2015

Signaling between Two Sensor Kinases Controls Biofilms and Host Colonization in a Bacterial Symbiont

Allison N. Norsworthy
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

Part of the Microbiology Commons

Recommended Citation
https://ecommons.luc.edu/luc_diss/1483

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 2015 Allison N. Norsworthy
ACKNOWLEDGEMENTS

There are a seemingly innumerable number of people that I could acknowledge for supporting me during my graduate school career. I must first thank my family, including my parents, for raising me to be an independent, hard worker, and my brothers for teaching me to stand up for myself. Additionally, I would like to thank my friends and colleagues for instilling confidence in me even during the most trying of times. Also, I must acknowledge my committee members, whose invaluable ideas and advice have functioned as blue prints that I can follow to become a great scientist.

Importantly, I would not be here today without the obliging guidance from Karen Visick, a talented and enthusiastic scientist who divulged priceless insight into microbial genetics. There are few labs in the world where I could have learned to be a better geneticist, and I can forever use this knowledge in my future scientific endeavours.
# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** iii

**LIST OF TABLES** vi

**LIST OF FIGURES** vii

**ABSTRACT** ix

## CHAPTER ONE: LITERATURE REVIEW

1. Introduction 1
2. Establishing the symbiotic relationship between *Vibrio fischeri* and *Euprymna scolopes* 4
   1. Introduction 4
   2. Initiating the symbiosis 7
   3. Maintaining a happy marriage 12
   4. Conclusion 15
3. Biofilm formation 15
   1. Introduction 15
   2. Initiation 17
   3. Maturation 19
   4. Dispersal 20
   5. Conclusion 21
4. Two-component signaling 21
   1. Introduction 21
   2. Sensor kinase 23
   3. Phosphorelays 42
   4. Response regulators 44
   5. Atypical TCS pathways 49
   6. Two-component regulation of biofilms in *V. fischeri* 56
   7. Conclusion 63

## CHAPTER TWO: MATERIALS AND METHODS

64

## CHAPTER THREE: EXPERIMENTAL RESULTS

76

1. SypF*: an active sensor kinase 76
   1. Introduction 76
2. SypF*: role as an SK 77
3. SypF*: function within the Syp signaling pathway 82
4. The role of SypF* in host colonization 86
5. Wild-type SypF: unusual role in biofilms 88
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>88</td>
</tr>
<tr>
<td>Requirement for SypF in RscS-induced biofilms</td>
<td>88</td>
</tr>
<tr>
<td>Function of SypF relative to SypG and SypE</td>
<td>90</td>
</tr>
<tr>
<td>Role for both RscS and SypF in biofilms</td>
<td>93</td>
</tr>
<tr>
<td>Requirement for SypF during host colonization</td>
<td>99</td>
</tr>
<tr>
<td>SypF as a negative regulator</td>
<td>101</td>
</tr>
<tr>
<td>CHAPTER FOUR: DISCUSSION</td>
<td>106</td>
</tr>
<tr>
<td>Introduction</td>
<td>106</td>
</tr>
<tr>
<td>SypF: surprising results</td>
<td>107</td>
</tr>
<tr>
<td>SypF* vs SypF</td>
<td>108</td>
</tr>
<tr>
<td>Evolution of SypF and its proposed function</td>
<td>110</td>
</tr>
<tr>
<td>Evolution of the Syp TCS pathway</td>
<td>113</td>
</tr>
<tr>
<td>Significance</td>
<td>114</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>122</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>130</td>
</tr>
<tr>
<td>VITA</td>
<td>170</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Tables</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <em>V. fischeri</em> strains used in this study</td>
<td>71</td>
</tr>
<tr>
<td>2. Plasmids used in this study</td>
<td>73</td>
</tr>
<tr>
<td>3. Primers used in this study</td>
<td>75</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figures</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Colonization of <em>Euprymna scolopes</em> by <em>Vibrio fischeri</em></td>
<td>5</td>
</tr>
<tr>
<td>2. Steps of biofilm formation</td>
<td>16</td>
</tr>
<tr>
<td>3. Canonical architecture of two-component signaling</td>
<td>22</td>
</tr>
<tr>
<td>4. Structural organization of selected sensor kinases</td>
<td>24</td>
</tr>
<tr>
<td>5. PDC structure and function</td>
<td>27</td>
</tr>
<tr>
<td>6. HAMP domain arrangement</td>
<td>31</td>
</tr>
<tr>
<td>7. Structures of cytoplasmic sensing domains: PAS and GAF</td>
<td>34</td>
</tr>
<tr>
<td>8. Structures of the core enzymatic regions of an SK</td>
<td>38</td>
</tr>
<tr>
<td>9. Response regulator domains</td>
<td>46</td>
</tr>
<tr>
<td>10. Lux TCS pathway controlling bioluminescence</td>
<td>50</td>
</tr>
<tr>
<td>11. Atypical TCS pathways, Gac and Rcs</td>
<td>53</td>
</tr>
<tr>
<td>12. Syp TCS pathway</td>
<td>57</td>
</tr>
<tr>
<td>13. Conserved domains found within Syp regulators</td>
<td>58</td>
</tr>
<tr>
<td>14. SypF sequence alignment with known functional SKs</td>
<td>78</td>
</tr>
<tr>
<td>15. SK activity of SypF*</td>
<td>80</td>
</tr>
<tr>
<td>16. Wrinkled colony phenotype and protein levels of SypF*-FLAG variants</td>
<td>83</td>
</tr>
<tr>
<td>17. Function of SypF* relative to SypE and/or SypG</td>
<td>85</td>
</tr>
<tr>
<td>18. Role of RscS in SypF*-induced biofilms</td>
<td>87</td>
</tr>
</tbody>
</table>
19. Role of SypF in RscS-induced phenotypes 89
20. Determining where SypF functions in the Syp pathway 92
21. Phenotypes of SypF phosphotransfer mutants 95
22. Function of single HPt domain of SypF 96
23. Determining whether RscS and SypF interact to promote biofilms 98
24. Mechanism by which SypF functions in vivo 100
25. RscS-SypF<sup>HPt</sup> interaction in vivo 102
26. Biofilms produced from single-copy expression of RscS or the chimera 104
27. Levels of SypH<sup>HA</sup> under different genetic conditions 120
28. Role for the HPt domain of RscS in biofilms 123
Organisms within all domains of life must acclimate to fluctuating environments to survive. To do this, cells utilize sensory circuits, which function to connect environmental stimuli to an intracellular response. One common sensory pathway utilized by bacteria is two-component signaling (TCS), composed of an environmental sensor (the sensor kinase, SK) and a cognate, intracellular effector (the response regulator, RR). The marine bacterium *Vibrio fischeri* uses an elaborate TCS phosphorelay containing a hybrid SK, RscS, and two RRs, SypE and SypG, to colonize its natural squid host, *Euprymna scolopes*. This TCS pathway regulates *V. fischeri*’s ability to form a biofilm, or a community of cells encased in an extracellular matrix, a process required to initiate colonization. Between the *sypE* and *sypG* genes lies *sypF*, which encodes another putative hybrid SK. Due to its location and predicted function, I hypothesized that *sypF* might also regulate biofilms. Indeed, I found that SypF was critical for biofilms by functioning downstream of RscS to directly control SypE and SypG. Surprisingly, although a mutant variant of SypF, SypF*, functioned as an SK both *in vitro* and *in vivo*, this did not seem to be the case for wild-type SypF. Specifically, wild-type SypF exhibited SK activity *in vitro*, but this activity was dispensable for colonization. In fact, only a single non-enzymatic domain within SypF, the HPt domain, was critical *in vivo*. Remarkably, this domain within SypF directly interacted with RscS, permitting a bypass of RscS’s own HPt domain and SypF’s enzymatic function. These findings represent the first *in vivo* example of a functional SK that exploits the enzymatic activity of another
SK, an adaptation that demonstrates the elegant plasticity in the arrangement of TCS regulators. This flexibility in TCS pathways likely permits bacteria to manage a vast repertoire of different environments, thus promoting their survival both inside and outside a host.
CHAPTER ONE

LITERATURE REVIEW

I. Introduction

Out of all three domains of life, the bacterial domain exhibits considerable phylogenetic diversity, the largest biomass, and the most extensive geographical spread. These single-celled organisms are incredibly successful forms of life to inhabit the earth, where they flourish in habitats ranging from deep-sea hydrothermal vents, to scum growing on a lakebed, and to the gastrointestinal tracts of humans (reviewed in (Orcutt et al, 2011; Rastogi et al, 2013; Redinbo, 2014)). Life as we know it would not exist if these microbes were not available to maintain the earth’s ecosystems, including the ecosystems within our own bodies (McFall-Ngai et al, 2013).

Bacteria and eukaryotes have co-existed since their phylogenetic split into two domains of life almost one billion years ago (Woese, 1987). Historically, microbial research focused on how a bacterium’s association with an animal leads to disease; however, it is now accepted that bacteria benefit most, if not all, animals by providing nutrients, proper training of the host’s immune system, and protective barriers against potentially pathogenic organisms (Chow et al, 2010). The importance of these microbiota-host interactions is made clear through the observation that a disruption of the microbiota can lead to disease, e.g. antibiotic treatment causing Clostridium difficile-induced enterocolitis (George et al, 1978; Lusk et al, 1978). To understand how these host-microbial interactions are established and maintained to promote the health of
animals, microbial research focuses on identifying and elucidating key molecular mechanisms in bacteria that are important for these processes.

To successfully associate with an animal, a microbe must cope with the multiple challenges presented by the host, such as fluctuations in pH (Bearson et al., 1996; Lober et al., 2006), nutrient sources (Lynch & Sonnenburg, 2012; Steiner & Malke, 2000), the presence of other microorganisms (Cosson et al., 2002; Costello et al., 2014; Li et al., 2014), antimicrobials (Shiloh et al., 2008; Strempel et al., 2013; Yang et al., 2013), and other components of the immune system (Jungnitz et al., 1998; Voyich et al., 2004). To manage these environments, bacteria have a large repertoire of signaling pathways that convert extracellularly derived signals into intracellular responses (Galperin, 2005). The most common signaling cascade used by bacteria is the two-component signaling (TCS) pathway, which consists of a signal-sensing sensor kinase and a cognate response regulator that elicits an output within the cell (Galperin, 2005). Bacteria often contain multiple TCS pathways (even as many as 150 per genome (Ulrich & Zhulin, 2010)), all of which have evolved to recognize a particular environmental signal to induce a specific intracellular response. These cascades are critical for host-association due to their efficiency in permitting a bacterial cell to gauge and respond to the extensive array of both favorable and lethal host-environments.

One important phenotype that promotes bacteria-host associations is the formation of a biofilm, or a community of cells encased in an extracellular matrix (Hall-Stoodley et al., 2004). Although many early studies of bacteria focused on their function as independent, planktonic cells, it is now appreciated that most, if not all, bacteria can form a biofilm under particular environmental conditions. These environments can
include host tissues, and many beneficial and pathogenic organisms use this process to invade or persist within an animal (Heindl et al., 2014; Joo & Otto, 2012; Percival & Suleman, 2014; Ramey et al., 2004; Yildiz & Visick, 2009). Importantly, cells within a biofilm are more tolerant towards antimicrobials (Anwar et al., 1989; Larsen & Fiehn, 1996; Nickel et al., 1985), and can shield themselves from host immune cells (Jesaitis et al., 2003; Leid et al., 2002; Vuong et al., 2004). Thus, understanding this bacterial developmental process provides important insights into the human-microbiota mutualism, and can advance the development of novel treatments for diseases caused by pathogenic bacteria.

Much research has probed how bacteria interact with a host. To ask in-depth questions about host-microbe interactions, researchers often study “simplified” model systems, in which only one or a few bacterial species successfully infect a host (McFall-Ngai et al., 2013). One model system used for this purpose is the symbiosis between the luminescent marine bacterium, *Vibrio fischeri*, and the nocturnal squid, *Euprymna scolopes*. In this symbiosis, *V. fischeri* is the only bacterium capable of colonizing a specialized symbiotic organ, the light organ (Wei & Young, 1989). This monospecific association permits researchers to ask deeply reductionist questions about bacteria-host interactions, and can provide insights into how both players contribute to the initiation and maintenance of this relationship. Indeed, the work presented in this dissertation has identified an unusual molecular mechanism that has evolved in *V. fischeri* to permit the exclusive interaction with *E. scolopes*. At the heart of this novel discovery is a critical TCS cascade in *V. fischeri* that controls biofilm formation, an early step of colonization that is required for efficient entry into *E. scolopes*. Below, I will review the literature that
focuses on three relevant areas of research: the *Vibrio*-squid symbiosis, biofilm development, and TCS pathways.

II. Establishing the symbiotic relationship between *Vibrio fischeri* and *Euprymna scolopes*

Introduction:

The symbiosis between *E. scolopes* and *V. fischeri* has been studied for almost 30 years (Wei & Young, 1989), and has expanded our knowledge about mutualisms between a bacterium and its host (Norsworthy & Visick, 2013; Nyholm & McFall-Ngai, 2004). This is due to the relative ease of manipulating both organisms in the laboratory: adult *E. scolopes* can be maintained within an artificial marine system, the genome of *V. fischeri* has been sequenced, and the bacterium is both easy to grow and genetically manipulate. Furthermore, colonization of *E. scolopes* by *V. fischeri* occurs within a few hours, and all steps of colonization can be directly visualized using fluorescence microscopy. This simple yet robust model system permits researchers the opportunity to develop and test hypotheses that may not be as easily examined or interpreted in other symbiosis models. Indeed, experiments performed with this system have exposed this seemingly simple symbiosis as being exceptionally complex.

Upon hatching, juvenile *E. scolopes* are aposymbiotic (Wei & Young, 1989), and must recruit *V. fischeri* from the surrounding seawater, which contains a complex consortium of bacterial species (Fig 1). Although *V. fischeri* represents only ~0.1% of the total bacterial population in seawater (Lee & Ruby, 1992; Ruby & Lee, 1998), it is the only organism capable of colonizing a specialized organ, known as the light organ, within
Figure 1. Colonization of E. scolopes by V. fischeri. (A) Image of a newly hatched juvenile squid that has not yet been colonized by V. fischeri. Dotted lines flank the light organ within the mantle cavity of the squid. (B, C, D) Cartoon representing an enlarged image of one half of the bilobed light organ. The ink sac is indicated in dark grey. Ciliated epithelial fields are yellow. Pores lead into the ducts (blue), antechambers (blue), and end at the crypt spaces (dark blue). Nitric oxide (NO) is produced at high levels during early stages of colonization, then these levels decrease once V. fischeri cells enter the light organ. (B) V. fischeri form a biofilm-like aggregate outside the light organ. (C) V. fischeri cells leave this aggregate, produce flagella, and traverse into and through the light organ to the crypt spaces. (D) V. fischeri cells grow to high cell density within the crypt spaces and bioluminesce (translucent blue halo). Figure adapted from (Nyholm & McFall-Ngai, 2004).
the squid. To recruit *V. fischeri* cells, *E. scolopes* ventilates seawater into its mantle cavity, which houses the light organ. This organ is bi-lobed and lined by ciliated epithelial fields, and each lobe contains three pores that lead into its interior (McFall-Ngai & Montgomery, 1990; Montgomery & McFall-Ngai, 1998; Sycuro *et al.*, 2006). Beyond the six pores lie ducts that open up into antechamber spaces that ultimately lead to the crypt spaces, or six finger-like projections that house *V. fischeri* (McFall-Ngai & Montgomery, 1990; Montgomery & McFall-Ngai, 1998) (Fig 1). One *V. fischeri* cell per crypt space is sufficient for colonization, where it then multiples to ~$10^{11}$ cells (Dunn *et al.*, 2006; Montgomery & McFall-Ngai, 1998). At this high cell density, *V. fischeri* cells bioluminesce, a phenotype that is important for the persistence of the symbiosis (Koch *et al.*, 2014; Visick *et al.*, 2000) (Fig 1). In exchange for providing a nutrient-rich environment, *E. scolopes* utilizes the light produced by *V. fischeri* to mask its silhouette cast by moonlight. This process, known as counterillumination, is hypothesized to protect the squid from predation while it hunts for food at night (Jones & Nishiguchi, 2004; Koch *et al.*, 2014; Wei & Young, 1989).

Although events that occur within the first few hours after the squid hatch are critical for promoting this specific host-microbe interaction, dynamic processes occur after initiation to further establish and maintain the symbiosis. For example, the squid are on a diel cycle, in which they bury themselves in the sand during the day, and then emerge at night to forage for food (Boettcher *et al.*, 1996). At dawn, when the squid sense environmental cues to rebury themselves, they expel ~95% of the *V. fischeri* cells, along with other light organ contents, into the surrounding seawater (Boettcher *et al.*, 1996; Nyholm & McFall-Ngai, 1998). Over the day, the remaining *V. fischeri* cells repopulate
the light organ. The dynamics of these processes are reflected by transcriptional studies of these organisms over the diel cycle: both *V. fischeri* and *E. scolopes* exhibit distinctive transcriptomes that correspond to different time points over a 24 hour cycle (Wier et al., 2010).

There are a few basic steps of colonization, each of which require that both *V. fischeri* and *E. scolopes* contribute mechanisms that establish this mono-association. In the following section, I will describe these key events that occur during the initiation and persistence stages, which establish this symbiosis for the remainder of the squid’s life.

**Initiating the symbiosis:**

Newly hatched squid are aposymbiotic and must acquire *V. fischeri* cells from the surrounding seawater. Bringing *V. fischeri* and *E. scolopes* together, however, can be considered a limiting factor in colonization. As mentioned above, *V. fischeri* represent as little as 0.1% of the total bacterial population in the Hawaiian seawater that is vented into the mantle cavity of the squid (Nyholm & McFall-Ngai, 2004) It has been estimated that a miniscule volume of seawater (1.3 µl) enters the mantle during each half-second ventilation, and thus only a few *V. fischeri* cells enter the mantle cavity during this short time period. Theoretically, *V. fischeri* cells would have to locate all six pores in a brief amount of time before they are expelled from this cavity (Nyholm & McFall-Ngai, 2004; Nyholm et al, 2000). So how do both the microbe and the host facilitate this specific interaction?

**The host’s role in initiation.** Immediately after hatching, *E. scolopes* is competent to interact with any bacterium found within seawater, not just *V. fischeri*. For example, ventilation by the squid brings water and any bacterial cells into the mantle
cavity where the light organ is located (McFall-Ngai & Montgomery, 1990). The surface of the light organ is lined with epithelial fields with cilia that circulate the water. This motion draws seawater towards the six pores that lead into the light organ, and similarly coaxes bacterial cells into the vicinity of these regions. The presence of microbe associated molecular patterns (MAMPs), including peptidoglycan, induces *E. scolopes* to secrete mucus on the surface of the light organ, which could facilitate its interaction with bacteria in the vented seawater (Nyholm *et al.*, 2002; Nyholm *et al.*, 2000).

1-2 hours after the squid hatch, *E. scolopes* and *V. fischeri* both utilize mechanisms to permit *V. fischeri* to outcompete the consortium of different bacterial species within seawater (Nyholm & McFall-Ngai, 2003; Nyholm *et al.*, 2000). *E. scolopes* secretes various antimicrobials that *V. fischeri* must combat. These molecules include reactive oxygen species (ROS) (Schleicher & Nyholm, 2011; Small & McFall-Ngai, 1999; Visick & Ruby, 1998) and nitric oxide (NO) (Davidson *et al.*, 2004). *V. fischeri*’s response to ROS seems to be a straight-forward story; for example, the bacterium encodes enzymes to detoxify ROS secreted by the squid (Ruby *et al.*, 2005; Schleicher & Nyholm, 2011; Visick & Ruby, 1998), although the exact interplay between ROS and these enzymes (and potentially other uncharacterized ROS-related enzymes in *V. fischeri*) remains unknown. Studies elucidating the role of NO in the symbiosis, however, revealed unexpected results. In the context of other host-microbe interactions, NO generally functions as an antimicrobial due to its participation in secondary reactions that produce reactive nitrogen species (RNS) (Bowman *et al.*, 2011); however, NO in the squid-*Vibrio* symbiosis seems to play a positive role for this interaction. During the initiation stage of colonization, many different species of bacteria, including *V. fischeri*, can congregate
outside the light organ. Interestingly, NO produced by the squid inhibits the accumulation of these non-\textit{V. fischeri} cells, and consequentially enriches \textit{V. fischeri} cells (Collins \textit{et al}, 2012; Davidson \textit{et al}, 2004). After successful colonization by \textit{V. fischeri}, the squid downregulate their expression of NOS, the enzyme which produces NO (Davidson \textit{et al}, 2004) (Fig 1). This decrease is attributed to the symbiont’s release of two MAMPs, lipopolysaccharide (LPS) and TCT, a component of the peptidoglycan (Altura \textit{et al}, 2011). Together, these data suggest that NO plays a key role in the cross talk between \textit{E. scolopes} and \textit{V. fischeri}.

**How \textit{V. fischeri} initiates colonization.** As mentioned above, \textit{V. fischeri} cells encode anti-ROS enzymes to potentially cope with ROS secreted by the squid, respond to NO produced by the squid, and release MAMPs that function a host-microbe signal; however, additional, key processes are implemented by \textit{V. fischeri} to establish the specific relationship with \textit{E. scolopes}. These processes include the formation of a biofilm-like aggregate on the surface of the light organ. Subsequently, \textit{V. fischeri} cells leave this aggregate, migrate through the pores against outward water currents produced by beating cilia, and traverse across the antechamber and into the crypt spaces in the organ (Nyholm & McFall-Ngai, 2004) (Fig 1). For these latter processes to occur, \textit{V. fischeri} cells utilize flagella for locomotion and chemotaxis proteins to alter the direction of movement towards their final destination.

One key phenotype of \textit{V. fischeri} required for colonization is the ability to form a biofilm-like aggregate on the surface of the light organ. This occurs after \textit{V. fischeri} interact with cilia and mucous on the host’s epithelial cells (Altura \textit{et al}, 2013; Nyholm \textit{et al}, 2000). After this initial contact with the host, \textit{V. fischeri} cells coalesce into an
aggregate that can include bacterial species other than *V. fischeri* (Fig. 1) (Nyholm & McFall-Ngai, 2003; Nyholm *et al*, 2000). Ultimately, *V. fischeri* cells dominate over other bacteria within the aggregate through an unknown mechanism (Altura *et al*, 2013; Nyholm & McFall-Ngai, 2003). This phenotype is controlled by a complex two-component signaling pathway known as Syp (Hussa *et al*, 2008; Yip *et al*, 2006; Yip *et al*, 2005), which functions to regulate the production of a polysaccharide that permits the cells to stick to each other and to the surface of the light organ. *V. fischeri* strains that fail to form this aggregate fail to colonize the squid; conversely, cells that form an enhanced aggregate exhibit an enhancement of colonization (Millikan & Ruby, 2002; Morris & Visick, 2013b; Nyholm *et al*, 2000; Yip *et al*, 2006). Aggregate formation represents a critical step of colonization, wherein cell-cell and/or cell-host attachment facilitates an enrichment of *V. fischeri* cells at the appropriate location to initiate colonization.

Once *V. fischeri* cells aggregate outside the squid’s light organ, they must leave this matrix-encased biofilm and begin their trip into and through intermediate regions of the light organ before they can establish their niche in the crypt spaces. To do this, *V. fischeri* cells have a tuft of 1-5 sheathed flagella at one pole that promotes their locomotion (McCarter, 2001). Studies have demonstrated that flagellar-dependent motility is required for early stages of host colonization, as non-motile or hypermotile strains exhibit colonization defects (Brennan *et al*, 2013; Graf *et al*, 1994; Millikan & Ruby, 2002; Millikan & Ruby, 2004; Wolfe *et al*, 2004). Interestingly, cells begin the colonization process flagellated, but they lose these appendages within the light organ, suggesting that motility is not important within this environment (Ruby & Asato, 1993). Once released from the light organ during venting from the squid at dawn, *V. fischeri*
cells again produce flagella (Ruby & Asato, 1993) presumably allowing *V. fischeri* to establish an interaction with a nascent host.

To identify and reach the colonization-permissive locations within *E. scolopes*, *V. fischeri* cells can chemotax, a mechanism that allows bacteria to sense and move towards attractants and away from repellants (see reviews (Manson *et al*., 1998; Sourjik & Wingreen, 2012; Wadham & Armitage, 2004)). This random biased movement up or down chemical gradients depends upon a complex signaling pathway in which receptors, or methyl-accepting chemotaxis proteins (MCPs), regulate the activity of a two-component signaling pathway that controls flagellar rotation. Importantly, *V. fischeri* cells require conserved chemotaxis genes for colonization (Deloney-Marino & Visick, 2012; Hussa *et al*., 2007). This suggests that *V. fischeri* cells respond to chemogradients, and that this promotes efficient host colonization. Indeed, *V. fischeri* can chemotax to squid-associated sugars *in vitro*: N-acetylneuraminic acid (NANA) and two chitin components, the monosaccharide, GlcNAc, and the disaccharide, (GlcNAc)₂ (DeLoney-Marino *et al*., 2003; Heath-Heckman & McFall-Ngai, 2011; Mandel *et al*., 2012; Nyholm *et al*., 2000). Importantly, disruption of the (GlcNAc)₂ gradient during colonization prevented *V. fischeri* from entering the squid’s ducts; the bacteria would form aggregates around the pore, but they rarely entered the light organ (Mandel *et al*., 2012). (GlcNAc)₂ is most likely one of many chemicals excreted by the squid that direct *V. fischeri* to the crypt spaces. Whether other chemogradients exist, or whether they are important for *V. fischeri* to navigate through different anatomical regions of the squid, remains to be addressed. Identifying host-derived signals that direct a bacterium’s movement towards colonization-permissive sites is an exciting research area that is awaiting exploration.
Maintaining a happy marriage

Although key events occur during the initiation of symbiosis to ensure *V. fischeri* and *E. scolopes* recognize one another, critical post-initiation pathways exist to promote this exclusivity for the remainder of the squid’s life. While *V. fischeri* cells successfully reach and populate the crypt spaces and bioluminesce, the squid undergoes morphological changes. Furthermore, the host’s immune system learns to tolerate *V. fischeri*, both by decreasing the expression of potential antimicrobials, such as NO (described above), and through preventing immune cells from eliminating *V. fischeri*.

