Molecular Investigations of Protein Assemblies Involved in Prokaryotic Virulence

Jordan M. Mancl

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of:

Doctor of Philosophy

in

Biological Sciences

July 30, 2018

Blacksburg, VA

Committee Chair: Florian D. Schubot, PhD
Committee Member: Birgit Scharf, PhD
Committee Member: Roderick Jensen, PhD
Committee Member: Zhaomin Yang, PhD

Keywords: Protein Assemblies, Virulence Regulation, Structural Biology

Copyright: Jordan Mancl, 2018
Protein complexes mediate a diverse range of behavior in prokaryotic cells, yet the exact molecular mechanisms explaining how many of these complexes assemble and function remain unknown. This work focuses on understanding the molecular mechanisms of two different protein assemblies responsible for regulating virulence in the opportunistic pathogen *Pseudomonas aeruginosa*. *P. aeruginosa* utilizes type IV pili (T4P) to adhere to, and move along, surfaces. Assembly of T4P is powered by a dedicated cytoplasmic ATPase, PilB. The structural study of PilB from a related system (chapter 2) resulted in the formulation of the first model describing the mechanism of force generation resulting from ATP hydrolysis, which explains how T4P are assembled. Chapter 3 focuses on the RetS/GacS interaction, which is responsible for globally regulating virulence in *P. aeruginosa*. A comprehensive structural study reveals a dynamics of a novel regulatory interaction and the discovery of a potentially universal transmembrane signaling mechanism.
Molecular Investigations of Protein Assemblies Involved in Prokaryotic Virulence

Jordan M. Mancl

GENERAL AUDIENCE ABSTRACT

Bacteria have threatened human health since the beginning of recorded history. With the development of antibiotics in the early twentieth century, the threat posed by bacterial infection was greatly lessened. However, decades of antibiotic mismanagement has led to the evolution of bacteria which are no longer vulnerable to these antibiotics. In order to combat this rising threat of resistant bacteria, we require a deeper understanding of how bacteria function and cause disease. Proteins play a crucial role in the diseases caused by bacteria, either by directly damaging host cells or regulating the expression of these damaging factors. By increasing our knowledge of the roles played by protein during bacterial infections, it will be possible to create new antibiotics while minimizing the risk of resistance. The work presented here grants a deeper understanding into how proteins work together to allow bacteria to survive inside the human body.
ACKNOWLEDGEMENTS

I would like to thank my advisor, Florian Schubot, for giving me the skills and knowledge necessary to be a successful independent scientist. You pushed me to do a lot of things that I really did not want to do, and it forced me to grow from someone who had no idea what Structural Biology was when I joined the lab to someone with the desire and confidence to pursue a career as a Structural Biologist.

I would like to thank my committee members, Birgit Scharf, Roderick Jensen, and Zhaomin Yang, for their assistance through my graduate career. I feel like your strengths complemented each other very well and allowed me to grow into a well-rounded scientist.

I would like to also thank my current and former lab members, Xing Jing, Cory Bernhards, Yi Xiao, Manisha Shrestha, and Kylie Ryan, for their assistance. Whether you know it or not, you taught me many valuable lessons about what I should, and sometimes what I should not be doing, in order to be successful.

I would like to thank my fellow members of the Microbiology program for their candid discussions and comradery, and letting me borrow stuff that Florian forgot to order. Don’t let Happy Hour die.

Lastly, I would like to thank my friends and family members. Especially my wife, Katrena, for dealing with my unpredictable schedule, my periodic misery when things weren’t going well, and for working three jobs to help support me though graduate school.
# TABLE OF CONTENTS

## CHAPTER 1: LITERARY REVIEW

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>TRANSCRIPTION</td>
<td>2</td>
</tr>
<tr>
<td>- Promoter structure</td>
<td>2</td>
</tr>
<tr>
<td>- Sigma factors</td>
<td>3</td>
</tr>
<tr>
<td>- Transcription initiation</td>
<td>4</td>
</tr>
<tr>
<td>- Transcription activation</td>
<td>5</td>
</tr>
<tr>
<td>- Transcription repression</td>
<td>6</td>
</tr>
<tr>
<td>TRANSLATION</td>
<td>7</td>
</tr>
<tr>
<td>MEMBRANE ASSOCIATED PROTEIN ASSEMBLIES</td>
<td>8</td>
</tr>
<tr>
<td>- Flagella</td>
<td>9</td>
</tr>
<tr>
<td>- Type IV pili</td>
<td>10</td>
</tr>
<tr>
<td>- Secretion systems</td>
<td>11</td>
</tr>
<tr>
<td>PSEUDOMONAS AERUGINOSA AS A PATHOGEN</td>
<td>13</td>
</tr>
<tr>
<td>- Type III secretion system</td>
<td>14</td>
</tr>
<tr>
<td>- Biofilm formation</td>
<td>15</td>
</tr>
<tr>
<td>REGULATION OF VIRULENCE IN PSEUDOMONAS AERUGINOSA</td>
<td>16</td>
</tr>
<tr>
<td>- Two-component systems</td>
<td>16</td>
</tr>
<tr>
<td>- cAMP signaling</td>
<td>17</td>
</tr>
<tr>
<td>- C-di-GMP signaling</td>
<td>18</td>
</tr>
<tr>
<td>- Quorum sensing</td>
<td>19</td>
</tr>
<tr>
<td>- The Gac/Rsm pathway</td>
<td>21</td>
</tr>
<tr>
<td>- The RetS/GacS interaction</td>
<td>21</td>
</tr>
<tr>
<td>- GacS</td>
<td>22</td>
</tr>
<tr>
<td>- RetS</td>
<td>23</td>
</tr>
<tr>
<td>GOALS OF THE STUDY</td>
<td>25</td>
</tr>
<tr>
<td>FIGURES</td>
<td>26</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>27</td>
</tr>
</tbody>
</table>
CHAPTER 2: CRYSTAL STRUCTURE OF A TYPE IV PILUS ASSEMBLY ATPASE: INSIGHTS INTO THE MOLECULAR MECHANISM OF PILB FROM THERMUS THERMOPHILUS

SUMMARY ......................................................................................................................... 36

INTRODUCTION .................................................................................................................. 37

RESULTS AND DISCUSSION ............................................................................................... 40

Structure of hexameric $Tt$PilB$_{ATP}$ at 2.65 Å resolution .................................................. 40

The tertiary structure of TtPilB$_{ATP}$ protomers closely resemble those of T2SS ATPases ........................................................................................................................................ 41

Conformational changes in a flexible inter-domain linker separate TtPilB$_{ATP}$ into open and closed forms .................................................................................................................. 42

Protomers in the closed conformation are poised for ATP hydrolysis .................................. 43

An arginine conserved in PilB/GspE enzymes but not PilT, Flal or AfGspE is critical for coordinating the γ-phosphate of the substrate in TtPilB$_{ATP}$ ........................................................................... 44

Open protomers display varying degrees of solvent accessibility to their active sites ................................................................................................................................. 46

TtPilB$_{ATP}$ employs a symmetric counterclockwise rotary mechanism of ATP hydrolysis ................................................................................................................................. 47

Force generation in TtPilB$_{ATP}$ ........................................................................................ 49

CONCLUDING REMARKS .................................................................................................. 52

EXPERIMENTAL PROCEDURES ....................................................................................... 53

Strain Construction ............................................................................................................. 53

Protein Purification .............................................................................................................. 53

Crystallization and Structure Solution ................................................................................. 54

Nucleotide binding assays .................................................................................................. 55

FIGURES .............................................................................................................................. 56

SUPPLEMENTAL FIGURES ............................................................................................... 64

AUTHOR CONTRIBUTIONS ............................................................................................... 68

ACKNOWLEDGEMENTS ...................................................................................................... 69

REFERENCES ...................................................................................................................... 69

CHAPTER 3: A HELIX CRACKING MECHANISM REGULATES INTERACTION BETWEEN THE SENSOR HISTIDINE KINASE GACS AND ITS NEGATIVE REGULATOR RETS IN PSEUDOMONAS AERUGINOSA

ABSTRACT .......................................................................................................................... 74
INTRODUCTION ................................................................................................................. 75

RESULTS .......................................................................................................................... 78

The DHp domains of RetS and GacS interact without breaking the GacS homodimer ............................................. 78

The crystal structure of the RetS kinase region reveals a uniquely asymmetric dimer interface and loss of the canonical ATP binding pocket .......................................................... 80

RetS_{HK}α1 residues 416-423 are required for inhibiting GacS signaling .............................................. 82

DISCUSSION ......................................................................................................................... 84

ONLINE METHODS ............................................................................................................ 88

Cloning .................................................................................................................................. 88

Protein Purification .............................................................................................................. 89

FRET Assays ....................................................................................................................... 91

Crystallization and Structure Solution ..................................................................................... 91

Bacterial Adenylate Cyclase Two-Hybrid Assay ......................................................................... 93

Crystal Violet Biofilm Assay ................................................................................................... 93

Quantification of RetS construct levels ...................................................................................... 94

FIGURES ............................................................................................................................... 96

SUPPLEMENTAL FIGURES ................................................................................................ 102

REFERENCES ....................................................................................................................... 113

CHAPTER 4: FINAL DISCUSSION

FINAL DISCUSSION .............................................................................................................. 117

REFERENCES ....................................................................................................................... 119

APPENDIX A: EFFECTS OF REMOVING THE RETS PERIPLASMIC SENSORY DOMAIN ON THE PSEUDOMONAS AERUGINOSA PAO1 TRANSCRIPTOME

INTRODUCTION ..................................................................................................................... 121

MATERIALS AND METHODS ............................................................................................... 122

RESULTS AND DISCUSSION ............................................................................................... 122

FIGURES ............................................................................................................................... 124

REFERENCES ....................................................................................................................... 125

APPENDIX B: NARX/GACS CHIMERA ALLOW FOR ARTIFICIAL CONTROL OF THE GAC/RSM PATHWAY

INTRODUCTION ..................................................................................................................... 126
Chapter 2 has been published as an article in the November 2016 issue of the journal *Structure*. *Structure* does not require that authors obtain permission to have their work reproduced in dissertations. Author contributions are as follows: J.M.M. conducted all the research and wrote the manuscript. W.B. developed the expression and purification protocols for the PilB protein. H.R. collected diffraction data. Z.Y. and F.D.S. supervised the research, assisted with data analysis, and wrote the manuscript.

As of ETD submission, chapter 3 has not been published. Author contributions are as follows: J.M.M. designed and conducted the research, analyzed data, and wrote the manuscript. W.K.R. performed the LC-MS work. R.F.H. supervised the LC-MS work. F.D.S designed and supervised the research, assisted with data analysis and wrote the manuscript.
INTRODUCTION

It can be argued that proteins are responsible for life as we know it. Proteins are responsible for crucial processes ranging from DNA replication to cellular division, inter-cell communication, and motility. The proteins involved in these processes often assemble into multimeric complexes. Whether homo- or heteromeric, these assemblies allow for additional functionality beyond what is possible for monomeric proteins. The evolution of multimeric protein assemblies is as fascinating as it is complex, and as such, remains one of the greatest mysteries in evolutionary biology. While the evolutionary history remains foggy, the effects that these protein assemblies exert on their current environments clearly embody the critical issues of modern day biology.

Traditionally, pathogenic bacteria have represented one of the greatest threats to human health. However, their danger was greatly blunted with the introduction of antibiotics in the first half of the 1900s. Unfortunately, “life, uh, finds a way,”¹ and soon after the introduction of new antibiotics, bacteria have developed resistance. The rate at which bacteria evolve resistance to antibiotics severely outstrips the rate at which new antibiotics are being developed, pushing society closer to a future where modern medicine is no longer able to treat bacterial infections. In fact, this future seems to be uncomfortably close, as cases have already been reported of bacterial infections where the pathogen demonstrates resistance to all available antibiotics². The CDC estimates that 2 million people in the United States suffer from bacterial infections every year, however, due to the historical success of antibiotics, only around 23,000 patients die from these infections. As the effectiveness of traditional antibiotics begins to fail in the face of increased resistance, it is easy to envision the mortality rate of bacterial infections skyrocketing. For this reason, combating the rise of antibiotic-resistant bacteria requires new approaches and is seen by many as the greatest challenge facing the medical world today³.
BACTERIAL TRANSCRIPTION

Before protein assemblies can be discussed, one must understand how proteins are made. The first step in creating a protein is transcription: the process by which DNA is copied, or transcribed, into messenger (mRNA) using the enzyme RNA polymerase (RNAP). RNAP is a multi-subunit enzyme, with the core enzyme consisting of five subunits (αββ’ω) (Fig. 1). Binding of the σ subunit completes the RNAP holoenzyme and is required for initiation of promoter-specific transcription. The DNA to be copied consists of genes which, in bacteria, are commonly organized into operons consisting of multiple genes under the control of one promoter. These operons are transcribed as a single unit, resulting in a polycistronic mRNA transcript. As all genes in an operon are under the control of the same promoter, the structure of a promoter plays a key role in the regulation of transcription.

Promoter structure

Generally, a promoter is a sequence of DNA upstream from the transcription start site of a gene or operon. The DNA sequence of a promoter contains two important hexamers: the -10 site, which is centered 10 bases upstream of the transcription start site and has a conserved sequence of 5’-TATAAT-3’, and the -35 site, which is centered 35 bases upstream of the transcription start site and has a conserved sequence of 5’-TTGACA-3’^4. The -10 and -35 sites are recognized by the σ subunit of the RNAP holoenzyme and are the most critical components for determining promoter strength and function^5. In addition to the traditional -10 and -35 sites, a promoter may contain an UP element, an extended -10 site, or a discriminator region. An UP element is an A-T rich stretch of DNA upstream of the -35 site to which the C-terminal domains of the two α subunits of RNAP bind^6. The extended -10 site is a short tetramer of consensus sequence 5’-TGTG-3’ typically found between 17 and 14 bases upstream of the transcription start site which can be bound by the σ subunit of the RNAP holoenzyme^5. The discriminator region is a G-C rich sequence found between the -10 and transcription start sites, usually from -6 to -4, where the σ subunit also binds. Strong binding between the σ subunit and the
discriminator region occurs when there is a guanine in the -5 position, with the sequence 5′-GGG-3′ producing optimal binding . Although every promoter will not contain all elements, each element serves to increase the affinity of RNAP for the promoter, thereby leading to a stronger promoter.

**Sigma factors**

Although the core RNAP is capable of non-specific DNA binding and transcription in certain cases , a σ factor (subunit) is required for specific RNAP-promoter binding and initiation of transcription. The number of different σ factors present in a cell can vary wildly between organisms, yet all σ factors fall into one of four groups. Group 1 σ factors are considered housekeeping σ factors, and tend to bind promoters controlling genes associated with growth and maintaining the “day-to-day” functions of a cell. Group 1 σ factors contain four major domains. The σ1 domain possesses an auto-inhibitory function unique to Group 1 σ factors, which prevents the σ factor from binding to DNA when not also bound to RNAP . All other domains are responsible for binding elements of the promoter. The σ2 domain is responsible for binding the -10 and discriminator sites. The σ3 domain contains a large α-helix which is responsible for binding the extended -10 site , and the σ4 domain contains a canonical helix-turn-helix motif which is responsible for binding the -35 site . Sigma factors in Group 1 are the only σ factors to contain all four domains. Sigma factors in Group 2, which consists of σ factors closely related to Group 1 yet not required for growth, only lack σ1. Group 3 consists of σ factors related to specialized functions such as sporulation and stress response. Sigma factors in Group 3 possess all domains except σ1. Group 4, also called the extracytoplasmic function subfamily (ECF) is the largest and most varied category of σ factors, containing σ factors which respond to extracellular signals. Group 4 σ factors only contain domains 2 and 4 . The genome of *P. aeruginosa* PAO1 was found to encode 24 putative σ factors, 19 of which have been annotated as belonging to the ECF subfamily.
Transcription initiation

Regardless of which group a σ-factor belongs to, all σ factors have the same function: to recruit RNAP to specific promoters and initiate transcription. Once bound to RNAP, σ factors are able to recruit the complex to a promoter based on the affinity between the various domains of the σ factor and the elements of the promoter. Only the -35 and extended -10 sites of a promoter are used to initially guide the RNAP to a promoter as they are recognized by the σ subunit in their double-stranded DNA form. The -10 and discriminator sites are recognized and bound in their single stranded form after the double helix of the promoter has melted. The UP element plays no role in σ-directed binding to a promoter as it binds the C-terminal domain of the RNAP α subunit directly. Initial binding of the RNAP holoenzyme to the promoter results in the formation of what is called the closed complex. The closed complex (Fig. 1) represents a critical step in the initiation of transcription, as it is this closed complex which is catalytically capable of breaking the DNA double helix of the promoter and exposing the bases for transcript synthesis. The unzipping of the promoter double helix marks an important transition, referred to as isomerization, from the closed complex to the open complex, the structure of which allows for the production of RNA transcripts. Roughly 13 bp of DNA are unzipped during isomerization, from the -10 site to just downstream of the transcription start site, forming a transcription bubble.

Once a stable open complex has been formed, nucleotide triphosphates (NTPs) are recruited and the formation of an RNA transcript can begin. As the RNA chain grows longer, stress is created as the RNAP tries to unzip additional downstream DNA, yet remains anchored to the promoter by the σ subunit. This stress can be overcome in two ways. The first way the stress can be overcome is through abortive transcription, where the RNA transcript is released prematurely, after only a few nucleotides have been incorporated into the growing RNA chain. The second method of overcoming stress is by releasing the σ subunit from the RNAP. Once the RNA chain reaches a length of 11 bp, the forces favoring chain extension are able to
overcome the promoter-σ interactions, resulting in the release of the σ subunit and successful transcription of the target gene 8.

Transcription activation

The method of transcription initiation described above is referred to as activator-independent initiation, as it requires no exogenous factors beyond the components of the RNAP holoenzyme. However, many promoters are incapable of the high affinity RNAP binding required for activator-independent initiation, as their expression needs to be tightly controlled in response to changes in the cell’s environment. These promoters require assistance from outside factors, referred to as transcriptional activators. There are two methods of activator-dependent initiation that bacteria use: RNAP-centric activation, and promoter-centric activation. RNAP-centric activation involves a direct interaction between a transcriptional activator and the RNAP itself, independent of promoter structure. The simplest, and most common, form of RNAP-centric activation involves the replacement of the housekeeping σ factor with an alternative σ factor which displays a high affinity for a different set of promoters 15.

Promoter-centric activation involves direct interaction between the transcriptional activator and the promoter itself. When discussing promoter-centric activation, promoters are classified into two groups. Class 1 promoters contain activator binding sites upstream of the -35 site, while class 2 promoters contain activator sites that overlap the -35 site 16. Class 1 promoters and the activators that bind them are exemplified by the cyclic AMP receptor protein (CRP). CRP binds to a promoter roughly 61 bases upstream of the transcription start site and serves to stabilize promoter RNAP interaction by also binding the CTD of the α subunit 4,17. Class 1 activators are capable of activating transcription from a wide variety of sites upstream of the promoter due to the flexible unstructured linker joining the C- and N-terminal domains of the α subunit of RNAP. Although CRP binds to a DNA binding domain of RNAP, different activators are capable of stabilizing promoter-RNAP interaction by binding to different domains of RNAP 18. These activators that bind multiple RNAP domains are considered class 2 activators, as their binding
sites must overlap the -35 site of the promoter in order to contact multiple regions of RNAP. Although divided into two classes, activators are not bound by their designation, as they may act differently depending on the promoter they are binding activating. Case in point, CRP acts as a class 1 activator when activating the lac operon promoter, yet when activating the gal operon promoter, CRP acts as a class 2 activator. Regardless of where an activator binds to the DNA or where it binds to the RNAP, all activators serve to stabilize the interaction between a promoter and the RNAP holoenzyme, ultimately resulting in a level of transcription higher than what is possible when the activator is not present.

**Transcriptional repression**

Just as there exist proteins capable of activating transcription, there also exist proteins capable of deactivating, or repressing transcription. These proteins are referred to as transcriptional repressors and may function in a variety of ways, the simplest of which is directly binding to the DNA in a location overlapping the promoter sequence, resulting in inhibition of RNAP binding through steric hindrance. Repression via steric hindrance is exemplified by the binding of the LacI repressor to the operator region of the lac operon. Although LacI directly binds to a site overlapping the promoter region to inhibit RNAP binding, it is not always necessary for a repressor to directly occlude the binding site of RNAP. Some repressors bind upstream of the promoter sequence and inhibit transcription by disabling an activator or by creating a scaffold for additional protein binding events which can prevent RNAP binding to a promoter.

Whereas steric hindrance can prevent RNAP from initially binding a promoter, other methods of repression can prevent active transcription while still allowing stable binding between RNAP and a promoter. It is possible for a repressor to allow RNAP binding to a promoter but prevent melting of the promoter and transition from closed to open complex as is the case for MerR, which binds the merT promoter on the opposite side of the DNA helix from RNAP, effectively preventing the unwinding of the double helix.
As stated above, transcription can be activated via the binding of a σ factor which displays a high affinity for a certain set of promoters. Transcription of those promoters can therefore be repressed by decreasing the availability of that specific σ factor to the RNAP. It has been shown that the cellular concentration of σ factors are significantly greater than the cellular concentrations of RNAP\textsuperscript{24}. As there is a potentially large excess of free σ factors in the cell, multiple σ factors will end up competing for the same RNAP binding site. Since σ factors naturally possess varying degrees of affinity for binding to RNAP, bacteria were forced to develop a method of regulation which would allow σ factors with a relative low binding affinity to overcome competition by σ factors with a higher binding affinity for RNAP. That regulation comes in the form of anti-σ factors. Anti-σ factors are proteins which bind to specific σ factors and sequester them, preventing them from binding to RNAP and effectively repressing transcription from the promoters normally recognized by the sequestered σ factors.

Similar to repression via anti-σ factors, anti-activators can be used to repress transcription from promoters enhanced by specific activators. Anti-activators act in the same manner as anti-σ factors, that is, they bind directly to an activator and sequester it, preventing the activator from functioning. Repression via anti-σ factors and anti-activators (anti-proteins) can be overcome by anti-anti-σ factors and anti-anti-activators (anti-anti-proteins), respectively. Anti-anti-proteins function by binding an anti-protein, releasing the previously bound protein (activator or σ factor) in the process and relieving repression by repressing the repressor\textsuperscript{25-27}.

**BACTERIAL TRANSLATION**

Translation is the process by which the genetic information carried in a mRNA strand, the end product of transcription, is converted, or translated, into an amino acid backbone, ultimately forming a structured protein. In bacteria, translation is performed by a 70S ribonucleoprotein complex called the ribosome. The bacterial ribosome consists of two subunits, called 50S and 30S, referencing their sedimentation coefficients, and is only comprised of roughly 1/3 protein, while the remaining 2/3 of the ribosome is made up of ribosomal RNA (rRNA)\textsuperscript{28}. The 70S
ribosome contains three distinct binding sites for transfer RNA (tRNA), termed the aminoacyl (A), peptidyl (P), and exit (E) sites. To initiate translation, the 30S subunit, with the assistance of three initiation factors (IF1, IF2, IF3) binds to the Shine-Dalgarno sequence of the mRNA. IF1 and IF3 then dissociate while IF2 recruits the 50S ribosomal subunit. During this process, the initiator tRNA (fMet-tRNA\textsubscript{Met}) is positioned into the P site after which IF2 is released, allowing for sustained translation of the mRNA target\textsuperscript{29}. To be incorporated into the growing chain, additional amino acids must first enter the ribosome through the A site, a process facilitated by the pairing of the charged tRNA anticodon with the proper mRNA codon. Once the correct charged tRNA has entered the A site, it transitions to the P site, where the ribosome catalyzes the formation of a peptide bond between the new aminoacyl tRNA and the growing peptide chain. Following peptide bond formation, the ribosome moves forward to the next codon, while shifting the now empty tRNA to the E site, where it can be ejected from the ribosome. This process continues until the ribosome reaches an in-frame stop codon, at which point the peptide chain is released from the ribosome. The peptide chain can then fold into a mature, structured protein, either on its own, or with the assistance of chaperonins.

