Computationally and Experimentally Exploring the Type IV Pilus Assembly ATPase for Antivirulence Drug Discovery

Jazel Mae Silvela Ramos

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

MASTER OF SCIENCE
in
BIOLOGICAL SCIENCES

Zhaomin Yang, Chair
Anne M. Brown, Co-chair
Stephen Melville

August 3, 2023
Blacksburg, Virginia

Keywords: Type IV Pilus, *Neisseria gonorrhoeae*, *Clostridioides difficile*, ATPase, Molecular Dynamics

Computationally and Experimentally Exploring the Type IV Pilus Assembly ATPase for Antivirulence Drug Discovery

Jazel Mae Silvela Ramos

ABSTRACT

Disease caused by antibiotic resistant (ABR) bacteria has become a widespread global public health issue as humanity's existing collection of effective antibiotics dwindles. ABR bacteria are responsible for approximately 5 million deaths worldwide annually, which is predicted to reach 10 million yearly by 2050. Antivirulence therapeutics have been explored in recent times as another approach to tackling the global ABR pandemic by disrupting the function of virulence factors that promote disease development. The bacterial type IV pilus (T4P) is a prevalent virulence factor in many ABR pathogens, contributing to bacterial pathogenesis by facilitating cell motility, surface adhesion, and biofilm formation. Critically, the T4P facilitates early stages of disease, providing a means to invade and colonize a host. T4P assembly is driven by the PilB/PilF motor ATPase that localizes to the cytoplasmic face of the inner membrane to drive pilus biogenesis by ATP hydrolysis. The thesis work here explores computational and experimental methods for the discovery of antivirulence therapeutics targeting the T4P assembly ATPase PilB. A computational model of *Chloracidobacterium thermophilum* PilB was generated by homology modeling and molecular docking was performed to analyze the binding characteristics of six anti-PilB inhibitory compounds identified in previous studies. Computational docking aligns with the existing body of work and reveals important protein-ligand interactions and characteristics, particularly involving the ATP binding domain of PilB. This work supports the use of PilB in

structure-based virtual screening to identify novel compounds targeting PilB. Additionally, through heterologous expression and chromatography methods, the ATPase core of *Neisseria gonorrhoeae* PilF was successfully expressed and purified as an active ATPase. This work optimized conditions for its ATPase activity *in vitro*. Additionally, this thesis documents the experimental attempt to express and purify *Clostridioides difficile* PilB as an active ATPase. Two of the seven *C. difficile* PilB variant proteins expressed led to soluble protein while one construct remains to be explored. The results of these studies provide insight for future methodology design for antivirulence therapeutic research targeting the T4P assembly ATPase using both *in silico* and *in vitro* methods.

Computationally and Experimentally Exploring the Type IV Pilus Assembly ATPase for Antivirulence Drug Discovery

Jazel Mae Silvela Ramos

GENERAL AUDIENCE ABSTRACT

Antibiotic resistant bacterial infections are responsible for nearly 5 million deaths worldwide every year. These infections are becoming increasingly more difficult to treat as bacterial pathogens acquire greater means to overcome our dwindling antibiotic repertoire. This has prompted researchers to explore alternative therapeutic strategies, including the antivirulence approach that aims to disable the function or production of bacterial virulence factors. Virulence factors serve as arms and armor that help bacteria cause disease, but they may be disrupted in such a way that renders potentially pathogenic bacteria harmless to humans. One major virulence factor in many antibiotic resistant bacteria is the type IV pilus (T4P), which is important in the early stages of host invasion by mediating adhesion and biofilm formation. This work explores both computational and experimental strategies to antivirulence drug discovery targeting the T4P, specifically the primary motor protein PilB/PilF. Newly identified PilB inhibitors were evaluated by molecular docking and molecular dynamics simulation to assess the use of PilB for drug discovery via virtual screening *in silico*. This revealed key characteristics and protein-ligand interactions that contribute to successful PilB inhibition and supports the use of *Ct*PilB for structure-based virtual screening. Additionally, the PilF motor protein from *Neisseria gonorrhoeae* was successfully purified and demonstrated to be active for inhibitor discovery in the future. This work also covers

efforts to establish *Clostridioides difficile* PilB as potential model enzyme for inhibitor discovery in the future.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Dr. Zhaomin Yang, for his guidance throughout the years. Thank you to everyone in the Yang lab both past and present, especially Keane Dye, Tori Shimozono, and Megan O'Hara, for all your support.