**Host factors.** Colonization of *E. scolopes* by *V. fischeri* induces significant morphological changes within the light organ of the squid (Foster & McFall-Ngai, 1998; McFall-Ngai & Ruby, 1991; Montgomery & McFall-Ngai, 1994). These changes are hypothesized to prevent subsequent infection by other bacteria and promote maintenance of the strains of *V. fischeri* that successfully colonized. For example, on the outer surface of the squid’s light organ, mucous shedding halts (Nyholm et al., 2002), and the ciliated epithelial cells irreversibly undergo apoptosis (Foster & McFall-Ngai, 1998; McFall-Ngai & Ruby, 1991; Montgomery & McFall-Ngai, 1994). This apoptosis leads to regression of the ciliated appendages. Within the light organ the pores and ducts constrict, and their associated epithelial cells fill with inclusions that have no known function (Claes & Dunlap, 2000; Kimbell & McFall-Ngai, 2004). Conversely, the epithelial cells lining the crypt spaces engorge, decorate their surfaces with additional microvilli, and increase mucous secretion (Lamarcq & McFall-Ngai, 1998; Montgomery & McFall-Ngai, 1994; Nyholm et al., 2000). Combined, these physiological changes mark the end of active accumulation of *V. fischeri* on the light organ surface, and ensure safekeeping of internal
*V. fischeri* in the light organ. Although growth of *V. fischeri* could lead to bacterial overcrowding and thus burden the squid, *E. scolopes* prevents this problem by venting ~95% of *V. fischeri* every day at dawn, allowing the remaining *V. fischeri* cells to repopulate the light organ (Ruby & Asato, 1993).

This host immune system also plays a role in maintaining *V. fischeri* within the light organ of the squid. Squid are invertebrates; thus, they do not have an adaptive immune system. Instead, *E. scolopes* utilizes innate immunity pathways including complement (Castillo *et al.*, 2009), ROS (Heath-Heckman & McFall-Ngai, 2011; Schleicher & Nyholm, 2011; Small & McFall-Ngai, 1999; Tomarev *et al.*, 1993; Weis *et al.*, 1996), and antimicrobial peptides (Nyholm & McFall-Ngai, 2004). It is unknown how these antimicrobials affect *V. fischeri* when the bacterium is within the light organ. The main immune-cell effector in *E. scolopes* is the hemocyte, a macrophage-like cell that circulates through the blood of the squid and can be found within the light organ of adults and juveniles (Koropatnick *et al.*, 2004; Nyholm & McFall-Ngai, 1998). The presumed roles of the hemocyte are to phagocytose cellular debris and potentially pathogenic organisms, while tolerating *V. fischeri*. Indeed, hemocytes isolated from squid that are colonized by *V. fischeri* readily bind to non-symbiotic species of bacteria, yet they fail to efficiently recognize *V. fischeri* (Nyholm *et al.*, 2009). This represents one method used by *E. scolopes* to permit persistence of *V. fischeri* within the light organ, but the exact mechanism of this tolerance remains unknown. *E. scolopes* encodes many regulators known to function within the NFκb pathway, a signaling cascade famous for controlling immune responses in eukaryotes (Goodson *et al.*, 2005). Perhaps the host hemocytes use
factors within this pathway to recognize an immune-suppressive MAMP secreted by \textit{V. fischeri}, thus allowing these hemocytes to ignore the presence of \textit{V. fischeri}.

**Bacterial factors.** When \textit{V. fischeri} finally reach the crypt spaces, they develop into small, spherically shaped cells and lose their flagella (Ruby & Asato, 1993). Importantly, they grow to high cell density and begin to bioluminesce (Boettcher \textit{et al}., 1996). Bioluminescence is a key component of the symbiosis; the squid supply a nutrient-rich niche, while the bacteria provide light to the squid that is used as camouflage to avoid predation (Jones & Nishiguchi, 2004; Ruby, 1996). The importance of this phenotype to the symbiosis was established when it was determined that mutants of \textit{V. fischeri} that were unable to produce light failed to persist in symbiosis (Visick \textit{et al}., 2000). Because bioluminescence is a critical component of the symbiosis, perhaps cells that cannot bioluminesce are expelled from the light organ, which would permit bioluminescence-competent (and thus symbiosis-competent) cells to proliferate within the squid.

Bioluminescence is clearly an important phenotype of \textit{V. fischeri} that is required for this organism to persist within the squid. Outside of bioluminescence, however, little is known about what other factors within \textit{V. fischeri} are integral for its long-term interaction with \textit{E. scolopes}. One study suggested that a \textit{flaA} mutant, which encodes the major flagellar structural subunit, is preferentially “kicked-out” of the light organ of the squid (Millikan & Ruby, 2004). Although the primary function of flagella lies in providing motility to a cell, they can also mediate adherence to a surface. Perhaps the flagella, or other similarly adherent appendages found on the surface of \textit{V. fischeri}, are important for attachment to the epithelium lining the light organ (Ruby \textit{et al}., 2005). In
fact, *V. fischeri* encodes 10 pilus loci that could facilitate surface attachment (Ruby *et al.*, 2005); however, the role of these or other appendages in persistence of the symbiosis remains unknown.

**Conclusion**

For such a seemingly simple symbiosis, the interaction between *V. fischeri* and *E. scolopes* requires numerous, complicated physiological processes, provided by both organisms, to promote host specificity and colonization. Astonishingly, *E. scolopes* contains an entire organ dedicated to this symbiosis that undergoes dramatic morphological changes after it recruits *V. fischeri*. *V. fischeri* cells, too, must adapt to the changing environments they encounter during the steps of colonization, which includes responding to signals that can either prevent or promote bacterial-host interactions. The roles of known pathways, and the discovery of new physiological processes, are active areas of research that will surely uncover new mechanisms that will expand the knowledge of how bacteria and a host establish a life-long, beneficial relationship.

**III. Biofilm formation**

**Introduction**

Rather than exist as independent, planktonic cells, bacteria in their natural environments often form complex cellular communities encased in an extracellular matrix. This important developmental process is known as biofilm formation, and is known to have environmental, industrial, and human-health implications (reviewed in (Beloin *et al.*, 2014; Cunliffe *et al.*, 2011; Dobretsov *et al.*, 2013). There are three basic steps involved in biofilm development: initiation, maturation, and dispersal (Davies *et al.*, 1998; Kolodkin-Gal *et al.*, 2010; Ma *et al.*, 2009; Sauer *et al.*, 2002) (Fig 2), and
Figure 2. Three major steps of biofilm formation. (A) Initiation: planktonic cells can reversibly attach to a surface. Under biofilm promoting conditions, cells irreversibly attach to this surface and begin to secrete matrix components (green). (B) Maturation: cells within the biofilm grow and divide, and/or planktonic cells can join the community. The biofilm grows into a 3D structure, with cells encased within the matrix. Cells within the biofilm are generally non-motile. (C) Dispersal: environmental conditions signal a subpopulation of cells to leave the biofilm. This involves matrix degradation, which permits motile cells to escape the community to begin the biofilm developmental process in a new location. Figure adapted from (Stoodley et al, 2002).
much research in the past decade has focused on how bacteria progress through these steps to both form and dissolve these communities.

**Initiation**

**Surface sensing.** Specific environmental signals sensed by a bacterium disfavor the independent, planktonic state and favor the assembly of a community (O'Toole & Kolter, 1998b; Pringle & Fletcher, 1986; Stanley, 1983). In the presence of biofilm-inducing signals, bacteria first reversibly attach to a surface, and, upon subsequent irreversible attachment to this surface, they begin the development of a community (Hinsa *et al.*, 2003; Vigeant *et al.*, 2002), although it should be noted that “irreversible attachment” does not imply that the biofilm is fixed in a particular location; in fact it has been demonstrated that some mature biofilms can move across surfaces (Klausen *et al.*, 2003; Tolker-Nielsen *et al.*, 2000). Often, initial attachment is mediated by structures on the surface of a bacterium, such as flagella (Cairns *et al.*, 2013), pili (Li *et al.*, 2012), LPS (Makin & Beveridge, 1996), or a combination of these molecules (O'Toole & Kolter, 1998a; Pratt & Kolter, 1998; Watnick & Kolter, 1999). There are examples of flagella and pili that act as mechanosensors which, when interacting with a surface, induce pathways involved in biofilm development (Cairns *et al.*, 2013; Li *et al.*, 2012). In *Bacillus subtilis*, for example, inhibition of flagellar rotation activates a TCS system, DegS-DegU, which differentially controls both flagellation and biofilm development according to the levels of phospho–DegU (Kobayashi, 2007; Stanley & Lazazzera, 2005). This regulatory process allows for the downregulation of flagellar-related genes and the upregulation of biofilm-related genes, an event that similarly occurs in other bacterial species after they commit to a stationary life. Additionally, some bacteria may sense surfaces in a manner
independent of cell membrane-associated appendages. For example, *E. coli* can utilize the CpxRA TCS system (originally studied for its role in detecting membrane stress (Danese *et al*, 1995; Pogliano *et al*, 1997)) to sense membrane perturbations caused by contact with hydrophobic terrain (Otto & Silhavy, 2002). Furthermore, CpxRA activation has been linked to pili synthesis (Hung *et al*, 2001); thus membrane perturbations sensed by CpxRA could promote the production of pili and therefore surface-adherence. Overall, these various surface-sensing mechanisms utilized by model organisms such as *B. subtilis* and *E. coli* likely apply to many other species of bacteria; however, it remains possible that different bacteria have evolved unique methods of surface-recognition that remain to be discovered.

**The biofilm matrix.** When a bacterial cell (A) receives the appropriate biofilm-inducing signals and (B) irreversibly attaches to a surface, it is poised to initiate the process of biofilm development. This process often begins with the production and secretion of extracellular polymeric substances (EPS), such as polysaccharide (Branda *et al*, 2006; Danese *et al*, 2000; Davies *et al*, 1993; Kaplan *et al*, 2004; Vuong *et al*, 2004; Wang *et al*, 2004; Watnick *et al*, 1999; Wozniak *et al*, 2003), protein (Branda *et al*, 2006; Diggle *et al*, 2006; Jurcisek & Bakaletz, 2007; Kaplan *et al*, 2004; Lasa & Penades, 2006; Romero *et al*, 2010; Zogaj *et al*, 2001), and extracellular DNA (eDNA) (Jurcisek & Bakaletz, 2007; Vilain *et al*, 2009; Whitchurch *et al*, 2002). Although the presence of a matrix is a defining characteristic of biofilms, they exhibit a variety of structures and compositions that vary according to particular environmental conditions. For example, individual matrix components are thought to have adhesive (Kaplan *et al*, 2004; Mack *et al*, 1994), structural (Berk *et al*, 2012; Danese *et al*, 2000; Hobley *et al*, 2013; Kaplan *et
al, 2004; Lawrence et al, 2007), protective (Jesaitis et al, 2003; Stewart & Costerton, 2001; Vuong et al, 2004), and/or signaling functions (Irie et al, 2012). The function and composition of the biofilm matrix, has been discussed extensively in a number of reviews (Branda et al, 2005; Flemming & Wingender, 2010).

Maturation

Although a biofilm begins as a thin layer of cells, it can eventually grow into a 3-D structure with a complex geography. This architecture is dictated by the arrangement of cells and matrix components, and can include channels and pores that are presumed to allow for the exchange of molecules such as water, nutrients, and oxygen (de Beer et al, 1994a; Guelon et al, 2012; Watnick & Kolter, 1999). Variegation within the structure of a biofilm causes the formation of microdomains, which contain different concentrations of environmental molecules (Damgaard et al, 2001; de Beer et al, 1994b; Huang et al, 1998; Wessel et al, 2014). Cells within each microdomain sense a unique repertoire of environmental ligands, thus causing populations of cells to exhibit distinctive gene-expression profiles (Rani et al, 2007; Sauer et al, 2002; Whiteley et al, 2001). This results in phenotypic heterogeneity, where different cells are though to perform different functions within the biofilm community determined, in part, by the particular environment they are experiencing (Lopez et al, 2009; Vlamakis et al, 2008). For example, subpopulations of cells can produce matrix components (Chai et al, 2008; Davies & Geesey, 1995; Lopez et al, 2009), become dormant persisters (Brooun et al, 2000; Rani et al, 2007), or become cannibals that consume nutrients from dead neighbors that they lysed with a toxin (Lopez et al, 2009). Phenotypic heterogeneity is a hallmark of biofilm development, and it is believed to allow for the formation of a community that
can withstand different stresses (Boles et al, 2004). One example of how phenotypic diversity benefits the community is the existence of the persister cell, a dormant cell that seems to withstand antimicrobial exposure, and can repopulate a community once the antimicrobial exits the environment.

**Dispersal**

Ultimately, a population of cells within the biofilm gain the ability to leave the biofilm, thus allowing them to reinitiate this developmental process in a new location, and bring the biofilm process full circle (Stoodley et al, 2001). It is believed that biofilm dispersal permits a bacterial cell to escape overcrowding or other poor environmental conditions in an aged biofilm and to search for a new, more acceptable location. Such environmental signals that induce dispersal include variations in nutrient concentrations (An et al, 2010; Delaquis et al, 1989; Gjermansen et al, 2005; James et al, 1995; Sauer et al, 2004; Sawyer & Hermanowicz, 2000), autoinducers or other bacterial-derived communication signals (Davies & Marques, 2009; Hammer & Bassler, 2003; Puskas et al, 1997; Rice et al, 2005), and other small molecules (Barraud et al, 2006; Barraud et al, 2009; Kolodkin-Gal et al, 2010). Dispersal signals are generally recognized by a subpopulation of cells (Purevdorj-Gage et al, 2005), where they induce genetic pathways that lead to the local destruction of the biofilm matrix. Matrix breakdown is mediated by secreted degradative enzymes (Allison et al, 1998; Boyd & Chakrabarty, 1994; Gjermansen et al, 2005; Johansson et al, 2008; Mann et al, 2009) surface-tension reducing surfactants (Boles et al, 2005), or cell-death from bacteriophage lysis (Garcia-Contreras et al, 2008; Rice et al, 2009). This “clearing” presumably permits a cell to
become motile and escape the biofilm, a process that likely allows a cell to escape a habitat that can no longer provide basic requirements that are required to survive.

**Conclusion**

Although bacteria were generally thought of as solitary, planktonic cells, this idea is now antiquated by the observation that they undergo complex developmental processes including biofilm formation. Cellular communities exist in all domains of life, a notion that is best represented by the rise of multi-cellular eukaryotic organisms. Perhaps cellular communities, rather than independently functioning cells, represent the standard lifestyle of organisms, with planktonic cells representing a newer development that allowed for the dissemination of an organism into new territories (Stoodley *et al*, 2002). The idea that communities represent the ancestral form rather than a nascent lifestyle changes our notion of how life as we know it evolved (discussed in (Stoodley *et al*, 2002; Vetsigian *et al*, 2006)).

**IV. Two-component signaling**

**Introduction**

For organisms to survive long enough to pass on their genes, they must have mechanisms to appropriately respond to the assorted environments they experience. To do this, they use signaling pathways that link environmental inputs with relevant intracellular outputs. One type of cellular circuitry found in most bacteria, some archaea, and a few eukaryotic species, is the two-component signaling (TCS) pathway (reviewed in (Stock *et al*, 2000; Wuichet *et al*, 2010)). The basic TCS architecture consists of two types of proteins: a sensor kinase (SK) and a response regulator (RR) (Fig 3). Typically, the SK senses an environmental signal, autophosphorylates on a conserved histidine, and
Figure 3. Canonical architecture of two-component signaling. (A, B, C) A sensor kinase (SK) often contains transmembrane regions that flank a periplasmic region. SKs contain a conserved HisKA (DHp) and HATPase_c (CA) domain within the cytoplasm. Response regulators (RR) are cytoplasmic and contain a conserved REC domain. Inactivated domains are in grey. Activated domains are in red. (A) An SK generally uses its periplasmic region to sense an environmental signal (signal: yellow sun). (B) Signal recognition activates the autokinase activity of the SK, wherein the HisKA and HATPase_c domains mediate autophosphorylation on a conserved histidine (His). This phosphate group is donated to a conserved aspartate (Asp) within the REC domain of a RR. (C) When phosphorylated, an RR generally elicits a response in a cell.
then donates this phosphoryl group to an aspartate on a downstream RR. Often the RR has an effector domain, such as a DNA binding or enzymatic domain, whose activity is activated or deactivated once it becomes phosphorylated (Galperin, 2010). This two-protein arrangement connected by a single his-asp phospho-transfer event remains the most common TCS architecture found in bacteria, although phosphorelays exist which consist of multiple phosphotransfer events (his-asp-his-asp) between two or more cognate TCS proteins. This section will cover basic TCS components with a focus on SKs, and then conclude with non-canonical arrangements and activities of atypical TCS pathways.

**Sensor kinase**

SKs sit at the top of the hierarchy in TCS pathways where they convert environmental information to an intracellular signal. On average, bacteria encode 25 to 30 unique TCS pathways (Whitworth, 2012) with SKs that presumably interact with different signals. The sensory domains within SKs exhibit little sequence similarities, which most likely reflects the diversity in ligand binding. Instead, SKs are identified based on the sequence conservation of autokinase domains which allow for an SK to function as both a kinase and a phosphatase. Although sequence and structural diversity exist even within the catalytic domains, and SKs can be grouped into different classes based on these similarities and differences (Grebe & Stock, 1999; Kim & Forst, 2001), the basic functions of a sensor kinase (sensing a signal, autophosphorylating, and serving as a phosphoryl donor to an RR) remain constant (Fig 4). The structure and function of domains that elicit these three activities will be discussed below.

**Sensing a signal.** SKs in the cell continually sample the environment in search for their cognate signal. Upon the discovery of an SK of interest, researchers, too, search for
this signal; however, the ligands that bind SKs remain the most elusive aspect of TCS. Although signals for some TCS pathways have been determined (e.g. (Bader et al, 2005; Freeman et al, 2000; Garcia Vescovi et al, 1996; Gilles-Gonzalez et al, 1991; Jourlin et al, 1996; Kaspar et al, 1999; Zientz et al, 1998)) most remain enigmatic. Furthermore, although some sensory domains exhibit similar tertiary structures, this observation provides little to no insight into what signal an SK may recognize. Some information can be gained from the location of the ligand-binding site. In principle, signals can be extracellular, intracellular, or membrane-associated. There are examples of SKs that have sensory domains, or associate with accessory signal binding proteins, within at least one of these locations (see below). However, regardless of the nature or location of the signal, the purpose of the ligand-SK interaction is to change the activity of the SK’s catalytic domains. Although, every ligand-SK interaction may have a unique, nuanced mechanism to control SK activity, recent structural analysis of sensory domain structures have elucidated broad methods by which signal binding could be translated into sensor kinase activation or deactivation. These structural analyses will be discussed below.

**Extracytoplasmic sensing domains.** ~90% of all SKs are membrane bound (Cock & Whitworth, 2007). Often, two transmembrane (TM) regions flank a protein segment that projects into the periplasm. In these structures, and in the absence of other known sensory domains, the periplasmic region is presumed to bind extracellular signals; this property has been demonstrated for a handful of SKs (Bader et al, 2005; Freeman et al, 2000; Garcia Vescovi et al, 1996; Gilles-Gonzalez et al, 1991; Jourlin et al, 1996; Kaspar et al, 1999; Zientz et al, 1998). Although the ligands for these domains remain mostly unknown, a growing number of studies have identified common structures within this
region. The best-studied periplasmic sensing domain is the PDC (PhoQ, DcuS, CitA) domain, which is named after the SKs in which it was first identified (Cheung et al., 2008; Cheung & Hendrickson, 2008; Sevvana et al., 2008) (this classification includes the previously described “cache” domain (Anantharaman & Aravind, 2000)). PDC structures contain a central five- to six-stranded \( \beta \)-sheet flanked by one long and one short \( \alpha \)-helices, which is very similar to a PAS fold (see “cytosolic signaling domains”) (Fig 5), although the debate continues about whether it is a bona fide PAS fold or a convergently-evolved motif (Cheung & Hendrickson, 2010; Moglich et al., 2009). Zhang et al. (Zhang & Hendrickson, 2010) performed structural analyses of five “double” PDC domains (two covalently attached PDC domains) from SKs isolated from an archaea (two from \textit{Methanosarcina mazei}) and different species of bacteria (\textit{Shewanella oneidenis}, \textit{Vibrio parahaemolyticus}, and \textit{Rhodopseudomonas palustris}) (Fig 5A, B). Although the signals were not known for these SKs, four out of five were bound to an adventitious ligand. The researchers of this study identified two mechanisms by which a signal-PDC complex could affect the downstream SK domain. They proposed that, after signal binding, a C-terminal \( \alpha \)-helix within a PDC monomer could undergo downward piston-like movements, sweeping rotational movements, or both types of movements, and any of these mechanisms could be sufficient to change the conformation of the C-terminal histidine kinase regions (Fig 5C). Zhang et al. further conclude that these mechanisms could be extrapolated to non-PDC or single PDC domains, especially considering similar conformational movements have been seen in other membrane-bound...
**Figure 5. PDC structure and function.** IM: inner-membrane (A) Ribbon diagram of putative double periplasmic PDC domain from a diguanylate cyclase, VF_0989, from *V. fischeri* (Altschul *et al*, 2005; Kelley & Sternberg, 2009; Ruby *et al*, 2005; Zhang & Hendrickson, 2010). (B) Topology cartoon of double PDC domain. (C) Two predicted mechanisms by which signal binding (yellow sun) can induce a conformational change in the double PDC domain. Purple: double PDC domain not bound to a signal. Orange: PDC domain bound to a signal. Grey arrows indicate the conformational change induced after ligand binding (Adapted from (Zhang & Hendrickson, 2010)).

Other than the PDC structure, few other periplasmic sensory regions within SKs exhibit conserved structural domains. Some domains consist entirely of alpha helices (Cheung & Hendrickson, 2009; Moore & Hendrickson, 2009), which, based on experiments performed with the aspartate receptor Tar (Yeh et al, 1996) and the SK NarX (Cheung & Hendrickson, 2009), could undergo piston-like movements that push into the membrane after binding a signal. Other extracytosolic sensory domains have structures akin to periplasmic binding proteins (Cheung et al, 2009), and could function similarly to LuxP/LuxQ, a periplasmic binding protein/SK complex, described below (Neiditch et al, 2006). However, an exact mechanism of how signal interaction could be propagated to an SK’s catalytic domain for most periplasmic sensory regions remains elusive. This is most likely due to the difficulty in identifying signals for an SK and in purifying full-length, membrane-bound SKs for crystallographic studies.

**Inner-membrane sensing domains.** Most SKs are associated with the membrane via 1 to 20 membrane-spanning regions (Mascher et al, 2006) (Fig 4). Transmembrane (TM) regions in SKs were originally assumed to function only as anchors for membrane-localization and for linking structural changes induced by signal binding in the signal domain to the histidine kinase domain (Cheung & Hendrickson, 2009; Goldberg et al, 2010; Hughson & Hazelbauer, 1996; Lowe et al, 2012). However, it has become appreciated that the membrane region can be a bone fide signal-sensing region. For example, the SK DesK found in *B. subtilis* utilizes its five TM regions as a thermometer (Aguilar et al, 2001; Albanesi et al, 2004; Albanesi et al, 2009). Cell membranes exhibit
a thicker, gel-like form at low temperature versus the thinner and more fluid-like form at high temperature. To sense temperature changes, the TM regions of DesK undergo conformational changes according the fluidity of the inner membrane. At low temperatures, the conformational changes in the TM regions induce a rotational shift in the structure of DesK’s catalytic domain thus promoting its kinase function (Aguilar et al., 2001; Albanesi et al., 2004; Albanesi et al., 2009). When this occurs, DesK induces expression of specific cold-shock proteins through activation of its RR, DesR. These cold-shock proteins desaturate membrane-embedded fatty acid chains to restore membrane fluidity. Higher temperatures or restoration of membrane fluidity changes the structure of DesK’s TM regions such that the phosphatase activity of DesK is favored, and the expression of the DesK-DesR regulon is shut off.

Besides DesK, there are but a few examples of SKs that utilize their TM domain(s) to directly (Voet-van-Vormizeele & Groth, 2008) or indirectly (Bogel et al., 2009) sense signals. There are also instances of histidine kinases with TM domains and no other obvious sensory regions, but whether they utilize their TM for sensory purposes is unknown (Mascher et al., 2006). It remains to be determined whether TM-sensing domains are truly rare, or whether they are simply underappreciated or undiscovered.

**HAMP domains.** Often, the sensory domains of SKs are not directly linked to the kinase domains. Instead, many SKs have a HAMP (Histidine kinases, adenylyl cyclases, methyl-accepting chemotaxis proteins, and some phosphatases) domain between these two regions. HAMP domains are not unique to SKs, but are found in a variety of signaling proteins in all domains of life (Aravind & Ponting, 1999), where they mediate signaling between an N-terminal input domain and a C-terminal output domain (reviewed
in (Parkinson, 2010)). It has been proposed that this domain could have many functions, such as amplifying the structural changes induced by signal binding, or fine-tuning the net kinase to phosphatase activity of the SK (Parkinson, 2010).