**MEMBRANE ASSOCIATED PROTEIN ASSEMBLIES**

For bacteria to survive in their respective environments, they must be able to sense and respond to changes in these environments, and sometimes alter their local environment to better suit their needs, either to extract nutrients or defend against competitors. One of the most basic survival mechanisms employed by bacteria is motility. Motile bacteria are able to leave harmful environments and seek out more favorable conditions, most commonly environments were nutrients are more plentiful. One of the most common, and best studied types of bacterial motility is swimming motility, a process wherein a bacterium utilizes a massive protein assembly, the flagellum to propel itself through an aqueous environment. While the ability to move into new, more favorable, environments is often a powerful boon to bacteria, the opposite is also true. Once a bacterium reaches a favorable environment, the ability to remain in that
environment, by adhering to a surface, prevents the cell from being forced into unfavorable conditions. Once again, a protein assembly plays a key role in this bacterial behavior. Type IV pili (T4P) are large protein complexes that span the entirety of the bacterial envelope and extend a fiber out into the environment, which allows the cell to adhere to surfaces. However, T4P are versatile assemblies; in addition to mediating surface attachment, T4P also allow bacteria to move along solid surfaces, via a process known as twitching motility. While the ability to find a prosperous environment is an essential survival tool, bacteria do not exist in a vacuum. Bacterial cells face tremendous competition for resources, whether from other bacteria, predatory eukaryotes, or a host’s immune system. Sometimes survival for bacteria means fighting back against these competitors, and this requires a bacterium to get its “weapons”, or virulence factors, into the environment where they can act upon external organisms without harming the bacteria producing them. Secretion of virulence factors, oftentimes proteins, is typically accomplished via large, dedicated, protein assemblies referred to as secretion systems.

Flagella

Many prokaryotic organisms utilize flagella to move through aqueous environments. The Gram negative flagellum is the most common and the best characterized example, therefore the subsequent discussion will focus on this type of flagellum. The bacterial flagellum is a massive macromolecular machine made up of more than thirty distinct structural proteins, which can measure up to a width of 60 nm and an astounding length of 10 µm, giving a single flagellum a molecular weight of approximately 1 billion Daltons. As with any massive cellular assemblage, the flagellum structure can be subdivided into distinct parts: a basal body, which contains the motor complex anchoring the flagellum into the bacterial membrane, the hook, and the filament. The basal body is the most complex of these domains and can be further subdivided into the MS ring, stator, rod, C ring, P ring, and L ring (Fig. 2). The MS ring is composed of a single protein, FliF, which forms the base for the flagellar basal body structure.
complex. The MS ring is anchored into the cytoplasmic membrane, allowing it to interact with the cytoplasmic C ring, which made up of three proteins, FliG, FliM, and FliN, and functions as a switch controlling rotation of the flagellum. The rod serves to connect the MS ring to the hook, passing through the P and L rings, and allows for the translocation of flagellar subunits. The P and L rings act as bushings, facilitating free rotation of the rod as it passes through the peptidoglycan and outer membrane, respectively. The final component of the basal body is the flagellar stator, which consists of two integral membrane proteins, MotA and MotB. These proteins utilize the proton motive force to cause conformational changes in the C ring which ultimately result in rotation of the flagellar filament\textsuperscript{34,35}. The flagellar filament is composed of a single protein, flagellin, and is assembled by dedicated protein export machinery homologous to the type III secretion system injectisome (see below)\textsuperscript{36,37}. The rotation of bacterial flagella is commonly regulated in response to chemical stimuli through a process called chemotaxis, which allows for a biased-random walk behavior of the bacterial cell\textsuperscript{38,39}.

**Type IV pili**

As mentioned above, T4P are versatile molecular machines capable of mediating cell attachment to surfaces as well as the subsequent movement along that surface. Similar to the flagellum, a single T4P assembly can be subdivided into its constituent parts: the alignment complex, the fiber, the platform, and the motor ATPase. Notation of the various pilus proteins varies by organism, so the most general annotations will be used from here on out. The alignment complex, formed by PilNOP, anchors the T4P machinery to the bacterial envelope, passing through the peptidoglycan, and interacting with the secretion pore, PilQ, at the outer membrane\textsuperscript{40}. The alignment complex is tethered to the inner membrane via interaction with PilM which, along with PilC, form the T4P platform, anchored into the inner membrane. Two cytoplasmic proteins, PilB and PilT, interact with the T4P platform and hydrolyze ATP to provide the energy necessary to extend and retract, respectively, the pilus fiber\textsuperscript{41-43}. Chapter 2 provides an in depth exploration into the mechanism by which ATP hydrolysis is coupled to pilus fiber
extension. Once assembled, the mature pilus fiber, primarily comprised of PilA subunits with varying degrees of minor pilin incorporation, extends out into the extracellular environment, where it can attach to a surface. Following surface attachment, the assembly ATPase, PilB, is replaced by the retraction ATPase, PilT, and the T4P rapidly disassembles the pilus fiber, an action that causes the bacterial cell to be pulled toward the point of attachment. The force generated during pilus retraction has been measured in excess of 100 pN, making the T4P complex the strongest biological motor currently known. By coordinating the retraction and extension activities of multiple T4P complexes, a bacterial cell is able to erratically move across solid surfaces in a process known as twitching motility.

Bacterial secretion systems

Many proteins are designed to function outside of a bacterial cell, whether they be virulence factors or metabolic enzymes degrading nutrient sources prior to cellular uptake. Bacterial cells possess dedicated protein assemblies, called secretion systems, to accomplish this task. The best studied secretion systems in Gram negative bacteria are the type II, III, V, and VI secretion systems. In addition to the secretion systems, bacteria utilize the general secretion (Sec) and twin arginine translocation (Tat) pathways to move proteins across or into the inner membrane. Both the Sec and Tat pathways are highly conserved, with organisms across all domains of life containing homologs. There are two ways proteins may be targeted to the Sec pathway in Gram negative bacteria, either post-translationally, with the assistance of the chaperonin SecB, or co-translationally, with the assistance of the signal recognition particle (SRP). Traditionally, the SRP pathway is the most common route for the insertion of proteins into the inner membrane. The Sec translocation machinery consists of a three protein complex made up of SecYEG, which forms a channel through the inner membrane. An adaptor ATPase, SecA, is also involved to provide the energy necessary to translocate proteins post-translationally, while translocation via the SRP pathway is powered via translation itself. Contrary to the Sec system, the Tat system translocates folded proteins across the inner membrane.
membrane utilizing a protein complex consisting of either two or three components and the energy from the proton motive force\textsuperscript{49}.

The type II secretion system (T2SS) is evolutionarily very similar to the type IV pilus machinery and represents the most general mechanism for protein secretion. T2SS substrates utilize the Sec system to cross the inner membrane into the periplasm, where they fully fold. The folded substrates are then pushed by a piston-like pseudopilus through a channel in the outer membrane out into the extracellular environment. Like the T4P machinery, the T2SS contains a dedicated inner membrane platform which interacts with a cytoplasmic ATPase responsible for powering secretion\textsuperscript{50}. The type III secretion system is unique, in that it secretes substrates from the bacterial cytoplasm, not into the extracellular environment, but directly into the cytoplasm of host cells, bypassing the host cell membrane completely. The type III secretion system represents one of the prominent virulence factors in \emph{Pseudomonas aeruginosa} and is discussed in more detail below. The type V secretion system (V) is unusual in that the substrates themselves are also the secretion system, leading to their designation as autotransporters. These autotransporters utilize the Sec system to cross the inner membrane. Once in the periplasm, autotransporters are able to spontaneously insert themselves into the outer membrane via specialized, pore-forming domains at their C-terminus. The N-terminus, referred to as the passenger domain, is then able to pass through the newly formed pore. These passenger domains are often cleaved, releasing them into the extracellular environment\textsuperscript{51}. The type VI secretion system (T6SS) is a relatively newly discovered system which utilizes a complex akin to the phage tail spike to inject cytotoxic effector proteins into the cytoplasm of eukaryotic and prokaryotic cells\textsuperscript{52}. The exact mechanism of this system is still being investigated, but it is known that the T6SS consists of two distinct complexes, and inner membrane complex, which serves to form a channel leading from the inner membrane to the outer membrane, and the tail spike which, in response to a currently unknown signal forcefully punches through the target membrane to deliver its payload directly to the target cytoplasm\textsuperscript{53}. 

\textsuperscript{12}
**PSEUDOMONAS AERUGINOSA AS A PATHOGEN**

*Pseudomonas aeruginosa* is a Gram-negative bacterium commonly found throughout a wide range of environments such as soil, plants, and humans\(^5\). The main reason that *P. aeruginosa* is able to survive in such varied environments is most likely due to the fact that *P. aeruginosa* devotes a significant amount of its genome to various regulatory elements that allow it to sense and respond to its environment in an appropriate way. For example, the genome of *P. aeruginosa* strain PAO1, the most commonly researched strain, consists of 6.3 million base pairs (Mbp). 9.4% of those base pairs have been annotated as coding for regulatory elements\(^5\). At the time its genome sequence was published, the 521 regulatory genes identified for *P. aeruginosa* was the largest proportion of regulatory genes ever found in a bacterial genome\(^5\).

As an opportunistic human pathogen, *P. aeruginosa* is able to cause disease in persons whose immune systems have previously been compromised, such as burn victims or patients with Cystic Fibrosis (CF)\(^6\). *P. aeruginosa* infections can be classified into two categories: acute and chronic. Acute infections can occur in any type of human tissue, and are characterized by the expression of the Type III Secretion System (T3SS), which functions as a molecular syringe capable of injecting cytotoxic effectors directly into the host cell’s cytoplasm\(^7\). Chronic infections are characterized by the formation of a biofilm, most commonly in the lungs of CF patients, which allows the bacteria to survive for extremely long periods of time while becoming functionally “immune” to the host’s immune system and antibiotic treatment. These long-term infections cause chronic inflammation and increase the likelihood of developing a more severe infection, which, should it occur, carries with it a very real chance of being fatal. *P. aeruginosa* has been identified as one of the main causes of mortality in patients with CF\(^8\), and since this disease is one of the most common hereditary genetic disorders, with an occurrence of roughly 1 in every 2,500 live births in the United States\(^9\), the importance of unraveling the mysteries of *P. aeruginosa* cannot be overstated.
Type III Secretion System

As previously mentioned, acute *P. aeruginosa* infections are characterized by the expression of the T3SS, which functions as a molecular syringe used by *P. aeruginosa* to inject toxic effector proteins directly into the cytoplasm of host cells. The T3SS is encoded by at least 42 different genes on the *P. aeruginosa* chromosome. The majority of these genes code for the structural components of the supramolecular needle complex. This complex can reach 120 nm in length while only measuring 10 nm in diameter. Following contact with a host cell membrane, three proteins, PopB, PopD and PcrV, are translocated to the tip of the needle complex where they form a pore in the host cell membrane, allowing for the entry of cytotoxic effector proteins: ExoS, ExoU, ExoT, and ExoY.

There exist four distinct effector proteins, yet it is unusual for any one strain to possess all four. ExoS and ExoT are similar, as both proteins contain a domain capable of ADP-ribosylation and a GTPase activating domain. The activity of these two effectors interferes with the host cell cytoskeleton and can lead to significant restructuring of actin filaments. ExoU is a phospholipase capable of destroying eukaryotic membranes, leading to rapid cell death. ExoY was previously considered to be an adenyl cyclase, but recent investigations have demonstrated it is capable of producing large amounts of cGMP, along with cCMP and cUMP, making ExoY a promiscuous nucleotidyl cyclase. It is known that ExoY alters host cell shape, preventing phagocytosis and causing cell rounding along with increasing the endothelial gaps, although the mechanism through which this occurs is still unclear.

Expression of the T3SS is principally regulated by the main activator of T3SS function in *P. aeruginosa*: ExsA. It has been demonstrated that ExsA can be activated as a result of direct contact with a host cell, as well as under low calcium conditions. A member of the AraC-family of transcriptional regulators, dimeric ExsA recognizes a binding site that overlaps the -35 site of target promoters and extends upstream roughly 34 bases. Currently, there are 10 promoters that have been identified as being regulated by ExsA, which include the structural components.
of the Type 3 secretion needle complex as well as the secreted effectors and regulatory proteins, including ExsA's own promoter. Under conditions unfavorable for the expression of the T3SS (non-inducing), ExsA is bound in an inactive state by the anti-activator ExsD. ExsA can be freed from ExsD-dependent repression by ExsC, which functions as an anti-anti-activator, binding ExsD and preventing interaction with ExsA. ExsC is typically held inactive through binding to ExsE in a 2:1 complex, however, under T3SS-inducing conditions, ExsE is secreted out of the cell through the T3SS needle, effectively freeing ExsC and ultimately ExsA from inhibition.

**Biofilm Formation**

While *Pseudomonas aeruginosa* boasts an impressive array of cytotoxins capable of causing acute infection, there are situations in which it would be detrimental to the bacterium to express such potent genes. In these situations, such as infection in the cystic fibrosis (CF) lung, gene expression is geared more toward long-term survival rather than actively attacking the host. Cystic fibrosis is the most common, lethal, genetic disease in humans, caused by a mutation in the cystic fibrosis transmembrane conductance regulator (CFTR) which results in a loss of regulation by cAMP. The deregulation of this protein ultimately causes secretion of abnormally thick mucus into the lung. This thick mucus impairs the traditional immune response and results in the colonization of the lung by numerous bacteria. Of the bacteria commonly found in the CF lung, *Pseudomonas aeruginosa* stands out, as nearly every patient with CF will become colonized by *P. aeruginosa* sometime during their life. Infection of the CF lung by *P. aeruginosa* leads to progressive pulmonary degradation and, ultimately, death.

Following the initial colonization of the CF lung, *P. aeruginosa* is capable of forming a biofilm, dramatically increasing its resistance to antibiotics and effectively hiding the bacteria from the host immune system. *Pseudomonas aeruginosa* is commonly studied in conjunction with biofilm formation as it is capable of forming highly structured, three-dimensional biofilms *in vitro*. Biofilms are formed through a step-by-step process beginning with attachment of cells to a
surface, followed by the formation of a microcolony, which then matures into a full-fledged biofilm. The biofilm cycle is then completed when the sessile cells become planktonic again and are released into the environment through a process known as dispersion. Bacteria in biofilms are encased in a thick matrix composed mostly of exopolysaccharides (EPS) along with extracellular DNA and some proteins. This matrix provides an effective buffer zone between the bacterial cells and their environment, contributing greatly to bacterial evasion of the human immune system. In Pseudomonas aeruginosa the principle components of the biofilm matrix are alginate and the Pel and Psl exopolysaccharides, although the contribution of each EPS to the matrix will vary based on strain. In mucoid strains, such as clinical isolates from CF patients, alginate appears to be the dominate EPS in biofilm matrices. While in environmental isolates, such as the lab strains PAO1 and PA14, two of the strains commonly used to study biofilm formation, Pel and Psl are the dominant matrix exopolysaccharides. The production of EPS is commonly used as an indicator of biofilm formation in laboratory assays as EPS can be easily stained using either Crystal Violet or Congo Red dyes.

REGULATION OF VIRULENCE IN PSEUDOMONAS AERUGINOSA

As a ubiquitous bacterium, Pseudomonas aeruginosa is able to survive and thrive in a variety of environments, including the human body. One reason P. aeruginosa can thrive in diverse environments is likely that it devotes a significant portion of its genome to regulatory factors. These regulatory factors primarily consist of two-component phosphorelay systems and small molecule signaling systems such as nucleotide based second messengers and quorum sensing systems.

Two-component systems

Two-component systems represent the most widespread mechanism of regulation in not only Pseudomonas aeruginosa, but the prokaryotic world as a whole. A canonical two-component
system consists of two parts: a histidine kinase and a response regulator. Histidine kinases are typically homodimeric proteins which frequently contain an additional sensory domain capable of detecting an environmental signal. Upon activation, commonly through ligand binding to the sensory domain, histidine kinases catalyze the hydrolysis of ATP and autophosphorylate in trans at a conserved histidine residue. The phosphate group can then be transferred from the histidine to an aspartate residue in the response regulator domain. Typically the response regulator domain will be located on a separate protein, often a transcriptional regulator. However, in the case of hybrid histidine kinases, the initial response regulator domain is a part of the same protein as the kinase domain. Hybrid histidine kinases typically require a histidine phosphotransfer (Hpt) domain, either as another domain of the hybrid histidine kinase or as a separate protein, as another step in the phosphotransfer reaction. The Hpt domain contains a histidine residue which accepts a phosphate group from the conserved aspartate residue in a response regulator domain. The Hpt domain can then transfer the phosphate group to a different response regulator, ultimately inducing an altered phenotype. Gao and Stock have published an excellent review detailing the mechanisms and importance of the two-component system in prokaryotes.

In *P. aeruginosa*, as well as the majority of gamma-proteobacteria, two-component systems play a vital role in regulating essential cell functions ranging from biofilm formation and virulence factor expression, to carbon metabolism and chemotaxis.

**cAMP signaling**

Nucleotide based signaling in prokaryotes was first discovered in the 1960s when cyclic AMP (cAMP) was found to be excreted in broth culture by *Brevibacterium liquefaciens*. Cyclic AMP is synthesized by adenylyl cyclases and is able to freely diffuse throughout the cell and bind to proteins homologous to the *E. coli* catabolite regulation protein (CRP). In *E. coli*, CRP is a global transcriptional regulator which requires cAMP as a cofactor and is capable of directly regulating, both positively and negatively, roughly half the operons in the *E. coli* genome, including its
own\textsuperscript{87,88}. While CRP is extensively studied in the context of metabolic control, its homologs possess extremely diverse activities, with many homologs, including Vfr in \textit{Pseudomonas aeruginosa}, implicated in the regulation of virulence factors \textsuperscript{89-91}.

One of the best examples of nucleotide based signaling in \textit{P. aeruginosa} is the cAMP/Vfr signaling pathway \textsuperscript{92}. Transcription of \textit{vfr} is regulated by the AlgZ/R two-component system, which will be discussed later\textsuperscript{93}. Vfr is a homolog of the \textit{E. coli} CRP, and acts by upregulating expression of the T3SS in response to the binding of cAMP \textsuperscript{94}. In \textit{P. aeruginosa}, cAMP is produced primarily by the adenylyl cyclase CyaB, whose activity is modulated by the Chp chemosensory system, presumably in response to an unknown signal \textsuperscript{95}. In general, high intracellular levels of cAMP are correlated with increased expression of acute virulence factors in \textit{P. aeruginosa} \textsuperscript{94}.

**C-di-GMP signaling**

Bis-(3´-5´)-cyclic dimeric guanosine monophosphate (c-di-GMP) is a signaling molecule found throughout the prokaryotic world which plays the role of a second messenger, similar to cAMP. Unlike cAMP, elevated c-di-GMP levels are associated with an increase in biofilm-related genes and a decrease in motility and virulence associated genes \textsuperscript{96}. Cyclic di-GMP is synthesized by diguanylate cyclases (DC), which contain a GGDEF domain, and degraded by specific phosphodiesterases (PDE) which contain an EAL or HY-DAP motif \textsuperscript{97-99}. By modulating the active levels of DCs and PDEs, the cell is able to finely control the concentration of c-di-GMP and elicit a desired response. Although a cell may transcriptionally regulate the levels of active DCs and PDEs, many of these proteins contain sensory domains, similar to histidine kinases involved in two-component systems, which allow them to sense and respond to signals and cause rapid, global changes in gene expression as a result of altered c-di-GMP expression \textsuperscript{100}. Cyclic di-GMP is recognized by an array of regulatory proteins which activate upon binding C-di-GMP and serve to alter gene expression.
While still not fully understood, c-di-GMP signaling has recently become an area of intense study due to its control of gene expression in *P. aeruginosa* and other prokaryotes.

Approximately 40 proteins have been identified in *P. aeruginosa*, so far, which possess the required motifs to participate in c-di-GMP signaling. Of those 40 genes, two of the best studied are SadC and WspR. The diguanylate cyclase WspR is part of a chemosensory pathway responsible for cell aggregation. As part of this pathway, WspR activity is increased via phosphorylation in response to cell contact with a surface.

The consequence of the increased WspR activity is increased production of the Pel/Psl exopolysaccharides. This self-produced EPS can then act as its own signal, stimulating the activity of SadC, leading to even higher intracellular levels of c-di-GMP, serving as a positive feedback loop, which can result in a rapid up-regulation of the expression of EPS in a relatively short timeframe. Translation of *sadC* mRNA into a functional protein has been found to be tightly repressed by RsmA, placing SadC activity under the control of the Gac/Rsm pathway. The importance of SadC isunderscored by the fact that it is required for the progression into the irreversible attachment stage of biofilm formation, meaning *sadC* mutants are incapable of forming mature biofilms.

**Quorum sensing**

First discovered in *Vibrio fischeri*, quorum sensing is essentially cell density dependent regulation of gene expression. The basic mechanism for how quorum sensing begins with a cell which is constantly producing some kind of signal, termed an autoinducer. In Gram negative bacteria autoinducers are typically N-acylhomoserinelactones (AHLs). These AHLs are able to freely pass through the cell membrane. As the number of cells producing AHL increases, so does the concentration of AHL in both the extracellular and intracellular environments. Once a threshold level of AHL is achieved, AHL receptor proteins are able to bind the autoinducer. These receptor proteins are typically transcriptional regulators, which are capable of modifying gene expression following activation by autoinducer binding.
Social behaviors in *P. aeruginosa*, such as biofilm formation, are often controlled by quorum sensing systems. *P. aeruginosa* possesses a hierarchical quorum sensing system consisting of three distinct pathways: Las, Rhl, and PQS. Chief amongst the three systems is the Lux-homologous Las system, as it positively regulates both the Rhl and PQS systems. The Las system is composed of LasI, the protein responsible for synthesizing the autoinducer 3-oxo-C12-HSL, and LasR, which functions as an autoinducer receptor and transcription factor. The Rhl system is composed of RhlI which produces the autoinducer C4-HSL, which is bound by RhlR. The PQS system is unique to pseudomonads, and involves an entire biosynthetic operon to produce the autoinducer 2-heptyl-3-hydroxy-4-quinolone (PQS), which is bound by PqsR. Although the Las system positively regulates both the Rhl and PQS systems, all three systems are interconnected through various networks of regulation, leading to an extremely complex quorum sensing system which is the focus of intense interest and research.

