2017

Utilizing Genetic Techniques to Identify Amino Acids Within the Putative Glycosyltransferase Sypq That Are Essential for Its Role in Biofilm Formation by Vibrio Fischeri

Mary Kathryn Flaherty
Loyola University Chicago

LOYOLA UNIVERSITY CHICAGO

UTILIZING GENETIC TECHNIQUES TO IDENTIFY AMINO ACIDS WITHIN THE PUTATIVE GLYCOSYLTRANSFERASE SYPQ THAT ARE ESSENTIAL FOR ITS ROLE IN BIOFILM FORMATION BY *VIBRIO FISCHERI.*

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

MASTER OF SCIENCE

PROGRAM IN MICROBIOLOGY AND IMMUNOLOGY

BY

MARY KATHRYN FLAHERTY

CHICAGO, ILLINOIS

AUGUST 2017

ACKNOWLEDGEMENTS

I would like to thank everyone who has helped me along my journey. In particular, I would like to thank my mentor Karen Visick for her guidance and encouragement. I would also like to thank the other members of my thesis committee, Drs. Wolfe and Alonzo, for challenging me push myself farther.

I would like to personally thank Cecilia Thompson for her guidance, knowledge, and friendship, as well as Alice (Tischeri) Tischler for sharing her incredible writing skills. I would be remiss if I did not thank all the members of the Visick lab (Squid Squad), past and current, for your help and camaraderie.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

LITERATURE REVIEW

Introduction

Polysaccharides are polymers made from sugar subunits arranged together in the form of a linear or branched chain. These polymers can range in length from only a few subunits to around 10,000 subunits in the case of some cellulose fibers. Polysaccharides are ubiquitous in nature and can be utilized by cells as nutrient sources [1], as structural polymers [2], as signaling polymers [3], and in industrial and clinical settings[4, 5].

The production of polysaccharides is carried out by a class of enzymes known as glycosyltransferases (GTs). GTs facilitate the formation of long and short chain polysaccharides by catalyzing glycosidic bond formation between sugar subunits. These enzymes are found in all domains of life and are necessary for a variety of functions. One of these functions in bacteria is the production of exopolysaccharide for biofilm. Bacterial biofilms are colonies of organisms that are encased in a sticky substance and often attached to a surface.

Vibrio fischeri is a marine bacterium that utilizes biofilm formation as a means to colonize its natural host, the *Euprymna scolopes* squid [6]. The 18 gene *syp* locus encodes the proteins that produce the exopolysaccharide component of the biofilm [7, 8]. Without this polysaccharide, biofilm cannot form and the bacteria cannot colonize the squid. Of the 18 genes

in the locus, 6 are predicted to encode glycosyltransferases. The goal of my thesis was to probe the structure and function of one of these GTs: SypQ. This enzyme functions in an unknown mechanism to control biofilm formation in *Vibrio fischeri.* To better understand the activity of SypQ, I used genetic approaches to identify amino acid residues important for its function. Here, I provide an overview of glycosyltransferases: in particular family 2 glycosyltransferases, the class of glycosyltransferase that includes SypQ. I also discuss exopolysaccharide biosynthesis pathways. Finally, I discuss biofilms, in particular *V. fischeri* biofilms and the exopolysaccharide contribution to these biofilms.

Glycosyltransferases

Glycosyltransferases (GTs) are enzymes responsible for catalyzing the formation of glycosidic bonds using an activated sugar as a substrate. GTs are found in all domains of life. The number of glycosyltransferases that an organism encodes directly correlates to the number genes present. Approximately 1-2% of the genes in an organism encode glycosyltransferases [9].

Classification.

There are currently over seven thousand sequenced glycosyltransferases. These enzymes are classified into several families (currently, more than 80) based on a number of factors. One level of classification is based on overall amino acid sequence [10]. Members of a GT family will often contain conserved functional motifs in their amino acid sequence. For example, the family 2 glycosyltransferases (GT-2s) contain a highly conserved DxD motif that has been shown to function in the active site to facilitate substrate binding and coordination of divalent cations that are essential for the enzyme's function.

Another level of classification is the catalytic mechanism of the enzyme. Glycosyltransferases catalyze the formation of a glycosidic bond between the sugar substrate and the receiving molecule or growing polysaccharide chain. The catalysis will result in one of two possible outcomes: the stereochemistry of the donor molecule will be retained or it will be inverted (Figure 1A).

The final level of classification is the three-dimensional fold of the protein. Nearly all of the GT structures that have been solved to date either have a GT-A fold or a GT-B fold (Figure 1B) [9]. Both folds are similar in that they are composed of two β/α/β Rossman-like folds. The GT-A and GT-B folds differ in how they contribute to the overall structure and catalytic activity of the protein. In the GT-A fold, the β-sheets of the two Rossman-like folds come in close contact with one another and create a continuous domain of β-sheets in the center of the enzyme. This large β-sheet contains the active site. The GT-B fold, on the other hand, adopts a different conformation. The two Rossman folds face each other, forming a cleft between them. The active site is located within this cleft region. These individual classifications combine to categorize a GT into a family. Despite the large number of known GT families, family 2 (which have a GT-A fold) and family 4 (which have a GT-B fold) are the most common families of GTs. Approximately 50% of all GTs belong to one of these two families [9].

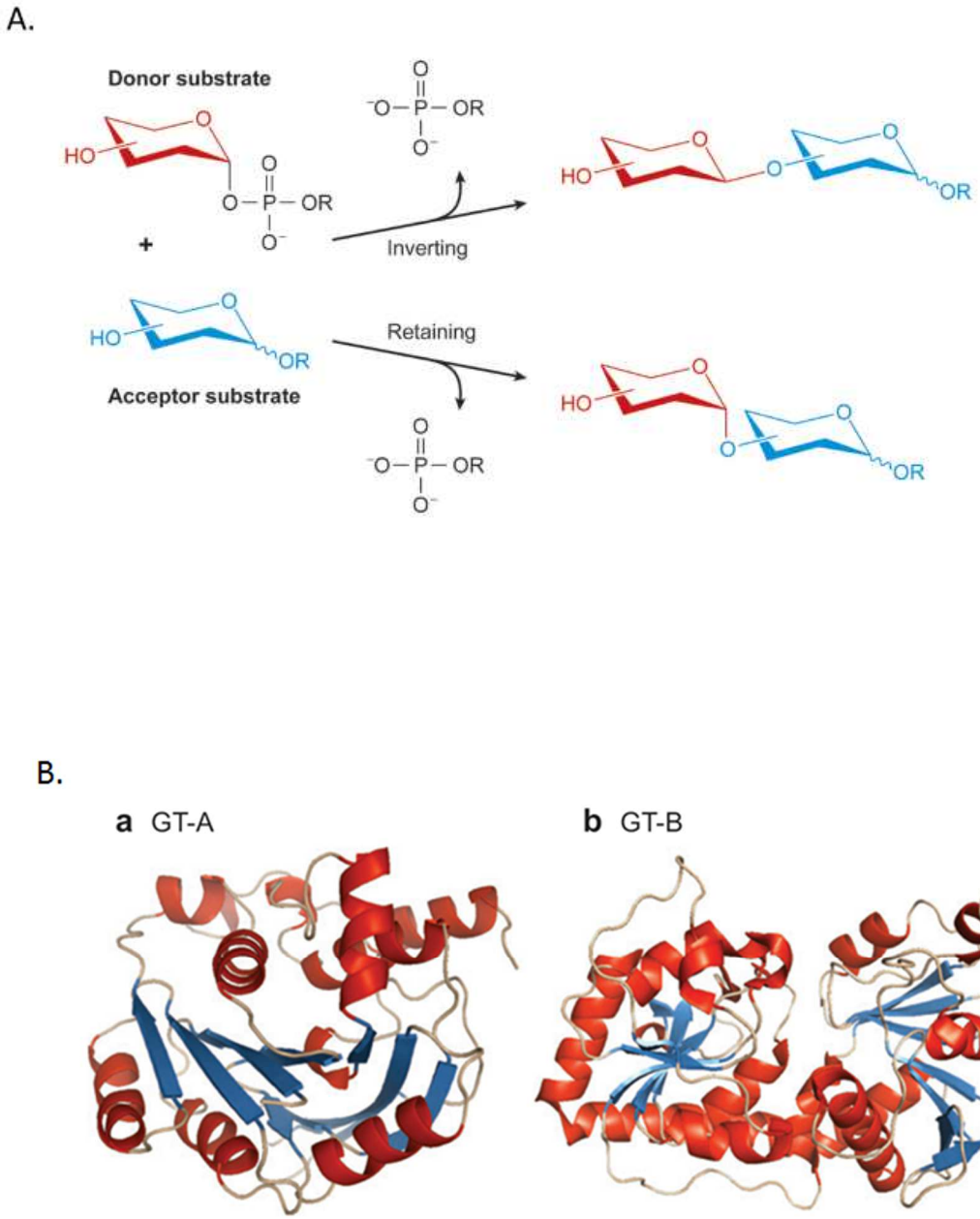

Figure 1. Structure and mechanisms of glycosyltransferases. A. Glycosyltransferases will catalyze the formation of glycosidic bonds via an inverting mechanism, in which the orientation of the anomeric carbon changes, or a retaining mechanism, where the anomeric carbon retains its orientation. **B.** Glycosyltransferases adopt one of two structural folds: a GT-A fold or a GT-B fold. Both folds are composed of two β/α/β folds. In the GT-A structure, the two folds come together to form a continuous β sheet at the base while in the GT-B structure, the two folds face each other and form a cleft between them [9]. Reprinted with permissions (Appendix B).

Family 2 glycosyltransferases.

The class of family 2 glycosyltransferases (GT-2s), which is one of the largest GT families, contains enzymes from all domains of life. While not all GT-2s are identical in sequence or in their substrate and product, they do all possess the same basic properties. All the known GT-2 enzymes adopt a similar 3-dimensional structure. While the size of the protein and number of transmembrane regions may differ, all GT-2s have a GT-A fold architecture. This similarity in structure leads to a similar mechanism of action as well. All members of this family of GTs catalyze the formation of glycosidic bonds via an inverting mechanism. In addition to structure and mechanism, GT-2s share some sequence similarities within the active region of the enzyme through the conservation of multiple functional motifs. This motif conservation is well characterized in one class of GT-2 enzyme in particular: the cellulose synthase catalytic subunit A (CesA) [11]. CesA and CesA-like enzymes have been thoroughly investigated in both plants (CesA) and in bacteria (BcsA). These enzymes catalyze the formation of the α(1-4) glycosidic bond between the donor sugar (a nucleotide diphosphate glucose) and the non-reducing end of the growing cellulose chain.

The catalytic domain of CesA enzymes consists of three variably spaced aspartic acids found within specific motifs in the primary amino acid sequence [12]. These conserved residues are flanked downstream by another motif, the Q/LxxRW motif (Figure 2). Together, these residues create the larger D, D, D, Q/LxxRW motif. The three aspartic acids are in the DDx, DxD, and TED motifs respectively [11, 12]. The number of amino acids that separates these motifs differs from enzyme to enzyme. Despite this variability, the functional motifs are often

arranged in the same location in the three-dimensional structure and the arrangement of these motifs is often conserved among CesA proteins from plants as well as bacteria [13, 14].

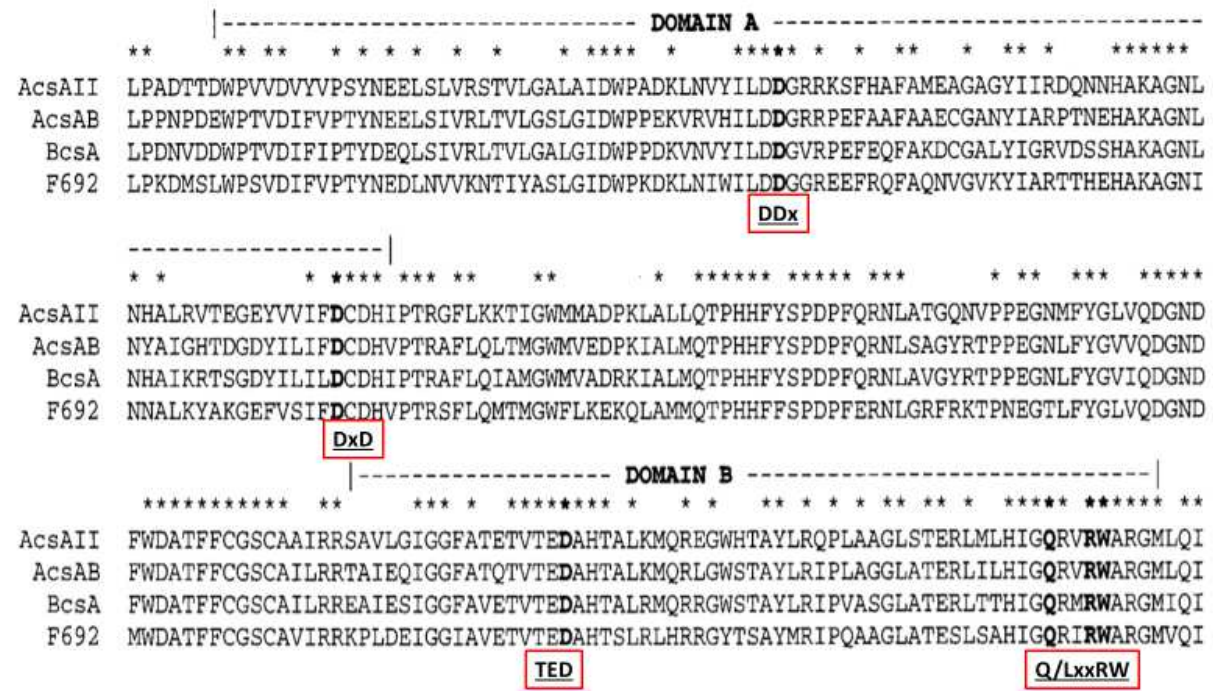

Figure 2. Alignment of GT-2s demonstrating conservation of the D, D, D, Q/LxxRW motif. As demonstrated by Saxena et. al., 1995, GT-2s share a specific conservation of certain motifs. Sequences from four processive GTs were aligned and the conserved residues indicated by an asterisk. The residues in bold type are those conserved in the functional motifs: DDx, DxD, TED, and Q/LxxRW, respectively [12].

BcsA. Many bacteria encode a CesA protein called BcsA (bacterial cellulose synthase). This protein has been well characterized in its mechanism as well as its structure [13, 15]. The BcsA sequence contains the motifs known to contribute to catalytic function. The first catalytically active aspartate in BcsA is the secondary aspartic acid in the DDx motif. The DDx motif is located adjacent to the DxD motif in the active site of the three-dimensional structure and the central aspartic acid interacts with the nucleotide bound to the donor sugar [16]. In other

GT-2 enzymes, the DDx motif has been indicated in bifunctionality within a single active site or the ability of the enzyme to catalyze the formation of both α- and β-glycosidic bonds [17].

The second essential aspartic acid is found in the DxD motif (in some GT-2s, the DxD is substituted with a DxH [18]). The role of the DxD functional motif has been well characterized in BcsA. These residues function in the active site where they interact with the phosphate of the nucleotide donor molecule and coordinate the divalent cations needed for the function of BcsA [11, 19, 20]. These cations stabilize the leaving group after the catalysis of the glycosidic bond. It has also been observed that the primary aspartic acid of this motif is flanked upstream by four hydrophobic residues [13, 21].

The third conserved aspartic acid residue in the BcsA catalytic active site is the terminal aspartate of the TED motif. This amino acid interacts with the acceptor molecule of the growing polysaccharide chain through hydrogen bonding. The aspartic acid then catalyzes the deprotonation of the growing chain at the non-reducing end [11, 20]. This allows the hydroxyl group of the donor molecule to bind to the growing polysaccharide chain.

While the conserved aspartic acid residues are involved in the catalytic activities of the BcsA active site, the Q/LxxRW motif is important for stabilizing the growing chain [14]. In BcsA, the tryptophan residue forms part of the transmembrane channel entrance. From here, the tryptophan interacts with the growing chain to keep the nonreducing end in proper position to interact with the aspartic acid residue from the TED motif. Keeping the two residues close to each other is essential for glycosidic bond formation [11].

Bacterial Exopolysaccharides

Functional Roles.

Bacterial exopolysaccharides (EPS) serve many functions. EPS that are bound to the cell contribute to capsule formation, which can protect the cells from desiccation and act as receptors for the cells [1]. Polysaccharides that are secreted into the extracellular space most often become the major component of the bacterial biofilm matrix. Within the matrix, the EPS polymers can have several functions. The polysaccharides facilitate adhesion of the cells to a surface, allowing for the initial stages of biofilm development. The EPS in biofilms can protect the bacteria from immune defenses during an infection (e.g. [22]) as well as confer resistance to antibiotic treatment (e.g. [23]). Thus, understanding how EPS is produced can lead to the development of effective treatment options.

Mechanisms of Production.

Despite the vast diversity of bacterial EPS, there are currently only four known mechanisms of production. The two most common pathways to EPS production are the Wzx/Wzy-dependent pathway and the ABC transporter-dependent pathway. Bacterial cells can also utilize the synthase-dependent pathway or produce the polysaccharide extracellularly through the use of a glycosyltransferase covalently bound to the cell surface [22].

Wzx/Wzy-dependent pathway. All polysaccharides made by the Wxz/Wxy-dependent pathway are heteropolymers, meaning that they are composed of multiple distinct sugar subunits. This makes the resulting polysaccharides incredibly diverse. This mechanism of EPS production is so named because all bacterial strains utilizing this method possess Wzx- and Wzy-like proteins. To begin the process, glycosyltransferases at the inner membrane generate individual

sugars. The proteins required for the polymerization of the polysaccharide chain are complexed at the inner membrane. There are three major players in the complex: Wzx, Wzy, and Wzz [24]. The Wzx protein is a flippase responsible for transferring the individual sugars to the periplasm, where they are polymerized into a chain by the Wzy protein. The length of the polysaccharide chain is controlled by the Wzz protein. These proteins have the capacity to oligomerize and, in the case of Wzz, this oligomerization contributes directly to its function [25]. If the polysaccharide is designated for capsule formation, the WaaL ligase protein will ligate the polysaccharide to a lipid oligosaccharide to form lipopolysaccharide. Once the polysaccharide chain is formed, two more proteins are needed to export the polymer out of the cell: a polysaccharide co-polymerase (PCP) and an outer membrane polysaccharide export (OPX) protein.

ATP binding cassette (ABC) transporter dependent pathway. The ATP-binding cassette transporter-dependent pathway is used primarily for the production of capsular polysaccharide. It is similar to the Wzx/Wzy-dependent pathway in that the process begins with the production of individual sugars at the inner membrane by membrane integrated glycosyltransferases. Unlike the Wzx/Wzy pathway, the sugar molecules are transported to the periplasm by a pump-like complex rather than a flippase. This complex is composed of transporter proteins and proteins from the PCP and OPX families. One major difference in the ABC transporter system, however, is that the entire polysaccharide chain is produced in the cytoplasm and is then transported to the pump complex by the ABC transporter complex [22].

Synthase-dependent pathway. There are noticeable differences in the synthase-dependent pathway compared to the Wzx/Wzy and ABC transporter pathways. The synthase-

dependent pathway produces homopolymers [26]. The polysaccharides produced in this pathway contain one kind of sugar molecule. Some examples of the polymers generated are curdlin, alginate, and hyaluronic acid. Another difference is that the polymerization and translocation occur simultaneously. A glycosyltransferase embedded in the inner membrane catalyzes the formation of the polysaccharide and exports it at the same time [22]. The regulation of polysaccharide production is often achieved by a co-polymerase [27]. This co-polymerase contains a PilZ domain and regulates the production of polysaccharide through c-di-GMP binding. The polysaccharide is stabilized in the periplasm by a scaffold protein and exported through a β-barrel porin protein. These export proteins are clearly distinct from the PCP and OPX protein common to the Wzx/Wzy and ABC transporter mechanisms.

Extracellular synthase pathway. Not all bacterial EPS is generated by a complex of proteins and not all exopolysaccharides are produced intracellularly. In the extracellular synthase pathway, the exopolysaccharide is produced by a glycosyltransferase that is covalently linked to the cell surface. Two examples of polysaccharide generated by this pathway are dextran and levan [28, 29].

Biofilm

Exopolysaccharides are the primary component of the bacterial biofilm matrix. Bacterial biofilms are sessile social communities that form on organic or inorganic surfaces and surround themselves with a thick sticky substance [30, 31]. This substance is composed primarily of exopolysaccharide, extracellular protein, and extracellular DNA. Biofilms and their communities are very diverse depending on the species (single species or multi-species) that make up the community and the environment in which the biofilm is formed [32]. Despite this heterogeneity,

bacterial biofilms possess some basic similar characteristics in their formation, matrices, and function. These similar characteristics will be described below.

Stages of Bacterial Biofilm Formation.

Environmental factors can induce bacterial biofilm formation. Environmental changes can lead bacteria to initiate biofilm formation by changing from a planktonic lifestyle to a sessile one [33]. The first stage of biofilm formation occurs when the planktonic cells attach to a surface. Extracellular appendages that aid in motility, such as flagella, enhance the bacteria's attachment to a surface [34].

Once the bacteria have attached to a surface, the transition to a sessile lifestyle is triggered by the upregulation of factors needed for the formation of the biofilm matrix [35, 36], beginning the process of biofilm maturation. The biofilm matrix accounts for approximately 90% of the biomass of the colony [37] and can enhance bacterial survival by trapping nutrients and water. Bacteria can adjust the matrix to suit changes in environment by secreting enzymes that alter the EPS component.

The final stage in bacterial biofilm development is the dispersal stage. Dispersal from the biofilm can be either active or passive. Passive dispersal occurs by physical disruption of the colony but can also occur in response to external stimuli such as a change in nutrient levels [38]. It has also been demonstrated that iron levels can trigger the dispersal of planktonic bacteria from mature biofilms [39].

Biofilm Matrix.

The matrix of bacterial biofilms accounts for the majority of the biomass [37]. The individual components of biofilm matrices can differ from biofilm to biofilm; however, the majority of bacterial biofilm matrices contain extracellular polysaccharides [(exopolysaccharides (EPS)], proteins, and DNA (eDNA).

The EPS is a key component of the biofilm matrix [32]. For example, in *V. cholerae,* 50% of the biomass of the biofilm is composed of the VPS polysaccharide [40-42]. Biofilm EPS is not always composed of a singular polymer. Often, the EPS is made of multiple different polysaccharides whose production can be regulated by separate pathways. The matrix EPS can interact with other components in the biofilm. Carbohydrate binding proteins can bind to EPS and form cross-links that can strengthen the matrix [37, 43]. Extracellular DNA (eDNA) is another integral part of the matrix and can also interact with the EPS. In *P. aeruginosa,* the eDNA interacts with the PSL polysaccharide to create a structure that supports to the biofilm [44].

The matrix has many functions in biofilms. Nutrients and water can become trapped in the matrix, which then supplies these much needed elements to the cellular community [45]. Cells within the matrix are held in close proximity via the cross-links formed by EPS and proteins [37]. This closeness allows for intercellular interactions and signaling. The matrix can also act like a recycling center. When cells die in the matrix, their components are kept close and the neighboring cells can utilize them.

Functional Roles of Biofilms.

Bacterial biofilms confer a number of benefits to the microbial communities that live within them. One major benefit to the bacterial cells is protection. Studies have demonstrated that the biofilm produced by *P. aeruginosa* can protect the cells within from antibiotics [23]. A recent study of *V. cholerae* biofilms demonstrated that the cells within the matrix were protected from invading bacterial cells [46]: motile *V. cholerae* cells could colonize the outside of the established biofilm but could not enter the matrix.

Clinical and Industrial Significance of Biofilms.

Approximately 80% of all human bacterial infections involve the formation of biofilm in some way. In a clinical setting, biofilm-associated infections are commonly seen with implanted devices such as catheters, artificial joint replacements, and implanted devices such as pacemakers [47, 48]. The biofilm-associated infections can lead to the ability of antibiotic resistant cells to persist in the body, which in turn can lead to chronic infection. In addition to creating problems in a clinical setting, bacterial biofilms have been shown to have a negative effect on industry, in particular the food industry where they can lead to food spoilage, contamination, and disease [49].

Vibrio fischeri

Symbiosis.

Vibrio fischeri is the natural symbiont of *Euprymna scolopes,* the Hawaiian bobtail squid. The *V. fischeri* bacteria produce light that allows the squid to evade predators [50]. The specificity of the symbiosis is such that, of the multitude of bacterial species in marine environments, only *V. fischeri* is capable of colonizing the light organ of the *E. scolopes* squid

[51]. Despite this specificity, *V. fischeri* must overcome many obstacles to achieve colonization, such as removal by phagocytes and reactive oxygen species [50]. *V. fischeri* biofilm allows the bacteria to attach to the surface at the opening to the light organ and colonize the squid.

The Contribution of *syp* Genes to *V. fischeri* Biofilm.

Previous genetic studies of *V. fischeri* have identified an 18-gene locus essential to the promotion of biofilm formation and symbiotic colonization [52]. This locus, the symbiotic polysaccharide (*syp)* locus, has been shown to contribute to the production and secretion of the exopolysaccharide component of the biofilm matrix [7]. All of the structural proteins encoded in the locus are important for colonization. Of the 18 genes in the locus, there are four that encode putative regulatory proteins, *sypA, E, F,* and *G,* and two that encode putative polysaccharide export proteins, *sypC* and *sypK*. Six genes, *sypB, D, L, M, O* and *R,* are thought to encode proteins that modify the *syp* polysaccharide or control the length of the growing polymer and the remaining genes, *sypH, I, J, N, P,* and *Q,* encode putative glycosyltransferases [8, 53]. A disruption to any one of these genes results in a reduction or total abrogation of biofilm formation *in vitro* as well as a reduction or loss of colonization of the *E. scolopes* by *V. fischeri.*

syp transcription is controlled through a network of regulatory proteins. The sensor kinase RscS is activated upon reception of an as of yet unknown signal [54, 55]. Once activated, RscS will activate the response regulator SypG. SypG is an essential transcriptional regulator that binds to the *syp* promotor regions to initiate σ^{54} transcription [56]. RscS also controls the activity of the response regulator SypE [57]. SypE has both positive and negative regulatory capabilities and regulates *syp* activation by controlling the activity of SypA, whose function is still unknown [58]. *In vitro,* biofilm formation can be induced in *V. fischeri* through the

overexpression of RscS. This overexpression induces the positive regulatory function of SypE by activating its C-terminal domain [57]. Biofilm can also be induced by overexpression of SypG in the absence of SypE [59]. Under the inducing conditions, *V. fischeri* forms wrinkled colonies (instead of smooth) and pellicles (biofilms at the air/liquid interface of static cultures).

