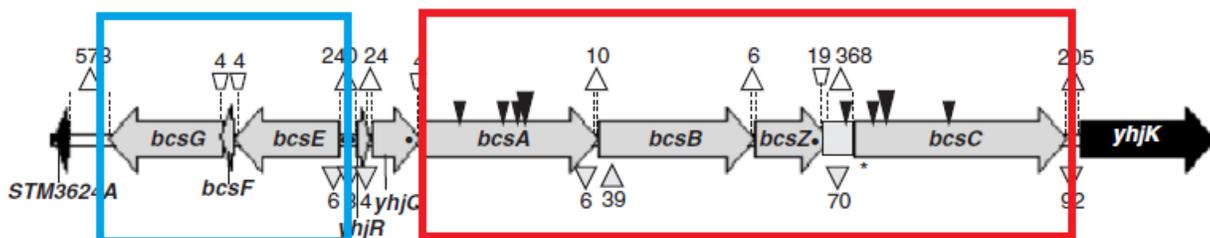
Figure 1.7: Depiction of the direction and arrangement of the two *bcs* operons from *Salmonella enteritidis*. The first operon, *bcsABZC* (outlined in red), contains the genes which encode for proteins BcsA, BcsB, BcsZ and BcsC, respectively. The second operon, *bcsEFG* (outlined in blue) runs both adjacent and divergent to the first. In *E. coli* and *Salmonella* species, both operons are required for cellulose biofilm formation (adapted from Römling, 2007).

Bacterial cellulose biosynthesis begins with BcsA and BcsB. These two proteins not only polymerize the growing cellulose chains, but also concomitantly transport it across the inner membrane. BcsA and BcsB form a complex through the inner membrane that extends into the periplasm (Zojag *et al.,* 2001; Morgan *et al.,* 2013). The three-dimensional (3D) model of the complex formed by BcsA and BcsB can be seen in Figure 1.8. BcsA has many domains which are ultimately responsible for cellulose synthesis and transport across the inner membrane. Polymerization of cellulose chains entails forming a β-1,4 linkage between UDP activated glucose (UDP-glucose) and a glucose residing on the end of the growing cellulose chain in the glycosyltransferase domain of BcsA (Morgan *et al.,* 2013). The activity of the glycosyltransferase domain is positively influenced through conformational changes that occur in a neighbouring PilZ domain of BcsA upon binding of the cellular messenger molecule, c-di-GMP (Morgan *et al.,* 2013; Omadjela, 2013). BcsA is also responsible for transferring the growing chain across the

18

inner membrane via a transmembrane polysaccharide channel that spans the inner membrane (Morgan *et al.,* 2013; Omadjela, 2013). At the periplasmic face of the inner membrane, the growing cellulose chain emerges where BcsB interacts with BcsA (Morgan *et al.,* 2013). BcsB contains two carbohydrate binding domains (CBD), CBD1 and CBD2. Structural studies indicate that CBD2 interacts directly with BcsA, while CBD1 is proposed to interact directly with the translocating polysaccharide (Morgan *et al.,* 2013).
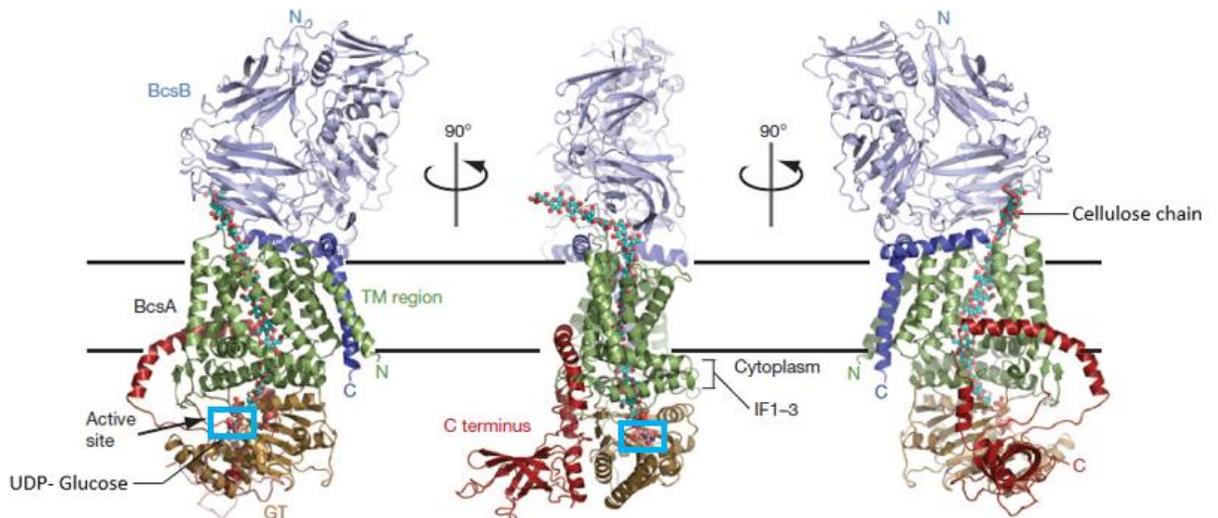
Figure 1.8: The BcsA/BcsB cellulose synthesis complex. Left panel - UDP-glucose molecules (blue box) are added to the growing glycan chain at the active site (brown area of BcsA) of BcsA in the cytoplasm. As each sugar molecule is added, the cellulose chain (blue and pink) advances through the transmembrane region (green area of BcsA) to the periplasmic space (seen in different orientations in all panels). BcsB (depicted in purple) consists of carbohydrate binding domains that assist in bringing the cellulose across the inner membrane (Modified from Morgan *et al.,* 2013).

The next protein in the *bcsABZC* operon that has yet to be introduced is BcsZ. In certain cellulose-producing bacteria, BcsZ homologues are not part of the cellulose

biosynthetic operon, such as the CMCax protein of *Acetobacter xylinum* (Standal *et al.,* 1994). However, in *E. coli*, *Salmonella enterica*, *Vibrio fisheri* and other Enterobacteriaceae, *bcsZ* has a conserved location among the *bcs* cluster of genes (Zogaj *et al.,* 2001). Mazur and Zimmer (2011) have demonstrated that BcsZ adopts an α_6/α_6-barrel fold and belongs to carbohydrate active enzymes (CAZy) glycoside hydrolase family 8. BcsZ is a predicted periplasmic protein and its endo-β-1,4-glucanase activity is required for optimal biosynthesis of cellulose by *E. coli* (Mazur and Zimmer, 2011). The specific role of BcsZ has been speculated to be involved in either the breakdown of cellulose that has accumulated within the periplasmic space and/or to cleave growing chains of cellulose into desirable lengths for proper microfibril formation in the biofilm once exported from the cell (Mazur and Zimmer, 2011).

Once synthesized and translocated into the periplasm, the cellulose chain is then ready for export from the cell. Although the initial stages in cellulose biosynthesis and transport are known, much less is understood with regards to the final steps involving export of the polymer from the cell. Much of the information with regard to the export of cellulose from bacteria has been garnered from genetic manipulation and sequence homology studies with the highly conserved *bcsC* gene in the *bcsABZC* operon. For example, studies with BcsC in *Acetobacter xylinus* indicate that this protein does not need to be present for cellulose synthesis to take place *in vitro*, but when absent *in vivo*, cellulose synthesis cannot be detected showing evidence that the cellulose is not being exported (Wong, 1990). Zogaj and colleagues (2001) found that when BcsC was mutated

in the *Salmonella typhimurium* MAE52 strain, cellulose could not be produced and there was a severe reduction in biofilm formation by this bacterium. Sequence homology to polysaccharide export proteins in other systems also suggests that *bcsC* encodes for a protein that plays an integral role in the export process (Keiski *et al.,* 2010). BcsC is predicted to consist of two distinct domains; an outer membrane β-barrel and a periplasmic tetratricopeptide repeat (TPR) region (Keiski *et al.,* 2010; Römling, 2007). The TPR region of BcsC is predicted to be involved in protein-protein interactions and possibly protein-carbohydrate interactions that facilitate transport of the cellulose polymer across the periplasm to the β-barrel portion in the outer membrane, which, in turn, facilitates export from the cell (Whitney and Howell, 2013; Römling, 2007). TPR motifs and β-barrel domains with high homology to BcsC are also present in analogous exopolysaccharide biosynthetic systems, but these domains are found in separate proteins rather than on a single polypeptide, like BcsC. For example, the alginate biosynthetic system in *P. aeruginosa* employs the β-barrel protein AlgE and the TPR protein AlgK for the export of alginate in biofilm formation. Similarly, Staphylococcal species and certain strains of *E. coli,* such as K12, have the ability to produce a poly-β-1,6-*N*-acetyl-D-glucosamine (PNAG) biofilm using the predicted export protein, PgaA which also encodes for β-barrel and TPR regions on a single polypeptide (Itoh *et al.,* 2008). The homology of BcsC with other export complexes in biofilm-forming systems is an important aspect for the elucidation of BcsC structure and function as research performed on these previously described systems can be used as a template for BcsC research.

Figure 1.9 demonstrates the homology of the *bcs* complex with the two other biofilm

forming systems, alginate and PNAG.
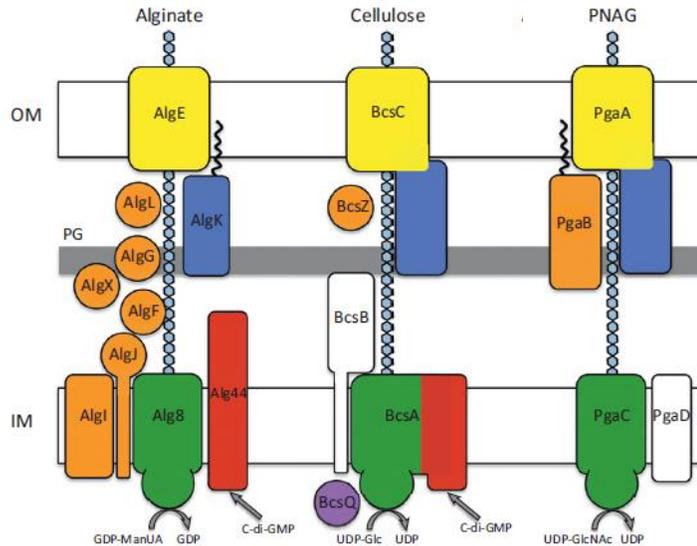
Figure 1.9: Polysaccharide synthesis and export. This is the proposed schematic for the alginate, cellulose and PNAG systems created by Whitney and Howell (2013). Each of these systems follows the same overall method by which they make (green and red modules) and export (blue and yellow modules) their respective polysaccharide. By following the protocols set out from previously characterized proteins in other systems, similar techniques can be used to define uncharacterized proteins.

1.8 Predicted Structural and Functional Characteristics of BcsC

BcsC has previously been hypothesized to be involved in cellulose export for

biofilm formation via the *bcs* system. As aforementioned, BcsC displays sequence

homology to other polysaccharide export proteins. These include AlgE/AlgK of the

alginate biosynthesis system in *P. aeruginosa* and PgaA of the PNAG biosynthesis

system in Staphylococcal species and certain strains of *E. coli*. Each of these systems

contain a TPR extending into the periplasmic space and a β-barrel spanning the outer membrane.

A TPR is typically composed of a 34 amino acid-long peptide sequence that forms two antiparallel α-helices in a helix-turn-helix motif (D'Andrea and Regan, 2003; Jernigan and Bordenstein, 2015). The spatial arrangement of the TPR motif forms an overall right-handed super helix containing a continuous helical groove (Das *et al.,* 1998). It is this helical groove that is characteristically involved in protein interactions. A TPR motif typically contains 3-16 TPR repeats with a final helix at the end, which is believed to be involved in stabilizing the protein (Jernigan and Bordenstein, 2015). TPRs have been found in all classes of life; however, bacteria and archaea have been found to be more enriched with TPR proteins in comparison to eukaryotes (Jernigan and Bordenstein, 2015).

Although the structural and functional analysis of the TPR domain of BcsC can be linked to previous studies performed using AlgK and PgaA there are a few key differences to note between the systems. The first obvious difference between the alginate and cellulose polysaccharide export systems is the fact that there are two separate proteins in the alginate system, the cellulose export complex is composed of only one protein (Figure 1.9). In order for AlgK to remain in close proximity to AlgE, and to be correctly oriented in the periplasm, Keiski and colleagues (2010) confirmed the presence of a lipidation site to anchor the protein to the outer membrane. The alginate

system also possesses a large number of accessory proteins (also seen in Figure 1.9) to aid in the translocation of the newly synthesized alginate strands to the outer membrane. These accessory proteins appear to be absent in the BCS system, thus offering a potential explanation as to why the BcsC TPR is longer than that of AlgK in the alginate system. The longer TPR domain on BcsC may be all that is needed to facilitate the transport of newly synthesized cellulose from the inner to the outer membrane in these bacteria. In comparison, PgaA is similar to BcsC in that it contains the outer membrane porin and TPR on the same polypeptide. However, export in this system also requires PgaB for partial d-*N*-acetylation of the PNAG chain (Little *et al.,* 2012) and so a shorter predicted TPR region may allow for the presence of this protein in the PNAG biosynthetic complex. In the BCS system researched in this study there are no accessory proteins involved in altering the chemical composition of cellulose, as seen in the PNAG system. Rather, the only alterations believed to occur are performed by BcsZ, as its predicted function is to cleave cellulose in the periplasmic space.

There are many things we can still learn from these previously studied proteins. In the alginate system, it was found that AlgK is responsible for facilitating AlgE localization to the outer membrane. When the lipidation site of AlgK was mutated, membrane fractionations found AlgE in both inner and outer membrane fractions (Keiski *et al.,* 2010). It is possible that the TPR domain of BcsC may also play a crucial role in localizing the β-barrel portion of BcsC to the outer membrane. With respect to the β-barrel porin, AlgE, of the alginate biosynthesis system has been found by Tan and

colleagues (2014) to contain two ordered loops that allow for the export of alginate, while preventing leaking of substrate into or out of the outer membrane. By mapping the residues of the β-barrel portion of BcsC to AlgE, we may be able to identify similar loops in the porin domain of BcsC.

2. RESEARCH NEED

Research into bacterial cellulose biofilms is of significance to many different areas within the medical field. A greater understanding of bacterial cellulose biofilms and bacterial cellulose production in general is likely to aid in the discovery of additional industrial/medical applications (e.g. acoustic membranes, wound dressings, implants and tissue regeneration scaffolds), while also improving the ability to combat biofilm-related infections. Since BcsC is believed to be the final protein involved in cellulose export for biofilm formation, it is a prime candidate for drug development in the prevention of food/water-borne diseases caused by *Salmonella* spp. and *E. coli*. According to the Centers for Disease Control and Prevention, nontyphoidal *Salmonella* spp. and *E. coli* (STEC) O157 infections hospitalize an estimated 19336 and 2138 people annually in the United States, respectively, from contaminated food (CDC, 2011). Furthermore, foodborne nontyphoidal *Salmonella* spp. cause an estimated 378 deaths in a year (CDC, 2011). Bean sprouts have been found to be a common vector for harmful strains of *Salmonella* spp. and *E. coli*. From 1995 to 2011, there was a reported 1000 cases of sprout-borne illnesses in eight separate outbreaks, the most devastating of which occurred in 2005 in Ontario where over 648 cases of *Salmonella* spp. poisoning were reported (Healthy Canadians, 2013). A BcsC targeted drug could help in preventing outbreaks from occurring by bacterial cellulose biofilm producing bacteria, such as the *E. coli* outbreak in Walkerton. Before any drugs can be designed to specifically target BcsC, the structure of BcsC must be determined and the function must be ascertained.

The present research has focused on determining the structure and function of BcsC as it has been hypothesized to play a vital role in cellulose biofilm formation by a number of important pathogens. Studying BcsC will provide a better understanding of bacterial cellulose because the *bcsC* gene is found in a broad range of cellulose-producing bacteria (e.g. members of the Enterobacteriaceae, Pseudomonadaceae and Xanthomonadales families), in comparison to proteins encoded in the *bcsEFG* operon which are only found in certain bacteria such as *E. coli* and *Salmonella* spp.. This research will contribute to knowledge on exopolysaccharide biosynthetic systems that are distinct from other well-characterized systems in bacteria, such as capsules and lipopolysaccharide biosynthesis, which will ultimately aid in combating biofilm-related bacterial infections, develop new biomedical devices and improve bioremediation approaches. Due to the homology of BcsC to other polysaccharide export proteins, the location of the *bcsC* gene within the *bcsABZC* operon, and the location of BcsC at the outer membrane, it has been hypothesized that BcsC is composed of an outer membrane β-barrel with a TPR extending into the periplasmic space, and that this protein is critical for proper cellulose export that ultimately aids in biofilm formation.

3. OBJECTIVES

3.1 Long Term Objectives

The overall objective of this research project was to determine the structure and function of BcsC. To investigate BcsC, the protein was separated into three main protein construct groupings (full, β-barrel and TPR) and each construct individually analyzed for protein structure and function. Provided these goals are met, information gained from BcsC characterization will be available for use in future studies on biofilm formation and to study cellulose biofilm producing bacteria.

3.2 Short Term Objectives

The more specific, short-term objectives of this research project are to:

1. Provide a general overview of the predicted characteristics of BcsC and provide a wider context of the homology of this protein to other known and unknown genes/proteins using bioinformatics tools. Bioinformatics are instrumental as a predictive tool for the design of specific constructs with particular domain boundaries.

2. Profile optimal expression and purification conditions of the various BcsC constructs

   - These studies are important for selecting which constructs provide a substantial amount of highly purified protein that is needed for further experiments pertaining to protein structure and function.

3. Perform structural analysis of BcsC

- Structural studies with BcsC provide insight into the overall fold, stability and critical residues involved in its function.

4. Perform functional analysis of BcsC

- Functional studies are focused on determining the roles of the predicted TPR and β-barrel domains of BcsC. These studies include probing for protein-carbohydrate interactions between the TPR domain of BcsC with soluble and insoluble forms of cellulose.

4. MATERIALS AND METHODS

4.1 Reagents and Media

Reagents used primarily for this study were obtained from Fisher Scientific, Bioshop and BioBasic. Reagents to make Lauria Bertani (LB) broth and super broth (SB) including tryptone powder, yeast extract and agar were purchased through BioBasic, NaCl from Fisher Scientific and isopropyl β-D-1-thiogalactopyranoside (IPTG) and kanamycin sulfate from Bioshop. Buffers made for purification and Western blot examination employed reagents including Tris purchased through Fisher Scientific, imidazole, urea, RNase A, glycine and primary and secondary antibodies were purchased through Bioshop, DNase I was purchased through Boehringer Mannheim and non-fat milk powder purchased from BioBasic. Protein folding detergent N,N-dimethyldodecylamine-N-oxide (LDAO) was purchased through Anatrace. Additional reagents required for preparing sodium dodecyl sulfate (SDS) sample buffer were purchased through Fisher Scientific, including glycerol, SDS and bromophenol blue. The remaining reagents used for SDS-PAGE, Coomassie stain and destain include bis-acrylamide from Bioshop and methanol, TEMED, coomassie R250 and acetic acid were purchased from Fisher Scientific. Cellulose products, including Avicel and carboxymethyl cellulose, were purchased through Sigma.

Media for the following experiments are common nutrient rich broths. LB, to grow starter cultures post transformation and for general culturing, consisted of tryptone (10 g/L), yeast extract (5 g/L) and NaCl (10 g/L). When solid media was used, 1% (w/v)

agar was added to the LB media. SB, used for large scale protein expression experiments, consisted of tryptone (32 g/L), yeast extract (20 g/L) and NaCl (10 g/L). Where necessary, kanamycin sulfate was added to the media to a final concentration of 50 µg/mL.

4.2 Objective 1: Bioinformatics Analysis of BcsC and Construct Generation

4.2.1 Bioinformatics

Preliminary research was performed on the BcsC amino acid sequence using publically available bioinformatics tools. The first of these tools was ProtParam, which provided basic information including the number of amino acids in the sequence, the molecular weight of the protein, the percentage of each amino acid in relation to the amino acid sequence and a predicted extinction coefficient (Gasteiger *et al.*, 2005). Regions of disorder within the amino acid sequence were then identified using DisEMBL through the DisProt website. Next, the cellular location of the protein based on its amino acid sequence was predicted using PSORT (Gardy *et al.*, 2005). SignalP and LipoP were then used to predict the presence of signal peptide cleavage sites and lipoproteins, respectively (Nordahl *et al.*, 20011; Juncker *et al.*, 2003). TMHMM2.0 was used to evaluate for the presence of any potential transmembrane regions of the protein TMHMM2.0 (Moller *et al.*, 2001). To determine regions that could make up TPR segments of the protein, the program TPRpred was employed (Karpenahalli *et al.*, 2007). Next, secondary structure was predicted using PSIPRED and Phyre[2] from the amino acid sequence of BcsC (Buchan *et al.*, 2013; Kelley *et al.*, 2015). Phyre[2] was also used to

generate a 3D model of the protein based on hidden markov homology to previously characterized proteins (Kelley *et al*., 2015). The 3D models created in Phyre[2] were then used in PyMOL to adjust the position of the model. PyMOL was also used to depict distances on the models using the wizard measure tool (PyMOL, 2011). Using the combined information from these tools, potential TPR and β-barrel domain boundaries were identified and used to create protein constructs composed of three main groups: TPR, β-barrel and full-length constructs. All constructs generated are listed in Table 4.1. All constructs discussed here in will be referred with the amino acid residue boundaries in superscript of BcsC (*i.e.* BcsC$^{24\text{-}742}$ will be used for the construct from amino acids 24 to 742).

Table 4.1: BcsC Constructs

Construct	Residues
Full	
	24-1141
	24-1154
	29-1141
β-Barrel	
	783-1141
	783-1154
	800-1141
	800-1154
	816-1154
TPR	
	24-781
	24-742
	24-813
	29-813

4.3 Objective 2: Profile Optimal Expression and Purification Conditions of the Various BcsC Constructs

4.3.1 Transformation and Expression

Chemically competent cells were created using *E. coli* BL21 cells (Novagen). A 3mL starter culture was grown in LB at 37°C for 18 h (200rpm). Using 0.5 mL of starter culture a 20 mL TYM broth (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.1M NaCl, 10 mM $MgSO_4$) flask was inoculated and incubated at 37°C with shaking at 200rpm until an OD_{600} of 0.6 was reached. The 20 mL culture was then transferred to a 100 mL TYM flask and grown to an OD_{600} of 0.6, incubating at 37°C with shaking at 200 rpm. Next, 400 mL of TYM broth was added to the culture flask and incubated at 37°C with shaking at 200 rpm until an OD_{600} of 0.6 was reached. The culture flask was then transferred to 4°C for 5-10 min to cool. Using pre-chilled centrifuge bottles, the cells were harvested by centrifugation at 5000 x *g* for 15 min. The supernatant was decanted and harvested cells were suspended in 100 mL of ice-cold TfB broth (10 mM MOPS (pH 7), 75 mM $CaCl_2$, 10 mM KCl, 15% (v/v) glycerol). Cells were harvested again by centrifugation at 5000 x *g* for 15 min. The supernatant was decanted and the harvested pellet was suspended in 20 mL of TfB broth. Suspended cells were then collected into 50 µL aliquots and stored at -80°C until required for transformations.

Frozen 50 µL stocks of chemically competent *E. coli* BL21 cells and frozen pET28a-*BcsC* ligated plasmid stock solutions were thawed at 4°C. With 5 µL of

Recombinant plasmid, competent cells were incubated at 4°C for an additional 5 min. Next, the competent cells were transformed with one of the various *bcsC* recombinant plasmids. Stock plasmid (typically 5-10 µL) was added to 50 µL of competent *E. coli* cells and incubated for 5 min. Next, transformation was induced via a standard method employing heat shock (*i.e.*, 2 min at 42°C). A period of 30 min on ice was provided to allow the cell membranes recovery time. Starter cultures were inoculated with newly transformed cells and grown in 500 µL LB for 1 h at 37°C with shaking at 240 rpm. LB broth (5 mL) with 50 µg/mL kanamycin were inoculated using the transformed starter culture and grown at 37°C with shaking at 200 rpm for 18 h. Overnight cultures (7.5 mL) were used to inoculate 1 L of SB and incubated at 37°C with shaking at 240 rpm until an optical density (OD_{600}) of 0.5-0.6 was reached. Variables tested to optimize expression included different concentrations of IPTG, varying temperature, and various times allowed for protein expression. For example, protein expression from the plasmid in the cells was testing using final concentrations of 1 mM and 2 mM IPTG. Once induction had begun, incubation temperatures of 15°C and 37°C were explored for protein productin. The expression times assayed were 4 h and 18 h after which cells were collected by centrifugation (5000 x *g* for 15 min). To monitor the progress of the induction, samples of 200 µL of culture were taken and cells from the samples were pelleted by centrifugation (5 min. at 5000 x *g)* and then suspended in SDS sample buffer (0.2 M Tris-HCl (pH 6.8), 50% (v/v) glycerol, 100 mg/mL SDS, 0.02% (w/v) bromophenol blue, 10 mM DTT). Expression levels were determined by SDS-PAGE

(outlined below). Loading of SDS-PAGE was performed with decreasing amounts of sample, starting with 20 µL of expression at time 0 and decreasing the amount loaded by 2.5 µL for each subsequent sample of hours expressed. Pellets from 1 L cultures were stored at -20°C until required for protein purification.

4.3.2 SDS-PAGE Analysis and Western Blot Analysis

This study involved the separation of proteins via SDS-PAGE with the Mini-PROTEAN Tetra cell apparatus (BioRad). Using a 12% (w/v) acrylamide solution (4 mL 30% (w/v) acrylamide with 8% (w/v) Bis, 3.35 mL MilliQ water, 2.5 mL 1.5 M Tris-HCl (pH 8.8, 100 µL 10% (w/v) SDS, 50 µL 10% (w/v) APS, 5 µL TEMED), separating gels were prepared using the BioRad gel casting apparatus with a 6% (w/v) acrylamide stacking solution (1.3 mL 30% (w/v) acrylamide with 8%(w/v) Bis, 6.1 mL MilliQ water, 2.5 mL 1.5 M Tris-HCl (pH 6.8), 100 µL 10% (w/v) SDS, 50 µL 10% (w/v) APS, 10 µL TEMED). All gels were run at 200 V for at least 45 min to ensure separation of various proteins in each sample without running samples off the gel. Gels were then stained using Coomassie (0.1% (w/v) Coomassie R250, 10% (v/v) acetic acid, 40% (v/v) methanol) or transferred to nitrocellulose paper (BioRad) for Western Blot analysis (detailed below). Stained gels were incubated for a minimum of 1 h followed by incubation in destaining solution (10% (v/v) acetic acid, 20% (v/v) methanol) until appropriate contrast was achieved for visualization. Each gel was run with 7 µL of Precision Plus dual colour protein standard as a molecular weight reference (BioRad).

Western blot transfers were performed using the Trans-Blot equipment from BioRad. Transfers were performed at 4°C in transfer buffer (12 mM Tris (pH 7.5), 96 mM glycine, 20% (v/v) methanol) for 2 h at 100V. Blots were blocked with 5% (w/v) skim milk powder in TBS buffer (10 mM Tris (pH 7.5), 150 mM NaCl) for a minimum of 1 h. Following the blocking, blots were washed twice in TTBS (20 mM Tris (pH 7.5), 2 mM NaCl, 0.05% (v/v) Tween) for 5 min each. Primary antibody, mouse anti-His, was diluted 1000 fold in 15 mL blocking buffer and applied to the blots to incubate for 45 min. Three subsequent washes were performed in TTBS, 5 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 5 min and once in TBS for 5 min before development. A minimal amount of 5-bromo-5-chloro-3-indolylphosphate (BCIP) substrate was then added to cover the nitrocellulose blot. The cleavage of BCIP by alkaline phosphatase on the conjugated secondary antibody forms nitroblue tetrazolium (NTB), revealing a purple precipitate that localizes on the blot near the detected His-tagged protein.

4.3.3 Purification

4.3.3.1 Denaturing Conditions:

Purification of insoluble constructs were tested using two harsh detergents to linearize insoluble proteins: 8 M urea and 6 M guanidine hydrochloride (GuHCl). For a typical sample preparation, three 1 L pellets were each suspended in 20 mL of denaturing buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 8 M urea or 6 M GuHCl) and incubated

at 4°C on a rocking platform for 18 h to permit lysis of the cells and solubilisation of the proteins therein. Cell lysates were then subjected to centrifugation (28000 x *g* for 45 min) to remove cellular debris and the clarified supernatant (containing denatured, but solubilized protein) was collected. Qiagen™ Ni-NTA agarose resin (typically 1mL per 1L of induced culture) was added to the clarified supernatant and left to incubate at 4°C on a multi-purpose rotator (model 151, Scientific Industries Inc.) for 5 h to permit interaction of the resin with His-tagged proteins. The mixture was then collected in a gravity-flow 10 mL column (BioRad) and washed two to three times with 50 mL of denaturing buffer to remove unbound proteins. Following washing of the column, His-tagged protein was eluted from the resin by adding 10 mL of denaturing elution buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 8 M urea or 6 M GuHCl, 1 M imidazole) and allowing the mixture to incubate for approximately 18 h. The elutate was then collected by gravity elution from the column. A second elution was sometimes performed using an additional 5 mL denaturing elution buffer with a 30 min incubation to ensure all His-tagged protein had disassociated from the resin. A 20 µL sample of the initial flow through the column, each wash and each elution was mixed with 20 µL of SDS sample buffer for SDS-PAGE and Western blot analysis to assess the presence of the recombinant protein and to provide an indication of the purity of the samples. Samples containing the desired recombinant protein were pooled and concentrated. After multiple trials, the concentration method that yielded the least amount of precipitated (unwanted) protein involved mixing 3 mL of eluted sample with 9 mL of denaturing buffer in a

Vivaspin 20 concentrator (Fisher Scientific) with a molecular weight cut off of 10 kDa. Concentrated samples were then kept at room temperature until use in refolding studies outlined below.

4.3.3.2 Native Purification Conditions:

A 1 L pellet was suspended in 20 mL of lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl) containing final concentrations of 0.0125 mg/mL DNase and 0.025 mg/mL RNase (added during suspension of the pellet). Initial purifications were performed with the presence of 10% (v/v) glycerol in the lysis buffer. Additional purifications were performed in the absence of glycerol to determine the necessity of having glycerol incorporated in the lysis buffer. Cell lysis was achieved with two passages through a Constant Systems Pressure machine (Constant Systems TS Series 0.75kW machine Pressure Biosciences) operating at 17 KPSI and 4°C. Lysate that remained inside the system was washed out with 50 mL of lysis buffer. The cell lysate was subjected to centrifugation (28000 x *g* for 45 min at 4°C) to separate soluble protein from remaining cellular contaminants (*i.e.,* whole cells and inclusion bodies). The supernatant was decanted and incubated for 1 h with 2 mL Qiagen™ Ni-NTA agarose resin at 4°C on a multi-purpose rotator. The mixture was then applied to a gravity-flow 10 mL column (BioRad) and washed with 20 mL lysis buffer. The column was then washed with at least 20 mL of wash buffer I (50 mM Tris pH 8.0, 300 mM NaCl, 20 mM imidazole) and sometimes with a third wash with 20 mL of wash buffer II (50 mM Tris pH 8.0, 300 mM NaCl, 50 mM imidazole). His-tagged protein was then eluted from the column by

applying 1 mL volumes of elution buffer (50 mM Tris pH 8.0, 300 mM NaCl, 250 mM imidazole) at a time for a total of 15 elutions. A 20 µL sample of the initial flow through the column, each wash and each elution was mixed with 20 µL of SDS sample buffer for SDS-PAGE and Western blot analysis to assess the presence of the recombinant protein and to provide an indication of the purity of the samples. All elutions were stored at -20°C until further use. Fractions containing BcsC were pooled and concentrated using a Vivaspin 20 concentrator with a molecular weight cut off of 30 kDa. Further purification was tested using a gel filtration ENrich SEC 650 column (BioRad) equilibrated in running buffer (20 mM Tris pH 8.0, 10 mM NaCl). For rapid removal of salt and imidazole from purified samples a PD-10 Desalting Column (GE Healthcare) was employed that was also equilibrated with the gel filtration running buffer. Following removal of salt/imidazole, samples were concentrated (as above) to a final concentration typically between 20 and 60 mg/mL and either used directly for assays or stored at -20°C until needed.

4.3.3.3 Refolding:

A gradual refolding process with decreasing amounts of urea in a favourable detergent were attempted. Briefly, concentrated denatured protein was incubated with refolding buffer I (6 M urea, 0.06% (w/v) LDAO, 20 mM Tris pH 8.0, 500 mM NaCl) for at least 18 h with a 1:3 ratio of protein sample to buffer added. The sample was then concentrated in a Vivaspin 20 concentrator (10 kDa molecular weight cut off). The 6M urea/0.06% (w/v) LDAO samples were diluted three fold in refolding buffer II (3 M urea,

0.06% (w/v) LDAO, 20 mM Tris-HCl pH 8.0, 500 mM NaCl) and incubated at 4°C for 8 h and then concentrated, as above. Samples continued to be diluted in refolding buffers III, IV, and V (0.06% (w/v) LDAO, 20 mM Tris-HCl pH 8.0, 500 mM NaCl containing 1.5 M urea, 0.75 M urea and 0.33 M urea, respectively), as described above. Finally, refolding buffer VI (0M urea, 0.06% (w/v) LDAO, 20 mM Tris-HCl pH 8.0, 500 mM NaCl) was added to the final concentrated sample to achieve a 10-fold dilution in this buffer. The sample was then incubated for 2 h at 4°C and concentrated to approximately 0.5 mg/mL. Samples (20 µL) were collected from each of the steps prior to concentration and mixed with 20 µL SDS sample buffer in order to assess the progress of the refolding method by SDS-PAGE (discussed above). Two additional samples were collected after concentrating in refolding buffer VI for analysis of the final concentrated sample as well as a heated sample, to determine the difference in migration of the folded protein and linearized, heat denatured, protein.

4.3.3.4 Protein Concentration Determination

Protein concentration was determined using a BioDrop DUO (purchased through Montréal Biotech Inc.). Sample information, including the molecular weight and extinction coefficient were programmed under the molar extinction function using the microlite setting. Samples were read for an absorbance at 280, whereby the internal computer calculated the concentration of sample in mg/mL.

4.4 Objective 3: Structural Analysis of BcsC

4.4.1 Circular Dichroism

Following the purification methods described above, protein samples were diluted to 1 mL at a concentration of 2 μM for circular dichroism (CD) analysis. BcsC[24-742] samples were analyzed in running buffer and also in the presence of carboxy-methyl cellulose (CMC) at 0.05, 0.1 and 0.5% (w/v). BcsC[24-781] samples were analyzed in the presence of 0.5% (w/v) CMC. Lastly, BcsC[24-813] samples were also analyzed without and with 0.1 and 0.5% (w/v) CMC. Each sample was analyzed using Far-UV CD from 260-198 nm at a 1 nm resolution on an Aviv spectropolarimeter (Aviv Biomedical, NJ) in running buffer. Each sample was measured using a 0.1-cm pathlength quartz cuvette and scanned 3 times. Samples were prepared in duplicate from each of two separate purifications (*i.e.*, using protein purified 10 and 17 days prior to CD analysis). An average of the three scans on each sample were calculated. Buffer and CMC controls were subtracted from the appropriate averaged data and the resultant data was smoothed and converted to molar elipticity for analysis. The data were then subjected to deconvolution programs, (CDSSTR (Compton and Johnson, 1986; Manavalan and Johnson, 1987; Sreerama and Woody, 2000), SELCON (Sreerama and Woody, 1993; Sreerama *et al.*, 1999) and K2D (Andrade *et al.*, 1993)) on the website DichroWeb for secondary structure content estimation (Whitmore and Wallace, 2008; Whitmore and Wallace, 2004; Lobley *et al.*, 2002).

Additional CD analysis was performed on a 2 µM sample of BcsC[24-742] in running buffer to obtain a melting curve. A spectral profile was obtained using a protein elipticity at 222 nm and increasing the temperature from 15°- 85°C, scanning at a 1°C resolution. The spectra obtained was normalized and fitted to the Hill equation:

(1)
$$y=V_{max}(x^n/(k^n+x^n))$$

Where y is the molar elipticity at 222 nm, V_{max} is the elipticity at maximum molar elipticity, x is the Hill coefficient and k is the temparature constant, whereby $k_{0.5}$ is the temperature at $\frac{1}{2}V_{max}$.