In *P. aeruginosa*, quorum sensing has been implicated in helping regulate expression of genes associated with the formation of a biofilm. It has been demonstrated that both the Las and Rhl systems are required for biofilm formation, and that they are also capable of activating production of the Pel/Psl exopolysaccharides. The Rhl system was originally named due to its control over the biosynthetic operon responsible for the production of rhamnolipids. Rhamnolipids have been shown to play an important role in forming the 3-dimensional structure of *P. aeruginosa* biofilms, and are also required for microcolony formation and the release of planktonic cells from the mature biofilm during dispersion.

Little work has been done on the relationship between the *P. aeruginosa* quorum sensing systems and the expression of the T3SS, yet it has been shown that the Rhl system represses the activity of the T3SS, demonstrated by increased secretion of ExoS in a *rhlI* mutant. While few studies investigate quorum-controlled expression of the T3SS, the Las and Rhl systems have been shown to be required for expression of a number of other acute virulence factors, including exoproteases, phenazines, and hydrogen cyanide.
The Gac/Rsm pathway

In *Pseudomonas aeruginosa*, the Gac/Rsm pathway (Fig. 3) combines almost every method of regulation described above to reciprocally regulate the expression of the T3SS and biofilm formation. The lynchpin of the Gac/Rsm pathway is the histidine kinase GacS, which is inhibited through a direct physical interaction with the putative sensor kinase RetS. This interaction with RetS prevents GacS from autophosphorylating and activating its cognate response regulator GacA through a canonical phosphotransfer reaction. GacA, when phosphorylated, is capable of directly activating the transcription of two genes, *rsmY* and *rsmZ*, which encode sRNAs. These sRNAs, when produced in large enough quantities, sequester the RNA binding protein RsmA. Free RsmA is able to post-transcriptionally inhibit the expression of genes associated with the T3SS and indirectly activate the expression of genes resulting in biofilm formation, most notably the Pel and Psl biosynthetic pathways. The Gac/Rsm pathway has also been shown to regulate the levels of C4-HSL produced by the Rhl system, presumably through RsmA, demonstrating that the Gac/Rsm pathway is able to tightly regulate EPS production via not only RsmA-mediated regulation of the *psl/pel* operons, but also modulation of the Rhl quorum sensing system.

The RetS/GacS interaction

The Gac/Rsm pathway acts as a master control switch between acute and chronic infection phenotypes (Fig. 3). GacS is a histidine kinase which is able to phosphorylate its cognate response regulator, GacA, which is activated upon phosphorylation. Once active, GacA is capable of binding to DNA to increase the transcription of the small regulatory RNAs (sRNA) *rsmZ* and *rsmY*, which functions to sequester the protein RsmA, an activator of T3SS expression and transcriptional inhibitor of the exopolysaccharides Psl and Pel, which, as mentioned earlier, are necessary components of the biofilm matrix.
The activity of GacS is modulated by counteracting regulation from two hybrid sensor kinases: LadS and RetS. LadS, responding to calcium levels, acts as an activator of the GacS/A pathway, shuttling phosphates through the Hpt domain of GacS, which ultimately promotes chronic infection, while RetS acts as an inhibitor of GacS, ultimately promoting acute infection. The activity of each of these hybrid sensor kinases is presumably dependent on their binding of a currently unknown signal, as they both possess periplasmic ligand binding domains. The exact mechanism of how RetS interacts with GacS has been proposed as a completely novel interaction between sensor kinases that requires physical contact between the kinase domains, yet does not require any of the conserved histidine and aspartate residues necessary for the classical function of two-component sensor kinases.

**GacS**

GacS is a canonical, hybrid histidine kinase of approximately 100 kDa in size. As an integral membrane protein, GacS possesses a periplasmic domain, the structure of which has recently been elucidated. GacS possesses the traditional residues necessary for it to function in a phosphorelay reaction, with the histidine at residue 284 being the site of autophosphorylation. Following the classical model, the phosphate group would then be passed to the aspartate residue at position 707 in GacS’ response regulator domain. From that aspartate, the phosphate group would then be transferred to a second histidine, at position 853 in the Hpt domain of GacS, the second histidine would finally transfer the phosphate group to GacA. Although it is assumed that phosphorelay would follow the classical model, there exists some evidence that, under certain cellular conditions, GacS is capable of bypassing the phosphorelay reaction and phosphorylating GacA directly from the initial histidine based on studies of homologous systems in other bacteria.

GacS, along with many other sensor kinases, possess a HAMP signal transduction domain. HAMP domains, named for their presence in histidine kinases, adenyl cyclases, methyl-
accepting chemotaxis proteins, and phosphatases, are responsible for transducing a signal from the periplasm to the cytoplasmic domains of a protein \(^{136}\).

**RetS**

RetS is an approximately 104 kDa protein, which is annotated as a hybrid sensor histidine kinase \(^{137}\), indicating that it would be expected to contain a histidine region, as well as a periplasmic sensory domain, and response regulator domain. However, RetS is unorthodox in that it contains two response regulator domains in tandem at its C-terminal region, a structural arrangement that is not commonly seen.

As stated above, RetS uniquely interacts with GacS and ultimately serves to promote expression of the T3SS and acute infection \(^{129}\). This interaction has been shown to only require the kinase core domain of RetS, with the response regulator domains not being necessary for formation of a RetS/GacS multimeric complex \(^{123}\). The kinase domain of RetS shares a high degree of homology with the kinase domain of traditional, functional histidine kinases, such as GacS, with the histidine in position 424 acting as the conserved putative site of autophosphorylation, although such activity has yet to be detected. A threonine in position 428 is conserved across histidine kinases as the critical residue required for phosphatase activity. RetS has been demonstrated to de-phosphorylate GacS \(^{123}\), although it has not been demonstrated that the threonine in position 428 is the site of that phosphatase activity.

The function of RetS' response regulators is currently unknown as there is no evidence of those domains participating in a phosphotransfer reaction. Likewise, the function of the RetS periplasmic sensory domain is currently unknown. Although the crystal structure of the sensory domain was previously solved in the lab \(^{138}\), and it was predicted to bind a carbohydrate ligand, the specific ligand has yet to be found.
Studies in *Pseudomonas fluorescens* strain CHAO, prior to the discovery of RetS in *P. aeruginosa*, found that *rsmZ* levels could be altered with the addition of a non-AHL, dichloromethane-soluble signal produced by the cells during stationary phase, which appeared to exert its effect through the Gac/Rsm pathway\(^{139}\). These researchers believed the signal to be sensed by GacS however, it is possible that the poorly characterized, unknown signal exerts its effect on GacS by first binding to RetS (and/or LadS) to alter the resulting interaction with GacS.

It has been theorized that the sensory domain of RetS binds a ligand that allows it to modulate its interaction with GacS, however, evidence in support of this theory is severely lacking. Aside from the fact that the identity of the presumed ligand is unknown, it has yet to be demonstrated that the sensory activity possesses any detectable biological function. Previous experiments in the lab have shown that RetS mutants lacking the periplasmic sensory domain fail to demonstrate a significant phenotypic difference when compared to controls.

In addition to its interaction with GacS, recent work has suggested that RetS is also capable of directly interacting with another hybrid sensor kinase of unknown function, PA1611\(^{140}\). This study found that expression of PA1611 resulted in a phenotype similar to a ΔretS mutant, which led the researchers to suggest a model in which PA1611 serves to bind to and sequester RetS in a heteromeric complex until a signal is received which causes dissolution of the complex and allows RetS to interact with GacS. A direct interaction between RetS and PA1611 has recently been claimed\(^{141}\), yet the described interface possesses no biological relevance, as the authors claimed disruption of the PA1611 homodimer and formation of a RetS/PA1611 heterodimer. The authors claimed the interaction interface was formed by the solvent-exposed β sheet of the RetS CA domain and the antiparallel α-helices of the PA1611 DHp domain which mediate homodimerization. However, to date, there is no evidence of histidine kinase homodimer dissociation.

Although PA1611 is a hybrid sensor kinase like RetS, unlike RetS, PA1611 has demonstrated the ability to participate in a phosphotransfer reaction. Starting with autophosphorylation, it was
found that PA1611 could pass a phosphate group to a response regulator (PA3346), with the protein HptB acting as an intermediary\textsuperscript{142}. Interestingly, an hptB mutant displays a phenotype similar to a retS mutant, suggesting an interaction of both proteins. However, it has been demonstrated that RetS and HptB signal through distinct pathways that merely intersect at GacA\textsuperscript{143}. Although evidence linking HptB and RetS is lacking, it is possible, based on similarities with PA1611, that RetS interacts with another \textit{P. aeruginosa} Hpt protein, yet to be identified.

**GOALS OF THE STUDY**

The goals of this study are to develop an in-depth understanding of the molecular mechanisms underlying protein assembly formation and subsequent biological activity. Two different protein assemblies were selected as the foci for this study. PilB is the ATPase responsible for powering T4P assembly. PilB assembles into a homo-hexamer and is a member of the AAA+ family of molecular machines which play critical roles in a number of essential cellular processes. The goal of studying PilB is to understand the exact mechanism by which PilB hydrolyzes ATP and how the chemical energy generated from ATP hydrolysis is converted into the mechanical energy necessary to assemble the type IV pilus fiber (Chapter 2). The second focus of this study is the RetS/GacS complex, which plays a critical role in regulating virulence in \textit{P. aeruginosa}. RetS and GacS assemble into a hetero-complex of unknown composition and represent the first documented hetero-oligomerization event between histidine kinases. The goal of studying this complex is to understand the exact mechanism that allows for interaction between RetS and GacS, as well as the RetS-mediated inhibition of GacS activity.
FIGURES

Figure 1: Schematic demonstrating binding of the RNA polymerase holoenzyme to a bacterial promoter and formation of the closed complex. Figure adapted from 143.

Figure 2: Single particle cryo-EM reconstruction of the bacterial flagella indicating the positioning of each ring body. Adapted from 30, see text for details.
**Figure 3:** Schematic of the Gac/Rsm pathway with the RetS/GacS interaction and type IV pili highlighted as the emphasis of this research.

**REFERENCES**

51. Leo, J.C., Grin, I. & Linke, D. Type V Secretion: Mechanism(s) of Autotransport Through the Bacterial Outer Membrane. Philosophical Transactions of the Royal Society B: Biological Sciences 367, 1088 (2012).
89. Lampidis, R., Gross, R., Sokolovic, Z., Goebel, W. & Kreft, J. The Virulence Regulator Protein of *Listeria ivanovii* is Highly Homologous to PrfA from *Listeria monocytogenes* and Both Belong to the Crp–Fnr family of Transcription Regulators. *Molecular Microbiology* **13**, 141-151 (1994).


Chapter 2: Crystal Structure of a Type IV Pilus Assembly ATPase: Insights into the Molecular Mechanism of PilB from Thermus thermophilus

1 Jordan M. Mancl, 1Wesley P. Black, 2Howard Robinson, 1Zhaomin Yang, 1.*Florian D. Schubot

1Virginia Polytechnic Institute and State University, Department of Biological Sciences, Life Sciences 1 (MC 0910), 970 Washington Street SW, Blacksburg, Virginia 24061; 2Biology Department, Brookhaven National Laboratory, Upton, NY 11973-5000.

* Corresponding author:

Virginia Polytechnic Institute and State University
Department of Biological Sciences
125 Life Sciences 1 (MC 0910)
970 Washington Street SW
Blacksburg, Virginia 24061
E-mail: fschubot@vt.edu
Phone: (540) 231-2393
Fax: (540)231-4043

Author Contributions

J.M. conducted all the research and wrote the manuscript. W.B. developed the expression and purification protocols for the PilB protein. H.R. collected diffraction data. Z.Y. and F.D.S. supervised the research, assisted with data analysis, and wrote the manuscript.

This work has been published as a standalone article in the November 2016 issue of Structure:

SUMMARY

Type IV pili (T4P) mediate bacterial motility and virulence. The PilB/GspE family ATPases power the assembly of T4P and type 2 secretion systems. We determined the structure of the ATPase region of PilB (PilB\textsubscript{ATP}) in complex with ATPγS to provide the first model of a T4P assembly ATPase and the first view of any PilB/GspE family hexamer at better than 3 Å resolution. Spatial positioning and conformations of the protomers suggest a mechanism of force generation. All six PilB\textsubscript{ATP} protomers contain bound ATPγS. Two protomers form a closed conformation poised for ATP hydrolysis. The other four molecules assume an open conformation but separate into two pairs with distinct active site accessibilities. We propose that one pair represents the post-hydrolysis phase, while the other pair appears poised for ADP/ATP exchange. Collectively, the data suggest that T4P assembly is powered by coordinating concurrent substrate binding with ATP hydrolysis across the PilB hexamer.
INTRODUCTION

Type IV pili (T4P) are prevalent cellular surface structures in the bacterial world, consisting of hair-like appendages which play a crucial role in the lifestyles of many bacteria including *Myxococcus xanthus* (Li et al., 2003), *Neisseria gonorrhoeae* (Merz and So, 2000), *Thermus thermophilus* (Friedrich et al., 2003), and *Pseudomonas aeruginosa* (Comolli et al., 1999). T4P fibers consist primarily of repeating PilA subunits, but variable amounts of several auxiliary pilins are also incorporated (Alm, 1996; Craig et al., 2004; Paranchych, 1979). They are involved in a diverse array of cellular functions including motility, adherence, and biofilm formation (Strom and Lory, 1993; Wang et al., 2013). T4P are of intense interest in the field of molecular motors because the T4P motor complex is capable of generating forces in excess of 100 pN (Maier et al., 2002), ranking them among the most powerful molecular machines. A detailed characterization of the structure and function of T4P is fundamental for improving our basic understanding of bacterial motility. Beyond addressing basic research questions, these studies also have the potential to inform development of novel antimicrobials because numerous bacterial pathogens require T4P for infection (Craig et al., 2004).

Bacterial twitching motility, a flagella-independent movement along surfaces, is mediated through two antagonistic motions, assembly and retraction of the T4P fiber (Bradley, 1980). During assembly, the pilin subunits polymerize to form a mature pilus fiber which protrudes from the cell and eventually attaches to a surface. Rapid disassembly causes the pilus to retract, thereby propelling the cell towards the attachment point. The T4P machinery responsible for polymerization and depolymerization of the filament assembles into a large multi-protein complex analogous to the bacterial type 2 secretion system (T2S) apparatus (Peabody et al., 2003). T4P assembly and retraction are powered by the ATPases PilB and PilT, respectively. Collective evidence from biochemical studies of T4P proteins and analogous processes in the T2S system suggests that PilB and PilT interact with PilM and PilC to relay forces generated by
these cytoplasmic ATPases across the membrane to the site of pilus assembly and disassembly (Bischof et al., 2016; Chang et al., 2016; Mattick, 2002; Whitchurch et al., 1991). The primary structural distinction that may be gleaned from a sequence comparison of PilB and PilT is the presence of a 200 residue long N-terminal domain in PilB, which is absent in PilT. This N-terminal extension is conserved among T4P pilus assembly ATPases as well as T2SS ATPases. For the latter enzymes, this domain has been shown to interact with the inner membrane platform proteins (Chen et al., 2005) and might serve a similar role in PilB (Chang et al., 2016). Recently, significant strides have been made toward understanding the structure and function of PilT-powered retraction, however, comparable advances in the understanding of PilB-powered pilus assembly have not been forthcoming.

PilB and its antagonist PilT are members of the AAA+ (ATPases associated with diverse cellular activities) superfamily of proteins (Iyer et al., 2004). Usually functional hexamers, AAA+‘s power a variety of cellular activities including protein degradation, DNA uptake, and motility (Vale, 2000; Wendler et al., 2012). The signature feature of AAA+ enzymes is a conserved, two-domain ATPase core capable of oligomerization which encompasses conserved Walker A and B boxes (Erzberger and Berger, 2006; Walker et al., 1982). AAA+ enzymes, including PilB and PilT, also contain conserved arginine residues in the N-terminal domain that play an important role during catalysis. These arginines stabilize the transition state and promote ATP hydrolysis (Bos et al., 2007) primarily through interactions with the γ-phosphate of ATP, similar to the role of the arginine finger in G-protein GAP complexes (Ogura et al., 2004).

While the chemical mechanism of ATP hydrolysis is well-understood, questions remain as to how ATP hydrolysis is coordinated across AAA+ hexamers and mechanical forces are produced. Oligomerization is essential for protein function, but cooperativity is not universally observed (Erzberger and Berger, 2006). Collectively, results from numerous studies suggest
that AAA+ ATPases do not use a single universal mechanism. Instead, some four different scenarios have been proposed: a symmetric rotary mechanism, an asymmetric rotary mechanism, concerted catalysis, and finally a stochastic mechanism (Cordova et al., 2014; Johnson and O’Donnell, 2003; Liu et al., 2014; Lyubimov et al., 2011; Martin et al., 2005; Ogura and Wilkinson, 2001). It appears that the mechanism of ATP hydrolysis correlates to the biological function of the AAA+ ATPase and the amount of energy required for the associated biological process. In this regard, the proteins that most closely mirror the biological function of PilB would be its counterpart PilT, the ATPases associated with the bacterial type II secretion system, and the ATPases that drive flagellar assembly and rotation in archaea.

Crystal structures have been determined for a number of these related enzymes. These include PilT from *P. aeruginosa* (*Pa*PilT) and *Aquifax aeolicus* (*Aa*PilT) with and without ATP analogs (Misic et al., 2010; Satyshur et al., 2007), the T2S ATPase from *Vibrio cholerae* (*Vc*GspE or EpsE) (Lu et al., 2013; Robien et al., 2003), the archaeal flagellar ATPase Flai from *Sulfolobus acidocaldarius* (*Sa*Flai) (Reindl et al., 2013), and the putative archaeal GspE2 from *Archaeoglobus fulgidus* (*Af*GspE) (Yamagata and Tainer, 2007). Recently, cryoelectron microscopy reconstructions of the nucleotide-bound and nucleotide-free states of the PilB protein from *Thermus thermophilus* (*Tt*PilB) have been published (Collins et al., 2013). Those images are consistent with biochemical studies showing *Tt*PilB to be hexameric both in absence and presence of nucleotide (Salzer et al., 2014). Comparison of the nucleotide-bound with the nucleotide-free states revealed a significant conformational change in response to nucleotide binding. Analysis of the *Tt*PilB reconstructions supported the notion of two distinct conformations, a “bound” and “unbound” state but the relatively low resolution and the six-fold symmetry constraints imposed during the reconstruction did not permit detailed insights into how ATP hydrolysis is coordinated to power T4P assembly.
Because no atomic resolution structure of a PilB enzyme has been reported to date, we crystallized and determined the structure of the PilB core ATPase domain of *Tt*PilB (*Tt*PilB<sub>ATP</sub>) in complex with ATPγS at 2.65 Å resolution. In particular, we sought to determine how ATP hydrolysis is coordinated across the PilB hexamer and how these coordinated events would generate force. Collectively, our results point toward a model wherein ADP/ATP exchange in two protomers is coordinated with ATP hydrolysis in another pair of protomers consistent with a counterclockwise symmetric rotary mechanism of catalysis across the hexamer. The direction of the resultant force is consistent with recently published models for T4P assembly.

**RESULTS AND DISCUSSION**

**Structure of hexameric *Tt*PilB<sub>ATP</sub> at 2.65 Å resolution**

*Tt*PilB<sub>ATP</sub>, encompassing amino acids 506 – 890, crystallized in the presence of the ATP analog ATPγS in the orthorhombic space group P22₁2₁. The structure was solved via molecular replacement and refined against x-ray diffraction data of up to 2.65 Å resolution. The asymmetric unit of the crystal contains six *Tt*PilB<sub>ATP</sub> molecules forming an elliptical, hexameric toroid (Fig. 1.A). The hexamer exhibits two-fold symmetry, with molecules A and D, B and E, and C and F forming equivalent pairs. All six *Tt*PilB<sub>ATP</sub> molecules appear well-ordered within the crystal, producing unbroken electron density for the backbone of each chain.

At the quaternary structure level, the *Tt*PilB<sub>ATP</sub> ring measures about 135 Å in length, 90 Å in width, and 65 Å in height. The largest dimension of the central pore measures 45 Å while it only measures 15 Å at its narrowest point (Fig. 1.A.). As expected, PilT and the T2SS ATPase GspE were found to be structurally most similar to *Tt*PilB<sub>ATP</sub>. Similar elongated hexamers with approximate two-fold symmetry have been observed for *Aa*PilT crystallized in the presence of ADP (pdb code: 2GSZ) (Satyshur et al., 2007) and a complex of a hexameric *Vc*GspE-Hcp1
fusion protein with AMP-PNP (pdb code: 4KSR) (Lu et al., 2013). However, a detailed molecular analysis was not possible in either case because the resolution of the diffraction data was limited to 4.2 Å. The higher resolution of the TtPilB\textsubscript{ATP} structure here enables us to correlate the quaternary structure of the ring with the catalytic mechanism.

**The tertiary structure of TtPilB\textsubscript{ATP}protomers closely resemble those of T2SS ATPases**

Individual protomers assume the characteristic AAA+ ATPase fold (Fig. 1.B) also observed for PilT (Misic et al., 2010; Satyshur et al., 2007) and GspE (Lu et al., 2013; Robien et al., 2003; Yamagata and Tainer, 2007), which consists of an N-terminal domain (NTD) connected to a C-terminal domain (CTD) by a flexible linker. The PAS-like NTD, formed by residues 506 to 612, is composed of a central six-stranded anti-parallel β-sheet that cradles two α-helices. The CTD, encompassing residues 628 to 890, consists primarily of an ASCE (additional strand catalytic E) domain. A twisted, five-stranded, parallel β-sheet forms the structural core of this domain (Story et al., 1992; Wendler et al., 2012). The structural topologies of the TtPilB\textsubscript{ATP} and other T2S/T4P ATPases deviate slightly from the canonical ASCE domain architecture, as ASCE domains are generally characterized by a β-strand topology of 5-1-4-3-2, while in this subfamily the β-strands are arranged in the order of 1-5-4-2-3. Key residues nevertheless align with the analogous residues in the canonical ASCE domain.

In TtPilB\textsubscript{ATP}, the core ASCE motif is extended by two large, antiparallel β-strands (β12/β13). Inserted between these two strands is a coordinated metal domain (CMD) wherein four cysteines coordinate a single Zn\textsuperscript{2+} ion. This CMD, found to be essential for maintaining the structural stability of the protein (Salzer et al., 2014), is also present in VcGspE but absent in all known PilT proteins. In TtPilB\textsubscript{ATP}, the CMD is followed by four C-terminal α-helices, which are conserved among GspE, PilB, and PilT proteins and collectively referred to as a C2 domain (C2D) (Fig. 1.B). α10 and α11, the first two helices of the C2D, line the inner pore of the
hexamer and contribute greatly to the asymmetric appearance of the pore (Fig. 1). As discussed below, these two helices may play a role in transforming the energy released by the ATPase activity into the mechanical work required for T4P assembly. Helices α12 and α13 are sandwiched between the CMD and strands β12 and β15 (Fig. 1B). The hypothesis that the CMD is important for the stability of the T4P/T2S assembly ATPases is supported by the observation that the two terminal helices produce clearly defined density in $Tt$PilB$_{ATP}$ but significantly weaker and partially broken electron density in the PilT structures that lack the CMD (Misic et al., 2010; Satyshur et al., 2007).