Of the 6 putative *syp* glycosyltransferases, one is of particular interest: SypQ. While *sypH, I, J, N, and P* encode putative GT-1 family glycosyltransferases, only *sypQ* encodes a putative GT-2 enzyme. The deletion of *sypQ* results in a total loss of biofilm as well as a light colony phenotype (Thomson and Visick unpublished data) that mimics that of mutations in the regulatory genes. Understanding SypQ and its role in the production of the *V. fischeri syp* polysaccharide is an essential step in understanding the mechanism by which this polysaccharide is made.

CHAPTER TWO

MATERIALS AND METHODS

Strains and Media

All bacterial strains utilized in this study are listed in Table 1. *V. fischeri* strains were derived via triparental conjugation into recipient *V. fischeri* strains KV5562 or KV4883. GT115 (Invivogen, San Diego, CA), and TAM1 competent *E. coli* cells were used for conjugation and cloning [6]. *E. coli* strain CC130, a derivative of the Mph44 [60] mutator strain, was used for random mutagenesis of *sypQ*. All *V. fischeri* strains were grown using Luria-Bertani salt (LBS) medium [61]. The following antibiotics were added to the LBS medium: Chloramphenicol (Cm) at a final concentration of 2.5 μg/ml, and/or Tetracycline (Tet) at a final concentration of 2.5 μg/ml. *E. coli* strains were grown in Luria-Bertani medium (LB). Ampicillin (Amp) at a final concentration of 100 μg/ml and Cm at a final concentration of 12.5 μg/ml were added to the LB media. Thymidine was added when appropriate at a final concentration of 0.003 M.

Transformation

Bacterial transformation was used to introduce plasmid DNA into *E. coli* strains. 5 μl of plasmid DNA was added to 50 μl of competent cells. The mixture was incubated on ice for 30 minutes. The cells were heat shocked at 42°C for exactly 30 seconds then rested on ice for 2 minutes. Cells were then incubated in Super Optimal Broth with catabolite repression (SOC) [62]

at 37°C for 1 hour. 100 µl of the cell culture was plated on LB plates containing Amp or Cm and incubated at 37°C overnight.

To make *E. coli* cells competent, cells were first incubated overnight in 5 ml of LB media at 37°C with shaking at 220 rpm. After incubation, 2.5 ml of the overnight culture was used to inoculate 250 ml of fresh LB. The cells were grown at 37°C to an optical density at 600 nm (OD_{600}) of 0.3. The culture was incubated on ice for 15 to 20 minutes. The cells were centrifuged at 2000xg for 15 minutes at 4°C. The cells were resuspended in 100 ml of 100mM $CaCl_2$ and incubated on ice for 20 minutes. The suspension was centrifuged for 15 minutes at 2000xg at 4°C. The pellet was resuspended in 100 ml of 100mM $CaCl_2$ and incubated on ice overnight at 4°C. After incubation, glycerol was added to a final concentration of 10% and the cells were aliquoted and frozen at -80°C

Conjugation

Triparental conjugation was used to introduce plasmid DNA into *V. fischeri* strains. *E. coli* strains containing the *sypQ* plasmid DNA and the helper *E. coli* strain pEVS104 [63] were grown overnight in LB broth containing the appropriate antibiotic at 37°C with shaking at 220 rpm. The recipient *V. fischeri* strain was grown overnight in LBS plus Tet at 28°C with shaking at 220 rpm. After incubation, a 50 µl aliquot of each overnight culture was diluted in 5 ml of liquid medium plus antibiotic and incubated at the appropriate temperature with shaking for ~2 hours. Post incubation, 250 µl of the donor *E. coli* strain, 250 µl of the pEVS104 helper *E. coli* strain, and 1 ml of the recipient *V. fischeri* strain were combined in a 1.5 ml Eppendorf tube and centrifuged at 13,000 rpm for 2 minutes. The supernatant was discarded and the pellet was resuspended in 10 µl of the remaining liquid. The cellular suspension was then spotted onto an

LBS plate and incubated at 28°C for a minimum of 3 hours but no longer than overnight. The spot was resuspended in 1 ml of LBS and 100 µl of the suspension was plated onto LBS plates containing Cm and Tet. For the random mutagenesis, the resuspended spot was diluted before plating. 2 µl of the suspension was added to 1 ml of LBS (a 1:500 dilution). The plates were incubated at 24°C for 48 hours. To isolate *V. fischeri* from the *E. coli* background, the *V. fischeri* colonies were streaked onto a fresh plate.

PCR Reactions

Primers used in this study are listed in Table 2. *sypQ* alleles containing either a FLAG epitope tag (pCKM1-4) or a single site-directed point mutation (pMKF2-13), were generated using a PCR SOEing (splicing by overlap extension) reaction [64]. Primers containing either the FLAG epitope tag or the specific point mutation were used to generate tagged/mutated fragments of *sypQ* with overlapping ends. 5 µl of 10X KOD polymerase buffer, 3 µl of 25 mM $MgCl_2$, 5 µl of 2 mM dNTPs, 1 µl each of the 20 µM required forward and reverse primers, 1 µl of the DNA template, and 0.4 µl of the KOD HiFi polymerase were added to 33.6 µl of dH_2O. The reaction was processed using a BioRad MJ mini personal thermo cycler or a T100 thermo cycler using the following parameters: a denaturing step at 98°C for 15 seconds, an annealing step at 55°C for 5 seconds, and an extension step at 72°C for 15 seconds. The cycles were repeated thirty times. The product was then cleaned and concentrated using the Zymogen (Irvine, CA.) Clean and Concentrate kit. The resulting fragments were annealed together without primers at the overlapping regions and the strand was extended with the overlap acting as the free 3' end. 5 µl of 10X KOD polymerase buffer, 3 µl of 25 mM $MgCl_2$, 5 µl of 2mM dNTPs, approximately 100 ng of each of the fragments from the previous PCR reaction, and 0.4 µl of the KOD HiFi

polymerase were added to dH_2O for a final volume of 50 µl. The reaction was completed using the previously stated parameters. The cycles were repeated twenty times. This mega-fragment was then amplified using primers 1963 and 1964. The reaction was completed using the previous parameters and running for thirty cycles. The amplified *sypQ* allele was then ligated into the pJET 1.2 blunt cloning vector (Thermo Scientific, Waltham, MA.) (see below).

The pMKF1 plasmid was generated by first amplifying the *syp* promoter from ES114 chromosomal DNA via a traditional PCR reaction [65]. Primers were used to first amplify the *syp* promotor. 0.5 µl of ES114 chromosomal DNA, 5 µl of 5X reaction buffer (Promega, Madison, WI.), 2 µl of 2.5 mM dNTPs, 1.5 µl of 25 mM $MgCl_2$, 1.5 µl of 20 µM forward primer (1991) and 1.5 µl of 20 µM of reverse primer (1992) were added to 12.875 µl of dH_2O and 0.125 µl of Taq polymerase (Promega). The reaction was completed using a BioRad MJ mini personal thermo cycler or a T100 thermo cycler with the following parameters: an initial denaturing step at 95°C or 2 minutes, a cycle denaturing step at 95°C for 1 minute, an annealing step at 55°C for 30 seconds, and an extension step at 72°C for 1.5 minutes. The cycle (steps 2-4) was repeated 30 times. Following the final cycle, a final extension step was performed at 72°C for 5 minutes. The final product was stored at 4°C until needed. The amplified *syp* promotor was ligated into the pJET cloning vector and transformed into TAM1 competent *E. coli.*

Digests and Ligations

Plasmids used in this study are listed in Table 3. For cloning of the site-directed *sypQ* mutants, 5 µl of the 2X pJET reaction buffer, 3.5 µl of dH_2O, 0.5 µl of the amplified PCR product, and 0.5 µl of the pJET blunt cloning vector were incubated with 0.5 µl of T4 DNA

ligase at room temperature for 5 minutes and then transformed into TAM1 competent *E. coli* cells. The cells were then plated on LB media with Amp and incubated at 37°C overnight.

DNA digestions were used to excise the target gene from the pJET cloning vector and to verify the presence of the insert in the pMKF1 plasmid backbone. 1 µL of 10X digest buffer (Thermo Fischer Scientific), 0.5 µl of the required fast digest restriction enzymes (SacI, SalI, and BglII) (Thermo Fischer Scientific), and 5 µl of plasmid DNA were added to dH_2O to a final volume of 10 µl. The mixture was incubated at 37°C for 30 minutes. The digestion product was resolved on a 1% agarose gel at 110V for 35-45 minutes. For cloning of the *sypQ* insert into the pMKF1 plasmid, the *sypQ* band was excised from the gel and the DNA was recovered using the Zymogen Gel DNA recovery kit. The extracted DNA was then ligated into the pMKF1 backbone. 3 µl of the digested pMKF1 backbone and 1 µl of the digested insert (a vector to insert ratio of 3:1) were incubated with 4.5 µl of dH_2O, 1 µl of 10X T4 ligation buffer, and 0.5 µl of T4 DNA ligase at room temperature overnight. The ligation product was then transformed into GT115 competent *E. coli* cells. To verify the presence of the insert, the plasmid was extracted using the Zymogen Plasmid Miniprep kit and sequenced using ACGT Inc. (Wheeling, IL).

Bioinformatics

Multiple sequence alignments were performed using the Clustal W server [66, 67]. To perform sequence analysis on the mutated s*ypQ* alleles, an NCBI BLAST analysis was performed [68]. To predict potential transmembrane regions of SypQ, the sequence was entered into both the PredictProtein [69] and Topcons [70] bioinformatic servers. The Phyre2 [71] sever was utilized to predict a possible structure for SypQ.

Site-directed Mutagenesis of *sypQ*

Mutants of *sypQ* containing specific point mutations were generated using a PCR SOEing reaction. The mutated allele of *sypQ* was ligated into the pMKF1 backbone and transformed into GT115 competent *E. coli* cells. The plasmid was introduced into *V. fischeri* strain KV5562 and KV4883 via triparental conjugation. A wrinkled colony assay was performed to determine if the *sypQ* allele containing the specific point mutation was capable of restoring biofilm formation to a *sypQ* deficient strain of *V. fischeri.*

Random Mutagenesis of *sypQ*

To introduce random mutations in *sypQ,* the pMKF1 plasmid was grown in the CC130 mutator *E. coli* strain. The plasmid was introduced into the cells via transformation as previously described. The mutated plasmid was then introduced into the KV5562 *V. fischeri* strain via triparental conjugation. The conjugation spot was incubated at 28°C for three hours. For each independent mutagenesis, the conjugation spot was resuspended in 1 ml of LBS broth and diluted by adding 2 µl of the suspension to 1 ml of LBS. 100 µl of the dilution was plated onto each of 7 LBS plates containing Cm and Tet (at final concentrations of 2.5 µg/ml for each antibiotic). The plates were incubated at 24°C for 5 days (~220 hours). After 72 hours, the plates were counted to determine the number of colonies/plate. After 220 hours, the plates were screened for any colonies with a mutated phenotype (smooth, delayed, dark, or architecturally abnormal). These isolates were streaked onto LBS plates containing Cm and Tet and incubated at 24°C for 48 hours. The isolates were restreaked to generate single colonies. A wrinkled colony assay (see below) was performed to determine whether the randomly mutated *sypQ* allele was capable of restoring biofilm to the *sypQ* deficient strain. To determine if the observed mutant

phenotype was due to a mutation in the *sypQ* allele on the plasmid, the plasmid was extracted using the quick boil method and transformed into GT115 competent *E. coli* cells. The plasmid was then introduced back in to the parent strain (KV5562) via triparental conjugation. A wrinkled colony assay was performed to determine if the original defective phenotype was retained, a result that would suggest that the *sypQ* allele on the plasmid carried a mutation. Any plasmid that resulted in a delayed or diminished phenotype was extracted from the corresponding GT115 cells using a Zymogen Plasmid Miniprep kit. The concentration of the extracted plasmid was determined using a nanodrop. Mutations were identified by sequence analysis with primers M13R and M13F-20 using ACGT Inc. (Wheeling, IL). For *sypQ* alleles that resulted in a smooth colony phenotype, the corresponding plasmids were transferred into KV4883 (see below).

Wrinkled Colony Formation Assay

To assess the ability of the mutant s*ypQ* alleles to restore biofilm formation to a s*ypQ* deficient strain of *V. fischeri* (KV5562)*, V. fischeri* strains containing the plasmids carrying the mutated *sypQ* allele were streaked onto LBS plates containing the appropriate antibiotic. Single colonies were collected and grown overnight in liquid LBS media with appropriate antibiotics and shaking at 220 rpm. In later experiments, the cultures were standardized to an OD_{600} of 1.0 at this stage. The overnight culture was then diluted by adding 50 µl of overnight culture to 5 ml of fresh media and incubated at 28°C with shaking at 220 rpm for ~1.5 hours. The subcultures were then standardized to an OD_{600} of 0.2. 10 µl of the standardized culture were spotted onto LBS plates containing Cm and Tet and incubated for 72 hours at 24°C. All plates used for the wrinkled colony assay contain exactly 25 ml of media. Images of the colonies were taken at 24, 48, and 72 hours using the Zeiss stemi 2000-C dissecting microscope.

KV4883-Based Wrinkled Colony Assay

Any plasmids that resulted in a null phenotype were introduced via conjugation into *V. fischeri* strain KV4883. This strain is a biofilm competent strain that contains a wild-type copy of *sypQ*. Once the mutated plasmid was introduced into KV4883, the strain was grown overnight in 5 ml of LBS containing Cm and Tet. The cultures were incubated at 28°C with shaking at 220 rpm. After incubation, the overnight strains were standardized to an OD_{600} of 1.0. 50 µl of diluted culture was used to inoculate 5 ml of LBS containing Cm and Tet. The inoculated cultures were incubated at 28°C with shaking at 220 rpm for ~2.5 hours. After incubation, the cultures were standardized to an OD_{600} of 0.2. A wrinkled colony assay was performed to determine if the mutated protein displayed an inhibitory effect on wild type SypQ. Pictures were taken every hour starting at 19 hours and continuing until 24 hours.

Cell Fractionation Assay

The cell fractionation protocol utilized is adapted from previously described methods [72, 73]. 1.5 ml of overnight culture were incubated on ice for 5 minutes and then pelleted at 16,000xg for 3 minutes at 4°C. the pellet was resuspended in 150 µl of spheroplast buffer (100 mM Tris pH 8, 0.5 mM EDTA, 0.5 mM sucrose, 20 ug/ml PMSF) and iced for 5 minutes. 50 µl was frozen for whole cell fraction. The rest was spun for 3 minutes at 16,000xg and 4°C. The supernatant was removed and the pellet was warmed to room temperature and resuspended by shaking in 100 µl of ice-cold dH_2O and incubated on ice for 45 seconds. 5 µl of 20 mM $MgCl_2$ was added and the sample was centrifuged for 3 minutes at 16,000xg at 4°C. The supernatant was frozen for the periplasm fraction. The pellet was resuspended in 150 µl of cold spheroplast

buffer with 15 µl of 2 mg/ml lysozyme and 150 µl of ice-cold dH_2O and incubated for 5 minutes on ice. After incubation, the sample was centrifuged at 16,000xg at 4°C for 3 minutes. The pellet was resuspended in 600 µl of 10 mM Tris pH 8 with 20 µg/ml of PMSF. The sample was then frozen and thawed at -80°C three times. 20 µl of 1M $MgCl_2$ and 6 µl of 1 mg/ml DNase was added and the sample was centrifuged at 16,000xg at 4°C for 25 minutes. The supernatant was frozen for the cytoplasmic fraction. The pellet was resuspended in 50 µl of 10 mM Tris pH 8 with 2% Triton X-100 and incubated at room temperature for 30 minutes. The sample was centrifuged at 16,000xg at 4°C for 25 minutes. The supernatant was frozen for the inner membrane fraction. The pellet was resuspended in 50 µl of 10 mM Tris pH 8 and frozen for the outer membrane fraction. The samples were then visualized via Western blot (see below).

Rough Membrane Preparation

To isolate a rough membrane fraction from *V. fischeri*, cells were grown overnight in LBS media with Tet and Cm at 28°C with shaking at 220 rpm. After incubation, the cells were centrifuged at 13,000xg for 2 minutes. The supernatant was discarded and the pellet was resuspended in 250 µl of sample buffer (described below) and lysed via sonication. The sample was centrifuged again at 16,000xg at 4°C for 25 minutes. The supernatant was removed and the pellet was resuspended in 200 µl of sample buffer. Both the resuspension and the supernatant were visualized via Western blot (see below).

Western Blot Analysis of *V. fischeri* Strains

V. fischeri strains carrying the mutated *sypQ* alleles were grown overnight at 28°C with shaking at 225 rpm. 1.5 ml, 5 ml, 10 ml, or 20 ml of the culture was centrifuged at 13,000 rpm for 2 minutes. Cells were lysed in 200 µl of 2X sample buffer (4% SDS, 10% β-mercaptoethanol,

0.0005% bromophenol blue, 0.1 M Tris pH7, and 2 ml of dH_2O). The proteins were separated by SDS-PAGE with 150V for 1 hour at 24°C. The samples were loaded onto a 12% acrylamide gel (8.6 ml dH_2O, 5.0 ml 1.5 M Tris pH 8.8, 200 µl 10% SDS, 9.2l ml 30% Acrylamide, 115 µl 10% APS, and 7.5 µl of TEMED). Proteins were transferred to PVDF membrane with 35V for 16 hours at 4°C. FLAG -tagged proteins were detected by western blot analysis using anti-rabbit-anti-FLAG primary antibody (Sigma-Aldrich, St. Louis, MO) followed by an anti-donkey-anti-IgG secondary antibody (Sigma-Aldrich, St. Louis, MO) conjugated to horseradish peroxidase (HRP) and visualized with SuperSignal West Pico Chemiluminescent Substrate (Thermo Fischer Scientific, Rockford, IL).

Table 1. Strains used in this study.

Strains	Plasmid(s)	Chromosomal Genotype	Reference or source
E. coli			
GT115		*F- mcrA D(mrr-hsdRMS-mcrBC) Φ80lacZDM15 DlacX74 recA1 endA1 Ddcm uidA(DMluI)::pir-116 DsbcC-sbcD*	InvivoGen
CC130		*F-, araD139, delta(ara-leu)7697 delta lac-74 galU galK rpsL*	[74]
KV5066	pEVS104	*lacIQ thi-1 supE44 endA1 recA1 hsdR17 gyrA462 zei-298::Tn10 DthyA::(erm-pir-116) (EmR TcR)*	[63]
V. fischeri			
KV4700	pARM7, pVSV105	WT	Andy Morris
KV4883	pARM7 (*RscS*)	WT	Andy Morris
KV5452	pARM7, pVAR45 ($sypG^{FLAG}$)	Δ*sypE*	Karen Visick
KV5562	pARM7	Δ*sypQ*	Satoshi Shibata
KV5597	pARM7, pVSV105	Δ*sypQ*	Satoshi Shibata
KV6348	pARM7, pKV471 ($sypM^{FLAG}$)	Δ*sypM*	Karen Visick
KV7800	pARM7, pCKM1 ($sypQ^{FLAG1}$)	Δ*sypQ*	This study
KV7801	pARM7, pCKM2 ($sypQ^{FLAG2}$)	Δ*sypQ*	This study
KV7802	pARM7, pCKM3 ($sypQ^{FLAG3}$)	Δ*sypQ*	This study
KV7803	pARM7, pCKM4 ($sypQ^{FLAG4}$)	Δ*sypQ*	This study
MF001	pMKF1 (pCKM1 *syp* promoter)	Δ*sypQ*	This study
MF002	pMKF2 (pMKF1 $sypQ^{W29A}$)	Δ*sypQ*	This study
MF003	pMKF3 (pMKF1 $sypQ^{W29E}$)	Δ*sypQ*	This study
MF004	pMKF4 (pMKF1 $sypQ^{D99A}$)	Δ*sypQ*	This study
MF005	pMKF5 (pMKF1 $sypQ^{D99E}$)	Δ*sypQ*	This study
MF006	pMKF6 (pMKF1 $sypQ^{D100A}$)	Δ*sypQ*	This study
MF007	pMKF7 (pMKF1 $sypQ^{D100E}$)	Δ*sypQ*	This study
MF008	pMKF8 (pMKF1 $sypQ^{D116A}$)	Δ*sypQ*	This study
MF009	pMKF9 (pMKF1 $sypQ^{D116E}$)	Δ*sypQ*	This study
MF010	pMKF10 (pMKF1 $sypQ^{D151A}$)	Δ*sypQ*	This study
MF011	pMKF11 (pMKF1 $sypQ^{D151E}$)	Δ*sypQ*	This study

MF012	pMKF12 (pMKF1 *sypQ*^{D236A})	Δ*sypQ*	This study
MF013	pMKF13 (pMKF1 *sypQ*^{D236E})	Δ*sypQ*	This study
MF1031	pMKF1031 (pMKF1 *sypQ*^{NM})[1]	Δ*sypQ*	This study
MF3241	pMKF3241 (pMKF1 *sypQ*^{NS})[2]	Δ*sypQ*	This study
MF3342	pMKF3342 (pMKF1 *sypQ*^{ES})[3]	Δ*sypQ*	This study
MF3451	pMKF3451 (pMKF1 *sypQ*^{K89Q})	Δ*sypQ*	This study
MF3741	pMKF3741 (pMKF1 *sypQ*^{S304P})	Δ*sypQ*	This study
MF4111	pMKF4111 (pMKF1 *sypQ*^{G322D})	Δ*sypQ*	This study
MF4841	pMKF4841 (pMKF1 *sypq*^{Y18H})	Δ*sypQ*	This study
MF5111	pMKF5111 (pMKF1 *sypQ*^{S153L})	Δ*sypQ*	This study
MF5141	pMKF5141 (pMKF1 *sypQ*^{ES})	Δ*sypQ*	This study
MF5241	pMKF5241 (pMKF1 *sypQ*^{ES})	Δ*sypQ*	This study
MF5551	pMKF5551 (pMKF1 *sypQ*^{L287Q})	Δ*sypQ*	This study
MF5711	pMKF5711 (pMKF1 *sypQ*^{S153L})	Δ*sypQ*	This study
MF5771	pMKF5771 (pMKF1 *sypQ*^{NS})	Δ*sypQ*	This study
MF5961	pMKF5961 (pMKF1 *sypQ*^{NM})	Δ*sypQ*	This study
MF6011	pMKF6011 (pMKF1 *sypQ*^{A216V})	Δ*sypQ*	This study
MF6021	pMKF6021 (pMKF1 *sypQ*^{NM})	Δ*sypQ*	This study
MF6121	pMKF6121 (pMKF1 *sypQ*^{D236N})	Δ*sypQ*	This study
MF6371	pMKF6371(pMKF1 *sypQ*^{ES})	Δ*sypQ*	This study
MF6431	pMKF6431 (pMKF1 *sypQ*^{Y193H})	Δ*sypQ*	This study
MF6441	pMKF6441 (pMKF1 *sypQ*^{V244A})	Δ*sypQ*	This study
MF6551	pMKF6551 (pMKF1 *sypQ*^{ES})	Δ*sypQ*	This study
MF6611	pMKF6611 (pMKF1 *sypQ*^{NS})	Δ*sypQ*	This study
MF6841	pMKF6841 (pMKF1 *sypQ*^{NS})	Δ*sypQ*	This study
MF7341	pMKF7341 (pMKF1 *sypQ*^{ES})	Δ*sypQ*	This study
MF7371	pMKF7371 (pMKF1 *sypQ*^{G307R})	Δ*sypQ*	This study
MF7511	pMKF7511 (pMKF1 *sypQ*^{NS})	Δ*sypQ*	This study
MF8041	pMKF8041 (pMKF1 *sypQ*^{K306E})	Δ*sypQ*	This study
MF8171	pMKF8171 (pMKF1 *sypQ*^{ES})	Δ*sypQ*	This study
MF8261	pMKF8261 (pMKF1 *sypQ*^{ES})	Δ*sypQ*	This study
MF8461	pMKF8461 (pMKF1 *sypQ*^{ES})	Δ*sypQ*	This study
MF8551	pMKF8551 (pMKF1 *sypQ*^{L26P})	Δ*sypQ*	This study
MF8611	pMKF8611 (pMKF1 *sypQ*^{D82N})	Δ*sypQ*	This study
MF8911	pMKF8911 (pMKF1 *sypQ*^{ES})	Δ*sypQ*	This study
MF8951	pMKF8951 (pMKF1 *sypQ*^{NM})	Δ*sypQ*	This study
MF9031	pMKF9031 (pMKF1 *sypQ*^{NM})	Δ*sypQ*	This study
MF9272	pMKF9272(pMKF1 *sypQ*^{G346E})	Δ*sypQ*	This study
MF9361	pMKF9361 (pMKF1 *sypQ*^{N127D})	Δ*sypQ*	This study
MF9461	pMKF9461 (pMKF1 *sypQ*^{H214Y})	Δ*sypQ*	This study
MF9511	pMKF9511 (pMKF1 *sypQ*^{K359E})	Δ*sypQ*	This study

MF9522	pMKF9522 (pMKF1 *sypQ*^{NS})	Δ*sypQ*	This study
MF9751	pMKF9751 (pMKF1 *sypQ*^{H214Y})	Δ*sypQ*	This study
MF9831	pMKF9831 (pMKF1 *sypQ*^{K306E})	Δ*sypQ*	This study
MF10041	pMKF10041 (pMKF1 *sypQ*^{S323L})	Δ*sypQ*	This study
MF10071	pMKF10071 (pMKF1 *sypQ*^{Y180H})	Δ*sypQ*	This study
MF10271	pMKF10271 (pMKF1 *sypQ*^{NS})	Δ*sypQ*	This study
MF10462	pMKF10462 (pMKF1 *sypQ*^{NM})	Δ*sypQ*	This study
MF10831	pMKF10831 (pMKF1 *sypQ*^{NS})	Δ*sypQ*	This study
MF11151	pMKF11151 (pMKF1 *sypQ*^{NS})	Δ*sypQ*	This study
MF11231	pMKF11231 (pMKF1 *sypQ*^{ES})	Δ*sypQ*	This study
MF11261	pMKF11261 (pMKF1 *sypQ*^{Y18C})	Δ*sypQ*	This study
MF11511	pMKF11511 (pMKF1 *sypQ*^{NS})	Δ*sypQ*	This study
MF11751	pMKF11751 (pMKF1 *sypQ*^{Y23H})	Δ*sypQ*	This study
MF11871	pMKF11871 (pMKF1 *sypQ*^{ES})	Δ*sypQ*	This study

[1]NM = no mutation found in gene
[2]NS = not sequenced
[3]ES = early stop codon found in gene

Table 2. Plasmids used in this study.