4.4.2 Crystallization Trials

Concentrated TPR constructs were screened using polyethylene glycol (PEG) and MCSG (1-4) crystal screens (Microlytic). Reservoirs of crystallization INTELLI-PLATEs (Art Robbins Instruments) were initially filled with 100 µL of the 96 conditions for each screen. Concentrated protein and crystallization condition were then combined at a 1:1 ratio in the top drop and a 1:2 ratio in the bottom drop for a total volume of 2 µL and 3 µL, respectively. Trays were incubated at 20°C for the duration of the experiment and observed regularly for the formation of protein crystals. Multiple trays were set up using protein concentration ranging from 5 mg/mL to 60 mg/mL (over a variety of TPR constructs). Potential protein crystals were assessed using Izit protein dye (Hampton) according to the manufacturers protocol and observed for fluorescence with UV light (Xtallight-100, Molecular Dimensions Inc.). These assessment methods were used as the

majority of protein crystals absorb the dye, turning dark blue, and/or fluoresce under UV light caused by excitation of tryptophan residues present in the protein.

Protein crystals were used to create seed stocks using seed beads (Hampton Research) via the contemporary seed stock method. Crystals were crushed using a probe. Next, 10 µL of the mother liquor from the reservoir was added to the crushed crystals and aspirated and re-dispensed to mix the solutions. The liquid was then transferred from the well to a pre-chilled tube containing one seed bead. This procedure was repeated until 50 µL of the mother liquor had been used (with mixing in the well before transferring to the seed bead tube). The seed bead tube with crushed crystals was then vortexed for 3 min. To grow new crystals using the newly formed seeds, a new tray was set up with 100 µL of mother liquor in the reservoirs. Freshly purified protein and seed stock were added to the drop well using 1.5 µL and 0.5 µL, respectively. Mother liquor from the reservoir was then added to the drop for trials using 0.5 µL, 1 µL, 1.5 µL, 2 µL and 2.5 µL in an attempt to assess different conditions.

4.5 Objective 4: Functional Analysis of BcsC

4.5.1 Fluorescence Analysis

Following purification, buffer exchange and concentrating, samples were stored at -20°C until needed (typically 3 d). CMC-dependent quenching was investigated using concentrations of CMC, ranging from 0 mg/mL to 18.5 mg/mL, in the presence of the TPR constructs BcsC[24-742], BcsC[24-781] and BcsC[24-813] in running buffer at final molar

concentrations of 3 µM, 3 µM and 2.5 µM, respectively. Fluorescence intensities were measured from 300 nm to 450 nm using an excitation wavelength of 295 nm on a Cary Eclipse Fluorescence spectrophotometer. The fluorescence intensity of CMC in running buffer at the tested concentrations was used to correct sample data for each respective sample. Association, disassociation and binding constants were determined using Stern-Volmer plots created using emission data collected at 335 nm for BcsC[24-742] and BcsC[24-781] and 325 nm for BcsC[24-813]. All trials were performed in triplicate to ensure consistent quenching.

Fluorescence analysis was also performed on bovine serium albumin (BSA) as a control, following the fluorescence conditions described by Moriyama and colleagues (1996). Using 10 µM BSA in running buffer, spectra were collected from 250 nm to 550 nm using a 300 nm excitation with CMC concentrations ranging from 0 mg/mL to 18.5 mg/mL. Data analysis followed the same procedure as for BcsC using data points collected for 350 nm.

Following the collection of quenching profiles for each construct Stern-Volmer plots were created based on the equation below that have been used for other quenching studies (Acharya *et al*., 2013; Beattie and Merrill, 1999):

(2) $$F_0/F = 1 + K_S [Q]$$

where F and F_0 are the fluorescence intensities of the protein in the presence and absence of CMC, respectively, and [Q] represents the concentration of CMC, and K_S is the Stern-

Volmer constant. Since the protein was incubated with CMC for 30 minutes it can be assumed that the association and disassociation of protein with CMC reached equilibrium. It can further be decided that the K_S can be considered as an association constant, K_a, since the reaction is at equilibrium. A linear relationship on a Stern-Volmer plot is indicative of a single binding site (Acharya *et al*., 2013) where CMC associates with BcsC. Furthermore, the relationship introduced in equation (2) results in a K_S (or K_a) equivalent to the slope of the trend. Additionally, a disassociation constant K_d can be determined by the inverse relationship with K_a using the following equation

(3) $$K_d = 1/K_a$$

4.5.2 Avicel Binding Assay

BcsC[24-742], BcsC[24-781,] BcsC[24-813] and BSA (as a control) were diluted to 0.5 mg/mL in Buffer A (20 mM Tris (pH 8.0), 20 mM NaCl), Buffer B (20 mM Tris (pH 7), 20 mM NaCl) and Buffer C (20 mM Na_2HPO_4 (pH 6), 20 mM NaCl). A 1 mL sample of each solution was then mixed and incubated with 0 mg, 5 mg, 10 mg, 15 mg, 20 mg, 25 mg and 50 mg of Avicel for 30 min (with the exception that there was not enough protein for BcsC[24-813] to run a 50 mg Avicel sample). Avicel-BcsC complex was pelleted by centrifugation (12000 x *g* for 5 min) the supernatant was removed and then the pellet was suspended in 50 µL SDS sample buffer. Avicel-BcsC complex in SDS sample buffer was incubated 10 min at 100°C and then subjected to centrifugation (12000 x *g* for 5 min)

prior to analyzing 20 µL of the supernatant by SDS-PAGE. Band intensities were determined using the TotalLab Quant gel scanning software.

5. BIOINFORMATICS ANALYSIS OF BcsC AND CONSTRUCT GENERATION:

5.1 Background

A substantial amount of information on a protein can be obtained using a variety of different bioinformatics tools. Publically available programs have been generated to determine nucleic acid and amino acid sequence homology between previously classified proteins and new proteins being studied. Through these bioinformatics programs we can gain valuable preliminary information on BcsC by analyzing its primary sequence.

BcsC has shown sequence homology to proteins in other biofilm-producing complexes that are responsible for polysaccharide export. Using a variety of bioinformatics tools, BcsC can be analyzed for similar structural domains and the domain boundaries. Preliminary information on the protein, including the molecular weight, extinction coefficient and the ratio of each amino acid, was obtained using ProtParam (Gasteiger *et al*., 2005). Inputting the amino acid sequence into PSORT provides a predicted cellular location for the protein (Gardy *et al*., 2005). LipoP and SignalP are two programs that compare the amino acid sequence to databases to determine the presence of lipidation and signal sites on the protein, respectively (Juncker *et al*., 2003; Nordahl Peterson *et al*., 2011). Using TPRpred predicted TPR regions can also be located using the polypeptide sequence (Karpenahalli *et al*., 2007). Next, the secondary structure of the protein construct can be predicted using PSIPRED (Buchan *et al*., 2013; Jones, 1999). The three dimensional (3D) model-prediction software Phyre[2] can be helpful in determining residue orientation of amino acids within their predicted domains (Kelley *et*

al., 2015). Finally, ClustalW is a sequence alignment program that pairs homologous amino acids between multiple sequences, such as the BcsC gene in *E. coli* and *Salmonella* spp. to find critical residues (Larkin *et al*., 2007).

Having obtained information using the various bioinformatics tools described above, various protein constructs can be developed based on the predicted domains of the protein as well as the termini of α-helices and β-sheets. Using the pET28a expression vector recombinant protein constructs can be tagged at the N- and/or C-termini of the protein. With the constructs ligated into the expression vector they are ready for transformation and expression experiments in BL21 cells.

5.2 Results

5.2.1 Bioinformatics

The culmination of using the above programs with the amino acid sequence of the *bcsC* gene from BL21 *E. coli* DNA (Figure 5.1) aided our prediction of a number of domain boundaries for the TPR and β-barrel regions of this protein (outlined in Figure 5.2). The constructs developed are described in Table 5.1 with the number of amino acids, molecular weight, extinction coefficient, PI and His_6 tags for each construct determined using ProtPram (Gasteiger *et al*., 2005). Furthermore, using the whole amino acid sequence, the first 23 amino acids were found to be a signal sequence using SignalP and no constructs were predicted to contain a lipidation site. When analyzed using PSORT the full length protein was predicted to be a periplasmic protein with a certainty

of 0.410. A secondary prediction indicated an outer membrane with a certainty of 0.174. PSORT predicted a cytoplasmic protein for both the TPR and β-barrel constructs with certainties of 0.470 and 0.327 when using BcsC$^{24\text{-}742}$ and BcsC$^{783\text{-}1141}$, respectively. These predictions support the hypothesis that BcsC is made up of periplasmic and outer membrane domains (the TPR and β-barrel domains, respectively). Furthermore, the TPR domain of BcsC is much larger than the β-barrel domain, which further supports the main prediction to indicate a periplasmic protein. When each domain location was predicted using PSORT both the TPR and β-barrel constructs were predicted to be cytoplasmic. This is likely due to the absence of the signal peptide removed from the N-terminus for the TPR construct. In contrast, the β-barrel has been predicted to be located in the cytoplasm as inclusion bodies (based on similar expression experiments with homologous proteins). TMHMM indicated that there are no helical transmembrane regions in the protein.

MRKFTLNIFTLSLGLTVMPMVEAAPTAQ**QQLLEQVRLGEATHREDLVQQSLYRLELIDP**
NNPDVVAARFRSLLRQGDIDGAQKQLDRLSQLAPSSNAYKSSRTTMLLSTPDGRQALQQ
ARLQATTGHAEEAVASYNKLFNGAPPEGDIAVEYWSTVAKIPARRGEAINQLKRINADT
PGNTGLQNNLALLLFSSDRRDEGFAVLEQMAKSNAGREGASKIWYGQIKDMPVSDASVS
ALKKYLSIFSDGDSVAAAQSQLAEQQKQLADPAFRARAQGLAAVDSGMAGKAIPELQQA
VRANPKDSEALGALGQAYSQKGDRANAVANLEKALALDPHSSNNDKWNSLLKVNRYWLA
IQQGDAALKANNPDRAERLFQQARNVDNTDSYAVLGLGDVAMARKDYPAAERYYQQTLR
MDSGNTNAVRGLANIYRQQSPEKAEAFIASLSASQRRSIDDIERSLQNDRLAQQAEALE
NQGKWAQAAALQRQRLALDPGSVWITYRLSQDLWQAGQRSQADTLMRNLAQQKPNDPEQ
VYAYGLYLSGHDQDRAALAHINSLPRAQWNSNIQELVNRLQSDQVLETANRLRESGKEA
EAEAMLRQQPPSTRIDLTLADWAQQRRDYTAARAAYQNVLTREPANADAILGLTEVDIA
AGDKAAARSQLAKLPATDNASLNTQRRVALAQAQLGDTAAAQRTFNKLIPQAKSQPPSM
ESAMVLRDGAKFEAQAGDPTQALETYKDAMVASGVTTTRPQDNDTFTRLTRNDEKDDWL
KRGVRSDAADLYRQQDLNVTLEHDYWGSSGTGGYSDLKAHTTMLQVDAPYSDGRMFFRS
DFVNMNVGSFSTNADGKWDDNWGTCTLQDCSGNRSQSDSGASVAVGWRNDVWSWDIGTT
PMGFNVVDVVGGISYSDDIGPLGYTVNAHRRPISSSLLAFGGQKDSPSNTGKKWGGVRA
DGVGLSLSYDKGEANGVWASLSGDQLTGKNVEDNWRVRWMTGYYYKVINQNNRRVTIGL
NNMIWHYDKDLSGYSLGQGGYYSPQEYLSFAIPVMWRERTENWSWELGASGSWSHSRTK
TMPRYPLMNLIPTDWQEEAARQSNDGGSSQGFGYTARALLERRVTSNWFVGTAIDIQQA
KDYAPSHFLLYVRYSAAGWQGDMDLPPQPLIPYADW

Figure 5.1: The amino acid sequence of BcsC in BL21 *E. coli.* The protein is predicted to contain a signal sequence, a TPR domain and a β-barrel domain. Brown letters represents the signal peptide, blue bolded letters represents the TPR domain, and purple letters represents the β-barrel domain.

Table 5.1: General Characteristics of BcsC Constructs

Construct	Residues (amino acid numbering)	# Amino Acids	Molecular Weight (kDa)	Extinction Coefficient $(M^{-1}cm^{-1})$	Predicted pI	His- Tag Termini
Full						
	24-1141	1152	127.2	213600	6.50	N and C-Terminal
	24-1154	1165	128.6	215090	6.35	N- Terminal
	29-1141	1146	126.6	213600	6.50	N and C-Terminal
β-Barrel						
	783-1141	392	43.7	127310	6.43	N and C-Terminal
	783-1154	405	45.2	134300	6.24	N- Terminal
	800-1141	375	41.9	118955	7.02	N and C-Terminal
	800-1154	388	43.4	125820	6.65	N- Terminal
	816-1154	352	39.4	120320	6.22	N- Terminal
TPR						
	24-781	792	87.1	86290	7.35	N and C-Terminal
	24-742	753	82.4	79300	7.35	N and C-Terminal
	24-813	824	90.7	94770	6.72	N- Terminal
	29-813	819	90.1	93280	6.42	N- Terminal

Secondary structural tools including TPRpred, PSIPRED and Phyre[2] were used next to analyse protein folding. Using TPRpred the full and TPR constructs were consistently found to contain 18 TPR segments each. The consistency between the full and TPR constructs indicates that the whole TPR domain has been conserved in each of these constructs. PSIPRED and Phyre[2] both provide a representative of which amino acids are found in helices, sheets and coils. Figures 5.2 and A1.2 (located in Appendix I) show consistency between the two prediction tools whereby each displays high helical

content at the N-terminus of the protein and β-sheets at the C-terminus. DisProt was used to define disordered regions of the protein sequence when creating the various constructs. Within the DisProt site, DisEMBL[TM] suggested that there is disorder located at the N-terminus from residues 13-24 as well as at the C-terminus from residues 1037-1157 (Linding *et al*., 2003). Using this information, the various constructs were designed to reduce disorder. For example, constructs beginning with residue 24 may have more highly dynamic peptides than those beginning with residue 29 and consequently, constructs ending with residue 1154 may have more disorder than those ending with residue 1141. TPR and β-barrel constructs use a variety of boundaries for the end and beginning of the constructs, respectively. This is due to three more disordered regions of the protein located from residues 741-779, 787-805 and 811-890 (Linding *et al*., 2003). Varying the length of the various constructs allows investigating the propensity of these disordered regions. Additionally, TPR constructs were created with by adding or removing TPR motifs as this may lead to stabilizing the protein. Next, Phyre[2] was used in providing a 3D model of each protein construct. Figure 5.3 shows an example of Phyre[2] predicted models for a full, TPR and β-barrel construct. Each of the figures displayed in Figure 5.3 were created using PyMOL. The wizard measure function allowed to see the distances the two domains will span as well as their width. This information was obtained to determine the physiological relevance of each of the models.

Figure 5.2: Phyre² secondary structural prediction of BcsC²⁴⁻¹¹⁴¹. Green coils represent α-helices, blue arrows represent β-strands, straight lines represent coils and the colour scale represents the disorder confidence whereby the red end of the spectrum represents a high confidence level and the blue end of the spectrum represents a low confidence level (Kelley *et al.*, 2015).

The full construct BcsC²⁴⁻¹¹⁴¹ was modeled after the crystal structure of the yeast

Ski complex, ski2-3-8, seen in Figure 5.3 Panel A (PBD: 4buj; Halbach *et al.*, 2013;

Kelley *et al.*, 2015). The Ski complex is responsible for a variety of functions including

RNA turnover, surveillance and interference for cytoplasmic exosomes (Halbach *et al*., 2013). Ski2 is an RNA helicase, Ski3 consists of 33 TPR motifs in two arms, an N-terminal and C-terminal arm, and Ski8 is composed of WD40 repeats. The repeating units of Ski3 and Ski8 are responsible for mediating protein-protein interaction (Halbach *et al*., 2013). The full-length constructs were likely primarily based off of the Ski3 protein with a high TPR content. The protein-protein interactions that Ski3 are involved in provide insight into the potential for interactions of the TPR domain of BcsC to be involved with other periplasmic proteins. One such potential interaction includes BcsC with BcsB. BcsB is a member of the cellulose synthesis complex with BcsA. Unlike BcsA, BcsB is located mainly in the periplasmic space (Morgan *et al*., 2013). BcsB also contains two carbohydrate-binding domains that facilitate the movement of the growing cellulose chain through the cellulose synthesis complex into the periplasmic space (Morgan *et al*., 2013). It is possible that the TPR of BcsC extends through the periplasm to associate with BcsB and the emerging cellulose chain.

The TPR construct BcsC$^{24-742}$ was modeled after the crystal structure of the human O-linked GlcNAc transferase (OGT), seen in Figure 5.3 Panel B (PDB: 1w3b; Jinek *et al*., 2004; Kelley *et al*., 2015). OGT is responsible for attaching *O*-GlcNAc to serine or threonine residues of intracellular proteins (Jinek *et al*., 2004). Figure 5.3 Panel B indicates the length of the TPR model created containing 11 TPR pairs. This shortened form of the TPR is approximately 105.4 Å long with a 24.4 Å wide center. The predicted number of TPRs within BcsC's N-terminal domain is 18, therefore, the full TPR domain

of BcsC would be approximately 160-170 Å long. The length of the periplasm from the outer membrane to the inner membrane in an *E. coli* cell can range from 100 Å to 165 Å (Hobot *et al*., 1984). Therefore, since this particular model is a shortened representative of the TPR it is likely that the wildtype protein will extend farther than the predicted model, thus covering more than enough distance to bridge the periplasmic gap between the inner and outer membranes, which the cellulose chain must pass through.

The β-barrel construct BcsC$^{783-1141}$ was modeled after the structure of the *P. aeruginosa* esterase estA, seen in Figure 5.3 Panel C (PDB: 3kvn; van den Berg, 2010; Kelley *et al*., 2015). Esterase estA is a β-barrel protein located in the outer membrane of the cell with the purpose of secreting virulence factors (van den Berg, 2010; Kelley *et al*., 2015). The function of the β-barrel domain of BcsC has been predicted to be responsible for cellulose export. Having been modeled based on homology with after another export protein supports this potential role of the β-barrel.

The published results using the 3D model of BcsA indicate that the channel through which the cellulose chain is formed and extruded across the inner membrane is 8 Å wide and 33 Å long (Morgan *et al*., 2013). The β-barrel portion predicted for BcsC$^{781-1141}$ is approximately 30.1-32.4 Å wide and 47.1 Å long. This size presumably provides a large enough pore for cellulose to be exported from the cell, and is also wide enough to lead to unwanted passage of molecules across the outer membrane. Research on the β-barrel protein AlgE may be able to provide some insight on this matter. AlgE of

the alginate biosynthesis system contains two ordered loops, one at each end of the pore, that form a seal (Tan *et al*., 2014). These loops have been called the periplasmic gate (P-gate) and the extracellular gate (E-gate). When alginate is being synthesized, it is predicted that the P-gate opens allowing for the chain to enter the pore, followed by the E-gate. When the end of the chain has moved past the P-gate it closes, while the E-gate remains open to pass the remainder of the alginate chain through. Once the chain has emerged, the E-gate closes and both gates are blocking the pore (Tan *et al*., 2014). Since the β-barrel was not modeled after a polysaccharide export protein, such as AlgE, it does not indicate the presence of these pores. However, due to homology between the β-barrel of BcsC and AlgE, and the hypothesized function is suggested based on this similarity, it is very possible that the cellulose system contains a similar mechanism to control leakage from the cell. Evidence for the E-gate to exist in BcsC is also supported in the 3D model generated for this protein as there are a number of extracellular loops that can be seen in Figure 5.3 Panel C. The control over the periplasmic face of the pore, however, may be under the control of the TPR domain of BcsC. The alginate export complex is composed of two separate proteins, AlgE and AlgK, whereas, the cellulose export complex is made up of one polypeptide forming these two domains. Rather than containing a P-gate, it is possible that the TPR domain may play a role in blocking the pore until cellulose is presented for export.

A: Full-Construct Prediction
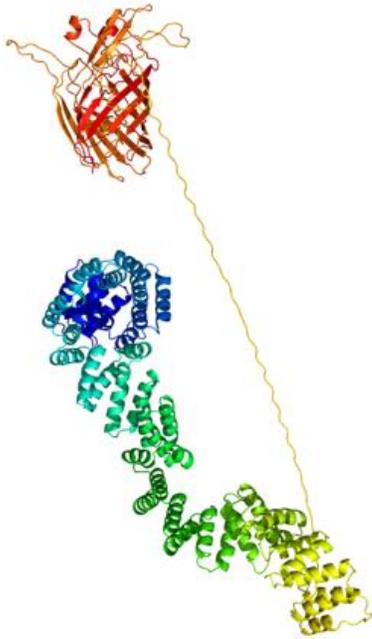
B: TPR Region Prediction
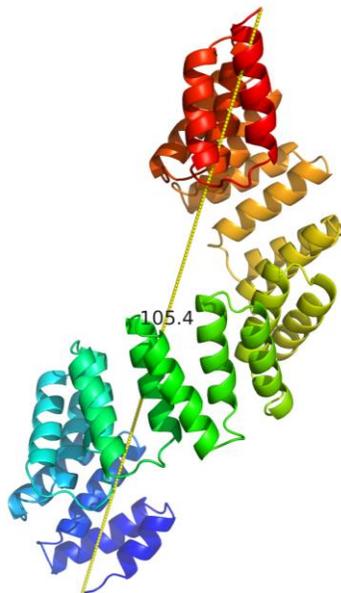
Side View   Top Down View
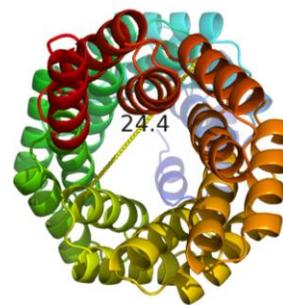
C: β-Barrel Region Prediction

Side View   Top Down View
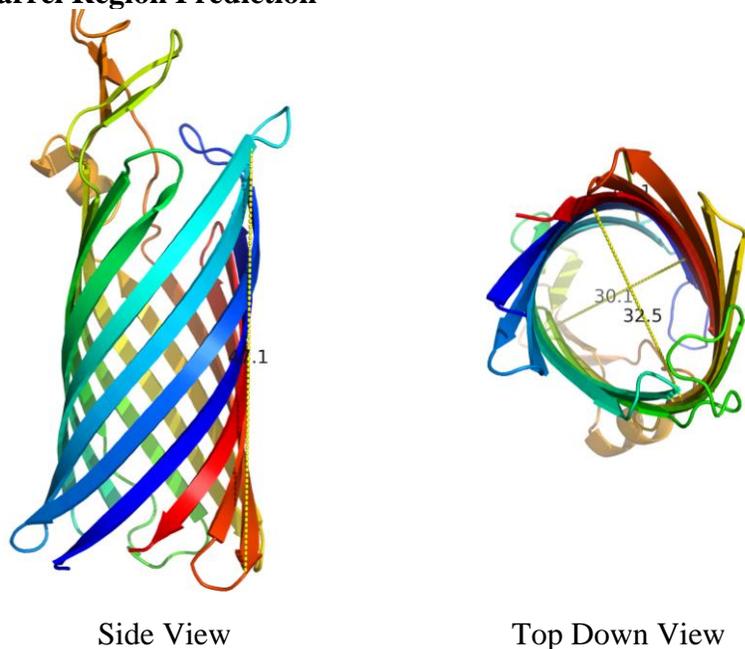
Figure 5.3: BcsC domain models as predicted by Phyre^2. BcsC$^{24-1141}$ (A) BcsC$^{24-742}$ using residues 32-671 (consisting of 11 TPR pairs) (B) and BcsC$^{781-1141}$ using residues 859-1134. Using this prediction software it is suggested that BcsC may fold into the 2 predicted domains: an N-terminal TPR spanning approximately 105.4 Å and 24.4 Å wide and C-terminal β-barrel spanning approximately 47.1 Å and 30.1-32.5 Å wide. Each model was produced by Phyre^2 on the crystal structure of the *Saccharomyces cerevisiae* ski2-3-8 complex; PDB: 4buj; Confidence level: 100; Identity: 22% (A) (Halbach *et al*., 2013; Kelley *et al*., 2015), human O-linked GlcNAc; PDB: 1w3b; Confidence level: 100; Identity: 18% (B) (Jinek *et al*., 2004; Kelley *et al*., 2015) and *P. aeruginosa* esterase estA; PDB: 3kvn; Confidence level: 96.8; Identity: 13% (C) (van den Berg, 2010; Kelley *et al*., 2015). All illustrated structures were created using PyMOL v1.5.0 and the distances were measured using the wizard measure function.

5.3 Conclusions

Various bioinformatics tools were used to collect background information based on the amino acid sequence of BcsC and homology searches with previously classified proteins. Homology to polysaccharide export proteins in other biofilm systems has led to

the prediction that BcsC is responsible for the export of newly synthesized cellulose chains for biofilm formation, and is composed of an outer membrane β-barrel and a periplasmic TPR. Using secondary and tertiary prediction software analysis on the amino acid sequence it has been verified that BcsC is likely composed of the two predicted domains; an N-terminal TPR and C-terminal β-barrel. The 3D models of BcsC have partly confirmed this hypothesis as they have been modeled on crystal structures that support these confirmations. For example, OGT, a TPR rich protein that is responsible for protein-protein interactions in order to glycosylate other proteins (Jinek *et al*., 2004) served as the model for the TPR portion of BcsC. Furthermore, the β-barrel domain has been modeled after the crystal structure of a homologous outer membrane protein (OMP), esterase estA; a protein that is involved in removing virulence factors from the cell through its β-barrel pore. Additionally, based on sequence and functional homology to the alginate export complex, it is possible that BcsC may contain a loop mechanism to control leakage from the cell.

6. PROFILE OPTIMAL EXPRESSION AND PURIFICATION CONDITIONS OF THE BcsC CONSTRUCTS

6.1 Background

Having determined domain boundaries using bioinformatics tools, a variety of gene constructs can be developed and cloned into an expression vector, such as pET28a which was used in this study. There are many benefits to using expression vectors. For example, ligation of the amplified gene segment into pET28a behind the T7 promoter and lac operator (seen in Appendix I) provides some useful features. Using the lac operator allows control over *bcsC* induction by using IPTG, a lactose analog, for induction. pET28a also provides the means for tagging the protein at either the N and/or C termini with histidine (His$_6$) tags. Histidine is an amino acid that has high affinity to nickel, therefore, His$_6$ tags can be effective in protein purification using nickel resin in an immobilized metal affinity chromatography (IMAC) column.

Prior to purification, bacterial cells must be lysed to release recombinantly expressed protein. Cell lysis can be achieved by a variety of methods. In denaturing conditions cells can be lysed in a native buffer where the protein of interest would form inclusion bodies and collected by centrifuge and linearized in a harsh detergent or harsh chemicals (8M urea or 6M guanidine HCl) are directly applied to lyse the cells. The use of high concentrations of urea or guanidine HCl will result in a flexible and disordered linear polypeptide that is soluble (Tsumoto *et al.*, 2003). In comparison, native techniques generally use a combination of enzymes (lysozyme), chemicals and some sort of

mechanical lysis to open the cells. We have access to a variety of mechanical lysis possibilities; sonication, the French pressure cell and a Constant Systems high pressure cell lysis machine. The Constant Systems instrument has been the preferred lysis method as it is quick, easy to use and minimizes the denaturation of protein that can be caused by other methods through heat generation and/or excessive shearing forces. The Constant System machine is a highly effective method that pumps the sample through a narrow canal at high pressure towards a solid surface. As the sample is forced through the canal the sample undergoes a sheering pressure which lyses the cells open. A cool metal plate at the end of the canal is used to cool the sample as a means to prevent denaturation of the protein.

Different cell lysis strategies must be employed based on the particular *BcsC* constructs used for protein expression. In the case of TPR-related constructs that express as soluble proteins, non-denaturing methods can be employed to acquire high concentrations of pure, folded protein. In contrast, any constructs that contain the membrane associated β-barrel (*i.e.*, β-barrel and full constructs) express this protein into inclusion body aggregates which must be linearized for purification using harsh detergents and refolded using more gentle protein detergents. Preliminary refolding experiments performed by project students in the lab have shown LDAO to be a promising detergent to provide the appropriate environment to refold the β-barrel constructs.

Both protein purification methods for native and denatured protein utilize the same isolation techniques. With His_6 tags recombinantly attached to each construct, the protein of interest can be selected for using nickel resin; taking advantage of the natural affinity between the imidazole ring of histidine and nickel (Qiagen, 2003). Passing a cleared lysate through a column containing nickel resin provides the means to capture the protein of interest in the column while all other protein contaminants flow through the column. Once the protein has been maximally purified using a variety of wash steps it can be eluted from the column using buffer containing a high concentration of imidazole which outcompetes with the His_6 tag, releasing the protein from the resin.

6.2 Results and Discussion

6.2.1 Expression of BcsC Constructs

Different expression conditions were tested using the β-barrel construct $BcsC^{816-1154}$ and the full construct $BcsC^{29-1154}$. It was determined that the expression of these two constructs did not change between the different concentrations of 1 mM and 2 mM IPTG. This was an indication that 1 mM IPTG was enough to induce expression for the cells present and as a means to save resources, all remaining expressions were performed using the standard 1 mM IPTG concentration. When protein expression was allowed to proceed for 18 h, there was not a substantial difference in the amount of protein produced from 4 h as seen in Figure 6.1 Panel A (lanes 6 and 7) for $BcsC^{816-1154}$ and Panel B (lanes 5 and 6) for $BcsC^{29-1154}$. Additionally, after 18 h of expression, the protein constructs showed some level of degradation in $BcsC^{29-1154}$, whereas the β-barrel domain showed consistent

degradation throughout the expression period, evidenced by a consistently thicker band (~30 kDa) below the expressing protein (39 kDa). The expression of each construct thereafter was given 4 h. Also, by reducing the sample load applied to the gel, the background proteins were reduced, allowing to optimally view the protein construct being expressed. Expression at two different temperatures were also tested, 15°C and 37°C, as other researchers in the lab had found 15°C to allow for more protein production and less protein degradation associated with other Bcs proteins studied in the lab. It was found that 37°C provided a more efficient temperature for protein expression as seen in the post purification temperatures in Figure 6.2. The choice of 37°C is also physiologically relevant as *E. coli* grow optimally at this temperature, therefore, the enzymes in this bacteria work efficiently at this temperature. This permits cellular transcriptional and translational equipment to work efficiently within the bacterial cell. All subsequent expressions were performed at 37°C for 4 h for all constructs used in this study.

A
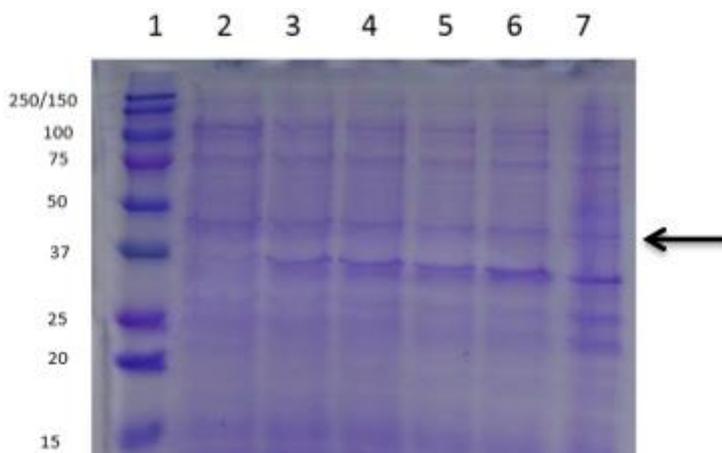
B
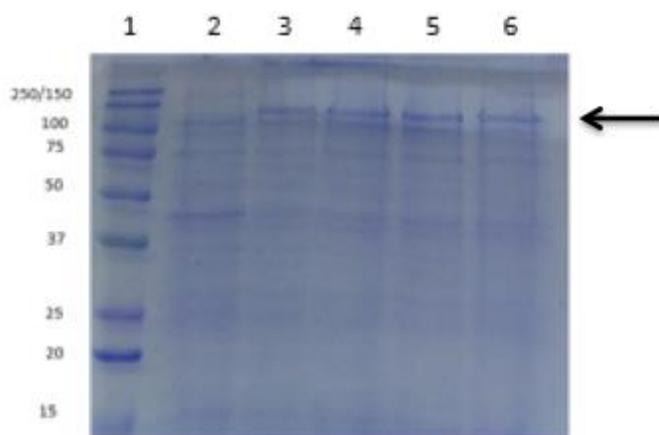
Figure 6.1: Protein expression of BcsC$^{816-1154}$ and BcsC$^{29-1154}$ from 0 to 18 h at 37°C using 1 mM IPTG. (A) Samples of BcsC$^{29-816}$ were loaded in the order of 1) protein standard, and expression at 2) 0 h, 3) 1 h, 4) 2 h, 5) 3 h, 6) 4 h, and 7) 18 h on a 12% SDS-PAGE, stained with Coomassie. Expression levels did not vary between 4 and 18 h as indicated by the arrow. Degradation is apparent throughout the entire expression suggested by the increasing intensity of a band at ~30 kDa. (B) Samples of BcsC$^{29-1154}$ were loaded in the order of 1) protein standard, and expression at 2) 0 h, 3) 1 h, 4) 2 h, 5) 3 h, and 6) 18 h on a 12% SDS-PAGE, stained with Coomassie. Expression levels remained visually the same from 3 to 18 h indicated by the arrow.

A

B
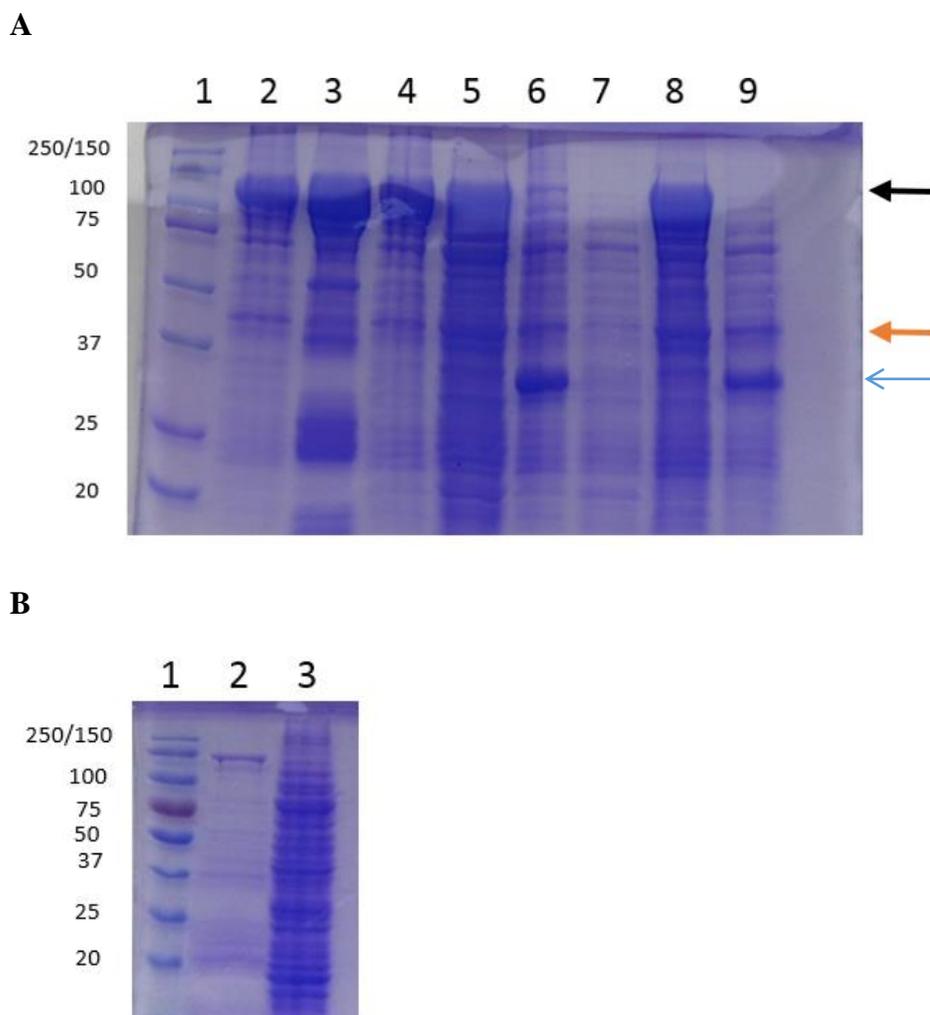
Figure 6.2: Determining the optimal expression and purification conditions for BcsC$^{29\text{-}1154}$ and BcsC$^{816\text{-}1154}$. (A) Expression temperatures tested at 15°C and 37°C. Purification conditions tested in urea and GuHCl. Samples loaded in the order of 1) protein standard, 2) BcsC$^{29\text{-}1154}$ in GuHCl expressed at 37°C, 3) concentrated BcsC$^{29\text{-}1154}$ in GuHCl, 4) BcsC$^{29\text{-}1154}$ in GuHCl expressed at 15°C, 5) BcsC$^{29\text{-}1154}$ in urea expressed at 15°C, 6) BcsC$^{816\text{-}1154}$ in GuHCl expressed at 37°C, 7) BcsC$^{816\text{-}1154}$ in urea expressed at 15°C, 8) BcsC$^{29\text{-}1154}$ in urea expressed at 37°C, and 9) BcsC$^{816\text{-}1154}$ in in urea expressed at 37°C on a 12% SDS-PAGE, stained with Coomassie. Full-length sample constructs are indicated by bands at the black arrow, β-barrel sample constructs are indicated by the orange arrow and a portion of the degraded β-barrel is shown by the blue arrow. B) Final condition tested was run on a second gel seen in lane 3 BcsC $^{816\text{-}1154}$ in GuHCl expressed at 15°C, to be compared against protein standard seen in lane 1. This sample was highly impure such that the construct could not be visualized apart from many contaminants.