**Conformational changes in a flexible inter-domain linker separate $Tt$PilB$_{ATP}$ into open and closed forms**

In $Tt$PilB$_{ATP}$, residues 612 to 628 form an inter-domain linker between the PAS-like NTD and the ASCE domain (Fig. 1). In the open and closed forms of $Tt$PilB$_{ATP}$, this linker assumes two distinct conformations, both of which are clearly defined in the electron density map. Chains A, C, D, and F all assume what has previously been described as an open conformation (Satyshur et al., 2007). The four molecules can be overlaid with a root mean square deviation (RMSD) of about 1 Å for all backbone atoms. Molecules B and E, are also readily overlaid with an RMSD value of approximately 0.5 Å for all backbone atoms, assume a closed conformation. The differences between the two conformations are best explained by a hinge motion of roughly $55^\circ$ about the linker region connecting NTD and CTD (Figures Fig. 2.C and Fig. 5). The aforementioned structures of the VcGspE-Hcp1-AMPPNP and AaPilT-ADP complexes (Lu et al., 2013; Satyshur et al., 2007) not only show the greatest similarities to the present structure at the quaternary structure level, but the conformations of the individual protomers also most closely resemble those observed for $Tt$PilB$_{ATP}$, especially in the case of VcGspE. The open conformation, observed for chains A, C, D, and F of $Tt$PilB$_{ATP}$, is structurally most similar to that
of chains A and B of the VcGspE hexamer with an RMSD of 3.3 Å for 364 aligned residues. Chains B and E of TtPilB\textsubscript{ATP}, representing a closed conformation, align well with chain C of the same VcGspE complex, producing an RMSD of 2.9 Å for the superposition of 364 aligned residues. The closed conformation of TtPilB\textsubscript{ATP} is also structurally similar to the monomeric structure of VcGspE (pdb code: 1P9R and pdb-code: 1P9W) (Robien et al., 2003), producing an RMSD of 2.6 Å for 369 aligned residues.

**Protomers in the closed conformation are poised for ATP hydrolysis**

As expected for an AAA+ ATPase, the nucleotide binding pocket of TtPilB\textsubscript{ATP} is located at the interface between NTD and CTD. The map showed well-defined electron density for ATP\textsubscript{γS} molecules in all six protomers (Fig. S1). The nitrogenous base and carbohydrate moieties of the bound nucleotides form identical interactions with the P-loop residues Gly652-Ser-Gly-Lys-Ser656 in all six protomers (Fig. S2). This loop also coordinates the α and β phosphate groups of the bound nucleotide. Differences amongst chains emerge with respect to the coordination of the γ-thio-phosphate group (Fig. 2). In the open conformation (molecules A, C, D, and F), the γ-thio-phosphate is only coordinated by the sidechain of Ser656. In chain F, the ligand density indicates the formation of a covalent bond between Ser656 and the γ-thio-phosphate. We have reconciled this anomaly by modeling a phosphoserine type link between Ser656 and the γ-thio-phosphate. We suspect that the bond was created through radiation damage incurred during data collection rather than being indicative of a step in catalysis.

In the closed conformation, represented by chains B and E, the γ-thio-phosphate is coordinated by additional residues. Most notably, two arginines from the NTD, Arg582 and Arg595 have swung into the binding pocket to form hydrogen bonds with the γ-thio-phosphate (Fig. 2.A). Additional electron density was visible just above Ser656 in the active sites of molecules B and E. Consistent with the expected mechanism of catalysis, we interpreted this
density to belong to a Mg$^{2+}$ ion. During catalysis, the Mg$^{2+}$ stabilizes the pentacovalent intermediate formed at the γ-phosphate (Ye et al., 2004). The Mg$^{2+}$ ion is accommodated by a 2.5 Å displacement of the γ-thio-phosphate in chains B and E compared to the other four chains. This displacement is facilitated through the hydrogen-bonding interactions between the γ-thio-phosphate and the Arg582 and Arg595. The same TtPilB$_{ATP}$ molecules possess additional density between the γ-phosphate and the sidechain of Glu681 (Fig. S1.B). The S$_{N}$2-type mechanism of catalysis predicts the presence of a water molecule in that position to act as a nucleophile during ATP hydrolysis. In related enzymes, glutamates equivalent to Glu681 in TtPilB$_{ATP}$ are predicted to polarize the water molecule to facilitate the nucleophilic attack on the ATP (Abrahams, 1994; Davidson et al., 2008). Such a mechanism is consistent with the previously proposed hydrolysis mechanism based on the structure of F1-ATPase (Abrahams, 1994).

An arginine conserved in PilB/GspE enzymes but not PilT, Flal or AfGspE is critical for coordinating the γ-phosphate of the substrate in TtPilB$_{ATP}$

In the open conformation, the side chains of Arg582 and Arg595 are positioned >8 Å away from the γ-phosphate (Fig. 2.B). The repositioning of the arginine fingers is accomplished by the aforementioned hinge motion. If one compares the conformations observed in the various structures of AAA+ ATPases (Lu et al., 2013; Masic et al., 2010; Reindl et al., 2013; Satyshur et al., 2007; Yamagata and Tainer, 2007), the relative positions of NTDs and CTDs vary significantly, suggesting that, at least in absence of ATP, the hinge region permits fluid domain motions. Even though the resolution of the structurally similar AaPilT and VcGspE hexamers was too low to provide detailed electron density around the active site, a comparison of the closed conformations of these molecules with that of TtPilB$_{ATP}$ revealed an interesting distinction with respect to the locations of the arginine fingers (Fig 3). Arg582 and Arg595 of
TtPilB_{ATP} are positioned on beta strands $\beta_4$ and $\beta_5$ of the PAS-like domain. TtPilB_{ATP} also contains a third arginine, Arg610 on $\beta_6$. However, Arg610 is not involved in coordinating the $\gamma$-phosphate of ATPγS in the closed form of TtPilB_{ATP}. Instead, it is part of the invariable subunit interface that creates the rigid unit of the hexamer discussed in a later section. All three arginines are conserved in VcGspE (Fig. 3.A), where Arg197, Arg210, and Arg225 correspond to Arg582, Arg595 and Arg610 of TtPilB_{ATP}, respectively. Like in TtPilB_{ATP}, Arg197 and Arg210 on $\beta_4$ and $\beta_5$ in the VcGspE-AMPPNP structure, are best positioned to interact with a $\gamma$-phosphate group of a bound nucleotide, although it appears that the molecule would have to still close further to facilitate effective interactions between arginine fingers and nucleotide. AaPilT does not have an arginine on $\beta_4$ that is structurally equivalent to Arg582 of TtPilB_{ATP} (Fig. 3.B). Instead, its Arg110, which corresponds to Arg610 on $\beta_6$ in TtPilB_{ATP}, appears to interact with the $\gamma$-phosphate along with Arg95, the equivalent to Arg595 on $\beta_5$ in TtPilB_{ATP}. A direct consequence of this difference is that the closing motion observed for AaPilT exceeds that of TtPilB_{ATP} thus explaining why the NTD of TtPilB_{ATP} does not overlay well with the NTD of AaPilT (Satyshur et al., 2007) in the closed conformation. All known PilB/GspE proteins appear to possess a residue equivalent to Arg582 while the known PilT proteins do not (Fig. 3.C). This suggests the identification of a further distinguishing feature between PilT and PilB/GspE enzymes, in addition to the extra N-terminal domain and the CMD domain in the bacterial PilB/GspE proteins.

With this in mind, it appears that the archaeal ATPases AfGspE and SaFlaI blur the line between PilB and PilT type enzymes, as they both lack a CMD as well as a conserved arginine on the $\beta_4$ strand, instead using the arginine finger configuration observed in PilT (Fig. 3.C) (Reindl et al., 2013; Yamagata and Tainer, 2007). The biological function of AfGspE, a putative T2S ATPase from A. fulgidus, remains to be experimentally determined. The ATPase SaFlaI is critical for the function of the archaeal flagellum or archaellum, a pilus-like structure, which is
evolutionally related to but functionally distinct from the T4P. Here, Flal ATPase is not only required for assembly of the archaellar filament but also for its rotation in both directions (Albers and Jarrell, 2015).

Open protomers display varying degrees of solvent accessibility to their active sites

Our structure of the *Tt*PilB<sub>ATP</sub> hexamer contains six bound nucleotides. This raised the question whether, under physiological conditions, *Tt*PilB protomers exist in the apo-form free of nucleotide or if ADP only leaves the active site when being forced out by an incoming ATP molecule. To address this question we performed nucleotide binding experiments. First we measured the dissociation constant for the interaction of *Tt*PilB<sub>ATP</sub> with the ATP analog 2′-(or-3′)-O-(Trinitrophenyl)-ATP (TNP-ATP), obtaining a value of 40 mM (Fig. 4.A). Subsequently, we used a competition assay to probe the affinity of *Tt*PilB<sub>ATP</sub> for ADP and found a K<sub>d</sub> of about 80 mM (Fig. 4.B). In *E.coli*, the cellular concentrations of ADP and ATP were reported at 116 mM and 3.5 mM, respectively (Buckstein et al., 2008). If these levels are similar in *T. thermophilus*, the K<sub>d</sub> for ADP lies below the cellular concentration of ADP. This would suggest that ADP could remain stably bound to PilB until it is replaced by the tighter binding and more abundant ATP.

A scenario where ADP is competitively displaced by ATP is also consistent with an interesting observation from the *Tt*PilB<sub>ATP</sub> structure. As described above (Fig. 1), the *Tt*PilB<sub>ATP</sub> hexamer is formed by two conformationally distinct sets of protomers, four of which, chains A, C, D, and F, assume an open conformation. Within the ring, the positions occupied by chains A and D are not equivalent to the positions of molecules C and F. The most significant distinction between the two sets of positions is the different accessibilities of the active sites. These differences may be visualized by calculating pores and channels that permeate the hexamer.
Large pores provide access to the active sites of molecules A and D from the inside and the outside of the ring. We speculate that these pores could be used during ADP/ATP exchange following catalysis by permitting ATP to enter from one side of the ring and ADP to exit through the other side. Our analysis revealed no such solvent channels through the active sites of chains C and F (Fig. 4.C). The binding pockets of these chains are only accessible from the outside of the ring meaning that incoming ATP and outgoing ADP would have to use the same channel if the exchange were to occur in the same phase of the catalytic cycle. This, in turn, would mean that ADP would have to depart the pocket before ATP can enter, but our binding data suggest that this is not likely. Therefore, assuming that the ring positions are associated with phases in the ATP hydrolysis process, this leads us to propose the following assignments. Chains A and D have an open conformation with a "wide open" active site, ideally poised for ADP/ATP exchange. Ring positions associated with the post-hydrolysis stage are represented by molecules C and F, which assume an open conformation but their active sites are partially occluded by neighboring molecules in the hexamer. It is noteworthy that a small channel remains open in these two protomers. These channels are lined with combined conserved arginine wire residues from chains C and A on one hand and chains D and F on the other, which are thought to escort the phosphate from the active site after hydrolysis (Fig. 4.D) (Satyshur et al., 2007).

**TtPilB_{ATP} employs a symmetric counterclockwise rotary mechanism of ATP hydrolysis**

Based on our observations in the *TtPilB_{ATP}* structure, we can now propose a mechanism for how catalysis is coordinated across the ring-shaped PilB hexamer. When a protomer pair poised for ATP hydrolysis (molecules B and E) transitions to the post-hydrolysis state by hydrolyzing ATP, the adjacent protomer pair (molecules C and F) is moved from the post-hydrolysis phase into the wide-open position. At the same time, the protomer pair in the wide-
open conformation (molecules A and D) binds ATP to assume the active conformation. This coordination of the catalytic cycle has been described as a symmetric sequential rotary mechanism (Lyubimov et al., 2011). When viewed from the N-terminal side of the ring, catalysis proceeds in a counterclockwise direction because active protomers are positioned to the left of those in the post-hydrolysis state and wide-open molecules are positioned to the left of the molecules in the active state.

A symmetric rotary model is also consistent with the conformations observed in the low resolution AaPilT and VcGspE-Hcp1 structures (Lu et al., 2013; Robien et al., 2003; Satyshur et al., 2007), suggesting that these enzymes might coordinate catalysis in the same manner. In the AaPilT structure, the three conformations were described as “wide-open”, “active” and, “resting”. Remarkably, the “wide-open” protomers were located on the most peripheral part of the ring, while the “resting” molecules were positioned centrally (Satyshur et al., 2007). In the TtPilB_{ATP} structure, the “wide-open” molecules are located centrally, whereas the molecules displaying an open conformation with a partially blocked active site are located on the periphery. This observation is interesting in the context of the opposing functions of PilB and PilT as it may impact interaction with the inner membrane platform and how generated force is transferred. For example, in TtPilB_{ATP}, the major constriction of the central pore is created by two helices from the active molecules in the closed conformation, whereas in AaPilT, the constriction is formed by the resting protomers.

Whether or not a rotary mechanism is a universal mode for directional force generation in T4P/T2S ATPases remains to be determined. For example, structural studies of a PaPilT-AMP-PCP complex suggest that this enzyme might use an alternative mechanism (Mitic et al., 2010). The PaPilT hexamer also contained three distinct conformers, although the three molecules show much smaller differences. The authors proposed a “Ready, Active, Release"
model for ATP hydrolysis reminiscent of RecA (Story et al., 1992). The authors speculated that the drastic structural changes they had observed previously in AaPilT might have been caused by crystal packing forces and suggested instead that PilT enzymes utilize a stochastic mechanism of ATP-hydrolysis. The structure of a variant of the related SaFlaI protein in complex with ADP, showed a hexamer with two-fold symmetry, but complementary SAXS studies suggested that SaFlaI forms a symmetric hexamer in the presence of ATP (Reindl et al., 2013), a result consistent with either a stochastic or a concerted mechanism of catalysis. Analysis of the TtPilB_{ATP} structure here support a sequential and rotary model, suggesting that the conformations previously observed in the AaPilT structure may indeed be biologically relevant.

**Force generation in TtPilB_{ATP}**

Despite the different models for ATP catalysis by hexameric ATPases, a clear consensus has emerged for how events in neighboring active sites are linked to create directional motion. While the intramolecular NTD-CTD interfaces are characterized by dramatic conformational shifts, the intermolecular NTD-CTD interactions are conserved across the hexamer appearing to form rigid blocks. These covalently linked, rigid building blocks have been described previously for GspE, PaPilT and SaFlaI, as well as other hexameric AAA+ ATPases, (Lu et al., 2013; Misic et al., 2010; Reindl et al., 2013; Robien et al., 2003; Satyshur et al., 2007; Yamagata and Tainer, 2007). In TtPilB_{ATP}, this rigid interface (Fig. S3) is formed between the NTD of one protomer and the CTD of the next protomer, moving in a clockwise direction when the ring is viewed from the N-terminal side. The rigid intermolecular NTD\textsubscript{A}-CTD\textsubscript{B}, NTD\textsubscript{B}-CTD\textsubscript{C}, NTD\textsubscript{C}-CTD\textsubscript{D}, NTD\textsubscript{D}-CTD\textsubscript{E}, NTD\textsubscript{E}-CTD\textsubscript{F}, NTD\textsubscript{F}-CTD\textsubscript{A} blocks can be superimposed within an RMSD of less than 1 Å. Building on the assumption that each molecule will cycle through the three phases represented by different molecules in the TtPilB\textsubscript{ATP} structure, the
conformational changes caused by ATP binding and hydrolysis may be viewed in terms of the rotational and translational repositioning of these rigid units. The motion of each block is tied to events in the active sites of two neighboring molecules (Fig. 5), for example: molecules A (Fig 5A, yellow) and B (Fig. 5A, red) in case of the NTD$_A$-CTD$_B$ block. According to this model, molecule A will bind ATP, while molecule B hydrolyzes ATP at the same time. As a result of these coordinated events (Fig. 5, stage 1), the NTD$_A$-CTD$_B$ block will experience a clockwise, $47^\circ$ rotation away from the center of the ring around the axis shown in Fig. 5.B. This rotation would be accompanied by a translation of 5 Å along the same axis (Fig. 5.B, white arrow). Overall, the center of mass of the NTD$_A$-CTD$_B$ block will shift by about 12 Å. Following this transition, molecules A and B will be in the “poised for hydrolysis” and “post-hydrolysis” phases, respectively. Hydrolysis in molecule A and the shifting of molecule B into a “wide-open” position will rotate the NTD$_A$-CTD$_B$ block counterclockwise by $24^\circ$ accompanied by a 3 Å translation (Fig. 5, stage 2). In the third stage, molecule A is moved into the “wide-open” position, while molecule B binds ATP. These events are accompanied by a $28^\circ$ counterclockwise rotation of the NTD$_A$-CTD$_B$ block accompanied by a small, approximately 2 Å, translation (Fig. 5, stage 3). As a result of the observed two-fold symmetry, equivalent blocks across the ring would be predicted to undergo the same set of motions, while the movements of the other four molecules would be offset plus or minus one phase, resulting in the concurrent hydrolysis of two molecules of ATP per hexamer.

Finally, we sought to place our model into its biological context to understand how the motions associated with substrate binding and catalysis are coordinated within the PilB hexamer to drive T4P assembly. Previous work has demonstrated that the N-terminal region of the T2SS ATPase XspE interacts with the cytoplasmic membrane protein XpsL (Chen et al., 2005), suggesting that the corresponding region of PilB would engage in similar interactions and thus be positioned closest to the membrane. Recent work showed that a ring structure formed by
the membrane protein PilM binds to PilB (Bischof et al., 2016; Chang et al., 2016). In this context, it is interesting that the NTDs of PilB remain quite stably positioned during the projected structural rearrangements. This manifests itself in the observation that the six NTDs alone are evenly spaced displaying almost six-fold symmetry (Fig. 1). It is the CTDs that demonstrate the most dynamic positioning. In particular, helices α10 and α11 experience extreme repositioning during the projected movements (Fig. 5.). In the static structure, the C-terminal region of the protein reaches up to 23 Å below the central plane of the ring in chains C and F. In chains A and D, the C-terminal region is shifted upwards 8 Å to roughly 15 Å below the plane of the ring. In chains B and E, the C-terminal region is shifted upwards even further, to about 7 Å below the plane of the ring. The two upward shifts are accompanied by shifts toward the center of the ring. Collectively, the observed changes appear to produce scooping-type motion up and through the center of the pore, before resetting via a single, large, outward swing. The motion is simulated in a movie that has been provided as a supplement (Fig. S4). The notion that the force is transmitted through the central pore is consistent with what has been observed for other AAA+ ATPases (Lyubimov et al., 2011). Viewed from the NTD side, the coordinate structural changes would move the central pore around the ring in a counterclockwise direction.

The proposed concurrent and directional coupling of ATP hydrolysis and ATP binding across the PilB hexamer distinguishes this model from previously ones. Considering their distinct functions, it is quite possible that TtPilB and these other enzymes use similar molecular mechanisms of force generation but simply differ in how the various phases of catalysis are coordinated across the ring. This is also borne out by the proposed directions of the generated forces. In case of SaFlaI, protomer rotation was proposed to generate a force parallel to and across the plane of the hexamer (Reindl et al., 2013), whereas we propose that concurrent ATP-hydolysis and ATP-binding in the TtPilB hexamer drive T4P assembly by pushing upward
through the central pore of the hexamer, generating force perpendicular to the plane of the hexamer.

**Concluding Remarks**

Oligomeric AAA+ ATPases utilize a variety of ATP hydrolysis mechanisms to convert chemical energy into mechanical force (Liu et al., 2014; Lyubimov et al., 2011). A number of AAA+ ATPases have been crystallized and shown to form hexameric structures, including those of the T2SS ATPase GspE from *V. cholerae* and the PilT enzymes from *A. aeolicus* and *P. aeruginosa*. Many of these structures have also been solved as complexes with ATP analogs or ADP. Some structures displayed six-fold symmetry, indicating perhaps a stochastic or a concerted mechanism of hydrolysis, while others showed two-fold or three-fold symmetry more indicative of a sequential series of events. With the exception of *Tt* PilB, no structural studies have been reported for any PilB protein. *Tt* PilB itself had previously been studied via cryo-EM. However, the resolution of the model was low and modelling was based on the underlying assumption that PilB forms a hexamer with six-fold symmetry (Collins et al., 2013). The *Tt* PilB<sub>ATP</sub>-ATP<sub>•</sub>S complex structure, reported here, shows that the substrate bound enzyme ring has only two-fold symmetry, providing evidence for a symmetric, rotatory mechanism of hydrolysis. To our knowledge, this structure provides the first direct structural evidence for this type of coordination in a T4P/T2S ATPase.

Our model is also consistent with the recently published cryo–electron tomography study of the T4P from *Myxococcus xanthus* (Chang et al., 2016). In the model, PilB was proposed to be anchored to a ring formed of the protein PilM at the base of the T4P. PilB has recently been shown to interact with both PilM and PilC (Bischof et al., 2016). The model derived from the tomography work posits that opposing PilB subunits in the hexamer contact an asymmetric PilC dimer. ATP hydrolysis by PilB was proposed to rotate the PilC dimer, through a sequential
mechanism, which in turn interacts with the growing pilus and incorporates new pilin subunits via a “scooping” motion. The dimeric state of the PilC protein fits nicely with the observed two-fold symmetry of our structure, supporting the idea that the movements of the PilB rigid bodies up and through the central pore of the hexamer cause the PilC dimer to rotate and scoop pilin subunits out of the membrane and into the growing pilus fiber (Chang et al., 2016).

Experimental Procedures

Strain Construction

For the heterologous expression of TtPilB in E. coli, the coding region for the equivalent of M. xanthus PilB, corresponding to residues 146-890 from T. thermophilus HB27, was PCR amplified and cloned into the BamHI and HindIII sites of pQE30 (Qiagen). The resulting plasmid, pWB750, expresses TtPilB with an N-terminal 6×His tag. For protein purification, pWB750 was transformed into the E. coli Rosetta strain containing pREP4 (Qiagen).

Protein Purification

An overnight culture of TtPilB expression strain was sub-cultured into 6 L of Lennox broth, supplemented with 10 g/L dextrose, 100 mg/L ampicillin, and 50 mg/L kanamycin, and grown at 37° C, with shaking, until the culture reached an OD<sub>600</sub> of approximately 0.700. Protein expression was induced through the addition of IPTG at a final concentration of 1 mM for 18 hours at 18° C with shaking. Cells were harvested via centrifugation at 5,000 g for 20 minutes and cell pellets were re-suspended in 5 ml buffer A (25mM Tris-HCl pH 7.4, 500 mM NaCl, 25 mM imidazole, 10% glycerol, 1 mM DTT), supplemented with 0.3 mM PMSF, per gram of cells. Cells were lysed via sonication and insoluble cell debris was spun down at 117,000 g for 40 minutes at 4° C. The resulting supernatant was heat treated at 60° C for 20 minutes and the insoluble precipitate was removed via centrifugation at 10,000 g for 20 minutes. The recovered
supernatant was filtered and loaded on a 30 ml Ni-NTA column (Qiagen, Valencia, CA) pre-equilibrated with buffer A. Weakly bound proteins were removed by washing to baseline with buffer A, His-tagged protein was eluted with a linear gradient of imidazole to 250 mM. Peak fractions were pooled and thermolysin buffer (1X: 10 mM Tris-HCl pH 8.0, 200 mM NaCl, 2 mM CaCl$_2$) was added. PilB was then digested with thermolysin (Roche Molecular Biochemicals), at a final concentration of 0.375 mg/ml, for 1 hour at 37° C, to yield a truncated PilB product (PilB$_{ATP}$) of approximately 40 kD, containing residues 506 – 890, as verified by mass spectrometry (data not shown). The proteolytic reaction was stopped via the addition of 2 mM EDTA, and PilB$_{ATP}$ was separated from other digestion products via size exclusion chromatography using a HiPrep 26/60 Sephacryl S-200 HR column (GE Healthcare) and eluted in buffer C (25 mM Tris-HCl pH 7.4, 135 mM NaCl, 2.5% Glycerol, 2 mM TCEP pH 8.0). Protein purity of >95% was assessed via SDS-PAGE.