Although they exhibit little sequence similarity, HAMP domains are categorized based on a common secondary structure. Specifically, they consist of two amphipathic helices with coiled-coil properties linked by an unstructured segment (Butler & Falke, 1998; Le Moual & Koshland, 1996; Singh et al, 1998) (Fig 6A). Bending of the unstructured linker at a conserved glycine residue allows the two alpha helices to interact (Hulko et al, 2006). HAMP domains can form homodimers, in which the helices of each protomer pack together to form a four-helix bundle (Hulko et al, 2006; Singh et al, 1998) (Fig 6B). There are multiple hypotheses that attempt to explain how the conformation of the HAMP domain controls C-terminal activities of a protein. The original hypothesis proposed a two-phase state, where domains switch between “on” or “off” conformations that involve either axillary rotation of each helix within the four helix bundle (Hulko et al, 2006), or a switch between a relaxed two-helix bundle and a compact four helix bundle (Airola et al, 2010). A later hypothesis, known as the biphasic dynamic bundle hypothesis, suggested that the conformation of the HAMP domain exhibits a wide-range of four-helix packing stabilities, where “too tight” or “too loose” conformations represent an “off” state, whereas the intermediate “sweet spot” generates the “on” state (Ames et al, 2014; Manson, 2009; Stewart, 2014; Zhou et al, 2009). Regardless, all hypotheses support a model in which signal binding leads to a structural change in the HAMP domain that is then propagated to the downstream portion of the protein, such as the catalytic region of an SK (Tao et al, 2002). Whether the hypotheses to explain HAMP
Figure 6. HAMP domain arrangement (A) Ribbon structure of a HAMP monomer from the SK, SypF, from *V. fischeri* (Altschul *et al.*, 2005; Darnell *et al.*, 2008; Kelley & Sternberg, 2009). The HAMP structure is rainbow-colored, where red represents the N-terminus, and blue the C-terminus. (B) Helical packing arrangement of the HAMP dimer as a four-helix bundle. $a$, $d$, and $x$ indicate key packing residues. Dotted lines represent intersubunit interactions, and solid lines indicate intrasubunit interactions. Upper packing layer: black; lower packing layer: gray. This figure was adapted from (Parkinson, 2010).
function are mutually exclusive or not, or whether HAMP domains may function
dissimilarly in different proteins, remains to be determined.

**Cytoplasmic sensing domains.** Some SKs have neither a periplasmic nor a TM
localized signaling region; instead, these proteins contain cytoplasmic domains that
function as the site of signal recognition. In general, cytoplasmic sensory domains in SKs
can be identified by the presence of one of two structures: PAS or GAF folds (Fig 4 and
Fig 7). Research over the past 20 years has detailed the mechanisms by which these two
large classes of protein domains can regulate signaling pathways in all domains of life.

The PAS (Per-ARNT-Sim) domain was originally named based on its discovery
in the *Drosophila* circadian rhythm factor (PERIOD), the aryl hydrocarbon nuclear
translocator (ARNT) in mammals, and the single-minded (SIM) protein that controls
development in *Drosophila* (Henry & Crosson, 2011; Huang *et al.*, 1993). Although it
was first described in eukaryotes, the PAS domain is also found in archea and bacteria
(Finn *et al.*, 2010) and can control a wide variety of signaling processes (reviewed in
(Taylor & Zhulin, 1999)). Generally, PAS domains within different signaling pathways
can either directly sense a signal (Henry & Crosson, 2011), change protein-protein
interactions (Lee *et al.*, 2008; Neiditch *et al.*, 2006), promote signal propagation (Oka *et
al.*, 2008), or affect protein localization (Chen *et al.*, 2005). All of these mechanisms are
thought to induce a conformational change in the PAS domain that affects the activity of
the covalently attached effector domain, such as a histidine kinase domain. PAS domains
do not exhibit substantial sequence homology, although domains that recognize similar
ligands or cofactors group together in structural comparisons (Henry & Crosson, 2011).
All PAS domains exhibit a characteristic PAS fold that consists of a central, five-stranded
antiparallel β sheet flanked by a variable number of α-helices: signal or cofactor binding
generally occurs in a pocket in the central β-fold (Moglich et al, 2009) (Fig 7A). In
bacteria, half of these domains are associated with an SK (Henry & Crosson, 2011). It has
been proposed that slight differences in the structure of the PAS domain allow for
binding of unique signals (An et al, 2014; Malpica et al, 2004) or cofactors that interact
with signals (Fu et al, 1994; Gong et al, 1998) and can also promote kinase dimerization
(Lee et al, 2008) to control SK function and therefore control the induction of TCS
cascades.

Like the PAS domain, the GAF (cGMP-specific phosphodiesterase, adenyly
cyclase, and FhlA) domain is found in all domains of life, but has been best characterized
in plants and cyanobacteria (Aravind & Ponting, 1997; Karniol et al, 2005). In bacteria,
GAFs are cytosolic and generally found attached to a variable output domain, such as a
histidine kinase domain. Its structure is similar to PAS, with a central five to six-strand
anti-parallel β-sheet surrounded by four or more α-helices, and a pocket within the core
where a signal or cofactor interacts (Ho et al, 2000; Podust et al, 2008) (Fig 7B). In
bacteria, GAFs are implicated regulating many different cellular physiologies including
the response to hypoxia (Ohno et al, 2003; Park et al, 2003), the regulation of toxin-
antitoxin systems (Florek et al, 2011), and sensing light (Lamparter et al, 2002; Narikawa
et al, 2008; Yeh et al, 1997). Their main function is to interact with an environmental
signal, with or without the help of cofactors, which stimulates a change in the
conformation of GAF that can be propagated to an attached output domain (Batchelor et
al, 2013; Heyes et al, 2012; Ho et al, 2000). Additionally, some GAFs are implicated in
mediating protein-protein interactions, although it is not always known whether this is
Figure 7. Structures of cytoplasmic sensing domains: PAS and GAF. (A, B)

Structures are rainbow-colored, where red represents the N-terminus, and blue the C-terminus. (A) Ribbon structure of the cytoplasmic PAS domain from the SK, RscS, from *V. fischeri* (Altschul *et al.*, 2005; Geszvain & Visick, 2008a; Kelley & Sternberg, 2009). (B) Ribbon diagram of the GAF domain from the hypothetical protein, VF_1278, from *V. fischeri* (Altschul *et al.*, 2005; Kelley & Sternberg, 2009; Ruby *et al.*, 2005).
signal dependent or independent (Gao et al, 2008; Martinez et al, 2002).

One well-studied SK with an attached GAF domain is the phytochrome, Cph1, from the cyanobacterium, *Synechococcus* sp. OS-B’. Phytochromes function as light receptors in both plants and microorganisms, and the canonical phytochrome structure consists of a PAS-GAF-PHY (phytochrome) complex attached to an output domain (Rockwell et al, 2006). In Cph1, the regulatory domain is a histidine kinase module. Light detection occurs through a bilin (open-chain tetrapyrrrole) chromophore covalently bound to the GAF domain. Bilin absorbs red and far-red wavelengths of light, which leads to an isomerization of specific carbon-carbon bonds within its molecular structure. This isomerization is sufficient to alter the intra- and inter-domain interactions of the PAS-GAF-PHY regions within Cph1, causing a conformation shift that is large enough to affect the enzymatic activity of the attached SK domain. Although much is known about the structure of Cph1, its exact physiological role in the cell remains unknown.

**Accessory sensing proteins.** Most SKs contain their own sensory region; however, some SKs utilize a separate protein to recognize a signal and induce the catalytic activity of an SK. The well-characterized example is the periplasmic binding protein, LuxP, which binds to the SK LuxQ, a quorum-sensing regulator (Bassler et al, 1994a; Chen et al, 2002) (Fig 4 and Fig 10). At low cell density, the quorum signal (AI-2) is low, and the LuxP-LuxQ complex functions as a kinase to inhibit quorum-regulated genes. At high cell density, LuxP interacts with AI-2 (Chen et al, 2002), causing an asymmetrical shift to occur in the associated LuxQ dimer (Neiditch et al, 2006). This disrupts the kinase activity of LuxQ, thus allowing for the expression of quorum-
regulated genes. Interesting, LuxQ contains a periplasmic loop with a PAS domain, but it utilizes this domain to interact with LuxP instead of with a signal (Neiditch et al., 2005).

Bacterial MCPs (methyl-accepting chemotaxis proteins) represent a large, well-studied class of accessory proteins. They are membrane-bound chemical sensing proteins involved in chemotaxis (reviewed in (Szurmant & Ordal, 2004)). MCPs control the activity of CheA, the SK that promotes phosphorylation of the RR, CheY, which, in turn, regulates whether a cell “runs” in a particular direction or tumbles by binding to the flagellar motor and changing its direction of rotation (Borkovich et al., 1989; Hess et al., 1988; Welch et al., 1993; Wolfe et al., 1987). Different MCPs in a cell are responsible for recognizing different chemotactic signals. Signal binding induces a conformation change in the sensory domain of the MCP, which is often propagated through one or more HAMP domains and then to CheA. This leads to the activation or deactivation of the kinase activity of CheA (Khursigara et al., 2008). One of the best-studied MBPs, Tar, has a periplasmic four-helix bundle that binds to an attractant, aspartate, in the environment (Clarke & Koshland, 1979; Yeh et al., 1996). Interaction with aspartate causes a piston-like shift in the helical region (Adase et al., 2013; Yeh et al., 1996). This structural change is presumably propagated through the C-terminal region of Tar, though a connector protein, CheW, and then to CheA. This event deactivates the CheA kinase, which subsequently leads to cells migrating up an aspartate gradient. Although Tar is an MCP, the structural mechanism by which it controls the function of CheA may be analogous to similar sensory structures in found SKs (Cheung & Hendrickson, 2009; Moore & Hendrickson, 2009).
Kinase activity. The purpose of all sensory domains within or associated with an SK is to control the most recognizable and arguably important feature of an SK: its enzymatic domains involved in kinase and phosphatase activity. An SK functions as a dimer and requires two regions on each protomer to autophosphorylate: a catalytic and ATP-binding (CA) domain, and the dimerization-and histidine transfer domain (DHp) (Fig 8). There are two enzymatic steps involved in this activity. First, the CA domain cleaves a bound ATP to generate ADP + Pi, and, secondly, the CA and DHp form an active site that promotes a covalent bond between the cleaved phosphate with the histidine residue in the DHp domain. These processes are induced upon signal binding, which promotes a conformational change in the SK such that the CA and DHp domains are aligned to favor the inter-domain interaction (Bilwes et al, 1999; Marina et al, 2005; Mechaly et al, 2014; Neiditch et al, 2006). It is believed that autophosphorylation occurs in trans (Ellefson et al, 1997; Ninfa et al, 1993; Yang & Inouye, 1991), although there are examples of SKs autophosphorylating via a cis mechanism (Casino et al, 2009; Pena-Sandoval & Georgellis, 2010). The CA and DHp domains, or the core structure, remain the best-characterized feature of an SK.

CA domain (Pfam: HATPase_c). The CA domain contains the enzymatic residues important for cleaving ATP. It is part of the GHKL (Gyrase B, Hsp90, Histidine Kinase, and MutL) ATPase family (Marina et al, 2001). These diverse proteins share a Bergerat fold (Bergerat et al, 1997) that consists of a two-layered α-β sandwich with a loop that appears as a “lid” over the ATP (Bilwes et al, 1999; Marina et al, 2001; Tanaka et al, 1998) (Fig 8A). In SKs there are four motifs that are important for the CA to function and whose names are derived from conserved amino acids located within these
Figure 8. Structures of the core enzymatic regions of an SK. (A, B) Structures are rainbow-colored, where red represents the N-terminus, and blue the C-terminus. (A) Ribbon structure of a monomer of the CA domain of SypF. (B) Ribbon structure of the monomeric form of the Dhp domain of SypF (Altschul et al, 2005; Darnell et al, 2008; Kelley & Sternberg, 2009).
motifs (Parkinson & Kofoid, 1992; Stock et al, 1988): the D-box (or G1-box) and N-box form hydrogen bonds with the ATP, the F-box forms part of the ATP lid, and G2 box represents the hinge of the lid (Bilwes et al, 1999; Hirschman et al, 2001; Marina et al, 2001; Tanaka et al, 1998). The catalytic activity of most SKs requires a metal ion, usually a magnesium ion, whose binding within the CA core is partially facilitated by interactions with the N and G box (Marina et al, 2001); however, some SKs require other metal cofactors such as calcium or manganese (Gamble et al, 1998; Haydel et al, 1999; Martinez-Wilson et al, 2008) to function.

**DHp:** The DHp domain serves two functions: it contains the conserved H box with the histidine that serves as the site of autophosphorylation (Hess et al, 1988) (Ninfa & Bennett, 1991; Roberts et al, 1994), and it is the site that mediates SK homodimerization (Ashenberg et al, 2011; Park et al, 1998). The DHp region from EnZ was the first to be resolved (Tomomori et al, 1999). This structure demonstrated that the DHp protomer consists of two $\alpha$-helices connected by a loop (a “hairpin” structure) (Fig 8B), and that the “hairpin” from the two protomers interact to form a four-helix bundle with the loop region located towards the bottom of the structure. The conserved histidine is located in the middle of a helix from each protomer, and it is outwardly exposed to allow for an interaction with the CA domain (Marina et al, 2005; Tomomori et al, 1999). More recent work has described a possible third function for the DHp domain: determining whether an SK autophosphorylates in cis or in trans via the structure of the loop (Ashenberg et al, 2013). The authors found that switching this loop between trans- or cis-acting SKs swapped this mechanism of phosphorylation. They reasoned that cis or trans phosphorylation could be predicted based on this structure, as the cis SKs seem to
have a left-handed loop, whereas trans SKs have a right-handed loop. The authors noted that, in opposition to their hypothesis, that the SK DesK from Bacillus subtilis has a left-handed loop; yet, it seems to autophosphorylate in trans (Trajtenberg et al., 2010). Thus, further work will determine whether this is a conserved function of the loop in most SKs, and what the exact structural change the alteration of this loop does to the SK to make it prefer either cis or trans activities.

**Phosphatase activity.** The function of TCS is to activate or deactivate particular intracellular pathways in response to the cell’s surroundings. However, environments fluctuate, and the TCS systems that govern the response to one environment may not be required to adapt to new surroundings. Although the kinase activity of an SK can be passively turned off based on fluctuations in the proximal concentration of a ligand, a high ratio of phosphorylated RR to unphosphorylated RR could still remain in the cell, interfering with adaptations to a new environment. Furthermore, RRs autophosphorylate using small molecules such as acetyl phosphate (Klein et al., 2007; Lukat et al., 1992) and non-cognate SKs (Haldimann et al., 1997; Verhamme et al., 2002; Zhao et al., 2002), causing them to be active even when no signal is received from their cognate SK. However, regulatory mechanisms have evolved within TCS systems to actively manage the phosphorylation state of an RR and to reset the system to pre-stimulus levels. For example, most TCS systems contain an SK that functions as both a kinase and a phosphatase (Igo et al., 1989; Keener & Kustu, 1988; Ninfa & Magasanik, 1986), and mutations of phosphatase regions within SKs can lead to phenotypic alterations (Chamnongpol et al., 2003; Depardieu et al., 2003; Kostakioti et al., 2009; Pioszak & Ninfa, 2003) presumably due to the inability to turn a particular regulatory pathway on or
off. The phosphatase activity of an SK generally exists in equilibrium with the kinase activity, where signal recognition can either push the equilibrium towards kinase (e.g. NarX (Williams & Stewart, 1997)) or phosphatase activity (e.g. PhoQ(Chamnongpol et al, 2003)) presumably through inducing structural changes that favors one catalytic activity over the other (Marina et al, 2005; Mechaly et al, 2014; Stewart & Chen, 2010).

There are two biochemical mechanisms by which a SK can remove phosphoryl groups from a RR: 1. using residues similar to those used for the kinase reaction (Igo et al, 1989; Keener & Kustu, 1988; Ninfa & Magasanik, 1986) or 2. using a REC domain that is often covalently attached to an SK, also known as a hybrid SK (described below) (Freeman et al, 2000; Georgellis et al, 1998; Pena-Sandoval & Georgellis, 2010; Uhl & Miller, 1996). The latter, known as signal decay, occurs when the phosphatase activity of a REC domain in a hybrid SK drives the reverse phosphotransfer in the TCS system; the N-terminal CA/DHp domains are not required for this mechanism either in vitro (Ault et al, 2002; Uhl & Miller, 1996) or in vivo (Freeman et al, 2000; Pena-Sandoval et al, 2005). The former mechanism utilizes amino acids that include or surround residues that are important for the kinase reaction. However, phosphatase activity is not simply the reverse biochemistry of autophosphorylation, as this reaction does not necessarily require the conserved site of autophosphorylation (Atkinson & Ninfa, 1993; Hsing & Silhavy, 1997). Evidence suggests that the DHp domain alone mediates phosphatase activity, as this domain is sufficient to remove phosphates from a cognate RR (Carmany et al, 2003; Jiang et al, 2000a; Zhu et al, 2000); however, the presence of a CA domain can enhance catalysis (Hsing & Silhavy, 1997; Jiang et al, 2000b). In an effort to identify the key residues that mediate phosphatase activity in an SK, Hunyh et al observed that the DHp
domain of some SKs contain a DxxxQ or an ExxN motif (Huynh et al., 2010), which are also found in other bacterial phosphatases such as CheZ (Zhao et al., 2002) and enzymes within the CheC family (Park et al., 2004; Pazy et al., 2010; Sircar et al., 2013); however, there has been conflicting evidence on whether these residues are a common requirement for all SK that exhibit phosphatase activity (Atkinson & Ninfa, 1993; Sheeler et al., 2005; Willett & Kirby, 2012).

**Phosphorelays**

The most common TCS pathways consist of an SK and RR linked by a single His-Asp phosphotransfer event; however, multi-step His-Asp-His-Asp phosphorelays exist. Phosphorelays often contain a hybrid SK that has one or two additional sites of phosphorylation within a receiver (REC) or histidine phosphotransferase (Hpt) domain, respectively. Alternatively, rather than be covalently attached to a hybrid SK, additional REC or HPt domains can exist as single-domain proteins that functions between the upstream SK and the downstream RR. It is believed that these extra phosphotransfer events represent checkpoints that control whether a cell initiates physiological changes under particular conditions (Appleby et al., 1996).

**Hybrid sensor kinase.** 25% of all SKs are hybrid SKs, which are identified based on the presence of a C-terminal REC domain covalently attached to the upstream CA/DHp regions (Zhang & Shi, 2005). The REC domain, which is identical to the REC domains found in RRs (see RR section), contains a second site of phosphorylation (Asp). Some hybrid SKs also possess a third site of phosphorylation, a histidine within a C-terminal HPt domain. In hybrid SKs, phosphotransfer generally occurs in a step-wise, consecutive manner, wherein phosphorylation of the downstream domains cannot occur
until the most N-terminal domain has become phosphorylated. This forward phosphotransfer is initiated via the autokinase reaction within the CA/DHp regions. In contrast, the phosphatase activity of hybrid SKs resides in the REC domain (see above “phosphatase” section). To date, most hybrid SKs with autokinase and phosphatase activity require these additional sites of phosphotransfer to effectively control the phosphorylation state of the downstream cognate RR (Hsu et al, 2008; Jourlin et al, 1997; Takeda et al, 2001; Tszuki et al, 1995; Uhl & Miller, 1996). Although both canonical and hybrid SKs are thought to function as “on/off” switches, there is evidence that hybrid SKs can exhibit rheostat-like control over their downstream signaling pathway (Cotter & Jones, 2003). This is best exemplified by the work done with the hybrid SK, BvgS, in Bordetella pertussis, in which BvgS induces distinct phenotypes (“low,” “intermediate,” and “high”) based on the increasing concentration of signals that bind BvgS (Deora et al, 2001; Prugnola et al, 1995; Scarlato et al, 1991). BvgS controls the phosphorylation state of BvgA, an RR that recognizes a particular DNA element to control the transcription of its regulon (Karimova et al, 1996; Roy & Falkow, 1991; Uhl & Miller, 1994; Zü et al, 1996). Each gene set that controls the high, intermediate, or low phenotype is within the BvgA regulon, but each of these gene sets have characteristic promoter binding affinities for phospho-BvgA (Karimova et al, 1996; Roy & Falkow, 1991; Zü et al, 1996). The range in binding affinities results in the induction of high, intermediate, or low phenotypes based on the levels of phospho-BvgA. Because BvgS controls the phosphorylation state of BvgA, different concentrations of signal recognition by BvgS controls the expression of “low, intermediate, high” genes (Deora et al, 2001). Furthermore, the multiple sites of phosphorylation within BvgS represent additional
checkpoints that control whether BvgA becomes phosphorylated. The BvgS rheostat mechanism of controlling gene expression may be a way for hybrid SKs to control a range of different phenotypes; however, the use of this mechanism in other SKs has not been defined.

**Histidine-containing phosphotransfer protein (Hpt).** All phosphorelays contain an Hpt domain that exists as either a single domain cytoplasmic protein or is attached to the C-terminal end of a hybrid SK. Although little sequence similarity exists among Hpt proteins, they fall into one of two structural categories: a monomeric four-helix bundle (Ikegami et al., 2001; Rogov et al., 2004; Xu & West, 1999) or a dimeric four-helix bundle (Fioravanti et al., 2012; Zhou et al., 1997). Interestingly, the dimeric form resembles the CA/DHp core found in SKs, but with a degenerate CA domain. Neither class of Hpt proteins contains enzymatic activity. The phosphorylatable histidine in both of these structures is generally located in a central portion of an α-helix with the side chain pointed outward. The position of the histidine side chain is believed to allow for phosphotransfer between an upstream hybrid SK and a downstream RR. Although the core four-helix bundle is common to all monomeric Hpt domains, most have additional structured or unstructured regions that surround the core. Furthermore, the electrostatic characteristics of amino acids surrounding the conserved histidine vary among Hpt proteins (Rogov et al., 2004). These distinguishing characteristics are believed to maintain interaction fidelity with cognate TCS regulators.

**Response regulators**

Although SKs are responsible for sensing an environmental signal, it is the cognate RR that elicits a response in a cell. Although some REC domains exist as single
cytoplasmic proteins, most RRs often have two or more domains: the conserved REC domain that catalyzes phosphotransfer from the upstream SK, and a variable effector domain whose activity is controlled via the phosphorylation state of the REC region. The most common effector domains are those with DNA binding ability, which allows an RR to function as a transcription factor. Other effector domains contain one of an assortment of other functions that are perfectly tailored for the needs of a particular TCS pathway.

**REC domain (Pfam: Response_reg).** RRs are identified based on the presence of a conserved REC domain, which is responsible for catalyzing phosphotransfer from the histidine in the SK to the aspartate in the REC domain, and for the reverse autodephosphorylation reaction. The basic structure of the REC domain consists of alternating β-strands and α-helices that fold into a five-stranded β-sheet surrounded by five α-helices (Baikalov et al, 1996; Birck et al, 1999; Muchova et al, 1999; Muller-Dieckmann et al, 1999; Sola et al, 1999; Stock et al, 1989; Volkman et al, 1995) (Fig 9A). The C-terminal end of the REC domain contains the active site with at least five conserved residues: three acidic amino acids (usually aspartates), a threonine or serine, and a lysine (Appleby & Bourret, 1998; Lukat et al, 1991). The acidic residues interact with a critical metal cofactor (usually magnesium), and one of these residues is a highly conserved aspartate that serves as the site of phosphorylation (Sanders et al, 1989; Stock et al, 1993). The oxygen molecules found within the incoming phosphate group form hydrogen bonds with the conserved threonine or serine side-chains and a salt bridge with the lysine side-chain, and also interact with the metal cofactor. These events coordinate the nucleophilic attack of the conserved aspartate with the phosphate atom, and it results in structural repositioning of key residues including the conserved lysine,
Figure 9. Response regulator domains. (A) Ribbon structure of the REC domain of SypG from *V. fischeri* (Altschul *et al*, 2005; Kelley & Sternberg, 2009). The REC structure is rainbow-colored, where red represents the N-terminus, and blue the C-terminus. (B) Chart representing the percentage of effector domain classes found associated with a REC domain (adapted from (Galperin, 2010)).
serine/threonine, and a less conserved phenylalanine/tyrosine residue. These residues are relocated to their original locations during the reverse autodephosphorylation reaction, in which a water molecule instigates the nucleophilic attack with the aspartate-bound phosphate molecule, presumably releasing $\text{HPO}_4^{\text{2-}}$ into the cellular milieu. The position of these key amino acids generally represents the “active” or “inactive” state of the RR; however, REC domains do not statically exist in an “on” or “off” structure. Instead, they fluctuate between active, inactivated, and intermediate states, with phosphorylation pushing the equilibrium towards activated (or inactivated) structures (McDonald et al., 2012; Volkman et al., 2001). These conformational changes can then be propagated to the effector domain (if present) to elicit a response. In addition to the observation that phosphorylation induces intramolecular conformational changes, phosphorylation can also control whether RRs form dimers or higher order oligomers (Bachhawat et al., 2005; Barbieri et al., 2013; Batchelor et al., 2009; Birck et al., 2003; Friedland et al., 2007; Paul et al., 2007; Toro-Roman et al., 2005), an event that is often required for an RRs to control downstream events.

For phosphotransfer to occur between an SK and an RR, residues from both regulators must form the active site; however, catalysis is mostly driven by residues within the REC domain. This enzymatic activity exhibited by the REC domain also allows an RR to autophosphorylate using high-energy, small molecule phosphodonors, such as acetyl-phosphate, phosphoramidate, carbamoyl phosphate, and phosphoimidazole (Lukat et al., 1992), although the rate of catalysis is generally slower than with an SK (Da Re et al., 1999). Physiological relevance for phosphorylation of an RR using acetyl-phosphate in vivo has been established (reviewed in (Wolfe, 2010)). Importantly,
phosphorylation of a REC domain *in vitro* has allowed researchers to study both the phosphorylated and unphosphorylated forms of a REC domain, and has permitted in-depth, mechanistic studies on both the biochemical and structural features important for the function of these critical regulatory proteins.

**Effector domain.** The function of the REC domain is to control the activity of an attached effector domain (when present). ~63% of all RRs with DNA-binding structures (Gao & Stock, 2009) (Fig 9B) fall into the OmpR (Martinez-Hackert & Stock, 1997), NarL (Baikalov et al, 1996), LyTR (Sidote et al, 2008), or NtrC families (Batchelor et al, 2008). DNA-binding RRs function as transcription factors, so that signal recognition by a SK can lead to a direct change in the expression of relevant genes. Other common effector domains include RNA-binding regions that exhibit antiterminator activity (Fox et al, 2009; O'Hara et al, 1999; Wilson et al, 1996), and enzymatic domains, such as GGDEF (Paul et al, 2004) or EAL domains (Slater et al, 2000) involved in the production or degradation of the second messenger c-di-GMP, respectively, adenylate cyclases and cAMP phosphodiesterases (Shaulsky et al, 1998), serine phosphatases (Bhuwan et al, 2012; Morris et al, 2011; van Schaik et al, 2005), serine kinases (Morris et al, 2011), and methylesterases (Hayashi et al, 1979) (Fig 9B). Additionally, some effector domains can promote protein-protein interactions (Francez-Charlot et al, 2009) (Fig 9B). There seems to be little to no limits placed on the class of cytosolic effector proteins that can be incorporated into an RR structure.