The expression of the TPR-related constructs led to consistently higher yields of protein than any of the other BcsC constructs (Figure 6.3). For example, distinct bands at the appropriate molecular weight were seen at 82 kDa and 87 kDa for BcsC$^{23-742}$ and BcsC$^{23-781}$, respectively. However, BcsC$^{23-813}$ showed high levels of degradation during expression, evidenced by the large band at 37 kDa when expected at 90 kDa. Although the degraded TPR construct was not ideal it was still used in the subsequent assays. BcsC$^{24-813}$ was the only TPR construct used in this study that contained only an N-terminal His$_6$ tag. Using the degraded construct permitted the exploration of the N-terminus of the TPR domain. When mapped out for the 37 kDa cut off, the degraded BcsC$^{24-813}$ construct collected constitutes approximately the first 342 amino acids in the sequence of this construct.

A

B

C
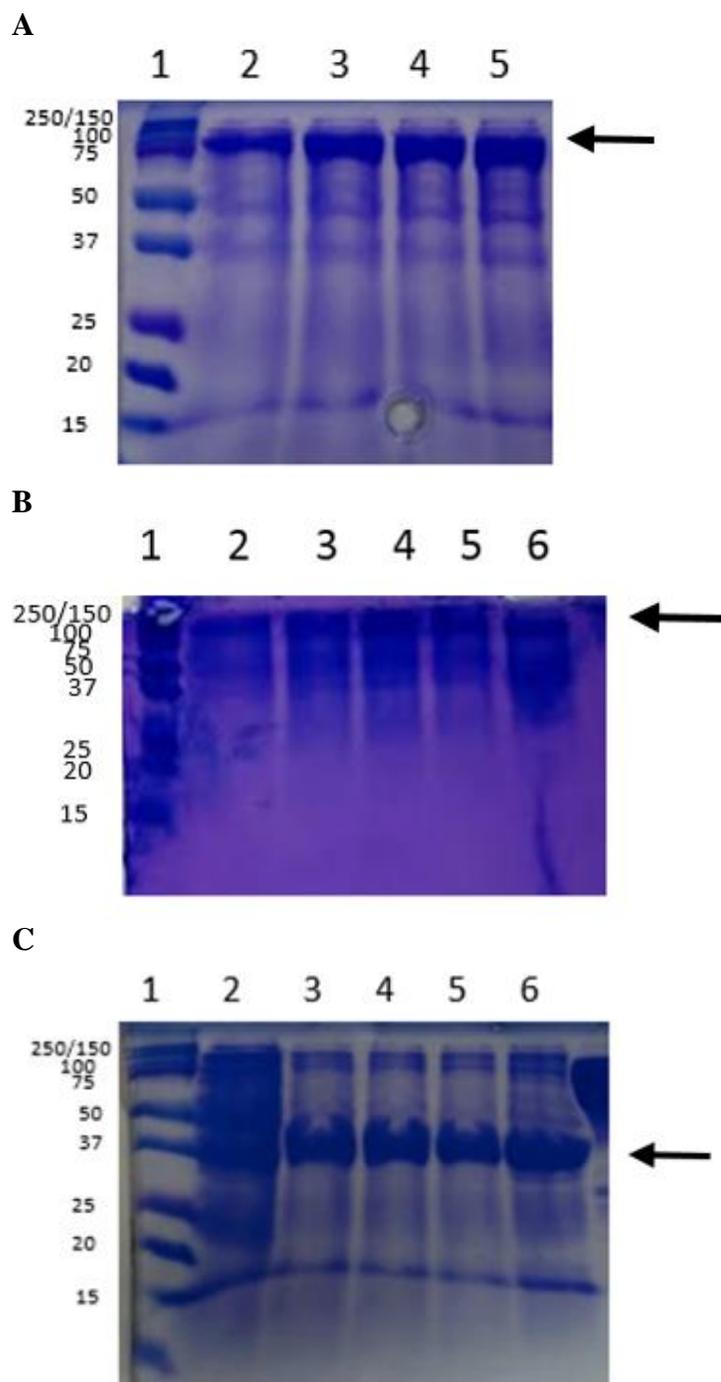
Figure 6.3: Expression of TPR constructs. Expression at 37°C for 4 h using 1 mM IPTG on a 12% SDS-PAGE, stained with Coomassie. (A) BcsC$^{24-742}$ indicated by the black arrow at 82 kDa; in the order of 1) protein standard, and expression at 2) 1 h, 3) 2 h, 4) 3 h 5) 4 h. (B) BcsC$^{24-781}$ indicated by the black arrow at 87 kDa; in the order of 1)

protein standard, and expression at 2) 0 h, 3) 1 h, 4) 2 h, 5) 3 h and 6) 4 h. (C) Degraded BcsC[24-813] indicated by the black arrow at 37 kDa. Intact protein would be expected at 95 kDa; in the order of 1) protein standard, and expression at 2) 0 h, 3) 1 h, 4) 2 h, 5) 3 h and 6) 4 h. Although there is a small band located at ~100 kDa which could be believed to be this full construct, it is not collected in purifications (seen later in Figure 6.11 B).

6.3.2 Purification of BcsC Constructs

BcsC has sequence homology to polysaccharide export proteins including AlgE/AlgK of the alginate biofilm synthesis complex. Using the research performed on these two proteins, purification methods were used as a template to purify the full and β-barrel constructs following the procedures used for AlgE (Whitney *et al*., 2009), and procedures used for AlgK (Keiski *et al*., 2007) were employed as a guide for TPR constructs.

6.2.1 Purification under Denaturing Conditions

All constructs containing a membrane associated region (full and β-barrel constructs) were subject to a harsh denaturant, testing urea and GuHCl, to allow for the solubilization of protein out of inclusion bodies. When found in inclusion bodies, the protein was self-associated and aggregated with other improperly folded proteins. Using these harsh detergents to solubilize and, thus, linearize all proteins present in the sample, the histidine tags became available (as they would have been buried within the inclusion bodies) and provided a means to capture the desired protein on a nickel column. An additional advantage to using denaturing conditions also provided a means to reduce the

effects of other proteins present in the cell, including proteases, as a means to reduce the amount of BcsC lost to degradation. Figure 6.2 displays a range of conditions tested post-purification as an indication of both the optimal expression conditions and solubilizing detergent for $BcsC^{29-1154}$ and $BcsC^{816-1154}$. The expression conditions were best seen using 37°C for both $BcsC^{29-1154}$ and $BcsC^{816-1154}$ as seen in Panel A lanes 2, 6, 8 and 9. In comparing the purity of the samples, it appears as though GuHCl resulted in less contaminants than urea (lanes 2 and 8, respectively) for $BcsC^{29-1154}$. Urea and GuHCl showed similar levels of purity (lanes 6 and 9 for GuHCl and urea, respectively). Both urea and GuHCl detergent trials were successful in providing a means to purify protein, however, urea was thought to have an advantage since it provided slightly higher amounts of desired protein. Remaining contaminants post purification can be removed using other purification techniques such as gel filtration and ion exchange, except, while initially promising, after a number of purification trials of the chosen protein constructs consistently led to high levels of degradation during purification as made evident in Figures 6.4 and 6.5 for $BcsC^{29-1154}$ and $BcsC^{816-1154}$, respectively. The aid of undergraduate thesis students was then employed to continue optimizing the remaining full-length and β-barrel constructs with some success (outlined in their research theses and summarized in Table 6.1) and begin protein folding trials. Meanwhile, purification studies were commenced for TPR constructs using native conditions.
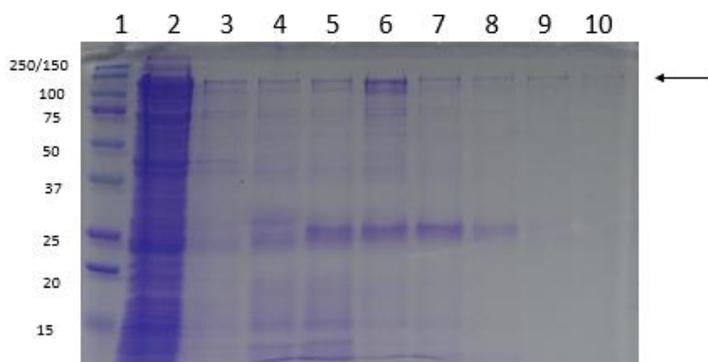
Figure 6.4: Purification of BcsC$^{29\text{-}1154}$ by NTA-Ni column. High levels of degradation are believed to occur based on the bands located at ~25 kDa. Samples were loaded in the order 1) protein standard, 2) flow through, 3) wash 1, 4) wash 2, 5) wash 3, 6) elution 1, 7) elution 2, 8) elution 3, 9) elution 4, 10) elution 5 on a 12% SDS-PAGE, stained with Coomassie. Arrow indicates the location of BcsC$^{29\text{-}1154}$.
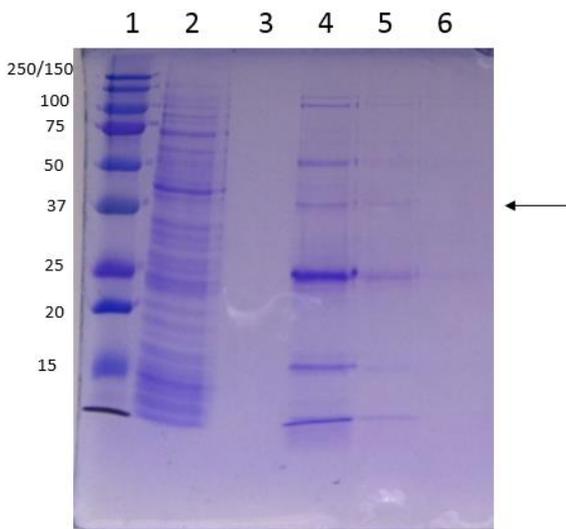
Figure 6.5: Purification of BcsC$^{816\text{-}1154}$ by NTA-Ni column. High levels of degradation are believed to have occurred based on the thick band at ~25 kDa. Samples are 1) protein standard, 2) wash, 3) elution 1, 4) elution 2, and 5) elution 3 on a 12% SDS-PAGE, stained with Coomassie. Very small amounts of BcsC$^{816\text{-}1154}$ were obtained. Arrow indicated the location of BcsC$^{816\text{-}1154}$.

After obtaining small protein yields and evidence of degradation this aspect of the project was handed off to two undergraduate thesis students (Deanna MacNeil followed by Nicole Trepanier) to continue to attempt purification optimization and refolding of each of the separate constructs. Following the thesis students research (specifics noted in their written thesis), purification (shown in Figure 6.6) and refolding (described in the refolding section) of BcsC$^{783\text{-}1141}$ was performed for the work of this thesis. Briefly, refolding involved using buffers with a mild detergent, capable of providing a suitable environment for protein folding, and decreasing concentrations of urea as folding progressed. While this study was only replicated once due to time constraints, it did lead to promising results with respect to purification of this variant of the β-barrel. Using the purification protocol outlined in the undergraduate thesis (Nicole Trepanier, 2015) provided higher purification yields than seen with BcsC$^{816\text{-}1154}$. This purification, however, may have been able to maximize the amount of protein collected by using additional elution steps. Furthermore, using an intermediate wash step with a low concentration of imidazole may aid in the removal of some contaminants found primarily in the first elution (Figure 6.6).
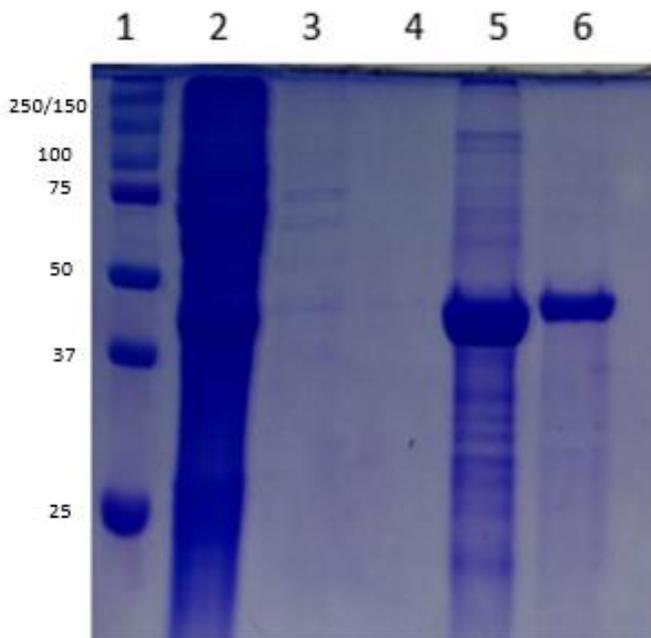
Figure 6.6: Purification of BcsC$^{783\text{-}1141}$ using previously optimized protocols. Lane 1 contains the protein standard. After collecting the flow through (lane 2), the column was washed twice with solubilization buffer (lanes 3 and 4). Two elution steps were performed (lanes 5 and 6) using 1 M imidazole. Samples were run on a 12% SDS-PAGE, stained with Coomassie

6.2.2 Native Purifications

Native protein purifications were performed in both the presence and absence of glycerol. Glycerol can aid in stabilizing protein during protein purification, however, in order to run further analysis (particularly CD), the glycerol needs to be removed. Glycerol has been known to improve protein stability by preferential hydration through electrostatic interactions, leading to more compact conformations (Vagenende *et al*., 2009), and as a result purifications performed with and without glycerol were necessary to keep the TPR proteins stable. Purifications were performed side by side to determine the importance of glycerol in the purification protocol of BcsC$^{24\text{-}742}$. Protein elutions from

both purifications were analyzed by SDS-PAGE (Figure 6.7; lanes 2-6 for the purification with no glycerol; lanes 7-10 for the purification with glycerol). Both purifications resulted in high concentrations of BcsC[24-742] as indicated by abundant bands at 82 kDa in each elution. There were also traces of contaminants in each purification protocol, mainly between the 20 and 25 kDa markers. These contaminants were removed in later steps using a 30 kDa MW cut-off concurrently with gel filtration or de-salting columns. Based on the identical results obtained from purifications with and without glycerol, it was determined that the presence of glycerol was unnecessary to obtain protein of similar puritiy. This was further assessed using CD analysis (presented in the Chapter 7).
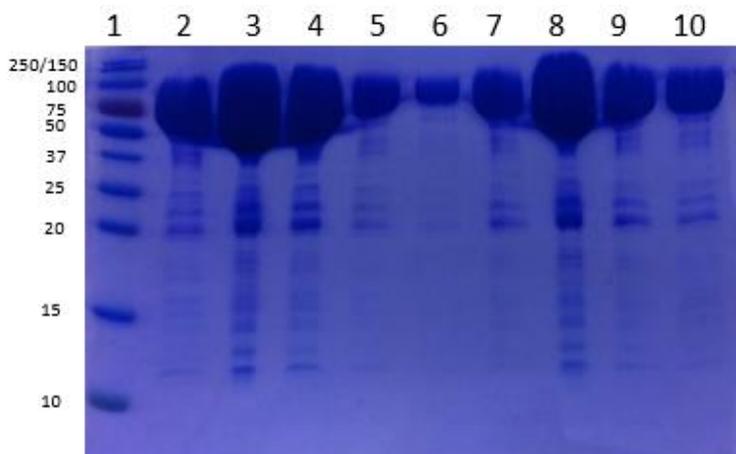
Figure 6.7: Purification of BcsC[24-742] using buffer with and without glycerol. Lane 1 contains protein standard. Lanes 2-6 were loaded with samples from elutions 1-5 in buffer containing no glycerol. Lanes 7-10 were loaded with samples from elutions 1-4 in buffer containing glycerol. Both purifications were successful in collecting large quantities of BcsC[24-742], indicated by thick bands at 82 kDa. They were also equally contaminated with other proteins, most of which are removed in further purification steps

including the use of 30 kDa MW cut-off concentrators and gel filtration or de-salting columns.

The optimization of protein purification for the TPR-related constructs also examined the optimal washing of the column (to remove unwanted contaminants) prior to elution of the desired protein. For example, Figure 6.8 portrays a purification with a wash step containing only the lysis buffer (*i.e.*, lacking imidazole) (lane 3; Figure 6.8) followed by a wash step with 20 mM imidazole in the same buffer (lane 4; Figure 6.8). For comparison, Figure 6.9 Panel A shows a similar purification, but elution of the protein following this wash led to a purer sample since it removed more of the contaminating proteins. Upon investigation using Western blot analysis, bands at 82 kDa were consistent in showing the presence of His_6-tagged BcsC protein. However, due to blotting conditions numerous background bands were also formed. Again, the additional wash provided a means to obtain a purer sample and, therefore, it was decided that the third wash would be a necessary trade-off to obtain a purer protein sample for downstream steps.
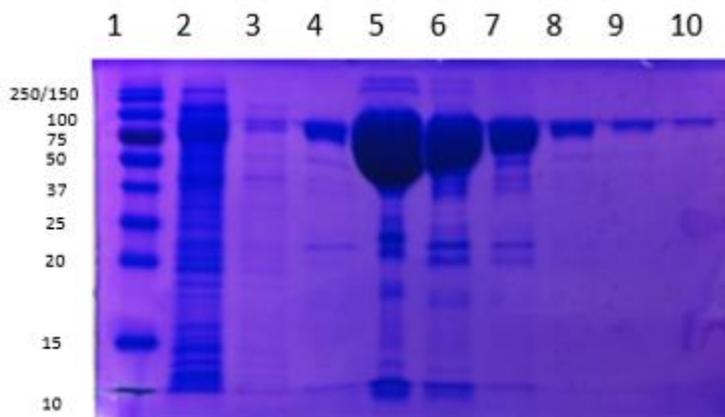
Figure 6.8: Purification of BcsC$^{24\text{-}742}$ using two washes on NTA-Ni resin. Lane 1 contains protein standard. Qiagen nickel beads were collected in a 10 mL column (BioRad) and flow through was collected (lane 2). Two wash steps were performed to clear any bound protein contaminants from the column. The first wash contained lysis buffer (lane 3) while the second wash contained 20 mM imidazole (in lysis buffer) to clear contaminants more tightly bound to the column (lane 4). Subsequent 1 mL elutions were collected and loaded in lanes 5-10 for elutions 1-6. Samples run on a 12% SDS-PAGE, stained with Coomassie.

A

B
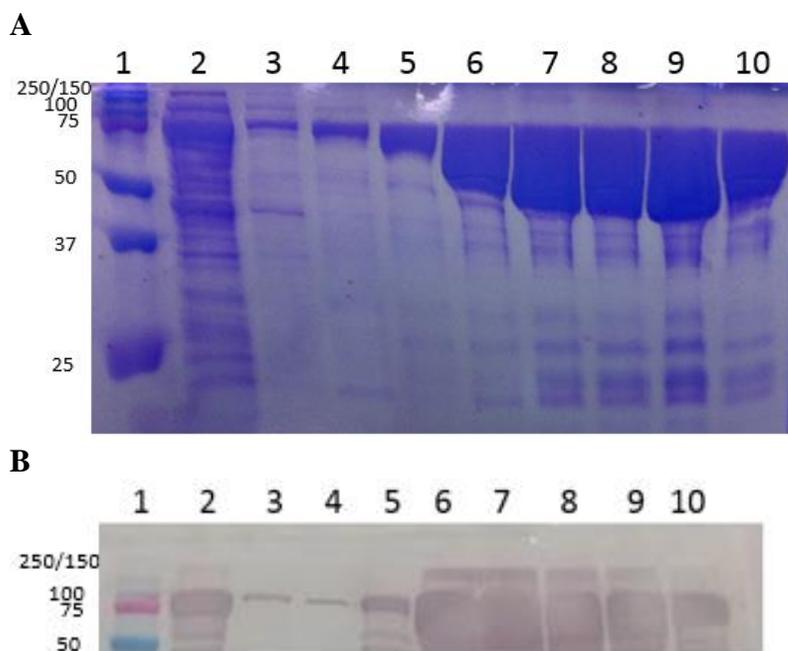
Figure 6.9: Purification of BcsC$^{24\text{-}742}$ using three washes. (A) Lane 1 contains protein standard. Qiagen beads were collected in a 10 mL column (BioRad) and flow through was collected (lane 2). Three wash steps were performed to clear any bound protein contaminants from the column. The first wash contained lysis buffer (lane 3), the second wash contained 20 mM imidazole (in lysis buffer) and the third wash contained 50 mM imidazole (in lysis buffer) to clear contaminants more tightly bound to the column (lanes 4 and 5, respectively). Subsequent 1 mL elutions were collected and loaded in lanes 6-10 for elutions 1-5. Samples were loaded on a 12% SDS-PAGE, stained with Coomassie. (B) Samples described in Panel A were also subject to Western blot analysis which consistently demonstrated the presence of BcsC$^{24\text{-}742}$ as indicated by thick bands at 82 kDa. High levels of background bands are apparent indicating protein overloading of the gel.

Gel filtration (size exclusion) chromatography was tested as a means to further purify samples, as well as remove excess salt and imidazole. Figure 6.10 depicts isolated BcsC$^{24\text{-}742}$ collected in samples from gel filtration (Panel B) following the separation of BcsC$^{24\text{-}742}$ from other contaminants observed on the accompanying chromatogram (Panel A). All samples collected containing BcsC$^{24\text{-}742}$ can be seen in lanes 5 through 8. These samples display a lesser amount of contaminants than the original purified from the

nickel column. Alternatively, de-salting columns coupled with concentration devices (30 kDa MW cut-off) were also employed as a faster means to remove imidazole and some of the contaminating proteins to a level that was similar to the gel filtration procedure. A comparison between gel filtration and de-salting columns with centricons can be made between lanes 5-8 of Figure 6.10 Panel B and lane 1 of Figure 6.12. Although the concentrated sample for BcsC$^{24-742}$ in Figure 6.12 appears to contain more contaminants than those found in Figure 6.10, the sample is still much purer than the product of the nickel column purification. De-salting procedures were routinely used with the PD10 De-salting columns to prepare samples for further experiments. All purification optimization for TPR constructs were performed on BcsC$^{24-742}$. Remaining constructs used the same conditions as the optimized procedure. Purification images of BcsC$^{24-781}$ and BcsC$^{24-813}$ can be seen in Figure 6.11. Concentrated protein for all three TPR constructs post de-salting column can be seen in Figure 6.12. The culmination of these studies suggested that the TPR constructs BcsC$^{24-742}$ and Bcsc$^{24-781}$ are the best candidates to continue working with since they both expressed and purified in high concentrations with little to no degradation apparent. Additionally, due to high amounts of degradation of BcsC$^{24-813}$, having only collected the first 37 kDa of this construct, this construct was continued to be analyzed throughout this thesis to investigate the N-terminus of the TPR domain.

A
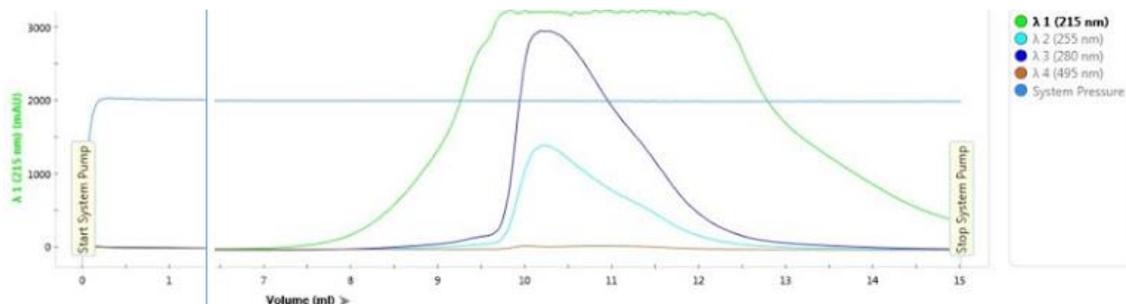
B
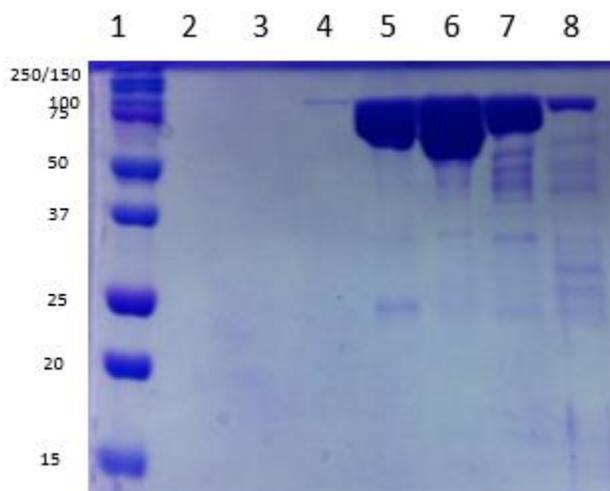
Figure 6.10: Gel Filtration of BcsC$^{24\text{-}742}$. (A) Adapted gel filtration chromatogram. The peak of the dark blue line indicates the presence of BcsC$^{24\text{-}742}$ as was determined when analyzed by SDS-PAGE. Shoulders beginning at 11 and 12 mL represent a heterogeneous mixture of BcsC protein and were therefore collected separately (original chromatogram can be seen in Figure A2.1 found in Appendix II. The vertical blue line indicates a break in the volume scale to better represent the data). (B) SDS-PAGE of gel filtration samples collected. Protein standard was loaded in lane 1. The samples collected consisted of the entire amount of solution to pass through the column, the first sample (lane 2) consisted of the first 7.5 mL to pass through the column, at this point the chromatogram indicated a small increase in the absorbance at 280 nm, thus, a second sample was collected from 7.5 to 8 mL (lane 3). Another small increase in the

chromatogram made for the collection of a sample from 8 mL to 9.5 mL (lane 4). Next, a dramatic increase in absorbance indicated the presence of BcsC$^{24-742}$ and, thus, a sample was collected from 9.5 to 11.3 mL (lane 5). The appearance of a shoulder forming caused for collection of a new sample from 11.3 to 12.1 mL (lane 6), and a second shoulder from 12.1 to 13 mL was collected separately (lane7). A final sample from 13 to 14.5 mL was collected to ensure all of BcsC$^{24-742}$ was captured. Lanes 5 to 8 indicate that the majority of BcsC$^{24-742}$ eluted off the column between 9.5 and 14.5 mL. Due to small amounts of contaminants, these samples were pooled and used for further analysis.

A

B
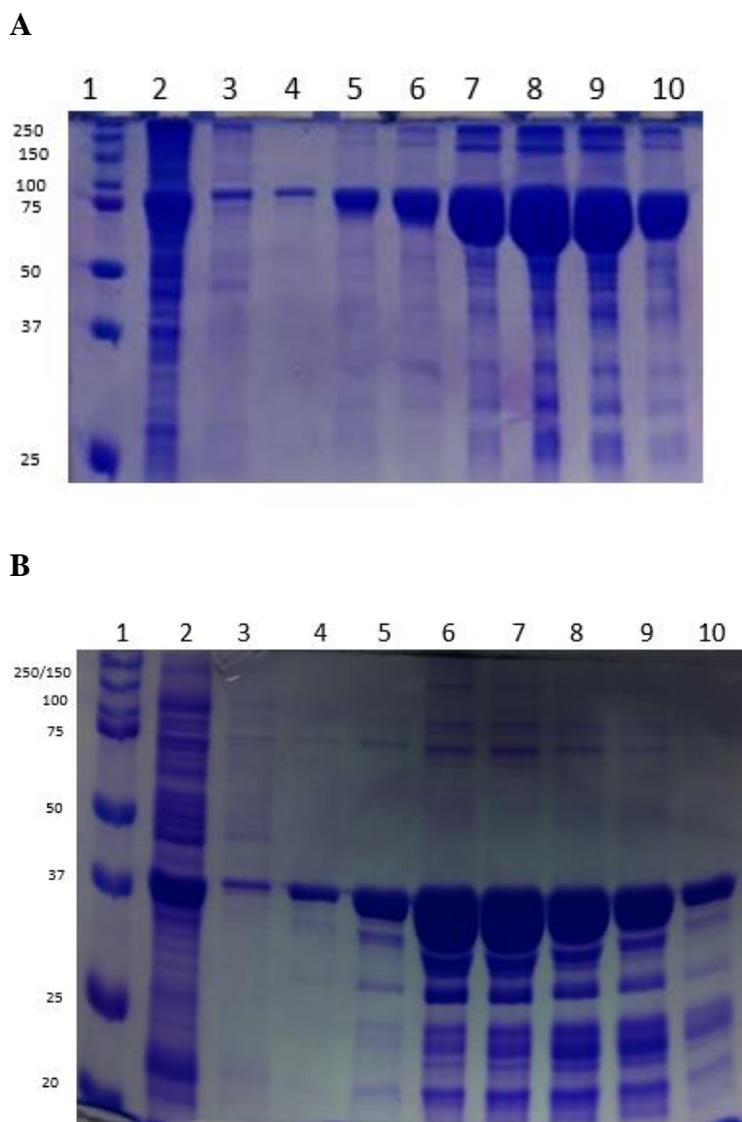
Figure 6.11: Purification of remaining TPR expressed protein constructs. Samples of (A) BcsC$^{24-781}$ and (B) BcsC$^{24-813}$ were loaded on a 12% SDS-PAGE, stained with Coomassie. Purifications were performed using the optimized procedure developed for BcsC$^{24-742}$. Both constructs were purified using nickel resin on a 10 mL column. Lane 1 contains protein standard. Flow through collected is represented in lane 2. Three washes were performed to release any bound contaminants with increasing concentrations of imidazole (in lysis buffer). Wash 1 contained no imidazole shown in lane 3. Imidazole concentrations were raised to 20 mM in wash 2 (lane 4), and to 50 mM in wash 3 (lane 5). Elutions were collected using 250 mM imidazole in 1 mL aliquots and loaded in lanes

6-10 for elutions 1-5. All purifications resulted in contaminated protein that was further cleared using additional purification techniques.
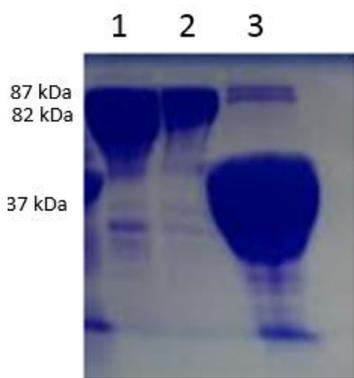
Figure 6.12: Concentrated TPR constructs post de-salting column. A majority of contaminants have been removed from each purified construct using the 30 kDa MW cut-off concentrators and de-salting columns. Lane 1 was loaded with sample from concentrated BcsC[24-742] (82 kDa), lane 2 with concentrated BcsC[24-781] (87 kDa) and lane 3 with concentrated BcsC[24-813] (37 kDa). Samples were loaded on a 12% SDS-PAGE, stained with Coomassie.

Following protein purification by nickel affinity purification and concentrators with de-salting columns or gel filtration, the purified samples were re-concentrated to a volume of ~200 μL. Before continuing with following experiments the concentration was measured for each preparation using a BioDrop spectrophotometer programmed with the appropriate molecular weight and extinction coefficient for each BcsC construct. The approximate concentration of a typical purification for each construct has been provided in Table 6.1. The TPR constructs were most successful in obtaining high quantities of protein; particularly purifications of BcsC[24-742] consistently resulted in the most amount of protein collected. All three constructs, however, collected enough protein to perform a full assay for subsequent experiments.

Table 6.1: Protein Constructs Purified To Date

Construct	Residues	Approximate Concentration (mg/mL)*	Additional Notes
Full**			
	24-1141	6.1	
	24-1154	1.2	Degradation
	29-1141	NYP	
β-Barrel**			
	783-1141	3.4	
	783-1154	2.2	
	800-1141	3.3	
	800-1154	5.4	
	816-1154	1.1	Degradation
TPR			
	24-781	40	
	24-742	60	
	24-813	20	Degradation
	29-813	NYP	

*As reported following concentration of the IMAC fractions for the full length and β-barrel constructs and after the desalting and concentration steps for the TPR constructs. Protein concentration was determined using the molecular weight and extinction coefficient for each respective construct (table 5.1) programmed into a BioDrop for direct quantification.

**Note: Full-length and β-barrel construct information obtained by Deanna MacNeil (2014 BI499 research thesis work).

NYP=Not Yet Purified

6.2.3 Refolding of BcsC β-barrel

Initial refolding experiments were performed by a thesis student in the lab using the full-length construct, BcsC[783-1141]. Due to calculation error, it was found that the concentration of the various components of each refolding buffer were not as initially expected. The original refolding conditions used refolding buffers containing 0.00036%

83

(w/v) LDAO, 2 mM Tris-HCl (pH 7.2), 500 mM NaCl and decreasing concentrations of urea (*i.e.* 6M, 3 M, 1.5 M, 0.75 M, 0.33 M and 0 M). The refolding detergent, LDAO, forms spherical micelles with a radius of 20.7 Å (Timmins *et al*., 1988) when used above the critical micelle concentration of 1 mM (0.023% (w/v)). In the previous undergraduate study, the concentration of 0.00036% (w/v) LDAO was far below the critical micelle concentration and therefore would not have provided a suitable environment for protein folding. LDAO has been widely used for solubilization, isolation, purification and crystallization of OMP's (Riguad *et al*., 2000, le Maire *et al*., 2000). Furthermore, the buffer concentration for Tris-HCl was not high enough to effectively control the pH of the refolding environment.

To determine if a protein is effectively folding from its linearized state, samples were analyzed by SDS-PAGE. The expected result for protein folding is indicated by a band migration further down the gel in comparison to the linearized sample. Folded proteins are more compact and therefore, better able to migrate through the acrylamide pores of the gel than the linear, unfolded counterpart. When analysed for band migration in the original refolding buffers used by the undergraduate thesis student, it was found that the protein constructs did show a folding pattern indicated by a band migration from 43 kDa to 23 kDa, as seen in Figure A2.2 in Appendix 2. Although, even with a band migration pattern, the protein refolding was only shown to be successful as far as the buffer containing 0.33 M urea. Further folding into the final folding buffer, in the absence

of urea, resulted in precipitation of all remaining protein when analyzed for protein concentration using a BioDrop.

As an attempt to remedy the mass amount of protein loss new buffers were used which contained higher concentrations of LDAO and Tris-HCl than the concentrations used in the previous undergraduate thesis experiment (consisting of 0.06% (w/v) LDAO, 20 mM Tris (pH 8.0), 500 mM NaCl, and decreasing concentrations of urea (*i.e.* 6M, 3 M, 1.5 M, 0.75 M, 0.33 M and 0 M)). These conditions provide a more suitable environment to maintain protein stability by increasing the Tris-HCl concentration and effectively fold the linearized protein by increasing the folding detergent concentration to be higher than the critical micelle concentration. Using refolding buffers with higher concentrations of Tris-HCl and LDAO, the refolding experiment was repeated for BcsC[783-1141] and analyzed by SDS-PAGE shown in Figure 6.13. In Panel A it appears as though the protein may be folding indicated by a shift from 43 kDa (white arrow) to 23 kDa (blue arrow) with an intermediate band located at ~30 kDa (orange arrow). Additionally there was an apparent migration upwards to ~70 kDa (black arrow). This may be indicative of oligomerization of the folded protein. For example if three monomers (which migrate to 23 kDa) were to form a trimer they should form a band around 69 kDa. Oligomerization of β-barrels folded in LDAO has been previously seen in another OMP, PhoE. In a study performed by Jensen and colleagues (2000), folded PhoE presented as a β-barrel using 0.2% (w/v) LDAO as a detergent. The outcome from the PhoE study resulted in the detection of a trimeric form for PhoE (Jensen *et al.*, 2000),

which is this protein's native form (de Cock *et al*., 1991). Since BcsC[783-1141] shows signs of oligomerization it is likely that we have obtained properly folded protein, although it is recommended to explore different concentrations of LDAO in an attempt to optimize protein folding. In the PhoE study, the LDAO concentration was well above that used in folding BcsC, and, thus, may be an indication that higher concentrations of LDAO may be required for optimal folding.