**Crystallization and Structure Solution**

PilB$_{ATP}$ was concentrated and ATPγS was added, resulting in a solution of 4.0 mg/ml PilB$_{ATP}$ and 5 mM ATPγS. The protein/ligand solution was then screened using commercially available screens. Initial crystals were optimized via the hanging drop method. Final crystals were obtained by mixing a solution of 4 mg/ml PilB$_{ATP}$ and 5 mM ATPγS with a solution of 0.4 M MgCl$_2$, 0.1 M Bis-Tris pH 6.5, and 10% PEG 3,350 in a 1:1 ratio. PilB$_{ATP}$ crystals were loop mounted and flash frozen in liquid nitrogen. Data was collected at beamline X-29A of the National Synchrotron Light Source (Brookhaven National Laboratory) using an ADSC Q315 CCD detector. Data was processed using iMOSFLM (Battye et al., 2011) and the CCP4 program suite (Winn et al., 2011). A homology model of $T$PilB$_{ATP}$ was constructed using EpsE (PDB ID: 1P9R) (Robien et al., 2003) as the template. For structure solution this model was broken into two parts, each representing one domain, and submitted as independent search
models for molecular replacement with the PHENIX suite (Adams et al., 2002). The conformations of chains B and D differed significantly from the search model and required manual building into the electron density map using COOT (Emsley and Cowtan, 2004). The structure was then refined using PHENIX, the relevant crystallographic statistics for which can be found in Table T1. The refined model was deposited in the protein data bank under the accession code 5IT5.

**Nucleotide binding assays**

To assess the ability of PilB<sub>ATP</sub> to bind nucleotides, a fluorescent assay utilizing 2'- (or 3')- O-(Trinitrophenyl) Adenosine 5'-Triphosphate (TNP-ATP) was employed. TNP-ATP has been shown to fluoresce strongly when located in the binding pocket of many proteins (Stewart et al., 1998). PilB<sub>ATP</sub> was prepared as described above, the final buffer (buffer C) was supplemented with 2 mM MgSO<sub>4</sub>, and titrated into a solution of 5 µM TNP-ATP. The competition experiment was performed by titrating ADP into a solution of 40 µM PilB<sub>ATP</sub> and 5 µM TNP-ATP. Data was collected using a Tecan Infinite F200 Pro fluorescent plate reader with an excitation wavelength of 405 nm and an emission wavelength of 550 nm. Data analysis and curve fitting was performed using XLFit.
Figure 1. The structure of ATPγS bound PilB<sub>ATP</sub>. (a) PilB<sub>ATP</sub> adopts an elongated hexamer displaying two-fold symmetry. The hexamer measures 135 x 90 x 75 Å. Analogous subunits share the same color. (b) Domain organization. All protomers consist of an N-terminal PAS-like domain attached to a CTD through a flexible linker. The CTD consists of ASCE, CMD, and C2D subdomains. ATPγS is bound at the interface between the NTD and CTD in all protomers.
Figure 2. Ligand binding pocket. (A) Two catalytic arginine residues (Arg582 and Arg595) are located >8 Å from the gamma phosphate in the open conformation (brown). (B) Following the
transition to the closed conformation, both Arg595 and Arg582 are capable of forming hydrogen bonds with the gamma phosphate of the bound nucleotide. In addition to the highly coordinated gamma phosphate, the closed conformation (cyan) also contains electron density which strongly suggests the presence of metal ion, modeled here as Mg$^{2+}$, that is absent in the four open protomers. (C)The movement of Arg582 and Arg595 is a result of a 55° NTD rotation during transition from the open to the closed conformation.
Figure 3. Conserved arginine residues. (A) Overlay of the NTDs from VcGspE (cyan, pdb code: 4ksr) and TtPilB (purple, pdb code: 5IT5). Both proteins contain three structurally conserved arginine residues on β4-6. (B) Overlay of the NTDs from AaPilT (green, pdb code: 2gsz) and TtPilB (purple). AaPilT contains only two of the arginine residues, lacking the arginine on β4 which correlates to Arg256 in TtPilB. (C) Sequence alignment demonstrating conservation of the three arginine residues on β4-6 amongst the PilB/GspE subfamily. Members of the PilT subfamily retain the β5 and β6 arginines (red) but lack the β4 arginine (blue). Residues are labeled such that the first number represents the residue in TtPilB and the second number represents the residue in the comparable protein.
Figure 4. Nucleotide binding. TNP-ATP fluorescences when bound to a protein. (A) A titration of PilB<sub>ATP</sub> into a solution containing 5 µM TNP-ATP resulted in a $K_d$ of 42 µM. A competition experiment using native ATP revealed the $K_d$ for the binding of native ATP to be statistically identical (data not shown). (B) A titration of ADP into a solution of 40 µM PilB<sub>ATP</sub> and 5µM TNP-ATP revealed that ADP directly competes for TNP-ATP for binding with a resulting $K_d$ of 80 µM.. Data is presented as the mean +/- standard deviation. (C) An analysis of solvent channels reveals large pores offering access to the ligand binding pocket in chains A and D. (D) Arginine wire residues predicted to escort $P_i$ from the binding pocket following hydrolysis.
Figure 5. Mechanism of force generation. Rigid bodies are formed between the NTD and CTD of adjacent subunits. The movement of these rigid bodies is predicted to generate the force necessary to assemble the pilus fiber. Panel A shows the make up of a rigid body (shown as gray shadow) formed by the NTD of subunit A (yellow) and subunit B (red). This rigid body is flanked by the ATP binding pockets for each subunit (ligand shown as blue spheres). As subunits A and B cycle through the three stages of catalysis (see text and Fig. S5 for more detail) they shift the position of the intersubunit rigid body through the combined actions of the events occurring in the adjacent binding pockets. Stage 1 represents a shift of the rigid body caused by the combined action of ATP binding and hydrolysis. Stage 2 represents a shift of the
rigid body caused by the combined action of ATP hydrolysis and “passive” repositioning. Stage 3 represents a shift of the rigid body caused by the combined action of “passive” repositioning and ATP binding. Panel B compares the position of the rigid body at the beginning of the stage (gray cartoon) to the position at the end of the stage (blue shadow) in relation to the overall hexamer. Two rigid bodies are shown since, as a result of the previously observed two-fold symmetry, it is expected that two rigid bodies will concurrently undergo the same movements, one on each side of the hexamer. Top (as if looking from the membrane) and side views (N-terminus at the top) are shown with the axis of rotation shown in blue. Alpha helices 10 and 11 are highlighted as they are primarily responsible for constricting the size of the central pore. More details are provided in the text.
Figure S1. Binding pocket density. In the open conformation (A) all electron density in the binding pocket is accounted for by sidechains and nucleotide. In the closed conformation (B) additional density exists which would suggest the presence of a catalytic magnesium ion (green sphere) as well as a catalytic water molecule (purple sphere). All electron density mesh is contoured at 1σ from the 2Fo-Fc map.
Figure S2. Coordination of bound nucleotide. Each panel visualizes the sidechain interactions with the bound nucleotide of the corresponding chain. The closed conformations represented by chains B and E demonstrate the highest degree of gamma phosphate coordination. Diagrams were generated using Ligplot (Laskowski and Swindells, 2011).
Figure S3. Subunit interface. Diagrams were generated using Dimplot which show the dimerization interface between chains A-B (A), B-C (B), C-D (C), D-E (D), E-F (E), and F-A (F).
**S4. Motion-Movie.** The movie shows the proposed concerted motions of the “rigid units” in the PilB hexamer as the protomers cycle through the stages of catalysis. The color coding marks these rigid blocks and NOT the discrete PilB molecules.

**Figure S5: Subunit transition cycle.** This figure shows the movements of a single subunit as it cycles through the three proposed mechanistic stages: ATP hydrolysis, “passive” repositioning, and ATP binding. The starting position is indicated in gray while the ending position is indicated in red. The movements shown here result in the rigid body movements described in figure 5. Figure 5 stage 1 results from a combination of the movements associated with ATP binding and ATP hydrolysis shown above. Figure 5 stage 2 results from a combination of the movements associated with ATP hydrolysis and “passive” repositioning shown above. Figure 5 stage 3 results from a combination of the movements associated with “passive” repositioning and ATP binding shown above.
Table 1. Crystallographic Data Collection and Refinement Statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>TtPilB&lt;sub&gt;ATP&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>65.29 - 2.65 (2.74 - 2.65)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Space group</td>
<td>P 2 2&lt;sub&gt;1&lt;/sub&gt; 2&lt;sub&gt;1&lt;/sub&gt;</td>
</tr>
<tr>
<td>Unit cell dimensions A, B, C (Å)</td>
<td>107.6, 133.6, 208.4</td>
</tr>
<tr>
<td>Total reflections</td>
<td>174,766 (16,469)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>87,967 (8,248)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>2.0 (2.0)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.95 (99.73)</td>
</tr>
<tr>
<td>&lt;I/σ(I)&gt;&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5.83 (0.83)</td>
</tr>
</tbody>
</table>

| Refinement                                           |                                      |
| Resolution (Å)                                       | 65.29 - 2.65 (2.74 - 2.65) |
| R<sub>merge</sub><sup>c</sup>                         | 0.1243 (0.8679) |
| CC<sub>1/2</sub><sup>d</sup>                          | 0.989 (0.409) |
| R<sub>work</sub>/R<sub>free</sub><sup>f</sup>         | 0.2336/0.2780 |
| RMS(bonds)                                            | 0.008 |
| RMS(angles)                                           | 0.91 |
| Average B-factor (Å<sup>2</sup>)                     | 40.9 |

<sup>a</sup>The values in parentheses relate to the highest resolution shell from 2.74 – 2.65 Å.

<sup>b</sup><I>/σ(I) reaches 2.0 at approximately 2.8 Å

<sup>c</sup>R<sub>merge</sub> = Σ|I| - <I>/ΣI, where I is the observed intensity and <I> is the average intensity obtained from multiple observations of symmetry-related reflections after the rejection of significant
outliers.

$^{\text{d}}\text{CC}_{1/2} = \text{Pearson correlation coefficient between random half-datasets}$

$^{\text{e}}R_{\text{work}} = \Sigma | | F_o | - | F_c | / \Sigma | F_o |$

$^{\text{f}}R_{\text{free}}$ defined by Brunger

Acknowledgments

Funding was provided by a grant from the National Science Foundation (MCB-1417726) to Z.Y. and F.D.S. Funding for data collected at beamline 29 NSLS is provided by Department of Energy Office of Biological and Environmental Research and NIH. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References


Paranchych, W. Biochemical Studies on Pili Isolated from Strain PAO. Canadian J. Microbiology, 25, 1175-1181.


Chapter 3: A helix cracking mechanism regulates the interaction between the sensor histidine kinase GacS and its negative regulator RetS in *Pseudomonas aeruginosa*

JORDAN M. MANCL, W. KEITH RAY, RICH F. HELM, AND FLORIAN D. SCHUBOT*

Virginia Polytechnic Institute and State University, Department of Biological Sciences, Derring Hall (MC 0910), 926 West Campus Drive, Blacksburg, Virginia 24061

* Corresponding author:

Virginia Polytechnic Institute and State University
Department of Biological Sciences
5002 Derring Hall (MC 0910)
926 West Campus Drive
Blacksburg, Virginia 24061
E-mail: fschubot@vt.edu
Phone: (540) 231-2393
Fax: (540) 231-4043

Author contributions: J.M.M. designed and conducted the research, analyzed data, and wrote the manuscript. W.K.R. performed the LC-MS work. R.F.H. supervised the LC-MS work. F.D.S designed and supervised the research, assisted with data analysis and wrote the manuscript.
Abstract

Sensor kinases are central regulators of gene expression utilized by all prokaryotic organisms. Recent discoveries have challenged the classical paradigms that defined how these enzymes modulate gene expression patterns in response to extracellular stimuli. Here we describe the structural and functional analysis of the kinase region of RetS from *Pseudomonas aeruginosa*, which functions as a direct inhibitor of the canonical sensor kinase GacS. Our initial analysis tested and disproved a longstanding model for the RetS/GacS binding interaction that posits the formation of a RetS/GacS heterodimer. We subsequently uncovered unique structural features within the RetS kinase region that suggest RetS uses an intriguing helix cracking mechanism to regulate its interaction with GacS. Follow-up *in vivo* studies experimentally corroborated the critical role that a structurally dynamic helical region in RetS plays in mediating interaction with GacS. Although RetS itself is no longer a functional kinase, yet the observed conformational changes, occurring in immediate proximity to the catalytic histidine, may well represent a new mechanism whereby classical sensor histidine kinases control autophosphorylation through extracellular signaling.
Introduction

Bacteria utilize phosphorylation-dependent signal transduction systems to regulate diverse cellular behaviors including the coordinated expression of major virulence mechanisms\textsuperscript{1-3}. The opportunistic pathogen \textit{Pseudomonas aeruginosa} uses more than sixty of these systems to orchestrate its remarkable adaptability during infection\textsuperscript{4}. In humans, \textit{P. aeruginosa} is a major cause of acute and chronic nosocomial infections\textsuperscript{5,6}. Acute infections are characterized by severe cytotoxicity caused by secretion of effector proteins via the type three secretion system (T3SS)\textsuperscript{7}. On the other hand, the chronic pulmonary \textit{P. aeruginosa} infections, which commonly befall patients with cystic fibrosis, result from the formation of a bacterial biofilm in host lungs\textsuperscript{8,9}. During biofilm formation, expression of virulence factors associated with acute infections, such as the T3SS, are down-regulated whereas exopolysaccharide production is strongly stimulated. This transition between acute and chronic infection modes is governed by the GacS/GacA phosphorelay system\textsuperscript{10-13}. Here, the hybrid sensor kinase GacS is composed of an N-terminal periplasmic sensory domain that is coupled to the cytoplasmic histidine-kinase (HK) region via a HAMP domain, followed by a receiver (REC) domain and a C-terminal histidine phosphotransfer (Hpt) domain\textsuperscript{14}. Signaling follows the pattern typically observed in phosphorelay systems: following ATP binding and autophosphorylation at a conserved histidine in the HK region, sequential phosphotransfer reactions channel the phosphate to an aspartate residue in the REC domain and from there to a histidine in the Hpt domain, which mediates the final transfer to an aspartate residue on the cognate response regulator GacA. Phosphorylation stimulates GacA activity, resulting in the increased expression of two small RNAs: RsmY and RsmZ, which function to
sequester the post-transcriptional regulator RsmA\textsuperscript{10,11,13,15,16}. A member of the CsrA family, RsmA impacts the expression of more than 600 different genes in \textit{P. aeruginosa}\textsuperscript{17,18}. Crucially, RsmA was found to inhibit biofilm formation while promoting expression of the T3SS\textsuperscript{17}. Therefore, GacS ultimately upregulates biofilm-associated processes and concomitantly downregulates mechanisms associated with acute infections. To effectively modulate these cell-wide responses, GacS is subject to intracellular regulation involving fascinating direct interactions the sensor signaling kinases LadS and RetS\textsuperscript{19,20}. The discovery of complex networked interactions between GacS and two other sensor histidine kinases fundamentally challenged our understanding of how bacterial two-component and phosphorelay systems operate. Prior, it had been assumed that evolution selects against crosstalk to maintain signaling specificity\textsuperscript{21}. However, mounting evidence paints a more complex picture suggesting that webbed interactions between signaling kinases may be rather common\textsuperscript{22}. The interaction network surrounding GacS constitutes an ideal model system for studying nontraditional mechanisms of regulation as the enzyme is subject to both positive and negative regulation by LadS and RetS, respectively. LadS enhances GacS signaling by shuttling additional phosphate groups into the GacS/Rsm pathwa via the GacS Hpt domain\textsuperscript{20,23}. RetS, on the other hand, inhibits the activity of GacS independent of RetS autophosphorylation\textsuperscript{10,13,19,24}. Because GacS autophosphorylation likely occurs \textit{in trans} within a GacS homodimer, RetS had originally been proposed to inhibit GacS by simply forming a RetS/GacS heterodimer\textsuperscript{19}. A recent study uncovered however, that RetS/GacS interactions are vastly more complex, involving two additional mechanisms by which RetS inhibits GacS signaling, beyond the previously established direct
interaction which was known to disrupt autophosphorylation \textsuperscript{25}. Competing with the first phosphorelay step, the C-terminal receiver domain of RetS was found to siphon away phosphates from the GacS\textsubscript{HK} region. In a separate mechanism the RetS\textsubscript{HK} region was shown to act as phosphatase that dephosphorylates the GacS REC domain, thus inhibiting the second step in the phosphorelay reaction. While we have a reasonable understanding of the molecular basis for the first two mechanisms because they derive from classic phosphotransfer reactions between HKs and receiver domains, open questions remain as to how RetS and GacS bind through their HK regions and how the this interaction might be regulated. Likely, this third mechanism also impacts the kinetics of the other two inhibitory processes, as formation of a stable RetS/GacS complex should dramatically increase the efficiency of the phosphotransfer reactions. Although all three mechanisms are required under physiological conditions, overexpression of a RetS(T428A,HDA) variant that lacks the phosphotransfer activities but does still bind to GacS can support \textit{P. aeruginosa} virulence in a \textit{Galleria mellonella} virulence model \textsuperscript{25}.

In the present study, we provide compelling experimental evidence that RetS does not, as originally proposed, prevent GacS homodimerization, but blocks GacS autophosphorylation through the formation of a higher order complex. We were able to demonstrate that the HK regions of the two proteins interact via their DHp domains. Our structural analysis of the RetS HK domain, in conjunction with \textit{in vivo} and \textit{in vitro} functional studies, supports a new model wherein the RetS/GacS interaction is modulated by the reversible unfolding of a short helical region in the RetS DHp domain. The conformational plasticity of a loop region in RetS, which mediates ATP binding in canonical HKs appears to modulate the equilibrium between the folded and unfolded states. We propose that these conformational changes are triggered by ligand binding to the RetS periplasmic sensory domain. Although RetS itself appears to have lost
kinase activity, the observed helix-cracking mechanism could represent a widely-employed mode for coupling periplasmic ligand binding to enzymatic activity in canonical sensor histidine kinases as it involves the ATP binding lid region and the Dhp domain helix containing the catalytic histidine\(^{26}\).

**RESULTS**

**The DHp domains of RetS and GacS interact without breaking the GacS homodimer**

Initially, we considered two possible mechanisms for how RetS binding might interfere with GacS \textit{trans}-autophosphorylation. The original, longstanding model posits the formation of a RetS/GacS heterodimer through interactions between their respective DHp domains, essentially mimicking homodimeric HKs\(^{19}\). However, it seemed equally plausible that RetS interferes with GacS activation by forming crosswise interactions with the GacS DHp or CA domains using its own DHp and CA domains\(^{27}\). Either mechanism could have possibly derived from the interactions commonly occurring in HKs during autophosphorylation. In order to evaluate both models, we examined RetS/GacS interactions using a bacterial adenylate cyclase two-hybrid (BACTH) assay\(^{28}\). GFP was added as a C-terminal fusion to all RetS variants to monitor their stable expression within the cell. All constructs expressed stably and controls gave the expected results. Consistent with the originally proposed model, the DHp domains of RetS and GacS were indeed found to interact and no crosswise interactions between the CA and DHp regions were observed (Fig. 1A). In order to directly assess whether
or not RetS binding disrupts GacS dimerization, we developed a Förster Resonance Energy Transfer (FRET)-based binding assay. The cytoplasmic region of GacS (GacS\textsubscript{cyt}) was purified and fluorescently labeled with either Alexa Fluor 488 (GacS\textsubscript{cyt}-488) or Alexa Fluor 555 (GacS\textsubscript{cyt}-555). Titration of GacS\textsubscript{cyt}-555 into a solution of GacS\textsubscript{cyt}-488 resulted in the formation of dually-labeled GacS\textsubscript{cyt} dimers as evident from the characteristic quenching of GacS\textsubscript{cyt}-488 fluorescence (Fig. 1B). Having established suitable conditions for obtaining a strong FRET signal, we used this GacS\textsubscript{cyt}-555:GacS\textsubscript{cyt}-488 ratio (25 pM:250 fM), in our subsequent competition experiments. We reasoned that, if the original model were correct, titration of unlabeled RetS\textsubscript{cyt} to the dually-labeled GacS\textsubscript{cyt} dimer should restore GacS\textsubscript{cyt}-488 fluorescence when the heterodimeric RetS\textsubscript{cyt}/GacS\textsubscript{cyt} complex forms. Yet, the addition of unlabeled RetS\textsubscript{cyt} had no measurable impact on the FRET signal (Fig. 1C), clearly demonstrating that RetS\textsubscript{cyt} does not break up the GacS\textsubscript{cyt} homodimer. To confirm that the lack of homodimer disruption did not result from a lack of RetS/GacS interaction under experimental conditions, we successfully titrated GacS\textsubscript{cyt}-488 with Alexa Fluor 555 labeled RetS\textsubscript{cyt} (RetS\textsubscript{cyt}-555) to verify interaction between GacS\textsubscript{cyt}-488 and RetS\textsubscript{cyt}-555 (Fig. 1D). A titration of the dually-labeled GacS\textsubscript{cyt} dimer with unlabeled GacS\textsubscript{cyt} served as positive control to show that the FRET signal is sensitive to the competitive targeting of GacS dimerization interface (Fig. 1E). Collectively, the FRET studies offer compelling evidence that RetS and GacS interact to form a higher order complex, possibly a dimer of dimers involving novel interactions between their DHp domains, rather than a heterodimer as the original model had proposed.
The crystal structure of the RetS kinase region reveals a uniquely asymmetric dimer interface and loss of the canonical ATP binding pocket

In canonical sensor HKs, residues pivotal for ATP binding and hydrolysis cluster in six boxes named H, N, G1, G2, G3, and F for their conserved features. Remarkably, many of these box residues are not conserved in RetS, resulting in an HK region that lacks kinase activity but has retained phosphatase activity. In order to determine if RetS has evolved distinct structural features relating to its unique inhibitory interaction with GacS, we solved the crystal structure of entire kinase region of RetS (RetS$_{HK}$), encompassing residues 416-649 at 2.6 Å resolution.