Not all RRs contain an effector domain (Fig 9B). ~2% of REC domains function as signal transduction proteins, often in the context of a hybrid SK (see “hybrid SK” section, above), that promote forward phosphotransfer to a downstream HPt protein;
however, there is one example of a hybrid SK in *Myxococcus xanthus*, FrzE, whose REC domain inhibits phosphotransfer to a downstream RR, FrzZ, to control motility (Inclan et al, 2008). Additionally ~15% are single domain, cytoplasmic proteins (Galperin, 2010). This latter class of RRs often mediates protein-protein interactions, such as the prototypical chemotaxis regulator CheY, which, when phosphorylated, undergoes a conformational change that allows it to directly interact with the flagellar apparatus to control the direction of flagellar rotation (Dyer & Dahlquist, 2006; Parkinson et al, 1983). Additional functions of single REC domains include controlling SK activity (Paul et al, 2008) and protein localization (Iniesta et al, 2006). It is unknown whether there is an advantage to separating the REC domain and its target effector into separate proteins. Perhaps single REC domains are better able to interact with multiple output proteins, or perhaps they represents a historical relic before the REC and effector domains merged into one protein.

**Atypical TCS systems**

Although TCS systems with a single SK-RR pair connected by one phosphotransfer event remain the most common architecture, complex variations exist. Presumably, these changes have endowed a particular pathway with exquisite, multifaceted control over a response to an environmental signal. Below, I will discuss three non-canonical TCS pathways: Lux, which controls luminescence in *Vibrio* species; Gac-Ret, involved in *P. aeruginosa* virulence; and Rcs, a conserved pathway in *Enterobacteriacea* that regulates a variety of physiological processes in a considerable number of bacterial species.
Figure 10: Lux TCS pathway controlling bioluminescence. At low cell density, the SKs LuxN, LuxP/Q, and CpsS exhibit net kinase activity leading to the phosphorylation of LuxU and subsequent phosphotransfer to LuxO. Phospho-LuxO inhibits bioluminescence. At high cell density, three distinct autoinducer (AI) molecules (indicated by orange squares, yellow triangles, or red circles) are produced by the cell or from neighboring organisms. AI molecules are at sufficient concentrations to switch the activity of the SKs from net kinase to net phosphatase. This leads to dephosphorylation of the downstream RR, LuxU and LuxO, wherein unphosphorylated LuxO no longer inhibits bioluminescence. Phosphatase activity is mediated by the REC domain of the SKs, which ultimately release inorganic phosphate into the cytoplasm.
**Lux.** Bacterial luminescence, or light production, is an energetically exhausting process (Bourgois *et al.*, 2001) controlled by a complex TCS pathway known as Lux, which has been described extensively in *Vibrio* species (for a review, see (Ng & Bassler, 2009)) (Fig 10). Lux is a quorum-sensing pathway, meaning its activity is mediated by the concentration of bacterial signals in the local extracellular milieu: these signals can be sensed via an autocrine or paracrine mechanism. Generally, an elevated density of bacteria often leads to higher local concentrations of secreted signaling molecules [often known as autoinducers (AI)] that can be sensed by quorum-sensing receptors, which then induce an intracellular response. These remarkable signaling pathways allow bacteria to alter their gene expression according to the number of surrounding neighbor cells, and permits group-wide activities that would never be observed in a bacterium outside of a community.

The Lux TCS pathway in *V. harveyi* is comprised of three hybrid SKs (LuxN, LuxQ, and CqsA) that each recognize a specific AI signal (Bassler *et al.*, 1994a; Freeman *et al.*, 2000; Henke & Bassler, 2004) (Fig 10). These SKs funnel phosphate groups toward the HPt protein LuxU, which donates them to the RR LuxO (Freeman & Bassler, 1999b). LuxO is a negative regulator of luminescence (Bassler *et al.*, 1994b). In contrast to the paradigm that signals activate an SK, the AI molecules that bind to the Lux SKs inhibit their kinase activity, and allow for phosphatase activity. Thus, at low cell density (when concentrations of AI are low), LuxQ, LuxN, and CqsA function as kinases, leading to the phosphorylation of LuxO, which inhibits luminescence. With high levels of AI at high cell density, the phospho-transfer pathway is reversed, with the SKs functioning as phosphatases (via their REC domains) to remove the phosphoryl group from LuxU and
ultimately LuxO (Henke & Bassler, 2004). Non-phosphorylated LuxO cannot inhibit luminescence (Freeman & Bassler, 1999a); thus, at high cell density, the community of *Vibrio* cells bioluminesces.

The Lux pathway is complex; it contains three SKs that all function within a phosphorelay to control the same phenotype. Furthermore, each SK recognizes a different AI. Presumably, this intricate architecture allows a cell to mediate bioluminescence according to the repertoire of AI signals found in the surrounding environment. Clearly, the cell is ensuring this energetically taxing phenotype is only induced under the appropriate environmental conditions.

**Gac.** *Pseudomonas aeruginosa* is a Gram-negative organism most famous for its ability to opportunistically colonize the lungs of cystic fibrosis patients. It has one of the largest repertoires of regulatory systems, representing almost 10% of its genome (Stover *et al*, 2000). This is presumably due to its ability to colonize a variety of different niches both inside and outside a host (Stover *et al*, 2000). One of the master regulators of virulence pathways in *P. aeruginosa* is the GacA-GacS TCS system. In this pathway, the hybrid SK, GacS, recognizes an unknown signal to activate the RR, GacA, which can control virulence genes that are directed towards a variety of distinctive hosts (Brinkman *et al*, 2001; Chieda *et al*, 2005; Clatworthy *et al*, 2009; Coleman *et al*, 2003; Cosson *et al*, 2002; Jander *et al*, 2000; Mahajan-Miklos *et al*, 1999; Rahme *et al*, 1997; Tan *et al*, 1999) (Fig 11A). GacS is a hybrid SK with both a REC and an HPt domain involved in phosphotransfer to GacA; however, there is evidence that two other hybrid SKs, LadS and RetS, also regulate the Gac pathway (Goodman *et al*, 2004; Ventre *et al*, 2006). LadS positively regulates the Gac system; however, it is missing a C-terminal HPt domain, so it
Figure 11. Atypical TCS pathways, Gac and Rcs. (A) The hybrid SK, GacS, autophosphorylates, undergoes intramolecular phosphotransfer events, and then donates phosphoryl groups to the REC domain within GacA (“on”). Under unknown environmental conditions, a second hybrid SK, RetS, inhibits GacS autophosphorylation by directly binding to GacS (“off”). RetS does not require conserved sites of phosphorylation to control GacA/GacS mediated gene expression. (B) Upon signal recognition, the autokinase activity of RcsC ultimately leads to the phosphorylation of its REC domain. This domain passes phosphates to RcsD, which serves as the phosphodonor to RcsB. The RcsB regulon includes genes controlled solely by RcsB, or by an RcsB-RcsA complex. The intermembrane (IM) protein, IgaA, and the outermembrane (OM) protein, RcsF, control RcsC activity through an unknown mechanism.
cannot directly donate phosphoryl groups to GacA. It is unknown whether LadS functions as a SK or whether it directly or indirectly affects the Gac cascade. In contrast to GacS and LadS, the third SK, RetS, negatively regulates the Gac virulence pathways (Goodman et al, 2004). This hybrid SK, like LadS, is missing a C-terminal HPt domain; regardless, no sites of phosphorylation within RetS were required to inhibit virulence gene expression (Goodman et al, 2009), nor could it autophosphorylate \textit{in vitro} (Hsu et al, 2008). This suggested that it does not function as a canonical hybrid SK. Instead, researchers observed that RetS heterodimerizes with GacS, an interaction that prevents GacS from donating phosphoryl groups to GacA (Goodman et al, 2009).

It is unknown which signals activate either the autokinase activity of GacS or LadS, or the dimerization state of RetS and GacS; however, it is likely that these different signals allow for \textit{P. aeruginosa} to coordinate the activation of virulence factors according to a particular environmental niche the cell is experiencing. Importantly, the novel RetS-GacS interaction represents an unusual mechanism for TCS regulation, but it may not be unique; perhaps TCS systems in other bacteria have similarly evolved to allow an SK-SK interaction to control a signaling pathway, rather than the canonical phosphotransfer mechanism.

**Rcs.** Rcs is a complex TCS system conserved in most Enterobacteriaceae (Erickson & Detweiler, 2006). It was first analyzed in \textit{E. coli}, where it plays a role in the mucoid phenotype, due to its involvement in regulating the expression of capsular synthesis genes (Gottesman et al, 1985). Res contains an unusual architecture with a hybrid SK, RcsC, a membrane-bound HPt protein, RcsD, and a downstream RR, RcsB, which functions as a transcription factor (Brill et al, 1988; Stout & Gottesman, 1990;
Takeda et al, 2001) (Fig 11B). Interestingly, RcsD, although very similar in structure to hybrid SKs, has a degenerative HisKA domain; instead, it utilizes a C-terminal HPt domain to receive phosphoryl groups from the REC domain of RcsC, which it then donates to RcsB (Takeda et al, 2001).

The Rcs pathway is also unusual in that it contains a number of accessory proteins that function as additional regulators of this cascade. One key regulator is RcsA, which, similar to RcsD, represents a degenerative TCS regulator, although RcsA appears more similar to a RR rather than an SK. Surprisingly RcsA and the bona fide RR, RcsB, must interact to control the expression of a number of genes within the Rcs regulon (Ebel & Trempy, 1999; Pristovsek et al, 2003), although not all of them (Davalos-Garcia et al, 2001; Majdalani et al, 2002). RcsA is not the only accessory protein within the Rcs pathway: two additional regulators exist, IgaA (an inner-membrane protein) and RcsF (a outermembrane lipoprotein). They both seem to control the enzymatic activity of the membrane bound SK, RcsC; however, the mechanism of regulation is unknown, although it is most likely direct (Castanie-Cornet et al, 2006; Dominguez-Bernal et al, 2004; Gervais & Drapeau, 1992; Majdalani et al, 2005; Mariscotti & Garcia-del Portillo, 2009; Tierrez & Garcia-del Portillo, 2004).

Although the presence of auxiliary proteins within TCS cascades has only recently been appreciated, the list of these accessory regulators in various TCS systems in distinct bacterial species continues to grow (reviewed in (Buelow & Raivio, 2010)). This suggests that perhaps auxiliary regulators represent a more common means of control over TCS than is currently appreciated, and that even exhaustively characterized TCS systems may have additional regulators that have yet to be recognized.
Two-component regulation of biofilm formation in *V. fischeri*

*V. fischeri* experiences a large repertoire of environments both inside and outside the squid. As described in the symbiosis section of this dissertation, *V. fischeri* contains a number of different signaling pathways, including TCS pathways, which allow it to appropriately respond to environmental cues. One TCS cascade that is critical for host colonization is the Syp cascade, which controls biofilm formation during the initiation of the symbiosis (Fig 12). Here, I will further describe Syp, as it is the focus of this dissertation.

Before entry into the light organ of *E. scolopes*, *V. fischeri* forms an aggregate on the ciliated surfaces of the organ. This aggregate *in vivo*, and wrinkled colony formation (an indicator of biofilm formation) *in vitro*, requires the production of a polysaccharide matrix, which is regulated by the Syp TCS pathway. The Syp regulatory cascade is complex, in that it includes two hybrid SKs, RscS and SypF, two RRs, SypE and SypG, and a non-TCS STAS-domain protein, SypA (Fig 13). Much work has elucidated how these regulators, together, coordinate biofilm formation.

At the top of the Syp hierarchy lies RscS, the hybrid SK that is hypothesized to sense an environmental signal to induce biofilm formation (Fig 12). RscS was first identified using a transposon screen generated to identify *V. fischeri* mutants that exhibit colonization defects (Visick & Skoufos, 2001). Similarly, mutants with mutations in the 18-gene syp locus, which encodes proteins predicted to be involved in regulating and producing a polysaccharide, also exhibited severe colonization defects (Shibata *et al*, 2012; Yip *et al*, 2005). A link between RscS and syp polysaccharide production was first
Figure 12. Syp TCS pathway. (A) Previous model: the hybrid SK, RscS, functions upstream of two RRrs, SypE and SypG, to promote biofilm formation on agar plates (depicted as a wrinkled colony) and biofilm formation during host colonization (represented by an image of a squid). Phospho-SypG functions as a transcription factor to activate the transcription of the syp locus at four promoters, and SypE inhibits SypA activity at a level below syp transcription. When phosphorylated, SypE is no longer inhibitory. The function of SypA is unknown. Another hybrid SK, SypF, controls biofilms through an unknown mechanism. sypA, sypE, sypF, and sypG are located within the syp locus. (B) Revised model: the C-terminal domain of SypF functions between RscS and the two RRrs, SypE and SypG, thus bypassing the requirement for the C-terminal domain of RscS. The faded colors indicate domains found to be non-essential for biofilm formation in culture and during colonization.
Figure 13. Conserved domains found within Syp regulators. RscS and SypF are predicted to be SKs based on the presence of conserved HisKA (DHp) and HATPase (CA) domains. The presence of two additional domains within RscS and SypF [REC (receiver) and HPt (Histidine Phosphotransferase) domains] designate them as hybrid SKs. SypG (blue) and SypE (purple) are predicted to be RRs due to the presence of a REC domain. AAA+ and HTH (helix-turn-helix) domains within SypG suggest that SypG functions as a σ54 dependent transcription factor. Although SypE contains a N-terminal HATPase domain, this domain functions as a serine kinase, while its C-terminal PP2C domain functions as a serine phosphatase. SypA (yellow) is a single domain STAS (sulfate transporter and antisigma factor antagonist) protein with unknown function. Amino acids predicted to be sites of phosphorylation are indicated.
determined when overexpression of RscS was found to induce wrinkled colony formation and, importantly, *syp* transcription in a SypG-dependent manner (Hussa *et al*, 2008; Yip *et al*, 2006). RscS is a squid specificity factor; its introduction into a non-colonizing strain of *V. fischeri* allows it to colonize the squid at levels similar to the squid-isolate, *V. fischeri* ES114 (Mandel *et al*, 2009). Furthermore, *V. fischeri* ES114 that overexpresses RscS outcompetes vector-control cells through the formation of an unprecedentedly large aggregate outside the light organ (Yip *et al*, 2006). Based on its predicted function as an SK due to the presence of conserved CA/DHp regions, RscS most likely senses an environmental signal to permit aggregate formation and thus colonization of the squid. RscS has two domains potentially involved in signal recognition: a 200 amino acid periplasmic region and a cytosolic PAS domain. Mutagenesis studies suggested that the PAS domain, but not the extracytoplasmic region was required for biofilm formation (Geszvain & Visick, 2008a). The PAS domain exhibits structural homology with NifL from *Azotobacter vinelandii*, which binds a FAD cofactor (Key *et al*, 2007), and mutations within amino acids in RscS predicted to bind FAD prevented RscS from inducing biofilms (Geszvain & Visick, 2008a). Together, these data suggest that RscS senses a signal through its PAS domain, presumably through a bound co-factor, such as FAD. The identity of this signal remains unknown.

RscS is a predicted hybrid SK, and genetic experiments suggested that its cognate RRs were the *syp*-encoded biofilm regulators, SypE and SypG (Hussa *et al*, 2008; Morris *et al*, 2011; Visick & Skoufos, 2001). RscS contains both REC and HPt domains, which were originally predicted to be critical for biofilm induction (Fig 13). In support of this, mutational analysis indicated that the REC domain was required for wrinkled colony
formation; conversely, the HPt domain was largely dispensable for this phenotype (Geszvain & Visick, 2008a). This suggests that, if the HPt does play a role in biofilms, it is a minor one (Geszvain & Visick, 2008a). It is biochemically impossible for the REC domain in RscS to transfer phosphates to another REC domain in an RR. Thus, RscS may not directly phosphorylate SypE and SypG, implying an unknown TCS protein functions between these regulators. The identity of this regulator remains unknown.

RscS-induced biofilm formation requires the expression of the syp locus. syp expression is positively controlled by SypG, a $\sigma^{54}$-dependent RR (Hussa et al., 2008; Yip et al., 2005). SypG is part of the NtrC-like class of RRs. These RRs have a characteristic three domain structure that permit them to function as $\sigma^{54}$-dependent transcription factors, with a REC region attached to AAA$^+$ ATPase domain and a C-terminal helix-turn-helix domain used to bind DNA (reviewed in (Bush & Dixon, 2012)) (Fig 13). The ATPase domain provides energy to $\sigma^{54}$, a unique sigma factor that, unlike the $\sigma^{70}$-class of regulators, requires energetic input from an accessory enzyme to promote gene transcription. Additionally, $\sigma^{54}$ interacts with a non-canonical promoter sequence, located at the -12 and -24 positions upstream of a transcription start site, rather than at canonical -10 and -35 elements. In support of the hypothesis that the syp genes are part of the $\sigma^{54}$ regulon in V. fischeri, $\sigma^{54} (rpoN)$ is critical for SypG to induce syp transcription (Yip et al., 2005). Furthermore, four $\sigma^{54}$-dependent promoters exist within the syp locus, and SypG directly binds to all of them (Ray et al., 2013; Yip et al., 2005).

Interestingly, putative SypG-binding sites are found elsewhere in the chromosome of V. fischeri, suggesting that the SypG regulon extends outside of the syp locus. Indeed, SypG could directly bind and induce transcription of these genes (Ray et al., 2014; Ray et
al, 2013). Three of these promoters were predicted to control the expression of three similar two-gene operons that encode a putative periplasmic protein (Bmp) and a lipoprotein (Bal) (Ray & Visick, 2014). Surprisingly, the bmp, but not the bal, genes were required for biofilm formation; a triple mutant deleted for all three bmp genes formed smooth colonies, whereas a triple deletion of bal exhibited no effect on this phenotype. Although both syp mutants and bmp mutants alone failed to permit wrinkled colonies induced by RscS, a mixture of these two mutant strains resulted in the formation of wrinkled colonies. This result suggested that both the syp and bmp mutants were still secreting components of the biofilm matrix, and that full biofilm development required the presence of all these components. Further experiments demonstrated that the Bmp proteins constitute part of the extracellular biofilm matrix, where they aid in the maturation of the 3D structure of a wrinkled colony (Ray & Visick, 2014). Thus, the Syp biofilm consists of at least two independently secreted macromolecules: the Syp polysaccharide and the Bmp proteins. Whether other molecules, such as extracellular DNA, exist within this matrix remains unknown.

Although most TCS architectures include one SK and one RR, the Syp pathway contains two RRs: SypG, as described above, and SypE. Like SypG, SypE is controlled by RscS; however, unlike SypG, SypE exerts both positive and negative control over biofilms at a level below syp transcription (Hussa et al, 2008; Morris et al, 2011; Morris & Visick, 2013a). These contrasting functions of SypE are mediated by its two effector domains with opposing functions: an N-terminal serine kinase, and a C-terminal serine phosphatase (Fig 13). The phosphorylation state of a central REC domain within SypE determines which of these enzymatic activities is favored: unphosphorylated SypE
functions as a serine kinase, while phosphorylated SypE is a serine phosphatase (Morris et al, 2011). The opposing functions of SypE’s two effector domains are directed towards a small STAS domain protein, SypA, which is encoded at the beginning of the syp locus (Morris & Visick, 2013b) (Fig 12 and Fig 13). When phosphorylated by SypE, SypA is inactive, whereas unphosphorylated SypA promotes biofilm formation. The exact mechanism by which SypA controls biofilm formation is unknown. SypE and SypA contain domains found within partner-switching systems (Bhuwan et al, 2012; Duncan & Losick, 1993; Hua et al, 2006; Kozak et al, 2005; Mercer & Lang, 2014; Morris & Visick, 2013b; Yang et al, 1996) with SypA predicted to be an anti-anti-sigma factor; however, the target of SypA regulation remains an active area of research.

RscS may not be the only SK that regulates Syp biofilms; another hybrid SK, SypF, may control this phenotype (Fig 12 and Fig 13). The sypF gene is conspicuously located between the sypE and sypG genes, a typical arrangement for TCS regulators that function within the same pathway (Capra & Laub, 2012; Yip et al, 2005). Indeed, overexpression of an active variant of SypF, SypF*, was sufficient to induce wrinkled colonies and syp expression (Darnell et al, 2008). In support the possibility that SypF functions within the Syp TCS cascade, both SypE and SypG were important for SypF*-induced biofilms. However, an additional RR, VpsR, was also required for the SypF* phenotype, thus raising the possibility that SypF* exhibits cross-talk with other TCS systems, or else that the SypF* phenotype is not physiologically relevant (Darnell et al, 2008). Further confounding the importance of SypF in biofilm regulation is the observation that neither overexpression of wild-type SypF nor an insertional mutation
within the \textit{sypF} gene exerted any Syp-dependent phenotype (Darnell \textit{et al}, 2008; Hussa \textit{et al}, 2008). SypF will be discussed further in the results section of this dissertation.

**Conclusion**

TCS systems are the most common signaling pathway found in bacteria. They can be readily identified based on the presence of conserved domains, such as the CA/DHp regions in an SK, and the REC domain in the RR. Although these core enzymatic regions exhibit sequence similarity, individual SKs and RRs contain distinct signal input and effector output domains, respectively. This suggests that TCS pathways have evolved to allow bacteria to appropriately respond to a seemingly infinite number of environments. Furthermore, TCS pathways can deviate from the canonical two-protein paradigm, such as the Rcs pathway, which contains multiple non-TCS auxiliary proteins, and the Syp pathway, which utilizes two SKs and two RRs. Clearly, there exists flexibility in the architecture of TCS cascades, thus allowing bacteria exquisite control over specific cellular physiologies. The plasticity of these systems most likely explains why they are almost ubiquitously found in members of the bacterial kingdom, and how they can control a limitless number of phenotypes, such as bioluminescence, virulence, and biofilm formation.
CHAPTER TWO

MATERIALS AND METHODS

Bacterial strains and media

The bacterial strains used in this study are listed in Table I and were derived from ES114, a wild-type *V. fischeri* strain isolated from *E. scolopes* (Boettcher & Ruby, 1990). *V. fischeri* cells were grown in Luria-Bertani Salt (LBS) media (Graf *et al.*, 1994), Seawater Tryptone (SWT) media (Boettcher & Ruby, 1990), or HEPES Minimal Media (HMM) (Ruby & Nealson, 1977). *E. coli* strains used for molecular genetics in this study include: ER2508 (NEB), TAM1 \( \lambda pir \) (Active Motif), \( \pi 3813 \) (Le Roux *et al.*, 2007), CC118 \( \lambda pir \) (Herrero *et al.*, 1990) and GT115 (Invivogen). *E. coli* strains were grown in LB (Davis *et al.*, 1980). Solid media contained 1.5% agar. For *V. fischeri*, antibiotics were added to the following concentrations when necessary: erythromycin at 5 \( \mu g \) ml\(^{-1}\), kanamycin (Kan) at 50 \( \mu g \) ml\(^{-1}\) tetracycline (Tet) at 5 \( \mu g \) ml\(^{-1}\) in LBS or 30 \( \mu g \) ml\(^{-1}\) in SWT and HMM, or chloramphenicol (Cm) at 2.5 \( \mu g \) ml\(^{-1}\). The following antibiotics were added to *E. coli* media where appropriate: Cm at 25 \( \mu g \) ml\(^{-1}\), Tet at 15 \( \mu g \) ml\(^{-1}\), Kan at 50 \( \mu g \) ml\(^{-1}\), or ampicillin (Ap) at 100 \( \mu g \) ml\(^{-1}\).

Bioinformatics

All amino acid sequences were obtained from the National Center for Biotechnology Information (NCBI) database or the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Conserved protein domains were identified using NCBI.
Multiple sequence alignments were performed using the ClustalΩ program provided by EBI (European Bioinformatics Institute) (http://www.ebi.ac.uk/Tools/msa/clustalo/).

**Plasmid construction**

Plasmids used in this study are indicated in Table II. Plasmids were generated using either restriction digest-based cloning or Gibson assembly cloning [New England Biolabs (NEB)]. In some cases, DNA sequences of interest were amplified via PCR using the indicated primers and inserted into the pJET1.2 cloning vector (Thermo Fisher Scientific). DNA sequences were subsequently subcloned into mobilization vectors using standard restriction digest techniques. Alternatively, sequences were amplified using the indicated primers and then directly inserted into a mobilization vector using the Gibson Assembly approach (NEB). For site-directed mutagenesis of sypF, sypG, or sypF*, either Gibson Assembly or the Quick-Change Site-Directed Mutagenesis Kit (Stratagene) with the primer(s) indicated in Table III was used. To confirm that the correct DNA sequences were inserted into mobilization plasmids, these constructs were subjected to sequencing analyses (ACGT, Wheeling, IL)

**Genetic techniques**

Conjugation was used to move plasmids into *V. fischeri* using triparental or tetrarental mating techniques as described previously (McCann *et al*, 2003; Visick & Skoufos, 2001). Overexpression vectors include pVSV105, pKV282, and pKV69. To move DNA sequences into the Tn7 site of *V. fischeri*’s chromosome, these sequences were inserted into pEVS107 or a P<sub>lacZ</sub>-containing derivative, pARM47, and then incorporated into the chromosome using previously published methods (McCann *et al*, 2003).
To delete genes from the chromosome of *V. fischeri*, a previously published counter-selectable mutagenesis approach was used (Le Roux *et al.*, 2007; Shibata *et al.*, 2012). Briefly, suicide vectors containing two ~500 base pair (bp) sequences that flank a gene of interest were cloned into pKV363 using the methods described above. For deletions that were more difficult to obtain (*e.g.* Δ*yypE* Δ*yypF*) the size of the flanking DNA was increased to ~2500 bps each. These plasmids were then introduced into *V. fischeri* using conjugation (described above). Cells were grown on Cm-containing plates to select for the single insertion of the plasmid into the chromosome. Cm-resistant *V. fischeri* cells were then grown in 0.2% arabinose to induce the expression of a *ccdB* toxin gene, which negatively selects for cells that did not undergo a second recombination event. Viable cells were plated on LBS-0.2% arabinose, and colony PCR was performed on single colonies to verify that the gene of interest was deleted. These colonies were streaked twice more onto LBS plates and the colony PCR was repeated to ensure that these strains represented a true deletion.