The band migration between 8 M and 0 M urea is not visible due to collecting sample prior to concentration, thus, the samples were too dilute. It is advised that further folding studies with this construct concentrate samples prior to collecting samples for SDS-PAGE analysis. Panel B shows the corresponding Western Blot from the folding study. The migrated bands from Panel A (black, orange and blue arrows) did not develop in the Western blot, rather the only bands to form were linearized protein (lanes 2 and 9). The lack of bands at 23, 30 and 70 kDa may be due to a lack of sensitivity. Another curious outcome of this experiment was the consistency between the heated and non-heated samples on the SDS-PAGE (Figure 6.13 Panel A lanes 9 and 8, respectively). By boiling one of the samples, it was expected that any folded protein would linearize and form a band at 43 kDa. Instead, the heat-denatured sample was identical to the folded sample on the SDS-PAGE. . Since these bands are consistent in both heated and non-heated samples, this may be evidence that the β-barrel is fairly stable and requires a longer denaturation time period to completely unfold all protein present in the sample. This outcome becomes more curious as the Western blot (Figure 6.13 Panel B) does not

show an identical banding pattern, whereby, it appears that the heated sample indicates

linear protein in the Western blot, which is not seen in the SDS-PAGE.

A
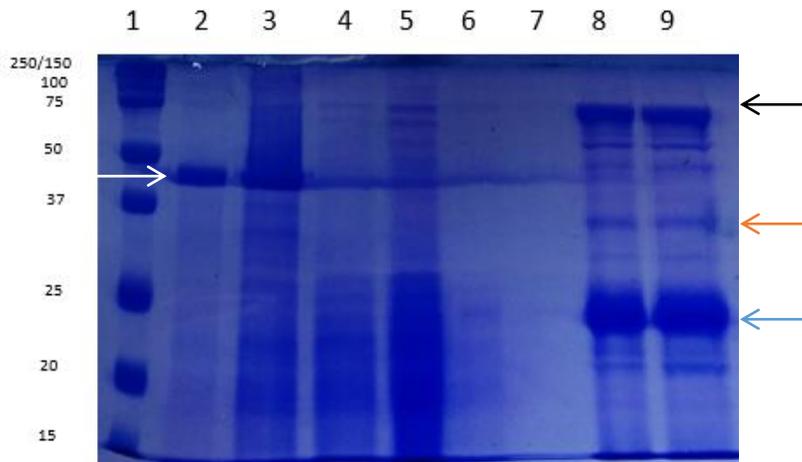
B
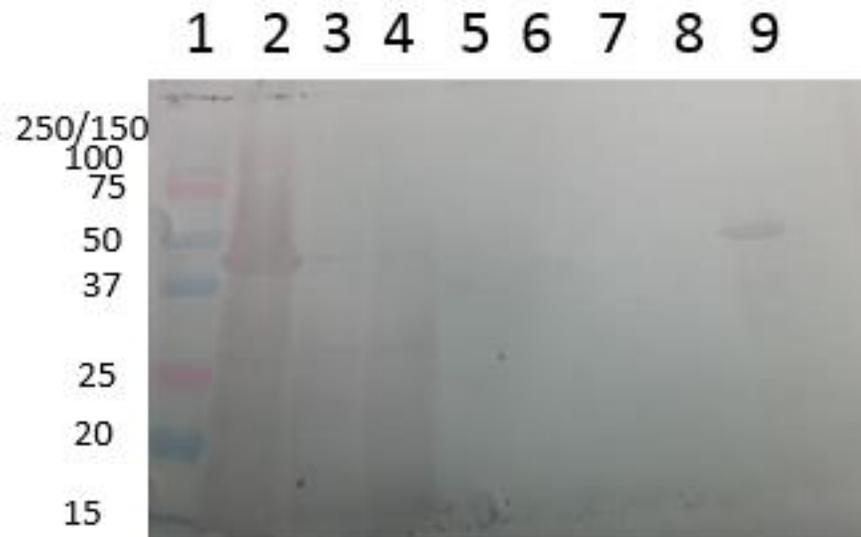
Figure 6.13: Protein folding performed on BcsC$^{783-1141}$ using 0.06% (w/v) LDAO. (A) Band migration of BcsC$^{783-1141}$ appears to be evident by SDS-PAGE. Lane 1 contains protein standard. Lane 2 contains sample from purified protein (elution 2 seen in Figure 6.6) to compare fully linearized protein with decreasing concentrations of urea. Lane 3 was loaded with sample containing 3 M urea, lane 4 with 1.5 M urea, lane 5 with 0.75 M urea, lane 6 with 0.33 M urea, and lanes 8 and 9 with 0 M urea without and with heat

denaturation, respectively. Protein folding may be evident based on a shift from a 43 kDa band (white arrow; lanes 2 and 3) to 23 kDa (blue arrow; lanes 8 and 9). There also appears to be an intermediate folding step shown at ~30 kDa (orange arrow; lanes 8 and 9). Additionally, a band has migrated upward to the 70 kDa area on the gel (black arrow; lanes 8 and 9), which may be indicative of oligomerization of three units shown at the 23 kDa mark. Samples were loaded on a 12% SDS-PAGE, stained with Coomassie. (B) Westen blot analysis of BcsC$^{783-1141}$ did not indicate band migration between linear and folded samples. Band formation was only observed in the 8 M urea elution sample (lane 2) and the heated 0 M urea sample (lane 9). Samples were loaded in the same order used for Panel A.

The protein concentration of each of the BcsC$^{783-1141}$ samples was monitored during the protein folding process using a BioDrop programmed with the extinction coefficient and molecular weight (found in Table 5.1). These numbers are tabulated in Table 6.2, which indicate that most of the protein purified was lost to aggregation as the protein folding experiment ensued. A similar trend in the results, as to what was seen visually on the SDS-PAGE gels (Figures A2.2 and 6.13), was observed, whereby both experiments showed a band migration between the linearized protein and protein in refolding buffers. While 0.5 mg of total protein was recovered after the procedure, high amounts of protein were lost at each step (about 92% total protein loss). Due to time constraints, this experiment could not be repeated, however, it would be advised that for future protein folding studies with BcsC, protein samples be taken after concentrating in new refolding buffers as opposed to before based on the inability to visualize the shift in bands due to diluted samples. Another aspect, as stated earlier, would be to increase the LDAO concentration which may help fold more protein, and, hopefully, reduce protein

loss. Additionally, longer incubation times in new refolding buffers may prevent protein loss. The addition of glycerol to protein folding buffers may also help to stabilize protein as it folds, preventing aggregation, and supporting the formation of ordered states of the folding protein. Finally, concentration steps may play a role in protein aggregation, whereby concentrating the protein too high causes the protein to precipitate out. Limiting the time/volume of concentrating the protein may play a role in obtaining higher yields of folded protein.

Table 6.2: BcsC$^{781\text{-}1141}$ Yield with Step-Wise removal of Urea

Concentration of Urea (M)	Protein concentration after using concentrators (mg/mL)	Approximate volume after concentrating (mL)	Approximate amount of protein present (mg)	% Loss
8	18.2	3	54.6	0
6	8.3	4	33.2	39.2
3	3.8	4.4	16.7	69.6
1.5	NA	1.5	NA	NA
0.75	1.6	2.2	3.5	93.6
0.33	0.6	0.6	0.36	99.3
0	0.1	0.2	0.5	99.9

Purification conditions for full length and β-barrel constructs were performed following the methods used in previous studies for the alginate export β-barrel protein,

Alg E (Whitney *et al*., 2009). This method acquired protein yields of ~20 mg which was then refolded in β-ᴅ-maltopyranoside using a drop-wise procedure. Provided the suggestions described above for folding optimization of BcsC are not effective, altering the current protein folding procedure to follow this technique may aid in refolding the β-barrel constructs in future studies. Purification of the TPR, on the other hand, did not correspond with the purification procedure used by Keiski and colleagues (2007) on AlgK. When cloned into its expression vector, a signal sequence was added to AlgK that resulted in exporting of the protein to the periplasm. Additionally, the buffers used to purify AlgK used a lower pH than the ones used for the TPR constructs of BcsC (pH 7.5 for AlgK as opposed to pH 8.0 for BcsC). $BcsC^{24-742}$ and $BcsC^{24-781}$ both have a pI of 7.35. If the pH of a protein buffer were to correspond to the pI of that protein, it will become less soluble in the buffer causing precipitation of the protein. The pH was, therefore, raised for experiments with BcsC.

6.3 Conclusions

Protein expression and purification of the various BcsC constructs were explored for optimization. Full and β-barrel ($BcsC^{29-1154}$ and $BcsC^{816-1154}$, respectively) constructs were used to determine optimal expression conditions. Both constructs, however, displayed degradation during purification trials. Thesis project students took over the β-barrel purification and refolding portion of this study. Upon re-investigation with the β-barrel construct, $BcsC^{781-1141}$, purification and refolding experiments were shown to be more promising, but still require further optimization. The TPR constructs $BcsC^{24-742}$ and

90

BcsC$^{24-781}$ proved to be the most promising constructs, resulting in high yields of protein that were relatively pure. By obtaining high quantities of pure samples obtained from a single purification, more time can be spent analyzing the protein for structural and functional characteristics. This is particularly important when performing crystal trials, which require very high purity and varying concentrations of protein (described in more detail in Chapter 7).

7. STRUCTURAL ANALYSIS OF BcsC

7.1 Background

Structural analysis of recombinant BcsC proteins were performed to assess their folded state. CD provides a means for characterizing the secondary structure of a protein (Greenfield, 2006). CD utilizes circular left-handed and right-handed beams of polarized light to obtain an absorbance profile, typically at the far UV spectra between 260 nm and 190 nm (Kelly *et al.,* 2005). When the two beams of light travel through the sample, each will have a different absorbance when compared to each other and an overall absorption can be calculated for each wavelength used (Kelly *et al.,* 2005). The absorbances obtained from 260 nm to 190 nm provide a profile for the protein under investigation, which can be compared to standard profiles that make up α-helices, β-strands and disordered regions. An example of the different profiles obtained using CD can be seen in Figure 7.1. CD can also be useful in determining protein stability, and has potential in analysis of substrate binding. Once a spectral profile has been obtained a number of deconvolution software programs can be used to analyze the data. These software programs provide useful information including the composition of the protein based on percentages of α-helices, β-strands and turns. To determine if a substrate is stabilizing or destabilizing a protein, the helical/strand/turn content can help determine what is happening to the secondary structure as substrate is added.
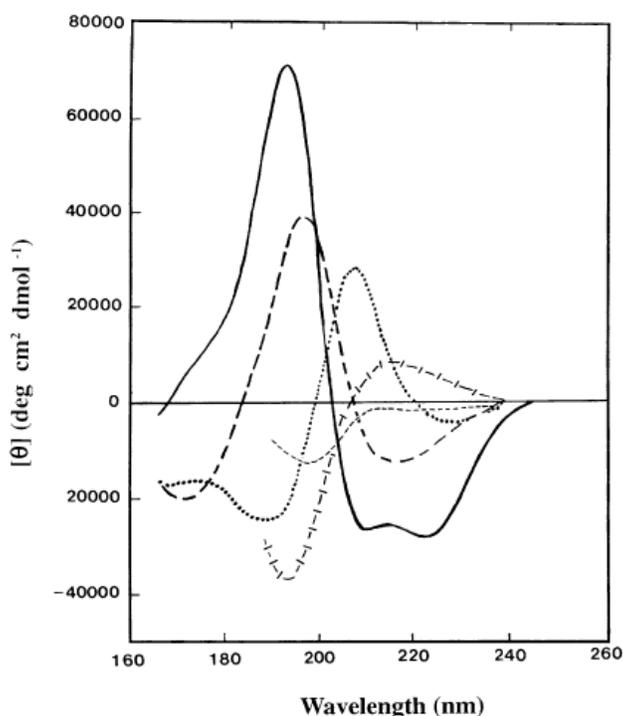
Figure 7.1: Circular Dichroism Profiles. Based on the profile collected from a far UV spectra between 190 nm and 260 nm the ratio of α-helices, β-strands and random coils can be determined providing secondary structural features of a protein. Solid line= α-helix; long dashed line= anti-parallel β-strand; dotted line= type I h-turn; cross dashed line= extended 31-helix or poly (Pro) II helix; short dashed line= irregular structure (Kelly *et al.,* 2005).

In order to obtain reliable data to construct a three dimensional-structural model

of a protein by X-ray crystallography, the protein construct must be capable of forming

protein crystals. Crystal trials are performed on highly pure and concentrated protein

using a variety of crystallizing conditions. These conditions are solutions that vary in pH,

salt, buffer and precipitant (Smyth and Martin, 2000). Crystal trays can be set up using a

variety of techniques, however, members of the Weadge lab prefer one of two methods.

The first method involves mixing small volumes (0.5-1 μL) of concentrated protein with

a crystalization condition in a small well located next to a larger reservoir containing a larger volume of the condition (typically 50-100 µL). The second method, called the hanging drop method, involves mixing the protein and condition on a plate that is suspended above the reservoir of the condition. Both of these techniques require sealing the wells from the external environment. With the crystallization solution in the reservoir at a higher concentration than the diluted crystallization solution (from mixing with concentrated protein in buffer), water is lost from the drop to the reservoir. As water is diffusing from the drop, the concentration of the crystallization solution slowly increases providing a shift towards the labile zone (Figure 6.1 Panel B). As described in Figure 7.1 Panel B, the labile zone is where protein nuclei are formed. The formation of protein nuclei allows for crystals to form. Protein crystals are made up of light atoms forming large unit cells, which results in a weak diffraction. To make up for weak diffraction, protein crystals must have a large volume (Smyth and Martin, 2000).

In addition to protein crystals, salt crystals may form in the plate. To differentiate between the two, crystals can be evaluated using a few simple methods. The first of these is to check for fluorescence with UV light. When UV light is shone on a protein crystal, the aromatic residues (tryptophan, tyrosine and phenylalanine) within the protein emit fluoresce caused by electron excitation within a chromophore of these residues. There are a variety of plastics that can also fluoresce when excited by UV light. To determine if plastic pieces are present (as well as determine if salt crystals are present) a protein dye, Izit, may be applied to the crystal. Izit works by entering the solvent channels of a protein

crystal, turning the crystal dark blue (Hampton Research, 2015). Salt crystals do not have solvent channels large enough for the dye to penetrate, thus the crystal does not get stained (Hampton Research, 2015), and any plastic present would not uptake the dye as it would not contain these channels either. The disadvantage to using the protein dye is that it renders the crystals unusable, and, therefore, would only be used in the case that multiple crystals were formed. A final method by which a crystal can be tested for its validity is by probing the crystal, if the crystal shatters it was protein, however, again the crystal would be lost.
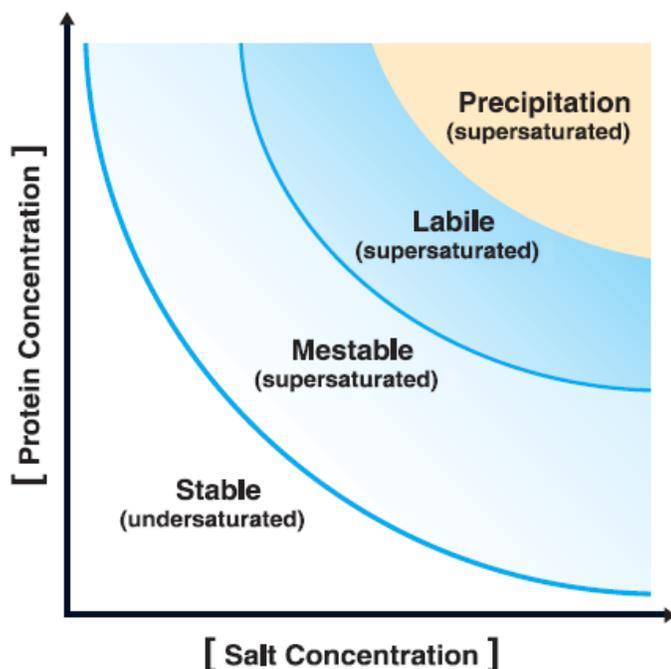
Figure 7.2: Diagram of zones of crystallization. To obtain protein crystals, nuclei must be formed in the labile zone. Here, the salt concentrations are high and the protein is supersaturated without causing precipitation. Once nuclei are formed, the conditions can be brought down to the metastable zone by decreasing either the protein concentration,

salt concentration or both. The metastable zone is also supersaturated but will not form nuclei. If the concentration of either the protein or the salt becomes too high, precipitate will form and crystal formation is impaired. If the concentration of either the protein or the salt are too low, the protein will not be saturated enough to form protein crystals effectively (Hampton Research 2013).

7.2 Results and Discussion

7.2.1 Circular Dichroism

Circular dichroism was performed on the various TPR constructs for the purpose of obtaining a secondary structure profile of each construct. Using the bioinformatics tools discussed in Chapter 4, the constructs that were created from the beginning portion of the gene were predicted to fold into a super-helical structure that makes up a TPR. The main purpose of performing CD was to confirm the predicted α-helical structure and ensure that the protein was properly folded in its current buffer for downstream experiments. Minima in the CD spectra located at 208 and 222 nm are indicative of an α-helical secondary structure (Kelly *et al*., 2005), and thus, would suggest that the TPR is mainly α-helical. For comparison, each experimental dataset was analyzed using deconvolution software (provided through the Dichroweb internet site) resulting in the profiles seen in Figures A3.1-A3.8 for TPR constructs BcsC [24-742] and BcsC[24-781]. As part of a stability comparison over time, some of the TPR constructs were incubated with CMC, which is a commercially available, soluble form of cellulose that may be a suitable version to facilitate binding to this TPR. This was attempted given the putative role of the BcsC TPR in facilitating the transport of cellulose across the periplasm (outlined in the

introduction). While these studies were geared towards analyzing the stability of the various TPR constructs, it did also suggest a functional role that will be discussed here, but more fully expanded upon in Chapter 8 of this thesis.

Each of the deconvolution programs used provided the percent helical content with the de-convoluted data. Included with the helical content given was the normalized root mean square displacement (NRMSD). An NRMSD of >0.3 is indicative of a good fit to structural deconvolution (Hall *et al*., 2014). Keeping this in mind, each TPR construct was analyzed with deconvolution programs CDSTTR (using reference sets 4 and 7), K2D and SELCON3 (using reference set 4). For both BcsC$^{24-742}$ and BcsC$^{24-781}$ based on the NRMSD of each sample, the deconvolution program that best fit the data was CDSTTR using reference set 7 (Table 7.2 for BcsC$^{24-742}$; Table 7.6 for BcsC$^{24-781}$), therefore, this analysis will focus on data obtained from CDSTTR (reference set 7). BcsC$^{24-813}$ was not successfully de-convoluted, therefore, data was not collected for this particular construct.

Based on the indicative minima located at 208 and 222 nm on Figure 7.3, BcsC$^{24-742}$ shows evidence of α-helical secondary structure. BcsC$^{24-742}$ was assessed using samples containing 0, 0.05%, 0.1% and 0.5% (w/v) CMC (0 CMC was 15 days old, the 0.05% (w/v) and 0.1% (w/v) CMC samples were 10 days old and 0.5% (w/v) CMC sample was 17 days old). The helical content as reported using CDSTTR (reference set 7) for protein sample containing no CMC reported a helical content of 80%. As CMC concentration increased the helical content decreased to 74%, 71 % and 67% for samples

containing 0.05%, 0.1% and 0.5% (w/v) CMC, respectively. Moreover, the spectral profiles of increasing CMC concentrations display a spectral change in intensity as CMC concentrations increase (Figure 7.3). Additional figures have been provided in Appendix 3, displaying the profile shifts for de-convoluted data created by each of the software programs used for analysis (Figures A3.1-A3.4 each of which correlate with Tables 7.1-7.4). The profile shift observed in Figure 7.3 may be reflective of the decrease in helical content as reported above. In another CD spectral study performed by Cliff and colleagues (2005), the addition of substrate affected the helical content of a TPR domain under investigation. In the study by Cliff *et al*. (2005) the protein under investigation was the TPR domain of protein phosphatase 5 (TPR-PP5) which had a helical content of 64%. Upon the addition of the ligand, pentapeptide MEEVD, the helical content increased to 98%. In the spectral profile of TPR-PP5 with and without pentapeptide MEEVD the profile shifted downward upon the addition of ligand. The similar observations in the TPR-PP5 study support the hypothesis that the TPR domain of BcsC is interacting with CMC.

From the spectra obtained for BcsC[24-742], samples run with varying CMC concentration it can be concluded that the spectra adopts an α-helical structure, in agreement with the secondary and tertiary structures predicted using PSIPRED and Phyre[2]. The highly helical conformation was an indication that the purification protocol was effective in providing an environment that was conductive to proper protein folding, in particular the absence of glycerol did not affect the stability of the protein (as

mentioned in the previous chapter). As CMC concentration increases, the helical content of the protein decreases coinciding with a spectral change in intensity which may be an indication that BcsC is capable of binding CMC.

Additional information provided by the deconvolution program CDSTTR (reference set 7) include the percentage of β-strands, turns and disorder (Table 7.2). As BcsC$^{24-742}$ incubated with increasing CMC concentrations, the β-strand content increased from 6% to 7.9% and 10% for 0, 0.05% (w/v) and 0.1% (w/v) CMC, respectively, followed by a decrease to 7% in the 0.5% (w/v) CMC sample. In contrast, there was very little change in the turn content. The sample containing no CMC showed a turn content of 7%, while all other samples had a turn content of 8%. Following the trend of the β-strands, protein disorder increased from 7% to 10% and 11% in samples containing 0, 0.05% (w/v) and 0.1% (w/v) CMC, respectively, and decreased to 8% disorder in the 0.5% (w/v) CMC sample. These changes in β-strand content and protein disorder were very minimal, thus, can be disregarded. Due to time constraints available on the spectropolarimeter only BcsC$^{24-742}$ was analyzed with different concentrations of CM, 0, 0.05% (w/v), 0.1% (w/v), 0.5% (w/v). These concentrations were chosen because cellulose absorbs in the far UV region (Min Dong and Gray, 1997), therefore, the concentration had to be limited to prevent interference with the secondary structural data.

CD analysis was performed on all three TPR constructs, however, BcsC$^{24-742}$ was the only construct analyzed using varying concentrations of CMC. Due to length of time

of data acquisition and limitation on the availability of the instrument, we elected to pursue fluorescence measurements with the TPR to analyze further interactions between CMC and the various constructs.
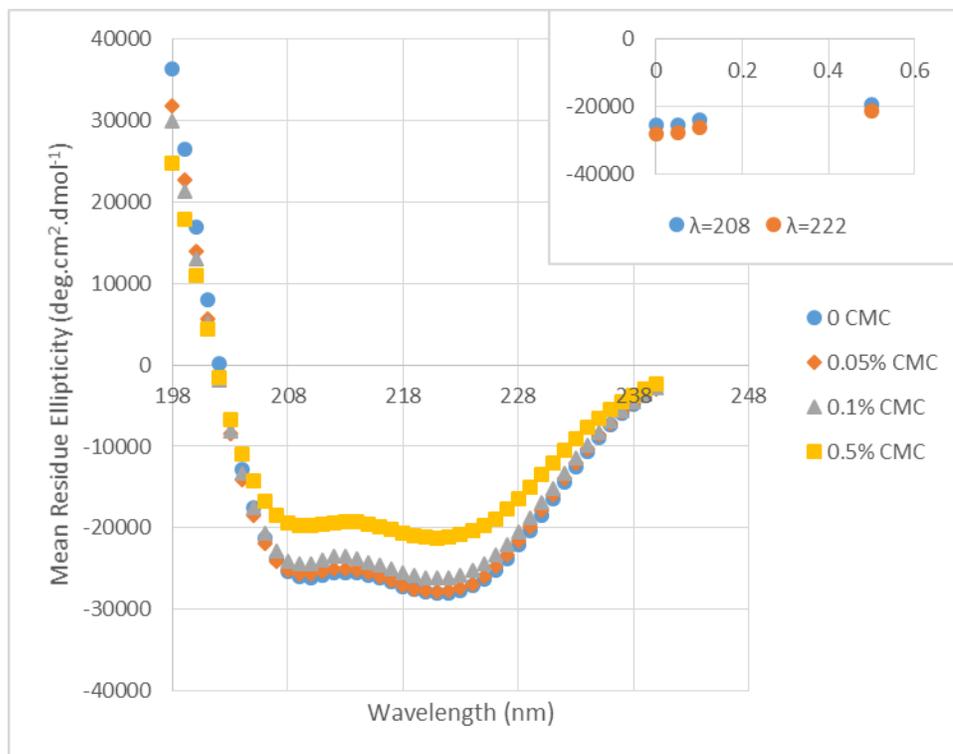
Figure 7.3: Experimental CD spectrum of BcsC$^{24\text{-}742}$ in the presence and absence of CMC. Data shown here was collected using protein purified and analyzed at different time points to one another. Protein samples were prepared in running buffer and diluted to a final concentration of 2 µM BcsC$^{24\text{-}742}$. Samples were run using concentrations of 0, 0.05%, 0.1% and 0.5% (w/v) CMC. As CMC concentration increases, the spectral profile shifts up, indicating an interaction between BcsC$^{24\text{-}742}$ and CMC causing a weaker signal collected in higher CMC concentrations.

Table 7.1: Deconvolution Analysis of BcsC$^{24\text{-}742}$ Using CDSSTR with Reference set 4

% (w/v) CMC	Helical Content (%)	β-Strand Content (%)	Turn (%)	Disorder (%)	NRMSD
0	77	12	2	9	0.003
0.05	67	12	9	12	0.005
0.1	69	10	10	11	0.008
0.5	62	11	10	17	0.068

Table 7.2: Deconvolution Analysis of BcsC$^{24\text{-}742}$ using CDSSTR with Reference set 7

% (w/v) CMC	Helical Content (%)	β-Strand Content (%)	Turn (%)	Disorder (%)	NRMSD
0	80	6	7	7	0.004
0.05	74	7.9	8	10	0.007
0.1	71	10	8	11	0.007
0.5	67	7	8	8	0.005

Table 7.3: Deconvolution Analysis of BcsC$^{24-742}$ using K2D

% (w/v) CMC	Helical Content (%)	β-strand Content (%)	Disorder (%)	NRMSD
0	87	0	13	0.029
0.05	86	0	14	0.045
0.1	81	0	19	0.22
0.5	71	3	26	0.068

Table 7.4: Deconvolution Analysis of BcsC$^{24-742}$ using SELCON3 with Reference set 4

% (w/v) CMC	Helical Content (%)	β-Strand Content (%)	Turn (%)	Disorder (%)	NRMSD
0	81	0	11	11	0.09
0.05	64	0.8	6	29	0.210
0.1	61	2	7	30	0.219
0.5	58	3	5	34	0.197

BcsC$^{24-781}$ was analyzed with 0.5% (w/v) CMC at two different time points; 10 days post purification and 17 days post purification (each stored at -20°C). Again, the spectral minima at 208 and 222 nm confirm a helical secondary structure (Figure 7.4). There was very little shift between the profiles, which indicates that the protein maintains its secondary structure over time (Figure 7.4, corresponding with Tables 7.5-7.8). CMC was added to the protein sample before running CD analysis with an incubation period

that allowed for the interaction between the protein and CMC to reach equilibrium. Again, due to time constraints with the spectropolarimeter, samples could not be repeated using a variety of CMC concentrations for this TPR construct. In order to analyze the protein with CMC the highest concentration used throughout this CD experiment was also used for this construct. Appendix 3 contains the individual deconvolution software programs with their respective profiles for both 17 and 10 day old samples in Figures A3.5-A3.8, coinciding with Tables 7.5-7.8.

As a means to consistently examine the TPR constructs, this analysis will focus on the deconvolution program CDSTTR (reference set 7) as it successfully produced a good fit to the reconstructed profile, reflected by the low NRMSD values (Table 7.6). Helical content of BcsC$^{24-781}$ was consistent between the two samples (73% α-helical for the 10 day old protein sample and 74% α-helical for the 17 day old sample). With a difference of only 1% between the two samples analyzed (and the older sample with a higher helical content) this outcome further supports the claim that the TPR construct BcsC$^{24-781}$ maintains its secondary structure over a week-long period. Additionally, the helical content of BcsC$^{24-781}$ is relatively close to the BcsC$^{24-742}$ 0.1% (w/v) CMC sample (which had a helical content of 71%). Since it was determined that the addition of CMC to BcsC$^{24-742}$ caused a decrease in the helical content of samples with increasing concentrations of CMC, it is possible that the helical content of BcsC$^{24-781}$ without CMC would be a higher value than 80% (which was the helical content of BcsC$^{24-742}$ without CMC) likely due to an extra TPR on the C-terminus of BcsC$^{24-781}$. This prediction is

103

based on the fact that the helical content of BcsC$^{24-781}$ in 0.5% (w/v) CMC is higher than that of BcsC$^{24-742}$ which had a lower CMC concentration at 0.1% (w/v) CMC.

Interestingly, the values for the older sample, purified 17 days prior to analysis, had a higher β-strand and turn content than the 10-day-old sample. There was a 2% difference seen between the two samples for both the amount of β-strands and turns. The disorder between the two samples remained consistent at 9% for both. The consistency between the two samples further support the concept that the protein retains its secondary structural characteristics over a week-long period of time. Furthermore, experiments performed on older protein would be beneficial in order to determine at which point the TPR begins to lose the integrity of its secondary structure.
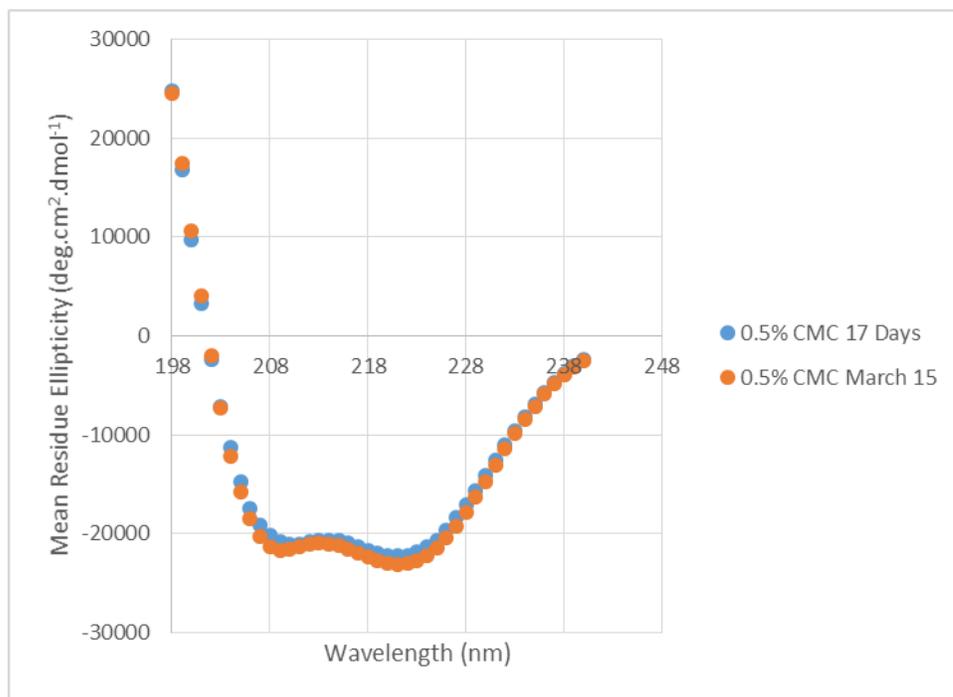
Figure 7.4: Experimental CD spectrum of BcsC[24-781] measured at two different time points post-purification. Data shown here was collected using protein purified and analyzed at different time points to one another, 17 days and 15 days post purification. Protein samples were prepared in running buffer and diluted to a final concentration of 2 μM BcsC[24-781]. Samples were run using a concentration 0.5% (w/v) CMC. The two samples analyzed displayed similar spectral profiles, and thus indicate that BcsC[24-781] does not lose the integrity of its secondary structure.

Table 7.5: Deconvolution analysis of BcsC[24-781] using CDSSTR with Reference set 4

Sample	Helical Content (%)	β-Strand Content (%)	Turn (%)	Disorder (%)	NRMSD
17 Days	61	12	11	15	0.011
10 Days	60	11	11	18	0.008

Table 7.6: Deconvolution analysis of BcsC$^{24\text{-}781}$ using CDSSTR with Reference set 7

Sample	Helical Content (%)	β-Strand Content (%)	Turn (%)	Disorder (%)	NRMSD
17 Days	74	12	9	9	0.008
10 Days	73	10	7	9	0.008

Table 7.7: Deconvolution analysis of BcsC$^{24\text{-}781}$ using K2D

Sample	Helical Content (%)	Disorder (%)	NRMSD
17 Days	78	22	0.105
10 Days	79	21	0.092

Table 7.8: Deconvolution analysis of BcsC$^{24\text{-}781}$ using SELCON3 with Reference set 4

Sample	Helical Content (%)	β-Strand Content (%)	Turn (%)	Disorder (%)	NRMSD
17 Days	58	3	5	34	0.097
10 Days	62	0.4	5	32	0.205

The experimental data for BcsC$^{24\text{-}813}$ (Figure 7.5) represents α-helical secondary structural characteristics indicated by minima located at 208 and 222 nm. When analyzed

using the same deconvolution software, however, the profile of the de-convoluted spectra portrayed flat lines across the graphs, meaning that the data could not be processed using this software. Referring back to the previous chapter, this construct showed high levels of degradation in protein expression, and is instead a 37 kDa piece of the N-terminus of the TPR. Since this construct (from amino acids 24-813) has been shown to be easily degraded it is possible that the protein fraction being analyzed has disordered regions within this protein. Although the spectra appears to be helical in nature, the experimental data collected during CD analysis must not have been consistent with the reference data sets used by the various deconvolution software, preventing the ability to produce useable reconstructed data. As a result of the inability to collect reconstructed data, no information was gained on the helical, sheet, turn or disorder content of these samples. Although the anticipated information was not collected from the Dichroweb software, there was a profile shift in the intensity consistent with that seen in the analysis of BcsC$^{24\text{-}742}$. Again, with the addition of higher concentrations of CMC, a spectral shift in the intensity occurred which further indicates that BcsC$^{24\text{-}813}$ is displaying an interaction with CMC. Therefore, it can be further implied that the TPR domain of BcsC will interact with cellulose. This hypothesis was further examined in Chapter 8 using cellulose binding assays.
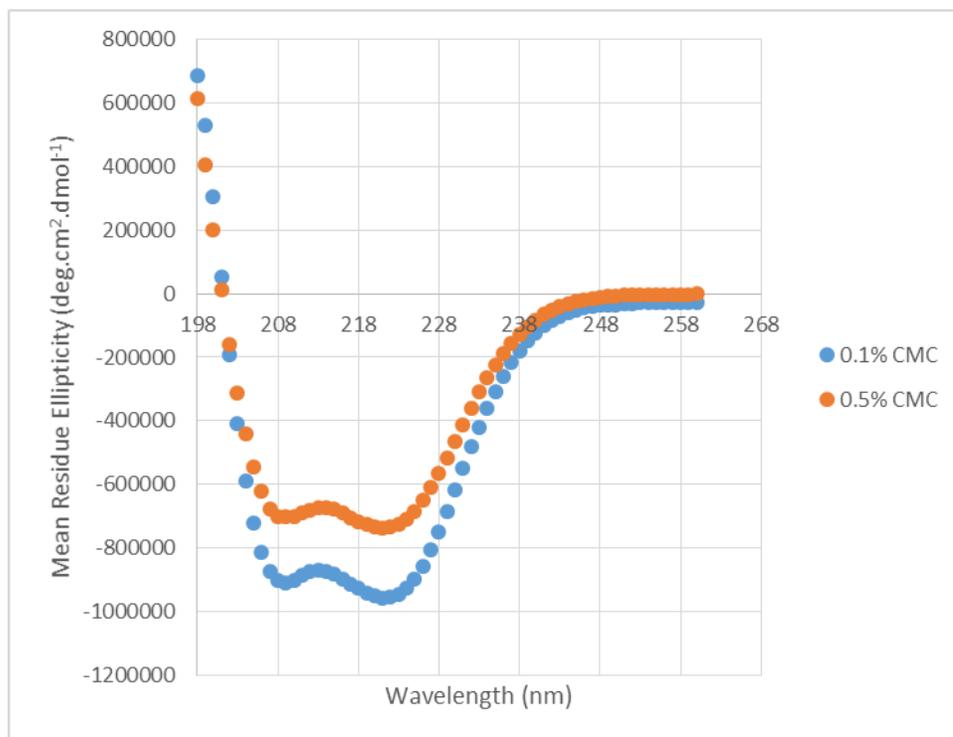
Figure 7.5: Experimental CD spectrum of BcsC[24-813] in the presence of CMC. Data shown here was collected using protein purified and analyzed at the same time point. Protein samples were prepared in running buffer and diluted to a final concentration of 2 μM BcsC[24-813]. Samples were run using concentrations of 0.1% and 0.5% (w/v) CMC. As CMC concentration increases, the spectral profile shifted up, indicating an interaction between BcsC[24-813] and CMC causing a weaker signal collected in higher CMC concentrations.