Plasmids	Description	Reference
pARM7	*rscS* overexpression plasmid; Tet	[59]
pVSV105	Cm Stable *Vibrio* expression vector	
pCKM1	pVSV105 containing epitope-tagged *sypQ* with tag at AA 36 (*lac* promotor only)	This study
pCKM2	pVSV105 containing epitope-tagged *sypQ* with tag at AA 123 (*lac* promotor only)	This study
pCKM3	pVSV105 containing epitope-tagged *sypQ* with tag at AA 223 (*lac* promotor only)	This study
pCKM4	pVSV105 containing epitope-tagged *sypQ* with tag at AA 292 (*lac* promotor only)	This study
pMKF1	pVSV105 with *sypQ*-FLAG (*syp* promotor and *lac* promotor)	This study
pMKF2	pMKF1 containing *sypQ W29A*	This study
pMKF3	pMKF1 containing *sypQ W29F*	This study
pMKF4	pMKF1 containing *sypQ D99A*	This study
pMKF5	pMKF1 containing *sypQ D99E*	This study
pMKF6	pMKF1 containing *sypQ D100A*	This study
pMKF7	pMKF1 containing *sypQ D100E*	This study
pMKF8	pMKF1 containing *sypQ D116A*	This study
pMKF9	pMKF1 containing *sypQ D116E*	This study
pMKF10	pMKF1 containing *sypQ D151A*	This study
pMKF11	pMKF1 containing *sypQ D151E*	This study
pMKF12	pMKF1 containing *sypQ D236A*	This study
pMKF13	pMKF1 containing *sypQ D236E*	This study

Table 3. Primers used in this study.

Gene / promoter	Sequence 5’ – 3’	Primer
sypA promotor F	GCATGCAGCTTCTTCCTTATAGTTATG	1991
sypA promotor R	GTCGACAATAAGCTCCTAGGGAATAATC	1992
sypQ Gibson F	GCCTGCAGGTCGACTCTAGAGAAGGAGCAAATCTATGAATC	1963
sypQ Gibson R	TATAGGGGCGAATTCGAGCTCTTATTTTTGAACTCGAGTCCACG	1964
sypQ-FLAG 1 F	GATTATAAAGATGATGATGATAAA AAAGCAATAAAGCCAACCTGTAG	1955
sypQ-FLAG 1 R	TTTATCATCATCATCTTTATAATC TATTGGGTGACGAGAAGCAAAC	1956
sypQ-FLAG 2 F	GATTATAAAGATGATGATGATAAA TTTGAACATAATCGAGGAAAAGTG	1957
sypQ-FLAG 2 R	TTTATCATCATCATCTTTATAATC ATCTAAAATTTCAAAGTGGGTATC	1958
sypQ-FLAG 3	GATTATAAAGATGATGATGATAAA TCAGATACCATTAATGATGACTTTA	1959
sypQ-FLAG 3 R	TTTATCATCATCATCTTTATAATC CTCTAGATGAATAAATAGCTCGG	1960
sypQ-FLAG 4 F	GATTATAAAGATGATGATGATAAA TACAAAGGAACAGCATTTACCTTT	1961
sypQ-FLAG 4 R	TTTATCATCATCATCTTTATAATC TTTAGGGTTAAACAAGGTAACTAG	1962
sypQ W29A F	ATACTATTACGCGCTTTTGCTTCTCGTCAC	1999
sypQ W29A R	GTGACGAGAAGCAAAAGCGCGTAATAGTAT	2000
sypQ W29F F	ATACTATTACGCTTTTTTGCTTCTCGTCAC	1997
sypQ W29F R	GTGACGAGAAGCAAAAAAGCGTAATAGTAT	1998
sypQ D99A F	GGATGTACTGCTGAC	2077
sypQ D99A R	GTAACAGTGTCAGCAGTACATCC	2078
sypQ D99E F	GGATGTACTGAAGAC	2075
sypQ D99E R	GTAACAGTGTCTTCAGTACATCC	2076
sypQ D100A F	GTACTGATGCCACTGTTACTATTGCTTATAAC	2001
sypQ D100A R	GTAACAGTGGCACTGTTACTATTGCTTATACC	2002
sypQ D100E F	GTACTGATGAAACTGTTACTATTGCTTATAAC	2003
sypQ D100E R	GTAACAGTTTCACTGTTACTATTGCTTATACC	2004
sypQ D116A F	GCTCTATGTGCAGCTACCCACTTTGAAATT	1995
sypQ D116A R	AATTTCAAAGTGGGTAGCTGCACCGATAGC	1996
sypQ D116E F	GCTCTATGTGCAGAAACCCACTTTGAAATT	1993
sypQ D116E R	AATTTCAAAGTGGGTTTCTGCACCGATAGC	1994

sypQ D151A F	GCTCTGAGTGCTGTTTCAGCC	2085
sypQ D151A R	GACGGCTGAAACAGCACTCAGAGC	2086
sypQ D151E F	GCTCTGAGTGAAGTTTCAGCC	2083
sypQ D151E R	GACGGCTGAAACTTCACTCAGAGC	2084
sypQ D236A F	CCATTAATGATGCTTTTATTTTAC	2081
sypQ D236A R	GTAAAATAAAAGCATCATTAATGG	2082
sypQ D236E F	CCATTAATGATGAATTTATTTTAC	2079
sypQ D236E R	GTAAAATAAATTCATCATTAATGG	2080

CHAPTER THREE

EXPERIMENTAL RESULTS

The marine bacterium *Vibrio fischeri* is the natural symbiont of the *Euprymna scolopes* squid. Initiation of symbiosis is dependent on the production of a biofilm matrix composed of the Syp exopolysaccharide (EPS)[53]. The gene locus in *V. fischeri* that encodes the machinery responsible for this EPS component is the symbiotic polysaccharide (*syp)* locus [8]. This locus contains 18 genes (*sypA* through *sypR)* that encode regulatory proteins, putative polysaccharide exporter proteins, and putative glycosyltransferases. In *V. fischeri,* SypQ has been identified as critical for the formation of wrinkled colonies as well as symbiotic colonization [8, 53]. The mechanism by which SypQ contributes to biofilm formation in *V. fischeri* is currently unknown. The goal of this study was to better understand the role of SypQ in biofilm formation. By using both a bioinformatics approach and a genetic approach, I intend to identify specific amino acids that are important to the ability of the protein to contribute to biofilm formation. If an amino acid is necessary for the function of SypQ, then a mutation changing that amino acid will result in a loss of function, as indicated by a smooth colony phenotype.

Bioinformatic Analysis of the *sypQ* Gene

SypQ is a putative glycosyltransferase [7, 8], but the specific function of this protein is unknown. There are numerous families of glycosyltransferases with different mechanisms of

action. Thus, to better understand the function of SypQ, I used a bioinformatic approach to determine the family of GTs to which SypQ belongs, identify the presence of conserved motifs, and generate a possible structure for SypQ. The *sypQ* sequence was first analyzed using NCBI BLAST [68], which compares the input sequence to sequences in multiple databases to identify regions of similarity. The result suggested that the *sypQ* gene encodes a family 2 glycosyltransferase (GT-2) with a putative DxD motif (Figure 3). The prediction also categorizes SypQ as a CesA-like protein. The CesA proteins are cellulose synthase-like proteins. These proteins all share the common characteristic of elongating polysaccharide chains. Finally, the program also indicated that SypQ belongs to the glycosyltransferase GTA-type superfamily (glyco_tranf_GTA_type). All of the members of this superfamily contain a specific structural similarity, namely, the GT-A fold. Most of the GTs that belong to this superfamily come from the family 2 glycosyltransferases.

I next investigated the structure of SypQ by submitting the amino acid sequence to the Phyre 2 bioinformatics server. Phyre 2 utilizes homology modeling, also known as comparative modeling. Homology modeling aligns the amino acid sequence of the submitted protein sequence, or query, with specific residues of a known and characterized structure. The structure on which the query sequence is threaded is chosen based on percent homology. If the query sequence has ≥50% homology to the target sequence, the accuracy of the prediction increases [75]. Once the query sequence has been submitted, the server will align the sequence to multiple target structures and then rank them based on a raw score that considers both sequence and secondary structure similarities.

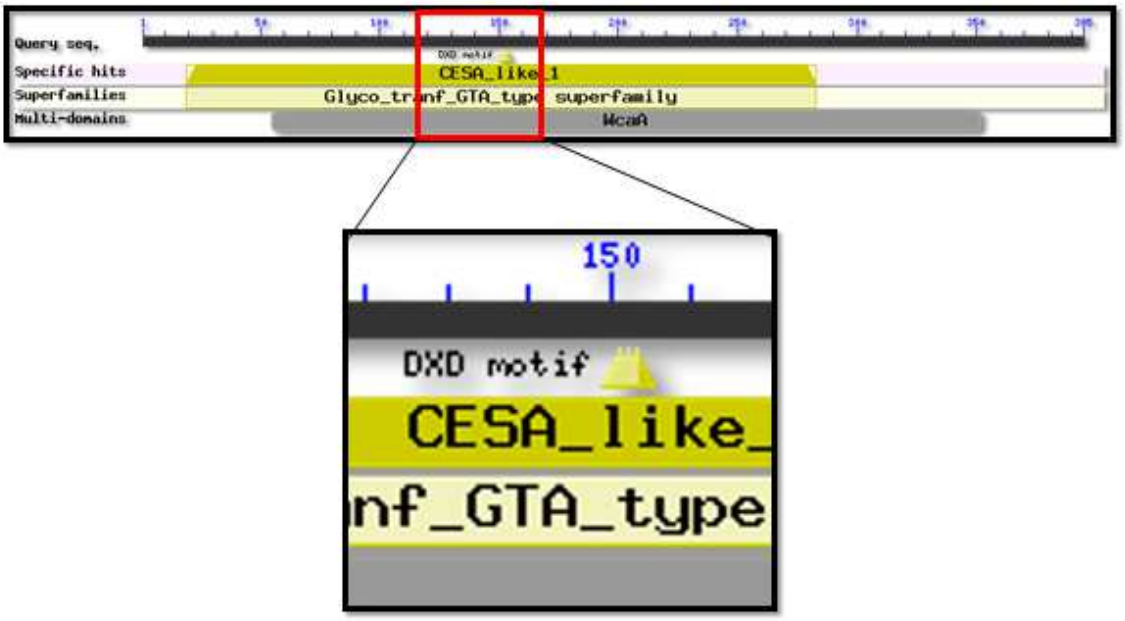

Figure 3. NCBI BLAST prediction. After submitting the *sypQ* sequence to the server, the results suggested that SypQ is likely a family two glycosyltransferase and contains a putative DxD motif. The region of SypQ that contains the putative DxD motif is boxed in red and enlarged to show detail [68].

The top three returns for the SypQ sequence were structures for glycosyltransferases: cellulose synthase subunit A (BcsA), mannosyl-3-phosphoglycerate synthase (MngA), and polypeptide *N*-acetylgalactosaminotransferase 1(GALNT1) (Figure 4). All three of the results had a confidence of 100.0. The confidence number indicates, on a scale from 1-100, the confidence that the query sequence and target structure are homologous. The protein structure that the highest percentage of the SypQ query sequence aligned with it was the cellulose synthase subunit. Although SypQ had an overall low percentage of sequence homology with all three results, this does not necessarily equate to an inaccurate structural prediction. Previous work has demonstrated that protein structure is more highly conserved than primary sequence [76]. This suggests that distantly related proteins with low sequence homology can still have

significant structural homology and thus it is likely that this modeling can reveal insight into SypQ.

The predicted ribbon structure resulting from the threading of the SypQ amino acid sequence onto the BcsA structure depicts a globular protein containing multiple alpha helices and a central region of comprised of beta sheets (Figure 5). The Phyre 2 structural analysis results combined with the NCBI BLAST prediction highly suggests that SypQ is a BcsA-like family 2 glycosyltransferase. Because this family of proteins elongates polysaccharide chains via an inverting mechanism, the similarity of SypQ to BcsA could indicate that SypQ functions in a similar manner to contribute to the production of the exopolysaccharide component of the *V. fischeri* biofilm matrix.

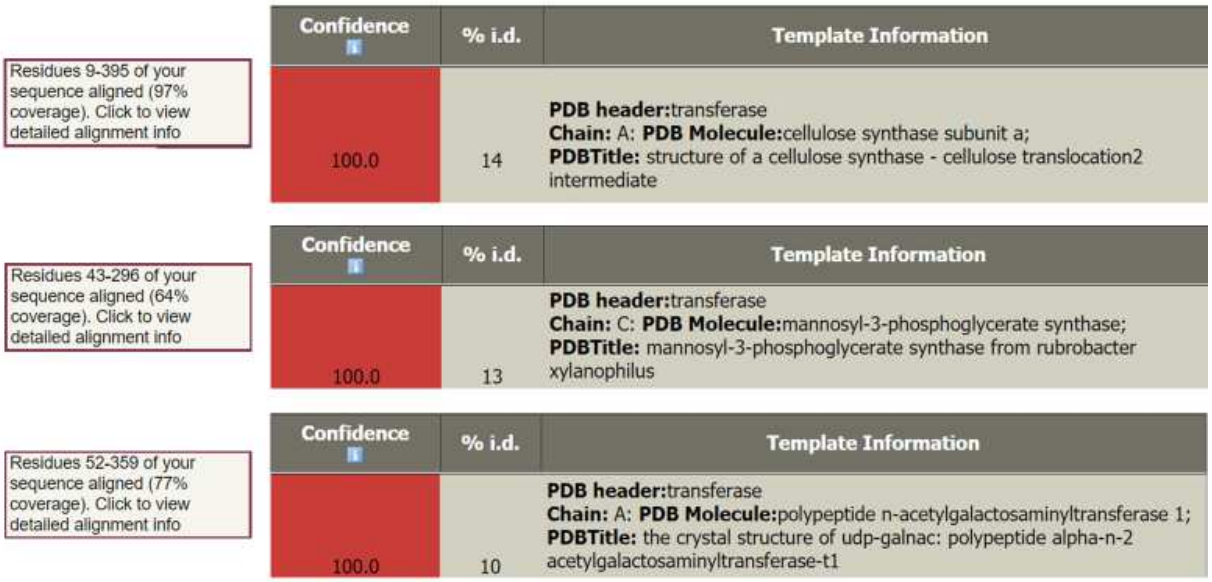

Figure 4. Phyre 2 prediction results. The top three results from the Phyre 2 software are indicated with the confidence number in the red box. The percent identity of the query sequence that aligns with the resulting sequence is in the second box. The protein to which the query sequence was aligned is in the box titled template information. Boxed on the left are the percentage of the query sequence that aligned with the structure of the resulting protein [71].

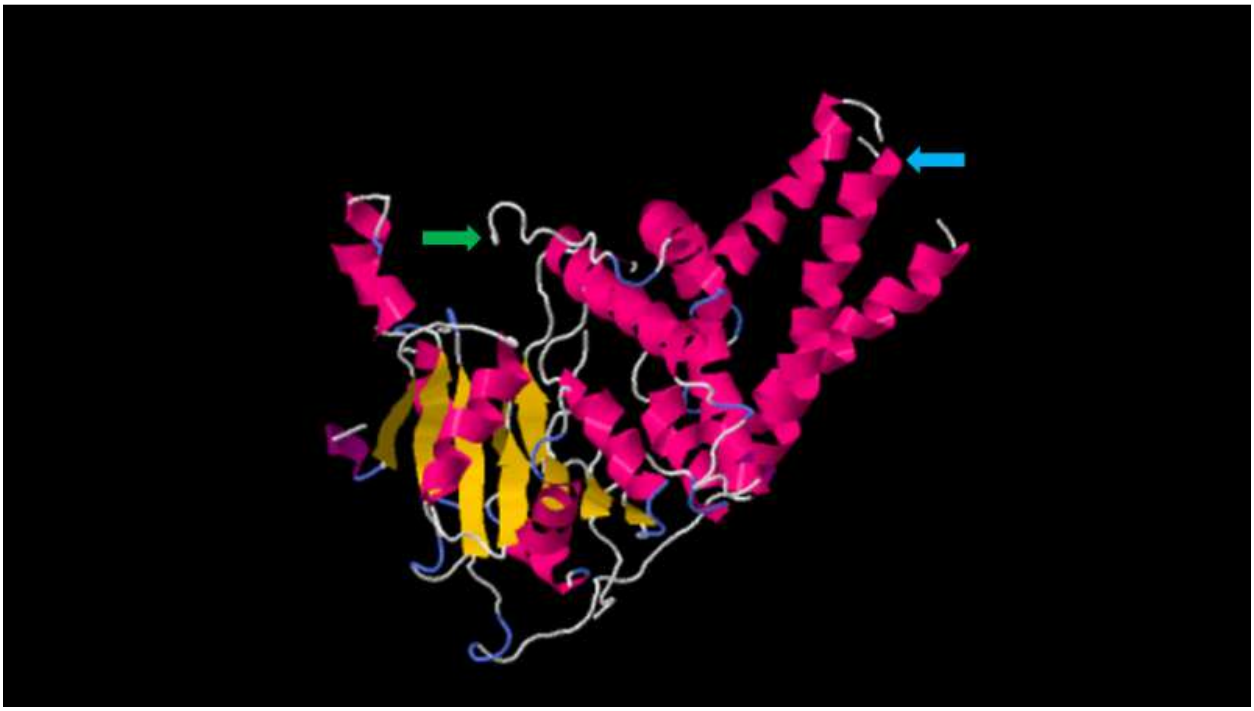

Figure 5. The predicted structure of SypQ obtained by submitting the SypQ amino acid sequence to the Phyre 2 bioinformatics server. The predicted alpha helices are depicted in pink while predicted beta sheets are depicted in yellow. The N-terminal end of the protein is indicated by the blue arrow and the C-terminal end is indicated by the green arrow [71].

To better understand SypQ, I compared it to BcsA. In *V. fischeri,* BcsA is composed of 870 amino acid residues and contains 9 putative transmembrane helices while SypQ contains 395 amino acids and an unknown structure. I utilized protein prediction bioinformatic programs to elucidate the topology of SypQ. First, I submitted the SypQ amino acid sequence to the PredictProtein software. PredictProtein is a sequence based prediction tool that compares the query sequence to multiple sequences from multiple servers to predict secondary structure and transmembrane regions within the protein. The prediction suggested that SypQ contains five transmembrane helices (Figure 6). Second, I submitted the SypQ sequence to the TOPCONS software. The TOPCONS software uses five different topology algorithms to generate a consensus prediction. The prediction generated by the SypQ amino acid sequence suggested that

the protein has four transmembrane helices (Figure 7). The bioinformatics results suggest that SypQ is a membrane-integrated GT-2 and that it potentially shares properties with the well-characterized GT-2 BcsA. Multiple studies have shown GT-2s to be integrated into both eukaryotic and bacterial membranes [77, 78]. Given the bioinformatics results and our knowledge that SypQ contributes to the formation of the *V. fischeri* biofilm exopolysaccharide, I hypothesize that SypQ executes its role in exopolysaccharide synthesis from a location in the inner membrane.

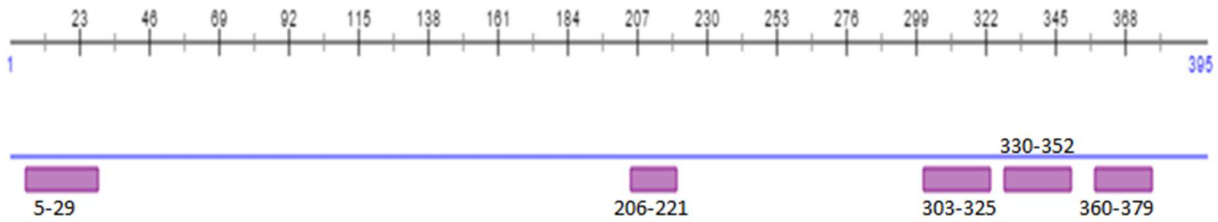

Figure 6. Putative SypQ transmembrane regions predicted by the PredictProtein server. The amino acid position is depicted at the top and the predicted transmembrane regions are depicted below as purple bars. The amino acids predicted to be in the TM regions are indicated [69].

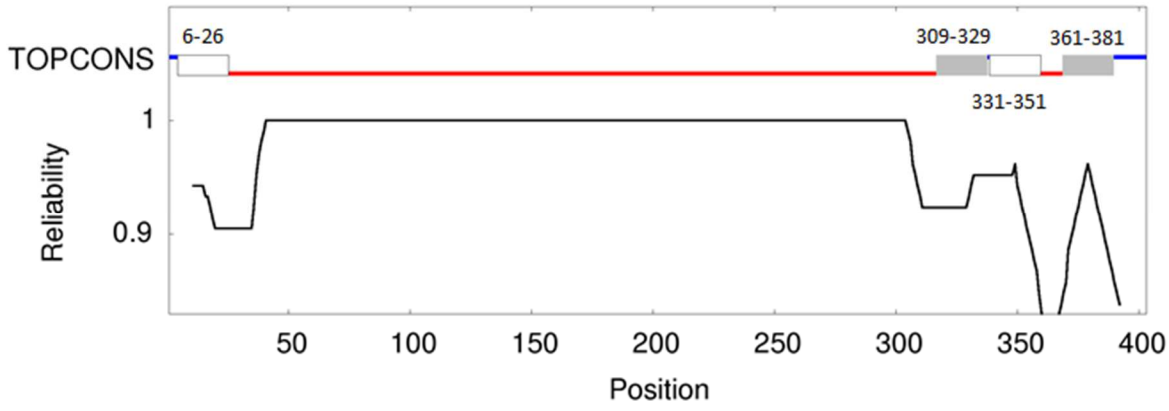

Figure 7. Putative SypQ transmembrane regions as predicted by the TOPCONS server. Regions of the protein predicted to be outside of the cytoplasm are depicted in blue while regions predicted to be inside the cytoplasm are depicted in red. Transmembrane regions thought to traverse the membrane in an outside to inside direction are shown as white boxes and those thought to traverse the membrane in an inside-out direction are shown as grey boxes. The amino acids predicted to be in the TM regions are indicated [70].

Visualization of Tagged SypQ Via Western Blot Analysis

The contribution that SypQ makes to the production of the exopolysaccharide component of the *V. fischeri* biofilm matrix is remains unknown. To more fully understand how SypQ promotes biofilm formation it is important to be able to visualize SypQ. Visualization of SypQ can answer a variety of questions about the protein. For example, what are the relative SypQ protein levels in the cell and does the amount of SypQ present correlate to a change in biofilm phenotype? Is the protein being processed (cleaved) in the cell and is the processing essential to its function? Is the protein interacting/oligomerizing with other proteins? Where is SypQ located in the cell? Thus, one of my goals was to develop tools to visualize SypQ via Western blot analysis.

To visualize SypQ by Western blot analysis, I first needed to generate a functional epitope-tagged version of SypQ. A previous attempt to tag the protein placed a FLAG epitope tag at the C-terminal end of the protein. This construct resulted in a protein that was unable to restore biofilm formation in a *sypQ*-deficient *V. fischeri* strain. As a result, it was necessary to insert the epitope tag at an alternate position, one that would permit SypQ to retain its function. The ProteinPredict-generated prediction of SypQ's topology (Figure 6) was used to identify regions of the protein that would be potentially permissive to insertion, *i.e.* regions that lacked putative transmembrane sequences, buried residues, and predicted protein-protein interactions. Four regions within the protein were identified as possible locations for an internal epitope tag: amino acid 36, 123, 223, and 292. *sypQ* constructs that fused a FLAG epitope tag at those positions were generated using PCR SOEing. Sequence analysis confirmed the presence of the epitope tag within *sypQ*.

To determine whether the constructs encoded functional proteins, they were tested for the ability to complement a *sypQ*-deficient strain of *V. fischeri*. If the epitope-tagged constructs encoded functional proteins, then they should restore biofilm formation to a *sypQ*-deficient strain. Specifically, I analyzed biofilm formation using the wrinkled colony assay. Only two of the four constructs were able to restore biofilm to the deficient strain (Figure 8). The constructs containing the tag at amino acids 223 and 292 resulted in a loss of function as indicated by a smooth colony phenotype. The constructs containing the epitope tag at amino acids 36 and 123 restored biofilm formation as indicated by the wrinkled colony phenotype and colony stickiness, a characteristic that has been attributed to the presence of Syp polysaccharide [7, 8, 79]. While the strain containing SypQ with the epitope tag at amino acid 123 had a noticeable delay in biofilm formation, the strain containing SypQ with the tag at amino acid 36 displayed a biofilm formation similar to that of the positive control. This suggests that the insertion of the FLAG epitope tag at this region of the protein has a minimal effect on the protein's ability to promote biofilm formation in *V. fischeri.* As a result, all experiments moving forward were completed using the construct with the tag at amino acid 36.

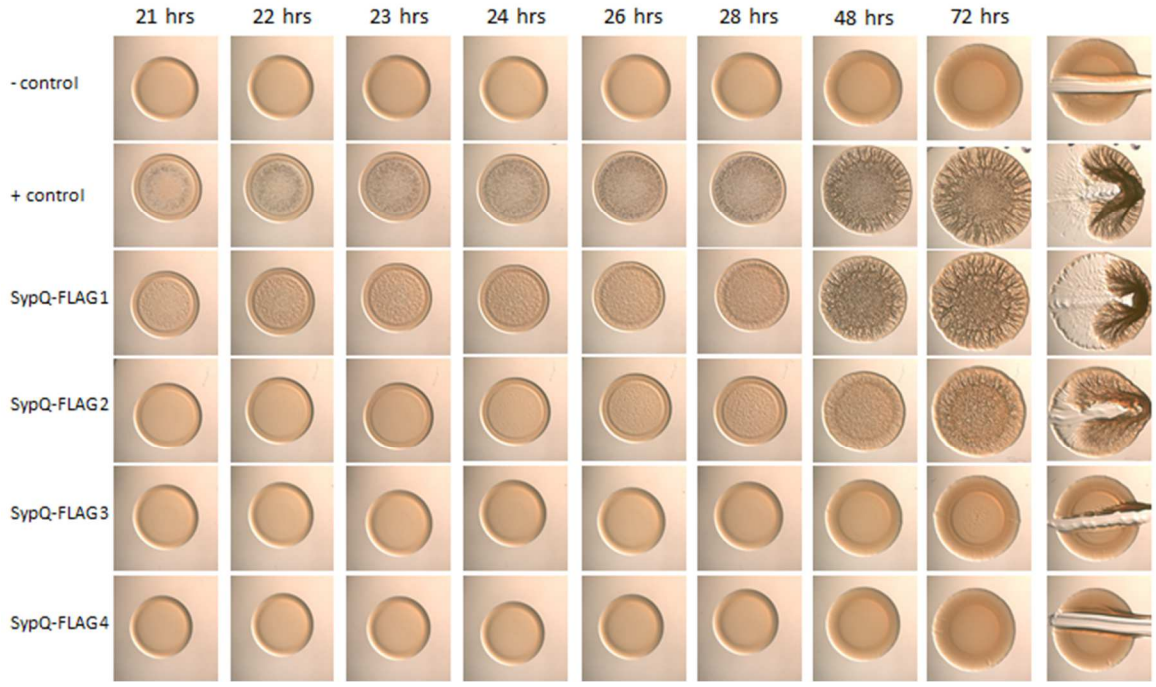

Figure 8. Wrinkled colony assay to determine functionality of tagged SypQ costructs. The four epitope-tagged *sypQ* constructs were introduced into *sypQ* deficient *V. fischeri* via triparental conjugation. 10 µl of diluted cell culture standardized to an OD_{600} of 0.2 was spotted onto LBS media containing Cm and Tet. The plates were incubated at 24°C and pictures were taken at the time points indicated. Only the *sypQ* constructs with the tag inserted at amino acid 36 (SypQ-FLAG #1) and 123 (SypQ-FLAG #2) restored biofilm formation to the pRscS *sypQ* deficient strain of *V. fischeri* (KV5562). The *sypQ* constructs with the tag inserted at amino acid 223 (SypQ-FLAG #3) and 292 (SypQ-FLAG #4) did not restore biofilm formation. The negative control is strain KV5597 and the positive control is strain KV5709. The experimental strains are 7800-7803 respectively.