**Wrinkled colony assay**

*V. fischeri* strains were freshly streaked on LBS-agar media containing the appropriate antibiotic and grown at 28°C. Single colonies were grown overnight with shaking at 28°C in LBS with the appropriate antibiotic and then subcultured and grown to early log-phase. The cell number in of each sample was standardized to an optical density of 600 nm (OD$_{600}$) of 0.2. 10 µl of the cultures were spotted onto LBS plates with the appropriate antibiotic to maintain plasmid selection. Spots were grown at room temperature (24°C) and images were captured at indicated time points using a Zeiss stemi 2000-C dissecting microscope.
**β-galactosidase measurements**

Indicated *V. fischeri* reporter strains were grown overnight in triplicate at 24°C with shaking in HMM with Tet. Cultures were back-diluted into fresh medium to an OD$_{600}$ of 0.2 and then grown for 24 h. 1 ml was removed, resuspended in Z-buffer, and lysed using chloroform. β-galactosidase activity was measured as described (Miller, 1972). Protein levels were assessed using previously described methods (Lowry *et al.*, 1951) and the data are reported as β-galactosidase activity/mg of protein. Alternatively, some data are reported as Miller Units, which represents the β-galactosidase activity as compared to OD$_{600}$ of the cells, rather than total mg of protein.

**Western blot procedure**

Overnight samples of *V. fischeri* cells were standardized by OD$_{600}$ and lysed with 2X sample buffer (100mM Tris pH 6.8, 4% SDS, 20% glycerol, 12% β-mercaptoethanol, 0.01% bromophenol-blue). When higher concentrations of cells were needed to assess SypF-FLAG levels expressed in single copy, samples were lysed with B-PER (Thermo Scientific) with 10 mg/mL DNase. Lysates were resolved on an SDS-polyacrylamide gel (10%, 29:1 acrylamide: $N$, $N'$-methylene-bis-acrylamide, 375 mM Tris pH 8.6, 0.1% SDS), transferred to a PVDF membrane, and subjected to standard western blot procedures using a rabbit anti-FLAG primary antibody (Sigma-Aldrich) and a anti-rabbit HRP-conjugated secondary antibody (Sigma-Aldrich). Proteins were visualized using SuperSignal West Pico Chemiluminescent Substrate (Thermo Fischer Scientific) with subsequent exposure to film.

**Protein expression**
Sequences encoding the REC domain of SypG and the cytoplasmic form of SypF were amplified by PCR and cloned into pMAL-c5x using Gibson Assembly to generate N-terminal maltose binding protein (MBP) fusion proteins. Plasmids were transformed into the ER2508 strain (NEB), a BL21 derivative that does not express native MBP. To purify cytoplasmic MBP-SypF (pANN48) and MBP-SypF* (pANN74), 1 l of Amp-containing LB was inoculated with the appropriate E. coli strain and grown to an OD of 0.7 at 37°C. Protein expression was induced with 0.1 mM IPTG at 18°C overnight. Cells were harvested by centrifugation (10,000 x g) for 10 min and lysed using B-PER detergent (Thermo Scientific) with 100 µl 20 mg/ml lysozyme (Thermo Scientific), 20 µl 10 mg/ml DNase (Sigma) and 50 µl 100 µM PMSF (Sigma). Lysates were cleared by centrifugation at 16,000 x g for 20 min. Supernatant was applied to an amylose-resin column (NEB), washed three times with 1X Phosphate Buffered Saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM NaH$_2$CO$_4$, 1.8 mM KH$_2$PO$_4$, pH 7.4), and eluted with 10 mM maltose. An Amicon 30k filter device (Millipore) equilibrated with storage buffer (50 mM Tris pH 8, 50 mM KCl, 50% glycerol) was used to exchange the elution buffer with storage buffer and to concentrate purified protein. To purify MBP-SypG-REC (pANN49), a similar approach as above was taken, except 500 ml of cells at an OD$_{600}$ of 0.5 were induced with 0.5 mM IPTG at 24°C overnight. To purify GST-SypE (pARM141) we modified the methods from Morris and Visick (Morris & Visick, 2013b) as follows: Briefly, pARM141 expressed from the ER2508 strain was used because this improved solubility of GST-SypE. This E. coli strain was grown to an OD$_{600}$ of 0.5 and then induced with 0.4 mM IPTG overnight. Cells were harvested and lysed with Bugbuster (Novagen), and the supernatants were applied to Glutathione Sepharose 4B
columns. Bound proteins were eluted with 10 mM glutathione. GST-SypE was concentrated and the elution buffer was exchanged with storage buffer using an Amicon 30k filter device (Millipore). Purified proteins were assessed by resolving samples on a 10% SDS-polyacrylamide gel with subsequent Coomassie Brilliant Blue R-250 protein staining (Thermo-Scientific) or western immunoblotting procedures as described above using anti-GST or anti-MBP primary antibodies (Sigma).

**in vitro assays**

Autokinase reaction: 2 μg/μL of purified MBP-SypF or MBP-SypF* were incubated in kinase buffer (50 mM Tris-HCl pH 8, 50 mM KCl, 5 mM MgCl₂, and 5 μCi [γ³²P]-ATP) for 30 minutes at 28°C. In reactions without radiolabeled ATP, the same volume of 2 mM of cold ATP was added. Samples were stopped with 5X sample buffer (250 mM Tris-HCl pH 6.8, 10% SDS, 20% glycerol, 3% β-mercaptoethanol, 0.01% bromophenol-blue), electrophoresed through a 10% SDS-polyacrylamide gel which was dried for 2 h and then exposed to film for 24-48 h. Phosphotransfer reactions: phospho-MBP-SypF or phospho-MBP-SypF* was obtained as described above. Equimolar concentrations of GST-SypE or MBP-REC were added and the reactions were incubated for 30 min. As a negative control, GST-SypE or MBP-SypG-REC were incubated in the same buffer conditions for 30 min but in the absence of a kinase. To assess levels of phosphorylated proteins, autoradiographs were generated as described above.

**Colonization assay**

*V. fischeri* strains were grown on SWT agar plates overnight and then inoculated and grown to early log phase in liquid SWT media without shaking at 28°C. Aposymbiotic juvenile squid were collected shortly after hatching and placed in artificial
sea water (ASW) (Instant Ocean, That Pet Place) that contained \textit{V. fischeri} strains at a concentration of 1000 cells per ml. Colonization was allowed to proceed for 18 h at which point individual \textit{E. scolopes} were homogenized in 70\% ASW. Serial dilutions of the homogenates were plated on SWT to determine the CFU of \textit{V. fischeri} per squid. Limit of detection is 14 CFUs of \textit{V. fischeri} per squid. Experiments involving \textit{E. scolopes} animals were carried out using approaches described in an Animal Component of Research Protocol (ACORP) approved by Loyola University’s Institutional Animal Care and Use Committee (IACUC) (LU #107314, 201297).
<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES114</td>
<td>Wild-type <em>V. fischeri</em></td>
<td>(Boettcher &amp; Ruby, 1990)</td>
</tr>
<tr>
<td>KV3246</td>
<td>attTn7::sypA-lacZ</td>
<td>(Morris &amp; Visick, 2013b)</td>
</tr>
<tr>
<td>KV4389</td>
<td>attTn7::ermR</td>
<td>(Morris et al, 2011)</td>
</tr>
<tr>
<td>KV4390</td>
<td>ΔsypE attTn7::ermR</td>
<td>(Morris et al, 2011)</td>
</tr>
<tr>
<td>KV4819</td>
<td>ΔsypE attTn7::sypE&lt;sup&gt;+&lt;/sup&gt;</td>
<td>(Morris et al, 2011)</td>
</tr>
<tr>
<td>KV4885</td>
<td>ΔsypE attTn7::sypE&lt;sup&gt;D192A&lt;/sup&gt;</td>
<td>(Morris et al, 2011)</td>
</tr>
<tr>
<td>KV5367</td>
<td>ΔsypF</td>
<td>This study</td>
</tr>
<tr>
<td>KV5478</td>
<td>ΔsypA attTn7::sypA</td>
<td>(Morris &amp; Visick, 2013b)</td>
</tr>
<tr>
<td>KV5481</td>
<td>ΔsypA attTn7::sypA&lt;sup&gt;SS6A&lt;/sup&gt;</td>
<td>(Morris &amp; Visick, 2013b)</td>
</tr>
<tr>
<td>KV5654</td>
<td>ΔsypA ΔsypG</td>
<td>This study</td>
</tr>
<tr>
<td>KV6092</td>
<td>sypO::pKV356</td>
<td>This study</td>
</tr>
<tr>
<td>KV6351</td>
<td>ΔrscS ΔsypF</td>
<td>This study</td>
</tr>
<tr>
<td>KV6439</td>
<td>ΔsypE ΔsypF</td>
<td>This study</td>
</tr>
<tr>
<td>KV6526</td>
<td>ΔsypF attTn7::sypG&lt;sup&gt;*&lt;/sup&gt;-FLAG</td>
<td>This study</td>
</tr>
<tr>
<td>KV6527</td>
<td>attTn7::sypG&lt;sup&gt;*&lt;/sup&gt;-FLAG</td>
<td>This study</td>
</tr>
<tr>
<td>KV6533</td>
<td>ΔrscS</td>
<td>This study</td>
</tr>
<tr>
<td>KV6586</td>
<td>ΔsypE ΔsypF attTn7::sypG&lt;sup&gt;*&lt;/sup&gt;-FLAG</td>
<td>This study</td>
</tr>
<tr>
<td>KV6659</td>
<td>ΔsypF attTn7::sypF-FLAG</td>
<td>This study</td>
</tr>
<tr>
<td>KV6692</td>
<td>ΔsypF attTn7::sypF&lt;sup&gt;D549A&lt;/sup&gt;-FLAG</td>
<td>This study</td>
</tr>
<tr>
<td>KV6896</td>
<td>ΔsypF attTn7::sypF&lt;sup&gt;H250Q&lt;/sup&gt;-FLAG</td>
<td>This study</td>
</tr>
<tr>
<td>KV6921</td>
<td>ΔsypF attTn7::ermR</td>
<td>This study</td>
</tr>
<tr>
<td>KV7013</td>
<td>ΔsypF ΔsypG sypH::pKV388</td>
<td>This study</td>
</tr>
<tr>
<td>KV7014</td>
<td>ΔsypE ΔsypF sypH::pKV388</td>
<td>This study</td>
</tr>
<tr>
<td>KV7074</td>
<td>ΔSE-A sypH::pKV388&lt;sup&gt;1&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>KV7085</td>
<td>ΔsypF attTn7::sypF&lt;sup&gt;H705Q&lt;/sup&gt;-FLAG</td>
<td>This study</td>
</tr>
<tr>
<td>KV7154</td>
<td>ΔsypF attTn7::sypF&lt;sup&gt;H705Q&lt;/sup&gt;-FLAG&lt;sup&gt;D549A&lt;/sup&gt;-FLAG</td>
<td>This study</td>
</tr>
<tr>
<td>KV7155</td>
<td>ΔsypF attTn7::sypF&lt;sup&gt;H705Q&lt;/sup&gt;-FLAG&lt;sup&gt;D549A&lt;/sup&gt;-FLAG&lt;sup&gt;H705Q&lt;/sup&gt;-FLAG</td>
<td>This study</td>
</tr>
<tr>
<td>KV7226</td>
<td>ΔsypF attTn7::HPt-FLAG</td>
<td>This study</td>
</tr>
<tr>
<td>KV7230</td>
<td>attTn7::sypG&lt;sup&gt;*&lt;/sup&gt;-FLAG P&lt;sub&gt;syp&lt;/sub&gt;-lacZ</td>
<td>This study</td>
</tr>
<tr>
<td>KV7231</td>
<td>ΔsypF attTn7::sypG&lt;sup&gt;*&lt;/sup&gt;-FLAG P&lt;sub&gt;syp&lt;/sub&gt;-lacZ</td>
<td>This study</td>
</tr>
<tr>
<td>KV7244</td>
<td>ΔsypA ΔsypF</td>
<td>This study</td>
</tr>
<tr>
<td>KV7276</td>
<td>sypH::pKV388</td>
<td>This study</td>
</tr>
<tr>
<td>KV7371</td>
<td>IG (yeiR-glmS): P&lt;sub&gt;syp&lt;/sub&gt;-lacZ&lt;sup&gt;2&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>KV7372</td>
<td>ΔsypF IG (yeiR-glmS): P&lt;sub&gt;syp&lt;/sub&gt;-lacZ&lt;sup&gt;2&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>KV7377</td>
<td>ΔsypF IG (yeiR-glmS): P&lt;sub&gt;syp&lt;/sub&gt;-lacZ attTn7::HPt-FLAG&lt;sup&gt;2&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>KV7386</td>
<td>ΔsypF IG (yeiR-glmS): P&lt;sub&gt;syp&lt;/sub&gt;-lacZ attTn7::sypF-FLAG&lt;sup&gt;2&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>KV7387</td>
<td>ΔsypF IG (yeiR-glmS): P&lt;sub&gt;syp&lt;/sub&gt;-lacZ attTn7::sypF&lt;sup&gt;H705Q&lt;/sup&gt;-FLAG&lt;sup&gt;2&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>KV7410</td>
<td>IG (yeiR-glmS): P&lt;sub&gt;syp&lt;/sub&gt;-lacZ-lacZ attTn7::ermR&lt;sup&gt;2&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>KV7412</td>
<td>ΔsypF IG (yeiR-glmS): P&lt;sub&gt;syp&lt;/sub&gt;-lacZ attTn7::ermR&lt;sup&gt;2&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>KV7413</td>
<td>ΔsypF IG (yeiR-glmS): P&lt;sub&gt;syp&lt;/sub&gt;-lacZ attTn7::HPt&lt;sup&gt;H705Q&lt;/sup&gt;-FLAG&lt;sup&gt;2&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>KV7485</td>
<td>ΔsypF attTn7::HPt&lt;sup&gt;H705Q&lt;/sup&gt;-FLAG</td>
<td>This study</td>
</tr>
<tr>
<td>KV7649</td>
<td>ΔsypF attTn7::rscS-sypF chimera</td>
<td>This study</td>
</tr>
<tr>
<td>KV7651</td>
<td>ΔsypF ΔrscS attTn7::rscS-sypF chimera</td>
<td>This study</td>
</tr>
<tr>
<td>KV7652</td>
<td>ΔsypF attTn7::rscS</td>
<td>This study</td>
</tr>
<tr>
<td>KV7653</td>
<td>ΔrscS attTn7::rscS</td>
<td>This study</td>
</tr>
<tr>
<td>Strain</td>
<td>Genetic Modification</td>
<td>Source</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------</td>
<td>--------</td>
</tr>
<tr>
<td>KV7654</td>
<td>ΔsypF ΔrscS attTn7::rscS</td>
<td>This study</td>
</tr>
<tr>
<td>KV7655</td>
<td>attTn7::rscS</td>
<td>This study</td>
</tr>
<tr>
<td>KV7656</td>
<td>ΔsypF ΔrscS attTn7::sypF</td>
<td>This study</td>
</tr>
<tr>
<td>KV7657</td>
<td>ΔsypF ΔrscS attTn7::ermR</td>
<td>This study</td>
</tr>
<tr>
<td>KV7658</td>
<td>attTn7::rscS-sypF chimera</td>
<td>This study</td>
</tr>
</tbody>
</table>

1 ΔSE: SypA enhancer; 27 bp removed
2 IG (yeiR-glmS) indicates the intergenic region between yeiR and glmS
Table 2. Plasmids used in this study

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Primers</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pANN9</td>
<td>pKV363 + 1.2 kb sequences flanking sypF sypG</td>
<td>910, 1160, 1222, 427</td>
<td>This study</td>
</tr>
<tr>
<td>pANN11</td>
<td>pKV282 + sypF</td>
<td>N/A</td>
<td>This study</td>
</tr>
<tr>
<td>pANN12</td>
<td>pKV282 + sypF&lt;sub&gt;D549A&lt;/sub&gt;</td>
<td>1295</td>
<td>This study</td>
</tr>
<tr>
<td>pANN15</td>
<td>pKV282 + sypF&lt;sub&gt;H1250Q&lt;/sub&gt;</td>
<td>1531</td>
<td>This study</td>
</tr>
<tr>
<td>pANN16</td>
<td>pKV282 + sypF&lt;sub&gt;H705Q&lt;/sub&gt;</td>
<td>1630</td>
<td>This study</td>
</tr>
<tr>
<td>pANN17</td>
<td>pKV363 + 3.8kb sequences flanking sypE sypF</td>
<td>1219, 519, 1249, 1375</td>
<td>This study</td>
</tr>
<tr>
<td>pANN20</td>
<td>pEVS107 + P&lt;sub&gt;lac&lt;/sub&gt;-sypF-FLAG</td>
<td>1609, 1563</td>
<td>This study</td>
</tr>
<tr>
<td>pANN21</td>
<td>pEVS107 + P&lt;sub&gt;lac&lt;/sub&gt;-sypF&lt;sub&gt;D549A&lt;/sub&gt;-FLAG</td>
<td>1609, 1563</td>
<td>This study</td>
</tr>
<tr>
<td>pANN24</td>
<td>pEVS107 + P&lt;sub&gt;lac&lt;/sub&gt;-sypF&lt;sub&gt;H1250Q&lt;/sub&gt;-FLAG</td>
<td>1609, 1563</td>
<td>This study</td>
</tr>
<tr>
<td>pANN34</td>
<td>pEVS107 + P&lt;sub&gt;lac&lt;/sub&gt;-sypF&lt;sub&gt;G*&lt;/sub&gt;-FLAG</td>
<td>1609 1438</td>
<td>This study</td>
</tr>
<tr>
<td>pANN45</td>
<td>pEVS107 + P&lt;sub&gt;lac&lt;/sub&gt;-sypF&lt;sub&gt;H705Q&lt;/sub&gt;-FLAG</td>
<td>1795, 1793, 1796, 1794</td>
<td>This study</td>
</tr>
<tr>
<td>pANN46</td>
<td>pEVS107 + P&lt;sub&gt;lac&lt;/sub&gt;-sypF&lt;sub&gt;H705Q&lt;/sub&gt; D&lt;sub&gt;549A&lt;/sub&gt; H&lt;sub&gt;705Q&lt;/sub&gt;-FLAG</td>
<td>1795, 1793, 1796, 1794</td>
<td>This study</td>
</tr>
<tr>
<td>pANN48</td>
<td>pMAL-c5x + SypF amino acids:</td>
<td>1828, 1829</td>
<td>This study</td>
</tr>
<tr>
<td>pANN49</td>
<td>pMAL-c5x + REC amino acids:</td>
<td>1809, 1810</td>
<td>This study</td>
</tr>
<tr>
<td>pANN50</td>
<td>pARM47 + P&lt;sub&gt;lac&lt;/sub&gt;-HPt-FLAG</td>
<td>1902, 1796</td>
<td>This study</td>
</tr>
<tr>
<td>pANN52</td>
<td>pANN34 + P&lt;sub&gt;lac&lt;/sub&gt;-&lt;sub&gt;lacZ&lt;/sub&gt;</td>
<td>N/A</td>
<td>This study</td>
</tr>
<tr>
<td>pANN58</td>
<td>pARM47 P&lt;sub&gt;lac&lt;/sub&gt;-HPt&lt;sub&gt;H705Q&lt;/sub&gt;-FLAG</td>
<td>1902, 1796</td>
<td>This study</td>
</tr>
<tr>
<td>pANN59</td>
<td>pMO8 + P&lt;sub&gt;lac&lt;/sub&gt;-&lt;sub&gt;lacZ&lt;/sub&gt;</td>
<td>1860, 1861</td>
<td>This study</td>
</tr>
<tr>
<td>pANN61</td>
<td>pKV69 + sypF&lt;sub&gt;D549A&lt;/sub&gt;</td>
<td>1295</td>
<td>This study</td>
</tr>
<tr>
<td>pANN62</td>
<td>pKV69 + sypF&lt;sub&gt;H705Q&lt;/sub&gt;</td>
<td>1569</td>
<td>This study</td>
</tr>
<tr>
<td>pANN65</td>
<td>pEVS107 + P&lt;sub&gt;lac&lt;/sub&gt;-sypF&lt;sub&gt;H705Q&lt;/sub&gt; D&lt;sub&gt;549A&lt;/sub&gt;-FLAG</td>
<td>1795, 1796</td>
<td>This study</td>
</tr>
<tr>
<td>pANN69</td>
<td>pCLD29 + rscS-sypF chimera-FLAG ²</td>
<td>1899, 1900, 1901, 1882</td>
<td>This study</td>
</tr>
<tr>
<td>pANN70</td>
<td>pCLD29 + sypF*-FLAG ²</td>
<td>1881, 1882</td>
<td>This study</td>
</tr>
<tr>
<td>pANN71</td>
<td>pCLD29 + sypF&lt;sub&gt;H705Q&lt;/sub&gt;-FLAG ²</td>
<td>1881, 1786, 1785, 1882</td>
<td>This study</td>
</tr>
<tr>
<td>pANN72</td>
<td>pCLD29 + sypF&lt;sub&gt;D549A&lt;/sub&gt;-FLAG ²</td>
<td>1295</td>
<td>This study</td>
</tr>
<tr>
<td>pANN73</td>
<td>pCLD29 + sypF&lt;sub&gt;S247F&lt;/sub&gt;-FLAG ²</td>
<td>1881, 1784, 1783, 1882</td>
<td>This study</td>
</tr>
<tr>
<td>pANN74</td>
<td>pMAL-c5x + SypF* amino acids:</td>
<td>1828, 1829</td>
<td>This study</td>
</tr>
<tr>
<td>pANN76</td>
<td>pCLD29 + sypF&lt;sub&gt;H705Q&lt;/sub&gt;-FLAG ²</td>
<td>1881, 1793, 1794, 1882</td>
<td>This study</td>
</tr>
<tr>
<td>pANN77</td>
<td>pARM47 plac-rscS-sypF chimera ¹</td>
<td>1908, 1907</td>
<td>This study</td>
</tr>
<tr>
<td>pANN78</td>
<td>pARM47 plac-rscS&lt;sub&gt;1&lt;/sub&gt;</td>
<td>1908, 1909</td>
<td>This study</td>
</tr>
<tr>
<td>pARM7</td>
<td>RscS overexpression construct; Tet&lt;sub&gt;R&lt;/sub&gt;</td>
<td>N/A</td>
<td>(Morris et al, 2011)</td>
</tr>
<tr>
<td>pARM9</td>
<td>pVSV105 + sypG</td>
<td>N/A</td>
<td>(Morris &amp; Visick, 2013a)</td>
</tr>
<tr>
<td>pARM47</td>
<td>pEVS107 + plac-sypE</td>
<td>N/A</td>
<td>(Morris et al, 2011)</td>
</tr>
<tr>
<td>pARM131</td>
<td>pEV107 + plac-sypA</td>
<td>N/A</td>
<td>(Morris &amp; Visick, 2013b)</td>
</tr>
<tr>
<td>pARM133</td>
<td>pEV107 + plac-sypA&lt;sub&gt;S56A&lt;/sub&gt;</td>
<td>N/A</td>
<td>(Morris &amp; Visick, 2013b)</td>
</tr>
<tr>
<td>pARM141</td>
<td>pGEX-5X-1 + sypE</td>
<td>N/A</td>
<td>(Morris &amp; Visick, 2013b)</td>
</tr>
<tr>
<td>pCLD29</td>
<td>pKV69 + sypF*</td>
<td>N/A</td>
<td>(Darnell et al, 2008)</td>
</tr>
<tr>
<td>pCLD46</td>
<td>pKV282 + rscS</td>
<td>N/A</td>
<td>(Hussa et al, 2008)</td>
</tr>
<tr>
<td>pCLD54</td>
<td>pKV69 + sypF</td>
<td>N/A</td>
<td>(Darnell et al, 2008)</td>
</tr>
<tr>
<td>pESY20</td>
<td>Ori6K suicide plasmid, mob, Erm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>N/A</td>
<td>(O'Shea et al, 2006)</td>
</tr>
<tr>
<td>pEV104</td>
<td>Conjugal helper plasmid (tra trb)</td>
<td>N/A</td>
<td>(Stabb &amp; Ruby, 2002)</td>
</tr>
<tr>
<td>Vector</td>
<td>Description</td>
<td>Resistance</td>
<td>Source</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------------</td>
<td>--------</td>
</tr>
<tr>
<td>pEVS107</td>
<td>Mini-Tn7 delivery plasmid, OriR6K, mob</td>
<td>N/A</td>
<td>(McCann et al, 2003)</td>
</tr>
<tr>
<td>pKV282</td>
<td>Mobilization vector, TetR</td>
<td>N/A</td>
<td>(Morris et al, 2011)</td>
</tr>
<tr>
<td>pJET1.2</td>
<td>Commercial cloning vector, ApR</td>
<td>N/A</td>
<td>Fermentas</td>
</tr>
<tr>
<td>pJMO8</td>
<td>Suicide vector with sequences flanking region adjacent to Tn7 site</td>
<td>N/A</td>
<td>(Ondrey &amp; Visick, 2014)</td>
</tr>
<tr>
<td>pKG63</td>
<td>pKV69 + rscS</td>
<td>N/A</td>
<td>(Geszvain &amp; Visick, 2008a)</td>
</tr>
<tr>
<td>pGI08</td>
<td>pKV69 + rsc51467Q</td>
<td>N/A</td>
<td>(Geszvain &amp; Visick, 2008a)</td>
</tr>
<tr>
<td>pKPQ17</td>
<td>pKV363 + 1 kb sequences flanking sypF</td>
<td>910, 1160, 1249, 271</td>
<td>This study</td>
</tr>
<tr>
<td>pKV69</td>
<td>Mobilization vector, CmR, TetR</td>
<td>N/A</td>
<td>(Visick &amp; Skoufos, 2001)</td>
</tr>
<tr>
<td>pKV276</td>
<td>pKV69 + sypG*</td>
<td>N/A</td>
<td>(Hussa et al, 2008)</td>
</tr>
<tr>
<td>pKV356</td>
<td>pESY20 + 400 bp sypO-HA</td>
<td>1150, 1151</td>
<td>This study</td>
</tr>
<tr>
<td>pKV363</td>
<td>Suicide plasmid, CmR</td>
<td>N/A</td>
<td>(Shibata et al, 2012)</td>
</tr>
<tr>
<td>pKV388</td>
<td>pESY20 + 1 kb sypH-HA</td>
<td>426, 1147</td>
<td>This study</td>
</tr>
<tr>
<td>pKV456</td>
<td>pKV363 + 1.5 kb sequences flanking rscS</td>
<td>1494, 1495, 1496, 1497</td>
<td>This study</td>
</tr>
<tr>
<td>pMAL-c5X</td>
<td>Commercial MBP tag protein expression vector; ApR</td>
<td>N/A</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>pVSV105</td>
<td>Mobilization vector; CmR</td>
<td>N/A</td>
<td>(Dunn et al, 2006)</td>
</tr>
<tr>
<td>pUX-BF13</td>
<td>Transposase expressing vector</td>
<td>N/A</td>
<td>(Bao et al, 1991)</td>
</tr>
</tbody>
</table>