An additional assay performed using CD involved heating BcsC[24-742] from 15°C to 85°C while collecting a spectral profile using the α-helical indicative wavelength, 222nm (Figure 7.13). The purpose of this assay was to obtain a melting temperature, T_m, the point at which half the sample has become denatured due to the increase in heat. Using the Hill equation the T_m was found to be 51°C. Additionally, when reviewing the

melting curve (Figure 7.13) it can be seen that the protein holds its conformational state at temperatures up to ~40°C. The relevance for the protein to remain folded up to 40°C supports the concept that this protein construct is likely in its native conformation, as it would most likely be found in the bacterial environment within a mammalian host. The main significance to this outcome provides information on the stability of this protein, allowing for further experiments to be designed within these parameters.
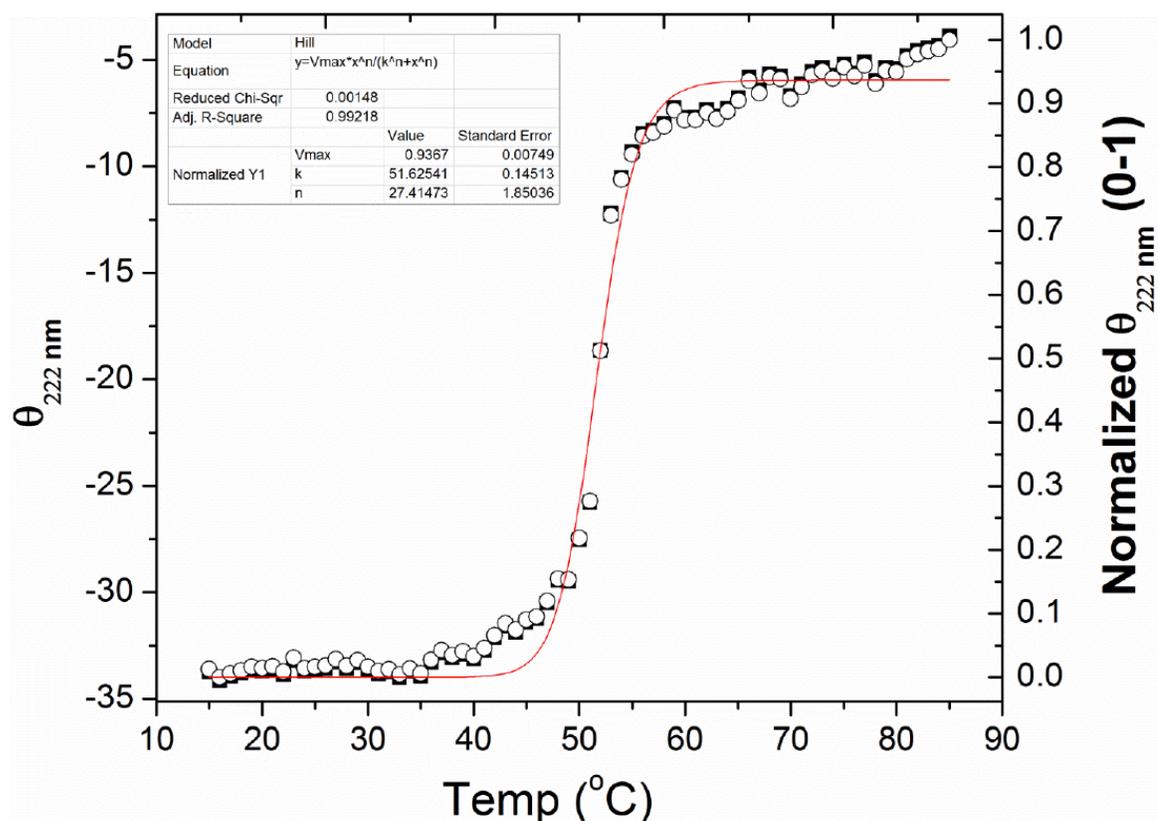
Figure 7.6: Melting curve for BcsC[24-742] using an elipticity at 222 nm to monitor changes in secondary structure. Melting temperature was obtained using the Hill equation (inset), T_m= 51°C.

7.2.2 Crystallization Trials

Crystal screens were performed on the various TPR constructs using the PEG and MCSG screens. Crystal screens were set up for the purpose of obtaining 3D models of the various TPR constructs, which would be used later on with X-ray crystallography. Multiple conditions appeared promising, providing crystals with clear, sharp edges. Upon inspection using UV, only one very small crystal (appearing in a condition with 0.1 M sodium acetate:HCl pH 4.6, 0.8 M lithium sulfate mother liquor) fluoresced when excited with UV light, seen in Figure 7.14. This fluorescence signal was a result of exciting tryptophan residues within the protein (this construct contains 9 tryptophan residues). Since this was the only crystal to emit light with UV it was not tested with Izit dye, rather it was used to create crystal seeds, however, no crystals appeared in the seeded conditions. The crystal was confirmed to be protein when the crystal seeds were created, and the crystal broke down upon probing, a characteristic of only protein crystals; thus disproving the possibility of the crystal being salt or plastic. All other conditions tested with Izit were suggested to contain salt crystals as they did not absorb the dye. While these trials are ongoing with different constructs and concentrations, there are no conditions that have led to crystal growth as of yet.

A   **B**
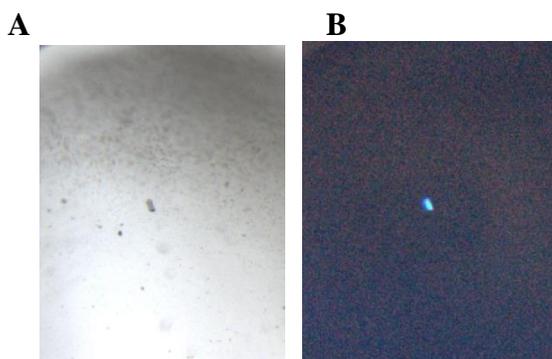
Figure 7.7: BcsC$^{24\text{-}742}$ crystals with fluorescence. (A) and (B) show the same crystal without and with UV light, respectively. Grown in 0.1 M sodium acetate:HCl pH 4.6, 0.8 M lithium sulfate with a concentration of 10 mg/ml BcsC$^{24\text{-}742}$ for 1 year.

During crystal trials, many crystals were further analyzed to determine if they were composed of protein as opposed to salt or inorganic material. Using the protein dye and UV fluorescence it was determined that only one of the crystals obtained was composed of protein (mentioned above). This and other methods used to obtain additional protein crystals were unsuccessful. Since protein crystal nuclei and protein crystals need different conditions to grow the seeds collected needed to be placed in a metastable environment (described above). It is clear that none of the trials using different concentrations of mother liquor were able to provide this environment. Additionally, this protein was grown using a relatively low concentration (10 mg/mL) that took a long time to form. It is, therefore, debatable as to whether the protein concentration used when seeding new crystals should have been 10 mg/mL, a higher concentration of protein with a lower concentration of mother liquor, or a lower concentration of protein with varying mother liquor concentrations, such that the end

111

result would place the seeds into a metastable environment. Additionally, the protein used for crystal seeding was from a newer purification than the original crystal. If the purity of the protein was not adequate for crystallization, then the environment within the drop would not permit the formation of more protein crystals. Further purification steps may be required to obtain highly pure protein, including using the 30 kDa MW cut-off concentrators and de-salting columns paired with gel filtration and/or using ion exchange chromatography. Finally, the optimization of growing protein crystals will require many more trials. Using the crystallization conditions utilized to grow the protein crystal represented in Figure 7.7, optimization of crystal formation could be performed by varying the pH of the solution and the concentration of the salts found therein.

In the study performed by Keiski and colleages (2010) the crystal structure of the TPR protein, AlgK, was obtained. Initial crystals obtained in this study were grown using a gene from *P. aeruginosa*, however, these crystals were not sufficient to use for X-ray crystallography. Rather, they used the gene from another species, *P. fluorescens*, to attempt to crystallize AlgK. By switching organisms they were able to obtain a crystal that obtained a resolution of 2.5 Å for AlgK (Keiski *et al*., 2010). Based on this simple solution, it may be worth investigating BcsC in a different organism such as *Salmonella* spp. or another strain of *E. coli*. It may also be worth investigating the conditions in which AlgK was grown; [10%(w/v) PEG 6000, 0.1 MES pH 6.0] (Keiski *et al*., 2007; Keiski *et al*., 2010). The crystal structure of AlgK showed a TPR protein that was made up of 9.5 TPR repeats (Keiski *et al*., 2010). Since the conditions used for AlgK provided

a suitable environment for this TPR protein, a variation of these conditions may be suitable to obtain a crystal suitable for X-ray crystallography for the TPR domain of BcsC.

7.3 Conclusions

This portion of the study successfully characterized the secondary structure of two BcsC TPR constructs. Having confirmed that these constructs are helical, the bioinformatics analysis of these constructs have been validated and purification protocols have been proven to be effective in providing a suitable environment for the protein. Furthermore, using increasing concentrations of CMC laid the ground work that there may be protein-carbohydrate interactions between the TPR and cellulose, shown by CD analysis. By analysing protein purified at different time points, we can also be confident that these protein constructs maintain their secondary structure, allowing more time to perform various experiments with the purified protein. Finally, the TPR was found to be a considerably stable protein with respect to temperatures that would be encountered in environmental conditions, including temperatures as high as 37°C, the internal gut temperature of a mammalian host.

Protein crystal trials were performed using the various BcsC TPR constructs and resulted in the production of a simple crystal. Seeding attempts were unsuccessful in growing new protein crystals. However, by using the crystallization condition used to grow this crystal, further trials in altering the concentrations of the condition components

may be more successful in obtaining multiple crystals. The possibility of growing BcsC TPR crystals using the gene from another organism is also another option in order to obtain a suitable crystal for X-ray crystallography, and ultimately determining the 3D structure of this TPR domain.

8. FUNCTIONAL ANALYSIS OF BcsC

8.1 Background

The primary experiment used in this study to determine the potential interaction of BcsC with cellulose was fluorescence spectroscopy. The main theory behind fluorescence involves the excitation of fluorophores in a protein, particularly the intrinsic fluorescence of aromatic residues: tryptophan, tyrosine and phenylalanine. Fluorescence operates by promoting an electron to an excited state in the fluorophore. When the electron returns to its ground state an emission of light occurs (and the protein fluoresces) (Möller and Denicola, 2002). Each aromatic residue has its own unique excitation and emission spectra. In water, tryptophan is excited from 295-305nm and has an emission maximum near 350-355 nm (295 nm is typically used in fluorescence experimetns), tyrosine is excited at 275 nm and has an emission maximum near 303 nm, and phenylalanine is excited near 258 nm with an emission maximum near 282 nm (Lakowicz, 2010). Based on the selected excitation wavelength, it is possible to select a specific residue to analyze. The emission is determined by the quantum yield, the ratio between the photons emitted and the total photons absorbed. A low quantum yield would indicate a low emission efficiency (Möller and Denicola, 2002). Phenylalanine has a low quantum yield and is therefore, not typically measured in fluorescence spectroscopy experiments. Additionally, tyrosine is often quenched in native proteins (Möller and Denicola, 2002). Therefore, tryptophan is the best-suited amino acid for fluorescence analysis. Tryptophan fluorescence is sensitive to its environment and when a protein

undergoes conformational changes, ligand binding or denaturation, the spectral profile of the protein changes (Möller and Denicola, 2002). This study is investigating the binding of the TPR domain of BcsC with a soluble cellulose product, CMC, by observing for a quenching pattern as the concentration of CMC added increases.

Tryptophan residues are relatively abundant in the TPR constructs, ranging from 9 to 11 in the TPR constructs alone, and a total of 29 tryptophan residues in the whole BcsC protein. Therefore, fluorescence quenching experiments offer a good way to monitor protein interactions. Furthermore, literature surveys have indicated that carbohydrate-binding modules (CBM) often contain at least two conserved tryptophan residues located in the binding sites of many of these proteins (Bolam *et al.*, 2001; Machovič *et al*., 2005; Machovič *et al*., 2006). By probing these sites we can learn more about BcsC, potentially leading to determining the binding mechanism of the TPR to cellulose and ultimately leading to better understanding of the mechanism by which the TPR facilitates the transport of cellulose through the periplasm towards the β-barrel for export. In the constructs studied in this project there are 9 tryptophan residues in $BcsC^{24-742}$, 10 in $BcsC^{24-781}$ and 11 in $BcsC^{24-813}$ (although this construct is degraded, and within the first 37 kDa of this construct only 4 tryptophan residues from the N-terminus are present).
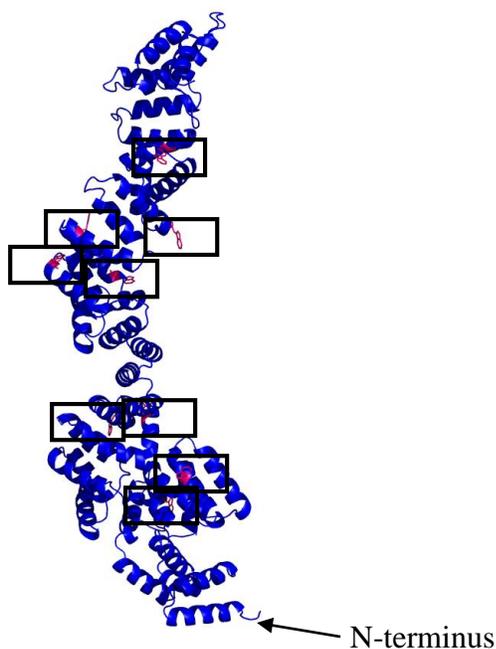
N-terminus

Figure 8.1: Highlighting tryptophan residues of BcsC$^{24\text{-}742}$. Tryptophan residues as seen in pink (outlined by boxes) are often found in binding sites of CBM proteins. There is a cluster of tryptophan residues located at the N-terminus of this model, which may indicate a binding site in this region. Conserved tryptophan residues from *E. coli* are located at amino acids 153, 342, 352, 496, 560 and 612 as determined using alignment software with *Salmonella* spp., *Yersinia* spp., *Pectobacterium* spp., *Klebsiella* spp. and *Pseudomonas* spp.. Additionally, 8 of the 9 tryptophan residues depicted on this model are located in helix segments, while the remaining tryptophan residue is located on a loop.

A secondary functional assay was performed using an insoluble form of cellulose, Avicel, to test for protein-carbohydrate interactions, as well as to determine if the TPR domain can bind varying forms of cellulose. The ability for the TPR to bind different forms of cellulose could further support the hypothesis that the TPR domain binds the emerging cellulose chain for export. Protein bound to insoluble Avicel can easily be separated from the unbound (soluble) fraction and subjected to SDS-PAGE analysis to

monitor the extent of the binding interaction. When analyzed by SDS-PAGE, the protein-carbohydrate interactions taking place between BcsC and Avicel will be evidenced by the presence of a band, if bands are consistent from low to high concentrations of Avicel the construct displays stronger binding than bands increasing in intensity.

8.2 Results and Discussion

8.2.1 Fluorescence Analysis

Fluorescence analysis of the various BcsC TPR constructs titrated with CMC are indicated in Figures 8.2, 8.3 and 8.4. Using an excitation wavelength of 295 nm, tryptophan residues were excited and emission spectra from 300 nm to 450 nm were measured. Maximum emission for the three constructs was at 335 nm for $BcsC^{24-742}$ and $BcsC^{24-781}$ and 325 nm for $BcsC^{24-813}$. Using increasing concentrations of CMC as a quenching agent the spectral profiles decreased consistently for each construct. Quenching profiles of $BcsC^{24-742}$, $BcsC^{24-781}$ and $BcsC^{24-813}$ can be seen in Figures 8.2, 8.3 and 8.4, respectively.

Focusing on the emission at the peak of the spectral profiles for $BcsC^{24-742}$, 335 nm, an overall decrease in emission of 37% occurred between samples containing 0 and 18.5 mg/mL CMC (Figure 8.2). The decrease in emission is indicative of tryptophan quenching by CMC, and thus confirms binding between the TPR and CMC. Since the samples consistently showed a peak at 335 nm, no spectral shift occurred when analyzing this construct. This is an indication that the protien did not change in its structure around

the tryptophans being probed. Furthermore, as stated above, the emission of tryptophan residues in water typically occurs around 350 to 355 nm. The emission peak of BcsC[24-742] shows a blue shift in the spectra (with a maximum at 335 nm) which is indicative of burried trypotphan residues, as opposed to being found on the surface where they would be exposed to the solvent (Lakowicz, 2010).
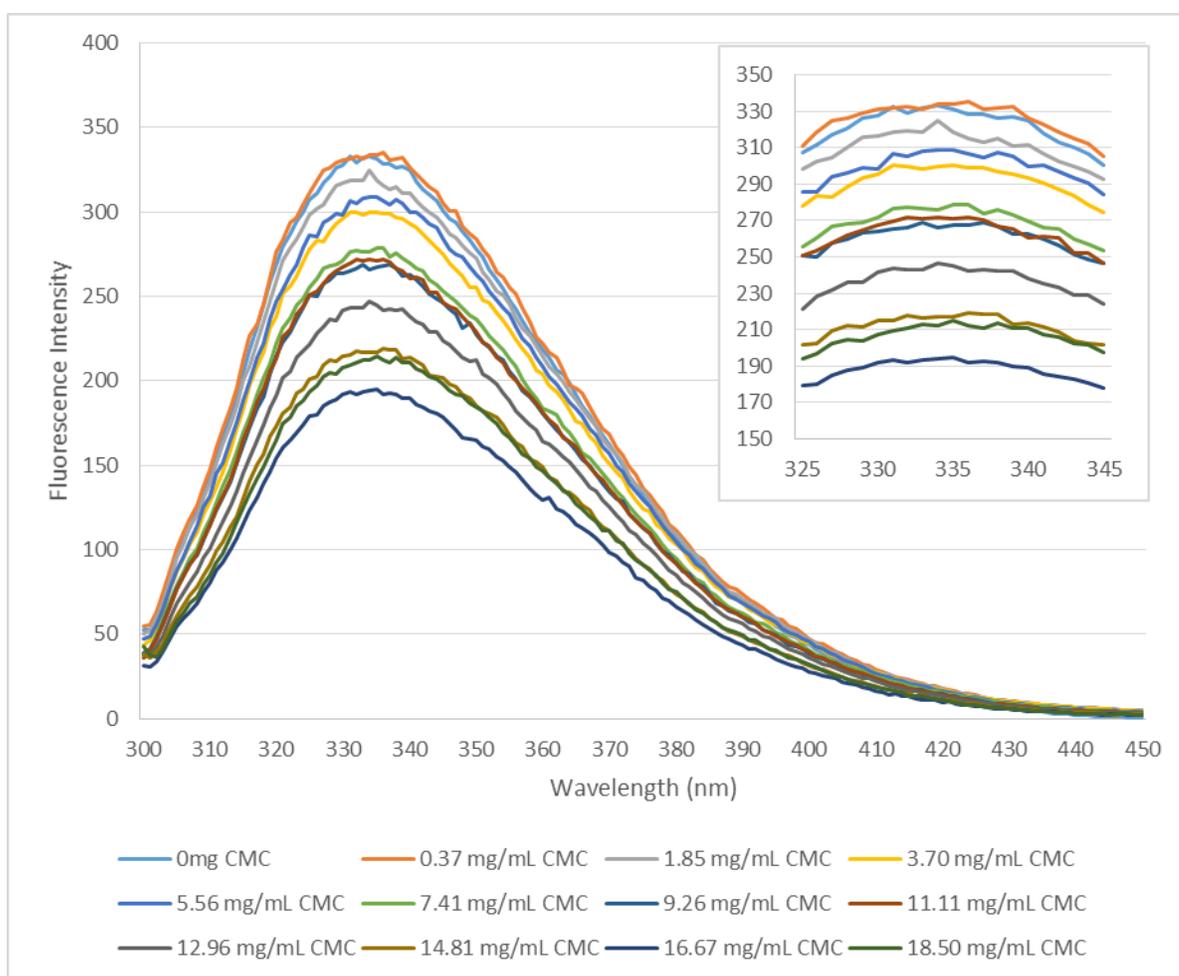
Figure 8.2: Fluorescence emission spectra of BcsC[24-742]. Increasing the concentration of CMC results in increasing quenching of tryptophan residues. Protein concentration at 3 µM of BcsC[24-742] was used.

BcsC[24-781] obtained a similar emission profile whereby the spectral peak occurred at 335 nm with decreasing intensities as the CMC concentration increases (Figure 8.3). One notable difference between the spectra for BcsC[24-781] and BcsC[24-742] is the large gap between the 16.67 mg/mL and 18.50 mg/mL samples of BcsC[24-781], whereas the difference between these same concentrations in BcsC[24-742] is more consistent with the quenching pattern seen thorough out the rest of the profiles. The decrease in emission for BcsC[24-781] from samples containing 0 and 18.5 mg/mL CMC was 53%. However, the difference from 0 to 16.67 mg/mL CMC was only 32%. The consistent decrease of emission is indicative of tryptophan quenching on BcsC[24-781], indicating that this construct is also binding CMC. As seen in BcsC[24-742], the emission peak of all samples for BcsC[24-781] from concentrations 0 to 18.5 mg/mL were consistently located at 335 nm. This strongly suggests that this construct also did not change in the local environment surrounding the tryptophans being assessed.
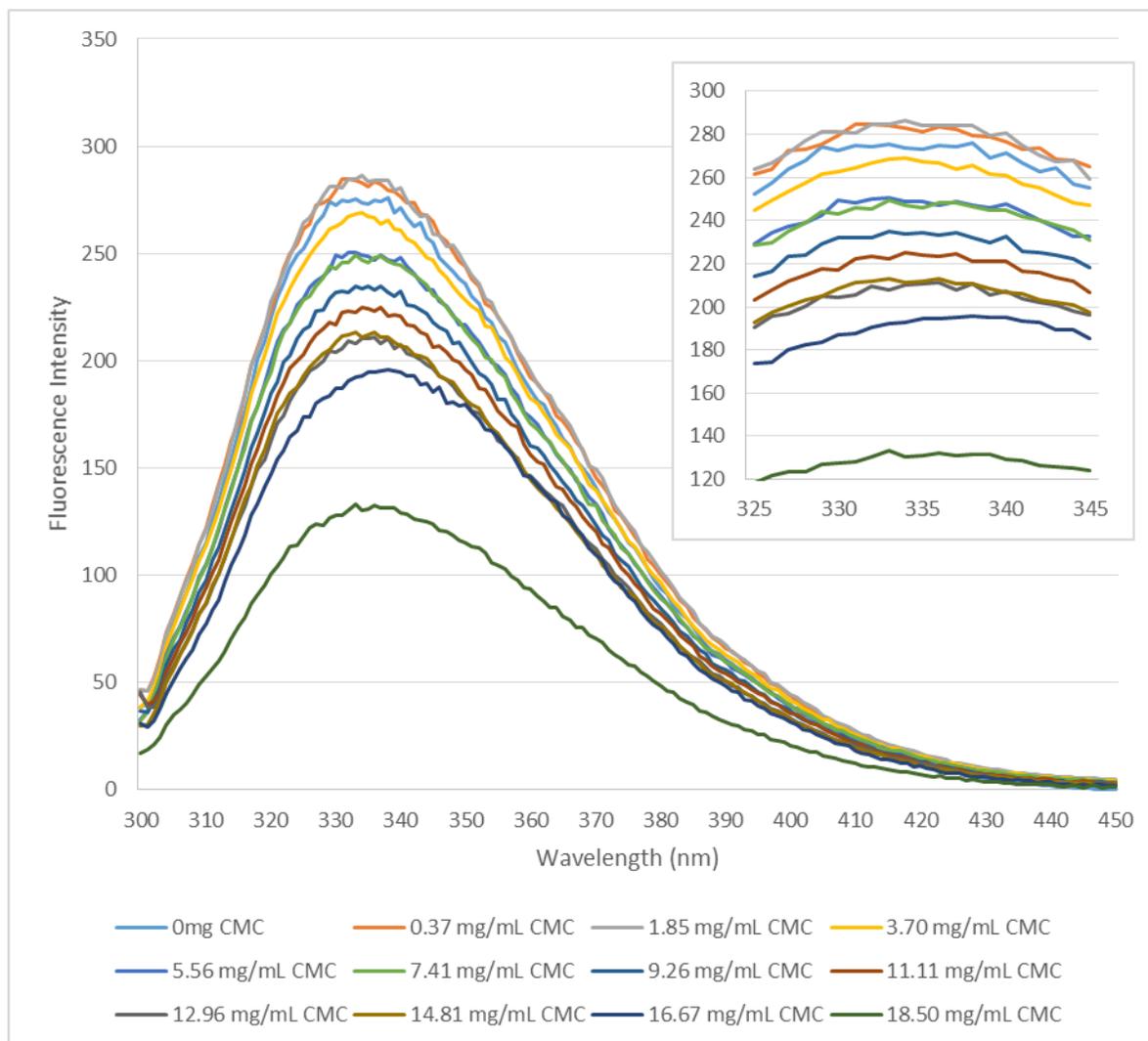
Figure 8.3: Fluorescence emission spectra of BcsC[24-781]. Increasing the concentration of CMC results in increasing quenching of tryptophan residues. Protein concentration at 3 μM of BcsC[24-781] was used.

The last construct to be assessed using fluorescence was BcsC[24-813]. This construct is a degraded form that makes up only ~40% of the entire construct. Using the emission peaks for each sample assessed at 330 nm, a decrease in emission of 43% was found between the 0 and 18.5 mg/mL samples. Having seen the quenching of the tryptophan residues in this degraded construct is indicative of an interaction between the TPR and CMC. Since this construct is composed of the first 37 kDa of the TPR domain of BcsC, it can furthermore be determined that the binding site of the TPR domain of BcsC is likely located at the N-terminal region. Furthermore, this construct portrayed a slight red-shift (about 2 nm) in the spectral profiles as the concentration of CMC increased. This indicates that the environment surrounding the tryptophan residues has changed, likely representing a change in the structure of the protein causing the environment around these tryptophan residues to become more aqueous (Lakowicz JR, 2010).
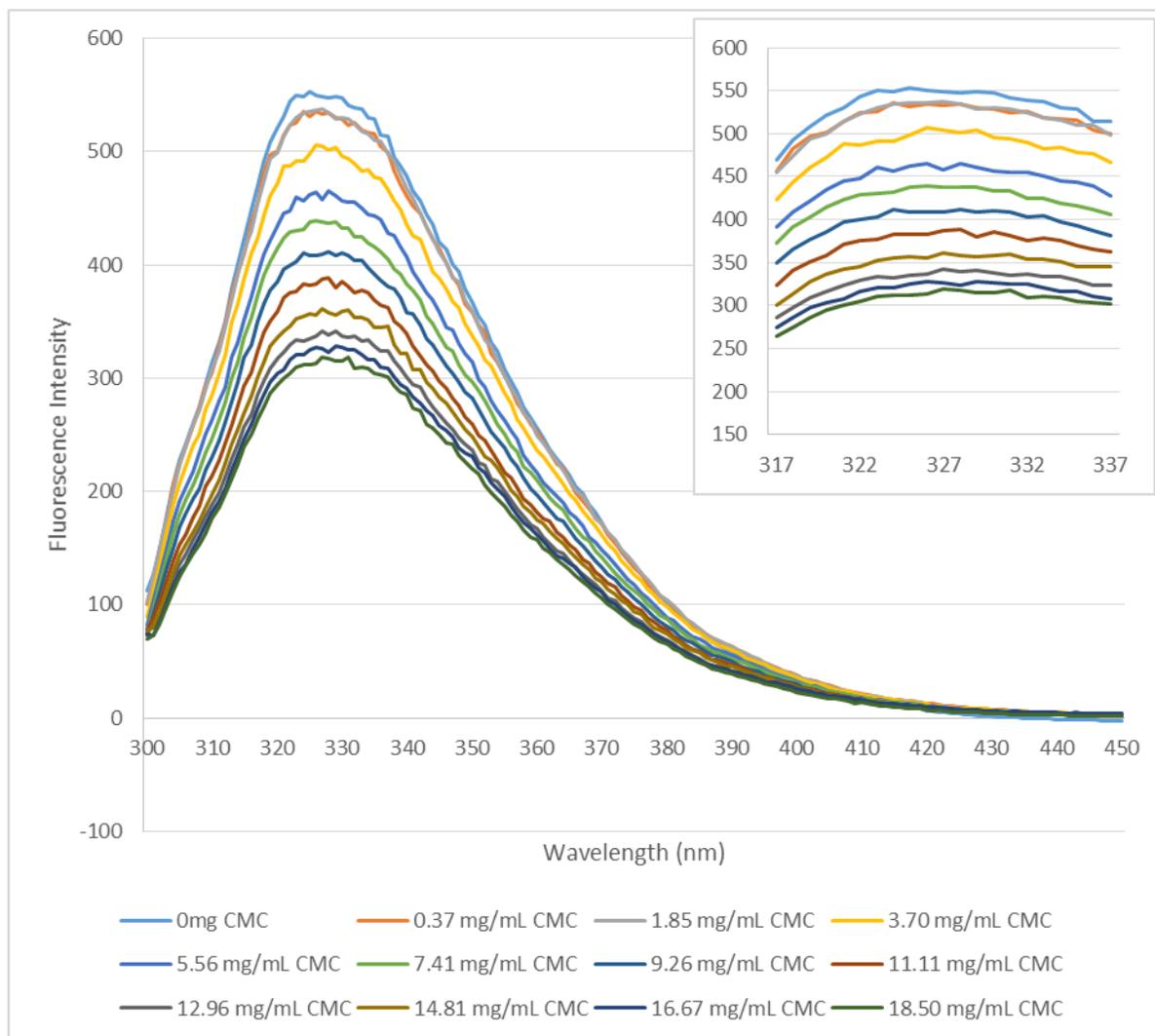
Figure 8.4: Fluorescence emission spectra of BcsC[24-813]. Increasing the concentration of CMC results in increasing quenching of tryptophan residues. Protein concentration at 2.5 μM of BcsC[24-813] was used.

All three fluorescence trials were normalized and collected in Figure 8.5 to compare the quenching of the TPR's. BcsC$^{24-813}$ appears to have the highest amount of quenching at -0.0238 compared to the other two. This is understandable since this degraded construct has only 4 tryptophan residues, all of which are in the predicted binding domain. Finally, BcsC$^{24-742}$ displayed the least amount of quenching at -0.022.

BcsC has been hypothesized to be involved in binding cellulose. This chapter has investigated the interaction of BcsC TPR constructs with both a soluble (CMC) and an insoluble (Avicel) form of cellulose. Using fluorescence spectroscopy, BcsC$^{24-742}$, BcsC$^{24-781}$ and BcsC$^{24-813}$ were investigated for tryptophan quenching using CMC. Figures 8.2-8.4 show the spectral profiles of each construct as the CMC concentration increases. Upon the addition of CMC, the fluorescence intensity decreased. This indicates that the CMC strands were interacting with the protein at or near tryptophan residues, thus blocking the availability for UV to excite the fluorophores of tryptophan residues at the binding site. Another reason for a decrease in fluorescence intensities would be due to potential conformational changes in the protein, thus, changing the environment surrounding the tryptophan residues. Using the peak at 335 nm (and 330 for BcsC$^{24-813}$), Stern-Volmer plots were created. It is important to note that this study is the first yet to assess protein-carbohydrate interactions between a TPR domain and its respective polysaccharide from a biofilm biosynthesis complex.
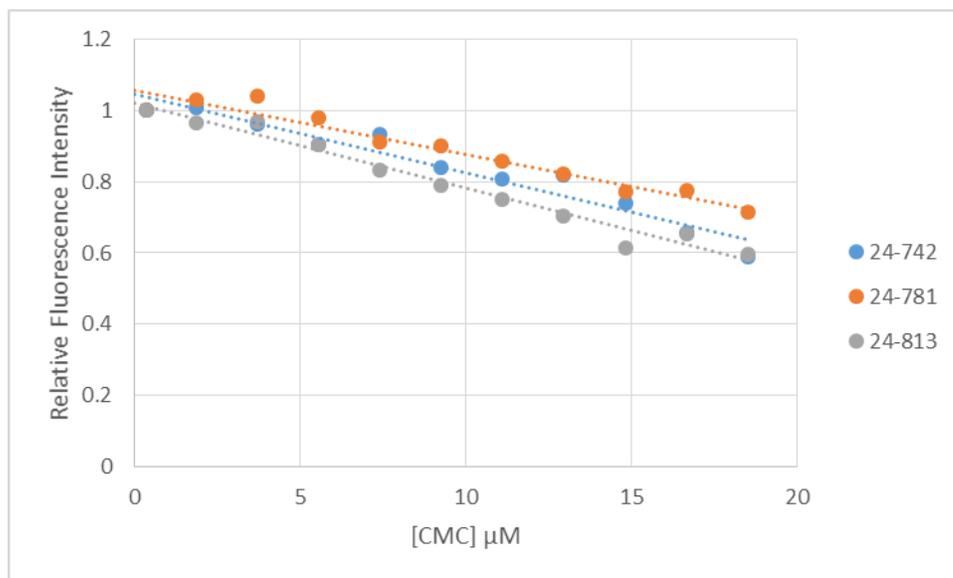
Figure 8.5: Relative fluorescence quenching of all three TPR constructs.   BcsC$^{24\text{-}742}$ (blue markers; y = -0.022x + 1.0463 R² = 0.9427), BcsC$^{24\text{-}781}$ (orange markers; y = -0.018x + 1.0582 R² = 0.9479), and BcsC$^{24\text{-}813}$ (grey markers; y = -0.0238x + 1.02 R² = 0.972).

Stern-Volmer plots created for each of the constructs indicated a linear relationship between F_0/F and the concentration of CMC which suggests a single binding sight (Acharya *et al*., 2013) on each of the TPR constructs analyzed. After confirming that only a single binding site is present, a binding constant, K_b, can be determined by plotting a modified Stern-Volmer plot, Log [(F0-F)/F] vs Log [Q]. From the linear relationship the intercept of the plot is equal to the Log K_b (Acharya, 2013), as seen in Figures 8.6 B, 8.7 B and 8.8 B. Therefore,

(4) $\text{Log } [(F_0\text{-}F)/F] = \text{Log } K_b + n \text{ Log } [Q]$

and

(5) $$K_b = 10^{\text{x-intercept}}$$

where n is the number of binding sites. The slope of the plot Log [(F0-F)/F] vs Log [Q] gives n (Acharya, 2013). Figures 8.6, 8.7 and 8.8 include Stern-Volmer plots each showing a linear relationship which can be indicative of a single binding site where CMC associates with BcsC. A list of the relative constants can be found in Table 8.1.

A Stern-Volmer plot was created for $BcsC^{24\text{-}742}$ to determine its association constant with CMC (Figure 8.6). The linear slope of the plot is indicative of a single binding site, and the K_a was determined form the slope to be 0.0032 μM^{-1} (Figure 8.6 A). In addition to the standard Stern-Volmer plot, a modified plot was created to establish the binding constant. The K_b was found to be 490.9μM^{-1}, using the x-intercept.