Next, I sought to visualize the epitope-tagged SypQ via Western blot using antibodies against the FLAG epitope. Having successfully generated a functional epitope-tagged *sypQ* allele, I first attempted to visualize the tagged SypQ protein using a Western blot approach. After introducing the plasmid containing the epitope-tagged construct into *V. fischeri* strain KV5562, which is a *V. fischeri* strain deficient in *sypQ,* the cells were grown overnight and lysed. Cells expressing either the epitope-tagged SypQ construct or, as controls SypM-FLAG or SypG-FLAG, were analyzed by Western blot using an anti-FLAG antibody (Figure 9). The epitope-tagged controls (SypG and SypM) were clearly visible (55 kD and 26 kD respectively); however,

SypQ, with an expected molecular weight of approximately 46 kD, was not visualized by this method. We hypothesized that low expression of the protein is a possible reason why SypQ was not visualized by Western blot. In this construct, *sypQ* transcription is driven solely by the *lac* promotor.

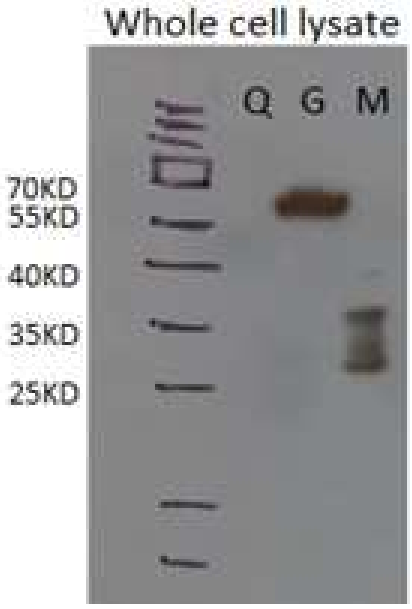

Figure 9. Western blot analysis of epitope-tagged SypQ construct. Tagged proteins SypQ, SypG, and SypM were separated using SDS-PAGE and transferred to a PVDF membrane. An immunoblot assay was performed to visualize the proteins using anti-FLAG primary and anti-IgG secondary antibodies. The epitope-tagged controls, SypM and SypG were clearly visible but the epitope-tagged SypQ protein was not. The position of the bands of the protein standards are indicated by lines and a box on the left-hand side of the blot, and specific sizes for a subset are indicated to the left of the blot.

To overcome the possible issue of low expression of SypQ, I inserted the *sypA* promotor upstream of the epitope-tagged *sypQ* allele. The *sypA* promotor is the first and best characterized of the four promotors within the *syp* locus. Transcription of the *sypA* promoter is controlled by the transcription factor SypG [54, 56]. This plasmid (pMKF1) was introduced into *V. fischeri* strain KV7720, which is deficient for both *sypQ* and the negative regulator *sypE*. The strain also overexpresses the *syp* transcription factor SypG from plasmid pCLD56. The absence of the

negative regulator and an abundance of the native transcription factor should increase the level of transcription from the *sypA* promoter, potentially resulting in higher levels of protein and an increase in signal. However, under these conditions. I was unable to visualize SypQ via Western blot (data not shown). It is known that SypQ participates in biofilm formation. I hypothesized that by using conditions that induce biofilm formation, I could increase the transcription of *sypQ* and, therefore, increase the levels of SypQ present in the cell. Thus, I induced biofilm formation in liquid cultures via the addition of 10 mM or 20 mM of $CaCl_2$ (Lie, Tischler, and Visick submitted). Cells within the *syp*-dependent clump that contained the SypQ-FLAG protein or the positive control, SypM-FLAG, were lysed and analyzed by Western blot using an anti-FLAG antibody. The epitope-tagged SypQ protein was not visible in any of the lanes, regardless of $CaCl_2$ concentration.

Another possibility for the lack of visible protein is that *sypQ* transcription may only occur early after biofilm induction, resulting in the production of protein at time points that are earlier than I have assessed to date. To determine if this were the case, the assay was repeated and samples were taken and processed at hourly intervals for eight hours after the addition of calcium (Figure 10). Similar to previous results, SypQ was not visible at any of the intervals.

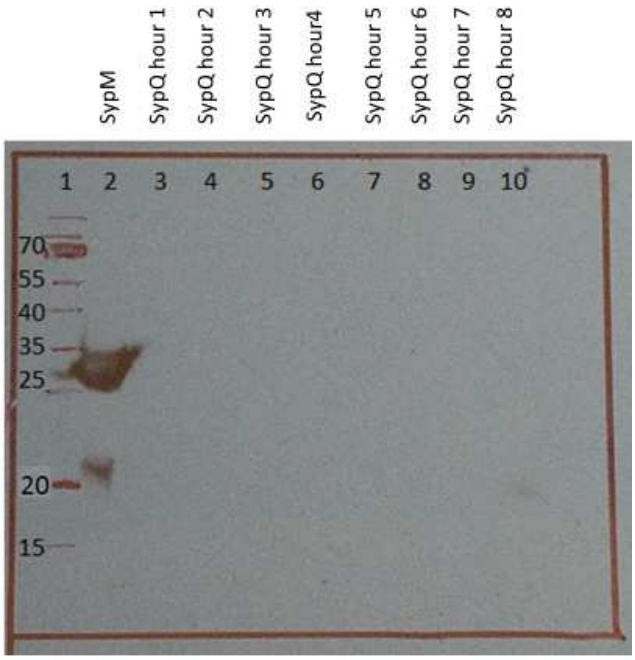

Figure 10. Western blot of epitope-tagged SypQ at hourly time points. *V. fischeri* cells carrying the epitope-tagged *sypQ* construct were treated with 20 mM calcium chloride. At one-hour intervals for 8 hours, 1.5 ml of cell culture was centrifuged and lysed via sonication. The proteins from the lysate were separated via SDS-PAGE and visualized by Western blot using an anti-FLAG primary antibody and an anti-IgG secondary antibody. The lanes are numbered one through ten at the top with the corresponding time point above. The position of the bands of the protein standards are indicated by lines and a box on the left-hand side of the blot, and specific sizes for a subset are indicated to the left of the blot.

Given my hypothesis that SypQ functions from a location in the inner membrane, I performed cell fractionation experiments to concentrate SypQ in the inner membrane fraction and thus amplify the signal. Using the fractionation technique described in the methods section, lysates of cells containing the epitope-tagged construct were separated into cytoplasmic, inner membrane, periplasmic, and outer membrane fractions, and then analyzed by Western blot. If SypQ is located in the inner membrane, it should be concentrated in that fraction and the signal will be amplified. This should allow for visualization. The epitope-tagged SypG and SypM proteins were used as a positive control for the cytoplasmic and membrane fractions respectively, but were instead visible in the wrong fractions (SypG was present in the inner membrane and

periplasmic fractions while SypM was detected in the periplasm), suggesting that the fractionation was not clean. Regardless, the epitope-tagged SypQ protein was still not visible in any of the fractions (Figure 11).

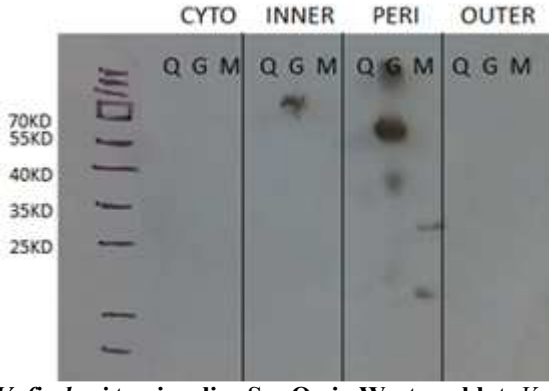

Figure 11. Cellular fraction of *V. fischeri* to visualize SypQ via Western blot. *V. fischeri* cells carrying the epitope-tagged *sypQ* allele were separated into the cytoplasmic, inner membrane, periplasmic, and outer membrane fractions using the described protocol. The fractions were resolved by SDS-PAGE and the proteins were visualized by Western blot using an anti-FLAG primary antibody and an anti-IgG secondary antibody. The lanes containing epitope-tagged SypQ, SypG, and SypM are labeled Q, G, and M respectively. The different cell fractions [cytoplasmic (cyto), inner membrane (inner), periplasm (peri), and outer membrane (outer)] are labeled above. The position of the bands of the protein standards are indicated by lines and a box on the left-hand side of the blot, and specific sizes for a subset are indicated to the left of the blot.

Finally, I attempted an enrichment technique to isolate SypQ: rather than separating the cells into the four fractions, I attempted to isolate a rough membrane fraction by utilizing a rough membrane extraction protocol. The simplified protocol could minimize the loss of product, leading to a more concentrated SypQ sample. The process was performed with 5 ml, 10 ml, and 30 ml of overnight culture. Extract from the lower 5 ml cell volume showed no signal on the gel, while extracts from the larger cell volumes displayed a large smear throughout the gel (Figure 12). In this assay, the tagged SypM positive gave no visible signal, suggesting that the cellular preparation was not optimal. Visualization of the tagged SypQ protein was unsuccessful. In the Discussion, I will address some possible reasons why the protein was not detectible by Western blot.

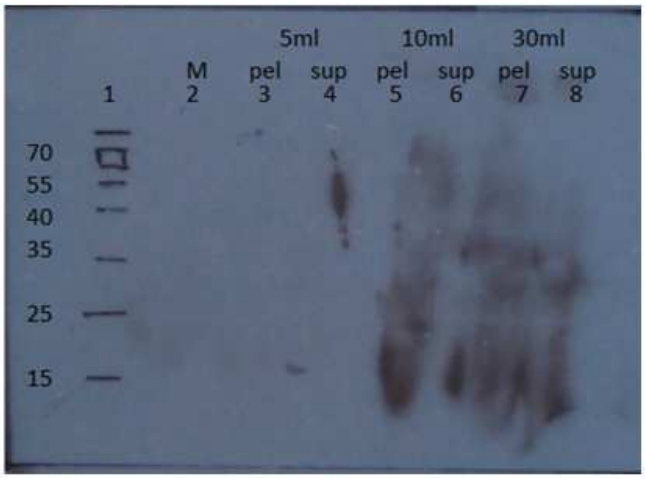

Figure 12. Western blot of rough membrane fraction to visualize SypQ. Different volumes of *V. fischeri* cells carrying the epitope-tagged *sypQ* allele (5 ml, 10 ml, or 30 ml) were separate into a pellet containing the rough membrane fraction and the supernatant. The proteins from both the pellet (pel) and the supernatant (sup) were separated by SDS-PAGE and visualized via Western blot using an anti-FLAG primary antibody and an anti-IgG secondary antibody. The epitope tagged SypM control is labeled M in lane 2. Lanes 3-8 contain proteins from the supernatant or pellet of the different volumes (indicated above). The position of the bands of the protein standards are indicated by lines and a box on the left-hand side of the blot, and specific sizes for a subset are indicated to the left of the blot.

Site-Directed Mutagenesis of *sypQ*

Identification of Putative Functional Motifs in SypQ.

Despite my inability to visualize SypQ, I proceeded to investigate this protein via a structure/function analysis. This approach will lay the groundwork for an understanding of SypQ's function. I began by examining the amino acid sequence for the presence of potential conserved residues. To identify essential amino acid residues in SypQ, I utilized the previous bioinformatics findings and primary sequence alignments with characterized GT-2 enzymes. The Phyre 2 results suggested that SypQ has a high amount of structural similarity with the family 2 glycosyltransferase protein BcsA, despite a low amount of sequence similarity. Previous research identified functional motifs that are essential to the function of family 2 glycosyltransferases [9, 13, 19]. There are four particular sequence motifs that contribute to the function of this family of

GTs: DDx, DxD/DxH, TED, and Q/LxxRW motifs each have roles critical to the function of the enzyme. I hypothesized that if SypQ is a GT-2, then it may contain some or all of these functional motifs. To determine if any of these motifs were present, I compared SypQ and BcsA. Specifically, I aligned the SypQ sequence from the *V. fischeri* strain ES114 with the BcsA sequence from *V. fischeri, E. coli,* and *S. typhimurium* (Figure 13) by submitting all of the amino acid sequences to the Clustal omega server [66]. I then examined the alignment to identify the motifs in BcsA and to determine if one or more of the motifs were present in SypQ. As described in more detail below, I identified regions of SypQ sequences that closely resembled known GT-2 functional motifs, including potential DDX, DxD/DxH, TED, and Q/LxxRW sequences. Although not fully conserved, the presence of these motifs is consistent with my hypothesis that SypQ functions as a GT-2.

DDx has been shown to be critical for the ability of GT-2s to catalyze the formation of both α- and β- glycosidic bonds [17]. The DDx motif is located adjacent to the DxD motif in the active site of the 3-dimensional structure, where it interacts with the anomeric side of the donor sugar. In my analysis of the SypQ sequence, I found a sequence, TDDT (residues 98-101, highlighted in yellow in Figure 13), that could serve as either a TED (TDD) or as a DDx (DDT) motif. While neither putative motif specifically aligns with the known motif of BcsA, the sequence is located very close to a known DDx motif in BcsA (DDG). As a result, I hypothesize that one or both of the aspartates (D99 and D100) may be important for SypQ function.

The role of the DxD functional motif has been well characterized in BcsA. This motif is present in the active site, where it interacts with the phosphate of the nucleotide donor molecule and coordinates the divalent cations needed for the function of BcsA [11, 19, 20]. As described

above, my original BLAST alignment identified a DxD motif (Figure 3). However, when I align SypQ with BcsA, the corresponding sequence motif in SypQ, located at amino acids D151-S153, is DVS (highlighted in blue in Figure 13). Although the sequence in SypQ is not an exact match to the corresponding sequence in the BcsA sequences, it is located at the same location and the primary aspartic acid is highly conserved. For these reasons, I hypothesize that the DVS sequence in SypQ is, in fact, functioning like the DxD motif. If so, then I predict that the aspartic acid residue (D151) will be essential to the function of SypQ. Experiments designed to test this possibility are described below.

As mentioned in chapter one, the DxD motif is sometimes replaced with a DxH. A putative DxH sequence is located from amino acids 116 to 118 (D116, T117, D118) in SypQ. While the DTH sequence in SypQ is not located at or near any known functional motifs in BcsA, when located on the Phyre-predicted ribbon diagram, the D116 residue is located very close to the beta sheets that make up the predicted active site (Figure 14). I hypothesize that, because of its closeness to the predicted active site, the SypQ DTH sequence may contribute to the function of the protein.

Figure 13. Alignment of SypQ with BcsA. The SypQ amino acid sequence from *V. fischeri* was aligned with the BcsA amino acid sequence from *V. fischeri* (Q5DZ42-1), *E. coli* (P37653-1), and *S. typhimurium* (Q93IN2-1). SypQ sequence with similarity to known GT-2 functional motifs are underlined. The amino acid residues targeted for mutagenesis are in bold. There are multiple amino acid residues in SypQ that are conserved in BcsA. These residues are designated by an asterisk (*). The known functional motifs in BcsA are highlighted in yellow (DDX), blue (DxD), grey (TED), and green (Q/LxxRW) with the name of the motif printed above underlined and in bold. All sequences attained from Uniprot [80].

The next motif of interest is TED. The TED motif interacts with the acceptor molecule growing polysaccharide chain through hydrogen bonding [11, 20]; this motif sequence is somewhat variable in cellulose synthase-like proteins, sometimes being replaced with TDD. I identified a sequence, NDDF (residues 234-237, highlighted in gray in Figure 13) that could function as either a TED motif (TDD) or a DDx motif (DDF). Although the TED/TDD sequence is not fully conserved, it is positioned at the same location as the known TED in BcsA. Due to

the alignment of the NDD sequence with the TED/TDD sequence of BcsA, I hypothesize that D236 is important to SypQ function.

Finally, the Q/LxxRW motif in BcsA is essential and may partner with the DxD motif residues to facilitate cellulose synthesis. In this motif, the tryptophan residue is located directly above the active site and interacts with the accepter end of the polymer chain while the arginine residue, like the aspartic acid in the DxD motif, is in contact with the donor molecule's phosphate group [13, 14, 20]. In BcsA, the motif is located towards the C-terminal end of the protein and is located above the active site in the 3-dimensional structure. No such sequence is present in the C-terminus of SypQ as SypQ ends prior to the established Q/LxxRW sequence in BcsA (Figure 13). However, a similar sequence, ILLRW (I25-W29), is located near the N-terminal end of the protein. Given the different position, it is not clear if this sequence in SypQ would function as it does in BcsA, however, the motif sequence is mostly conserved: the arginine and tryptophan match the sequence of the known motif and there is a conservative change of the leucine residue to an isoleucine. Due to the relative conservation of the sequence, I chose to investigate the possibility that this sequence in SypQ represents a true Q/LxxRW motif. If this is a true Q/LxxRW motif, then I hypothesize that the tryptophan residue contributes to the function of SypQ.

I then aligned the SypQ amino acid sequence with sequences from other processive GT-2s (Figure 14). I aligned SypQ with NodC from *Rhizobium meliloti* and AscA from *Komagataeibacter xylinus.* NodC is a glycosyltransferase essential for chitin synthase and AscA is utilized for cellulose synthase. As with BcsA, the known sequence motifs are present in both NodC and AcsA.

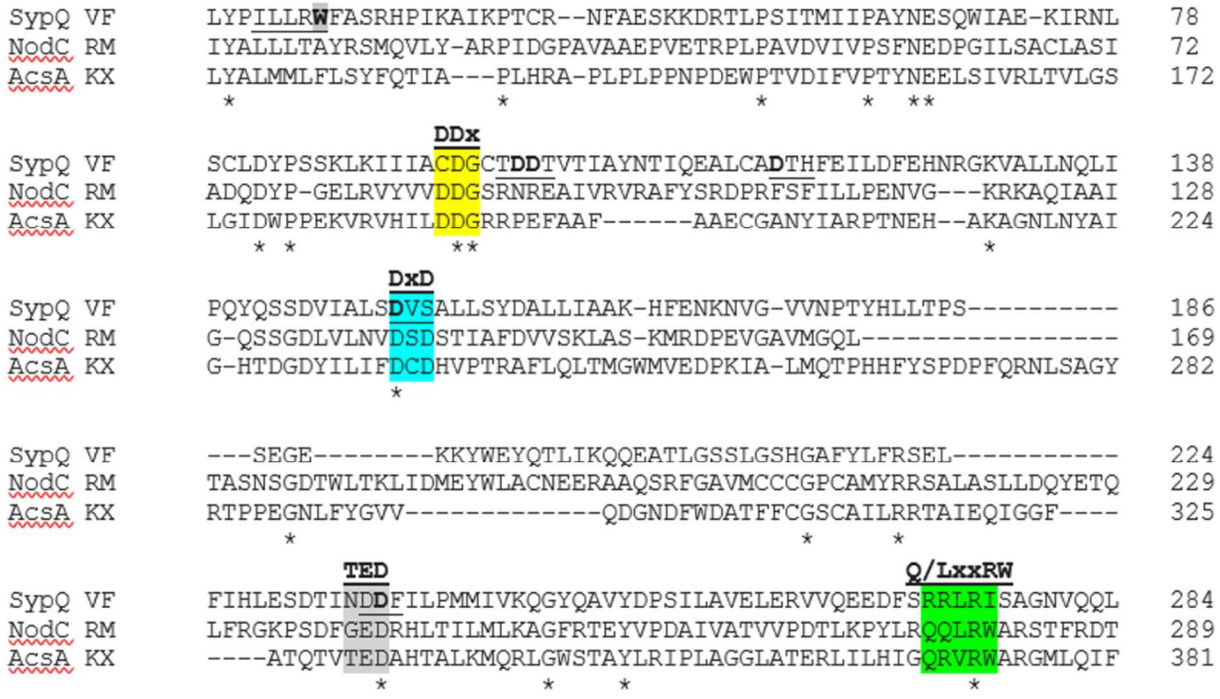

Figure 14. Alignment of SypQ with NodC and AcsA. The SypQ amino acid sequence from *V. fischeri* was aligned with the AcsA amino acid sequence from *K. xylinus* (P0CW87), and NodC from *R. meliloti* (P04341). SypQ sequence with similarity to known GT-2 functional motifs are underlined. The amino acid residues targeted for mutagenesis are in bold. There are multiple amino acid residues in SypQ that are conserved in AcsA and NodC. These residues are designated by an asterisk (*). The known functional motifs in BcsA are highlighted in yellow (DDX), blue (DxD), grey (TED), and green (Q/LxxRW) with the name of the motif printed above underlined and in bold. All sequences attained from Uniprot [80].

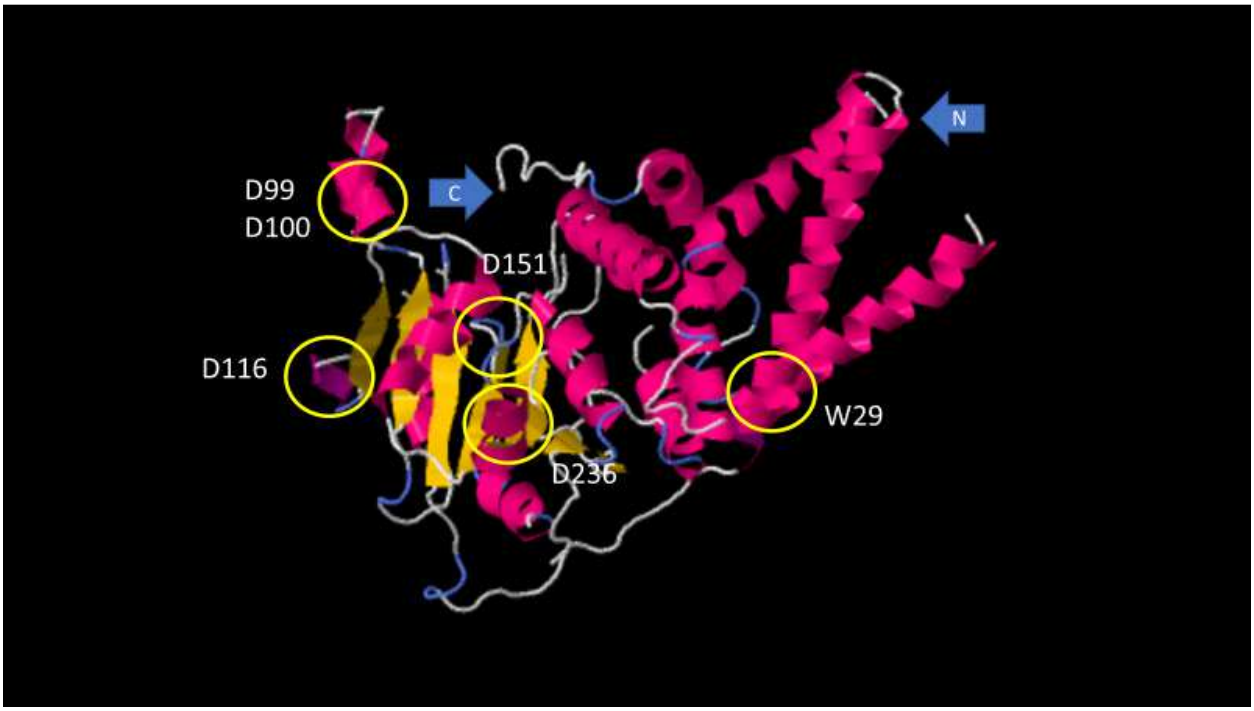

Figure 15. Location of putative functional motifs on predicted SypQ structure. The predicted alpha helices are depicted in pink while predicted beta sheets are depicted in yellow. The N-terminal and C-terminal ends are indicated by arrows. The location of the putative functional motifs is circled in yellow and the targeted amino acids are identified in white text.

I tested my hypotheses that the putative motifs identified in SypQ do, in fact, represent functional motifs by identifying residues to be targeted for site-directed mutagenesis and then designing and generating SypQ mutants. To determine if the targeted residues in the putative functional motifs are essential to the function of SypQ, I generated *sypQ* alleles containing single point mutations at the indicated amino acid residues (Figure 15). For the targeted aspartic acid residues, I hypothesized that the charge of the residue is necessary for the function of the protein. If so, then changing the aspartic acid to glutamic acid should result in little or no change in biofilm formation while an alanine substitution should result in a severe biofilm defect. Therefore, I designed my experiments to change each targeted aspartic acid to an alanine and a

glutamic acid. For the targeted tryptophan residue, I hypothesized that the large side chain of tryptophan may be important to its function. If that is true, then changing it to phenylalanine (which also has a large side chain) would result in little to no loss of biofilm formation, while an alanine substitution would result in partial to total loss of biofilm. Thus, for W29, I designed experiments to generate both alanine and phenylalanine substitutions.