Overall, RetS$_{HK}$ assumes a canonical HK fold containing an N-terminal DHp domain connected to a C-terminal CA domain via a short linker region (Fig. 2). The CA domain adopts the classical α/β sandwich fold, with three α-helices packed against five antiparallel β-strands. The diverging sequences in the N, G1, G2, and F box regions indeed manifested structurally in the putative ATP binding pocket of the CA domain. Although RetS$_{HK}$ was crystallized in the presence of ATPγS, no ligand electron density was observed in the predicted ATP binding site. In fact, the binding pocket appears to have collapsed especially around the region that usually binds to the beta and gamma phosphate units where the N-terminal end of helix α6 has been pushed into the pocket (Fig. S1), demonstrating, not only, that RetS is no longer a functional kinase, but also that nucleotide binding also plays no role in regulating the biological activity of RetS.

The RetS DHp region consists of two large anti-parallel α-helices with α1 containing the highly conserved “catalytic” histidine, H424. As expected, RetS$_{HK}$ is dimeric with the
DHp domain helices forming a four-helix bundle at the interface. These four helices are arranged in a right-handed manner, which, according to recent predictions regarding HK autophosphorylation propensities, would suggest that any potential CA and DHp domain interactions would occur in trans\textsuperscript{30,31} (Fig. S2). The overall asymmetry of the dimer is readily visualized when the two DHp domains are superimposed. Here, the CA domains are rotated at a 34° angle (Fig. S3). Conformational flexibility is essential and well-documented among HKs as autophosphorylation requires dramatic and distinctly asymmetric domain motions\textsuperscript{27,31,32}. However, the CA domains and connecting loop regions between the CA and DHp domains are well-ordered and structurally identical in all four RetS\textsubscript{HK} dimers in the asymmetric unit (Fig. S4) suggesting a rather rigid conformation that differs from previously published histidine kinase structures\textsuperscript{33,34}. RetS\textsubscript{HK} is also distinctly asymmetric at the dimerization interface (Figs. 2D and S5). In one molecule, DHp domain helix α1 starts at residue E417 and is kinked at position S423 just adjacent to the conserved histidine H424. In the second molecule, the N-terminus of helix α1 has shifted to H424, while the E417-S423 region is now unfolded and points away from the dimerization interface. This displacement is accompanied by the movement of a loop region, encompassing residues 588-603, from the other RetS\textsubscript{HK} molecule in the dimer. Connecting helices α5 and α6, this loop region is referred to as the ATP lid loop in sensor HKs and typically functions to seal ATP in the binding pocket of the CA domain. In the absence of bound nucleotide, the ATP lid region is commonly disordered\textsuperscript{29}. In the RetS\textsubscript{HK} dimer, despite the absence of nucleotide, the lid region assumes two well-defined conformations. On one side, the ATP lid residues 599-603 form a short helix that packs against the remainder of the CA domain (Figs. 2D and S5).
On the other, the ATP lid loop has swung toward the dimer interface by about 10 Å and displaced the N-terminal seven residues of α1. Here, loop residues 593-596 form a short helical turn that appears to compensate for the unfolded region of α1. All four RetS₇HK dimers in the asymmetric unit of the crystal display the same ATP lid loop movements (Fig. S4). Therefore, we hypothesized that the partial unfolding of α1 and associated movements of the adjacent ATP lid loop may have mechanistic significance, perhaps constituting a means by which binding of an extracellular ligand to the RetS sensory domain relieves inhibition of GacS.

**RetS₇HK α1 residues 416-423 are required for inhibiting GacS signaling**

The regulated unfolding of α-helices in response to signal sensing, or helix cracking has been shown to be an effective means of controlling enzymatic activity, primarily in adenylyl and diguanylate cyclases. In order to determine if helix cracking regulates RetS/GacS binding, two RetS variants were created and tested. In RetS₇HKΔlid, 16 residues of the ATP lid loop were replaced with a nine residue glycine linker (RetS₇HKΔ588-603::gly9). The rationale for this construct is based on the assumption that the unfolding of α1 requires a compensatory motion of the ATP lid loop to retain the stability of the RetS dimer interface. The fact that the RetS DHp domain, when expressed in the absence of the CA domain, readily binds to GacS suggests that partial unfolding of α1 is not required for GacS binding (Fig. 1A). Therefore, a RetS₇HKΔlid dimer was expected have two binding sites for GacS. For the second variant, RetS₇HKΔhelix, residues 416-423 from the N-terminus of the DHp α1 helix were replaced
with glycine residues (RetS<sub>HKΔ416-423::gly8</sub>). In the RetS<sub>HKΔhelix</sub> dimer, both α1 helices lack the helical section preceding the kink. Therefore, this variant was predicted to represent a functional state of RetS where it no longer binds to GacS. The impact of the introduced mutations was examined in two assays. First, direct interactions were probed with the already established BACTH assay (Fig. 3). Second, to determine the impact of the mutations on RetS function in vivo, using the crystal violet biofilm assay<sup>36</sup>, the same mutations were also constructed in full-length RetS and tested for their ability to rescue a retS deletion in <i>P. aeruginosa</i> PAK. As before, RetS BACTH constructs were expressed as GFP fusion proteins to visualize stable protein expression. Consistent expression of all variants in the PAK<sup>ΔretS</sup> strain was corroborated using LC-MS (Fig. S6). As anticipated, RetS<sub>HKΔlid</sub> was found to interact with both GacS<sub>HK</sub> and GacS<sub>Dhp</sub> (Fig. 3). In the crystal violet biofilm assay, RetS<sub>Δlid</sub> was also found to fully complement a retS deletion in PAK (Fig. 4). In fact, the strain expressing the RetS<sub>Δlid</sub> variant produced significantly less biofilm than a strain expressing wild-type RetS, a finding consistent with the prediction that removal of the ATP lid region creates two viable interaction interfaces per RetS dimer instead of one.

RetS<sub>HKΔhelix</sub>, on the other hand, did not interact with either GacS<sub>Dhp</sub> or the full GacS<sub>HK</sub> in the BACTH assay, but retained its ability to dimerize with wild-type RetS<sub>HK</sub> (Fig. 3). Excitingly, the RetS<sub>Δhelix</sub> variant also failed to complement a retS deletion in the biofilm assay. The construct retained some residual activity, suggesting that disruption of RetS<sub>HK</sub>-GacS<sub>HK</sub> binding interactions does not fully abolish the other inhibitory mechanisms relating to RetS phosphatase activity. This observation is not surprising as this mutation should not affect regions associated with phosphotransfer-mediated
inhibition of GacS\textsuperscript{25}. Because the construct failed to stably express in the PAK strains, we were unable to establish if RetS\textsubscript{DHp} alone, in absence of the CA domain, is sufficient for any degree of GacS inhibition, however, it was noted that expression of a RetS construct lacking all resides C-terminal to the HK region was unable to fully complement a \textit{retS} deletion (data not shown), a finding consistent with previously published results\textsuperscript{24,25}.

Collectively, the binding and functional data support our hypothesis that the observed partial unfolding of the N-terminal DHp helix in the RetS\textsubscript{HK} dimer structure plays a critical role in a helix cracking mechanism responsible for regulating the RetS-mediated inhibition of GacS, perhaps in response to periplasmic ligand binding.

\textbf{DISCUSSION}

Sensor kinases are central players in the regulation of gene expression in prokaryotic organisms. We have gained a deep understanding into the phosphate transfer mechanisms underlying two-component and phosphorelay systems. Important recent progress has shed light into the molecular basis for the reversal of some of these reactions when phosphatase activity outstrips kinase activity\textsuperscript{37}. The interactions of GacS with RetS and LadS were discovered some time ago and represent the paradigm for networked interactions between signaling HKs, but only very recently have we begun to understand the molecular basis for these interactions\textsuperscript{19,20,25}. The phosphorylation of the GacS Hpt domain through the LadS REC domain constitutes an important precedent for crosstalk between kinases that was previously believed to be improbable to occur\textsuperscript{20,21,23,38}. Even more recently, the Porter group has elegantly demonstrated the
complexity of the RetS-mediated inhibition of the GacS/A system involving three distinct mechanisms all targeting distinct steps of the phosphorelay\textsuperscript{25}. The first two mechanisms, the dephosphorylation of GacS\textsubscript{HK} by the REC-2 domain of RetS and the dephosphorylation of the GacS REC domain by the HK region of RetS, are derived from classical interactions observed in phosphorelay systems. However, there exists no known mechanistic equivalent for the third mechanism, mediated by the direct binding interaction between the HK regions of RetS and GacS. Our FRET binding studies convincingly show that the cytoplasmic regions of RetS and GacS do not form heterodimers. This finding is also supported by the observation that the RetS\textsubscript{HK\textDelta helix} variant is able to homodimerize yet unable to interact with GacS in our two-hybrid assays, as this demonstrates that the interfaces involved in RetS dimerization are distinct from those mediating GacS binding. The observed binding between the DHp domains of RetS and GacS does suggest that a single RetS dimer could interfere with GacS autophosphorylation by blocking access to the catalytic histidine in the bound DHp domain. At the same time, RetS could also sterically obstruct movement of the GacS CA domain to prevent formation of the catalytic pocket, and subsequent autophosphorylation, in the other GacS subunit. We are now pursuing a structural model of the complex to determine how exactly RetS\textsubscript{HK} prevents GacS autophosphorylation.

At this point, the other major unresolved question is how the RetS/GacS interaction is regulated in response to outside stimuli. The documented role of the N-terminus of α1, immediately upstream of H424, in mediating interaction with GacS suggests that such regulation may occur via the transmembrane and sensory regions of
RetS, which precede this dynamic region. In such a scenario, the binding of a currently unidentified ligand to the sensory domain of RetS would disrupt GacS binding by unfolding the N-terminal region of RetS α1. Conformational changes in the RetS ATP lid and concomitant movement of the CA domain would yield a symmetric RetS dimer with two partially unfolded DHp helices. GacS autophosphorylation activity would lead to an increase in the expression of virulence mechanisms associated with chronic infections. We propose that the movements of the ATP lid loop, associated with α1 unfolding, function to stabilize the protein following transition. Currently, only our structural data and the modest, but significant, difference in biofilm levels between strains complemented with wild-type RetS versus the RetSΔlid variant, provide supporting evidence for such a model. Likely the picture is significantly more complex. GacS also possess a periplasmic sensory domain, which could modulate RetS/GacS interactions by inducing conformational changes in the GacS-HK region. Additional control points may involve the modulation of RetS, GacS, and LadS expression levels or localization. Moreover, RetS has also been shown to be inhibited through interactions with another sensor HK PA1611; and the HptB protein is thought to reduce RetS inhibitory activity by feeding additional phosphate groups into the pathway.

Outside its specific role in the RetS/GacS system, the conspicuous location of the structurally dynamic region just N-terminal of the catalytic histidine and the involvement of the ATP lid loop suggests that the proposed helix-cracking mechanism may have originally evolved as means for regulating autophosphorylation in other members of the 7TMR-DISMED2 family of sensor kinases, which RetS belongs to, such as LadS. LadS is closely related to RetS, displaying 44% sequence similarity, and
possesses demonstrable autokinase activity. Additionally, LadS has been found to respond to changes in Ca\(^{2+}\) levels via its periplasmic sensory domain\(^{11,12,23}\).

The diversity of these signaling systems is rapidly becoming recognized as an increasing number of HKs which possess biological activity beyond autophosphorylation continue to be discovered\(^{22,37}\). \textit{P. aeruginosa} and more specifically, the Gac/Rsm pathway, has emerged as a model system for studying atypical regulation of histidine kinases, and this study seeks to lay a foundation for our understanding of this type of emerging regulatory model.

In summary, the work presented here seeks to redefine our understanding of the unique regulatory interaction between RetS and GacS. RetS-mediated inhibition of GacS functions through three different mechanisms\(^ {25}\). The present study focused on the molecular basis of the direct binding interaction. We examined the molecular basis for the novel direct binding interactions between the HK regions of RetS and GacS that inhibit GacS autophosphorylation. Rather than the formation of an inactive heterodimer, RetS and GacS were found to form a higher order complex mediated by the DHp domains. A structurally dynamic region within the RetS DHp domain was found to be critical for GacS binding and possibly play a key role in regulation of the interaction (Fig. 1). The role of the RetS CA domain remains to be established. Because a RetS construct lacking the CA domain did not express stably in the PAK strain, we could not establish whether this domain is required for inhibiting GacS autophosphorylation. The RetS CA domain was observed to swing 34° toward the RetS dimer interface as its ATP lid region replaces part of α1 of the other RetS molecule at the interface. This movement may well cause the release of GacS to facilitate biofilm formation. In order to
directly test this model and, in general, the question as to whether or not the proposed helix cracking mechanism controls the RetS-mediated inhibition of GacS the elusive ligand of the RetS periplasmic domain will need to be found.

**ONLINE METHODS**

**Cloning**

Expression plasmids for the cytoplasmic regions of RetS (pQE60RetS<sub>cyt</sub>) and GacS (pQE60GacS<sub>cyt</sub>) were generously gifted by Steven Porter (University of Exeter). All cloning primers are listed in table S1. The kinase region of RetS, corresponding to amino acids 416-649, was amplified from pQE60RetS<sub>cyt</sub> and inserted into the expression vector pDestHisMBP via pDONR201 using the Gateway cloning system (ThermoFisher). The resulting plasmid was inserted into BL-21(DE3) RIL cells for expression.

The RetS and GacS BACTH constructs were synthesized by GENEWIZ. Briefly, the region encoding residues 270-509 of GacS was inserted into pUT18C using the XbaI and KpnI sites. Additionally, the region encoding residues 416-649 of RetS was fused to the N-terminus of eGFP via a linker region (GSAGSAAGSGEF)<sub>40</sub> and inserted into the XbaI and KpnI sites of pKT25. During this process, a silent SacI site was inserted into the beginning of the eGFP gene to simplify future mutagenesis. The RetS<sub>HKΔlid</sub> and RetS<sub>HKΔhelix</sub> variants were generated via single overlap extension (SOE) PCR using the primers shown in table S1 and inserted via the SacI and XbaI sites. The RetS and GacS Dhp constructs were created via site-directed mutagenesis (SDM).
using the Quikchange SDM kit (Agilent Technologies) and the primers listed in table S1. RetS and GacS CA domain constructs, as well as the eGFP control, were generated via standard cloning procedures using the primers listed in table S1 and inserted into their respective plasmids as described for the full kinase regions. For the BACTH studies, the kinase region of RetS was cloned into the KpnI and XbaI sites of pUT18C. Full length RetSΔlid and RetSΔhelix constructs were synthesized by Genewiz and expressed under the control of an arabinose-inducible promoter from pHERD20T.

**Protein Purification**

A 6 L culture of BL-21(DE3)RIL pDESTHisMBP-RetS\textsubscript{HK}, in Lennox broth supplemented with 10 g/L glucose, was grown at 37° C, with shaking at 225 rpm. At an OD\textsubscript{600}=0.7, protein production was induced, via the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), for 18 hours at 18° C, and shaking was slowed to 125 rpm. Cells were harvested via centrifugation and pellets were resuspended in 5 ml buffer HK\_A, supplemented with 0.3 mM phenylmethanesulfonyl fluoride (PMSF), per gram of cells. Cells were lysed via sonication and insoluble cell debris was spun down at 117,000 g for 40 min at 4° C. Soluble HisMBP-RetS\textsubscript{HK} was then purified via Ni-NTA affinity chromatography and eluted into buffer HK\_B. Fusion protein was then digested with TEV protease\textsuperscript{42}. Following cleavage, and dialysis to remove excess imidazole, tag-less RetS\textsubscript{HK} was separated from HisMBP and uncleaved fusion product via Ni-NTA affinity chromatography. RetS\textsubscript{HK} was further purified via anion exchange chromatography using a HiTrap Q HP column (GE Life Sciences) pre-equilibrated with buffer HK\_C and eluted with a linear gradient of buffer HK\_D. The final purification step
consisted of size exclusion chromatography using a Superdex 26/60 S200 column (GE Life Sciences). Protein was eluted into buffer HK_E. Sample purity of >98% was assessed via SDS-PAGE. Buffer composition can be found in Table S2.

Selenomethionine-substituted RetS<sub>HK</sub> was produced by suppressing methionine biosynthesis through the addition of exogenous selenomethionine into growth medium. Initially, a 1 L overnight culture of the expression strain was grown in LB medium supplemented with 100 mg/L ampicillin and 30 mg/L chloramphenicol. The cells were washed twice in 100 mL of M9 selenomethionine growth media (Shanghai Medicilon Inc.). The cells were resuspended in 100 mL of M9 selenomethionine growth media and used to inoculate six 1 L cultures containing M9 selenomethionine growth media and antibiotics. These cultures were grown with shaking at 37°C and protein production was induced with 1 mM IPTG once the cultures reached an OD<sub>600</sub> of 0.2. Following the addition of IPTG, the temperature was lowered to 18°C for 18 hours. Cells were harvested and the protein was purified following the same protocol as the native protein.

The cytoplasmic regions of RetS and GacS were expressed in JM109(DE3) cells from a pQE60 expression vector following the protocol used for RetS<sub>HK</sub>. Cell harvesting and lysis also followed the RetS<sub>HK</sub> protocol, however, RetS<sub>cyt</sub> and GacS<sub>cyt</sub> were resuspended in buffer Cyt_A rather than HK_A. RetS<sub>cyt</sub> and GacS<sub>cyt</sub> were purified using a Ni-NTA column pre-equilibrated in buffer Cyt_A and eluted into buffer Cyt_B. The final purification step consisted of size exclusion chromatography using a Superdex S200 26/60 column (GE Life Sciences). Proteins destined to be fluorescently labeled were
eluted into labeling buffer, otherwise proteins were eluted into stabilization buffer.

Protein purity was assessed via SDS-PAGE. Buffer compositions are listed in table S2.

**FRET Binding Assays**

Protein samples were concentrated to 10 mg/ml and fluorescently labeled with amine-reactive Alexa Fluor 488 or Alexa Fluor 555 (Life Technologies) as directed by the manufacturer. Unbound fluorophore was removed via buffer exchange chromatography, using a HiPrep 26/10 desalt column (GE Life Sciences), into stabilization buffer (Table S2). Protein concentrations were determined using absorbance at 280 nm and verified via densitometry comparison to known standards following SDS-PAGE. The degree of labeling was determined as directed by the manufacturer. In each case, labeled proteins were found to contain >1 mole fluorophore per mole protein. Fluorescent readings were taken using a Tecan M200 fluorescent plate reader using an excitation wavelength of 430 nm and an emission wavelength of 522 nm. Controls for background fluorescence consisted of all experimental components present in the well minus the Alexa Fluor 488 labeled protein. Background-subtracted data was then normalized to the positive ($F_{\text{max}}$) and negative ($F_{\text{min}}$) controls.

**Crystallization and Structure Solution**

$\text{RetS}_{HK}$ crystals were grown via the hanging drop method by mixing a solution containing 10 mg/ml $\text{RetS}_{HK}$ (in HK_E buffer) and 5 mM ATPyS (50 mM stock, prepared in water) with the crystallization condition at a 4:1 ratio and suspending the drops over a reservoir solution consisting of 70% crystallization condition, 30% water. The optimized crystallization solution consisted of 2.7 M NaCl, 90 mM HEPES pH 7.5, 9 mM CoCl$_2$
hexahydrate. Crystals grew over the course of one month at 4°C. Once crystal growth abated, the reservoir solution was increased to 100% crystallization condition via three step-wise increases (80%, 90%, 100%) over three weeks to dehydrate the crystals. The dehydrated crystals were loop mounted and flash frozen. SeMet-derived protein crystals were grown under identical conditions.

Anomalous x-ray diffraction data for the crystals of the selenomthione-substituted sample was collected at beamline 17-ID-1 (AMX) of the National Synchrotron Light Source (NSLS-II) at Brookhaven National Laboratory. Data was processed using FastDP\textsuperscript{43} in combination with the CCP4 program suite\textsuperscript{44}. Data for crystals containing the native protein was collected at LRL-CAT beamline 31-ID at Argonne National Laboratory. Data was processed using the CCP4 program suite. The structure was solved using MAD phasing and refined against native data to 2.6 Å using the Phenix program suite\textsuperscript{45}.

Additional electron density observed in the map between histidine side chains was interpreted as Co\textsuperscript{2+} ions, based on the composition of the crystallization condition and the documented propensity of histidines for coordinating Co\textsuperscript{2+} ions\textsuperscript{46,47}. Several of these Co\textsuperscript{2+} ions mediated packing contacts between protein molecules. However, a 3.8 Å RetS\textsubscript{HK} structure obtained from a crystal grown in the absence of cobalt was essentially identical to the higher resolution structure with an RMSD 0.7 Å for all atoms; thus demonstrating that the Co\textsuperscript{2+} coordination stabilizes contacts but does not significantly impact the protein conformation.
Diffraction data, and refinement statistics are provided in Table 1. Final coordinates are available from the protein data bank under the accession numbers 6DK7 (Co$^{2+}$) and 6DK8 (Co$^{2+}$-free).

**Bacterial Adenylate Cyclase Two-Hybrid Assay**

Strains and vectors were purchased from Euromedex as part of the bacterial adenylate cyclase two-hybrid assay system kit (EUK001). Constructs were inserted into BTH101 cells via heat-shock transformation. Strains were grown on LB agar supplemented with 50 µg/ml kanamycin and 100 µg/ml ampicillin. Single colonies were grown overnight in LB broth supplemented with the proper antibiotics. To assess protein-protein interactions, 2 µl of an overnight culture of the desired strains (OD$_{600}$ ≈ 2.0) were plated on MacConkey agar (supplemented with 50 µg/ml kanamycin, 100 µg/ml ampicillin, 1% maltose, 0.5 mM IPTG) and incubated at 32 C for 40 hours. Positive interactions were indicated by the acidification of the media, visualized by a red color. eGFP fluorescence was visualized by imaging the plate using the Alexa-fluor 488 setting of a Bio-Rad Chemidoc MP imaging system. Full results, including additional controls are shown in figure S6.

**Crystal Violet Biofilm Assay**

*P. aeruginosa* PAK strains, carrying the pHERD20T plasmid with indicated variants, were plated on LB agar plates supplemented with 300µg/ml Carbenicillin. Single colonies were picked and grown O/N in LB broth at 37° C. The next morning, cells were sub-cultured into modified M63 media$^{36}$ and grown O/N at 37° C. Once the cultures had
reached stationary phase, cells were diluted to OD$_{600}$ = 0.05 in fresh media and 100 µl was inoculated into 96-well assay plates (Corning #2797). Plates were sealed and incubated at 37° C for 24. Following incubation, media was removed and the attached cells were washed 2x. Adhered cells were stained for 15 min using 0.1% crystal violet and then washed 4x. Adherence was quantified by destaining the cells in 200 µl of 30% acetic acid. The solution was transferred to a fresh 96-well plate and the absorbance of this solution was then measured at 600 nm using a Tecan M200 plate reader.