A   **B**
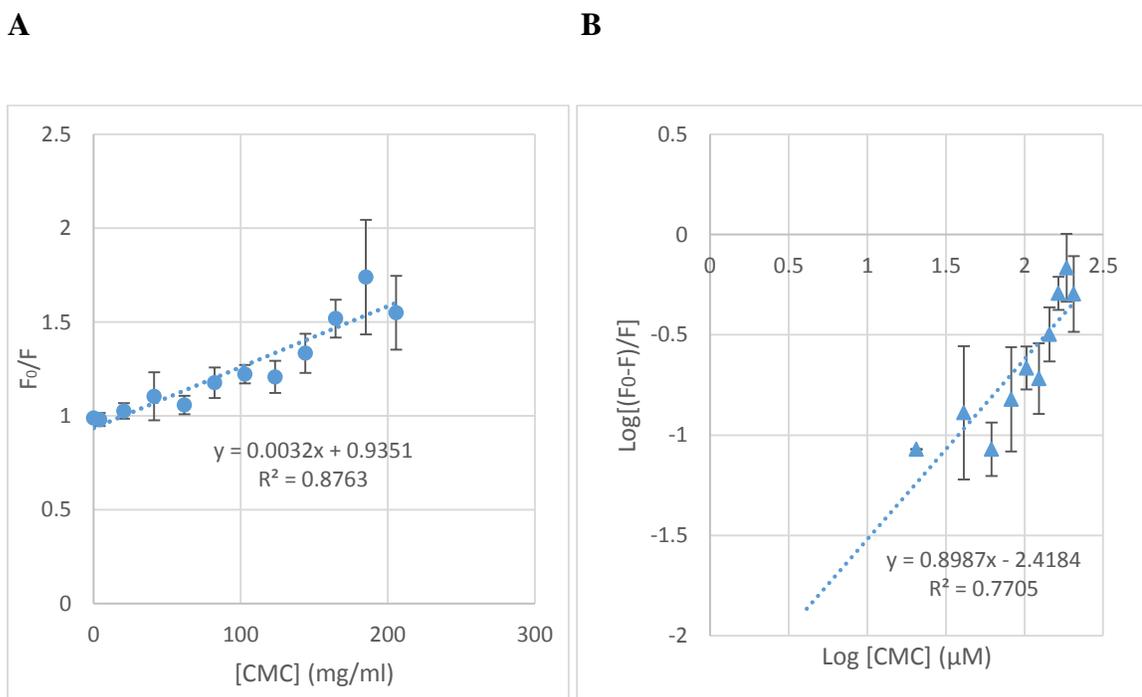
Figure 8.6: Stern-Volmer and binding constant plots of BcsC[24-742]. (A) The slope of the Stern-Volmer plot indicates a K_a of 0.0032 μM^{-1}. The linear relationship indicates a single binding site. (B) Using a modified Stern-Volmer plot, the Log K_b can be determined by solving for the x-intercept. The K_b for BcsC[24-742] is 490.9 μM^{-1}. Standard error bars were created based on values calculated for three trials.

An association constant with CMC was determined for BcsC[24-781] using the slope of the Stern-Volmer plot created (Figure 8.7 A). The slope of the plot indicated that the K_a of BcsC[24-781] is 0.002 μM^{-1}, and furthermore, confirmed a single binding site on the protein. The K_b was found to be 1033.6 μM^{-1} by using the x-intercept of the modified Stern-Volmer plot (Figure 8.7 B).

A

B
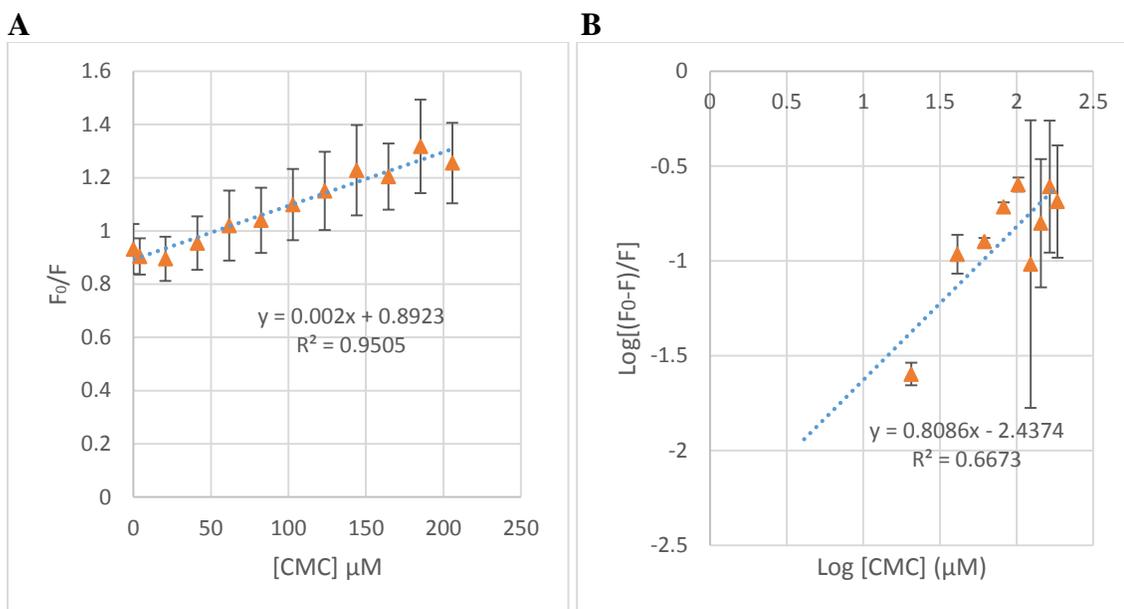
Figure 8.7: Stern-Volmer and binding constant plots of BcsC$^{24-781}$. (A) The slope of the Stern-Volmer plot indicates a K_a of 0.002 μM^{-1}. The linear relationship indicates a single binding site. (B) Using a modified Stern-Volmer plot, the Log K_b can be determined by solving for the x-intercept. The K_b for BcsC$^{24-781}$ is 1033.6 μM^{-1}. Standard error bars were created based on values calculated for three trials.

BcsC$^{24-813}$ was also assessed for association and binding constants. The slope of the Stern-Volmer plot for this construct indicated that BcsC$^{24-813}$ has a K_a of 0.0036 μM^{-1} with CMC and a single binding site (Figure 8.8 A). The binding constant, K_b, was found to be 573.1 μM^{-1} between BcsC$^{24-813}$ and CMC.

A

B
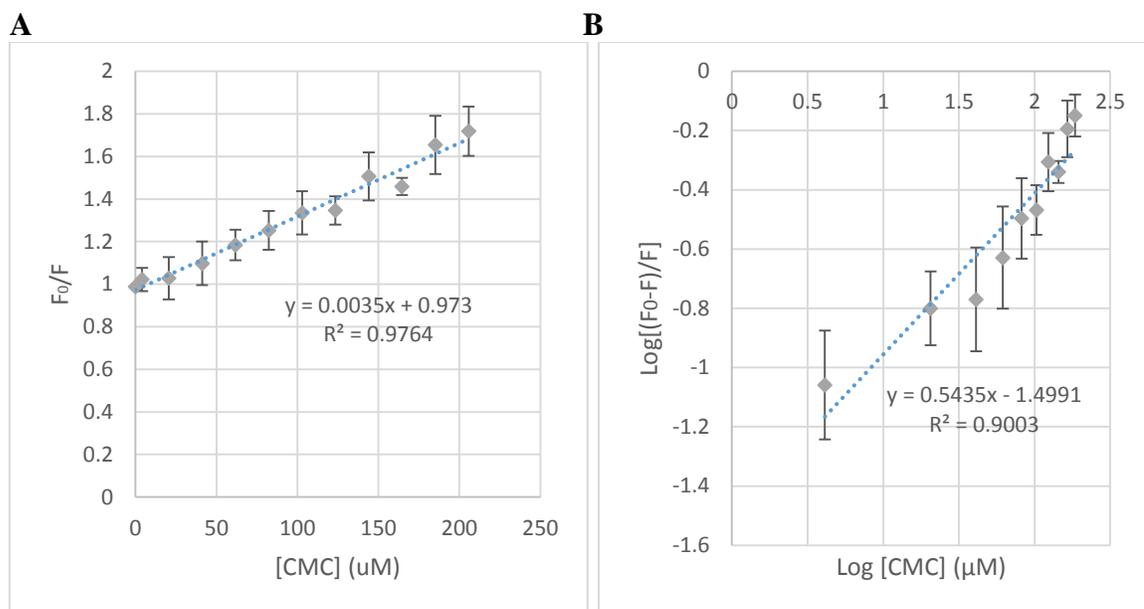
Figure 8.8: Stern-Volmer and binding constant plots of BcsC$^{24\text{-}813}$. The slope of the Stern-Volmer plot indicates a K_a of 0.0035 μM^{-1}. The linear relationship indicates a single binding site. (B) Using a modified Stern-Volmer plot, the Log K_b can be determined by solving for the x-intercept. The K_b for BcsC$^{24\text{-}813}$ is 573.1 μM^{-1}. Standard error bars were created based on values calculated for three trials.

Each construct tested indicated a single binding site, suggested by the linear relationship on the plot F_0/F vs [CMC] in Figures 8.6-8.8 A for each construct. Additionally, using the linear relationship the Stern-Volmer constant, K_{sv}, was found. It was determined that the K_{sv} could also be considered an association constant, K_a, since the system reached equilibrium. Likewise, a binding constant, K_b, was determined for each constant by plotting a modified Stern-Volmer, Log $[(F_0\text{-}F)/F]$ vs Log [Q] seen in Figures 8.6-8.8 B. Using the inverse of the association constant the disassociation constant was found for each construct to be 312.5 μM, 500 μM and 285.71 μM for BcsC$^{24\text{-}742}$, BcsC$^{24\text{-}781}$ and BcsC$^{24\text{-}813}$, respectively. The construct with the highest

129

association and binding constants was $BcsC^{24-813}$; as a result this construct also had the lowest disassociation constant. This is likely due to the fact that there were no additional tryptophan residues beyond those found in the hypothesized binding site of this degraded construct, meaning that the additional 5 tryptophan residues that found in $BcsC^{24-742}$ and $BcsC^{24-781}$ would have contributed to extra fluorescence if they were not located in the binding site. The next highest binding constant occurred in the $BcsC^{24-781}$ sample; however, $BcsC^{24-742}$ had a higher association constant. The K_d is a measure of the affinity by the protein while the K_b is a measure of the binding and unbinding of the protein at equilibrium, thus, a protein with high affinity to a ligand that binds well will have a lower K_d value and a higher K_b value. Based on the high K_d values of all three constructs, it can be determined that the TPR constructs did not display a high affinity to CMC. Moreover, when compared with other studies, the K_b values obtained in this research are quite low. Having determined that the TPR constructs have a low affinity and low binding to CMC it is determined that the TPR domain is likely involved in weak binding with cellulose. Table 8.1 outlines these various constants with their respective R^2 values as well as reports the calculated number of binding sites for each construct. Since the Stern-Volmer plots were linear, it was determined that there was a single binding site interacting with CMC. The n values calculated for each construct further support the presence of only one binding site.

In order for the TPR portion of BcsC to be able to bind cellulose and direct the chain to exit through the β-barrel domain, it must not bind the cellulose chain too tightly.

130

A strong binding between BcsC and the emerging cellulose chain would prevent export of cellulose from the cell via BcsC. This may explain the low association constants and high disassociation constants, K_d, where K_a and K_d maintain an inverse relationship. Additionally, when comparing the binding constants obtained in this experiment they are much lower than those found in other studies. For example, the association of resveratrol with sodium caseinate in a study performed by Acharya and colleagues (2013) used similar techniques to determine the binding constant of sodium caseinate with resveratrol at different temperatures, whereby, the binding constants ranged from 3.67 X 10^5 to 5.11 X 10^5 μM^{-1} indicating a highly favourable interaction (Acharya *et al.*, 2013). In comparison, the TPR constructs for BcsC range from 2.88 X 10^2 to 4.50 X 10^2 μM^{-1}, a dramatically lower value. Again, this helps to support the hypothesis that BcsC binds cellulose in a weak interaction.

In another study performed by Boraston *et al* (2000), the affinity of the CBM domain of Cellulase 5A from *Clostridium cellulovorans* was evaluated on a variety of carbohydrates including a variety of cellulose products such as cello trios, cellotetraose, cellopentataose and CMC (Boraston *et al.*, 2000). In the Cellulase 5A binding study, the researchers found that the binding domain for this protein showed high affinity to the various cellulose products (with K_a values from 1.2 X 10^3 to 4.3 X 10^3), however showed weak binding with CMC (Boraston *et al.*, 2000). Based these results, it may be worth analyzing the TPR constructs of BcsC with a variety of cellulose products to confirm weak binding.

Additionally, BSA was run as a control to determine that the BcsC TPR constructs were actually binding CMC. Figures A2.1 and A2.2 located in Appendix III indicate the quenching pattern and Stern-Volmer plots of the fluorescence of BSA with CMC. The quenching pattern shows an increase of emission from 0 mg CMC up until 25 mg CMC, after which emission spectra became lower than that of 0 mg CMC. When plotted on the Stern-Volmer, the plots are much more scattered than those produced for the various BcsC constructs indicating that the BSA failed to bind CMC. The use of the BSA as a control provided a means to assure protein-carbohydrate binding between the BcsC TPR constructs and CMC. This control was effective in reassuring an interaction between BcsC and CMC by displaying a lower confidence level ($R^2=0.68519$) due to plots distributed in a scattered form far from the trend line on the Stern-Volmer plot.

Table 8.1: Constants Obtained from Fluorescence Data using Stern-Volmer Plots

Construct	$K_S = K_a$ (μM^{-1})	R^2	K_d (μM)	K_b (μM^{-1}) (x-intercept)	R^2	n
24-742	0.0032	0.8763	312.5	490.9 (2.7)	0.7705	0.8987
24-781	0.002	0.9505	500	1033.6 (3.0)	0.6673	0.8086
24-813	0.0036	0.9765	285.7	573.1 (2.8)	0.9003	0.5435

8.2.2 Avicel Binding Assay

Upon the addition of higher concentrations of Avicel there was an increasing association with the TPR constructs in all three pH conditions tested, which can be seen in Figures 8.9, 8.10 and 8.11 for BcsC$^{24-741}$, BcsC$^{24-781}$ and BcsC$^{24-813}$, respectively. To gain numeric representation of the interactions, band intensities of the TPR constructs were collected using TotalLab Quant gel scanning software and normalized against the 50 mg Avicel samples for BcsC$^{24-741}$ and BcsC$^{24-781}$, and the 25 mg Avicel sample for BcsC$^{24-813}$. Normalized band intensity values have been graphically represented in Figures 8.9 (D), 8.10 (D) and 8.11 (D).

Avicel binding with BcsC$^{24-742}$ occurred the strongest in pH 7.0. Figure 8.9 depicts the binding patterns between the TPR construct and Avicel in pH 6.0, 7.0 and 8.0 for figure Panels A, B and C, respectively. An indication of stronger binding is seen when the normalized slope of the band intensities is low, indicating that pH 7.0 provided the best binding conditions for this construct with a slope of 0.002, seen in Panel D.
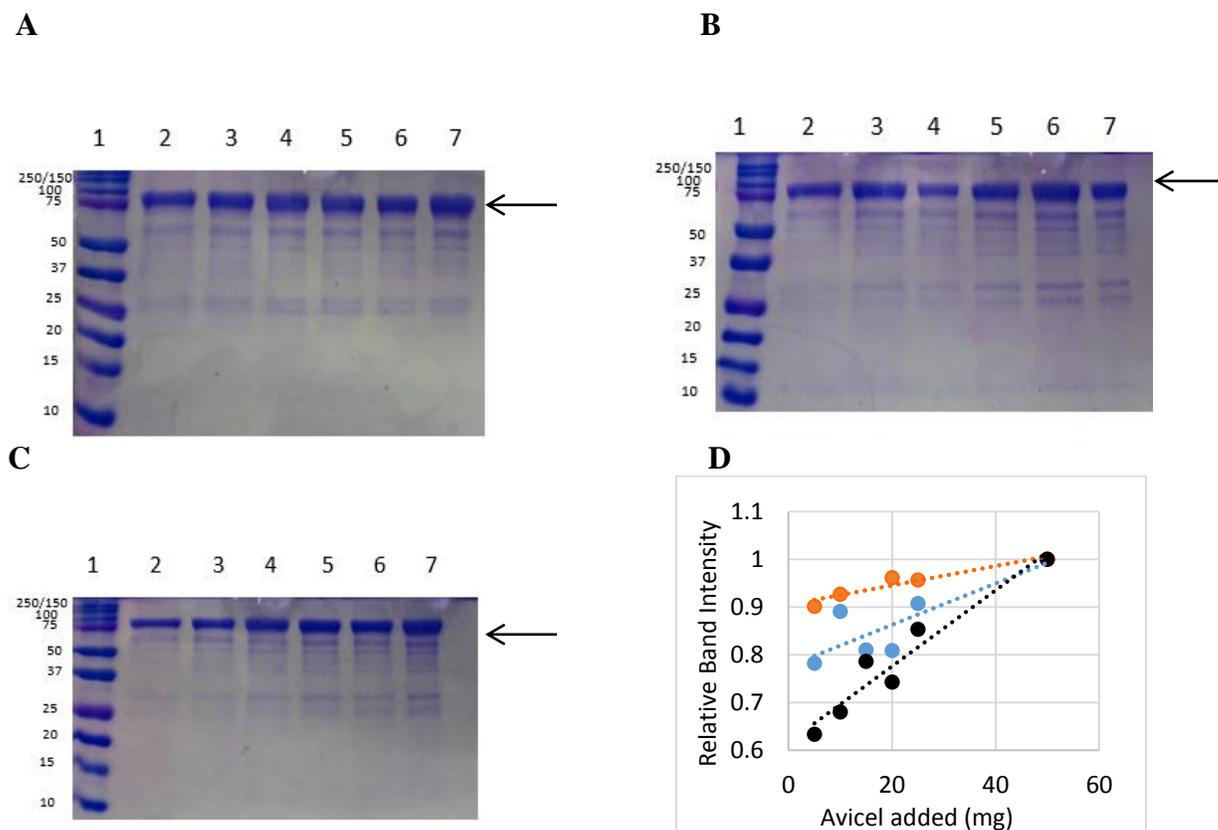
Figure 8.9: BcC$^{24\text{-}742}$ obtained from BcsC-Avicel binding complex. Buffers used in this assay varied by the pH, pH conditions tested were (A) pH 6.0, (B) pH 7.0, (C) pH 8.0. Panel (D) displays the normalized intensity for each pH tested: Blue represents pH6 (y = 0.0043x + 0.776 R² = 0.7094); Orange represents pH 7.0 (y = 0.002x + 0.9046 R² = 0.9117) and black represents pH 8.0 (y = 0.008x + 0.6161 R² = 0.9309). Based on the lowest slope value, pH 7.0 showed the highest binding to Avicel. Due to time constraints this assay was only performed once on this construct. Samples were loaded on the gels in the following order: lane 1) protein standard, lane 2) 5 mg Avicel, lane 3) 10 mg Avicel, lane 4) 15 mg Avicel, lane 5) 20 mg Avicel, lane 6) 25 mg Avicel, and lane 7) 50 mg Avicel. Black arrows are indicative of BcsC$^{24\text{-}742}$ that bound to the Avicel.

The differences in slopes for BcsC$^{24\text{-}781}$ are much less noticible than those seen in the BcsC$^{24\text{-}742}$ assay. Furthermore, the slopes for BcsC$^{24\text{-}781}$ are higher, indicating that BcsC$^{24\text{-}781}$ is less efficient in binding Avicel than BcsC$^{24\text{-}742}$. Of the three pH conditions tested pH 7.0 showed the highest level of binding, followed closely by pH 8.0 and pH 6.0

with slopes of 0.0125 and 0.0138, respectively (Figure 8.10). This experiment was only performed with one trial for each TPR construct in each pH tested. By repeating these trials, we may gain a better perspective on which conditions best support protein binding.
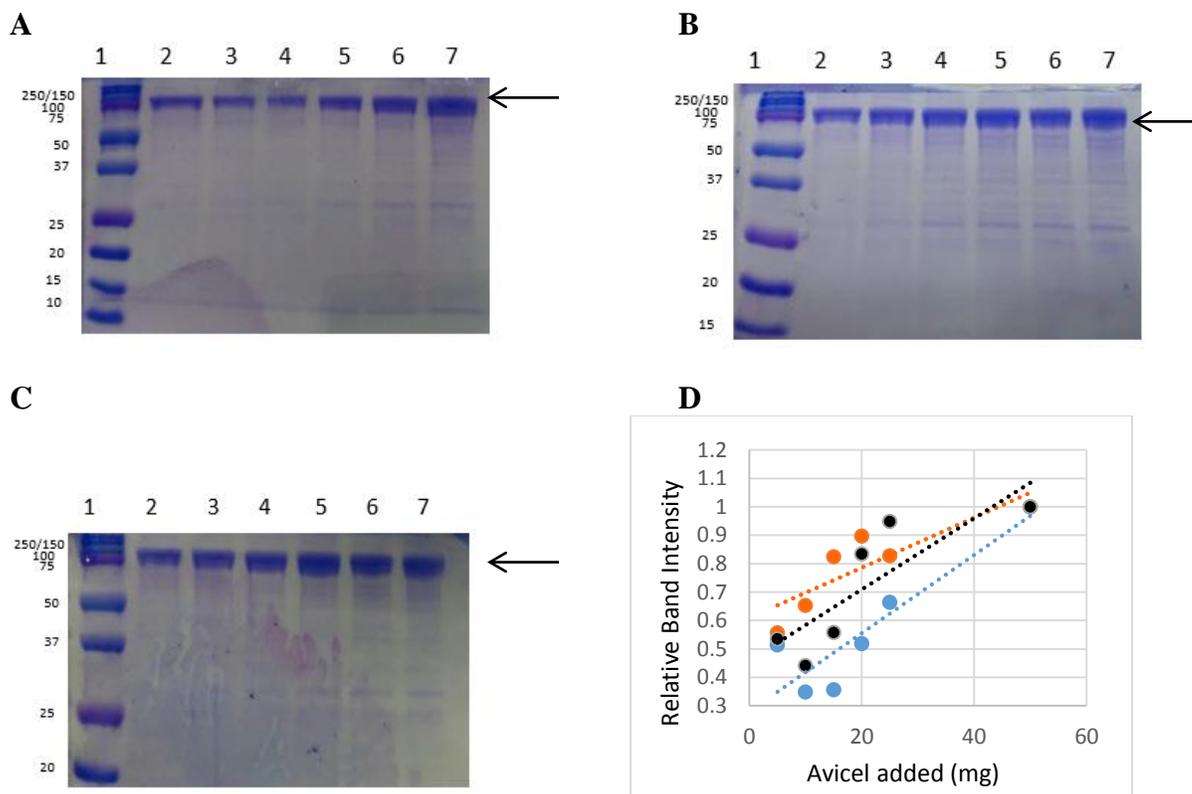
Figure 8.10: BcC$^{24-781}$ obtained from BcsC-Avicel binding complex. Buffers used in this assay varied by the pH, pH conditions tested were (A) pH 6.0, (B) pH 7.0, (C) pH 8.0. Panel (D) displays the normalized intensity for each pH tested: Blue represents pH6 (y = 0.0138x + 0.2798 R² = 0.8194); Orange represents pH 7 (y = 0.0088x + 0.6101 R² = 0.7491) and black represents pH 8.0 (y = 0.0125x + 0.4585 R² = 0.7079). Based on the lowest slope value, pH 7.0 showed the highest binding to Avicel. Due to time constraints this assay was only performed once on this construct. Samples were loaded on the gels in the following order: lane 1) protein standard, lane 2) 5 mg Avicel, lane 3) 10 mg Avicel, lane 4) 15 mg Avicel, lane 5) 20 mg Avicel, lane 6) 25 mg Avicel, and lane 7) 50 mg Avicel. Black arrows are indicative of BcsC$^{24-781}$ that bound to the Avicel.

Due to a shortage of protein, BcsC[24-813] was analyzed using samples up to 25 mg Avicel (Figure 8.11) as opposed to the 50 mg of Avicel used for the other two TRP constructs. Under the conditions tested, the condition that best provided an environment for this construct to bind Avicel was pH 6.0. The slope for this sample set was 0.0078. Additionally, this construct was the only one to portray a negative slope for one of its sample sets. This is due to the band of the 25 mg sample from the pH 8.0 buffer conditions displaying a much lower intensity in comparison to the other samples present. One explanation for this could be that the Avicel for this sample may not have been fully resuspended, trapping some of the protein released by heat among the insoluble Avicel chains, rather than releasing it to the sample buffer.
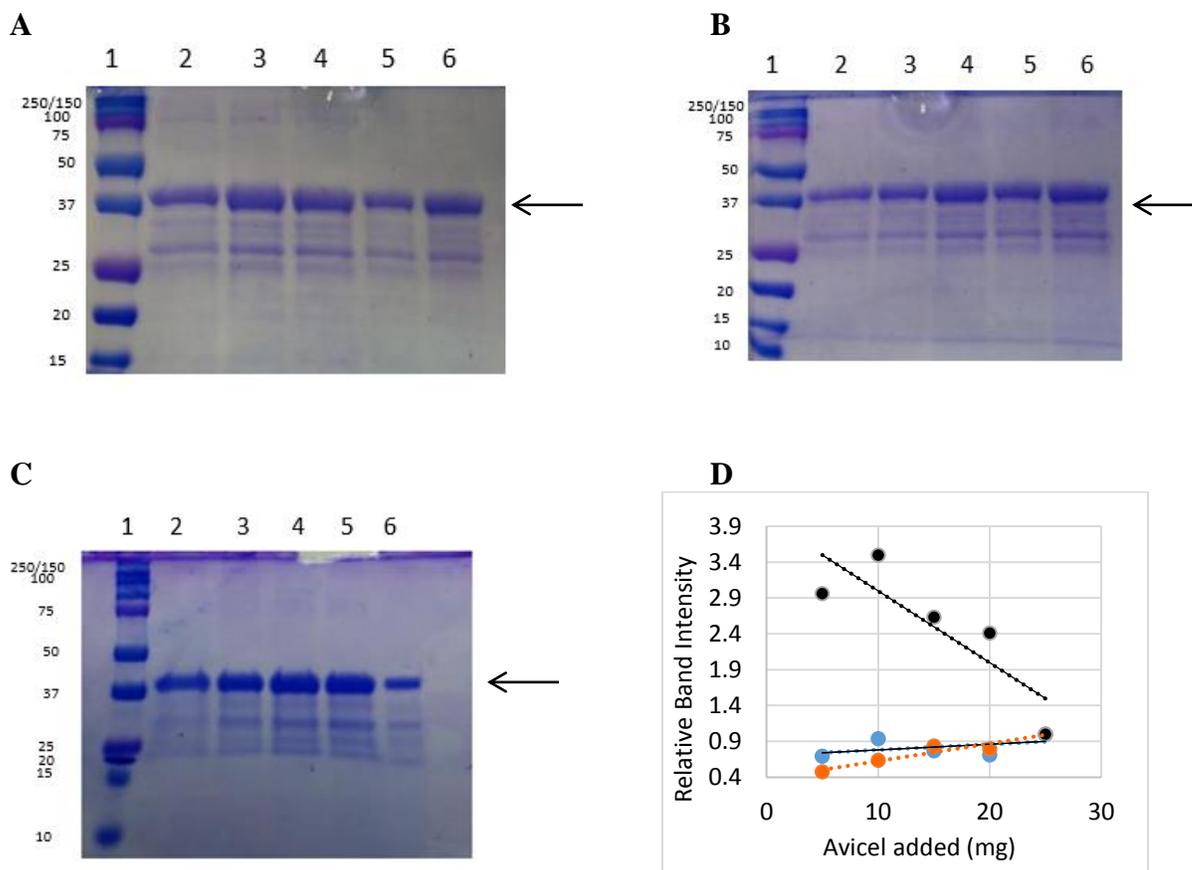
Figure 8.11: BcC$^{24\text{-}813}$ obtained from BcsC-Avicel binding complex. Buffers used in this assay varied by the pH, pH conditions tested were (A) pH 6.0, (B) pH 7.0, (C) pH 8.0. Panel (D) displays the normalized intensity for each pH tested: Blue represents pH 6.0 (y = 0.0078x + 0.7031 R² = 0.1996); Orange represents pH 7.0 (y = 0.0243x + 0.3836 R² = 0.9115) and black represents pH 8.0 (y = -0.1001x + 4.0002 R² = 0.7198). Based on the highest slope value, pH 6.0 showed the highest binding to Avicel. Due to time constraints this assay was only performed once on this construct. Samples were loaded on the gels in the following order: lane 1) protein standard, lane 2) 5 mg Avicel, lane 3) 10 mg Avicel, lane 4) 15 mg Avicel, lane 5) 20 mg Avicel and lane 6) 25 mg Avicel. Black arrows are indicative of BcsC$^{24\text{-}813}$ that bound to the Avicel.

Upon visual comparison of the gels, binding to Avicel was demonstrated at all

pHs tested for all of the three TPR constructs. While the difference is minor, binding

appeared to be best at pH 7.0 for BcsC[24-742], pH 7.0 for BcsC[24-781] and pH 6.0 for BcsC[24-813]. However, this experiment was only conducted once and further trials are needed to confirm this hypothesis. The slopes of the normalized band intensities across Avicel concentations also revealed similar trends amongst the three constructs. Most of the slopes were approximitely 0.1 in magnitude (with the exception of a few noted above). Additionally, future experiments should include a 0 mg Avicel sample to indicate the presence or absence of precipitated protein.

BSA was also used as a control to inspect the Avicel binding assay for non-specific binding. Since this assay was looking for weak binding of BcsC there were no wash steps performed as they could remove a fraction of the bound protein. When assayed with BSA, the Avicel samples appear to have bound this protein as well, however, the intensities were not as strong overall (especially at the low Avicel concentrations) as seen in the BcsC samples (Figure A4.3). The normalized plot of the BSA samples was also much more scattered in comparison to the BcsC samples, and obtained a very low R^2 value at 0.102. BSA is a helical protein responsible for binding fatty acids (Curry *et al*., 1999; Brown and Shockley, 1982), and is, therefore, not an entirely appropriate protein to use as a control. Nonetheless, BSA is a readily available, inexpensive protein was for these reasons was explored for use as a control in this study.

8.3 Conclusions

Through the use of fluorescence spectroscopy and Avicel binding assays the TPR domain of BcsC has been shown to be involved in protein-carbohydrate interactions. The two assays performed used a soluble (CMC) and an insoluble (Avicel) form of cellulose, showing evidence that BcsC can bind a variety of cellulose products. These interactions further support the hypothesis that the TPR domain of BcsC interacts with the emerging cellulose strands from the cellulose synthesis complex at the inner membrane, and facilitates the transport of the newly synthesized cellulose towards the outer membrane for export.

9. DISCUSSION

Biofilms can be composed of a wide variety of EPSs. In certain strains of *E. coli*, *Salmonella* spp. and other Enterobacteriaceae, cellulose may form the EPS surrounding the bacterial cells. Cellulose is produced by the BCS complex located at the cell membrane, this complex is primarily composed of BcsA and BcsB for cellulose synthesis, and BcsC which, as a result of homology searches, is believed to be responsible for the export of cellulose across the outer membrane. BcsC has been hypothesized to consist of an outer membrane β-barrel and a periplasmic TPR (Whitney and Howell, 2013). Bioinformatics results have confirmed the potential structure of BcsC to consist of an N-terminal TPR and a C-terminal β-barrel, as seen in other exopolysaccharide export proteins. Interestingly, however, the TPR and β-barrel in certain systems are divided into two separate polypeptides like AlgK and AlgE of the alginate system, and have remained one peptide in other systems such as the cellulose and PNAG systems (Whitney and Howell, 2013). AlgE and AlgK have had the most extensive research performed thus far and were a major contributor to the advancement of the research displayed in this study. For example, the 3D models produced for the β-barrel have shown a typical pore, however, with homology to AlgE, it may in fact also contain a blocking mechanism to prevent cellular leakage that was not evident in the 3D model structured from the crystal structure of esterase estA, which contains a central helix that prevents the pore from leakage that is lacking in BcsC (van den Berg, 2010). During expression and purification trials, it was found that the β-barrel containing

constructs were highly susceptible to degradation. This was evident by thick bands below the expected size of the protein constructs. One attempt to purify and refold BcsC[781-1141] was partially successful, however, the procedure still requires a large amount of optimization (*i.e.* collecting more elutions in purification, using longer incubation times for refolding, concentrating the protein to a lower amount to prevent precipitation, and adding glycerol to buffers to help stabilize the protein). Provided this construct can be effectively refolded, structural and functional analysis can be performed to characterize the β-barrel domain of BcsC. Crystallization should be the main focus in order to determine the process by which cellulose is exported from the cell, and investigating the potential control mechanism associated with preventing cellular leakage through the pore (similar to the P-gate and E-gate loops of AlgE) as well as determining key residues involved in the function of BcsC.

The TPR domain of the protein was modeled using the crystal structure of OGT (PDB:1w3b). OGT is a TPR protein that is conformationally flexible, for a predicted purpose of target protein recognition (Jinek *et al*., 2004). The flexibility of the TPR in OGT may be an indicator that the TPR domain of BcsC may also be flexible, allowing for the recognition of cellulose and/or other periplasmic proteins (*i.e*. BcsB). This may also support the idea of cellulose leading to conformational changes in the TPR domain, as was indicated in the CD analysis. As CMC concentration increased within the protein environment, the spectral profile displayed an upward shift and the helical content decreased. It is important to note, however, that cellulose absorbs in the far UV region,

which limited concentrations used in this study and may be a contributing factor to seeing a decrease in the helical content. This limitation also prevented the ability to analyse the interaction between the TPR constructs and CMC at higher concentrations, such as those used in the fluorescence assays (*i.e*. greater than 0.5% (w/v) CMC).

The TPR construct BcsC$^{24\text{-}813}$ showed high levels of degradation during protein expression. The protein was still used in the same assays and has provided some information on protein structure and function for the N-terminal region of the protein (since this is the portion of the degraded construct that would have been captured in the nickel column). This TPR construct may have still retained its structural integrity, as indicated by the highly α-helical spectra represented by minima at 208 nm and 222 nm collected using CD. Also, when analysed using fluorescence spectroscopy, BcsC$^{24\text{-}813}$ showed quenching with the addition of higher concentrations of CMC. This may be an indication that the binding site where the TPR domain interacts with cellulose is closer to the N-terminus (as this degraded construct is approximately 37 kDa in comparison to the full length of this construct which was ~91 kDa). Additionally, the fluorescence spectra portrayed an emission peak shift towards the higher wavelengths, indicating a red shift (Chaudhuri *et al*., 2010). This denotes an increase in exposure of the tryptophan residues to a more aqueous environment, likely due to a conformational change upon binding CMC. Together, the CD and fluorescence results indicate that this construct is still correctly folded well enough to maintain its activity. The active site localized to the N-terminus of the TPR paired with the physiologically relevant length of the TPR modeled

(Figure 5.3 Panel B) further supports the hypothesis that the TPR domain stretches through the periplasm to interact with the growing cellulose chain from the cellulose synthesis complex at the inner membrane.

Aside from the change in intensity in the CD spectra as the CMC concentration increased, all profiles were indicative of helical structures based on spectral minima at 208 nm and 222 nm. This result confirmed that the protein constructs remain in their folded state from the purification procedure. Additionally, knowing that the TPR constructs are properly folded provides a higher confidence for the functional assays performed. Moreover, in the CD assay where BcsC$^{24\text{-}742}$ was titrated with increasing amounts of CMC, the helical content of the protein decreased according to the deconvolution software. This is indicative of a potential conformational change in the protein, a characteristic often observed when proteins are activated by a substrate/ligand. Finally, using BcsC$^{24\text{-}781}$ it was determined that the TPR constructs maintain their secondary structure over time. This is useful as the more purification techniques are used, the longer the protein will be sitting before assayed in experiments. Using this information, we can be certain that our protein remains in its folded state long enough to perform our various assays.

One aspect that makes this research so unique is that it is the first time the export protein from biofilm production machinery has been tested to confirm protein-carbohydrate interactions with the polysaccharide. Fluorescence spectroscopy was used

to analyze the interaction between the TPR domain of the protein and cellulose (using CMC) and found that there is a weak interaction between the two. When paired with the CD results, it would appear as though the addition of CMC may in fact cause the protein to undergo a conformational change when the binding occurs. Again, when comparing this data with the 3D model structured from OGT which is a flexible TPR protein, it is possible that upon the binding of cellulose in the periplasm the protein may undergo a conformational change to transport the cellulose towards the outer membrane, causing the protein to become less stable, thus allowing it to release the cellulose chain and return to its resting state, repeating the process until there is no more cellulose remaining to transport. Additionally, the TPR constructs were examined for their ability to bind an insoluble form of cellulose, Avicel. The consistency of increasing band intensity and Avicel concentration is indicative of a protein-carbohydrate interaction. Therefore, the TPR domain of BcsC is capable of binding both soluble and insoluble forms of cellulose. Having shown to have the ability to bind these commercial brands composed of long, polymeric strands of cellulose, it is very feasible for the TPR to bind the single strands of cellulose emerging from the cellulose biosynthesis complex at the inner membrane.

10. FINAL CONCLUSIONS AND FUTURE DIRECTIONS

The research presented in this study focused on determining the structural and functional characteristics of BcsC. Initial work performed on BcsC involved the construction of three main construct groups: full-protein, TPR and β-barrel. TPR constructs proved to be the most successful from expression and purification trials and, therefore, became the main focus of this research. Upon further investigation with the TPR the secondary structure supported the hypothesis that the N-terminus of BcsC would be highly helical, to allow for the formation of the TPR structure. Crystallization trials to produce high quality protein crystals of BcsC for structure determination are as of yet unsuccessful, but are still ongoing.