1. restriction enzymes were used to remove the original sypE sequence but maintain P_lac to drive expression of inserted DNA sequences
2. the original sypF* sequence was removed from pCLD29 using restriction enzymes before the insertion of indicated DNA sequences
Table 3. Primers used in this study

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5' - 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>271</td>
<td>CTCGGGCATACCTTCTTTAC</td>
</tr>
<tr>
<td>426</td>
<td>TCCCTTCAAGCCTGGTACG</td>
</tr>
<tr>
<td>427</td>
<td>TAATACGCTTGTGCTTTG</td>
</tr>
<tr>
<td>519</td>
<td>GGGTGGTTTACCTGGTAC</td>
</tr>
<tr>
<td>910</td>
<td>GTGGTGTAATCATGGCTAGG</td>
</tr>
<tr>
<td>1147</td>
<td>TTATTTGCTATACCTTCTT</td>
</tr>
<tr>
<td>1150</td>
<td>CTATAATTGGTCCATTTCC</td>
</tr>
<tr>
<td>1151</td>
<td>CTATAATGATCTAGTTGAC</td>
</tr>
<tr>
<td>1160</td>
<td>TACCTATTGTAATAATGGATAG</td>
</tr>
<tr>
<td>1219</td>
<td>TCATCTAAGTGCGGCCGCTAGAAGCCTATGAAGAATCGGAATGATG</td>
</tr>
<tr>
<td>1222</td>
<td>CATACTAAGTGCGGCCGCTAGAAGCCTATGAAGAATCGGAATGATG</td>
</tr>
<tr>
<td>1249</td>
<td>CATACTAAGTGCGGCCGCTAGAAGCCTATGAAGAATCGGAATGATG</td>
</tr>
<tr>
<td>1295</td>
<td>(P)GACCTTATTTTCATGGCTATATCTATGCCTGAAATGGATGGCATGACGCCC</td>
</tr>
<tr>
<td>1375</td>
<td>TCATCTTCCGATTTTCCTCAG</td>
</tr>
<tr>
<td>1438</td>
<td>AAAAGGTAGCTTATTTATCATCATCATCTTTATAATATCTTTGGATTTTTCATAGGCTTCCCA</td>
</tr>
<tr>
<td>1494</td>
<td>TACCTAGTGCCCTCGTGTG</td>
</tr>
<tr>
<td>1495</td>
<td>GGGTAGCTCCAAAGATCAG</td>
</tr>
<tr>
<td>1496</td>
<td>TAGGCGGCCGCACTTATAATGATGAGGCTATAGGCTTXAAAGAATCGGAATGATG</td>
</tr>
<tr>
<td>1497</td>
<td>CATATTAGTGCGGCCGCTAGAAGCCTATGAAGAATCGGAATGATG</td>
</tr>
<tr>
<td>1563</td>
<td>ACCCGGGTTATTTATCATCATCATCTTTATAATCTTTGAGAAACCTTTGATATTTT</td>
</tr>
<tr>
<td>1569</td>
<td>(P)GCATTAGGTGTTGAAAGCCAAACATTAGGAGCATGTCATTTAAG</td>
</tr>
<tr>
<td>1609</td>
<td>AACTAGTGGCCAGCACAGTTTCCCGAC</td>
</tr>
<tr>
<td>1783</td>
<td>GTAAAAGTGTATTTTATGCCTTTGATGAGTAC</td>
</tr>
<tr>
<td>1784</td>
<td>GGGTGGCTTATTTATCATCATCATCTTTATAATCTT</td>
</tr>
<tr>
<td>1785</td>
<td>CAATATGATATTCTATGAGCTATATCTATTGAGC</td>
</tr>
<tr>
<td>1786</td>
<td>CATATTAGTGCCCTCGTATAGATAGCCATGAAAATAAG</td>
</tr>
<tr>
<td>1793</td>
<td>AATGTCATGCTTCTATAGTTTGTGCTTCAAAACTT</td>
</tr>
<tr>
<td>1794</td>
<td>ATGATTAGGTGTTTGAAGCCAAACATTAGGAG</td>
</tr>
<tr>
<td>1795</td>
<td>GATCTACTAGTGCGGAGTAACGCCTGAGGTGTTTCGGAC</td>
</tr>
<tr>
<td>1796</td>
<td>CGAGTCGTCGTTGATAGTCCTTTAAGGCTGTATCTTTATCAGT</td>
</tr>
<tr>
<td>1809</td>
<td>TCATCTAAGTGCGGCCGCTAGAAGCCTATGAAGAATCGGAATGATG</td>
</tr>
<tr>
<td>1810</td>
<td>CAGGGAATTCGGGACAGTTCATTTGAGAATTTGAGG</td>
</tr>
<tr>
<td>1818</td>
<td>TAGCATTAGGTGTTTGAAGCCAAACATTAGGAG</td>
</tr>
<tr>
<td>1828</td>
<td>TCAATAGTGGCCAGCACAGTTTCCCGAC</td>
</tr>
<tr>
<td>1829</td>
<td>CAGGGAATTCGGGACAGTTCATTTGAGAATTTGAGG</td>
</tr>
<tr>
<td>1860</td>
<td>AACACTATAAGTGCGGCCGCTCTTTATCTAGCTATGGATAGATA</td>
</tr>
<tr>
<td>1861</td>
<td>GAGAGACAATATGAGCAGCTCTTTGAGAATTTGAGG</td>
</tr>
<tr>
<td>1881</td>
<td>GCATCTACTGCTGAGTGCTAGT</td>
</tr>
<tr>
<td>1882</td>
<td>CGAGCTGGCTACCCAGGATCTTATATCCCTATCATCATATTAT</td>
</tr>
<tr>
<td>1899</td>
<td>GCTTGCAATGCGCTGAGCATTTGCAGGTAATCTCACCATGAAATAAGC</td>
</tr>
<tr>
<td>1900</td>
<td>TGAATCTACCTTATCATCATCATCTATATTTGAGAATTTGAGG</td>
</tr>
<tr>
<td>1901</td>
<td>TCTTCAATCAGTGGAGTATTGAATGAGCTATTTAATAGTA</td>
</tr>
<tr>
<td>1902</td>
<td>GATAGCCGAAGCTTCGATGAGGAGCTAATACATAGGATAATGAGCTATTTAATAGTA</td>
</tr>
<tr>
<td>1907</td>
<td>CCAGTCTAGGTGTAGCTGATCTGACGAGCGT</td>
</tr>
<tr>
<td>1908</td>
<td>GATAGCCGAAGCTTCGATGAGGAGCTAATACATAGGATAATGAGCTATTTAATAGTA</td>
</tr>
<tr>
<td>1909</td>
<td>CCAGTCTAGGTGTAGCTGATCTGACGAGCGT</td>
</tr>
</tbody>
</table>
CHAPTER THREE
EXPERIMENTAL RESULTS

1. SypF*: an active sensor kinase

Introduction

Successful colonization requires that \textit{V. fischeri} cells form and disperse from a biofilm-like aggregate to enter the symbiotic organ, known as the light organ (Nyholm \textit{et al.}, 2000; Yip \textit{et al.}, 2006). Formation of this aggregate depends on the production of the symbiosis polysaccharide (Syp) generated by proteins encoded by the 18-gene \textit{syp} locus (Shibata \textit{et al.}, 2012; Yip \textit{et al.}, 2006). Control over Syp production occurs via a complex TCS cascade. Previous work indicated that the hybrid SK (RscS) senses an unknown signal that leads to the phosphorylation of two downstream RRs (SypE and SypG) (reviewed in (Visick, 2009)) (Fig 12). Both \textit{sypE} and \textit{sypG} are located within the \textit{syp} locus, whereas \textit{rscS} is located elsewhere in the chromosome (Visick & Skoufos, 2001; Yip \textit{et al.}, 2005). Encoded between the RR genes \textit{sypE} and \textit{sypG} is an additional hybrid SK, SypF (Fig 13), a genetic configuration that is typical for TCS proteins that function together, suggesting that SypF, too, might regulate aggregate formation. Indeed, isolation of an active variant of SypF, SypF* (see below), supported a role for SypF in biofilms.

\textit{V. fischeri} aggregates must form on the surface of the light organ for this bacterium to colonize \textit{E. scolopes}. This \textit{in vivo} phenotype is directly correlated with \textit{V. fischeri}'s ability to form a wrinkled colony on agar plates, an \textit{in vitro} indicator of biofilm formation. The Syp TCS pathway controls both these phenotypes. In contrast to host
colonization, wrinkled colony formation by *V. fischeri* must be induced upon artificial overexpression of Syp regulators. Two of these regulators are SKs: RscS and SypF* (Darnell *et al.*, 2008; Hussa *et al.*, 2008; Yip *et al.*, 2006) (Fig 13). SypF* is a mutant variant of wild-type SypF; its overexpression can induce wrinkled colony formation, whereas overexpression of wild-type SypF cannot (Darnell *et al.*, 2008). Within SypF*, two mutations exist (S247F and V439I). The former is located 3 residues away from the predicted site of autophosphorylation (H250), a position that suggested SypF* could function as an SK within the Syp cascade to control biofilms. Indeed, SypF* could induce transcription of the *syp* locus, and SypF*-induced wrinkled colonies required *sypG* (Darnell *et al.*, 2008). However, many questions about SypF* remain unanswered, such as whether it functions as a bona fide SK, whether it requires both mutations to exhibit the “active” phenotype, where in the Syp signaling pathway might it function, and what is its role in host colonization. The goal of this dissertation was to understand how SypF regulates biofilms *in vitro* and *in vivo*, thus expanding our knowledge of how an intricate TCS pathway may control complicated phenotypes.

**SypF*: role as an SK

Bioinformatics readily identify SKs based on the conservation of key enzymatic motifs within the CA and DHp domains (Parkinson & Kofoid, 1992; Stock *et al.*, 1988). To determine whether SypF might function as an SK, I aligned these key SK motifs within SypF with these regions in SKs from other TCS pathways (Fig 14). The SKs I chose for the alignment have been demonstrated to function as SKs both *in vitro* and within the cell: EnvZ from *E. coli*, one of the most comprehensively studied canonical
Figure 14: SypF sequence alignment with known functional SKs. Colors indicate degrees of conservation: red = 100% conserved, blue = highly conserved, brown = moderately conserved, black = not conserved. The H-box region is found in the HisKA, or DHp, domain of SKs and contains the conserved site of histidine autophosphorylation. The N, D, F, and G boxes are found in the HATPase_c, or CA, domain where they bind to ATP and/or a metal co-factor (often Mg$^{2+}$) required for enzymatic activity.
SKs (Delgado et al, 1993; Roberts et al, 1994; Tanaka et al, 1998; Tomomori et al, 1999; Zhu et al, 2000); ArcB from *Escherichia coli*, a model for the function of hybrid SKs (Georgellis et al, 1998; Iuchi, 1993; Kwon et al, 2000; Pena-Sandoval & Georgellis, 2010; Pena-Sandoval et al, 2005; Tsuzuki et al, 1995); and VieS (Martinez-Wilson et al, 2008), a hybrid SK from the closely related bacterium, *Vibrio cholerae*. As seen in Fig 14, although these SKs did not exhibit high sequence identity, there were clear sequence similarities within the critical SK motifs, including identity with the conserved amino acids for which the names of each motif were coined. This suggested that SypF functions as an SK.

I next asked whether SypF* exhibited enzymatic activity. The SK autokinase reaction requires ATP, wherein the γ-phosphate group from an ATP molecule is hydrolyzed and covalently attached to the histidine side chain within the site of phosphorylation. To test whether SypF exhibits this catalytic activity, I used an *in vitro* approach. I purified a recombinant, cytoplasmic portion of SypF*, incubated it with [γ-32P]-ATP, resolved the reaction on a SDS-PAGE protein gel, and then generated an autoradiograph. Under these *in vitro* conditions, I found that SypF* became radiolabeled, demonstrating that SypF* exhibits autokinase activity (Fig 15A).

Although the above data established that SypF* can autophosphorylate *in vitro*, I next asked whether this activity was important within the context of a *V. fischeri* cell. To assess this, I tested whether the SypF* overexpression plasmid (pSypF*), which is sufficient to induce wrinkled colonies in *V. fischeri* (Darnell et al, 2008), required conserved amino acid residues to promote this biofilm phenotype. SypF encodes a hybrid SK, with three domains that each contain a site of phosphorylation: the DHp (H250),
Figure 15. SK activity of SypF*. (A) MBP-tagged SypF* or MBP-tagged wild-type SypF were incubated with either cold ATP or $[\gamma-32P]$-ATP for 30 minutes, electrophoresed on an SDS-polyacrylamide gel, and then exposed to film for 24 hours. (B) Colony morphology of wild-type *V. fischeri* cells expressing vector control (VC), pSypF*, or mutant derivatives of pSypF* after 24 h. Plasmids are as indicated: VC (pKV69) pSypF* (pCLD29), pSypF*$^{D549A}$ (pANN61), pSypF*$^{H705Q}$ (pANN62).
REC (D549), and HPt (H250) domains (Fig 13). In other hybrid SKs, all of these sites of phosphorylation are required for SK activity. To test whether these residues were important for SypF* activity, I mutated two putative sites of phosphorylation (D549 and H705) within pSypF*, and assessed the wrinkled colony phenotypes induced by these mutant plasmids. Whereas pSypF* induced biofilms, pSypF*D549A and pSypF*H250Q failed to do so (Fig 15B). These data support the hypothesis that SypF* functions as an SK in *V. fischeri.*

One caveat to the above experiment was that SypF*D549A and SypF*H250Q variants exhibit a negative phenotype. This could be due to 1. conserved amino acids being required for SypF* to function, or 2. the mutant proteins not being expressed. To confirm that the SypF* variants were produced, I generated and expressed FLAG epitope-tagged versions of SypF*, SypF*D549A and SypF*H705Q. I also generated an additional SypF*H250Q FLAG-tagged mutant to test whether the site of autophosphorylation, H250, was similarly required for biofilms. I then used western blot analysis to assess the levels of these proteins and colony morphology to assess their ability to induce biofilm formation. Importantly, I found that the steady-state levels of all these SypF variants were similar (Fig 16B). However, the FLAG tag diminished the biofilm-inducing activity of SypF* (Fig 16A, compare pSypF* to pSypF*-FLAG). Regardless, the H250Q, D549A, and H705Q mutants failed to induce the formation of biofilms. Together, these data confirm that SypF* functions as a hybrid SK within *V. fischeri.*

Although overexpression of SypF* is sufficient to induce wrinkled colony formation, overexpression of wild-type SypF induces no phenotype (Darnell *et al*, 2008). This seemed surprising, considering SypF* only contains two mutations (S247F and
V439I), yet exhibits a dramatically altered activity. To begin to elucidate how SypF* may function differently than wild-type SypF, I asked which mutation(s) within SypF* was sufficient to generate the active phenotype. Previous work demonstrated that the V439I mutation alone failed to permit SypF to induce biofilms (Darnell et al., 2008); thus, I asked whether the other mutation, S247F, could beget the active activity to SypF. To test this, I introduced the S247F mutation into otherwise wild-type SypF overexpressed from a plasmid, and assessed this mutant’s ability to induce wrinkled colonies. SypF\textsuperscript{S247F} also contained a FLAG tag, which allowed me to assess its protein levels via western blot; indeed, this protein was produced (Fig 16B). Furthermore, I found that the SypF\textsuperscript{S247F} mutant promoted wrinkled colony development to approximately the same extent as SypF*-FLAG (Fig 16A), demonstrating that this substitution is sufficient to produce SypF* activity.

**SypF*: function within the Syp signaling pathway**

SypF* clearly exhibits catalytic activity that is similar to canonical hybrid SKs. However, the purpose of this autokinase activity is to obtain phosphate groups that can be donated to a downstream RR. Thus, which RR(s) receives these phosphoryl moieties from SypF*?

The genetic arrangement of the *sypE, sypF,* and *sypG* genes suggests that these TCS regulators function within the same pathway (Fig 12). Indeed, it was previously demonstrated that biofilms induced by SypF* required *sypG*. This indicated that SypF* functioned upstream of SypG. To test whether SypF* also functioned upstream of SypE, I used a genetic approach. For Syp-dependent biofilms to form, SypE must be phosphorylated (Morris *et al.,* 2011). I reasoned that, if SypF* induced Syp-dependent
Figure 16. Wrinkled colony phenotype and protein levels of SypF*-FLAG variants.

(A) Wild-type *V. fischeri* cells expressing indicated plasmids were spotted on an agar plate and wrinkled colony formation was assessed after 40 h. (B) Anti-FLAG western blot analysis of wild-type *V. fischeri* strains expressing indicated plasmids. Plasmids are as indicated: pSypF* (pCLD29); pSypF*-FLAG (pANN70); pSypF*H250Q-FLAG (pANN71); pSypF*D549A-FLAG (pANN72); pSypF*H705Q-FLAG (pANN76); pSypF*S247F-FLAG (pANN73).
biofilms, then SypF*-induced wrinkled colonies should also require that SypE be phosphorylated. Therefore, expressing a SypE mutant in which site of phosphorylation (D192) (Fig 13) is mutated should prevent SypF* from promoting wrinkled colonies. In support of this prediction, SypF* was unable to induce wrinkled colony formation in a sypE deletion strain expressing SypE<sup>D192A</sup>, whereas reintroduction of wild-type SypE into the ΔsypE strain permitted the SypF* phenotype (Fig 17A). Combined, these data suggest that SypF* functions above both SypE and SypG.

To more directly assess the ability of SypF* to interact with and control SypG and SypE, I evaluated whether SypF* could donate phosphates to these RRs <i>in vitro</i>. To test this, I purified the REC domain of SypG (a previous attempt to purify full-length SypG was unsuccessful) and the full-length form of SypE, and added these purified proteins to reactions containing phosphorylated SypF*. If SypF* can directly interact with and phosphorylate these RRs, then SypE and SypG should become phosphorylated in the presence of phospho-SypF*. In support of the genetic data, I detected phosphorylated forms of SypE and SypG after incubation with phospho-SypF* (Fig 17B). These data indicate that SypF* can directly interact with and phosphorylate these two RR proteins.

If SypF* directly donates phosphoryl group to SypE and SypG to control biofilms, then what is the role of the other biofilm-inducing SK, RscS, in SypF*-induced wrinkled colonies? I hypothesized that RscS and SypF* could function at the same level, (<i>e.g.</i> both serve as phosphoryl donors to SypE and SypG) or that SypF* could work below RscS in the Syp pathway. To test whether RscS is important for SypF*-induced biofilms, I deleted <i>rscS</i> from the chromosome, moved pSypF* into this strain, and assessed wrinkled colony formation. I expected that SypF* biofilms could be 1.
Figure 17. Function of SypF* relative to SypE and/or SypG. (A) Indicated strains were spotted on agar plates and SypF*-induced (pCLD29) wrinkled colonies were assessed after 43 h. SypE and the D192A variant were expressed in single copy from the chromosome. (B) Phospho-MBP-SypF* was incubated with GST-SypE or MBP-REC for 30 minutes and then subjected to autoradiography. “REC” represents the REC domain from SypG.
unchanged (if SypF functions at or below RscS) or 2. diminished (if RscS and SypF function at the same level to control biofilms). I found that SypF* biofilms were indistinguishable between WT and ΔrscS cells, demonstrating RscS is expendable for this phenotype (Fig 18A). This result suggests either that SypF* and RscS function at the same level, or that SypF* functions below RscS in the Syp TCS cascade.

The role of SypF* in host colonization

Syp regulators are required for both wrinkled colony formation and squid colonization. Due to the role of SypF* in inducing wrinkled colonies, I predicted that it would also be involved in promoting host association. As mentioned above, both RscS and SypF* can induce wrinkled colony formation. Previously, it was published that RscS overexpression enhanced squid colonization (Yip et al., 2006); therefore, I reasoned that SypF* might also augment host association. To test this, I compared the colonization phenotype of WT cells expressing either vector or pSypF*. I found that cells with pSypF* colonized *E. scolopes* to higher CFU than vector containing cells at the 18 h time point (Fig 18B). These results suggest that SypF* enhances biofilm formation in vivo, which in turn facilitates efficient colonization.

Above, I determined that SypF* could promote wrinkled colony formation even in an rscS deletion strain. To assess whether SypF* could function at or below RscS during biofilm formation in vivo, I determined the colonization phenotype of the ΔrscS strain expressing pSypF*. rscS is absolutely required for wild-type cells to colonize *E. scolopes* (Visick & Skoufos, 2001); however, if SypF* functions below RscS, then pSypF* should permit an rscS mutant to colonize *E. scolopes*. Indeed, SypF* expression in ΔrscS cells mostly suppressed the colonization defect of this strain: ΔrscS cells
Figure 18. Role of RscS in SypF*-induced biofilms. (A) Wrinkled colony phenotypes of wild-type or rscS deletion strains expressing either vector control (VC) or pSypF* (pCLD29). Strains were spotted on agar plates and grown for 38 h. (B) WT or rscS deletion strains of V. fischeri expressing either VC or pSypF* were incubated with newly hatched squid. CFU of V. fischeri per squid was assessed after 18 h. Apo: aposymbiotic. Black bars represent the average CFU per squid. Limit of detection: 14 CFU.
expressing SypF* colonized the squid to similar CFUs as wild-type cells (Fig 18B). Combined these data suggest that SypF* is important for biofilms both \textit{in vitro} and \textit{in vivo}, and that it likely functions after RscS to control these phenotypes.

2. Wild-type SypF: unusual role in biofilms

Introduction

Wrinkled colony formation can be induced by two SKs: RscS and SypF*. The previous section focused on SypF*-induced phenotypes, which determined that SypF* functions as an SK to directly phosphorylate SypE and SypG to induce biofilms both \textit{in vitro} and \textit{in vivo}. One potential caveat to these sets of experiments, however, was that I interpreted the phenotype of an active variant of SypF, SypF*. The reason I utilized SypF* was because it exhibited a phenotype; its overexpression could induce wrinkled colonies, whereas overexpression of wild-type SypF had no observable phenotype. To determine whether wild-type SypF has any importance in biofilm formation, I studied its requirement for biofilms induced by the other Syp regulator, RscS, and unearthed rather unexpected results.

Requirement for SypF in RscS-induced biofilms

To determine whether wild-type SypF served a role in Syp biofilms induced by other regulators, I asked whether SypF was important for RscS-induced phenotypes. To do this, I deleted \textit{sypF} from the chromosome and assessed whether this affected the ability of RscS to induce wrinkled colony development. Whereas RscS induced the formation of wrinkled colonies by the wild-type strain, it failed to do so in the \textit{sypF} mutant, which formed smooth colonies indistinguishable from the vector control (Fig 19A). Complementation of the \textit{sypF} deletion with a wild-type copy of \textit{sypF} in single
Figure 19. Role of SypF in RscS-induced phenotypes. RscS-induced (pARM7) (A) biofilm formation and (B) syp transcription were assessed by overexpressing RscS (pARM7) in wild-type (WT) and ΔsypF strains, or in the ΔsypF strain containing various sypF alleles in single copy as indicated. (A) Cultures of the indicated strains were spotted on an agar plate and wrinkled colony formation was assessed after 39 h. (B) P_{sypA}-lacZ reporter activity in the indicated strains was assessed after 24 h of growth.
copy restored wrinkled colony formation. These data suggest that \textit{sypF} is required for RscS to promote wrinkled colonies, and that \textit{sypF} most likely works below RscS in the regulatory hierarchy.

Because RscS-induced wrinkled colony formation required \textit{sypF}, I asked whether RscS-induced \textit{syp} transcription would similarly require \textit{sypF}. To do this, I generated a \(P_{\text{sypA}}\)-\textit{lacZ} reporter that could be expressed from single copy in the chromosome, and then I evaluated the impact of a \textit{sypF} deletion on the activity of this reporter. In wild-type background, RscS induced expression of the \(P_{\text{sypA}}\)-\textit{lacZ} reporter relative to the vector control. In the \textit{sypF} deletion background, however, RscS failed to induce the reporter (Fig 19B). Finally, provision of the wild-type \textit{sypF} allele in single copy complemented the defect. Based on these results, I concluded that RscS requires \textit{sypF} to induce \textit{syp} transcription and wrinkled colony formation, and proposed a model wherein SypF functions downstream of RscS in the Syp TCS pathway (Fig 12).

\textbf{Function of SypF relative to SypG and SypE}

Above, I demonstrated that SypF* directly regulated both SypE and SypG to promote wrinkled colony formation, a hypothesis first introduced by Darnell \textit{et al.} (Darnell \textit{et al}, 2008) To ask whether this was similarly the case under RscS-inducing conditions, I asked where SypF functioned relative to SypE and SypG when RscS was overexpressed. Because RscS required \textit{sypF} to promote \textit{syp} transcription, I first asked if SypF functioned above SypG, the direct transcriptional activator of the \textit{syp} locus (Ray \textit{et al}, 2013). If so, then it should be possible to bypass the requirement for \textit{sypF} by expressing an active SypG variant that no longer required activation by an SK. Therefore, I generated strains that expressed SypG*, an active SypG protein in which the conserved
site of phosphorylation (aspartate 53) was converted to a glutamate (Fig 13). This mutation mimics the phosphorylated state of other RRs (Freeman & Bassler, 1999a; Sanders et al, 1992; Sanders et al, 1989) and has been shown to increase the activity of SypG (Hussa et al, 2008). Indeed, single-copy expression of SypG* was sufficient to induce syp transcription in the wild-type background and in the absence of sypF (Fig 20A). These data support a model in which SypF functions between RscS and SypG to control syp transcription (Fig 12).