I would like to express my heartfelt thanks to Dr. Anne Brown for giving me the opportunity to explore collaborative computational research. Thank you to Ian Hicklin, who contributed to the initial homology modeling, molecular docking, and molecular dynamics work in Chapter 2. I would also like to thank Dr. Stephen Melville for all of the valuable feedback I have received.

I want to extend my thanks to everyone in the Microbiology Teaching Labs, especially Katrina Lasley, and other fellow TAs for fostering a great teaching environment. Thank you for letting me be the TA I always wanted as an undergraduate.

Thank you to my fiancé, Sean McDonald, for sticking with me throughout this journey.

We acknowledge National Science Foundation, Lay Nam Chang Dean's Discovery Fund, and the Center for Emerging, Zoonotic, and Arthopod-borne Pathogens (CeZAP) for partially supporting the research presented in this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 - Literature Review

Antibiotics and Antibiotic Resistance.

Antibiotics have been the primary tool against bacterial infections since penicillin was first introduced into clinical use in the 1940s [1, 2]. Penicillin ushered in the golden age of antibiotic discovery from the 1940s to 1970s, during which most major antibiotic classes were identified. Although antibiotics were instrumental to advances in modern medicine, excessive use in healthcare and agriculture resulted in decreased effectiveness as bacteria quickly acquired mechanisms to overcome the effects of antibiotics [2-4]. With the decline in novel discoveries, antibiotic resistance (ABR) continues to threaten human health on a global scale. It is estimated that 4.9 million deaths are due to resistant pathogens worldwide, with ABR directly responsible for 1.27 million of those deaths [5]. Disease by six ABR pathogens account for nearly 75% of those 1.27 million, underscoring the severity of the current global antibiotic resistance pandemic. These six organisms, *Escherichia coli*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae*, *Acinetobacter baumannii*, and *Pseudomonas aeruginosa*, are among the priority pathogens identified by the World Health Organization (WHO) due to their multidrug resistance [5, 6]. Additionally, 90% of the organisms reported as critically antimicrobial resistant by the Centers for Disease Control and Prevention (CDC) are bacteria, most of which were also identified by the WHO report [1, 6]. Without intervention by effective novel therapeutics, it is estimated that up to 10 million deaths yearly will be attributed to ABR-related disease by 2050 [7].

All major classes of antibiotics target cellular processes that are essential for sustaining life, including biosynthesis of nucleic acids, proteins, or the cell envelope and

cell wall and membrane maintenance [8, 9]. Disruption of these processes is either bacteriostatic, inhibiting bacterial growth, or bactericidal, causing cell death. Antibiotics are fundamentally broad-spectrum drugs because these life-sustaining processes are universal conserved in bacteria, whether they are pathogenic or not. This is central to the ABR problem as antibiotics act as an evolutionary pressure that selects for resistant organisms [10]. Resistance may be intrinsic, existing naturally in those producing antibiotics, lacking targeted cell structures, or by drug exclusion. Gram-negative bacteria have an additional layer of resistance against certain antibiotics, owing to the low permeability of their outer membrane [4, 11]. Acquired resistance, in contrast, occurs by spontaneous mutation or horizontal gene transfer of resistance genes and is of greater clinical concern [4]. For example, *Acinetobacter* species, including the highly ABR *A. baumannii*, are naturally competent and are able to take up exogenous DNA with great efficiency [12]. Transmission of resistance genes can also be enhanced in high density communities, such as biofilms where extracellular DNA can be readily found in the biofilm matrix [4, 11, 13]. Thus, antibiotic resistance can easily and rapidly spread throughout bacterial communities, underscoring the urgent need for alternative therapeutics.