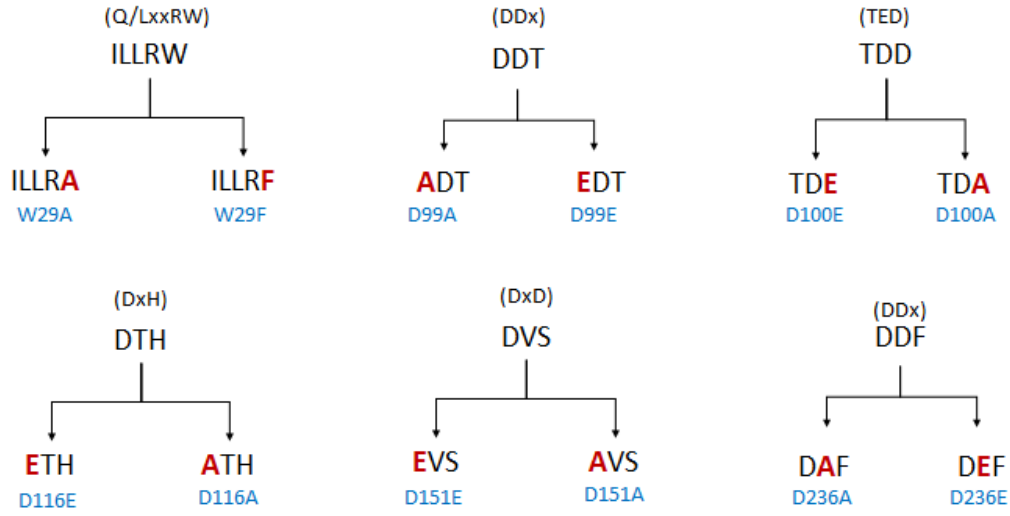

Figure 16. List of amino acids targeted for site-directed mutagenesis and proposed mutations. At the top of each pyramid is the SypQ sequence similar to known GT-2 functional motifs with its corresponding motif sequence in parentheses. Below that, the suggested residue substitutions are indicated in red and the name of the mutation is in blue.

Generation and Evaluation of Site-Directed Mutants.

To introduce the proposed amino acid substitutions into the epitope-tagged *sypQ* allele, I designed primers containing the codon changes and used PCR SOEing to generate the mutated alleles. Then, I ligated the mutated sequences into pMKF1 in place of the wild type *sypQ* and introduced the resulting plasmid into the KV5562, a *sypQ* mutant strain that also contains plasmid pARM7, which overexpresses the positive biofilm regulator RscS: the two plasmids are

fully compatible. Finally, I determined whether the mutated *sypQ* allele could restore biofilm formation to the *sypQ* deficient strain using the wrinkled colony assay (Figure 16). Wrinkled colony architecture and colony cohesion (stickiness) are used as a readout for biofilm formation and polysaccharide production respectively.

I used two positive controls: (i) a *sypQ* mutant that carries pARM7 (for overexpressing RscS) and pMKF1, which contains unmutated *sypQ* (strain MF001), and (ii) a wild type *V. fischeri* strain that carries pARM7 and pVSV105 (the vector control corresponding to pMKF1) (strain KV4700). For a negative control, the *sypQ* mutant overexpressing RscS and carrying the pVSV105 vector was used (strain KV5597). Both positive control strains displayed wrinkled colony formation within 24 hours, and exhibited high cohesiveness at the last time point (Figure 16A). In contrast, and as expected because *sypQ* had been previously shown to be critical for wrinkled colony formation [7], the negative control failed to produce any colony architecture for the duration of the time course, and it lacked cohesiveness at the final time point.

Of the six putative functional motifs in SypQ, only mutations to two of them failed to affect wrinkled colony formation (Figure 16B). These two were changes to D99 and D100, the putative TED and C-terminus DDx motifs that overlap at these two aspartic acid residues. Both the alanine and the glutamate mutations to these amino acids failed to affect biofilm formation. With these amino acid substitutions, SypQ restored normal wrinkled colony formation and polysaccharide production to the *sypQ* deficient *V. fischeri.* This suggests that the D99 and D100 amino acid residues are not essential to the function of SypQ and that this region of SypQ primary sequence does not represent a functional motif.

In contrast, mutations to the putative DxH motif had a noticeable effect on both the timing of biofilm formation and the cohesiveness of the colony. The D116A and D116E mutations to SypQ resulted in a diminished biofilm phenotype when the mutated protein was expressed in a *sypQ*-deficient *V. fischeri.* However, the D116A mutation was slightly more deleterious than the D116E; D116A resulted in a lack of surface architecture while still facilitating a mild amount of colony cohesion, whereas, D116E resulted in diminished (but present) surface architecture and colony cohesion.

The LxxRW motif in SypQ displayed two different phenotypes when mutated. When the tryptophan was mutated to a phenylalanine (W29F), there was no effect on the formation of biofilm in either timing or polysaccharide production. The alanine substitution (W29A), however, resulted in a null phenotype. The colony lacked both surface architecture and cohesiveness. Since the phenylalanine substitution is a conservative change, the difference in phenotype suggests that the tryptophan residue is not essential but the structure to which it contributes is essential. I will discuss possible reasons for the contrasting phenotypes in the Discussion section.

Mutations to the remaining two possible SypQ motifs resulted in a complete abrogation of biofilm formation. Glutamic acid and alanine substitutions to the aspartic acid (D151) in SypQ's putative DxD motif resulted in a smooth colony with no polysaccharide. The same result was achieved when the same substitutions were made to the aspartic acid residue (D236) in the C-terminal TED/DDx motif in SypQ. The severity of the phenotype regardless of residue substitution suggests that these amino acids are critical to the function of SypQ and could be located within functional motifs.

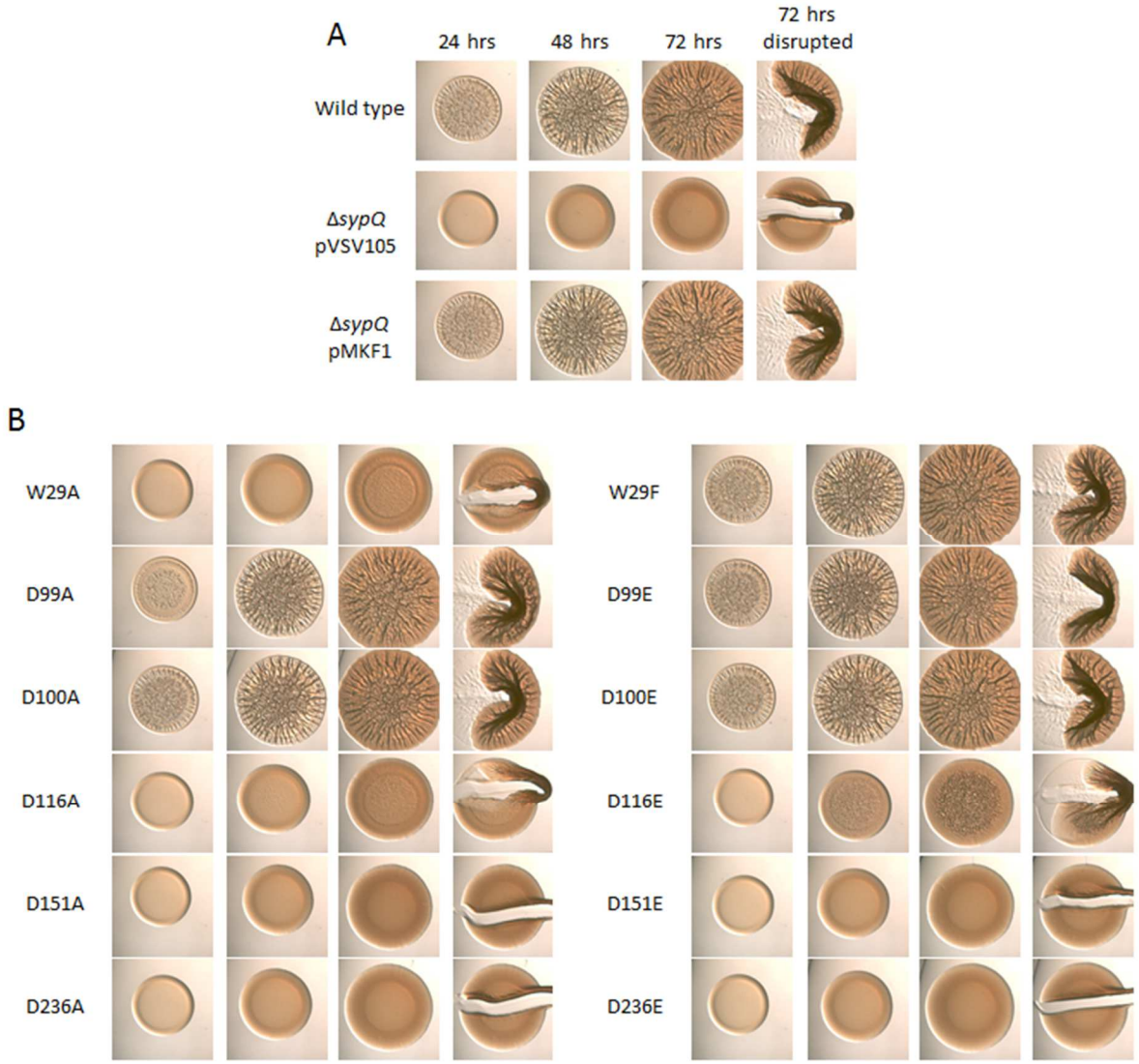

Figure 17. Wrinkled colony assay of site-directed *sypQ* mutants. A. Two positive controls were used. The wild type control is a *V. fischeri* strain that expresses a wild type SypQ from the chromosome, overexpresses RscS, and carries the vector control plasmid. The other positive control is a *sypQ* deficient strain that overexpresses RscS and carries the unmutated pMKF1 plasmid. **B**. Plasmids carrying the *sypQ* allele with the site-directed mutations were introduced into a *sypQ* deficient strain of *V. fischeri.* A wrinkled colony assay was performed over a 72-hour time course. At 72 hours, the spots were disrupted to assay for the presence of polysaccharide.

Random Mutagenesis of SypQ

The results from the site-directed mutagenesis identified four amino acid residues that are essential for the function of SypQ. However, SypQ contains 395 amino acids. To globally

identify which amino acids in SypQ are essential for its function, I randomly mutated the epitope-tagged *sypQ* allele. To generate random mutations into the *sypQ* allele, I passaged the pMKF1 plasmid through the mutator *E. coli* strain CC130. I then introduced the mutated plasmid into KV5562, a *sypQ* deficient strain of *V. fischeri* that overexpresses the sensor kinase RscS. Finally, I performed wrinkled colony assays to identify possible *sypQ* mutants and identified plasmid-based mutations (potentially in *sypQ*) by backcrossing into the parent strain the plasmid from any strains that displayed an altered biofilm phenotype. To reduce the incidence of identifying sibling mutants, I performed 148 independent mutageneses. From those experiments, I isolated 52 original mutants that retained their mutant phenotype after back-crossing: in the Discussion section, I will elaborate upon the possible reasons for the relatively poor yield of mutants following back-cross.

I next categorized the resulting mutants into one of three groups: delayed phenotype, diminished phenotype, and smooth phenotype. The delayed phenotype mutants demonstrated a delay in the onset of biofilm formation at 24 hours compared to the positive control. By 72 hours, however, the delayed mutants achieved wrinkling and colony stickiness at levels comparable to the positive control. The diminished phenotype mutants demonstrated both a delay in biofilm formation and a reduction in wrinkling and colony cohesion compared to the positive control. The smooth phenotype mutants showed a complete loss of biofilm indicated by a lack of both wrinkles and polysaccharide, giving them a phenotype similar to the negative control. Of the confirmed mutants, 5, 24, and 26 were classified as delayed, diminished, and smooth, respectively. As described in more detail below, I sequenced a subset of these mutants: when mutants with similar phenotypes were obtained from the same original mutagenesis, I

chose to sequence only one, to avoid re-sequencing possible siblings. Furthermore, knowing that the delayed and diminished mutants made at least some SypQ protein, I focused my initial sequencing on these mutants rather than the smooth, anticipating that the smooth mutants may not produce the SypQ protein for any of a variety of reasons.

To determine the nature of the mutation in both the delayed phenotype and diminished phenotype mutants, I sequenced the mutated alleles using ACGT (Wheeling, IL). The five delayed phenotype mutants had differing mutations (Figure 17). The most informative were mutants 4111, 4841, and 6441 which all had single point mutations: G322D, Y18H, and V244A respectively. In contrast, the other two were less informative: mutant 11871 had a single guanine to adenosine transition mutation in codon 79, resulting in a tryptophan being mutated to a stop codon, while 5961 had no discernable mutation in *sypQ* (Table 4). I conclude that G322, Y18, and V244 make contributions to SypQ function and /or stability. I will discuss possible explanations for the other delayed phenotype mutations in the Discussion section.

Table 4. Randomly generated delayed mutants

Name	**Mutagenesis**[1]	**Plate**[2]	**Isolate**[3]	**Mutation**
4111	41	1	1	g-a mutation = G322D
4841	48	4	1	t-c mutation = Y18H
5961	59	6	1	no genuine mutations found in gene
6441	64	4	1	t-c mutation = V244A
11871	118	7	1	g-a mutation = W41UGA early stop codon

[1]Mutagenesis refers to the specific number of the independent conjugation from which each mutant was obtained.
[2]Plate refers to the number of the plate from the independent conjugation.
[3]Isolate refers to the number of the isolate from the individual plate.

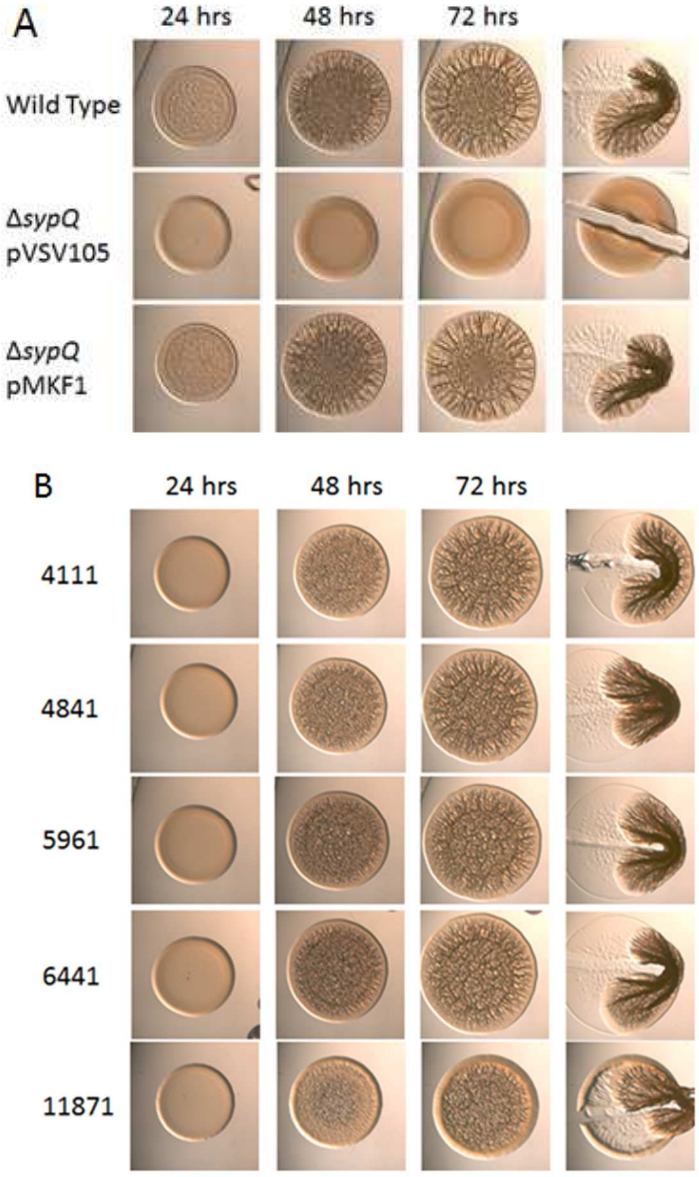

Figure 18. Wrinkled colony assay of the randomly generated delay phenotype mutants. **A**. Two positive controls were used. The wild type control is a *V. fischeri* strain that expresses a wild type SypQ from the chromosome, overexpresses RscS, and carries the vector control plasmid. The other positive control is a *sypQ* deficient strain that overexpresses RscS and carries the unmutated pMKF1 plasmid. **B**. Plasmids carrying randomly mutated *sypQ* allele were introduced into a *sypQ* deficient strain of *V. fischeri.* A wrinkled colony assay was performed over a 72-hour time course. Time points at which the pictures were taken are indicated above. At 72 hours, the spots were disrupted to assay for the presence of polysaccharide.

24 randomly generated mutants displayed a diminished phenotype. These mutants displayed both decreased colony cohesion and wrinkled colony architecture at 72 hours (Figure

18). Of these, nine contained single point mutations resulting in informative codon changes. Two of the mutants (10041and 11751) had a point mutation in predicted transmembrane regions (S323L and Y23H respectively) while two others (9831 and 9511) had point mutations near potential transmembrane regions (K306E and K359E respectively). The remaining five had mutations in different regions of the predicted large exposed region of the protein. Specifically, mutants 8611, 9361, and 5551 encoded SypQ proteins with codon changes D82N, N127D, and L287Q respectively. Mutants 9461 and 9751 each contained an independent mutation in H214Y (Table 5).

The remaining 15 mutants were much less informative. Three contained no mutations within the *sypQ,* while eight contained a nucleotide insertion that resulted in an early stop codon, similar to the phenomenon observed for the delayed mutants. Of the stop codon mutants, five mutants, (5141, 5241, 6371, 7341, and 11231) all had an adenosine insertion in a run of adenosines, while two mutants (8171 and 8261) each had the same cytosine insertion at nucleotide 105. Mutant 6551 had a cytosine insertion at nucleotide 1456. In the Discussion section, I will speculate on the possible reasons why the mutants containing stop codons were able to produce diminished biofilms and why three of the mutants had no visible mutations in the *sypQ* gene.

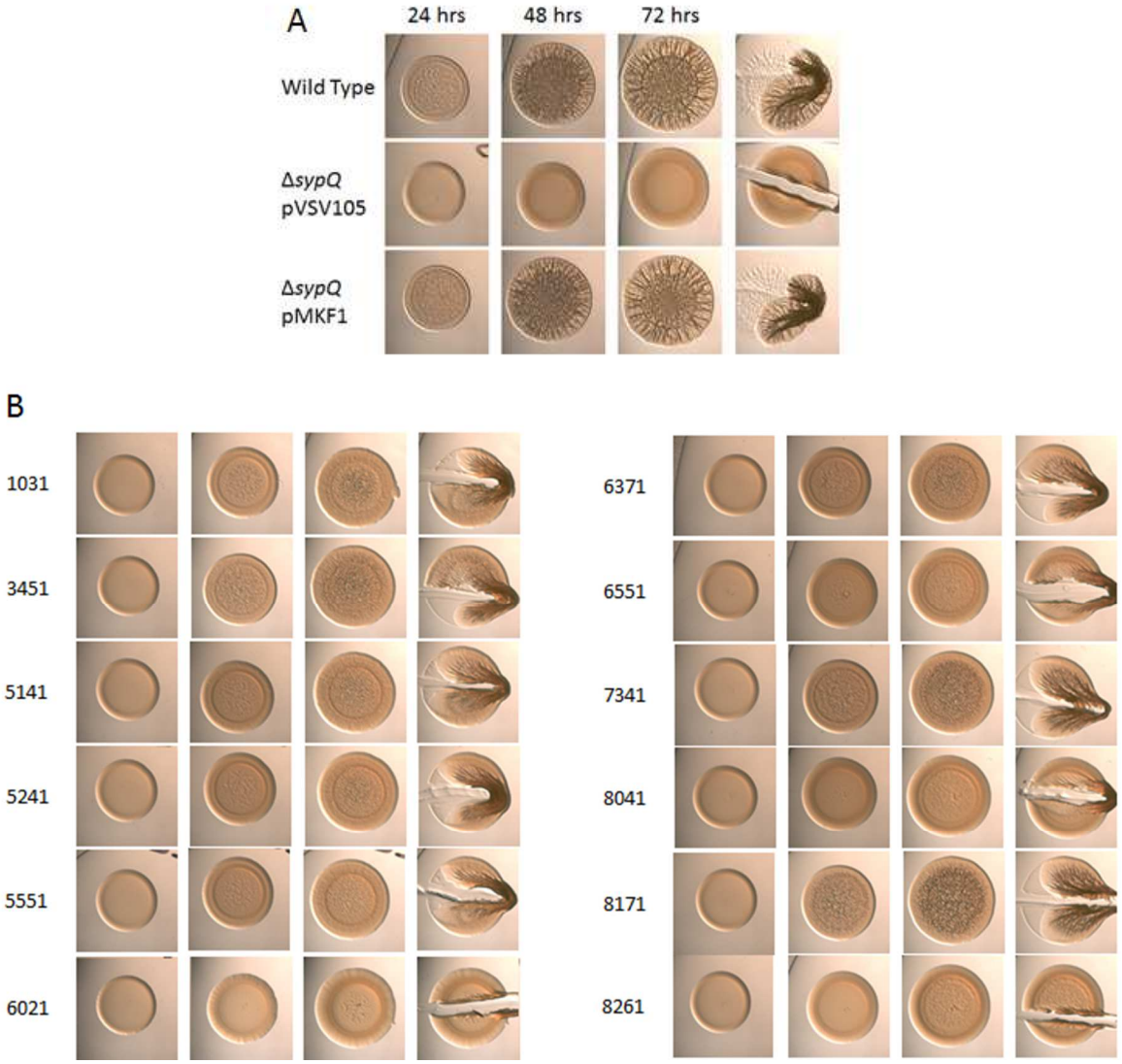

Figure 19. Wrinkled colony assay of randomly generated diminished phenotype mutants. A. Two positive controls were used. The wild type control is a *V. fischeri* strain that expresses a wild type SypQ from the chromosome, overexpresses RscS, and carries the vector control plasmid. The other positive control is a *sypQ* deficient strain that overexpresses RscS and carries the unmutated pMKF1 plasmid. **B**. Plasmids carrying randomly mutated *sypQ* allele were introduced into a *sypQ* deficient strain of *V. fischeri.* A wrinkled colony assay was performed over a 72-hour time course. Time points at which the pictures were taken are indicated above. At 72 hours, the spots were disrupted to assay for the presence of polysaccharide.

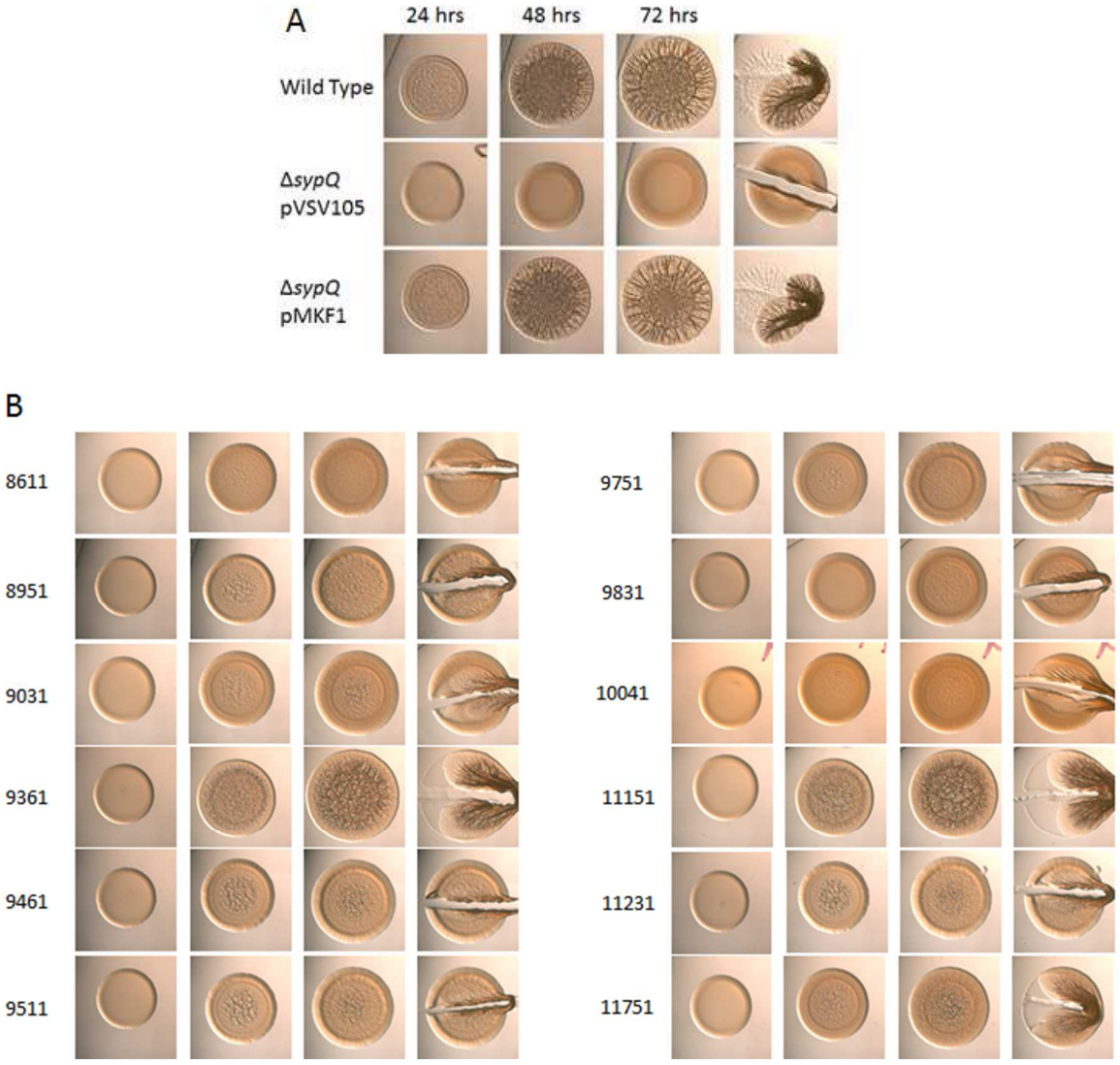

Figure 19 (cont.). Wrinkled colony assay of randomly generated diminished phenotype mutants. A. Two positive controls were used. The wild type control is a *V. fischeri* strain that expresses a wild type SypQ from the chromosome, overexpresses RscS, and carries the vector control plasmid. The other positive control is a *sypQ* deficient strain that overexpresses RscS and carries the unmutated pMKF1 plasmid. **B**. Plasmids carrying randomly mutated *sypQ* allele were introduced into a *sypQ* deficient strain of *V. fischeri.* A wrinkled colony assay was performed over a 72-hour time course. Time points at which the pictures were taken are indicated above. At 72 hours, the spots were disrupted to assay for the presence of polysaccharide.

Table 5. Randomly generated Diminished mutants.

Name[1]	Mutagenesis	Plate	Isolate	mutation
1031	10	3	1	No mutation found in gene
3451	34	5	1	a-c mutation = K89Q
5141	51	4	1	a insertion at nucleotide 171 = early stop codon
5241	52	4	1	a insertion at nucleotide 171= early stop codon
5551	55	5	1	t-a mutation = L287Q
6021	61	2	1	no mutations found in gene
6371	63	7	1	a insertion at nucleotide 171 = early stop codon
6551	65	5	1	c insertion at nucleotide 987 = early stop codon
7341	73	4	1	a insertion at nucleotide 131 = early stop codon
8041	80	4	1	a-g mutation = K306E
8171	81	7	1	c insertion at nucleotide 105 = early stop codon
8261	82	6	1	c insertion at nucleotide 105 = early stop codon
8611	86	1	1	g-a mutation = D82N
8951	89	5	1	no mutations found in gene
9031	90	3	1	no mutations found in gene
9361	93	6	1	a-g mutation = N127D
9461	94	6	1	c-t mutation = H214Y
9511	95	1	1	a-g mutation = K359E
9751	97	5	1	c-t mutation = H214Y
9831	98	3	1	a-g mutation = K306E
10041	100	4	1	c-t mutation = S323L
11231	112	3	1	a insertion at nucleotide 171 = early stop codon
11751	117	5	1	t-c mutation = Y23H

[1]Mutagenesis, plate, and isolate terms are as defined in Table 4.