To determine if SypF also functioned upstream of SypE, I evaluated whether expressing SypG* in the sypF deletion strain restored RscS-induced wrinkled colony formation. I predicted two possible outcomes: 1. biofilm formation, which would imply that SypF only controlled SypG or 2. no biofilm formation, which would suggest that SypF controlled another regulator, such as SypE. As controls, I evaluated the production of wrinkled colonies by sypF+ cells expressing SypG*. As predicted from previous work (Hussa et al, 2008), single copy expression of SypG* in this otherwise wild-type background failed to induce wrinkled colony formation due to inhibition by SypE; however, expression of both RscS and SypG* in wild-type cells induced wrinkled colony formation (Fig 20B). This demonstrates that, in wild-type cells, RscS expression is sufficient to turn off the inhibitory activity of SypE. In contrast, expression of both RscS and SypG* in the sypF mutant failed to induce this phenotype (Fig 20B). This observation suggests that sypF had an additional role in promoting biofilms, potentially by inactivating SypE. Indeed, a sypE sypF double mutant formed wrinkled colonies with RscS and SypG* expression (Fig 20B). I infer from these data that RscS works through SypF to control the activities of both SypG and SypE (See model in Fig 12B).
Figure 20. Determining where SypF functions in the Syp pathway. (A) SypG*-induced $P_{sypA}$-lacZ reporter activity in WT or $\Delta sypF$ strains after 24 h. Error bars represent standard deviation. (B) Wrinkled colonies of WT *V. fischeri* strains expressing RscS from a plasmid (pARM7) or vector control (VC, pKV282) and/or SypG* in single copy from the chromosome and compared to $\Delta sypF$ and $\Delta sypE \Delta sypF$ strains expressing both RscS and SypG*. Cultures were spotted and wrinkled colony morphology was assessed after 19 h. (C) pSypG* expression with wild-type SypA or Syp$^{AS56A}$ in $\Delta sypA$ or $\Delta sypA \Delta sypF$ strains. Wrinkled colony formation was assessed after 20 h.
To confirm that SypF functioned above SypE during RscS-inducing conditions, I asked whether the direct target of SypE’s regulatory activity, SypA, similarly functioned downstream of SypF. SypA is a small STAS domain protein (Fig 13). When unphosphorylated by SypE, SypA promotes biofilm formation (Morris & Visick, 2013b). Thus, I asked whether a mutation within the conserved serine (S56A), which prevents SypA from being phosphorylated, could suppress the biofilm defect of a \( sypF \) deletion. To do this, I generated a \( sypF \) deletion strain that expressed both SypG* (to induce \( syp \) transcription) and SypA\(^{S56A}\). In wild-type cells, expression of both of these regulators allowed for wrinkled colony formation, as expected (Fig 20C). Similarly, expression of these regulators induced biofilms in a \( sypF \) deletion strain. Altogether, this evidence suggests that SypF functions above key Syp regulators: SypG, SypE, and the target of SypE, SypA.

**Role for both RscS and SypF in biofilms**

The above evidence indicated that RscS functions through SypF to control the activity of SypG and SypE. This is an unusual regulatory set-up for TCS systems; thus, the mechanism by which SypF functioned after RscS to control biofilms remained unclear. Specifically, I wondered if wild-type SypF exhibited SK activity like SypF* and, if so, if that SK activity was necessary for RscS-dependent activation of the pathway. To ask the first question, I purified the cytoplasmic portion of wild-type SypF and assessed whether it could autophosphorylate \textit{in vitro}. Indeed, in the presence of radiolabeled ATP, SypF exhibited autokinase activity (Fig 15A).

To determine whether RscS-induced biofilm formation required SypF to function as an SK, I generated mutations in each predicted site of phosphorylation within wild-type
SypF, and then assessed whether the mutant proteins could complement the \textit{sypF} deletion strain for wrinkled colony formation. As shown above (Fig 19A), overexpression of RscS in the \textit{sypF} mutant failed to induce biofilm formation, but this defect could be restored with a wild-type copy of \textit{sypF} expressed in single copy from the chromosome.

Surprisingly, mutating the first conserved histidine (H250Q), the conserved aspartate (D549A), or both together (H250Q D549A), did not negatively impact complementation: strains with these alleles retained their ability to form wrinkled colonies (Fig 21A).

However, a SypF mutant disrupted for all three putative sites of phosphotransfer (H250, D549, and H705) failed to promote wrinkled colony formation, indicating that the last site of phosphotransfer may be required under these conditions. Indeed, SypF$^{H705Q}$, which contained a single substitution in the conserved site of phosphorylation within the HPt domain, did not complement the \textit{sypF} deletion (Fig 21A). Analogous results were seen when assessing whether this mutant allele could complement a \textit{sypF} deletion for \textit{syp} transcription (Fig 19B). Finally, I observed similar steady-state levels for epitope-tagged versions of the wild-type and mutant SypF proteins via western blotting (Fig 21B). Thus, the negative results for SypF$^{H705Q}$ and the triple mutant were not due to gross protein instability. Together, these data indicate that SypF does not function as a canonical hybrid SK under these conditions. Instead, it appears to require only H705 within the HPt domain.

Because RscS-induced biofilm formation and \textit{syp} transcription only required H705 in SypF, but not H250 or D549, I wondered whether the domain that contains H705, the HPt domain, was sufficient to promote these phenotypes. Indeed, SypF in other \textit{Vibrio} species often exists as a single HPt domain rather than a full-length SK (Fig 22A).
Figure 21. Phenotypes of SypF phosphotransfer mutants. (A) RscS-induced (pARM7) wrinkled colony phenotype in wild-type cells, or in sypF deletions strains expressing the indicated sypF alleles in single-copy from the chromosome. Indicated cultures were spotted on agar plates and wrinkled colonies were assessed after 39 h. (B) Anti-FLAG western blot analysis of SypF and SypF mutant proteins expressed in selected strains from (A).
Figure 22. Function of single HPt domain of SypF. (A) Cartoon image of the *sypF* gene (red) and the surrounding DNA sequences from other *Vibrio* species. Beige arrows represent ORFS that have replaced genes found within the *syp* locus in *V. fischeri*. The blue line indicates the 5’ portion of *sypF* that is missing in the indicated genomes.

*Aliivibrio salmonicida* LFI1238 (VSAL-II0307) (Holland *et al.*, 1997), *V. nigruplicultritaudo* (VIBNI_A1485) (Goudenenge *et al.*, 2013), (Vibrio sp. EJY3 (VEJY3_08720) (Roh *et al.*, 2012), *V. vulnificus* YJ016 (VV1628) (Chen *et al.*, 2003), *V. splendidus* LGP32 (VS_1526), *V. campbellii* ATCC BAA-1116 (VIBHAR_02229) (Wang *et al.*, 2013),Vibrio sp. Ex25 (VEA_003532), *V. parahaemolyticus* RMID 2210644 (VP1472) (Makino *et al.*, 2003), and *V. alginolyticus* NBRC 15630 = ATCC 17749 (VAL01S_15_00550). (B) Rscs-induced wrinkled colony phenotype of *sypF* deletion strains expressing the indicated alleles of *sypF*. Strains were spotted on an agar plate and wrinkled colonies were assessed after 39 h.
I thus cloned this domain and assessed complementation. I found that the HPt protein alone permitted RscS-induced biofilm formation (Fig 22B) and *syp* transcription in a *sypF* deletion mutant (Fig 19B). In contrast, expression of the HPt domain with a mutation in the site of phosphorylation did not complement the *sypF* deletion (Fig 22A and Fig 19B). These data suggested that the HPt domain in SypF is the sole domain to engage in phosphotransfer reactions that controls biofilm formation induced by RscS.

The requirement for only the HPt domain of wild-type SypF was surprising because single domain HPt proteins do not have enzymatic activity. Therefore, they must receive a phosphoryl group from an upstream protein to donate phosphoryl groups to downstream RRs. Interestingly, previous data suggested that RscS, a hybrid SK with three predicted sites of phosphorylation, did not require the last site of phosphorylation in its HPt domain to promote biofilms (Geszvain & Visick, 2008a). Thus, I hypothesized that RscS donates phosphoryl groups to the HPt domain of SypF, which then passes phosphoryl groups to the two downstream RRs, SypG and SypE (Fig 12B). To test this hypothesis, I generated a chimeric protein that contained the N-terminal portion of RscS (lacking its HPt domain) and the C-terminal HPt domain of SypF. I introduced the plasmid expressing this chimera into wild-type and *sypF* deletion backgrounds, and then assessed whether the chimeric protein was sufficient to induce biofilms even in the absence of *sypF*. In accordance with my hypothesis, the chimera induced wrinkled colonies in both backgrounds (Fig 23). Together, these data suggest that neither RscS nor SypF require the full complement of their own phosphotransfer domains, but instead rely on each other for the signal transduction that leads to biofilm formation.
Figure 23. Determining whether RscS and SypF interact to promote biofilms.

Wrinkled colony formation of WT or sypF deletion cells expressing RscS (pARM7) or the RscS-SypF<sup>HPt</sup> chimera (pANN69). Indicated strains were spotted and grown for 22 hours.
Requirement for SypF during host colonization

Our ability to assess the function of SypF in culture depends on the plasmid-based expression of regulators such as RscS and SypF*. Use of those two different regulators, however, yielded conflicting results about how SypF regulates biofilms. More specifically, SypF* required all three sites of phosphorylation to induce biofilms, whereas RscS-induced biofilms only required a single conserved site of phosphorylation within the HPt domain of wild-type SypF. I thus wanted to define a clear role for SypF and its putative enzymatic domains during biofilm formation using a more physiologically relevant approach. To do this, I assayed the importance of sypF and its conserved sites of phosphorylation for V. fischeri to colonize its squid host. Importantly, colonization is an in vivo phenotype that requires biofilm formation, but does not rely on the overexpression of regulatory proteins.

I first assessed the requirement of sypF for this phenotype by incubating the sypF deletion mutant with aposymbiotic squid for 18 h and then determining the number of V. fischeri cells in each squid. As expected, wild-type V. fischeri could colonize; however, the sypF mutant exhibited a severe colonization defect that could be complemented by providing wild-type sypF in single copy in trans (Fig 24). This evidence indicated that sypF is required for host colonization.

I next identified the domains/amino acids within SypF that were important for host colonization. I found that, similar to the RscS-induced biofilm experiments, cells that expressed SypF^{H250Q} or SypF^{D549A} successfully colonized the squid whereas cells expressing SypF^{H705Q} did not (Fig 24). Additionally, expression of the HPt domain of SypF alone allowed V. fischeri to colonize E. scolopes unless the HPt domain contained a
Figure 24. Mechanism by which SypF functions *in vivo*. Indicated strains of *V. fischeri* were incubated with aposymbiotic (Apo) juvenile squid for 18 h, and colonization of *E. scolopes* was determined by calculating colony-forming units (CFU) of *V. fischeri* in each squid. (Limit of detection, CFU = 14).
mutation within the site of phosphorylation (Fig 24). These results indicated that SypF does not function as a SK to promote colonization, and that the RscS-induced biofilm formation phenotypes are more physiologically relevant than the SypF*-induced phenotypes.

Finally, to confirm my findings that RscS and SypF function in an unusual phosphorelay to promote biofilm formation, I asked whether the RscS-SypF chimera, expressed from the chromosome of a double \textit{rscS sypF} mutant, was proficient to promote colonization. Because \textit{rscS} and \textit{sypF} are individually required for colonization (Visick & Skoufos, 2001) (Fig 24), it was not surprising that the \textit{rscS sypF} mutant failed to colonize the squid, and introducing either \textit{rscS} or \textit{sypF} alone into this strain did not restore host colonization (Fig 25). However, in support of our model for RscS-controlled biofilm regulation, the chimeric allele mostly complemented the \textit{rscS sypF} mutant for colonization (Fig 25). Together, these data confirm that the HPt domain of SypF functions between RscS and SypG/SypE to control biofilms, and that the enzymatic activity of SypF is largely dispensable for this signaling cascade during host colonization.

\textbf{3. SypF as a negative regulator}

For colonization experiments described above, I cloned the genes that encode RscS and the RscS-SypF chimera to express them in single copy expression from the chromosome of \textit{V. fischeri}. Although I initially tested the roles of these single-copy expression constructs in host colonization, I also assessed whether they could promote wrinkled colony formation similarly to their overexpression counterparts. Indeed, when expressed in single copy in wild-type cells, both the chimera and RscS could induce wrinkled colony formation, albeit not as robustly as when they are overexpressed (data...
Figure 25. RscS-SypF interaction *in vivo*. Juvenile aposymbiotic (Apo) *E. scolopes* were incubated with indicated *V. fischeri* strains for 18 h. Each squid was homogenized and the homogenate was plated to assess the colony forming units (CFU) of *V. fischeri* per squid. 14 CFU = limit of detection.
not shown); biofilms only formed after ~2 days, and they did not exhibit a robust wrinkled colony phenotype.

Because I used a double rscS sypF mutant strain to assess whether the chimera or RscS could promote host colonization, I also assessed whether single-copy expression of these regulators induced wrinkled colonies in this strain background. In the rscS sypF mutant, single-copy rscS expression did not allow this strain to form biofilms, supporting the requirement for sypF in RscS-induced wrinkled colonies (Fig 26A). Surprisingly, expression of the chimera in an rscS sypF double mutant allowed for rapid and robust wrinkled colony formation (Fig 26A). To test whether it was the loss of rscS or the loss of sypF that accounted for this enhanced biofilm development, I assessed the wrinkled colony phenotype of the chimera when it was expressed in either a sypF or an rscS single deletion strain. I found that loss of sypF, but not rscS, allowed for single-copy chimera expression to induce robust biofilms, and re-introduction of sypF into this strain restored the weak wrinkled colony phenotype (Fig 26B). These data suggest that the presence of SypF inhibited chimera-induced biofilms.

I next asked which domains of SypF were responsible for inhibiting biofilms induced by single-copy expression of the chimera. Because hybrid SKs can utilize sites of phosphorylation to function as a phosphatase, I assessed whether these residues in SypF were important for biofilm inhibition. To test this, I introduced SypF phosphotransfer mutants into the sypF deletion strain expressing the chimera from the chromosome, and then I assessed wrinkled colony formation. I found that both SypF\textsuperscript{H250Q} and SypF\textsuperscript{D549A} prohibited chimera-induced biofilm formation, suggesting that these residues are not important for the negative activity of SypF (Fig 26). SypF\textsuperscript{H705Q,D712-766} (containing both a
Figure 26. Biofilms produced from single copy expression of RscS or the chimera.

Indicated cultures were spotted on agar plates and wrinkled colonies assessed after (A) 44 h or (B) 48 h. (A) Single copy expression of RscS or the chimera in a double \( rscS \) \( sypF \) deletion strain or in single \( rscS \) or \( sypF \) deletion strains. (B) Wrinkled colony phenotypes of strains with \( sypF \) alleles introduced into a \( sypF \) deletion strain expressing the chimera in single copy. Plasmids are as indicated: pSypF (pANN11), pSypF\(^{H250Q}\) (pANN15), pSypF\(^{D549A}\) (pANN12), pSypF\(^{H705Q\ D712-766}\) (pANN16).
H705Q mutation and a truncation beginning 7 amino acids downstream of this residue) also inhibited these biofilms, although the inhibitory phenotype was less severe than the SypF\(^{H250Q}\) SypF\(^{D549A}\) constructs: wrinkled colonies eventually formed after 48 hours of growth (Fig 26B). These observations suggest that SypF does not utilize its autokinase or REC domains to inhibit chimera biofilms, but the HPt domain (assuming SypF\(^{H705Q.\Delta712-766}\) is expressed at levels similar to wild-type SypF) may promote reverse-phosphotransfer reactions within \textit{V. fischeri}.
CHAPTER FOUR

DISCUSSION

Introduction

The goal of my dissertation was to advance our knowledge of the complex regulatory mechanisms that *V. fischeri* utilizes to ensure a mono-specific association with its natural host, *E. scolopes*. I focused on a key event that occurs during the initiation of this symbiosis: aggregate formation on the surface of the symbiosis-specific light organ that *V. fischeri* ultimately colonizes. For over a decade, it was known that the intricate Syp TCS pathway, consisting of two hybrid SKs and two RRs, regulated not only this aggregation phenotype, but also the ability of *V. fischeri* to form wrinkled colonies (an indicator of biofilm formation) in the laboratory. Based on the integral role of these regulators for *in vitro* and *in vivo* biofilms, I decided to further elucidate how these proteins functioned together to control these exciting and complex phenotypes. Through the completion of this dissertation, I successfully ascertained the function of these regulators *in vitro*, in the context of wrinkled colony formation, and in the context of host colonization. Through these methods, I was able to determine that this complicated TCS cascade controls biofilms through a unique mechanism never before seen in any previously studied TCS pathway. Below, I will briefly summarize my findings, discuss how my results demonstrate the novel mechanism by which the Syp pathway functions, and speculate on how it may have evolved to allow *V. fischeri* to expand its niche to a new environment, the squid host. Where appropriate, I will suggest additional
experiments that can further our understanding of how the Syp pathway controls biofilms, and extrapolate on what the results mean in a broader context that extends outside of the squid-Vibrio symbiosis model.

**SypF: surprising results**

Much work has focused on how RscS and the two downstream RRs, SypE and SypG, control biofilm formation (Geszvain & Visick, 2008a; Geszvain & Visick, 2008b; Hussa et al, 2008; Hussa et al, 2007; Mandel et al, 2009; Morris et al, 2011; Morris & Visick, 2013a; Morris & Visick, 2013b; Shibata & Visick, 2012; Visick & Skoufos, 2001; Yip et al, 2006; Yip et al, 2005). However, previous evidence indicated that another protein, SypF, also regulated this phenotype due to the conspicuous location of its gene within the syp locus, and the generation of an active variant of SypF, SypF*, that could induce wrinkled colony formation on an agar plate (Darnell et al, 2008). Based on these observations, I hypothesized that SypF functions as an SK within the Syp TCS pathway to control the activity of these RRs. Indeed, my genetic and/or biochemical studies demonstrated that SypF can autophosphorylate, and that it functions directly above both SypG and SypE, thus confirming its role in the Syp regulatory cascade. Complicating these results, however, was the irrefutable evidence that another hybrid SK, RscS, also controls biofilms, an uncommon arrangement for TCS cascades.

I initially proposed that SypF and RscS acted as separate inputs into the downstream RRs, a mechanism described in the Vibrio harveyi luminescence (Lux) and Bacillus subtilis sporulation pathways (Henke & Bassler, 2004; Jiang et al, 2000a). This hypothesis was based on the observation that in culture, overexpressing either RscS or the
mutant allele of SypF, SypF*, induced wrinkled colony formation (Darnell et al., 2008; Yip et al., 2006), and that both SypF* (Fig 15) and RscS (Geszvain & Visick, 2008a) required sites of autophosphorylation to induce this phenotype. However, although SypF* could function as a hybrid SK in the cell, this activity seemed not to be physiologically relevant. In particular, only the single, non-enzymatic HPt domain of wild-type SypF was required to promote host colonization, an in vivo phenotype that does not require the artificial overexpression of regulators. This result was corroborated by the observation that RscS-induced biofilms required the HPt domain of SypF, but did not require the HPt domain of RscS (Geszvain & Visick, 2008a). Combined with my data that an RscS-SypF chimera is sufficient to promote colonization, I concluded that (1) SypF does not function as an SK under physiologically relevant biofilm-promoting conditions, (2) SypF* activity is not physiologically relevant, and (3) SypF functions downstream of RscS and thus RscS and SypF do not provide separate inputs into the Syp pathway. I propose a mechanism in which RscS bypasses its own HPt domain and preferentially hijacks the HPt domain of SypF to affect the activity of the downstream RRs, SypE and SypG, to control biofilms (Fig 12B).

**SypF* vs SypF**

I hypothesize that SypF*, but not wild-type SypF, functions as a hybrid SK within the cell under conditions used in the laboratory. This is supported by the observations that 1. overexpression of SypF*, but not wild-type SypF induced biofilm formation, 2. SypF*, but not SypF, required all conserved sites of phosphorylation to promote biofilms, and 3. SypF* bypassed the requirement for the enzymatic activity of RscS in host colonization. The mechanism that explains the difference between SypF* and SypF activities,
however, remains unresolved. One clue lies in the observation that introducing the S247F mutation into wild-type SypF generated the SypF* phenotype. Thus, understanding how SypF* functions requires an understanding of how this mutation alters the structure of SypF.

Serine 247 is located within the DHp domain, and it is three amino acids away from the site of phosphorylation (H250). Because the DHp domain controls both the homodimerization and kinase activity of hybrid SKs, I reason that the S247F mutation could influence either of these functions. I would predict that S247F may not affect dimerization, because the mutation is located towards the top of a DHp α-helix, whereas the region that determines homodimerization specificity is normally located at the bottom of the DHp structure near the loop that connects the two α helices within the DHp domain (Ashenberg et al, 2011). Thus, S247F most likely affects the autokinase activity of SypF.

As described in the introduction of this document, SK activity requires that the ATP hydrolyzing domain (CA) and the DHp domain are properly aligned to permit phosphorylation of the conserved histidine. This event is often instigated upon signal recognition by the SK, which leads to an alteration of the location or stability of the α-helices within the DHp structure, which is then propagated to the CA domain (Dago et al, 2012; Diensthuber et al, 2013; Rivera-Cancel et al, 2014). This conformational change properly orients the CA and/or DHp domains to permit kinase activity. Based on the location of S247F within an α-helix of the DHp domain of SypF, I hypothesize that this mutation could place the CA, DHp, or both domains in the orientation that promotes
phosphorylation of the conserved histidine. Changing a small serine side chain for a bulky phenylalanine side chain could, for example, position the histidine side chain in a location that readily permits phosphorylation by the CA domain. Alternatively, perhaps this mutation propagates a conformational change to the CA domain, placing it in the kinase “on” location. If the bulkiness of the phenylalanine side chain generates an active SypF protein, then perhaps exchanging the serine for another bulky amino acid, such as tryptophan or tyrosine, would be sufficient to generate the SypF* phenotype. Furthermore, one could imagine that other amino acid substitutions within the DHp or CA domains could sufficiently reorient the CA/DHp interaction to generate a protein with kinase activity similar to or even greater than SypF*. This could be tested by randomly mutagenizing the CA and DHp domains of SypF, and then assessing whether these full-length SypF mutants are sufficient to induce wrinkled colony formation. To irrevocably determine the structural mechanism that generates the differences between SypF and SypF*, however, requires structural studies comparing the orientation of the CA and DHp in both SypF and SypF*.

One unresolved observation about the role of SypF in V. fischeri is that it autophosphorylates in vitro, yet only the nonenzymatic HPt domain is required in vivo. However, if SypF can function as kinase and/or phosphatase, then these catalytic activities should affect the phosphorylation state of the HPt domain, and thus cause a dysregulation within the Syp TCS pathway. It is unlikely that the HPt domain is “shielded” from the enzymatic activities of the CA/DHp and REC domains due to their close proximity within the same protein. Additionally SypF* utilizes all three of these domains to control biofilms, suggesting that these regions are functional and can undergo
intramolecular phosphotransfer. One explanation for this conundrum is that these enzymatic activities are too weak to exert any effect within the cell. Secondly, the enzymatic regions within SypF could play a role in fine-tuning the Syp phosphorelay. Although this mechanism is not apparent in my *in vitro* experiments, the observation that expression of the HPt domain of SypF alone does not permit wild-type levels of colonization suggests a role for these N-terminal regions *in vivo* (see “Evolution of SypF and its Proposed Function”). A third explanation is that the enzymatic activities of SypF within *V. fischeri* (ES114) are inhibited.

There are two general mechanisms to inhibit SypF SK catalysis: either this function depends on domains within the SypF protein, or on molecules outside the SypF protein (or both). For example, in well-studied SKs, the domains located N-terminally to the CA/DHp regions control the net kinase to phosphatase activity. Within SypF, these domains include the putative signal-binding region located within the periplasm, the TM regions, and the HAMP domain within the cytosol. Intriguingly, these regulatory domains were not included in the purified SypF protein that was used for *in vitro* studies, suggesting that perhaps the loss of these regions may relieve the inhibition of SypF’s kinase activity. Furthermore, when SypF homologues from *Vibrio* species are aligned, the periplasmic region exhibits more sequence divergence as compared to domains containing sites of phosphorylation (data not shown). Thus, it is possible that the periplasmic region of SypF within the strain of *V. fischeri* used in this study (ES114) has acquired amino acid mutations that preclude SypF from functioning as a kinase and phosphatase, perhaps by preventing the binding of an activating signal or through preventing the disassociation of an inhibitory signal. Similarly, amino acid substitutions
within the TM or HAMP domains of SypF could hinder SypF’s enzymatic activity, most likely through inhibiting signal transfer from the signal-binding region to the CA/DHp domains. If mutations in SypF (ES114) prevent this enzyme’s activity, then overproducing SypF homologues in the ES114 strain could potentially induce wrinkled colony formation, assuming these homologues do not contain mutations that inhibit catalytic activity. Subsequently, one could generate chimeras of SypF (ES114) with homologues that exhibit kinase activity to identify amino acids that inhibit or promote kinase function. It is possible, however, that none of these homologues will induce biofilm formation if the cell tightly regulates SypF’s enzymatic activity to ensure signaling fidelity. Although SypF* can induce wrinkled colonies when overexpressed, this phenotype requires both SypG and, surprisingly, an unlinked RR, VpsR (Darnell et al, 2008). Perhaps the mutation within SypF* generates a constitutive kinase that interacts with VpsR in a non-physiologically relevant manner. Thus, SypF variants, which are mostly likely not constitutive, may not induce biofilms unless under the appropriate conditions which are generally not provided under standard laboratory conditions.