Functionally, the TPR domain was predicted to be involved in protein-protein and/or protein-carbohydrate interactions. This study confirmed the ability of the TPR domain to interact with cellulose in both soluble and insoluble forms. Additionally, by continuing experiments using a degraded protein, it was confirmed that the interaction between the TPR and cellulose is in the N-terminal region of the TPR. It was also determined that as BcsC binds cellulose a conformational change occurs. Moreover, the TPR has been shown to have a T_m of approximately 51°C. It would be interesting to see if the conformational change identified upon cellulose binding causes a change in protein stability. This could be accomplished using the same techniques used in this study, however, with the addition of CMC. If the T_m is higher, it will indicate that CMC is stabilizing the protein, a feature often seen by proteins with their ligands.

145

There is still much to learn about BcsC and its roles in polysaccharide export. Firstly, it would be interesting to determine the ability of the TPR domain to interact with other periplasmic proteins. This could be confirmed using gel filtration techniques to observe binding between the two proteins, this has been shown to work using D-phosphoglycerate dehydrogenase and phosphoserine aminotransferase (Mishra *et al*., 2012). Secondly, the β-barrel domain has promising potential in refolding experiments. Once a folded β-barrel construct is obtained it can be subjected to crystallization trials in order to produce protein crystals suitable for structure determination. These results would be essential to confirming or eliminating the idea of loops used as a control mechanism to prevent cellular leakage, as seen with the homologous AlgE (Tan *et al*., 2014). If crystallography trials continue to fail for both domain constructs, small-angle x-ray scanning (SAXS) could be used to obtain an overall outline of the protein, which could be used to fit the models created into the structure. Next, both domains could be analysed for critical residues. Tryptophan residues in particular would be a convenient place to start as their participation in cellulose binding can be easily analyzed by repeating fluorescence spectroscopy as seen in this study. Finally, it would be interesting to determine if the TPR is responsible for directing the β-barrel to the outer membrane using the same protocol as was outlined for the alginate system. This was mainly done by deleting the gene for the TPR protein (AlgK) and observing for the localization of AlgE in either the inner and/or outer membrane. When AlgK was present, AlgE was only found in the outer membrane (Keiski *et al.,* 2010).

11. INTEGRATIVE NATURE TO THIS RESEARCH

This research integrated a wide range of fields including microbiology, molecular biology, biochemistry and bioinformatics. A basic understanding of microbiology and aseptic techniques were helpful in obtaining optimal expression of the various protein constructs by maximizing the bacterial growth of the BL21 *E. coli* cells used and minimizing contaminating bacteria. Once BcsC was expressed a variety of purification protocols used molecular techniques such as protein capture via his$_6$ tags. When the various BcsC constructs from columns were purified and ready to undergo the various assays, biochemical associations were being observed and analyzed for protein structure and function. Finally, the bioinformatics tools used to design the protein constructs based on the two domains of BcsC eliminated helped to provide a basis for exploration in construct development.

This research also has an effect with regards to the impact it can have. By studying BcsC and its structural and functional aspects a greater understanding can be made on polysaccharide export and the many export proteins involved in forming various different kinds of biofilms. This research also plays a role in the healthcare field. The more that is understood about biofilm formation, the better infections can be prevented and treated. Additionally, biofilms play an important role in the bioremediation industry. Bacteria living in biofilms are responsible for breaking down harmful pollutants. By understanding more about biofilms and how they are formed, bioremediation sites can be

treated with biofilm promoting factors, encouraging microbial communities to utilize the

harmful chemicals present.

REFERENCES

Abbot DW, Boraston AB. 2012. Quantitative approaches to the analysis of carbohydrate-binding module function. Methods in Enzymology. 510: 211-231

Acharya D, Sanguansri L, Augustin MA. 2013. Binding of resveratrol with sodium caseinate in aqueous solutions. Food Chemistry. 141: 1050-1054

Ackers ML, Mahon BE, Leahy E, Goode B, Damrow T, Hayes PS, Bibb WF, Rice DH, Barrett TJ, Hutwagner L, Griffin PM, Slutsker L. 1998. An outbreak of *Escherichia coli* O157:H7 infections associated with leaf lettuce consumption. The Journal of Infectious Disease. 177: 1588-1593

Andrade MA, Chacón P, Merelo JJ, Morán F. 1993. Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network. Protein Engineering. 6: 383-390.

Banerjee S, Joshi SR. 2013. Insights into cave architecture and the role of bacterial biofilm. Proceedings of the National Academy of Sciences, India Section B: Biological Sciences. 83: 277-290

Barnett Foster D. 2013. Modulation of the enterohemorrhagic *E. coli* virulence program through the human gastrointestinal tract. Virulence. 4: 315-323

Beattie BK, Merrill AR. 1999. A fluorescence investigation of the active site fo *Pseudomonas aeruginosa* exotoxin A. The Journal of Biological Chemistry. 274: 15646-15654

Biscola FT, Abe CM, Guth BEC. 2011. Determination of adhesion gene sequences in, and biofilm formation by, O157 and non-O157 shiga toxin-producing *Eschericia coli* strains isolated from different sources. Applies and Environmental Microbiology. 77: 2201-2208

Bodin A, Bäckdahl H, Fink H, Gustafsson L, Risberg B, Gatenholm P. 2007. Influence of cultivation conditions on mechanical and morphological properties of bacterial cellulose tubes. Biotechnology and Bioengineering. 97: 425-434

Bokransz W, Wang X, Tschäpe H, Römling U. 2005. Expression of cellulose and curli fimbriae by *Escherichia coli* isolated from the gastrointestinal tract. Journal of Medical Microbiology. 54: 1171-1182

Bolam DN, Xie H, White P, Simpson PJ, Hancock SM, Williamson MP, Gilbert HJ. 2001. Evidence for Synergy between Family 2b Carbohydrate Binding Modules in Cellulomonas fimi Xylanase. Biochemistry. 40: 2468-2477

Boraston AB, Chiu P, Warren RAJ, Kilburn DG. 2000. Specificity and affinity of substrate binding by a family 17 carbohydrate-binding module from *Clostridium cellulovorans* Cellulase 5A. Biochemistry. 39: 11129-11136

Boudeau J, Barnich N, Darfeuille-Michaud A. 2001. Type 1 pilimediated adherence of *Escherichia coli* strain LF82 isolated from Crohn's disease is involved in bacterial invasion of intestinal epithelial cells. Molecular Microbiology. 39: 1272-1284

Brenner T, Weadge JT. 2015. Structure-function characterization of BcsE; a protein involved in cellulose biosynthesis by *Escherichia coli*. Undergraduate thesis, Wilfrid Laurier University. Waterloo ON.

Brown, J. R. & Shockley, P. 1982. Lipid protein interactions P. C. Jost & O. H. Griffith (ed) Wiley, New York. Vol. 1. Pages 25–68

Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT. 2013. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Research. 41: W340-W348.

CDC. 2011. CDC estimates of foodborne illness in the United States. Accessed online June 23, 2014 from: http://www.cdc.gov/foodborneburden/2011-foodborne-estimates.html

Chaudhuri A, Haldar S, Chattopadhyay A. 2010. Organization and dynamics of tryptophans in the molten globule state of bovine α-lactalbumin utilizing wavelength-selective fluorescence approach: comparisons with native and denatured states. Biochemical and Biophysical Research Communications. 394:1082-1086

Cliff MJ, Williams MA, Brooke-Smith J, Barford D, Ladbury JE. 2005. Molecular recognition via coupled folding and binding in a TPR domain. Journal of Molecular Biology. 346: 717-732

Compton LA, Johnson WC Jr. 1986 Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. Analytical Biochemistry. 155: 155-167.

Conthia C, Lesk AM. 1986. The relation between the divergence of sequence and structure in proteins. The EMBO journal. 5: 823-826

Costerton JW, Stewart PS, Greenberg EP. 1999. Bacterial biofilms: A common cause of persistent infections. Science. 284: 1318-1322

Czaja W, Krystynowicz A, Bielecki S, Brown RM Jr. 2006. Microbial cellulose- The natural power to heal wounds. Biomaterials. 27: 145-151

Czaja WK, Young DJ, Kawecki M, Brown RM, Jr. 2007. The future prospects of microbial cellulose in biomedical applications. American Chemical Society. 8: 1-12

Das AK, Cohen PTW, Barford D. 1998. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. The EMBO Journal. 17: 1192-1199

D'Andrea LD, Regan L. 2003. TPR proteins: the versatile helix. TRENDS in Biochemical Sciences. 28: 655-662

Danhorn T and Fuqua C. 2007. Biofilm formation by plant-associated bacteria. Annual Review of Microbiology. 61: 401-422

Davies DG, Chakrabarty AM, Geesy GG. 1993. Exopolysaccharide production in biofilms: substratum activation of alginate gene expression by *Pseudomonas aeruginosa*. Applied and Environmental Microbiology. 59: 1181-1186

Davies DG, Geesey GG, 1995. Regulation of the alginate biosynthesis gene *algC* in *Pseudomonas aeruginosa* during biofilm development in continuous culture. Applied and Environmental Microbiology. 61: 860-867

de Cock H, Quaedvlieg N, Bosch D, Scholten M, Tommassen J. 1991. Glycine-144 is required for efficient folding of outer membrane protein PhoE of *Escherichia coli* K12. FEBS Letters. 279: 285-288

Delmer DP. 1999. Cellulose biosynthesis: Exciting times for a difficult field of study. Annual Review of Plant Physiology and Plant Molecular Biology. 50: 245-276

Fang X, Ahmad I, Blanka A, Schottkowski M, Cimdins A, Galperin MY, Römling U, Gomelsky M. 2014. GIL, a new c-di-GMP-binding protein domain involved in regulation of cellulose synthesis in enterobacteria. Molecular Microbiology. 93: 439-452

Flemming HC, Wingender J. 2010. The biofilm matrix. Nature Reviews Microbiology. 8: 623-633

Fontana JD, De Souza AM, Fontana CK, Torriani IL, Moreschi JC, Gallotti BJ, De Souza SJ, Narcisco GP, Bichara JA, Farah LFX. 1990. *Acetobacter* cellulose pellicle as a temporary skin substitute. Applied Biochemistry and Biotechnology. 24/25: 253-264

Gardy JL, Laird MR, Chen F, Rey S, Walsh CJ, Ester M, Brinkman FSL. 2005 PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. Bioinformatics 21:617-623

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005. Protein identification and analysis tools on the ExPASy server. John M. Walker (ed): The Proteomics Protocols Handbook, Humana press. Pages 571-607

Girón RM, Domingo D, Buendfa B, Antón E, Ruiz-Velasco LM, Ancochea J. 2005. Nontuberculosis mycobacteria in patients with cystic fibrosis. Archivos de Bronconeumología. 41: 560-565

González Barrios AF, Zuo R, Hashimoto Y, Yang L, Bentley WE, Wood TK. 2006. Autoinducer 2 controls biofilm formation in *Escherichia coli* through a novel motility quorum-sensing regulator (MqsR, B3022). Journal of Bacteriology. 188: 305-316

Greenfield NJ. 2006. Using circular dichroism spectra to estimate protein secondary structure. Nature Protocols. 1: 2876-2890

Grimson MJ, Haigler CH, Blanton RL. 1996. Cellulose microfibrils, cell motility, and plasma membrane protein organization change in parallel during culmination in *Dictyostelium discoideum*. Journal of Cell Science. 109: 3079-3087

Halbach F, Reichelt P, Rode M, Conti E. 2013. The yeast Ski complex: crystal structure and RNA channeling to the exosome complex. Cell. 154: 814-826

Hall V, Sklepari M, Rodger A. 2014. Protein secondary structure prediction from circular dichroism spectra using a self-organizing map with concentration correction. Chirality. 26: 471-482

Hampton Research. 2013. Seed bead user guide. Hampton Research Corp. Aliso Viejo, CA

Hampton Research. 2015. Izit crystal dye user guide. Hampton Research Corp. Aliso Viejo, CA

Healthy Canadians. 2013. *Risks of Eating Raw Sprouts*. Accessed online June 23, 2014 from http://www.healthycanadians.gc.ca/recall-alert-rappel-avis/hc-sc/2012/15041a-eng.php

Henrissat B, Deleury E, Coutinho PM. 2002. Glycogen metabolism loss: a common marker of parasitic behaviour in bacteria? TRENDS in Genetics. 18:437-440

Hobot JA, Carlemalm E, Villiger W, Kellenberger E. 1984. Periplasmic gel: new concept resulting from the reinvestigation of bacterial cell envelope ultrastructure by new methods. Journal of Bacteriology. 160: 143-152

Hrudey SE, Payment P, Huck PM, Gillham RW, Hrudey EJ. 2003. A fatal waterborne disease epidemic in Walkerton, Ontario: comparison with other waterborne outbreaks in the developed world. Water Science and Technology. 47:7-14

Hung C, Zhou Y, Pinker JS, Dodson KW, Crowley JR, Heuser J, Chapman MR, Hadjifranglskou M, Henderson JP, Hultgren SJ. 2013. *Escherichia coli* biofilms have an organized and complex extracellular matrix structure. mBio. 4: 1-10

Hura GL, Menon AL, Hammel M, Rambo RP, Poole FL II, Tsutakawa SE, Jenney FE, Jr, Classen S, Frankel KA, Hopkins RC, Yang S, Scott JW, Dillard BD, Adams MWW, Tainer JA. 2009. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nature Methods. 6: 606-612

Hu J, Ma L, Wang S, Yang J, Chang K, Hu X, Sun X, Chen R, Jiang M, Zhu J, Zhao Y. 2015. Biomolecular Interaction Analysis Using an Optical Surface Plasmon Resonance Biosensor: The Marquardt Algorithm vs Newton Iteration Algorithm. PLoS One. 10: 1-12

Itoh Y, Rice JD, Goller C, Pannuri A, Taylor J, Meisner J, Beveridge TJ, Preston JF III, Romeo T. 2008. Roles of *pgaABCD* genes in synthesis, modification and export of the *Escherichia coli* biofilm adhesin poly-β-1,6-*N*-acetyl-ᴅ-glucosamine. Journal of Bacteriology. 190:3670-3680

Janesn C, Heutink M, Tommassen J, de Cock H. The assembly pathway of outer membrane protein PhoE of *Escherichia coli*. European Journal of Biochemistry. 267: 3792-3800

Jernigan KK, Berdenstein SR. 2015. Tandem-repeat protein domains across the tree of life. PeerJ. 3: e732

Jinek M, Rehwinkel J, Lazarus BD, Izaurralde E, Hanover JA, Conti E. 2004. The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin α. Nature Structural and Molecular Biology. 11: 1001-1007

Jonas K, Tomenius H, Kader A, Normark S, Römling U, Belova LM, Melefors Ö. 2007. Roles of curli, cellulose and BapA in *Salmonella* biofilm morphology studied by atomic force microscopy. BMC Microbiology. 7: 70-79

Jones DT. 1999. Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology. 292: 195-202.

Jungbauer A, Kaar W. 2007. Current status of technical protein refolding. Journal of Biotechnology. 128: 587-596

Juncker AS, Willenbrock H, von Heijne G, Nielsen H, Brunak S, Krogh A. 2003. Prediction of lipoprotein signal peprides in Gram-negative bacteria. Protein Science. 12: 1652-1662

Keiski CL, Yip P, Robinson H, Burrows LL, Howel PL. 2007. Expression, purification crystallization and preliminary X-ray analysis of *Pseudomonas fluorescens* AlgK. Acta Crystallographica. 63: 415-418

Keiski CL, Harwich M, Jain S, Neculai AM, Yip P, Robinson H, Whitney JC, Riley L, Burrows LL, Ohman DE, Howell PL. 2010. AlgK is a TPR-containing protein and the periplasmic component of a novel exopolysaccharide secretin. Structure. 18: 265-273

Karpenahalli MR, Lupas AN, Söding, J. 2007. TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. BMC Bioinformatics. 8: 1-8

Kelley LA, Muzulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre[2]web portal for protein modeling, prediction and analysis. Nature Protocols. 10: 845-858

Kelly SM, Jess TJ, Price NC. 2005. How to study proteins by circular dichroism. Biochimica et Biophysica Acta. 1751: 119-139

Klemm D, Schumann D, Udhardt U, Marsch S. 2002. Bacterial synthesized cellulose - artificial blood vessels for microsurgery. Progress in Polymer Science. 26: 1561-1603

Kline KA, Fälker S, Dahlberg S, Normark S, Henriques-Normark B. 2009. Bacterial adhesins in host-microbe interactions. Cell Host & Microbe. 5: 580-592

Lakowicz JR. 2010. Chapter 16: Quenching of Fluorescence, *Principles of Fluorescence Spectroscopy* (3rd ed), (p.529-575). USA (NY): Springer.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. 23: 2947-2948

Leimbach A, Hacker J, Dobrindt U. 2013. *E. coli* as an all-rounder: the thin line between commensalism and pathogenicity. Current Topics in Microbiology and Immunology. 358:3-32

le Maire M, Champeil P, Møller JV. 2000. Interaction of membrane proteins and lipids with solubilizing detergents. Biochemica et Biophysica Acta. 1508: 86-111

Le Quéré B, Ghigo JM. 2009. BcsQ is an essential component of the *Escherichia coli* cellulose biosynthesis apparatus that localizes at the bacterial cell pole. Molecular Microbiology. 72: 724-740

Lewis K. 2001. Riddle of biofilm resistance. Antimicrobial Agents and Chemotherapy. 45: 999-1007

Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. 2003. Protein disorder prediction: implications for structural proteomics. Structure. 11: 1453-1459

Liu J, He D, Li X, Gao S, Wu H, Liu W, Gao X, Zhou T. 2010. γ-Polyglutamic acid (γ-PGA) produced by Bacillus amyloliquefaciens C06 promoting its colonization on fruit surface. International Journal of Food Microbiology. 142: 190-197

Liu J, Ma X, Wang Y, Liu F, Qiao J, Li X, Gao X, Zhou T. 2011. Depressed biofilm production in *Bacillus amyloliquefaciens* C06 causes γ-Polyglutamic Acid (γ-PGA) overproduction. Current Microbiology. 62: 235-241

Lobley A, Whitmore L, Wallace BA. 2002. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. Bioinformatics. 18: 211-212.

Macarisin D, Patel J, Bauchan G, Giron JA, Ravishankar S. 2013. Effect of spinach cultivar and bacterial adherence factors on survival of *Eschericaia coli* O157:H7 on spinach leaves. Journal of Food Protection. 76: 1829-1837

Machovič M, Svensson B, E. Ann MacGregor EA, Janeček Š. 2005. A new clan of CBM families based on bioinformatics of starch-binding domains from families CBM20 and

CBM21. FEBS Journal. 272: 5497-5513

Machovič M, Janeček Š. 2006. The evolution of putative starch-binding domains. FEBS Letters. 580: 6349 -6356

MacNeil D, Weadge JT. 2014. Purification and characterization of the β-barrel domain of BCsC to investigate its role in cellulose export from *Escherichia coli*. Undergraduate thesis, Wilfrid Laurier University. Waterloo ON.

Mah TFC, O'Toole GA. 2001. Mechanisms of biofilm resistance to antimicrobial agents. TRENDS in Microbiology. 9: 34-39

Manavalan P, Johnson WC Jr. 1987. Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. Analytical Biochemistry. 167: 76-85.

Märtson M, Viljanto J, Hurme T, Laippala P, Saukko P. 1999. Is cellulose sponge degradable or stable as implantation material? An in vivo subcutaneous study in the rat. Biomaterials. 20: 1989-1995

Mazur O, Zimmer J. 2011. Apo- and cellopentaose-bound structures of the bacterial cellulose synthase subunit BcsZ. The Journal of Biological Chemistry. 286: 17601-17606

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: 39-50

Merghi M, Lee VT, Hyodo M, Hayakawa Y, Lory S. 2007. The second messenger bis-(3-5)-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in *Pseudomonas aeruginosa*. Molecular Microbiology. 65: 876-895

Min Dong X, Gray DG. 1997. Induced circular dichroism of isotropic and magnetically-oriented chiral nematic suspensions of cellulose crystallites. American Chemical Society. 13: 3029-3034

Mishra V, Kumar A, Ali V, Nozaki T, Zhang KYJ, Bhakuni V. 2012. Novel protein-protein interactions between *Entamoeba histolytica* ᴅ-phosphoglycerate dehydrogenase and phosphoserine aminotransferase. Biochemie. 94: 1676-1686

Mizrachi E, Mansfield SD, Myburg AA. 2012. Cellulose factories: advancing bioenergy production from forest trees. New Phytologist. 194: 54-62

Moller S, Croning MDR, Apweiller R. 2001. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics. 17: 646-653

Morgan JLW, Strumillo J, Zimmer J. 2013. Crystallographic snapshot of cellulose synthesis and membrane translocation. Nature. 492: 181-187

Nakamura K, Misawa N, Kitamura K. Sequence of a cellulase gene of *Cellulomonas uda* CB4. Journal of Biotechnology. 4: 247-254

Nordahl Petersen T, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods. 8: 785-786

Omadjela O, Naraharia A, Strumillob J, Mélidac H, Mazurd O, Bulonec V, Zimmer J. 2013. BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis. PNAS. 110: 17856–17861

Petersen N, Gatenholm P. 2011. Bacterial cellulose-based materials and medical devices: current state and perspectives. Applied Microbiology and Biotechnology. 91:1277-1286

Pier GB, Coleman F, Grout M, Franklin M, Ohman DE. 2001. Role of alginate O acetlylation in resistance of mucoid *Pseudomonas aeruginosa* to opsonic phagocytosis. Infection and Immunity. 69:1895-1901

Pierce MM. Raman CS, Nall BT. Isothermal titration calorimetry of protein-protein interactions. Methods- A companion to methods in enzymology. 19: 213-221

Potera C. 1999. Forging a link between biofilms and disease. Science. 283: 1837-1839

Potera C. 1999. Forging a link between biofilms and disease. Science. 283: 1827-1839

Prakash B, Veerogowda BM, Krishnappa, G. 2003. Biofilms: a survival strategy of bacteria. Current Science. 85: 1299-1307

Pratt LA, Kolter R. Genetic analyses of bacterial biofilm formation. Current Opinion in Microbiology. 2: 598-603

Razvi E, Weadge JT. 2014. Structural and functional characterization of the role of BcsG in bacterial cellulose biosynthesis. Undergraduate thesis, Wilfrid Laurier University. Waterloo ON.

Rigaud JL Chami M, Lambert O, Levy D, Ranck JL. 2000. Use of detergents in two-dimensional crystallization of membrane proteins. Biochemicha et Biophysica Acta. 1508: 112-128

Römling, U. 2007. Chapter 7: Cellulose biosynthesis in enterobacteriaceae. Cellulose: molecular and structural biology. 107-122. 3300 Aa Dordrecht, Netherlands, Springer

Ross P, Mayer R, Benziman M. 1991. Cellulose biosynthesis and function in bacteria. Microbiological Reviews. 55: 35-58

Saldaña Z, Xicohtencatl-Cortes J, Avelino F, Phillips A, Kaper JB, Peunte JL, Girón JA. 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: 992-1006

Seery M. 2013. Saving paper. Education in Chemistry. 50: 22-25

Singh R, Paul D, Jain RK. 2006. Biofilms: implications in bioremediation. TRENDS in Microbiology. 14: 389-397

Smyth MS, Martin JHJ. 2000. X-ray crystallography. Journal of Clinical Pathology: Clinical Molecular Pathology. 53: 8-14

Solano C, García B, Valle J, Berasain C, Ghigo JM, Gamazo C, Lasa I. 2002. Genetic analysis of *Salmonella enteritidis* biofilm formation: critical role of cellulose. Molecular Microbiology. 43: 793-808

Sonia A, Priya Dasan K. 2013. Chemical, morphology and thermal evaluation of cellulose microfibers obtained from *Hibiscus sabdariffa*. Carbohydrate Polymers. 92: 668-674

Sreerema N, Woody RW. 1993. A self-consistent method for the analysis of protein secondary structure from circular dichroism. Analytical Biochemistry. 209: 32-44

Sreerema N, Venyaminov SY, Woody RW. 1999. Estimation of the number of helical and strand segments in proteins using CD spectroscopy. Protein Science. 8: 370-380.

Sreerama N, Woody RW. 2000. Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set. Analytical Biochemistry. 287: 252-260

Tan J, Rouse SL, Li D, Pye VE, Vogeley L, Brinth AR, El Arnaout T, Whitney JC, Howell PL, Sansom MSP, Caffrey M. 2014. A conformational landscape for alginate secretion across the outer membrane of *Pseudomonas aeruginosa*. Acta Crystallographica. D70: 2054-2068

Taormina, PJ, Beuchat LR, Slutsker L. Infections Associated with Eating Seed Sprouts: An International Concern. Emerging Infectious Diseases. 5: 626-634

The PyMOL Molecular Graphics System, Version 1.5 Schrödinger, LLC. 2011

Timmins PA, Leonhard M, Weltzien HU, Wacker T, Welte W. 1988. A physical characterization of some detergents of potential use for membrane protein crystallization. FEBS Letters. 238: 361-368

Qiagen. 2003. The QIAexpressionist: A handbook for hig-level expression and purification of 6xHis-tagged proteins 5th edition. Accessed from www.qiagen .com

Trepanier N, Weadge JT. 2015. Refolding optimization of the BcsC β-barrel domain: researching the export of cellulose in *Escherichia coli*. Undergraduate thesis, Wilfrid Laurier University. Waterloo ON.

Tsumoto K, Ejima D, Kumagai I, Arakawa T. 2003. Practical considerations in refolding proteins from inclusion bodies. Protein Expression and Purification. 28: 1-8

Vagenede V, Yap MGS, Trout BL. 2009. Mechanisms of Protein Stabilization and Prevention of Protein Aggregation by Glycerol. Biochemistry. 48: 11084-11096

van den Berg B. 2010. Crystal structure of a full-length autotransporter. Journal of Molecular Biology. 396: 627-633

Walter S, Schrempf H. 2003. Oligomerization, membrane anchoring, and cellulose-binding characteristics of AbpS, a receptor-like Streptomyces protein. Journal of Biological Chemistry. 278: 26639-26647

White D. 2007. Microbial development and physiological adaptation: Varied responses to environmental cues and intercellular signals. The Physiology and Biochemistry of Prokaryotes 3rd Edition. 467-598. New York, Oxford University Press

Whitmore L, Wallace BA. 2004. DICHROWEB: an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Research 32:W668-673

Whitmore L, Wallace BA. 2008. Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and Reference Databases. Biopolymers 89: 392-400.

Whitney JCC, Neculai AM, Ohman DE, Howell PL. 2009. Expression, refolding, crystallization and preliminary X-ray analysis of *Pseudomonas aeruginosa* AlgE. Acta Crystallographica. 65: 463-466

Whitney JC, Howell PL. 2013. Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. Trends in Microbiology. 21: 63-72

Zogaj X, Nimtz M, Rohde M, Bokranz W, Römling 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

APPENDIX I

Cloning:

BcsC constructs were designed using the bioinformatics preliminary research discussed above and generated using *E. coli* BL21 DNA as a template. Polymerase chain reaction (PCR) amplification and ligation of the specific products into the pET28a expression vector was performed by other members of the lab using primers listed in Table A1.1. The constructs generated are summarized in Table 4.1.

Table A1.1: Primers for the Amplification of BcsC Constructs

Primer Site *	Primer	Restriction Enzyme Site underlined in Primer
24-	5´-GCCA TATGGCACCAACCGCTCAGCAACAG-3´	NdeI
29-	5´-GCACCTACCCA TATGCAACAGTTGCTG-3´	NdeI
-742	5´- GCA AGCTTACCGGATGCGACCATGGCG -3´	HindIII
-781	5´- GCA AGCTTCTGGCGATAGAGGTCCGCC -3´	HindIII
783-	5´-CGCA TATGGATCTTAACGTCACCCTTGAGCACG-3´	NdeI
800-	5´- GCCA TATGGGTTACTCCGATCTGAAAGCGCAC-3´	NdeI
-813	5´-GCCCA AGCTTTTACGGCGCATCC-3´	HindIII
816-	5´-GCCTCA TATGGACGGGCGGATG-3´	NdeI
-1141	5´- GCA AGCTTCTGCCATCCGGCGG -3´	HindIII
-1154	5´-GCGCA AGCTTTTACCAGTCGGCGTAAG-3´	HindIII

Note: Restriction enzyme site underlined and cut site indicated by a space
*Numbers denote the amino acid that initiates (e.g., 24-) or terminates (e.g., -742) the primer.

Repeat Begin	Alignment	End	P-value
TPR 28	QQQLLEQVRLGEATHREDLVQQSLYRLELIDPNN	61	6.6e-02
TPR 62	PDVVAARFRSLLRQGDIDGAQKQLDRLSQLAPSS	95	4.1e-07
TPR 105	MLLSTPDGRQALQQARLQATTGHAEEAVASYNKL	138	1.7e-01
TPR 147	DIAVEYWSTVAKIPARRGEAINQLKRINADTPGN	180	3.9e-02
TPR 181	TGLQNNLALLLFSSDRRDEGFAVLEQMAKSNAGR	214	9.9e-05
TPR 269	AFRARAQGLAAVDSGMAGKAIPELQQAVRANPKD	302	4.8e-06
TPR 303	SEALGALGQAYSQKGDRANAVANLEKALALDPHS	336	2.3e-11
TPR 351	YWLAIQQGDAALKANNPDRAERLFQQARNVDNTD	384	1.9e-06
TPR 385	SYAVLGLGDVAMARKDYPAAERYYQQTLRMDSGN	418	3.5e-09
TPR 419	TNAVRGLANIYRQQSPEKAEAFIASLSASQRRSI	452	1.2e-01
TPR 461	NDRLAQQAEALENQGKWAQAAALQRQRLALDPGS	494	6.8e-06
TPR 495	VWITYRLSQDLWQAGQRSQADTLMRNLAQQKPND	528	4.1e-05
TPR 529	PEQVYAYGLYLSGHDQDRAALAHINSLPRAQWNS	562	2.7e-02
TPR 567	LVNRLQSDQVLETANRLRESGKEAEAEAMLRQQP	600	4.4e-01
TPR 603	TRIDLTLADWAQQRRDYTAARAAYQNVLTREPAN	636	4.2e-06
TPR 637	ADAILGLTEVDIAAGDKAAARSQLAKLPATDNAS	670	6.5e-03
TPR 671	LNTQRRVALAQAQLGDTAAAQRTFNKLIPQAKSQ	704	1.1e-04
TPR 711	AMVLRDGAKFEAQAGDPTQALETYKDAMVASGVT	744	2.7e-04

Figure A1.1: TPR sequence alignment as predicted by TPRpred. Probability for TPR structure is 100%.

```
Conf: ]████▄▄▄▄████████████▄▄▄▄▄▄▄▄▄██▄▄███████████[
Pred:   ───────◖──────────◖──────────◖──────────────
Pred: CCCCCHHHHHHHHHHHHCHHHHCCCCHHHHHHHHHHHHHH
  AA: MRKFTLNIFTLSLGLTVMPMVEAAPTAQQQLLEQVRLGEA
              10        20        30        40

Conf: ]▄█████████████▄██████████████████████▄▄██[
Pred:   ───◖──────────◖─────────◖──────────────────
Pred: CCCHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHCCCHHH
  AA: THREDLVQQSLYRLELIDPNNPDVVAARFRSLLRQGDIDG
              50        60        70        80

Conf: ]██████████▄███████████████▄▄▄▄███████████[
Pred:   ─────────────◖─────────────◖───────────────
Pred: HHHHHHHHHHHHCCCCHHHHHHHHHHHHHHCCCCHHHHHHHH
  AA: AQKQLDRLSQLAPSSNAYKSSRTTMLLSTPDGRQALQQAR
              90       100       110       120

Conf: ]███████▄███████████▄▄██████▄▄███████████▄▄█[
Pred:   ─────◖───────────────◖──────────────◖──────
Pred: HHHHCCCHHHHHHHHHHHHHHCCCCCCHHHHHHHHHHHHHHHCC
  AA: LQATTGHAEEAVASYNKLFNGAPPEGDIAVEYWSTVAKIP
             130       140       150       160

Conf: ]████████████████▄██████████████▄▄██████▄▄█[
Pred:   ─◖──────────────◖─────────────◖────────────
Pred: CCHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHCCCCCHHH
  AA: ARRGEAINQLKRINADTPGNTGLQNNLALLLFSSDRRDEG
             170       180       190       200

Conf: ]██████████▄███▄▄██████████████▄▄█████████████[
Pred:   ─────────────◖──────────────◖──────────────
Pred: HHHHHHHHHHCCCCCHHHHHHHHHHHHHHHCCCCCHHHHHHHHHH
  AA: FAVLEQMAKSNAGREGASKIWYGQIKDMPVSDASVSALKK
             210       220       230       240

Conf: ]██████▄▄▄█████████████████▄▄███████████████[
Pred:   ───────◖──────────────◖────────────────────
Pred: HHHHCCCCHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHH
  AA: YLSIFSDGDSVAAAQSQLAEQQKQLADPAFRARAQGLAAV
             250       260       270       280
```

```
Conf: ]▐▋▋▐▐▋▋▋▋▋▋▋▋▋▋▐▋▋▐▋▋▋▋▋▋▋▋▋▋▋▋▐▋▋▐▋▋▐[
Pred:    ─◗        ◖▬▬▬▬▬▬▬▬◗      ◖▬▬▬▬▬▬◗      ◖◗─
Pred: HCCCHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHCCCHH
  AA: DSGMAGKAIPELQQAVRANPKDSEALGALGQAYSQKGDRA
         ·         ·         ·         ·
        290       300       310       320

Conf: ]▋▋▋▋▋▋▋▋▋▐▋▋▋▋▋▋▐▋▋▋▋▋▐▐▋▋▐▋▋▋▋▋▋▋[
Pred:    ◖▬▬▬▬▬▬▬▬▬▬◗     ◖▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬◗
Pred: HHHHHHHHHHHCCCCCCCCHHHHHHHHHHHHHHHHHHHHHH
  AA: NAVANLEKALALDPHSSNNDKWNSLLKVNRYWLAIQQGDA
         ·         ·         ·         ·
        330       340       350       360

Conf: ]▋▋▋▐▋▋▋▋▋▋▋▋▋▋▋▐▋▋▐▋▋▋▋▋▋▋▋▋▋▋▐[
Pred: ◖▬▬◗  ◖◗  ◖▬▬▬▬▬▬▬▬▬◗   ◖▬▬▬▬▬▬▬▬◗  ─
Pred: HHHCCCHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHCCC
  AA: ALKANNPDRAERLFQQARNVDNTDSYAVLGLGDVAMARKD
         ·         ·         ·         ·
        370       380       390       400

Conf: ]▋▋▋▋▋▋▋▋▋▋▋▋▐▋▋▋▋▋▋▋▋▋▋▋▋▋▐▋▋▋▋▋▋▋[
Pred: ◖◗▬▬▬▬▬▬▬◗  ◖◗▬▬▬▬▬▬▬▬▬▬▬◗◖◗▬▬▬▬▬
Pred: CHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHCCHHHHHH
  AA: YPAAERYYQQTLRMDSGNTNAVRGLANIYRQQSPEKAEAF
         ·         ·         ·         ·
        410       420       430       440

Conf: ]▋▋▋▋▋▐▋▋▋▋▋▋▋▐▋▋▋▋▋▋▋▋▋▋▋▋▋▋▐[
Pred: ◖▬▬▬▬▬◗ ◖◗▬▬▬▬▬▬▬▬▬▬▬▬▬▬◗ ◖◗─
Pred: HHHHHHHHHCCCHHHHHHHHHHHHHHHHHHHHCCCHHHHH
  AA: IASLSASQRRSIDDIERSLQNDRLAQQAEALENQGKWAQA
         ·         ·         ·         ·
        450       460       470       480

Conf: ]▋▋▋▋▋▋▋▋▐▋▋▋▋▋▋▋▋▋▋▋▋▐▋▋▋▋▋▋▋▋▋▋[
Pred: ◖▬▬▬▬▬▬◗ ◖◗▬▬▬▬▬▬▬▬▬▬◗ ◖◗▬▬▬▬▬▬▬
Pred: HHHHHHHHHHCCCCCHHHHHHHHHHHHHHHCCCHHHHHHHHHH
  AA: AALQRQRLALDPGSVWITYRLSQDLWQAGQRSQADTLMRN
         ·         ·         ·         ·
        490       500       510       520

Conf: ]▋▋▋▐▋▋▋▋▋▋▋▋▋▋▋▐▋▋▋▋▋▋▋▐▐▋▐▋▋▋▋[
Pred: ◖◗   ◖◗▬▬▬▬▬▬▬◗  ◖◗▬▬▬▬◗─
Pred: HHHCCCCCHHHHHHHHHHHCCCCHHHHHHHHHHCCCCCCC
  AA: LAQQKPNDPEQVYAYGLYLSGHDQDRAALAHINSLPRAQW
         ·         ·         ·         ·
        530       540       550       560
```

Figure A1.2: PSIPRED secondary structural prediction of the full length BcsC polypeptide. Pink cylinders represent α-helices, yellow arrows represent β-strands, straight lines represent coils and the intensity of the blue bands above the sequences represent the confidence level of the prediction (Buchan *et al*., 2013; Jones, 1999).