The remainder (24) of the randomly generated mutations resulted in a smooth colony phenotype (Figure 19). The smooth mutation phenotype mutants are the most informative in that they result in the most defective phenotype; thus, knowing the nature of the mutations in the smooth mutants can give us the most insight into the function of SypQ. However, the null phenotype of the smooth mutants could also be due the loss of SypQ production through several possible mechanisms. For example, the mutation introduced into the gene by the CC130 mutator *E. coli* cells could have been a point mutation resulting in an early stop codon. The mutation could be in the promotor region resulting in loss of transcription. Another possibility is that the mutation is a point mutation that affects the stability of the protein.

Usually, Western blots are used to distinguish between mutations that affect protein production/stability and those that do not. However, given my inability to visualize the epitope-tagged SypQ protein via Western blot, this approach was not possible for SypQ. Therefore, I attempted another method to determine if the protein was being produced. I hypothesized that if a mutant SypQ protein is, in fact, made, then the presence of the defective protein could potentially interfere with the function of a wild-type protein (Figure 20). If that were the case, then the mutant protein could potentially cause a delay in biofilm formation in an otherwise biofilm-competent strain. Such a result would provide evidence that the mutated protein is indeed being produced.

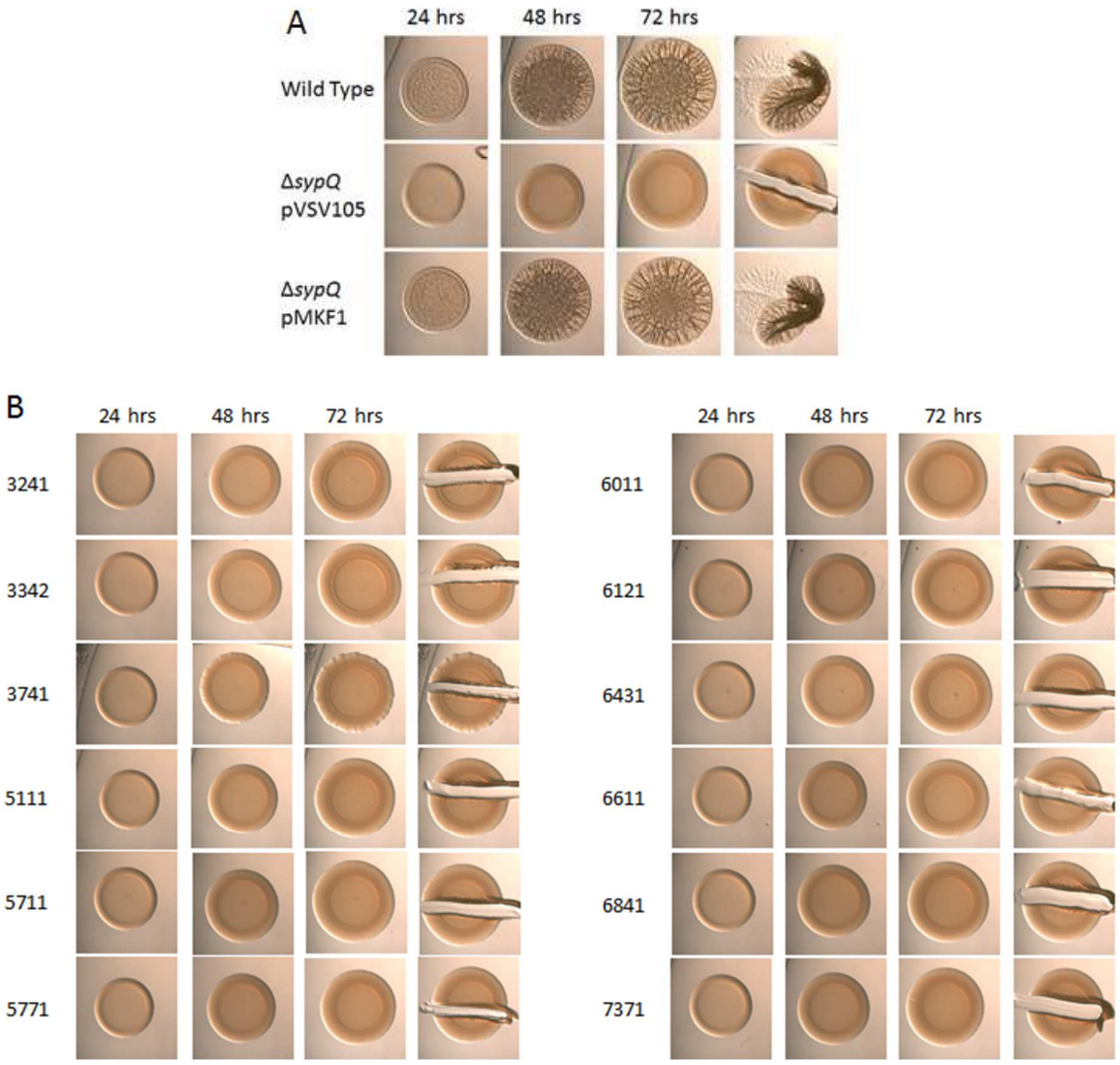

Figure 20. Wrinkled colony assay of randomly generated smooth phenotype mutants. **A**. Two positive controls were used. The wild type control is a *V. fischeri* strain that expresses a wild type SypQ from the chromosome, overexpresses RscS, and carries the vector control plasmid. The other positive control is a *sypQ* deficient strain that overexpresses RscS and carries the unmutated pMKF1 plasmid. **B**. Plasmids carrying randomly mutated *sypQ* allele were introduced into a *sypQ* deficient strain of *V. fischeri.* A wrinkled colony assay was performed over a 72-hour time course. Time points at which the pictures were taken are indicated above. At 72 hours, the spots were disrupted to assay for the presence of polysaccharide.

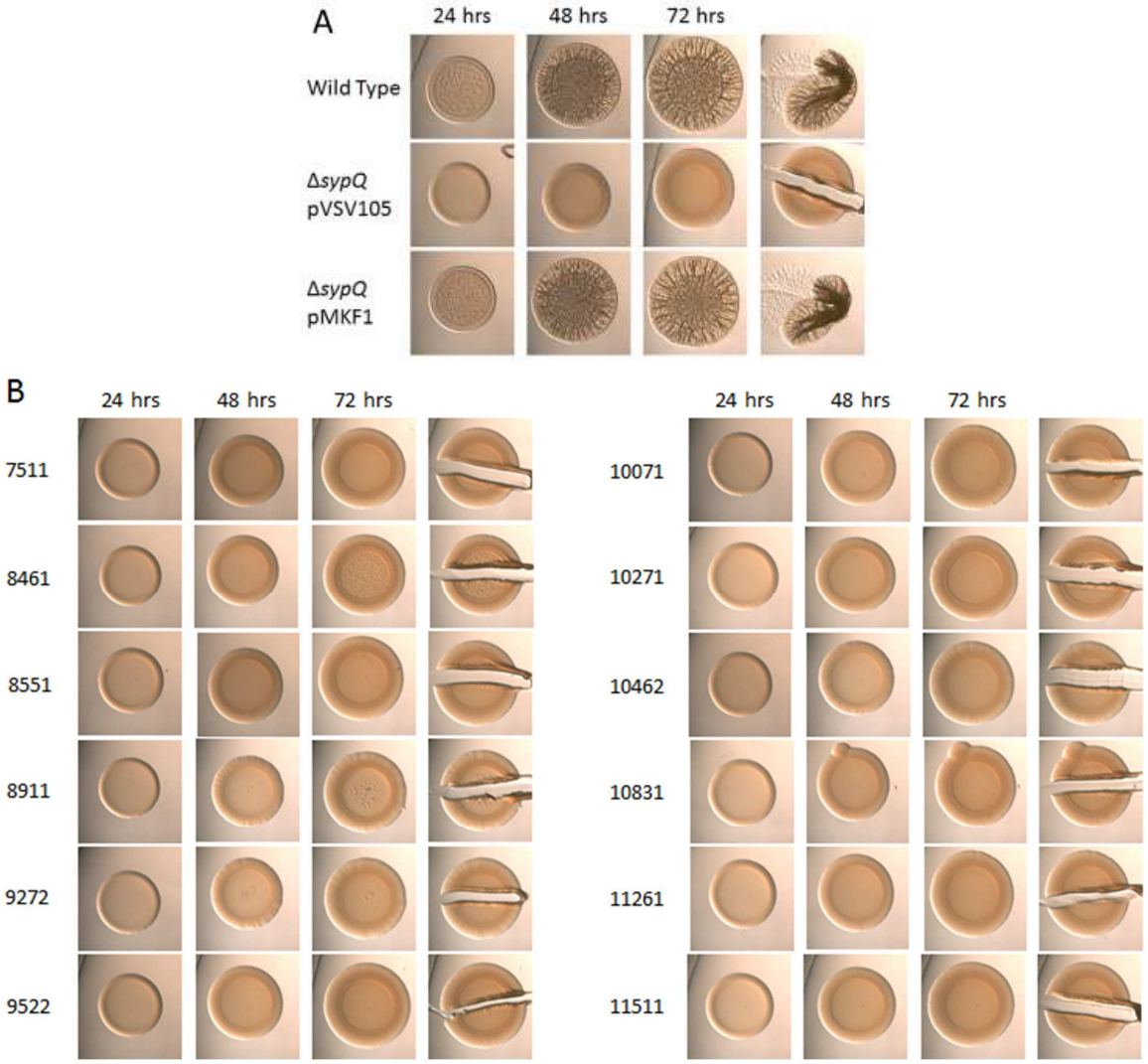

Figure 20 (cont.). Wrinkled colony assay of randomly generated smooth phenotype mutants. **A**. Two positive controls were used. The wild type control is a *V. fischeri* strain that expresses a wild type SypQ from the chromosome, overexpresses RscS, and carries the vector control plasmid. The other positive control is a *sypQ* deficient strain that overexpresses RscS and carries the unmutated pMKF1 plasmid. **B**. Plasmids carrying randomly mutated *sypQ* allele were introduced into a *sypQ* deficient strain of *V. fischeri.* A wrinkled colony assay was performed over a 72-hour time course. Time points at which the pictures were taken are indicated above. At 72 hours, the spots were disrupted to assay for the presence of polysaccharide.

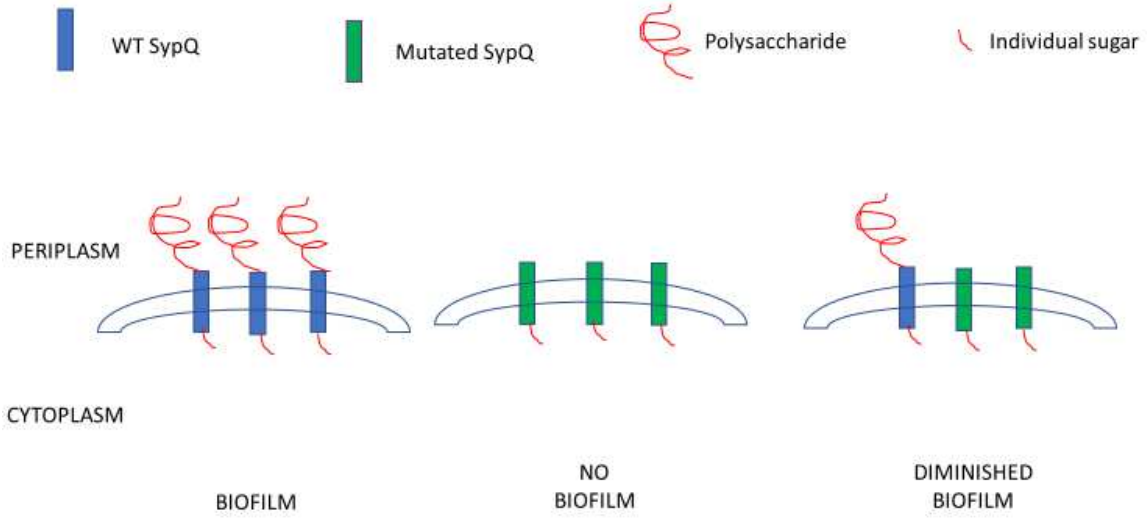

Figure 21. Depiction of SypQ co-expression assay. To determine if the mutated protein is being produced, the randomly mutated plasmids from smooth phenotype mutants were introduced to a *V. fischeri* strain carrying a wild type chromosomal *sypQ* allele. If the mutated protein is being made, then it will compete with the wild type SypQ protein and lead to a delay in biofilm formation.

To determine if any of the smooth phenotype mutants produced a stable SypQ protein that was present in the cell, I introduced plasmids with mutated *sypQ* alleles into KV4883. This strain contains a wild type *sypQ* allele in its native position in the chromosome. As proof of principle, I began by assessing the site-directed mutants, as I already knew the nature of their mutations (and that none of the alleles encoded proteins containing an early stop codon). I then assessed biofilm formation via the wrinkled colony assay (Figure 21). I found that eight of the 12 mutations caused a small delay in biofilm formation compared to the positive controls. Specifically, at 21 hours, the positive controls and four of the mutants (W29F, W29A, D99E, and D100A) began to wrinkle, while the others remained smooth. The D99A, D151A, D116E,

D116A, D236A, and D100E mutations all resulted in moderate, ~1 hour, delay in the onset of biofilm formation. The D151E and D236E mutations resulted in a more severe, ~2-3 hour, delay. These mutants did not display wrinkling until 23 and 24 hours, respectively. From these data, I conclude that the mutations with a severe delay (D151E and D151A) are producing protein at high enough levels as to interfere with the wild type protein. The lack of delay in the W29F, W29A, D99E, and D100A mutants could be due to a few reasons: it is possible that no protein is made, that it is made but unstable and easily degraded, or that it is made but is defective in substrate binding or protein-protein interaction.

Next, I introduced into KV4883 the mutated plasmids from 24 randomly generated smooth phenotype strains. 14 of the 26 demonstrated delayed biofilm formation relative to the controls (Figure 22). 12 of the mutants began wrinkling at 21 hours similar to the positive controls, thus I cannot make any conclusions about whether SypQ protein is made. In contrast, ten of the mutants had a modest, ~ 1 hour, delay and four had a slightly more severe, ~2 hour, delay. Based on these results, I conclude that these latter 14 may be producing a mutant SypQ protein that is inhibiting the action of the wild type SypQ. To identify the nature of the mutation in these mutants, I submitted the plasmids for sequencing. As shown in Table 6, two of the mutants (8461 and 8911) had an adenosine deletion at a run of adenosines that resulted in an early stop codon. One mutant, 3342, had both a point mutation (L228P) and an early stop codon as a result of a tyrosine insertion at nucleotide 928. Mutant 3342 had no apparent mutation found within the gene. The remaining ten mutants had single point mutations at various regions within *sypQ* (Table 6). Three of the mutations, G346, L26P, and Y18H, were located in putative transmembrane regions (9271, 8551, and 11261 respectively), while mutants 3741 and 7371 had

mutations located near transmembrane regions (S304P and G307R, respectively). Mutants 6011, 6431, and 10071 all had point mutations located in the central region of the protein (A216V, Y193H, and Y180H, respectively). Interestingly, two of the mutants contained point mutations in two of the identified putative functional motifs in SypQ (S153L and D236N, respectively). Possible reasons why these mutations contribute to a null phenotype are discussed to a greater extent in the Discussion section.

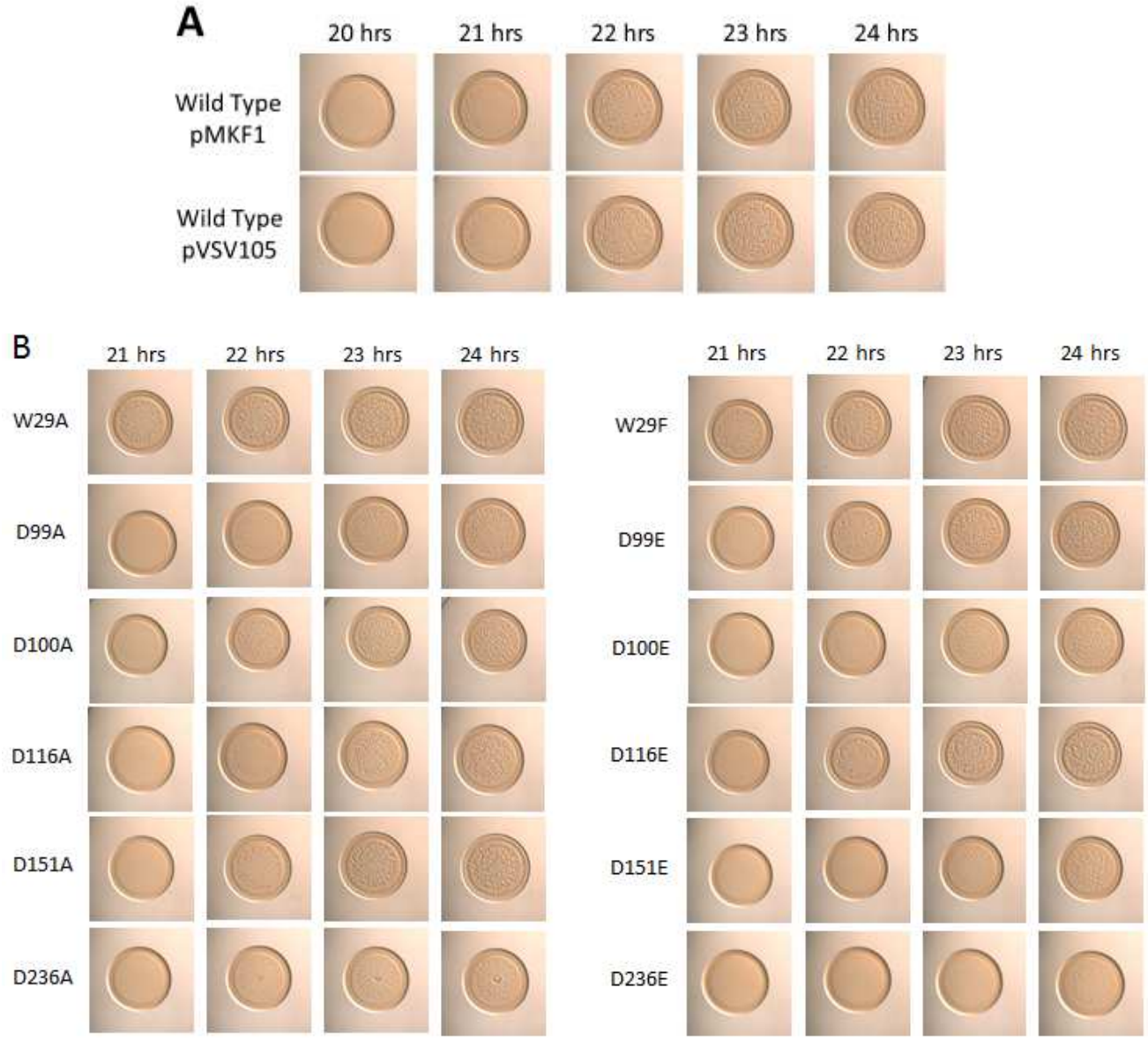

Figure 22. Expression of site-directed mutants in wild type *V. fischeri.* A. Two positive controls were used: a wild type *V. fischeri* overexpressing RscS and carrying the pVSV105 vector control plasmid and a wild type *V. fischeri* overexpression RscS and carrying the unmutated pMKF1 plasmid. **B.** The plasmid containing the site-directed mutations were introduced into wild type *V. fischeri* overexpressing RscS and a spotting assay was done. The identity of the point mutation is indicated to the left side while the time at which the pictures were taken is indicated above.

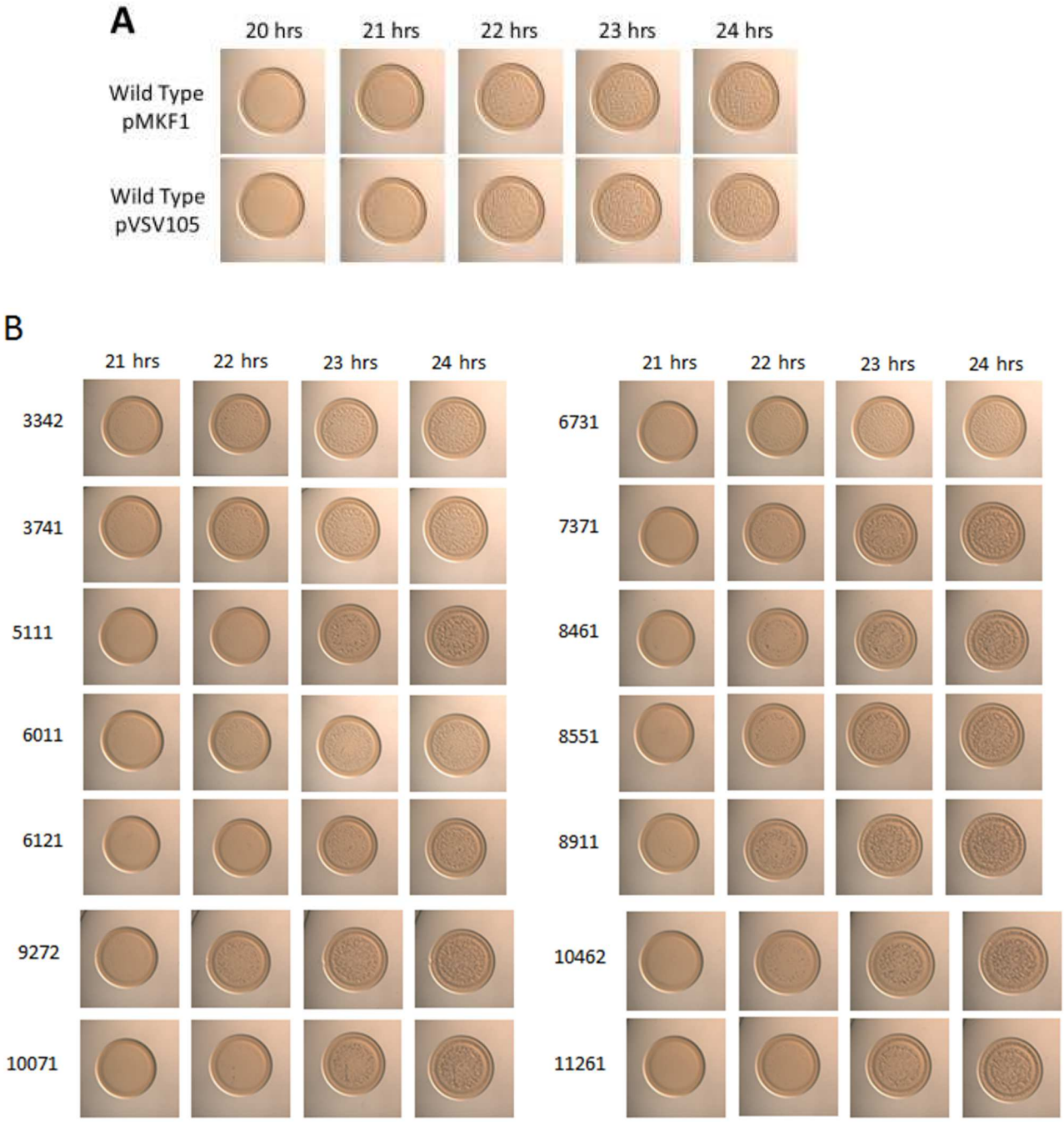

Figure 23. Expression of randomly generated smooth phenotype mutants in wild type *V. fischeri*. A. Two positive controls were used: a wild type *V. fischeri* overexpressing RscS and carrying the pVSV105 vector control plasmid and a wild type *V. fischeri* overexpression RscS and carrying the unmutated pMKF1 plasmid. **B.** The plasmid containing the randomly generated mutations were introduced into wild type *V. fischeri* overexpressing RscS and a spotting assay was done. The identity of the random mutant is indicated to the left side while the time at which the pictures were taken is indicated above.

Table 6. Randomly generated smooth phenotype mutants

Name	Mutagenesis	Plate	Isolate	mutation
3342	33	4	2	t-c mutation = L228P/ t insertion = early stop codon at amino acid 327
3741	37	4	1	t-c mutation = S304P
5111	51	1	1	c-t mutation = S153L
5711	57	1	1	c-t mutation = S153L
5771	57	7	1	Not sequenced
6011	60	1	1	c-t mutation = A216V
6121	61	2	1	g-t mutation = D236N
6431	64	3	1	t-c mutation = Y193H
6611	66	1	1	Not sequenced
6841	68	4	1	Not sequenced
7371	73	7	1	g-a mutation = G307R
7511	75	1	1	Not sequenced
8461	84	6	1	a deletion at nucleotide 131 = early stop codon at amino acid 37
8551	85	5	1	t-c mutation = L26P
8911	89	1	1	a deletion at nucleotide 131= early stop codon at amino acid 37
9271	92	7	1	g-a mutation = G346E
9522	95	2	2	Not sequenced
10071	100	7	1	t-c mutation = Y180H
10271	102	7	1	Not sequenced
10462	104	6	2	No mutation in gene
10831	108	3	1	Not sequenced
11261	112	6	1	a-g mutation = Y18C
11511	115	1	1	Not sequenced

[1]Mutagenesis, plate, and isolate terms are as defined in Table 4.

CHAPTER FOUR

DISCUSSION

The work discussed herein investigated the role of SypQ in the production of *syp* polysaccharide in *V. fischeri.* Site-directed and random mutagenesis identified amino acids that are critical for the function of the protein as well as the presence of likely conserved motifs that suggest SypQ functions in a GT-2-like fashion. These results contribute to a clearer understanding SypQ's contribution to the formation of biofilm by *V. fischeri.*

Bioinformatics Analysis

Bioinformatics suggests that SypQ is likely a GT-2 glycosyltransferase. Not only does SypQ possess a possible DxD motif, characteristic of this enzyme family, but it is predicted to have a large amount of structural similarity to the well-characterized GT-2 protein BcsA. Furthermore, because GT-2 proteins are inner membrane proteins and because two prediction programs indicated the presence of transmembrane segments, I hypothesize that SypQ likely functions from a location in the inner membrane of *V. fischeri.*

The topology of SypQ remains unclear. Two separate bioinformatics programs (Topcons and Predictprotein) suggested that SypQ has four or five transmembrane regions, respectively. Topological mapping assays could be utilized to address the discrepancy [81]. SypQ's orientation in the membrane could be determined with a *phoA* and *lacZ* reporters inserted into

the protein. Because β-galactosidase and PhoA are active only in the cytoplasm and periplasm [74], respectively , the positions of the active fusions would indicate the topology of SypQ. This information would provide insight into where the active site of the enzyme is positioned.