**Quantification of RetS construct levels in vivo**

Protein isolated from each cell type was treated with DTT and iodoacetamide to reduce and alkylate cysteine residues, then sequentially digested using LysC and trypsin as previously described$^{48}$. After desalting, protein digests were analyzed using an Orbitrap Fusion Lumos equipped with an Easy-nLC 1200 UPLC, and an Easy Spray nanospray source (Thermo Scientific). The column utilized for peptide separation was a PepMap RSLC C18 (2 µm particles, 75 µm x 25 cm column, ES802; Thermo Scientific). Samples were analyzed in duplicate and each analysis utilized a 110 minute gradient from 2 to 25% solvent B followed by a 45 minute ramp to 40% B where solvent A was 0.1% (v/v) formic acid in water and solvent B was 20:80 water:acetonitrile containing 0.1% (v/v) formic acid. The column temperature was maintained at 55°C and the ion transfer tube at 275°C. Other mass spec. conditions were as follows: 10 µl injection, 300 nl/min flow rate, 2200 V ion spray voltage, MS scans utilizing the orbitrap at 120000 resolution for m/z 400-1500, an agc target of 4e5 with 50 ms max inject time, profile, positive, and RF lens at 30%. The MS scan was then followed by MS/MS scans for multiply charged precursors (z = 2-5) with an isotopic distribution resembling a
typical peptide and an intensity above 2e4. Maximum cycle time was limited to 4 sec and dynamic exclusion limited MS/MS on the same mass (±10 ppm) to once every 15 sec. MS/MS scans utilized the orbitrap at 15000 resolution with an agc target of 1e5 and a max inject time of 200 ms. Activation type utilized was HCD (CE = 30±3) and the isolation window was 1.4 m/z.

Data were processed and proteins identified using Proteome Discoverer 2.2 (Thermo Scientific) searching the Pseudomonas aeruginosa (PAO1) reference proteome (UP000002438) from UniProt with entries corresponding to the altered RetS sequences appended. Protein false discovery rate was limited to less than 1%, peptide mass tolerance was ±10 ppm and peptide fragment mass tolerance was ±0.05 Da. Quantitative values for technical replicates were averaged and values were normalized to total peptide amount. Protein amount was estimated using the Top3 method (sum of 3 most intense peptides) and moles of RetS (Q9HUV7) is expressed as a ratio to moles of 60 kDa chaperonin groL (P30718). Peptides used for quantification were: 403-TSEAAHTAELQTK-415, 810-ILVAEDNSISTK-821, and 904-LVGMDGHMAKPVELSQR-921 for RetS and 16-MLVGVNVLADAVK-28, 143-AIAQVGTISANSDESIGQIIAEAMEK-168, and 405-AAVEEGVVPGGGVALVR-421 for the chaperonin
Figure 1: RetS/GacS Interaction Studies. (A) BACTH assay probing RetS/GacS domain interactions. All tested constructs are listed above the panels. A GFP tag was added to the C-terminus of all the RetS constructs to verify stable expression levels in vivo (top panel). RetS constructs and eGFP were expressed from pKT25 while GacS constructs were expressed from pUT18C. Strains were plated on MacConkey agar.
Positive interactions give colonies a bright red color (bottom panel). Images show representative results of triplicate measurements. (B) GacS homodimerization was verified by titrating GacS<sub>cyt</sub>-555 into a solution of 250 fM GacS<sub>cyt</sub>-488. Complete FRET quenching was observed near a ratio of 100:1 GacS<sub>cyt</sub>-555:GacS<sub>cyt</sub>-488. (C) RetS<sub>cyt</sub> binding does not disrupt the GacS<sub>cyt</sub>-488:GacS<sub>cyt</sub>-555 dimer, as an addition of unlabeled RetS<sub>cyt</sub> has no measurable effect on the observed FRET signal. The first data bar represents fluorescence of a solution of 250 fM GacS<sub>cyt</sub>-488. The second data bar represents the quenched fluorescence of a solution containing 250 fM GacS<sub>cyt</sub>-488 and 25 pM GacS<sub>cyt</sub>-555. The concentration of unlabeled RetS<sub>cyt</sub> runs from 0 to 8.7 nM. (D) FRET interactions between RetS<sub>cyt</sub>-555 and GacS<sub>cyt</sub>-488 demonstrate RetS<sub>cyt</sub>-GacS<sub>cyt</sub> complex formation under the given experimental conditions. (E) Titration of unlabeled GacS<sub>cyt</sub> into a solution of 250 fM GacS<sub>cyt</sub>-488 and 25 pM GacS<sub>cyt</sub>-555 disrupts the dually labelled GacS<sub>cyt</sub>-488:GacS<sub>cyt</sub>-555 dimer. The first data bar represents fluorescence of a solution of 250 fM GacS<sub>cyt</sub>-488. The second data bar represents the quenched fluorescence of a solution containing 250 fM GacS<sub>cyt</sub>-488 and 25 pM GacS<sub>cyt</sub>-555. The concentration of unlabeled GacS<sub>cyt</sub> ranged from 0 to 125 pM. Experiments were performed in triplicate and error bars represent the standard error calculated from those triplicate experiments.
Figure 2: Crystal structure of the RetS kinase region. Relevant crystallographic statistics can be found in Table 1. RetS_{HK} crystallized in spacegroup P4_{1}2_{1}2. Four identical homodimers were observed within the asymmetric unit of the crystal (see Fig. S4). (A) RetS_{HK} adopts the traditional histidine kinase fold, see text for details. (B)
Secondary structure visualization for RetSHK, connects discussed secondary structure and sequence elements. (C) Two-views of the domain architecture and quaternary structure of the RetSHK dimer. (D) The RetSHK dimer is asymmetric. On one side the ATP lid (shown in pink) of one RetSHK has become an integral part of the dimer interface (left panel), while this space is occupied by the now more extended and kinked α-1 helix occupies this position on the other side of the dimer (right panel).

Figure 3: Probing the importance of the ATP lid loop and N-terminus of the RetSHK helix α-1 RetS_{HKΔlid} interacts strongly with GacS_{HK} and the GacS_{DHP} domain alone but RetS_{HKΔhelix} failed to bind either GacS construct but retained its ability to dimerize with RetS_{HK}. Stable expression was verified via C-terminal eGFP tags, and proper folding was determined by the variant’s ability to dimerize with wild-type RetS_{HK}. Images show representative results, experiments were conducted in triplicate.
**Figure 4: in vivo activity of RetS variants.** The biological activity of RetS variants was tested using the crystal violet biofilm assay. No statistically significant difference was observed between the WT and complemented strain. The RetS$\Delta$lid variant fully complemented a $retS$ deletion, producing a significantly lower amount of biofilm than the WT strain. The RetS$\Delta$helix variant produced significantly more biofilm than the WT strain but did not quite phenocopy the $retS$ deletion strain indicating residual activity. * indicates a strain produced a statistically significant less amount of biofilm than strains expressing the WT $retS$ gene. ** indicates a strain produced a statistically significant less amount of biofilm than the $retS$ knockout, but significantly more biofilm than the strains expressing the WT $retS$ gene. Statistical significance assessed via Student’s two-tailed T-test where $p<0.05$. 
Table 1. Crystallographic Data Collection and Refinement Statistics

<table>
<thead>
<tr>
<th></th>
<th>RetS\textsubscript{peak}</th>
<th>RetS\textsubscript{inflection}</th>
<th>RetS\textsubscript{remote}</th>
<th>RetS\textsubscript{native_Co}</th>
<th>RetS\textsubscript{native_Co_free}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9792</td>
<td>0.9793</td>
<td>0.9755</td>
<td>0.9793</td>
<td>0.9793</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>29.81 – 3.55</td>
<td>29.80 – 3.54</td>
<td>29.76 – 3.72</td>
<td>56.78 – 2.60</td>
<td>56.78 – 3.80</td>
</tr>
<tr>
<td></td>
<td>(3.65 – 3.55)\textsuperscript{b}</td>
<td>(3.63 – 3.54)\textsuperscript{b}</td>
<td>(3.82 – 3.72)\textsuperscript{b}</td>
<td>(2.693 – 2.60)\textsuperscript{b}</td>
<td>(3.936 – 3.80)\textsuperscript{b}</td>
</tr>
<tr>
<td>Space group</td>
<td>P422</td>
<td>P422</td>
<td>P422</td>
<td>P4\textsubscript{1}2\textsubscript{1}2</td>
<td>P4\textsubscript{1}2\textsubscript{1}2</td>
</tr>
<tr>
<td>Unit cell dimensions</td>
<td>157.43</td>
<td>157.34</td>
<td>157.10</td>
<td>158.708</td>
<td>158.71</td>
</tr>
<tr>
<td>A, B, C (Å)</td>
<td>244.78</td>
<td>244.78</td>
<td>244.53</td>
<td>243.826</td>
<td>243.83</td>
</tr>
<tr>
<td>Total reflections</td>
<td>505,571 (33,230)</td>
<td>511,885 (32,481)</td>
<td>444,536 (28,629)</td>
<td>1,369,925 (174,109)</td>
<td>793,980 (109,715)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>37,643 (2,645)</td>
<td>38,069 (2,697)</td>
<td>32,772 (2,309)</td>
<td>96,024 (13,834)</td>
<td>31,211 (4,341)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>13.4 (12.6)</td>
<td>13.4 (12.0)</td>
<td>13.6 (12.4)</td>
<td>14.3 (12.6)</td>
<td>26.2 (25.3)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.6 (97.1)</td>
<td>99.6 (96.6)</td>
<td>99.6 (96.9)</td>
<td>99.9 (99.9)</td>
<td>100.0 (100.0)</td>
</tr>
<tr>
<td>&lt;I/\sigma(I)&gt;\textsuperscript{b}</td>
<td>14.7 (2.3)</td>
<td>16.6 (2.3)</td>
<td>16.5 (2.7)</td>
<td>9.3 (1.3)</td>
<td>8.1 (1.2)</td>
</tr>
<tr>
<td>CC\textsubscript{1/2}\textsuperscript{c}</td>
<td>99.9 (84.1)</td>
<td>99.9 (83.7)</td>
<td>99.9 (84.2)</td>
<td>99.7 (27.3)</td>
<td>99.9 (30.2)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>56.78 – 2.60</td>
<td>56.78 – 3.8</td>
<td>(2.693 – 2.60)</td>
<td>(3.936 – 3.80)</td>
<td></td>
</tr>
<tr>
<td>R\textsubscript{merge}\textsuperscript{d}</td>
<td>0.06761 (0.7454)</td>
<td>0.404 (3.620)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R\textsubscript{work}/R\textsubscript{free}\textsuperscript{f}</td>
<td>0.2275/0.2526</td>
<td>0.2183/0.2500</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMS(bonds)</td>
<td>0.008</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMS(angles)</td>
<td>0.97</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average B-factor (Å\textsuperscript{2})</td>
<td>62.01</td>
<td>54.01</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}All anomalous datasets were collected from a single crystal.

\textsuperscript{b}The values in parentheses relate to the highest resolution shell.

\textsuperscript{c}CC\textsubscript{1/2} = Pearson correlation coefficient between random half-datasets\textsuperscript{49}

\textsuperscript{d}R\textsubscript{merge} = \Sigma |I| - <I>/\Sigma I, where I is the observed intensity and <I> is the average intensity obtained from multiple observations of symmetry-related reflections after the rejection of significant outliers.

\textsuperscript{e}R\textsubscript{work} = \Sigma |F\_o| - |F\_c| / \Sigma |F\_o|

\textsuperscript{f}R\textsubscript{free} defined by Brunger\textsuperscript{50}
**SUPPLEMENTAL FIGURES**

Figure S1: Loss of the RetS ligand binding pocket. (A) Putative ligand binding pocket in RetS<sub>HK</sub> subunit A (green) superimposed on the CA domain of EnvZ (magenta) which was crystallized with bound ATP. (B) Putative ligand binding pocket in RetSHK subunit B (green) superimposed on the CA domain of EnvZ (magenta). In both subunits, the RetS<sub>HK</sub> putative ligand binding pocket is severely compressed. (C) A series of conserved residues have been identified as being critical for ATP in histidine kinases. These residues are highlighted in yellow. GacS and LadS are functional HKs that also participate in the Gac/Rsm pathway. KinB, WalK, EnvZ, and DesK are related functional HKs with known structures which have been the subject of intense structural studies<sup>31,32,34,51</sup>. RetS has acquired a number of mutations in these residues (indicated by *). Comparison of the RetS “binding pocket” to the EnvZ binding pocket reveals that
two mutations appear to be primarily responsible for the lack of RetS nucleotide binding. The first N box asparagine plays an important role in coordinating the catalytic Mg\textsuperscript{2+} ion and, by extension, the phosphate groups of ATP. In RetS, this residue has been mutated to a serine (Ser-537). The mutation of the second G2 box glycine to a histidine (His-604) is an extremely disruptive mutation which results in the insertion of a bulky sidechain directly into the binding pocket. Additionally, due to the repurposing of the ATP lid, the peptide backbone of the ATP lid compresses the binding pocket to the point that it would clash with any bound nucleotide.
Figure S2: Directionality of potential autophosphorylation. The spatial arrangement of DHp helices serves as the primary indicator of whether a histidine kinase will autophosphorylate in cis or in trans\textsuperscript{30}. HKs whose DHp helices arrange themselves in a clockwise manner, such as EnvZ, undergo trans-autophosphorylation, whereas HKs whose DHp helices arrange in a counter-clockwise manner, such as WalK, undergo cis-autophosphorylation. Analysis of the RetS DHp bundle revealed that the helices arrange in a clockwise manner, suggesting that the evolutionary ancestor of RetS was capable of undergoing trans-autophosphorylation.
Figure S3: CA domain movement resulting from ATP lid/DHp interaction.

Superposition of the DHp domains from RetSHK subunits A and B reveal that in subunit B (blue) the CA domain is rotated 34° compared to subunit A (green). This rotation appears to result from the “swinging in” of the ATP lid in subunit B, as described in the text. Centers of mass for each domain are indicated by spheres: DHp (Black), subunit A CA (green), subunit B CA (blue). Centers of mass were calculated using PyMol. 

52
**Figure S4: Asymmetric unit of the RetS\textsubscript{HK} crystal.** (A) RetS\textsubscript{HK} crystallized with eight separate peptide chains in the asymmetric unit, forming 4 homodimers. (B) These homodimers were found to be essentially identical, overlaying with an RMSD of $\approx 0.4$ Å. It was observed that the ATP lid of subunit B unfolded the N-terminal region of the first DHp helix from subunit A. This interaction was observed in each of the four dimers. To confirm that this structural reorganization was not an artifact resulting from crystal packing, the packing contacts were analyzed using Contact\textsuperscript{44}. (C) The residues involved in packing contacts are listed by chain. Each chain was found to adopt either a conformation where the DHp helix is partially unfolded or a conformation where the DHp helix is fully folded. Therefore, each residue involved in packing contacts was mapped to its respective subunit of the RetS\textsubscript{HK} homodimer (residues colored in pink). Residues within the ATP lid were found to form packing contacts only in subunits where the ATP lid (colored in yellow) did not displace the DHp helix from the adjacent subunit. Residues within the displaced DHp helix (colored in dark blue) were found to form packing contacts only in subunits where the helix was fully folded. Thus, it can be confirmed that the observed interaction between the ATP lid and DHp helix is not a result of artificial packing contacts and represents a novel observation of a biologically significant structure, as supported by Fig. 4.
Figure S5: RetS\textsubscript{HK} dimerization interface. The RetS\textsubscript{HK} homodimer exhibits a novel dimerization interface due to the previously discussed ATP-lid/DHp interaction. As a result of this interaction, the ATP-lid (magenta) from subunit B (cyan) compensates for the unfolding of the N-terminal DHp helix in subunit A (green).
Figure S6: *In vivo levels of RetS variants*. Stable expression of RetS variants was analyzed via LC-MS. No significant levels of RetS peptides were detected in the PAKΔretS strain. Data bars represent the average of duplicate samples, error bars represent the standard deviation of those duplicates. Data is expressed as moles of RetS per moles of the chaperonin GroL.
Figure S7: Expanded BACTH results. Results from figure 1 are summarized together along with additional controls. (A) Fermentation of maltose, indicating interaction between tested constructs, is visualized by the presence of a bright red color on
MacConkey agar. (B) Fluorescence of eGFP, which was added as a C-terminal fusion to all RetS constructs, was visualized. Aside from RetS<sub>CA</sub> and RetSD<sub>Hp</sub>, all RetS mutations were constructed in the RetS kinase region.

<table>
<thead>
<tr>
<th>Table S1: Cloning primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>RetS&lt;sub&gt;416_ATT_TEV_F&lt;/sub&gt;</td>
</tr>
<tr>
<td>RetS&lt;sub&gt;649_ATT_R&lt;/sub&gt;</td>
</tr>
<tr>
<td>pKT25_RetS_Xbal_F</td>
</tr>
<tr>
<td>pKT25_RetS_SacI_R</td>
</tr>
<tr>
<td>Ret&lt;sub&gt;Δ588-603::gly8_F&lt;/sub&gt;</td>
</tr>
<tr>
<td>Ret&lt;sub&gt;Δ588-603::gly8_R&lt;/sub&gt;</td>
</tr>
<tr>
<td>pKT25_RetS_CA_Xba_F</td>
</tr>
<tr>
<td>Ret&lt;sub&gt;Δ416-423::gly8_F&lt;/sub&gt;</td>
</tr>
<tr>
<td>pKT25_RetS_Δ416-423::gly8_Xba_F</td>
</tr>
<tr>
<td>pUTC18_GacS_KpnI_R</td>
</tr>
<tr>
<td>pUTC18_GacS_CA_Xba_F</td>
</tr>
<tr>
<td>pKT25_GFP_Xbal_F</td>
</tr>
<tr>
<td>pKT25_GFP_KpnI_R</td>
</tr>
<tr>
<td>pUTC18_RetS_Xba_F</td>
</tr>
<tr>
<td>pUTC18_RetS_KpnI_R</td>
</tr>
<tr>
<td>pHERD20T_RetS_649_stop_F</td>
</tr>
</tbody>
</table>
Table S2: Buffer compositions

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Composition</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>HK_A</td>
<td>50 mM Tris pH 7.4, 500 mM NaCl, 25 mM imidazole pH 7.4, 5% glycerol, 1 mM DTT</td>
<td>Cell resuspension, Ni column binding</td>
</tr>
<tr>
<td>HK_B</td>
<td>50 mM Tris pH 7.4, 500 mM NaCl, 250 mM imidazole pH 7.4, 5% glycerol, 1 mM DTT</td>
<td>Ni column elution</td>
</tr>
<tr>
<td>HK_C</td>
<td>50 mM Tris pH 7.4, 50 mM NaCl, 10% glycerol, 1 mM DTT</td>
<td>IEC column binding</td>
</tr>
<tr>
<td>HK_D</td>
<td>50 mM Tris pH 7.4, 500 mM NaCl, 10% glycerol, 1 mM DTT</td>
<td>IEC column elution</td>
</tr>
<tr>
<td>HK_E</td>
<td>25 mM Tris pH 7.4, 125 mM NaCl, 2.5% glycerol, 2 mM TCEP pH 8.0</td>
<td>SEC column elution, protein storage and crystallization</td>
</tr>
<tr>
<td>HK_Ni_dialysis</td>
<td>50 mM Tris pH 7.4, 500 mM NaCl, 5% glycerol, 1 mM DTT</td>
<td>Volume was adjusted so that the imidazole concentration in the protein sample reached 25 mM following dialysis</td>
</tr>
<tr>
<td>HK_IEC_dialysis</td>
<td>50 mM Tris pH 7.4, 10% glycerol, 1 mM DTT</td>
<td>Volume was adjusted so that the NaCl concentration in the protein sample reached 50 mM following dialysis</td>
</tr>
<tr>
<td>Cyt_A</td>
<td>50 mM HEPES pH 7.5, 500 mM NaCl, 25 mM imidazole</td>
<td>Cell resuspension, Ni</td>
</tr>
</tbody>
</table>
pH 7.4, 5% glycerol, 1 mM DTT

Cyt_B
50 mM HEPES pH 7.5, 500 mM NaCl, 250 mM imidazole pH 7.4, 5% glycerol, 1 mM DTT

Ni column elution

Labeling
150 mM sodium bicarbonate pH 8.3
SEC column elution, fluorescent labeling

Stabilization
25 mM HEPES pH 7.5, 150 mM NaCl, 2 mM TCEP pH 8.0
SEC column elution, storage, FRET assays

REFERENCES


FINAL DISCUSSION

Many of the most essential aspects of life, as we know it, are performed by multimeric protein assemblies. Therefore, unraveling the structure and function of these assemblies is of critical importance to many fields of study. Those who study human health and disease are particularly interested in understanding the function of protein assemblies, as many virulence factors take the form of multimeric proteins. *Pseudomonas aeruginosa* has emerged as a model organism for studying protein assemblies as they relate to human health. As an opportunistic human pathogen, *P. aeruginosa* is able to cause both chronic and acute infection in immunocompromised individuals\(^1\). Acute infections are characterized by the expression of the type III secretion system, a massive protein assembly which functions as a molecular syringe, allowing the bacterium to inject cytotoxic effector proteins directly into the host cell cytoplasm\(^2\).

Conversely, chronic infections by *P. aeruginosa* are characterized by the formation of a bacterial biofilm. The formation of a bacterial biofilm is a complex process, which involves numerous essential protein assemblies including flagella and type IV pili\(^3\). While the functional role of protein assemblies in basic cellular functions has long been studied, it has recently been determined that protein assemblies also play a vital role in the regulation of cellular behavior. The goal of this work is to understand the role that protein assemblies play in the regulation of cellular processes and how that mechanism of regulation affects global virulence. Towards that goal, two systems were chosen as foci for the study, the assembly of type IV pili, and the RetS/GacS interaction, which globally regulates the transition between acute and chronic virulence in *P. aeruginosa*.

The assembly of type IV pili is powered by the dedicated ATPase, PilB. The work presented here (chapter 2) produced the first crystal structure of the catalytic core domain found in all PilB-type enzymes and allowed for the generation of the first model explaining how PilB is able to convert the chemical energy, resulting from ATP hydrolysis, into the mechanical energy
necessary to drive pilus assembly and extension into the extracellular environment. PilB is a
tantalizing drug target, as many bacteria rely on type IV pili during the initial stages of host
colonization, and many mutants which cannot produce pili have been found to be avirulent. PilB
belongs to the AAA+ superfamily of proteins, which includes many essential enzymes
conserved across all kingdoms of life, therefore, to develop any potential therapeutics targeting
PilB, it is essential that the structure and mechanism of action is fully understood to maximize
the effectiveness of any potential therapeutic while minimizing the side effects caused by cross-
reactivity with related enzymes\textsuperscript{4}.

Like PilB, RetS also represents an ideal drug target. RetS is a sensor histidine kinase which
belongs to the 7TMR-DISMED2 family of enzymes, a family which is not present in higher
eukaryotes. RetS has been demonstrated to globally regulate virulence by inhibiting the
autophosphorylation activity of a related sensor kinase, GacS, through a direct binding
interaction\textsuperscript{5-7}. The results discovered during the course of this study (chapter 3) refute the
previously accepted model in the field and suggest a new mechanism of interaction mediated by
the regulated unfolding of a structured $\alpha$-helix, or helix cracking, in response to ligand binding.
This unfolding event was found to occur directly upstream of the catalytic histidine residue
conserved across all histidine kinases. This observation suggests that helix cracking could
represent a novel, universal means of transmembrane signal transduction in this family of
enzymes. Although purely speculation at this point, these results have the possibility to redefine
the current understanding of transmembrane signaling and represent an exciting avenue of
future exploration.