Figure A1.3: Overall structure of esterase estA. Esterase estA consists of an outer membrane β-barrel with a helical structure running through the central pore. The arrow indicates the connection between the β-barrel and the α- helix.

Cloning

All cloning and ligation into pET28a for all constructs generated were performed

by Jesika Strinjaric and Christina Notte. Figure A1.1 displays ligation of various BcsC

constructs into the pET28a plasmid.

A

B
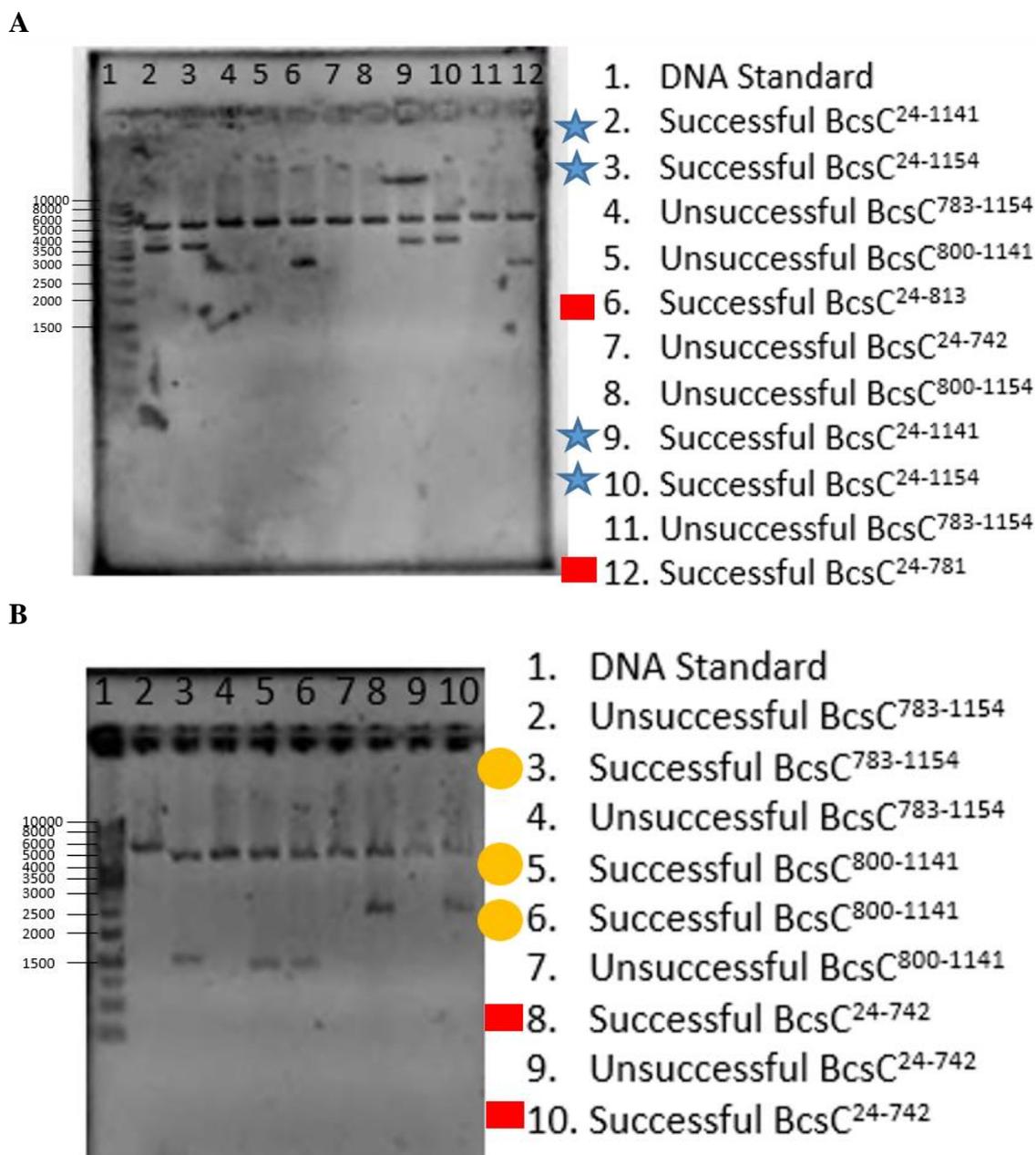
Figure A1.4: Ligation of various BcsC constructs into the pET28a expression vector. Red squares indicate successful ligation of TPR constructs, yellow circles indicate successful ligation of β-barrel constructs and blue stars indicate successful ligation of full constructs.

APPENDIX II
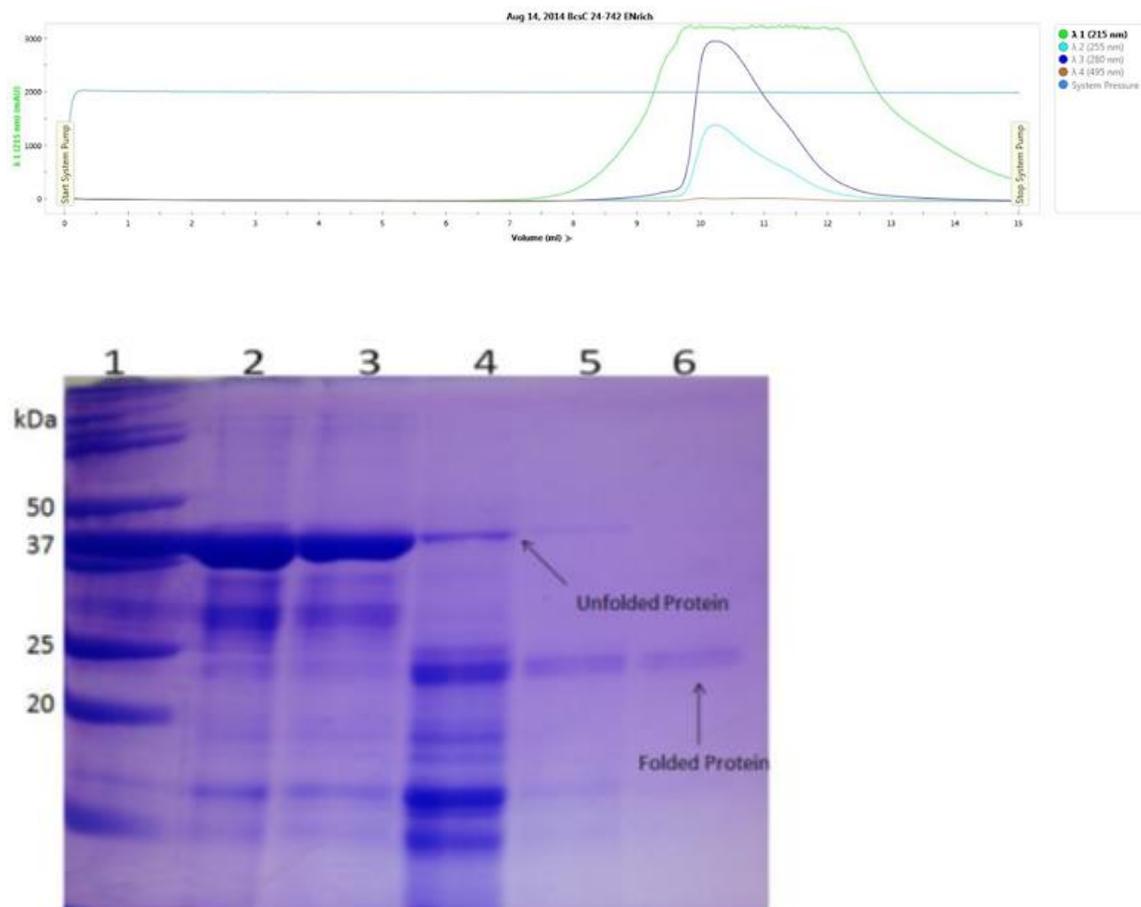
Figure A2.2: Protein folding performed on BcsC$^{783-1141}$ using 0.00036% (w/v) LDAO (Trepanier and Weadge Unpublished, 2015). Protein standard was loaded in lane 1. Samples decreased in concentration from 6 M urea (lane 2) to 3 M urea (lane 3), 1.5 M urea (lane 4), 0.75 M urea (lane 5) ending in 0.33 M urea (lane 6). No sample was run for 0 M urea due to no protein found when taking concentration readings. Band migration from 43 kDa in 6 M urea to ~25 kDa in 0.33 M urea is indicative of protein folding. Samples were loaded on a 12% SDS-PAGE, stained with Coomassie.

APPENDIX III
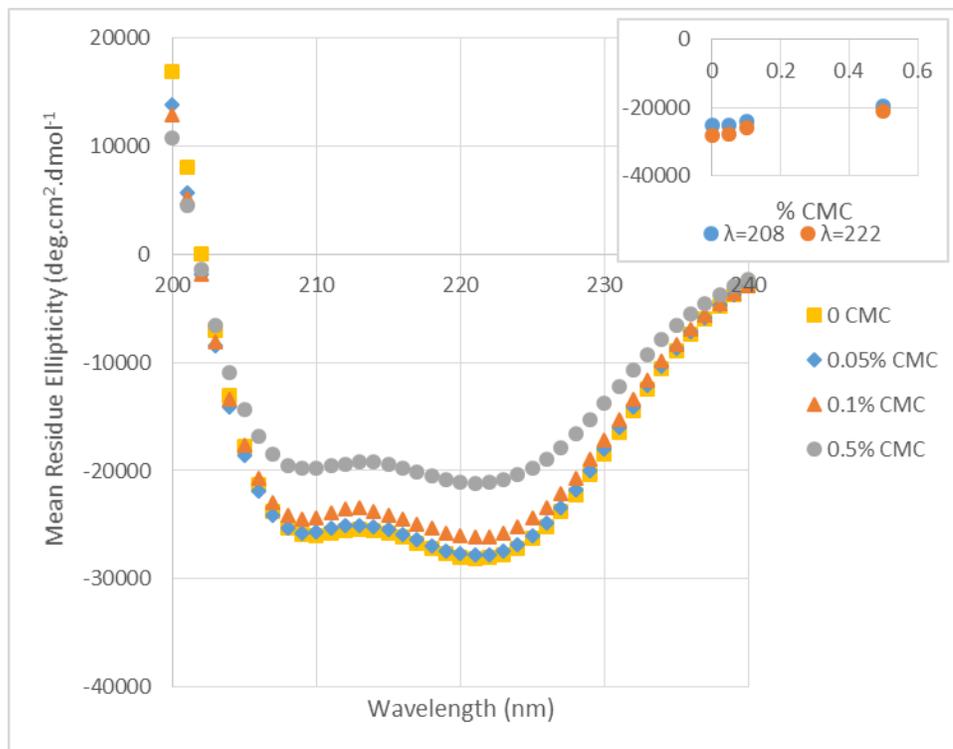
Figure A3.1: CD spectral profile of BcsC[24-742] in the presence and absence of CMC. Deconvolution was performed with CDSSTR using reference set 4 (Dichroweb). Data shown here was collected using protein purified and analyzed at different time points to one another. Data found in Table 7.1 correspond to the reconstructed data displayed here.
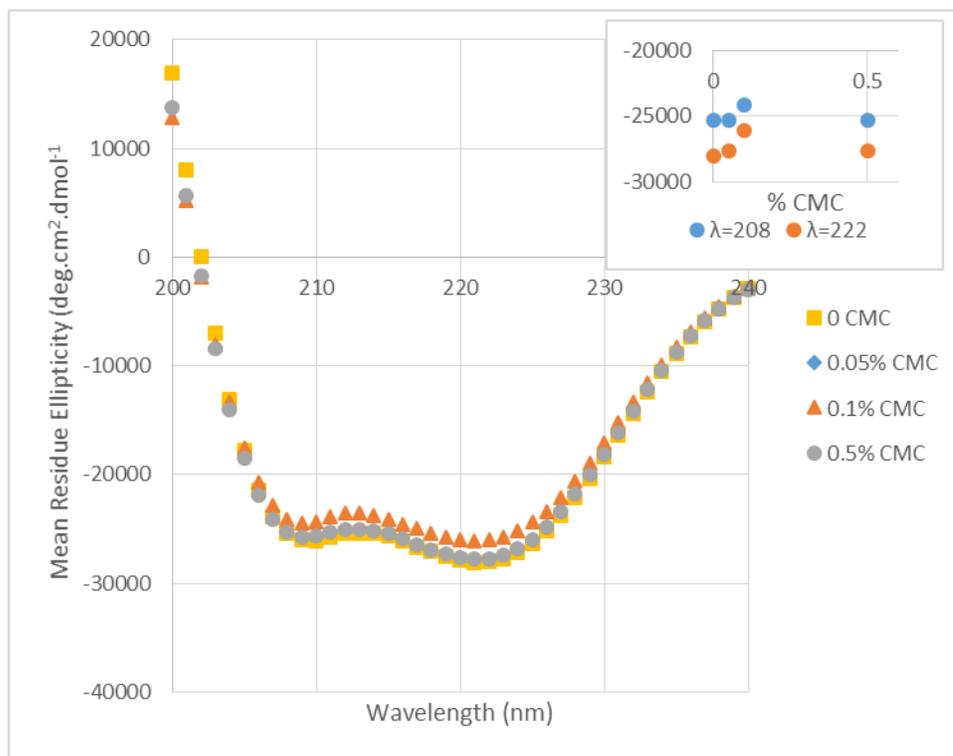
Figure A3.2: CD spectral profile of BcsC[24-742] in the presence and absence of CMC. Deconvolution was performed with CDSSTR using reference set 7 (Dichroweb). Data shown here was collected using protein purified and analyzed at different time points to one another. Data found in Table 7.2 correspond to the reconstructed data displayed here.
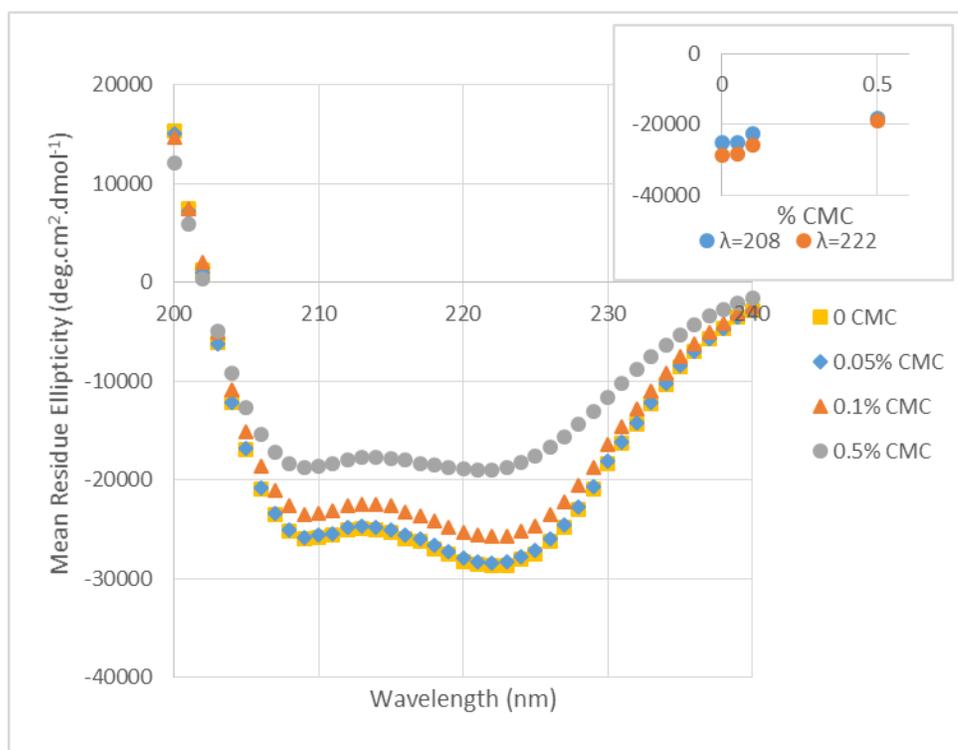
Figure A3.3: CD spectral profile of BcsC[24-742] in the presence and absence of CMC. Deconvolution was performed with K2D (Dichroweb). Data shown here was collected using protein purified and analyzed at different time points to one another. Data found in Table 7.3 correspond to the reconstructed data displayed here.
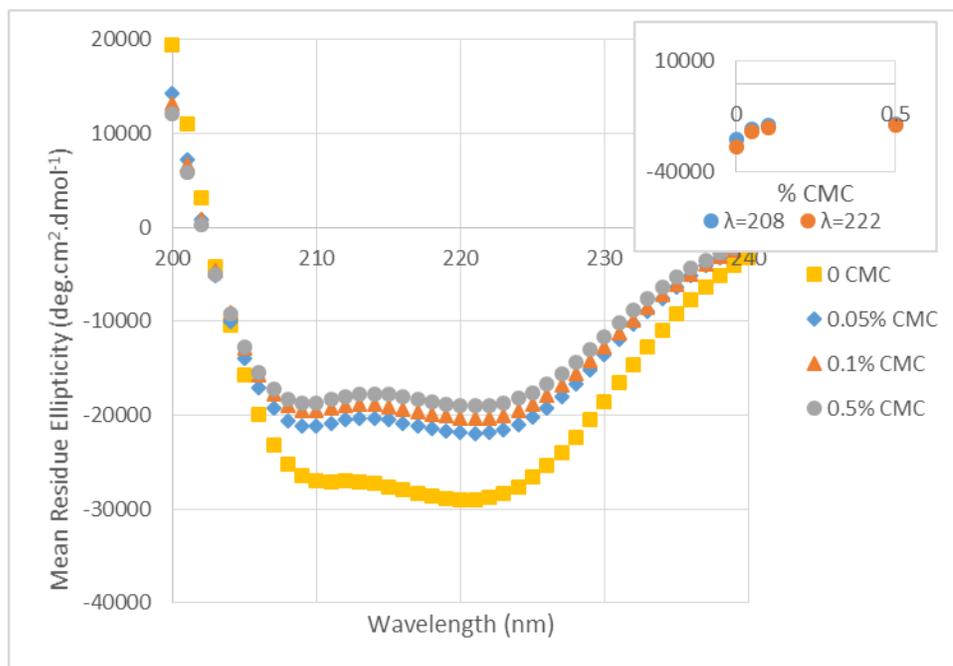
Figure A3.4: CD spectral profile of BcsC[24-742] in the presence and absence of CMC. Deconvolution was performed with SELCON3 using reference set 4 (Dichroweb). Data shown here was collected using protein purified and analyzed at different time points to one another. Data found in Table 7.4 correspond to the reconstructed data displayed here.
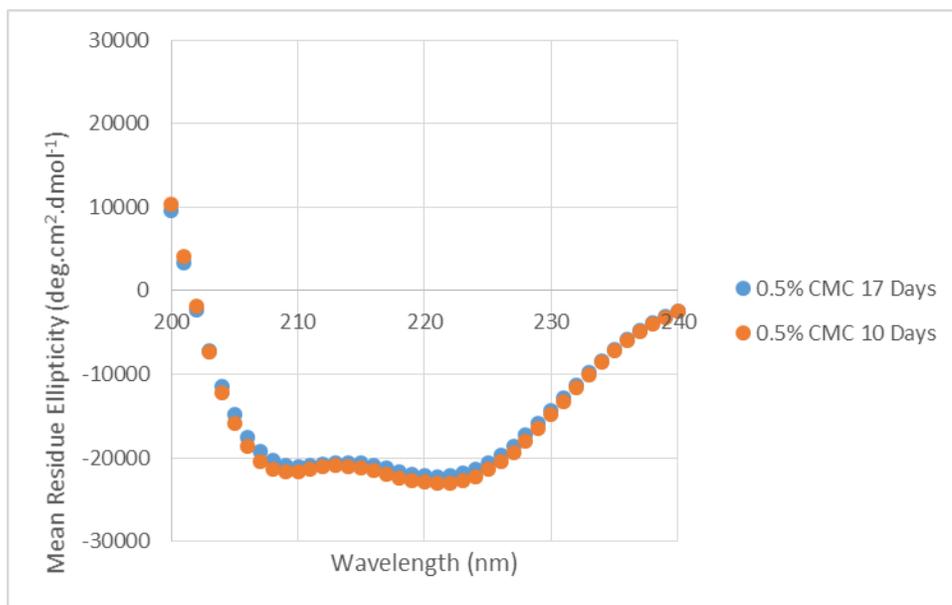
Figure A3.5: CD profile of BcsC[24-781] in the presence of 0.5% (w/v) CMC. Deconvolution was performed with CDSSTR using reference set 4 (Dichroweb). Data shown here utilize protein samples that have been purified at different time periods to one another; 10 days and 17 days prior to analysis. Data found in Table 7.5 correspond to the reconstructed data displayed here. Negative peaks at 208 and 222 nm are indicative of an α-helical secondary structure, as supported by data found in Table 7.5. Additionally, the minimal spectral shift between the two data sets reveals that the secondary structure remains intact after a week-long difference.
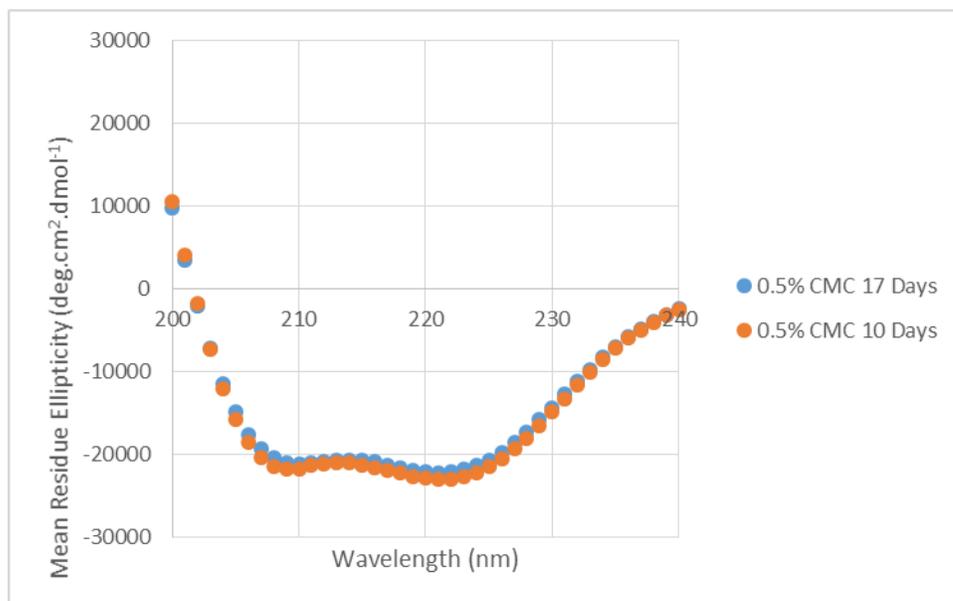
Figure A3.6: CD profile of BcsC[24-781] in the presence of 0.5% (w/v) CMC. Deconvolution was performed with CDSSTR using reference set 7 (Dichroweb). Data shown here utilize protein samples that have been purified at different time periods to one another; 10 days and 17 days prior to analysis. Data found in Table 7.6 correspond to the reconstructed data displayed here. Negative peaks at 208 and 222 nm are indicative of an α-helical secondary structure, as supported by data found in Table 7.6. Additionally, the minimal spectral shift between the two data sets reveals that the secondary structure remains intact after a week-long difference.
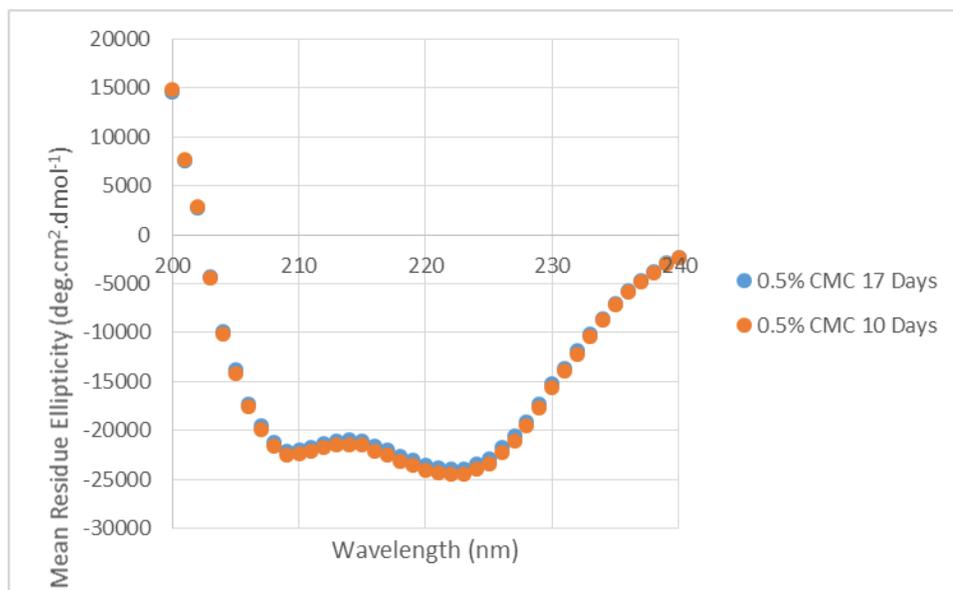
Figure A3.7: CD profile of BcsC[24-781] in the presence of 0.5% (w/v) CMC. Deconvolution was performed with K2D (Dichroweb). Data shown here utilize protein samples that have been purified at different time periods to one another; 10 days and 17 days prior to analysis. Data found in Table 7.7 correspond to the reconstructed data displayed here. Negative peaks at 208 and 222 nm are indicative of an α-helical secondary structure, as supported by data found in Table 7.7. Additionally, the minimal spectral shift between the two data sets reveals that the secondary structure remains intact after a week-long difference.
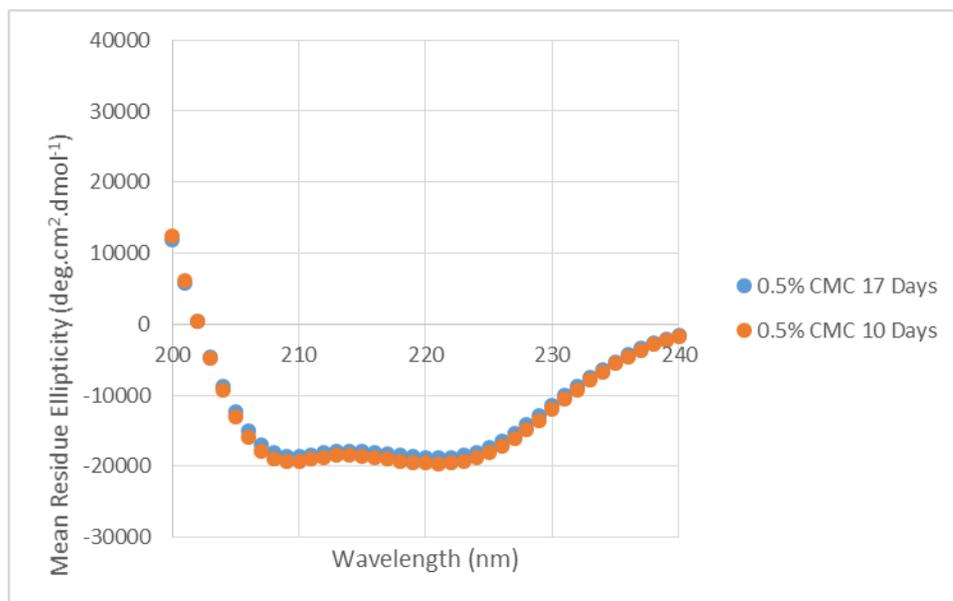
Figure A3.8: CD profile of BcsC[24-781] in the presence of 0.5% (w/v) CMC. Deconvolution was performed with SELCON3 using reference set 4 (Dichroweb). Data shown here utilize protein samples that have been purified at different time periods to one another; 10 days and 17 days prior to analysis. Data found in Table 7.8 correspond to the reconstructed data displayed here. Negative peaks at 208 and 222 nm are indicative of an α-helical secondary structure, as supported by data found in Table 7.8. Additionally, the minimal spectral shift between the two data sets reveals that the secondary structure remains intact after a week-long difference.

APPENDIX IV
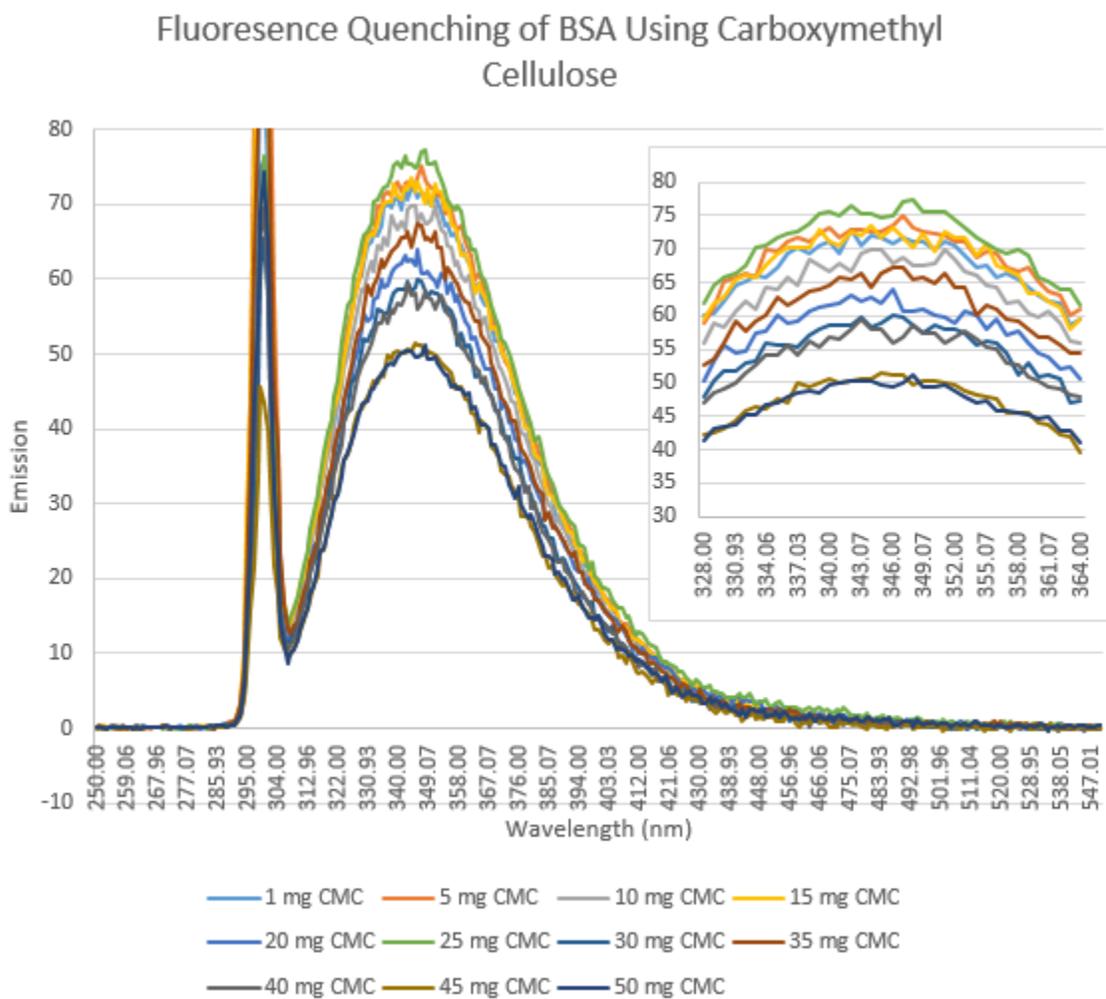
Figure A4.1: Fluorescence quenching of BSA with CMC. The emission profile was inconsistent, as CMC was added the emission increased up until the 25 mg sample, after which the spectra indicated quenching.

A

B
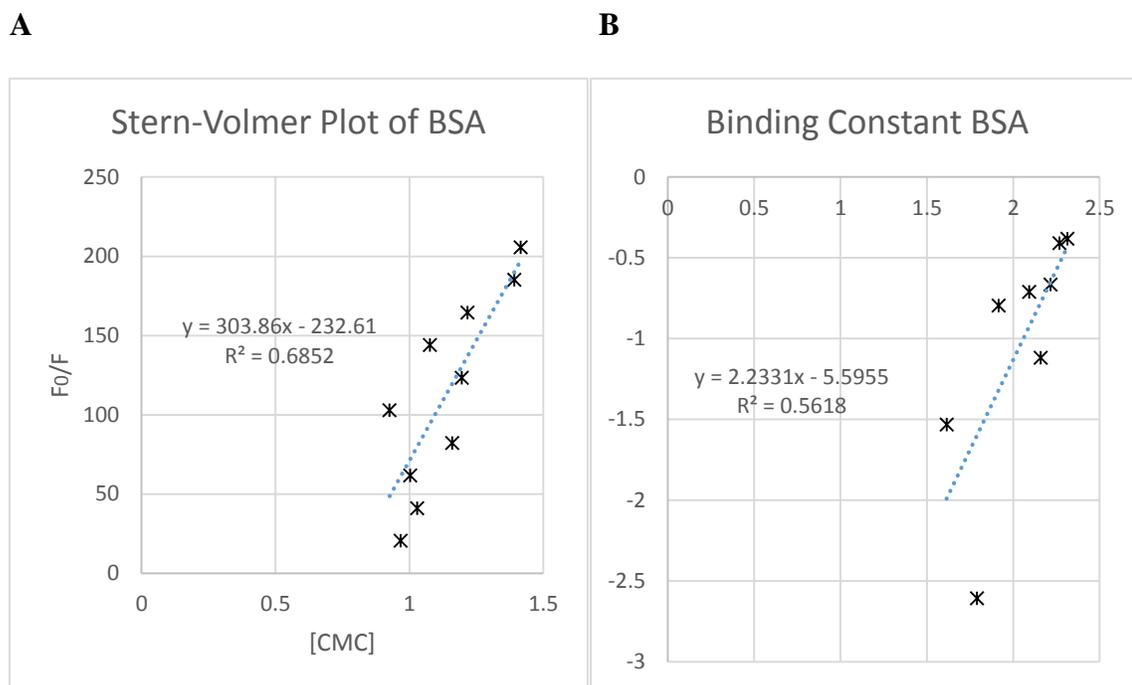
Figure A4.2: Stern-Volmer and binding constant plots of BSA. Plots are far from the trend line indicating that binding is not occurring.


1. Protein Standard
2. 0 mg Avicel
3. 1 mg Avicel
4. 5 mg Avicel
5. 10 mg Avicel
6. 15 mg Avicel
7. 20 mg Avicel
8. 25 mg Avicel
9. 50 mg Avicel
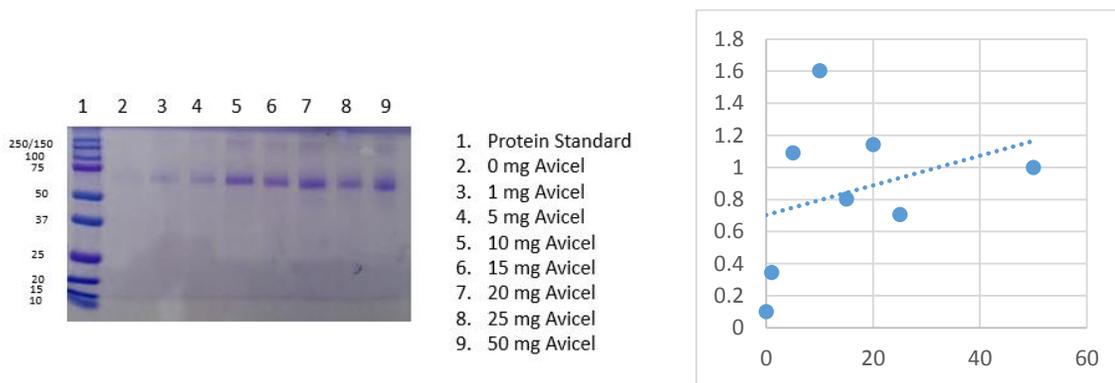
Figure A4.3: Avicel binding assay using 2 µM BSA y = 0.0092x + 0.7039 R^2 = 0.102