Lack of SypQ Visualization

Visualizing SypQ will provide insight into protein levels and whether these levels correlate to altered biofilm phenotypes. To facilitate this visualization, I inserted a FLAG epitope tag at four different internal locations within SypQ, generating four individual constructs. I found that SypQ is permissive to the insertion of a FLAG epitope tag immediately following amino acid residue 36. The resulting protein is functional and able to restore biofilm to a *sypQ* deficient strain. The other three constructs had resulted in either a diminished phenotype (aa 123) or a null phenotype (aa 223 and 292). There are many possible reasons for the defective phenotypes. Some of these possibilities include the inserted tag separating the essential aspartates, changing the conformation of the protein, inhibiting protein-substrate interactions, or protein degradation. One way to determine if the SypQ is made is by performing a co-expression assay (in strain KV4883).

Unfortunately, the protein carrying the epitope tag after residue 36 remained undetectable via Western blot after several attempts. This could be due to the protein being completely embedded in the membrane, thus occluding the epitope and making it inaccessible to the antibody. Isolating the inner membrane fraction via ultracentrifugation could concentrate the protein enough for optimal visualization via Western blot analysis. Alternatively, SypQ could be visualized by generating and using SypQ-specific antibody.

SypQ Mutagenesis

The goal of this work was to develop a deeper understanding of SypQ function by performing a structure-function analysis. Through site-directed and random mutagenesis, I identified a large number of residues that are critical for the function of the protein. The locations of all generated amino acid substitutions are depicted in Figure 23.

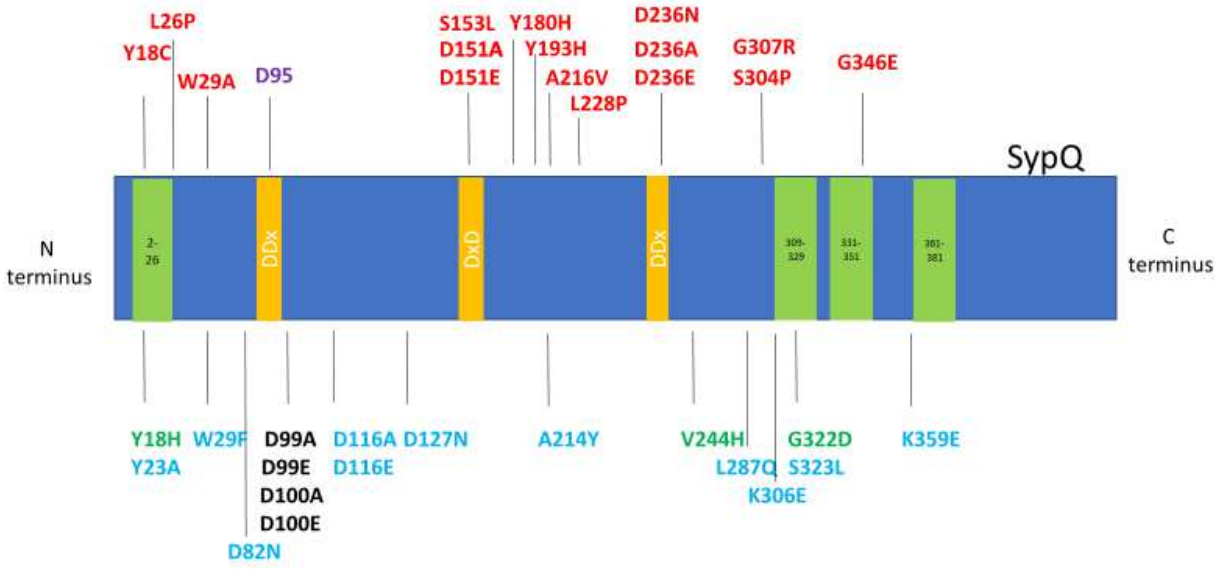

Figure 24. Depiction of generated amino acid substitutions in SypQ. The *sypQ* gene is depicted as a blue bar. The green and yellow vertical bars represent the predicted transmembrane regions and functional motifs, respectively. Mutations resulting in null phenotypes are depicted in red above the gene while substitutions resulting in delayed, diminished, and no phenotypes are indicated below the gene in green, blue and black, respectively. The residue predicted to be the essential aspartic acid in the DDx motif is identified by bold purple type.

Mutations to four SypQ residues stood out as the particularly interesting. Substitutions to D82, D151, S153, and D236 resulted in a severe biofilm defect. Additionally, three of the

residues (D151, S153, and D236) are positioned in regions corresponding to known conserved motifs in BcsA (DxD and TED). Another compelling factor is the high level of conservation of the specific residues (D82, D151, and D236) to those found in BcscA. Finally, the available data are consistent with the conclusion that these mutant proteins are made: I see a delay in the onset of biofilm when the mutated plasmids were introduced to a *V. fischeri* carrying a wild-type SypQ allele in the chromosome. The requirement for these conserved motifs strengthens my conclusion that SypQ is likely a GT-2.

In addition to these four substitutions, 12 more occurred in putative non-membrane regions of SypQ. The four mutations that I found most interesting were either tyrosine to histidine substitutions or histidine to tyrosine substitutions [H214Y (occurring in two independent mutants), Y193H, and Y180H). Histidine residues can participate in many interactions, including those with divalent cations [82]. If SypQ is a GT-2 and utilizes a mechanism like BcsA, it could require divalent cations to stabilize the nucleotide diphosphate of the sugar donor. These histidine substitutions may interfere with the divalent cation's ability to stabilize the donor sugar resulting in a null phenotype. The loss of histidine could conversely lead to a loss of cation interaction.

The random mutagenesis also generated substitutions in putative transmembrane regions. Eleven of the amino acid changes occurred in predicted transmembrane regions or directly adjacent to predicted transmembrane regions. These mutations included proline substitutions, lysine to glutamic acid substitutions, and, in one case, the loss of a large side chain. Proline has a cyclic structure, making it rigid relative to other amino acids. While prolines naturally occur in membrane proteins, evidence suggests that proline substitutions are poorly tolerated by

membrane proteins [83]. Thus, I hypothesize that the proline substitutions introduced structural changes that either destabilized the protein or cascaded to the active site, affecting function. K to E substitutions alter the charge from positive to negative, which could disrupt specific interactions or alter the stability of the protein in the membrane by increasing its tendency to migrate away from the cytoplasm [84]. The W29A substitution also contributed to a severely diminished phenotype. This defect could be attributed to many possibilities. One possibility is that the substitution interfered with membrane stability by removing the large aromatic side chain since a phenylalanine substitution at the same residue allowed biofilm formation.

The remaining mutants were less informative and, therefore, not as compelling. Site-directed mutations to D99 and D100 resulted in no phenotypic change compared to wild-type SypQ, suggesting that they are neither essential nor located in critical motifs. 12 mutants had early stop codons but still permitted some biofilm (delayed or diminished phenotypes). Because some biofilm was produced, there must be some SypQ protein present. I hypothesize that protein was produced by some level of translational frameshift such as a stop codon readthrough or a +1-ribosomal frameshift [85-87]. The least informative mutants had no amino acid substitutions present in the gene despite having a defective biofilm phenotype. One possible explanation is the presence of a mutation elsewhere in the plasmid, such as a change that decreases the copy number of the plasmid, resulting in less protein being made.

Model for SypQ Function

From all my data combined, I conclude that SypQ is a GT-2 enzyme that contains DDx, DxD, and TED motifs and that residues within these motifs play specific roles in the catalytic action of the enzyme (Figure 25A). The D151 and D236 residues are essential to the function of

SypQ and correspond to the second and third aspartic acids in the D, D, D, Q/LxxRW motif (as discussed in chapter one), respectively. The DxD motif has been shown to be essential to BcsA function by coordinating a divalent cation needed for activity [13, 16]. I hypothesize that the corresponding motif in SypQ, DVS, functions in a similar manner. I have depicted in Figure 25C how these residues, although not fully conserved, could function like the DxD residues in BcsA (Figure 25C). The TED motif has been shown to be responsible for deprotonation of the non-reducing end of the growing polysaccharide chain [16]. I, therefore, hypothesize that the corresponding motif in SypQ, NDD, functions in a similar manner. I have depicted in Figure 25D how these residues, particularly D236, could function like the TED motif in BcsA (Figure 25D). If SypQ is a family 2 GT, as I hypothesize, then another aspartic acid residue, positioned upstream of the DxD motif, will also prove essential for the function of the enzyme. The D95 residue in SypQ may serve that essential function: it corresponds to the conserved aspartate in the DDx motif of BcsA and may interact with the nucleotide diphosphate of the donor sugar molecule (Figure 25B). If amino acid substitutions to this residue result in a loss of function, then it would further suggest SypQ functions as a GT-2. I did not initially focus on the D95 residue, as the other aspartate flanking it in the DDx motif of BcsA is not conserved. However, given the conservation of the aspartate residue in SypQ relative to BcsA, its position relative to the other functional motifs, and the known requirements for the three sets of aspartate residues, I now conclude that D95 warrants further investigation. Finally, I further hypothesize that a motif with a function analogous to that of the Q/LxxRW motif is present in SypQ. Residues in the Q/LxxRW motif interact with the donor sugar molecule and the growing polysaccharide chain [88]. SypQ does not have a true Q/LxxRW motif at the same location as BcsA, AcsA, or NodC.

However, it does have two highly conserved arginine residues (R273 and R275) present at the location of the known Q/LxxRW motif. I hypothesize that the R275 residue is likely interacting with the nucleotide diphosphate of the donor sugar (Figure 25E) and should be targeted for future mutagenesis.

Using the Phyre 2 software, I highlighted all of the amino acid residues whose substitution resulted in null phenotypes as well as D95 and R275 on a 3-dimensional ribbon diagram of SypQ (Figure 26). The residues predicted to be in GT-2 motifs appear to be centered in or around the predicted active site supporting my hypothesis that SypQ is a family 2 glycosyltransferase. I then utilized the Phyre 2 software to identify more residues to target for mutagenesis (Figure 27). The Phyre 2 investigator tool provides amino acid preference and mutational sensitivity predictions for each residue in SypQ. Using this software, I identified residues that had high mutational sensitivity profiles as targets for future mutagenesis. The targeted amino acids are listed in Table 7. Mutagenesis to these residues can further our understanding of how SypQ contributes to the formation of *syp* polysaccharide in *V. fischeri.*

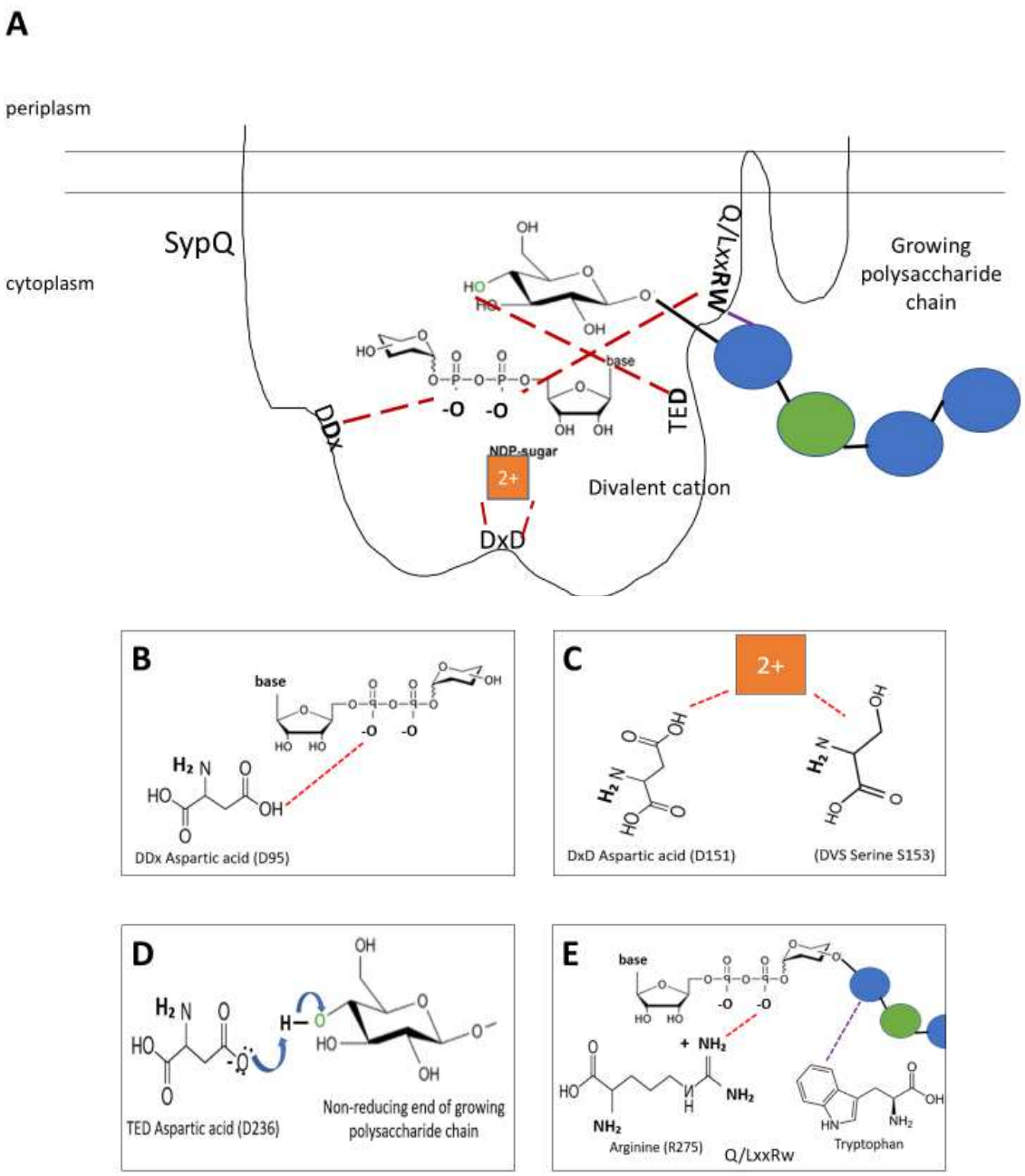

Figure 25. Predicted model of active site mechanism and role of identified residues. A. A depiction of the role of the GT-2 motifs in SypQ. Predicted hydrogen bonding interactions are indicated by red dashed lines, the movement of electrons are indicated by blue arrows, and predicted van der Waals forces are indicated by the purple dashed lines. The corresponding residues in SypQ are in parentheses. **B**. Predicted contribution of the D95 residue in SypQ. **C.** The predicted role of the D151 and S153 residues in SypQ. The divalent cation is represented by the orange square. **D.** The predicted mechanism of action of residue D236 in SypQ. **E.** Predicted role of R275 in SypQ and the know function of the tryptophan residue in the Q/LxxRW motif in processive GT-2s. These predictions are based on interactions known to occur in the GT-2 enzyme BcsA [13, 15, 16].

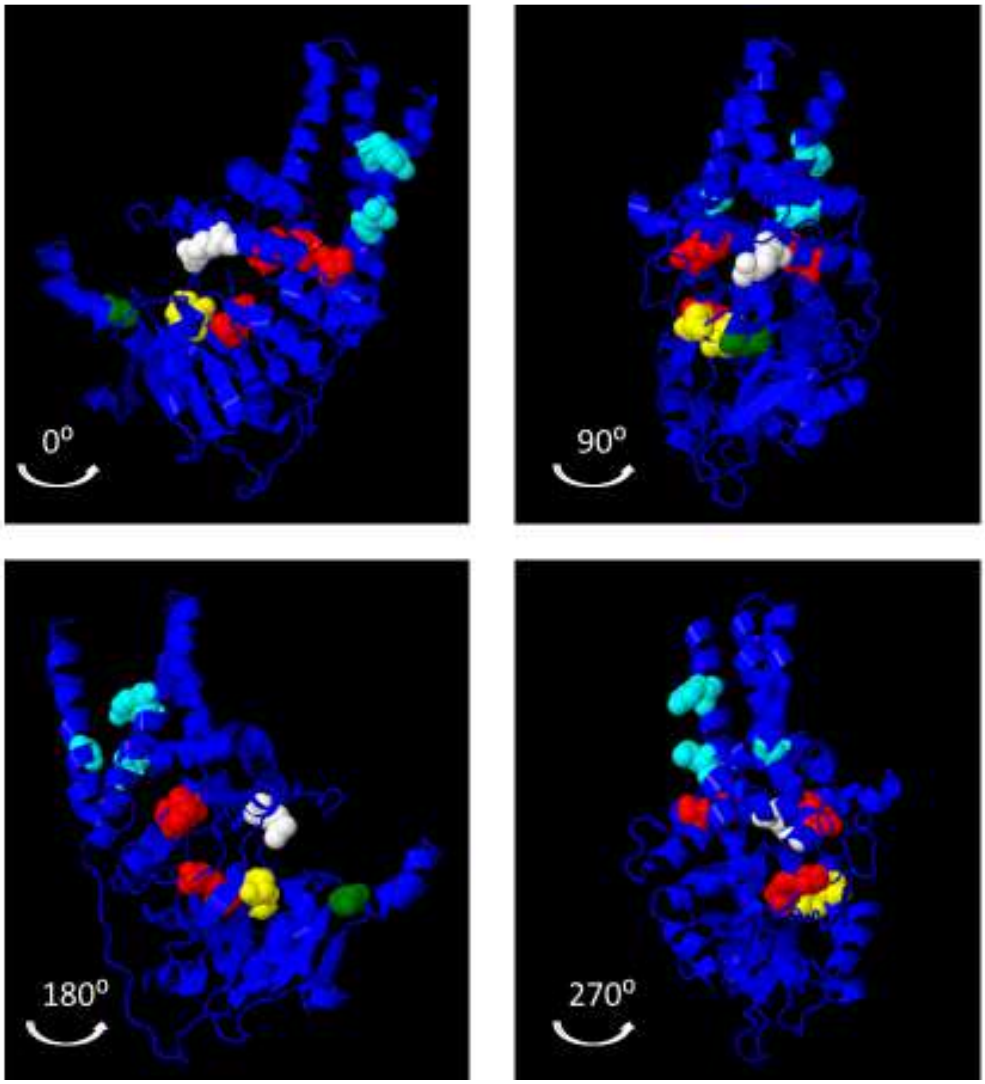

Figure 26. Model depicting locations of residues contributing to null phenotypes. The blue ribbon diagram indicates the SypQ structure as predicted by the Phyre 2 software. Residues located in predicted transmembrane regions are colored cyan. Residues predicted to be in the DxD and TED motifs are colored yellow. Residues that are neither in motifs or transmembrane regions are colored red. The white residue indicates the location of R275 while the green residue indicates the location of D95 [71].

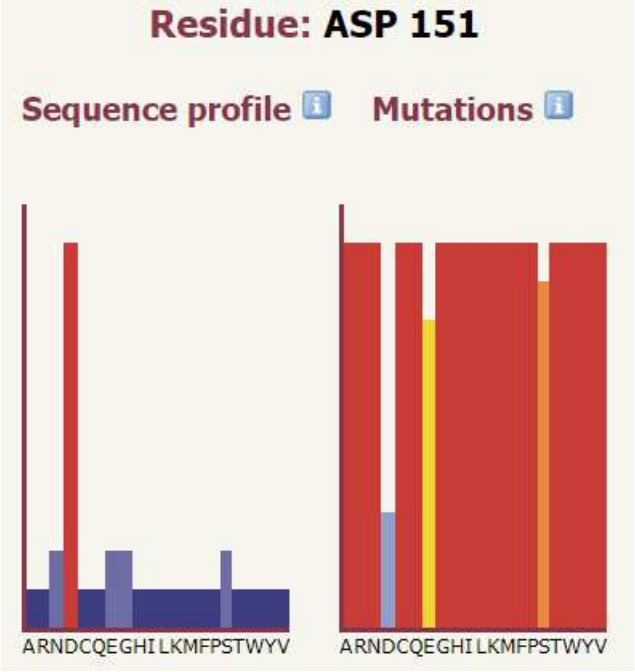

Figure 27. Mutational sensitivity profile predicted by the Phyre 2 software for SypQ residue D151. The profile for D151 is depicted as an example of high mutational sensitivity. The sequence profile predicts the residue preference at each location. High and red bars indicate high preference while low and blue bars indicate low preference. The mutation profile indicates the sensitivity of the residue at that location to substitutions by different amino acids. High and red bars indicate a high sensitivity to mutation while lower blue bars indicate less sensitivity [71].

Table 7. Residue predictions and suggested mutations.

Residue in SypQ	Predicted role	Expected phenotype
P56	Structural contribution	P56A = null
I58	Unclear – located in β sheet in putative active site	I58D = null
I61	Unclear – located in β sheet in putative active site	I61A = null
I62	Unclear – located in β sheet in putative active site	I62D = null
P63	Structural contribution	P63A = null
A64	Unclear – located in β sheet in putative active site	A64D =null
I71	Unclear – located in α helix near putative active site	I71D = null
I90	Unclear – located in β sheet in putative active site	I90D = null
D95	Interaction with the	D95A/D95E = null

	nucleotide diphosphate of the donor sugar	phenotype
S153	Coordination of divalent cations	S153A/S153E = null phenotype
R273	Stabilization of the penultimate sugar in the polysaccharide chain	R273K =diminished phenotype R273A = null phenotype
R275	Stabilization of the penultimate sugar in the polysaccharide chain	R273K =diminished phenotype R275A = null phenotype

SypQ is not likely to act alone in synthesizing the Syp polysaccharide. As discussed in appendix A, the other five Syp GTs (SypH, I, J, N, and P) encode putative family 1 glycosyltransferases (GT-1) that, I predict, generate the activated sugars necessary to produce Syp polysaccharide. I propose that SypQ functions in conjunction with the other Syp structural proteins (SypC, SypD, SypK, SypL, and SypO) to polymerize the individual sugars supplied by the five GT-1 proteins into a larger polysaccharide chain that gets modified and exported out of the cell. I further hypothesize that SypQ is likely functioning as a polymerase in a Wzx/Wzy-dependent pathway. In support of this proposal, specific Syp structural proteins (SypC, SypK, SypL, and SypO) have sequence homology to Wzx/Wzy-dependent proteins (Wza, Wzx, WaaL, and Wzz, respectively) [7, 8]. If these proteins do indeed function like other Wzx/Wzy-based systems, then it is likely that SypQ polymerizes a polysaccharide in the cytoplasm that is transferred to the periplasm via the putative flippase SypK. As discussed in chapter one, the polysaccharides produced by the Wzx-Wzy pathway are almost exclusively heteropolymers. Given the predicted function of SypQ as a GT-2 family protein, the presence of Wzx/Wzy-like Syp structural proteins, and the presence of multiple GT-1 family proteins each predicted to yield

activated sugars, I conclude that the polysaccharide produced by SypQ is most likely a heteropolymer.

This conclusion, however, contradicts that of a recent study that investigated the role of SypQ in *Vibrio parahaemolyticus* [89]. Ye et. al. determined that *sypQ* is critical for biofilm formation, as I and others have demonstrated to be the case in *V. fischeri* [7, 52]. The *V. parahaemolyticus* study concluded that SypQ facilitates the formation of the homopolymer PNAG (poly-N-acetyl glucosamine) and functions in a PgaC-like manner. In *E. coli,* PgaC is part of a synthase-dependent EPS biosynthesis pathway rather than a Wzx/Wzy-dependent pathway [22, 90]. Thus, this conclusion suggests that SypQ is polymerizing and translocating the polysaccharide, PNAG, simultaneously. This pathway of homopolymer polysaccharide production seems unlikely, given that (1) GT-2s functioning within Wzx/Wzy-dependent pathways produce heteropolymers and (2) the *syp* locus in *V. parahaemolyticus* is largely conserved relative to the *V. fischeri* locus and also encodes Wzx/Wzy-like structural proteins as well as GT-1 family proteins [52]. These conclusions, along with the data I have presented in this thesis, highlight the need for further investigation into the specific function of SypQ in polysaccharide production.

Appendix A
Syp glycosyltransferases

To better understand the *syp* polysaccharide and how SypQ contributes to it, I used bioinformatics to look more closely at the other five syp glycosyltransferases: H, I, J, N, and P. I submitted the amino acid sequences for the GTs to the Phyre program to elucidate potential structures for these proteins.

GTs H, I, J, and N all had the same predicted structure. They all were matched to the structure for the apo sucrose phosphate synthase (SPS) (Figure A1). The SPS protein is a family 1 glycosyltransferase that consists of a GT-B fold consisting of two β/α/β domains that face each other to form a cleft in which lies the active site (Figure A2). In addition to the GT-B structural fold, the family 1 glycosyltransferases utilize a retaining mechanism of action.

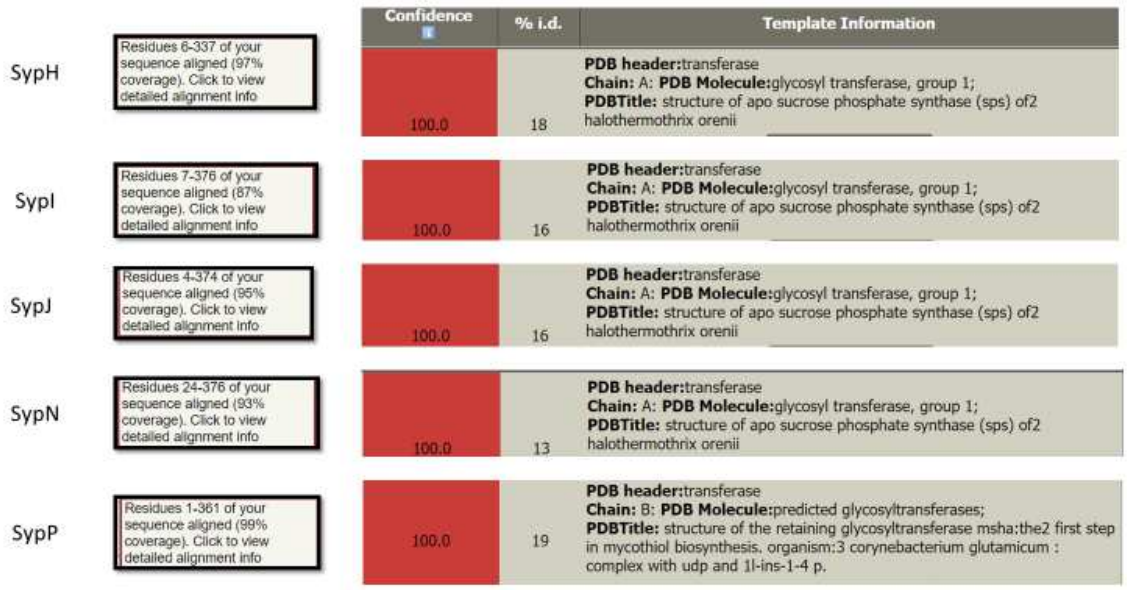

Figure A1. Phyre results for Syp glycosyltansferases H, I, J, N, and P. The top results from the Phyre 2 software show the confidence number in the red box. The percent identity of the query sequence that aligns with the resulting sequence is in the second box. The protein to which the query sequence was aligned is in the box titled template information. Boxed on the left are the percentage of the query sequence that aligned with the structure of the resulting protein [71].