A noteworthy stipulation in stating that mutations within SypF ES114 inhibit its enzymatic activity is that these alterations must not affect the functionality of the HPt domain. If true, this would imply that mechanisms that control the upstream enzymatic regions of SypF do not sufficiently regulate the HPt domain. This has the surprising consequence of permitting HPt domains within hybrid SKs to have additional roles that are independent of the enzymatic function(s) of the SK. Perhaps HPt proteins execute
unappreciated regulatory functions outside of being passive intermediates between active TCS regulatory enzymes.

Alternatively, factors may function *in trans* to regulate the enzymatic activity of SypF. This is true in *E. coli*, where the RcsF (Castanie-Cornet *et al.*, 2006) and IgaA (Cano *et al.*, 2002) proteins function as accessory regulators of the Rcs TCS pathway. In *V. fischeri*, similar factors could function within the Syp pathway to inhibit either the CA/DHp or REC activity of SypF. Presumably, inhibiting the REC domain would be the simplest solution to prevent both the kinase and phosphatase activity of an SK. This would represent a unique finding, as no precedence exists for a protein to regulate the REC domain of a hybrid SK.

If there is an accessory factor that hinders the kinase and phosphatase activities of SypF, then SypF* must have the capability to overcome this inhibition. It is easy to imagine that the S247F mutation within the DHp domain may prevent an auxiliary protein from binding in this region. However, it is more difficult to imagine how the S247F mutation would prevent an accessory factor from inhibiting the distally located REC domain (which contains phosphatase activity). As of yet, auxiliary proteins that function within the Syp pathway have not been found using genetic screens. This may not be surprising, as auxiliary proteins are generally small, and thus less likely to be targeted by mutagenesis methods. Perhaps RNA-seq or microarray analysis of biofilm-inhibiting or biofilm-promoting conditions would uncover a differentially expressed RNA that encodes a small regulator of SypF’s enzymatic function.

**Evolution of SypF and its proposed function**
What is unprecedented about the Syp pathway is that wild-type SypF apparently relies on the enzymatic activity of a different SK as a source of its phosphoryl group in vivo. This result is especially surprising considering the evidence that SypF exhibits autokinase activity in vitro. Similarly, the Eps pathway in Myxococcus xanthus contains a hybrid SK, EpsC, that exhibits SK activity in vitro, but does not require residues involved in autophosphorylation in vivo (Jagadeesan et al, 2009). In vitro evidence suggested that another hybrid SK, EpsA, could weakly phosphorylate the REC domain of EpsC, but whether this mechanism occurs in vivo remains to be determined. Together, SypF and EpsC contradict the accepted assumption that an enzymatically-competent SK must function as so in vivo. Furthermore, the fact that SypF instead uses the enzymatic activity of RscS is a unique result. I propose that this may be a mechanism more common than is currently appreciated; there are other examples of SKs that do not require all sites of phosphorylation to promote a phenotype [e.g., (Chand et al, 2011; Laskowski & Kazmierczak, 2006)], but it remains to be tested whether they have histidine kinase activity or whether an interacting partner exists to supply phosphoryl groups.

If V. fischeri does not require SypF to function as an SK to promote biofilms, then why is full-length sypF maintained in the genome? This question is especially perplexing given the observation that the syp locus in other species of Vibrio often encodes SypF as a single HPt domain (Fig 22). One explanation is that, in V. fischeri, sypF is fated toward degeneracy, but the 5’ sequences have not had sufficient time to be lost through neutral or negative selection. This hypothesis could potentially be addressed by analyzing the sypF sequence after passing V. fischeri multiple times through the squid. Presumably, if there is no positive selection for the enzymatic activity of SypF, then these N-terminal regions
of the protein may be lost after serial passaging. If this is \textit{syp}\textsubscript{F}’s fate, then the Syp TCS would end up similar to the Rcs pathway in \textit{E. coli}, where the hybrid SK, RcsC, donates phosphoryl groups to the HPt domain in a degenerate SK, RcsD (Takeda \textit{et al}, 2001) (Fig 11). If it were possible to peer into the evolutionary past of the \textit{E. coli}, perhaps RcsC and RcsD would function analogously to RscS and SypF, wherein RcsD, like SypF, would exhibit autokinase activity. The Syp pathway could represent an early evolutionary branch point that occurs after a TCS pathway acquires a SK, but before selective pressures have directed the exact fate of SKs within the cascade.

Secondly, many SKs exhibit both kinase and phosphatase activity (Aiba \textit{et al}, 1989; Casino \textit{et al}, 2009; Freeman \textit{et al}, 2000; Huynh \textit{et al}, 2010; Yang & Inouye, 1993), so SypF could utilize its N-terminal enzymatic domains to remove phosphoryl groups from downstream regulators, thus turning off biofilm formation and resetting the TCS system. However, this is not supported by my data studying the putative phosphatase activity of SypF; similar to wild-type SypF, SypF variants with mutations in two key enzymatic domains (Syp\textsuperscript{F\textsubscript{H250Q}} and Syp\textsuperscript{F\textsubscript{D549A}}) inhibited biofilms induced by single copy expression of the RscS-SypF chimera (Fig 26). The D549A result is particularly surprising, considering this conserved residue in other hybrid SKs drives reverse signal decay in TCS systems. Furthermore, expression of Syp\textsuperscript{F\textsubscript{H705Q-trunk}} also hindered these biofilms, but it did so at a lesser level; however, the relative stability of this protein remains unknown. If it is produced at levels similar to wild-type SypF, then these data suggest that the HPt domain of SypF can remove phosphoryl groups from SypE and SypG, or that it can prevent the HPt domain within the chimera from donating phosphate groups to these RRs. Regardless of the mechanism, these results further support the hypothesis that,
within the cell, the HPt domain is the critical region of SypF that controls biofilms both positively, and perhaps negatively. Whether or not SypF or the HPt domain can directly remove phosphoryl groups from SypE/SypG, or whether the HPt domain can directly interact with the chimera, remains unknown. These questions could be answered using biochemical experiments, such as assessing whether purified RscS can donate phosphoryl groups to the HPt domain, and whether the HPt domain can remove phosphoryl groups from SypE and SypG. Furthermore, if the RscS and the HPt domain of SypF interact, it would be possible to genetically determine which residues within the HPt domain are important for this interaction. This could be done by mutating residues within the HPt domain of SypF, and then assessing whether these mutations relieve the inhibition of biofilms induced by the RscS-SypF chimera. Secondary structure prediction of SypF predicts the HPt domain to be a four-helix bundle. In other, characterized HPt domains, residues around the site of phosphorylation are important for maintaining specificity between TCS regulators, so these residues in SypF would likely be important for the ability of RscS to interact with SypF. Interestingly, the HPt domain of SypF contains an additional N-terminal α-helix that was required for wrinkled colony formation (an initial HPt expression construct did not contain this α-helix, and it did not promote biofilm formation. Data not shown). Thus, the role of this helix in interacting with RscS could be assessed using this genetic approach. Many HPt domains have auxiliary secondary structures in addition to the four-helix bundle; however, the potential role of these regions remains unknown. Thus, characterizing which amino acids within the HPt domain of SypF are important for interacting with RscS may uncover general mechanisms that TCS proteins use to maintain specific interactions within a particular pathway.
Lastly, conditions found in later stages of colonization or outside of squid colonization could require that SypF utilize its hybrid SK domains. *V. fischeri* is a marine organism that can proliferate within a large repertoire of environments, such as ocean sediment and in association with a number of aquatic animals besides *E. scolopes* (Haygood, 1993; Lee & Ruby, 1992; Mandel et al, 2009; Ortigosa et al, 1994; Ramesh et al, 1989; Ruby & Lee, 1998; Ruby & Nealson, 1976; Yetinson & Shilo, 1979). Perhaps in these other contexts SypF functions as a bona fide SK to induce formation of the Syp or a Syp-like biofilm. To address whether particular environments may require SypF to function as an SK, one could compare the *sypF* sequence among metagenomes from different environmental reservoirs, with the expectation that distinct settings may negatively or positively select for particular domains within SypF. Whether or not both RscS and SypF, or just one of these SKs, must utilize SK activity to promote biofilms in diverse environments remains unknown, but this TCS arrangement raises the intriguing possibility that domains within the same signaling network could exert discrete roles depending on environmental conditions surrounding the cell.

**Evolution of the Syp TCS pathway**

Continued research into TCS systems has unveiled an increasing number of unusual TCS architectures [e.g., (Goodman et al, 2009; He et al, 2013; Kong et al, 2013; Schramm et al, 2012)]; however, the environmental pressures that selected for any of these TCS arrangements remain unknown. Conversely, *V. fischeri* has given researchers some clues as to how the complex Syp pathway may have evolved. In *V. fischeri*, there are at least two genetic loci required for *in vivo* biofilms: the *syp* locus, and *rscS*. The *syp* locus is found in all *V. fischeri* strains and other *Vibrio* species, such as *Vibrio vulnificus*...
and *Vibrio parahaemolyticus*. However, *rscS* is not well conserved amongst *Vibrios*; it is only found in *V. fischeri*, but not all strains of *V. fischeri*. *rscS* was proposed to be horizontally acquired, partly due to its seemingly random location between glycerol metabolism genes (Mandel *et al.*, 2009). Furthermore, when producing a phylogenetic tree of *V. fischeri* strains from different environmental reservoirs, it was possible to pinpoint a progenitor within the *V. fischeri* lineage that acquired *rscS*. Interestingly, the *V. fischeri* strains that derived from this progenitor gained the ability to colonize *Euprymna* squid (Mandel *et al.*, 2009). This suggested that the acquisition of RscS eventually granted *V. fischeri* access to the light organ of *E. scolopes*. Perhaps RscS acquired the ability to recognize a squid signal, and then functionally replaced the enzymatic activity of SypF. Rather than gain the ability to directly interact with both SypE and SypG, perhaps the more straightforward evolutionary trajectory generated the single RscS-SypF interaction, while maintaining the SypF-SypE/SypG interactions. Additionally, the HPt domain of SypF could also be positively selected for to provide an additional regulatory checkpoint. If only a small number of environments require the Syp biofilm, then it seems reasonable that this intricate TCS arrangement evolved to prevent inappropriate activation of a complex developmental process.

It may be possible to track the evolutionary trajectory of SypF and RscS. Thus far, many *V. fischeri* strains have been isolated from various environmental and animal reservoirs and organized according to phylogeny (Mandel *et al.*, 2009). Additionally, it was determined which strains contained the *syp* locus and/or *rscS* (Mandel *et al.*, 2009). Additional *rscS* and *sypF* genes (if present) could be sequenced and their amino acid residues aligned according to the phylogenetic relationship of the *V. fischeri* strain from
which they were isolated. Through this method, one could roughly trace the evolution of the sypF sequence both before and after V. fischeri acquired rscS, and potentially identify regions of these genes that exhibit either positive or negative selection. Using sequence information gained from this study, one could also identify amino acids within SypF and RscS that have coevolved (Burger & van Nimwegen, 2008; Weigt et al, 2009). This would permit the identification of residues in both these proteins that are critical for a direct SypF-RscS interaction. Upon identification, the requirement for these co-evolved amino acids could be characterized both in vitro and in vivo to further describe the mechanism of this unusual interaction between two SKs.

Conclusion

My dissertation work has elucidated the unusual mechanism by which the Syp TCS pathway functions in V. fischeri to control biofilm formation and host colonization. I have uncovered that the SK, SypF, exhibits an atypical role in Syp, and have therefore unveiled a previously unknown mechanism that may be applicable to SKs within other TCS pathways. The observation that two unlinked SKs within Syp have seemingly co-evolved to exert control over this cascade also reveals the plasticity of TCS pathways. This flexibility demonstrates that nature, rather than reinvent the wheel, reuses existing factors to more readily permit an organism to adapt to a new environment. Thus, the Syp pathway represents an exceptional model to understand how TCS can evolve to permit bacteria to successfully exist in almost any environment presented to them.

Significance

All domains of life contain organisms that encode TCS regulators. In bacteria, these pathways have become the dominant means to connect extracellular signals to an
intracellular response. This expansion has allowed bacteria to control a variety of different phenotypes, such as biofilm formation, and conquer a gamut of distinct environments, such as host tissues; in fact, bacteria that colonize a larger diversity of ecological niches often encode more TCS pathways (Alm et al, 2006; Galperin, 2005). A characteristic of these pathways that explains their prevalence is the fact that their architectures exhibit plasticity. This flexibility in the arrangement of TCS pathways allows all domains of life to precisely regulate their physiology to manage a vast repertoire of environments. The unique architecture of Syp, for example, has allowed V. fischeri to expand its niche to include the light organ of E. scolopes, thus outcompeting all other bacterial strains found in the local environment. My work with the Syp pathway demonstrates not only the plasticity of TCS pathways, but also provides a potential model for how a bacterium may adapt to conquer new environments and guarantee proliferation of its progeny.

This dissertation has unveiled a unique mechanism by which a symbiont utilizes a TCS signaling pathway to initiate colonization of a host. This remarkably complex TCS pathway, known as Syp, controls whether V. fischeri forms a biofilm on epithelial cells of a host organ, a process required for colonization. The Syp pathway is unique, in that it contains two SKs and two RRs, including the SK, SypF, which could seemingly provide a variety of different functions depending on environmental conditions. Although the function of this TCS system is thus far unique, perhaps similar TCS pathways exist outside of V. fischeri. Perhaps they, too, are involved in forming biofilms on the epithelial cells that line host organs, a cell type in animals that bacteria notoriously utilize as a site of colonization (McFall-Ngai & Ruby, 1991). Thus, determining how V. fischeri
employed this unique TCS pathway to form a biofilm and colonize a host can reveal mechanisms that other bacteria use for similar purposes. The exciting findings unveiled within this study could therefore be extrapolated to guide research involving more complex pathogenic or symbiotic systems, such as the beneficial or detrimental interactions humans exhibit with microbes.

Due to their prevalence in bacterial genomes, and their absence in mammals, TCS pathways are an ideal target for new antimicrobials developed to combat the rise of deadly bacterial strains that exhibit antibiotic resistance (Blackledge et al., 2013; Gotoh et al., 2010; Watanabe et al., 2008; Worthington et al., 2013). Progress with these pharmaceuticals can be accelerated through the discoveries in basic science that elucidate how TCS regulators function on a molecular level, such as the research presented in this dissertation. Presumably, broad-range antimicrobials would target conserved enzymatic regions within an RR or an SK, allowing for the inhibition of a wider range of TCS regulators in different pathogenic bacteria. Antimicrobials that potentially execute the most damage could target the REC domain, considering this domain is found in both hybrid SKs and RRs. In the Syp pathway, this antibiotic could hinder three key regulators: RscS, SypE, and SypG, since they all have REC domains required for colonization. However, the Hpt domain in TCS pathways could also serve as a site of antimicrobial recognition, especially the cytoplasmic HPt proteins that represent degenerate DHp/CA domains. This antibiotic would have the potential of targeting both single domain HPt proteins and the DHp/CA domain, again, killing two birds with one stone.
APPENDIX

ADDITIONAL STUDIES
1. Elements that control \textit{sypG} expression

SypG is a critical regulator of Syp biofilms and host colonization (Hussa \textit{et al}, 2008; Hussa \textit{et al}, 2007; Yip \textit{et al}, 2005). It is the direct transcriptional activator of the \textit{syp} locus, which encodes proteins that produce the polysaccharide component of the biofilm matrix and also regulatory proteins (Ray \textit{et al}, 2013; Yip \textit{et al}, 2006; Yip \textit{et al}, 2005). SypG is one of the regulators encoded within \textit{syp}. Although it may seem intuitive to contain \textit{sypG} within the locus that it controls, this genetic arrangement raises the question of how the SypG protein is produced if this protein is required for its own transcription. Interestingly, the open reading frames of \textit{sypG} and the downstream gene, \textit{sypH} (which is also required for biofilm formation), overlap, which suggests that they are transcriptionally and translationally coupled. Therefore, the expression of these two key biofilm regulators is likely controlled via similar mechanisms. One hypothesis to explain the conundrum surrounding \textit{sypG-sypH} expression is that these genes are controlled by SypG-independent regulatory elements.

To test whether or not SypG controls \textit{sypG-sypH} expression, I asked whether SypG induced the production of the SypH protein. Overexpression of SypG from a plasmid (pSypG) is sufficient to induce \textit{syp} transcription (Hussa \textit{et al}, 2008; Yip \textit{et al}, 2005); thus, I asked whether pSypG could induce \textit{sypH} transcription and thus increase SypH protein levels, expecting that it would not if this gene was not within the SypG regulon. To assess SypH protein levels, I introduced a Campbell mutation into the \textit{sypH} gene. This insertional mutation results in a merodiploid with a truncated copy of \textit{sypH} and a full-length copy that encodes a FLAG tagged SypH protein. This Campbell mutation is not predicted to exhibit polar effects on the downstream genes, as these genes
are controlled by a promoter distinct from \( P_{sypA} \); indeed, biofilm formation in this background was not affected (results not shown). I next introduced pSypG into the SypH-FLAG expressing strain and assessed SypH protein levels (as an indicator of both SypG and SypH production). As a control, I assessed the protein levels of SypO-FLAG, a protein that is predicted to be induced by pSypG; indeed, I found that overexpression of SypG increased SypO-FLAG levels over that produced by the vector control (Fig 27A). Conversely, pSypG did not alter the expression of SypH-FLAG, suggesting that the expression of \( sypH \) and \( (sypG) \) are controlled by regulatory mechanisms distinct from SypG overexpression (Fig 27A). Furthermore, overexpression of RscS, the SK that also induces \( syp \) transcription (Yip et al., 2006), induced expression of SypO-FLAG, but did not alter the relative amount of the SypH protein (Fig 27A), confirming that levels of \( syp \) transcription do not correlate with SypH or SypG protein levels.

To further test whether known Syp regulatory mechanisms similarly control \( sypG \)-\( sypH \) expression, I reasoned that if I delete a key genetic element that controls \( syp \) expression, the SypG and SypH proteins might still be produced. One key region that controls expression of the \( syp \) locus is the \( sypA \) promoter, which is presumed to control expression of the \( sypA \)-\( sypH \) genes. This promoter contains a SypG enhancer site (SE) required for expression of these \( syp \) genes. Thus, I generated a \( \Delta SE \) strain that expressed SypH-FLAG (as described above), and asked whether the loss of SE affected SypH protein levels. I found that SypH protein was produced regardless of the presence of the SE site upstream of \( sypA \), although the protein levels were reduced as compared wild-type cells, (Fig 27B). This suggests that \( sypG \)-\( sypH \) expression may be partially regulated by this promoter region, but an additional regulatory element must still be present to
Figure 27. Levels of SypHA under different genetic conditions. (A, B, C) Anti-HA (hemagglutinin) western blot analysis of cell lysate from indicated V. fischeri strains. (A) Levels of SypHA compared to levels of SypOHA under pRscS (pCLD46), pSypG (pARM9), or vector control (VC, pVSV105) inducing conditions. (B) Levels of SypHA with or without the sypA enhancer sequence (SE). B.1 and B.2 indicate independent replicates. (C) Levels of SypHA in wild-type, ΔsypE ΔsypF, or ΔsypF ΔsypG strains. C.1 and C.2 represent independent replicates.
promote expression of these genes.

The evidence described thus far demonstrates SypH protein levels are regulated via a mechanism that is distinct from those that control the expression levels of other Syp proteins. I reasoned that, if an unidentified genetic element controls sypG-sypH expression, then it would most likely exist upstream of these genes. Thus, I deleted regions upstream of SypH, and assessed SypH protein levels using the same approach described above. As seen in Fig 27C, removal of sypF-sypG sequences did not diminish the levels of SypH. However, deleting sypE-sypF sequences reduced the steady state level of SypH. This suggested that a regulatory element may exist either within sypE or within the intergenic region between sypE and sypF. Alternatively, deletion of this region could affect the stability of the sypG-sypH mRNA, which would similarly result in a decrease in SypH levels. To determine which of these hypotheses may be correct, q-RT-PCR (quantitative reverse-transcription PCR) could be utilized to determine whether or not mRNA levels change with the deletion of sypE-sypF.

These preliminary results demonstrate that SypG expression is controlled through a mechanism that is distinct from those that control other Syp regulators. These pilot observations provide strong rationale for performing additional, more quantitative experiments to better elucidate the mechanism by which SypG is produced. Further experiments could include identifying the transcriptional start site(s) for sypG-sypH, which could allow for the discovery of a promoter. The function and relative contribution of this newly identified promoter could then be assessed both during wrinkled colony formation and squid colonization. Additionally, it remains to be determined whether sypG and sypH are truly transcriptionally and translationally coupled, a question whose
answer would further our understanding of Syp regulation. Importantly, the identification of how SypG expression is controlled may uncover a unique regulatory mechanism that could expand the microbiologist’s view of the numerous ways a bacterium can manage the expression of any gene located within an organism’s DNA.

2. Role for the HPt domain of RscS

My dissertation revealed that the HPt domain of RscS was not critical for biofilm formation; instead, RscS utilized the HPt domain of another SK, SypF, for this phenotype. However, there exists some evidence to contradict the conclusion that this domain within RscS is completely dispensable. This is partly based on the observation that a mutation at the site of phosphorylation within the HPt domain of RscS causes a slight delay in the ability of RscS to induce biofilms, suggesting that this domain has a function (Fig 28) (Geszvain & Visick, 2008a). Additionally, I found that overexpression of RscS in a sypF deletion strain (which normally does not form wrinkled colonies) could eventually induce biofilms under certain conditions, albeit these biofilms were delayed and not robust (Fig 28). Combined, I reasoned that, perhaps under certain conditions, the HPt domain of RscS could be utilized to induce biofilms rather than the HPt domain of SypF.

To test whether Syp biofilms require that RscS utilize its HPt domain, I asked whether the site of phosphorylation (H867) (Fig 13) within the HPt domain of RscS was required for wrinkled colonies to form in a sypF deletion strain. To do this, I introduced a plasmid that expressed RscS\textsuperscript{H867Q} into a sypF deletion mutant, and assessed its ability to induce wrinkled colonies. As was previously published, wild-type cells containing this
Figure 28. **Role for the HPt domain of RscS in biofilms.** Time course of wrinkled colony formation induced by pRscS (pKG63) or pRscS$^{H867Q}$ (pKG108) in wild-type or ΔsypF deletion strains.
plasmid exhibited a slight delay in biofilm formation as compared to a plasmid that contains non-mutated RscS. In a sypF deletion wild-type RscS induced weak wrinkled colonies at a late time point (41.5 hrs) (Fig 28). Conversely, expression of RscS$^{H867Q}$ failed to induce these biofilms. This suggests that the HPt domain of RscS can replace the requirement for SypF during wrinkled colony formation.

Although RscS can induce biofilms in a sypF deletion, this phenotype is most likely not physiologically relevant, as these biofilms are severely delayed and noticeably weak. Whether or not this domain in RscS is important for squid colonization remains unknown. Although these results may be an artifact of RscS overexpression, it suggests the possibility that RscS has the ability to interact with the downstream RRs, SypE and SypG, a role that is currently assigned to SypF. To test whether RscS can donate phosphoryl groups to SypE and SypG, an in vitro assay could be performed similar to the method used to assess whether SypF* could donate phosphates to these RRs.
REFERENCES


Castillo MG, Goodson MS, McFall-Ngai M (2009) Identification and molecular characterization of a complement C3 molecule in a lophotrochozoan, the Hawaiian bobtail squid *Euprymna scolopes*. *Dev Comp Immunol* **33**: 69-76


Cheung J, Hendrickson WA (2009) Structural analysis of ligand stimulation of the histidine kinase NarX. Structure 17: 190-201


He K, Marden JN, Quardokus EM, Bauer CE (2013) Phosphate flow between hybrid histidine kinases CheA(3) and CheS(3) controls Rhodospirillum centenum cyst formation. *PLoS Genet* 9: e1004002


Holland LZ, McFall-Ngai M, Somero GN (1997) Evolution of lactate dehydrogenase-A homologs of barracuda fishes (genus *Sphyraena*) from different thermal environments: differences in kinetic properties and thermal stability are due to amino acid substitutions outside the active site. *Biochemistry* **36**: 3207-3215


Huang ZJ, Edery I, Rosbash M (1993) PAS is a dimerization domain common to Drosophila period and several transcription factors. *Nature* **364**: 259-262


Kong W, Chen L, Zhao J, Shen T, Surette MG, Shen L, Duan K (2013) Hybrid sensor kinase PA1611 in *Pseudomonas aeruginosa* regulates transitions between acute and chronic infection through direct interaction with RetS. *Mol Microbiol* **88**: 784-797


McDonald LR, Boyer JA, Lee AL (2012) Segmental motions, not a two-state concerted switch, underlie allostery in CheY. *Structure* 20: 1363-1373


Montgomery MK, McFall-Ngai M (1994) Bacterial symbionts induce host organ morphogenesis during early postembryonic development of the squid *Euprymna scolopes*. *Development* 120: 1719-1729

Moore JO, Hendrickson WA (2009) Structural analysis of sensor domains from the TMAO-responsive histidine kinase receptor TorS. *Structure* **17**: 1195-1204


Ray VA, Eddy JL, Hussa EA, Misale M, Visick KL (2013) The *syp* enhancer sequence plays a key role in transcriptional activation by the σ54-dependent response regulator


Ruby EG, Nealson KH (1976) Symbiotic association of *Photobacterium fischeri* with the marine luminous fish *Monocentris japonica*; a model of symbiosis based on bacterial studies. *Biol Bull* **151**: 574-586


VITA

The author, Allison Nichole Norsworthy was born in Lily Lake, Illinois, to Perry and Peggy Norsworthy. In 2009, she received her Bachelors of Science degree from North Park University in Chicago, where she obtained her first lab experience studying cytochrome production in *Bacillus subtilis*. 

In 2009, Allison was accepted to the Integrated Program in Biomedical Sciences graduate program at Loyola University, where she researched the symbiotic relationship between the bioluminescent bacterium, *Vibrio fischeri*, and its squid host, *Euprymna scolopes*. After receiving her PhD, she will be pursuing postdoctoral studies in the lab of Melanie Pearson at New York University, New York.