In an effort to understand how ligand binding affects the interaction between RetS and GacS,
nitrate-responsive GacS chimera were created (appendix B). While the constructs demonstrated
a definite nitrate response, the results have been difficult to be consistently replicated. It has
long been known that \textit{P. aeruginosa} can spontaneously produce mutants, termed rugose small
colony variants (RSCVs). These RSCVs display a hyper-biofilm phenotype, and have been theorized to carry mutations promoting growth in the cystic fibrosis lung environment\(^8\). In a laboratory setting, generation of RSCVs creates a heterogenic bacterial population, which makes analyzing effects on biofilm production troublesome, due to the RSCV hyper-biofilm phenotype. It has been demonstrated that mutation of certain genes increases the rate of RSCV generation\(^9\). Based on experimental observations, it would appear the mutation of \(\text{gacS}\) promotes generation of RSCVs, providing a possible explanation for the difficulty in reproducing the NarX-GacS chimera results. The use of nitrate as an activation signal is also complicated by the fact that nitrate metabolism has been shown to modulate biofilm production in \(P. \text{aeruginosa}\) on its own\(^10\). The fact that the functional assay is performed under oxygen-limiting conditions likely magnifies the natural effect of nitrate on biofilm production, making it likely that other, non-metabolic signals would give more consistently reproducible results.

In conclusion, \(P. \text{aeruginosa}\) utilizes a diverse array of protein assemblies to regulate its myriad of virulence factors. The results presented here grant a deeper understanding of the structure and molecular mechanism of critical protein assemblies and should serve as a solid foundation for future translational studies geared towards developing novel therapeutics and improving the quality of life for individuals at risk of infection by \(P. \text{aeruginosa}\).

REFERENCES


Appendix A: Effects of removing the RetS periplasmic sensory domain on the 
*Pseudomonas aeruginosa* PAO1 transcriptome

**INTRODUCTION**

RetS, GacS, and LadS form a tripartite signaling system responsible for controlling the expression of over 600 different genes across the *Pseudomonas aeruginosa* genome\(^1\). All three of these proteins are membrane bound sensor kinases, yet, at the beginning of this study, the ligand(s) sensed by these proteins were unknown. The crystal structure of the periplasmic sensory domain of RetS (RetS\(_{\text{peri}}\)) has previously been solved in the lab\(^2\). Based on this structure, it was hypothesized that RetS\(_{\text{peri}}\) bound a carbohydrate ligand. Unpublished data in the lab suggest that RetS\(_{\text{peri}}\) is able to bind the PSL polysaccharide, a major structural component of the *P. aeruginosa* biofilm matrix. Synthesis of the PSL, and its related polysaccharide, PEL, is controlled by the Gac/Rsm pathway, suggesting that the binding of PSL to RetS\(_{\text{peri}}\) could form a type of regulatory feedback loop. While the data suggest that RetS\(_{\text{peri}}\) may bind PSL, the resultant effects of that interaction are unknown. With that in mind, we endeavored to investigate the effect of RetS ligand binding on *P. aeruginosa*. A \(\Delta\)retS mutant displays a hyper-biofilm phenotype, yet when the sensory domain of RetS is deleted, the resulting strain yields an intermediate phenotype, producing more biofilm than the WT strain, yet less biofilm than a \(\Delta\)retS mutant. This intermediate phenotype holds true across biofilm assays as well as transcriptional assays from Gac/Rsm controlled promoters (data not shown). Having observed this intermediate phenotype, it was understood that identification of a clear molecular phenotype would greatly aid further investigations aimed at better understanding the physiological effects of RetS ligand binding. In an effort to find the new phenotype, we conducted an RNA sequencing experiment, where we compared the transcriptomes of three retS knockout strains, one of which expressed the wild-type retS gene *in trans*, one of which
expressed a mutant retS gene lacking the periplasmic sensory domain in trans, and one carrying a vector control, serving as our ΔretS background.

MATERIALS AND METHODS

Three stains, PAO1ΔretS pSK-retS (complemented strain), PAO1ΔretS pSK-retSΔperi (strain lacking the RetS periplasmic sensory domain), and PAO1ΔretS pSK (retS deletion strain), were streaked onto LB agar plates supplemented with 300 µg/ml carbenicillin and grown overnight at 37°C. Two single colonies were picked from each plate and grown in 5 ml LB broth supplemented with 300 µg/ml carbenicillin overnight at 37°C with shaking. The following day, the strains were subcultured into 5 ml Jensen’s defined medium, supplemented with 300 µg/ml carbenicillin, and grown overnight with shaking at 37 °C. After 18 hours of growth, each strain was further subcultured into 5 ml fresh Jensen’s defined medium, supplemented with 300 µg/ml carbenicillin, at an OD$_{600}$ ≈0.05 and grown until OD$_{600}$ ≈0.60. RNA was extracted using the RNeasy extraction midikit and further purified using the RNeasy minikit (Qiagen). RNA quality was assessed via RIN >8.0. Samples were then submitted to Genewiz for sequencing. Reads were mapped to the P. aeruginosa PAO1 genome using Geneious. To confirm sequencing results, expression of target genes was analyzed via qRT-PCR using the Sybr Green iQ supermix, following manufacturer’s instructions. Pseudomonas quinolone signal extraction and quantification was performed as described.

RESULTS AND DISCUSSION

Following transcriptome analysis, it was determined that two operons were significantly upregulated in the strain lacking only the RetS periplasmic sensory domain (Table 1). One of these operons consisted of hypothetical proteins, however, the other operon was determined to produce the metabolic machinery necessary for the synthesis of Pseudomonas Quinolone Signal (PQS). PQS is a signaling molecule unique to Pseudomonads that forms a third quorum
sensing system, alongside the Las and Rhl systems\textsuperscript{5,6}. The PQS quorum sensing system is responsible for regulating ~140 genes, although its regulon overlaps with the other quorum sensing systems\textsuperscript{7}. Believing that RetS ligand binding may play a role in regulating PQS production, the expression of the \textit{pqsABCDE} operon was monitored via qRT-PCR using the same strains as the RNA-seq experiment. Unfortunately, the results of the qRT-PCR did not match the results of the RNA-seq (Fig.). Due to the instability associated with RNA samples leading to inconsistent qRT-PCR results, it was decided to quantify the stable end product, instead. PQS was extracted directly from cells using acidified ethyl acetate and analysed via thin-layer chromatography. Natural fluorescence of PQS allows for visualization of the end product under UV light, however, no differences in the expression of PQS were observed among the different strains (data not shown). It was later determined that the final step in PQS synthesis is heavily oxygen dependent\textsuperscript{8}, and the variations in gene expression observed in the RNA-seq data were likely a result of altered oxygen availability amongst the samples during growth. In support of this idea, the second operon found to be upregulated in the absence of the RetS periplasmic domain is predicted to contain genes related to anaerobic metabolism.

Therefore, it is likely that removing the RetS periplasmic sensory has no physiological effect on \textit{P. aeruginosa} cells during exponential growth. This supposition is constant with the idea that PSL serves as the natural ligand for RetS. PSL is produced when \textit{P. aeruginosa} transitions to a sessile lifestyle, a transition which occurs late in the growth cycle. It is likely that the cells used in the RNA-seq experiment were harvested at a growth stage that was too early to observe a phenotype caused by the lack of the RetS periplasmic domain. It was later determined that PSL is able to stimulate its own production, a finding which would suggest that PSL binding to RetS would stimulate dissociation from GacS, allowing GacS to promote PSL synthesis as a positive feedback cycle\textsuperscript{9}.
Table 1: RNA sequencing results. Table is ordered by increase in expression in the RetSΔperi strain versus the RetS deletion strain. For the sake of brevity, complete results have been omitted.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Name</th>
<th>S1Expression</th>
<th>S4Expression</th>
<th>Comp avg</th>
<th>S3Expression</th>
<th>S6Expression</th>
<th>KO avg</th>
<th>S5Expression</th>
<th>S2Expression</th>
<th>KO/Comp</th>
<th>Peri/Comp</th>
<th>Peri/KO</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA4823</td>
<td></td>
<td>12.86</td>
<td>6.72</td>
<td>9.79</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>16.07</td>
<td>8.03</td>
<td>0.00</td>
<td>0.02</td>
<td>DIV/0/0</td>
</tr>
<tr>
<td>PA4856</td>
<td>retS</td>
<td>174960.39</td>
<td>170172.01</td>
<td>172566.20</td>
<td>773.84</td>
<td>985.86</td>
<td>879.85</td>
<td>262743.37</td>
<td>278847.38</td>
<td>0.01</td>
<td>1.57</td>
<td>307.77</td>
</tr>
<tr>
<td>PA4211</td>
<td>phtB</td>
<td>51.45</td>
<td>53.74</td>
<td>52.59</td>
<td>59.11</td>
<td>66.77</td>
<td>62.94</td>
<td>1000.06</td>
<td>662.14</td>
<td>1.20</td>
<td>15.80</td>
<td>13.20</td>
</tr>
<tr>
<td>PA0999</td>
<td>pqsD</td>
<td>351.55</td>
<td>446.69</td>
<td>399.12</td>
<td>300.94</td>
<td>2772.98</td>
<td>1536.96</td>
<td>17748.10</td>
<td>17887.84</td>
<td>3.85</td>
<td>44.39</td>
<td>11.53</td>
</tr>
<tr>
<td>PA1001</td>
<td>phtA</td>
<td>636.79</td>
<td>661.64</td>
<td>650.22</td>
<td>553.51</td>
<td>2631.58</td>
<td>1592.55</td>
<td>18093.50</td>
<td>18461.10</td>
<td>2.45</td>
<td>28.11</td>
<td>11.48</td>
</tr>
<tr>
<td>PA0998</td>
<td>pqsC</td>
<td>184.35</td>
<td>325.78</td>
<td>255.07</td>
<td>198.83</td>
<td>2776.91</td>
<td>1487.87</td>
<td>16053.21</td>
<td>17122.94</td>
<td>5.83</td>
<td>65.03</td>
<td>11.15</td>
</tr>
<tr>
<td>PA0997</td>
<td>pqsB</td>
<td>111.47</td>
<td>147.78</td>
<td>129.62</td>
<td>91.36</td>
<td>2156.33</td>
<td>1123.84</td>
<td>12486.73</td>
<td>12330.55</td>
<td>8.67</td>
<td>95.73</td>
<td>11.04</td>
</tr>
<tr>
<td>PA1000</td>
<td>pqsE</td>
<td>325.83</td>
<td>349.29</td>
<td>337.56</td>
<td>349.30</td>
<td>2120.98</td>
<td>1235.14</td>
<td>11980.67</td>
<td>12413.90</td>
<td>3.66</td>
<td>36.13</td>
<td>9.88</td>
</tr>
<tr>
<td>PA4213</td>
<td></td>
<td>25.72</td>
<td>53.74</td>
<td>39.73</td>
<td>26.87</td>
<td>51.06</td>
<td>38.97</td>
<td>421.71</td>
<td>268.56</td>
<td>0.98</td>
<td>8.69</td>
<td>8.86</td>
</tr>
<tr>
<td>PA2193</td>
<td>hcnA</td>
<td>25.72</td>
<td>43.66</td>
<td>34.69</td>
<td>48.36</td>
<td>58.92</td>
<td>53.64</td>
<td>465.89</td>
<td>449.14</td>
<td>1.55</td>
<td>13.19</td>
<td>8.53</td>
</tr>
<tr>
<td>PA3479</td>
<td></td>
<td>360.12</td>
<td>342.57</td>
<td>351.35</td>
<td>155.84</td>
<td>255.30</td>
<td>205.57</td>
<td>1815.37</td>
<td>1639.13</td>
<td>0.59</td>
<td>4.92</td>
<td>8.40</td>
</tr>
<tr>
<td>PA2194</td>
<td>hcnB</td>
<td>420.15</td>
<td>446.69</td>
<td>433.42</td>
<td>462.15</td>
<td>624.51</td>
<td>543.33</td>
<td>4498.27</td>
<td>4408.07</td>
<td>1.25</td>
<td>10.27</td>
<td>8.20</td>
</tr>
<tr>
<td>PA2195</td>
<td>hcnC</td>
<td>441.58</td>
<td>557.52</td>
<td>499.55</td>
<td>451.41</td>
<td>620.58</td>
<td>535.99</td>
<td>4582.62</td>
<td>3977.45</td>
<td>1.07</td>
<td>8.57</td>
<td>7.99</td>
</tr>
<tr>
<td>PA0996</td>
<td>pqsA</td>
<td>201.50</td>
<td>245.18</td>
<td>223.34</td>
<td>161.22</td>
<td>3479.98</td>
<td>1820.60</td>
<td>14257.92</td>
<td>14664.24</td>
<td>8.15</td>
<td>64.75</td>
<td>7.94</td>
</tr>
<tr>
<td>PA4129</td>
<td>hypo</td>
<td>107.18</td>
<td>184.72</td>
<td>145.95</td>
<td>161.22</td>
<td>243.52</td>
<td>202.37</td>
<td>1751.11</td>
<td>1435.40</td>
<td>1.39</td>
<td>10.92</td>
<td>7.87</td>
</tr>
<tr>
<td>PA1002</td>
<td>hypo</td>
<td>351.55</td>
<td>322.42</td>
<td>336.99</td>
<td>274.07</td>
<td>840.54</td>
<td>557.30</td>
<td>4004.27</td>
<td>4306.20</td>
<td>1.65</td>
<td>12.33</td>
<td>7.46</td>
</tr>
<tr>
<td>PA4130</td>
<td>hypo</td>
<td>557.34</td>
<td>601.18</td>
<td>579.26</td>
<td>714.73</td>
<td>1084.06</td>
<td>899.39</td>
<td>6791.59</td>
<td>5727.71</td>
<td>1.55</td>
<td>10.81</td>
<td>6.96</td>
</tr>
<tr>
<td>PA1249</td>
<td>hypo</td>
<td>5894.89</td>
<td>6055.50</td>
<td>5975.20</td>
<td>972.67</td>
<td>1296.15</td>
<td>1134.41</td>
<td>7986.45</td>
<td>7797.47</td>
<td>0.19</td>
<td>1.32</td>
<td>6.96</td>
</tr>
<tr>
<td>PA4128</td>
<td>hypo</td>
<td>51.45</td>
<td>77.25</td>
<td>64.35</td>
<td>150.47</td>
<td>235.66</td>
<td>193.07</td>
<td>1562.35</td>
<td>1041.82</td>
<td>3.00</td>
<td>20.24</td>
<td>6.74</td>
</tr>
<tr>
<td>PA4131</td>
<td>hypo</td>
<td>274.38</td>
<td>399.67</td>
<td>337.03</td>
<td>424.54</td>
<td>608.80</td>
<td>516.67</td>
<td>3474.11</td>
<td>3241.23</td>
<td>1.53</td>
<td>9.96</td>
<td>6.50</td>
</tr>
</tbody>
</table>
REFERENCES

Appendix B: NarX/GacS chimera allow for artificial control of the Gac/Rsm pathway

INTRODUCTION

The Gac/Rsm pathway controls the expression of over 600 genes across the *Pseudomonas aeruginosa* genome\(^1\). As discussed above, the activity of GacS, the lynchpin of this pathway, is directly modulated by several adaptor proteins, all of which are sensor kinases. Much work has focused on identifying the signals sensed by these adaptor proteins, yet ligand binding to the GacS periplasmic sensory domain has, until recently, mostly been overlooked. It is likely that this oversight can partially be attributed to the fact that bioinformatics annotation of the GacS domain structure long claimed that GacS did not possess a distinct periplasmic domain. However, it has been demonstrated that GacS does, in fact, possess a structured periplasmic sensory domain, as the structure of this domain has recently been solved via NMR spectroscopy\(^2\). As part of this study, it was determined that the GacS periplasmic region was required for proper GacS function, suggesting that GacS senses some sort of activation signal, however, the identity of this activating signal remains unknown.

The discovery of GacS signal sensing adds yet another layer to the regulation of the GacS/Rsm pathway. It is now known that the flow of phosphates through the pathway is modulated by GacS, LadS, RetS, and PA1611, all of which presumably respond to the presence of an unknown extracellular signal\(^3\-^5\). It is extremely difficult to study the physiological ramifications of regulatory interactions when the regulatory molecules are unknown. In an effort to overcome this obstacle, it has been determined that signaling modules from systems with known activating signals can be fused to proteins of interest to create chimeric constructs, the activity of which can be modulated via an artificial signal. The NarX nitrate sensing system from *Escherichia coli* represents one of the best characterized signal sensing systems to date\(^6\). In the NarX system, signals stemming from periplasmic ligand binding are transduced across the inner membrane.
via structural rearrangement of a HAMP domain. HAMP domains, due to their prevalence in histidine kinases, adenylyl cyclases, methyl-accepting chemotaxis proteins, and phosphatase, have become a subject of intense structural study focusing on understanding the exact mechanism by which they are able to transduce transmembrane signals. Coincidently, GacS also possesses a HAMP domain, making it likely that GacS and NarX utilize similar mechanisms of transmembrane signal transduction. It has previously been demonstrated that by fusing the sensory and HAMP domains of NarX to the cytoplasmic domain of a separate, similar, histidine kinase, the resulting chimera could be artificially activated in vivo via the exogenous addition of nitrate. A single point mutation in the NarX sensory domain (G51R) was also identified, which was found to reverse the signaling, wherein the wild-type NarX may activate in response to nitrate binding, the NarXG51R mutant was found to deactivate.

A series of GacS-NarX chimera were created in an effort to understand the role of GacS signal sensing on biofilm formation in *P. aeruginosa*. Of particular interest was determining the role of RetS-mediated inhibition in modulating the activity of GacS in response to ligand binding. The overall goal of this project is to establish a regulatory hierarchy between GacS signal-mediated activation and RetS-mediated inhibition.

**MATERIALS AND METHODS**

Four NarX-GacS chimera were created. Two of them (NacX and NacX G51R) contained the periplasmic sensory and HAMP domains of NarX fused to the cytoplasmic region of GacS following the GacS HAMP domain, while the remaining two (NacS and NacS G51R) contained the periplasmic sensory domain of NarX fused to the cytoplasmic region of GacS prior to the GacS HAMP domain (Fig. 1). Both the wild-type and mutant (G51R) NarX sensory domains were included in an effort to exert finer control over the Gac/Rsm pathway. The chimeric constructs, along with the native gacS gene, were synthesized by Genewiz, and inserted into
pHERD20T, which allowed for expression of the constructs under the control of an arabinose inducible promoter. Vectors were inserted into PAO1ΔgacS and PAO1ΔgacSretS strains via electroporation. Strains were tested for their effect on biofilm formation using the crystal violet biofilm assay\(^1\). Briefly, single colonies were grown overnight in LB broth supplemented with 300 µg/ml carbenicillin. Strains were then subcultured into modified M63 medium (see chapter 3) supplemented with 300 µg/ml carbenicillin. Following overnight growth, each strain was subcultured into fresh modified M63 medium w/carbenicillin, with or without 1 mM KNO\(_3\), at an OD\(_{600}\)=0.05. The growth, staining, and measurement of resultant biofilms is identical to the procedure described in chapter 3.

**RESULTS AND DISCUSSION**

Results are shown in figure 2. In the presence of RetS (PAO1ΔgacS background), strains expressing two constructs, NacX and NacS were found to produce substantially more biofilm in the absence of nitrate than they did in the presence of nitrate. This result suggests two possibilities. Either that nitrate binding to these constructs inhibits GacS activity, or that the constructs adopt a conformation that is predisposed to inhibition via RetS. If nitrate binding to NacX and NacS, both of which contain the wild-type NarX sensory domain, resulted in inhibition of GacS, it would be expected that strains expressing Nac\(_X\)\(_{G51R}\) and Nac\(_S\)\(_{G51R}\), constructs carrying a mutation which reverses signaling, would be activated in the presence of nitrate. However, in the presence of RetS, neither Nac\(_X\)\(_{G51R}\) nor Nac\(_S\)\(_{G51R}\) demonstrated any statistically significant nitrate response. This suggests that, upon nitrate binding, the NacX and NacS constructs adopt a conformation which is susceptible to inhibition by RetS.

In the absence of RetS (PAO1ΔgacSretS background), analysis of the biological effect of the GacS chimera becomes more convoluted. First, it appears that wild-type GacS does not properly function in the absence of RetS. A retS deletion strain produced a clear hyper-biofilm
phenotype. It would be expected that expressing native GacS in a ΔgacSretS background would also result in the strain displaying a hyper-biofilm phenotype. However, expression of GacS in PAO1ΔgacSretS was found to produce significantly less biofilm than the strain lacking GacS, a trend also observed for the NacX and NacX_{G51R} constructs. Conversely, the strains expressing NacS and NacS_{G51R} appeared to complement to wild-type level, and both strains seemed to generate a significant nitrate response.

Due to the complementation issue in the PAO1ΔgacSretS background, further analysis will focus on NacS and NacS_{G51R}. First, it must be said that vector-based expression of native GacS in the PAO1ΔgacS background might have resulted in overexpression of all constructs, therefore the complemented strain (PAO1ΔgacS pHERD20TgacS) will be used as the positive control and reference rather than the wild-type strain. When viewed in this light, NacS_{G51R} produced biofilm at levels consistent with native GacS, regardless of the presence of nitrate. Yet, while in the absence of nitrate, the NacS expressing strain produced biofilm levels comparable to native GacS. Upon addition of nitrate, this strain produced significantly less biofilm, suggesting that upon binding nitrate, NacS becomes especially susceptible to inhibition, presumably by RetS. It appears that biofilm formation in the PAO1ΔgacSretS background is globally modulated by the addition of nitrate. It has been shown that nitrate-based metabolism under oxygen-limiting conditions plays a role in regulating pathogenicity of *P. aeruginosa*\textsuperscript{11}, indicating that nitrate-responsive chimera are likely not the best tool for studying the Gac/Rsm pathway, an observation which has caused the lab to explore alternative ligand-based chimera. However, attempting to correct to the global nitrate bias (Fig. 3) it can be observed that in the absence of RetS, both NacS and NacS_{G51R} produce elevated levels of biofilm to the point that there is no longer an observable nitrate response.

Taken together, these results suggest that the RetS-mediated inhibition of GacS takes priority over any activation signal that may be sensed through the GacS periplasmic domain. However,
this statement discounts any contribution by LadS or PA1611 to the regulation of GacS. In short, the regulation of GacS is incredibly complex, with a number of factors to consider. To fully understand the effect of artificially-controlled GacS chimera, one may have to remove all other regulatory factors, including LadS and PA1611, in which case the survivability of the resultant strain is called into doubt.

FIGURES

Figure 1: Chimeric construct schematic. Constructs were synthesized by Genewiz and inserted into the arabinose-inducible vector pHERD20T. See text for description of each construct. Sequencing results for each construct verifying proper design was provided by Genewiz.
Figure 2: Crystal violet biofilm assay results. Results shown represent the average of 6 replicates. Error bars represent the standard error of those replicates. See text for description.

Figure 3: Nitrate adjusted biofilm production. The data from figure 1 was adjusted for the effect on the background strain by normalizing the biofilm produced in the presence and
absence of nitrate by the NacS and NacS_{G51R} variants to the biofilm produced by the PAO1ΔgacS.retS strain under the same conditions.

REFERENCES