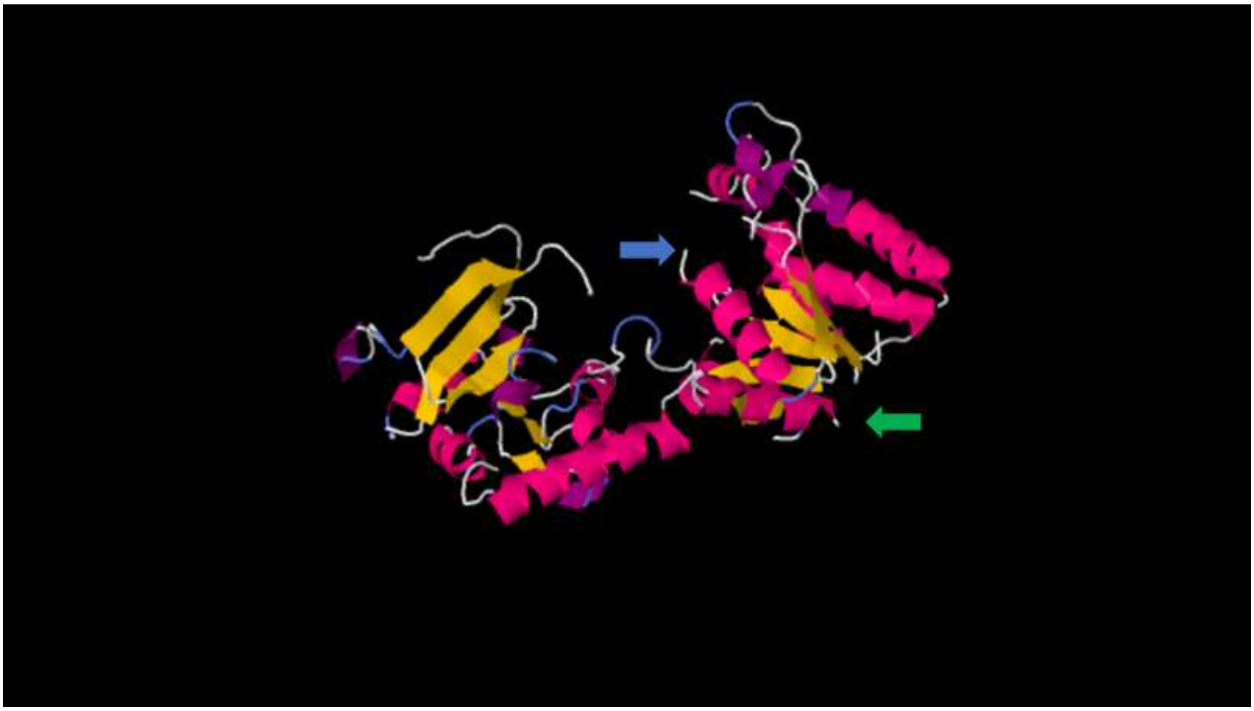

Figure A2. Structure predicted for Syp glycosyltransferases H, I, J, and N by Phyre. The predicted alpha helices are depicted in pink while predicted beta sheets are depicted in yellow. The N-terminal end of the protein is indicated by the blue arrow and the C-terminal end is indicated by the green arrow [71].

The SPS enzyme is responsible for the first step in the synthesis of sucrose: transferring a glucose molecule to fructose-6-phosphate to make sucrose-6-phosphate [91]. After this step, a sucrose phosphatase (SPP) protein removes the phosphate resulting in sucrose.

The SypP protein sequence matched to a different structure: MshA (Figure A1 and A3). The MshA protein is a member of the family 4 glycosyltransferases (GT-4) which, like the GT-1s, utilize a retaining mechanism of action. MshA also has a GT-B structural fold.

Figure A3. Predicted structure for SypP by Phyre. The predicted alpha helices are depicted in pink while predicted beta sheets are depicted in yellow. The N-terminal end of the protein is indicated by the blue arrow and the C-terminal end is indicated by the green arrow [71].

The MshA enzyme catalyzes the first step in mycothiol biosynthesis [92]. This enzyme produces GlcNAc-Ins using UDP-GlcNAc as a donor and 1L-Ins-1-P as an acceptor. This molecule will then undergo four more reactions to eventually become mycothiol. Mycothiol is a thiol commonly produced and used by mycobacteria.

While these data give some insight into the possible contribution of these five Syp GTs to the production of the *syp* polysaccharide, it still leaves many questions unanswered. What do these Gts produce? What are their substrates? What is the significance that the SypP structure is predicted to differ from the other Syp GTs? I hypothesize that these GTs are responsible for producing individual sugars that are then polymerized into a longer polysaccharide chain by SypQ.

Appendix B

Permissions

REFERENCE LIST

1. Nwodo, U.U., Green, E., and Okoh, A. I., *Bacterial exopolysaccharides: functionality and prospects.* Int J Mol Sci, 2012. **13**(11): p. 14002-15.

2. Festucci-Buselli, R.A., Otoni, Wagner C., and Joshi, Chandrashekhar P., *Structure, organization, and functions of cellulose synthase complexes in higher plants.* Brazilian Journal of Plant Physiology, 2007. **19**: p. 1-13.

3. Yang, N.J. and I.M. Chiu, *Bacterial signaling to the nervous system through toxins and metabolites.* J Mol Biol, 2017. **429**(5): p. 587-605.

4. Amspacher, W.H. and A.R. Curreri, *Use of dextran in control of shock resulting from war wounds.* AMA Arch Surg, 1953. **66**(6): p. 730-40.

5. Zannini, E., Waters, D. M., Coffey, A., and Arendt, E. K., *Production, properties, and industrial food application of lactic acid bacteria-derived exopolysaccharides.* Appl Microbiol Biotechnol, 2016. **100**(3): p. 1121-35.

6. Boettcher, K.J. and E.G. Ruby, *Depressed light emission by symbiotic* Vibrio fischeri *of the sepiolid squid* Euprymna scolopes. J Bacteriol, 1990. **172**(7): p. 3701-6.

7. Shibata, S., et al., *Roles of the structural symbiosis polysaccharide (syp) genes in host colonization, biofilm formation, and polysaccharide biosynthesis in* Vibrio fischeri. J Bacteriol, 2012. **194**(24): p. 6736-47.

8. Yip, E.S., Grublesky, B. T., Hussa, E. A., and Visick, K. L., *A novel, conserved cluster of genes promotes symbiotic colonization and sigma-dependent biofilm formation by* Vibrio fischeri. Mol Microbiol, 2005. **57**(5): p. 1485-98.

9. Lairson, L.L., Henrissat, B., Davies, G. J., and Withers, S. G., *Glycosyltransferases: structures, functions, and mechanisms.* Annu Rev Biochem, 2008. **77**: p. 521-55.

10. Campbell, J.A., Davies, G. J., Bulone, V., and Henrissat, B., *A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities.* Biochem J, 1997. **326 (Pt 3)**: p. 929-39.

11. McNamara, J.T., Morgan, Jacob L. W., and Zimmer, Jochen, *A molecular description of cellulose biosynthesis.* Annual Review of Biochemistry, 2015. **84**: p. 895-921.

12. Saxena, I.M. and R.M. Brown, Jr., *Identification of a second cellulose synthase gene (acsAII) in* Acetobacter xylinum*.* J Bacteriol, 1995. **177**(18): p. 5276-83.

13. Saxena, I.M., Brown, R. M., Jr., and Dandekar, T., *Structure--function characterization of cellulose synthase: relationship to other glycosyltransferases.* Phytochemistry, 2001. **57**(7): p. 1135-48.

14. Saxena, I.M. and R.M. Brown, *Identification of cellulose synthase(s) in higher plants: sequence analysis of processive β-glycosyltransferases with the common motif 'D, D, D35Q(R,Q)XRW'.* Cellulose, 1997. **4**(1): p. 33-49.

15. Morgan, J.L., McNamara, J. T., Fischer, M., Rich, J., Chen, H. M., Withers, S. G., and Zimmer, J., *Observing cellulose biosynthesis and membrane translocation in crystallo.* Nature, 2016. **531**(7594): p. 329-34.

16. Omadjela, O., Narahari, A., Strumillo, J., Melida, H., Mazur, O., Bulone, V., and Zimmer, J., *BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis.* Proc Natl Acad Sci U S A, 2013. **110**(44): p. 17856-61.

17. May, J.F., Levengood, M. R., Splain, R. A., Brown, C. D., and Kiessling, L. L., *A processive carbohydrate polymerase that mediates bifunctional catalysis using a single active site.* Biochemistry, 2012. **51**(6): p. 1148-59.

18. Hagen, F.K., Hazes, B., Raffo, R., deSa, D., and Tabak, L. A., *Structure-function analysis of the UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. Essential residues lie in a predicted active site cleft resembling a lactose repressor fold.* J Biol Chem, 1999. **274**(10): p. 6797-803.

19. Bi, Y., Hubbard, C., Purushotham, P., and Zimmer, J., *Insights into the structure and function of membrane-integrated processive glycosyltransferases.* Curr Opin Struct Biol, 2015. **34**: p. 78-86.

20. Morgan, J.L., McNamara, J. T., and Zimmer, J., *Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP.* Nat Struct Mol Biol, 2014. **21**(5): p. 489-96.

21. Wiggins, C.A. and S. Munro, *Activity of the yeast MNN1 alpha-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases.* Proc Natl Acad Sci U S A, 1998. **95**(14): p. 7945-50.

22. Whitney, J.C. and P.L. Howell, *Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria.* Trends Microbiol, 2013. **21**(2): p. 63-72.

23. Ceri, H., Olson, M. E., Stremick, C., Read, R. R., Morck, D., and Buret, A., *The Calgary Biofilm Device: new technology for rapid determination of antibiotic susceptibilities of bacterial biofilms.* J Clin Microbiol, 1999. **37**(6): p. 1771-6.

24. Islam, S.T. and J.S. Lam, *Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway.* Can J Microbiol, 2014. **60**(11): p. 697-716.

25. Daniels, C. and R. Morona, *Analysis of* Shigella flexneri *Wzz (Rol) function by mutagenesis and cross-linking: Wzz is able to oligomerize.* Mol Microbiol, 1999. **34**(1): p. 181-94.

26. Schmid, J., Sieber, V., and Rehm, B., *Bacterial exopolysaccharides: biosynthesis pathways and engineering strategies.* Front Microbiol, 2015. **6**: p. 496.

27. Merighi, M., Lee, V. T., Hyodo, M., Hayakawa, Y., and Lory, S., *The second messenger bis-(3'-5')-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in Pseudomonas aeruginosa.* Mol Microbiol, 2007. **65**(4): p. 876-95.

28. Dols, M., Remaud-Simeon, M., Willemot, R. M., Vignon, M., and Monsan, P., *Characterization of the Different Dextransucrase Activities Excreted in Glucose, Fructose, or Sucrose Medium by Leuconostoc mesenteroides NRRL B-1299.* Applied and Environmental Microbiology, 1998. **64**(4): p. 1298-1302.

29. Ceska, M., *Biosynthesis of levan and a new method for the assay of levansucrase activity.* Biochemical Journal, 1971. **125**(1): p. 209-211.

30. Hobley, L., Harkins, C., MacPhee, C. E., and Stanley-Wall, N. R., *Giving structure to the biofilm matrix: an overview of individual strategies and emerging common themes.* FEMS Microbiol Rev, 2015. **39**(5): p. 649-69.

31. Costerton, J.W., Cheng, K. J., Geesey, G. G., Ladd, T. I., Nickel, J. C., Dasgupta, M., and Marrie, T. J., *Bacterial biofilms in nature and disease.* Annu Rev Microbiol, 1987. **41**: p. 435-64.

32. Branda, S.S., Vik, S., Friedman, L., and Kolter, R., *Biofilms: the matrix revisited.* Trends Microbiol, 2005. **13**(1): p. 20-6.

33. O'Toole, G., Kaplan, H. B., and Kolter, R., *Biofilm formation as microbial development.* Annu Rev Microbiol, 2000. **54**: p. 49-79.

34. O'Toole, G.A. and R. Kolter, *Flagellar and twitching motility are necessary for* Pseudomonas aeruginosa *biofilm development.* Mol Microbiol, 1998. **30**(2): p. 295-304.

35. Bhomkar, P., Materi, W., Semenchenko, V., and Wishart, D. S., *Transcriptional response of* E. coli *upon FimH-mediated fimbrial adhesion.* Gene Regul Syst Bio, 2010. **4**: p. 1-17.

36. Inagaki, S., Kuramitsu, H. K., and Sharma, A., *Contact-dependent regulation of a* Tannerella forsythia *virulence factor, BspA, in biofilms.* FEMS Microbiol Lett, 2005. **249**(2): p. 291-6.

37. Flemming, H.C. and J. Wingender, *The biofilm matrix.* Nat Rev Microbiol, 2010. **8**(9): p. 623-33.

38. Sauer, K., Cullen, M. C., Rickard, A. H., Zeef, L. A., Davies, D. G., and Gilbert, P., *Characterization of nutrient-induced dispersion in* Pseudomonas aeruginosa *PAO1 biofilm.* J Bacteriol, 2004. **186**(21): p. 7312-26.

39. Rowe, M.C., Withers, H. L., and Swift, S., *Uropathogenic* Escherichia coli *forms biofilm aggregates under iron restriction that disperse upon the supply of iron.* FEMS Microbiol Lett, 2010. **307**(1): p. 102-9.

40. Fong, J.C., Syed, K. A., Klose, K. E., and Yildiz, F. H., *Role of* Vibrio *polysaccharide (*vps*) genes in VPS production, biofilm formation and* Vibrio cholerae *pathogenesis.* Microbiology, 2010. **156**(Pt 9): p. 2757-69.

41. Yildiz, F., Fong, J., Sadovskaya, I., Grard, T., and Vinogradov, E., *Structural characterization of the extracellular polysaccharide from* Vibrio cholerae *O1 El-Tor.* PLoS One, 2014. **9**(1): p. e86751.

42. Teschler, J.K., Zamorano-Sanchez, D., Utada, A. S., Warner, C. J., Wong, G. C., Linington, R. G., and Yildiz, F. H., *Living in the matrix: assembly and control of* Vibrio cholerae *biofilms.* Nat Rev Microbiol, 2015. **13**(5): p. 255-68.

43. Branda, S.S., Chu, F., Kearns, D. B., Losick, R., and Kolter, R., *A major protein component of the* Bacillus subtilis *biofilm matrix.* Mol Microbiol, 2006. **59**(4): p. 1229-38.

44. Wang, S., Liu, Xi, Liu, Hongsheng, Zhang, Li, Guo, Yuan, Yu, Shan, Wozniak, Daniel J., and Ma, Luyan Z., *The exopolysaccharide Psl–eDNA interaction enables the formation of a biofilm skeleton in* Pseudomonas aeruginosa*.* Environmental microbiology reports, 2015. **7**(2): p. 330-340.

45. Conrad, A., Suutari, M. K., Keinanen, M. M., Cadoret, A., Faure, P., Mansuy-Huault, L., and Block, J. C., *Fatty acids of lipid fractions in extracellular polymeric substances of activated sludge flocs.* Lipids, 2003. **38**(10): p. 1093-105.

46. Nadell, C.D., Drescher, K., Wingreen, N. S., and Bassler, B. L., *Extracellular matrix structure governs invasion resistance in bacterial biofilms.* Isme j, 2015. **9**(8): p. 1700-9.

47. Deva, A.K., Adams, W. P., Jr., and Vickery, K., *The role of bacterial biofilms in device-associated infection.* Plast Reconstr Surg, 2013. **132**(5): p. 1319-28.

48. Pittet, D., Hulliger, S., and Auckenthaler, R., *Intravascular device-related infections in critically ill patients.* J Chemother, 1995. **7 Suppl 3**: p. 55-66.

49. Van Houdt, R. and C.W. Michiels, *Biofilm formation and the food industry, a focus on the bacterial outer surface.* Journal of Applied Microbiology, 2010. **109**(4): p. 1117-1131.

50. Ruby, E.G., The Euprymna scolopes-Vibrio fischeri symbiosis*: a biomedical model for the study of bacterial colonization of animal tissue.* J Mol Microbiol Biotechnol, 1999. **1**(1): p. 13-21.

51. McFall-Ngai, M.J. and E.G. Ruby, *Symbiont recognition and subsequent morphogenesis as early events in an animal-bacterial mutualism.* Science, 1991. **254**(5037): p. 1491-4.

52. Yip, E.S., et al., *A novel, conserved cluster of genes promotes symbiotic colonization and sigma-dependent biofilm formation by* Vibrio fischeri. Mol Microbiol, 2005. **57**(5): p. 1485-98.

53. Shibata, S., Yip, E. S., Quirke, K. P., Ondrey, J. M., and Visick, K. L., *Roles of the structural symbiosis polysaccharide (*syp*) genes in host colonization, biofilm formation, and polysaccharide biosynthesis in* Vibrio fischeri. J Bacteriol, 2012. **194**(24): p. 6736-47.

54. Hussa, E.A., Darnell, C. L., Visick, K. L., *RscS functions upstream of SypG to control the syp locus and biofilm formation in* Vibrio fischeri. J Bacteriol, 2008. **190**(13): p. 4576-83.

55. Visick, K.L. and L.M. Skoufos, *Two-component sensor required for normal symbiotic colonization of Euprymna scolopes by* Vibrio fischeri. J Bacteriol, 2001. **183**(3): p. 835-42.

56. Ray, V.A., Eddy, J. L., Hussa, E. A., Misale, M., and Visick, K. L., *The* syp *enhancer sequence plays a key role in transcriptional activation by the sigma54-dependent response regulator SypG and in biofilm formation and host colonization by* Vibrio fischeri. J Bacteriol, 2013. **195**(23): p. 5402-12.

57. Morris, A.R., C.L. Darnell, and K.L. Visick, *Inactivation of a novel response regulator is necessary for biofilm formation and host colonization by* Vibrio fischeri. Mol Microbiol, 2011. **82**(1): p. 114-30.

58. Morris, A.R. and K.L. Visick, *The response regulator SypE controls biofilm formation and colonization through phosphorylation of the syp-encoded regulator SypA in* Vibrio fischeri. Mol Microbiol, 2013. **87**(3): p. 509-25.

59. Morris, A.R., Darnell, C. L., and Visick, K. L., *Inactivation of a novel response regulator is necessary for biofilm formation and host colonization by* Vibrio fischeri. Mol Microbiol, 2011. **82**(1): p. 114-30.

60. Michaelis, S., Inouye, H., Oliver, D., and Beckwith, J., *Mutations that alter the signal sequence of alkaline phosphatase in* Escherichia coli. J Bacteriol, 1983. **154**(1): p. 366-74.

61. Stabb, E.V., Reich, K. A., and Ruby, E. G., *Vibrio fischeri genes hvnA and hvnB encode secreted NAD(+)-glycohydrolases.* J Bacteriol, 2001. **183**(1): p. 309-17.

62. Hanahan, D., *Studies on transformation of* Escherichia coli *with plasmids.* J Mol Biol, 1983. **166**(4): p. 557-80.

63. Stabb, E.V. and E.G. Ruby, *RP4-based plasmids for conjugation between* Escherichia coli *and members of the* Vibrionaceae. Methods Enzymol, 2002. **358**: p. 413-26.

64. Ho, S.N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R., *Site-directed mutagenesis by overlap extension using the polymerase chain reaction.* Gene, 1989. **77**(1): p. 51-9.

65. Saiki, R.K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A., *Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.* Science, 1988. **239**(4839): p. 487-91.

66. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Mol Syst Biol, 2011. **7**: p. 539.

67. Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R., *A new bioinformatics analysis tools framework at EMBL-EBI.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W695-9.

68. Altschul, S.F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.

69. Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Honigschmid, P., Schafferhans, A., Roos, M., Bernhofer, M., Richter, L., Ashkenazy, H., Punta, M., Schlessinger, A., Bromberg, Y., Schneider, R., Vriend, G., Sander, C., Ben-Tal, N., and Rost, B., *PredictProtein--an open resource for online prediction of protein structural and functional features.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W337-43.

70. Tsirigos, K.D., Peters, C., Shu, N., Kall, L., and Elofsson, A., *The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides.* Nucleic Acids Res, 2015. **43**(W1): p. W401-7.

71. Kelley, L.A., Mezulis, Stefans, Yates, Christopher M., Wass, Mark N., Sternberg, and Michael J. E., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nat. Protocols, 2015. **10**(6): p. 845-858.

72. Rigg, G.P., Barrett, B., and Roberts, I. S., *The localization of KpsC, S and T, and KfiA, C and D proteins involved in the biosynthesis of the* Escherichia coli *K5 capsular polysaccharide: evidence for a membrane-bound complex.* Microbiology, 1998. **144 (Pt 10)**: p. 2905-14.

73. Woolridge, K.G., Morrissey, Julie A., and Williams, Peter H., *Transport of ferric-aerobactin into the periplasm and cytoplasm of* Escherichia coli *K12: role of envelope-associated proteins and effect of endogenous siderophores.* Microbiology, 1992. **138**(3): p. 597-603.

74. Manoil, C. and J. Beckwith, *TnphoA: a transposon probe for protein export signals.* Proc Natl Acad Sci U S A, 1985. **82**(23): p. 8129-33.

75. Chothia, C. and A.M. Lesk, *The relation between the divergence of sequence and structure in proteins.* Embo j, 1986. **5**(4): p. 823-6.

76. Kaczanowski, S., Zielenkiewicz, and Piotr, *Why similar protein sequences encode similar three-dimensional structures?* Theoretical Chemistry Accounts, 2010. **125**(3): p. 643-650.

77. Schachter, H., Jabbal, I., Hudgin, R. L., Pinteric, L., McGuire, E. J., and Roseman, S., *Intracellular localization of liver sugar nucleotide glycoprotein glycosyltransferases in a Golgi-rich fraction.* J Biol Chem, 1970. **245**(5): p. 1090-100.

78. Keenleyside, W.J., Clarke, A. J., and Whitfield, C., *Identification of residues involved in catalytic activity of the inverting glycosyl transferase WbbE from* Salmonella enterica *serovar* borreze*.* J Bacteriol, 2001. **183**(1): p. 77-85.

79. Ray, V.A., Driks, A., and Visick, K. L., *Identification of a novel matrix protein that promotes biofilm maturation in* Vibrio fischeri. J Bacteriol, 2015. **197**(3): p. 518-28.

80. Pundir, S., Martin, M. J., and O'Donovan, C., *UniProt Protein Knowledgebase.* Methods Mol Biol, 2017. **1558**: p. 41-55.

81. Oglesby, L.L., Jain, S., and Ohman, D. E., *Membrane topology and roles of* Pseudomonas aeruginosa *Alg8 and Alg44 in alginate polymerization.* Microbiology, 2008. **154**(Pt 6): p. 1605-15.

82. Liao, S.-M., Du, Qi-Shi, Meng, Jian-Zong, Pang, Zong-Wen, and Huang, Ri-Bo, *The multiple roles of histidine in protein interactions.* Chemistry Central Journal, 2013. **7**: p. 44-44.

83. Yohannan, S., Yang, D., Faham, S., Boulting, G., Whitelegge, J., and Bowie, J. U., *Proline substitutions are not easily accommodated in a membrane protein.* J Mol Biol, 2004. **341**(1): p. 1-6.

84. von Heijne, G., *Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule.* J Mol Biol, 1992. **225**(2): p. 487-94.

85. Wang, R., Xiong, J., Wang, W., Miao, W., and Liang, A., *High frequency of +1 programmed ribosomal frameshifting in* Euplotes octocarinatus*.* Sci Rep, 2016. **6**: p. 21139.

86. Xie, P., *Dynamics of +1 ribosomal frameshifting.* Math Biosci, 2014. **249**: p. 44-51.

87. Baggett, N.E., Zhang, Yan, and Gross, Carol A., *Global analysis of translation termination in* E. coli. PLoS Genetics, 2017. **13**(3): p. e1006676.

88. Dorfmueller, H.C., et al., *A structural and biochemical model of processive chitin synthesis.* J Biol Chem, 2014. **289**(33): p. 23020-8.

89. Ye, L., Zheng, X., and Zheng, H., *Effect of sypQ gene on poly-N-acetylglucosamine biosynthesis in* Vibrio parahaemolyticus *and its role in infection process.* Glycobiology, 2014. **24**(4): p. 351-8.

90. Itoh, Y., Rice, J. D., Goller, C., Pannuri, A., Taylor, J., Meisner, J., Beveridge, T. J., Preston, J. F., 3rd, and Romeo, T., *Roles of* pgaABCD *genes in synthesis, modification, and export of the* Escherichia coli *biofilm adhesin poly-beta-1,6-N-acetyl-D-glucosamine.* J Bacteriol, 2008. **190**(10): p. 3670-80.

91. Chua, T.K., Bujnicki, Janusz M., Tan, Tien-Chye, Huynh, Frederick, Patel, Bharat K., and Sivaraman, J., *The structure of sucrose phosphate synthase from* Halothermothrix orenii *reveals its mechanism of action and binding mode.* The Plant Cell, 2008. **20**(4): p. 1059-1072.

92. Newton, G.L., Ta, P., Bzymek, K. P., and Fahey, R. C., *Biochemistry of the initial steps of mycothiol biosynthesis.* J Biol Chem, 2006. **281**(45): p. 33910-20.

VITA

Mary Kathryn Flaherty was born in Quincy, IL to Michael and Rita Vilsick. She earned a Bachelor of Arts degree in Philosophy from Quincy University in May 2001 and a Bachelor of Science degree in Biology with a Molecular Biology emphasis from Loyola University in May 2015.

In August 2015, Mary entered the Masters of Science program for Microbiology and Immunology at Loyola University Chicago. She joined the lab of Dr. Karen Visick in 2016 where the focus of her research was to increase understanding of the role of the putative glycosyltransferase SypQ in *Vibrio fischeri* biofilm formation.

After graduation, Mary will continue to feed her passion of microbiology by pursuing a career in the microbiology field.