Antibiotic Resistance Mechanisms

Resistance mechanisms fall into four major categories: decreased influx and cell permeability, modification of the target, modification or inactivation of the antibiotic, and increased efflux by multidrug efflux pumps [4, 11, 14]. A single antibiotic class may encounter any or all of these mechanisms. This is evident for the β-lactam family, for example, which includes penicillin and carbapenem. β-lactams are a broad class of

antibiotics that target peptidoglycan biosynthesis by irreversible inactivation of penicillin-binding proteins (PBPs) responsible for peptidoglycan crosslinking in the cell wall [15]. To access their targets, antibiotics must be imported through porins on the outer membrane of Gram-negative bacteria. *Acinetobacter*, *Pseudomonas*, and *Neisseria* species downregulate porin expression or produce mutant porins to limit permeability [16]. The loss of OprD porins is implicated in carbapenem resistance in *P. aeruginosa* [3, 11]. β-lactams can be inactivated by β-lactamases that hydrolyze the β-lactam amide bond. PBPs may be modified to be less susceptible to β-lactams [15, 17]. Penicillin-resistant clinical strains of *S. pneumoniae* express a PBP2b protein containing mutations in catalytic domains, rendering it unable to bind β-lactams [17]. Strains of *N. gonorrhoeae* express multiple mutant PBPs that have lower affinity for extended-spectrum cephalosporins, a class that includes ceftriaxone [18, 19]. All bacteria express efflux pumps on their cell membrane that can actively remove or extrude antibiotics from the cell. The resistance-nodulation-division (RND) class of efflux pumps are closely associated with highly resistant Gram-negative bacteria. The rise of ABR *A. baumannii* is partially attributed to the expression of three RND pump systems, AdeABC, AdeFGH, and AdeJHI [20]. The breadth of these mechanisms indicates that no class of antibiotics are immune to resistance.

Since these mechanisms are a result of mutations or require maintenance of atypical systems, ABR mechanisms come with a substantial fitness cost when compared to bacteria without ABR [21]. Secondary mutations may arise that compensate for the fitness disadvantages associated with maintaining mechanisms of resistance, although with the trade-off of decreased resistance against some antibiotics [22-24]. Studies of

clinical isolates of *E. coli* reveal a balancing act between fitness costs and resistance, with a negative correlation between the prevalence of clinical isolates and the fitness costs of their resistance mechanisms [25, 26]. Ultimately, the benefits of resistance are environment-dependent and mutations with lower fitness costs in antibiotic-free conditions are more likely to persist [24].

Antivirulence Therapeutics (AVT)

Antivirulence is an alternative approach that inhibits bacterial pathogenesis instead of essential processes, as in the case of antibiotics. Specifically, antivirulence therapeutics (AVTs) target the synthesis or function of components and systems known as virulence factors that facilitate the disease or the infection process. A variety of virulence factors are expressed during bacterial infection, including adhesins, quorum sensing systems, secretion systems, and toxins [27, 28]. These factors facilitate entry into and colonization of a host, immune system evasion, and nutrient acquisition [27, 29]. Key to this strategy is that virulence factors are necessary for bacterial pathogenesis but benign for survival under conditions other than host infections [29]. Unlike antibiotics, which inhibit bacterial life processes indiscriminately, AVTs more selectively target pathogens expressing specific virulence factors without bactericidal or bacteriostatic effects. Due to this, AVTs theoretically exert fewer or minimal selective pressures on organisms, decreasing the odds of resistance development by bacterial pathogens.

A number of AVTs are in the drug development pipeline, with most at the preclinical and clinical trial stages. A few of these are already approved by the US Food and Drug Administration (FDA). These include monoclonal antibodies that target the botulism and anthrax toxins produced by *Clostridium botulinum* and *Bacillus anthracis* respectively

[29]. Bezlotoxumab, another monoclonal antibody targeting the TcdB toxin, was approved for high risk, reoccurring *Clostridioides difficile* infections [29, 30]. Though all three target toxins specifically, AVTs are not limited to toxins. Virstatin, a small organic compound in the preclinical stage, was found to attenuate virulence related to the toxin-coregulated pilus in *Vibrio cholerae* and decreases piliation and biofilm formation in *A. baumannii* [31, 32]. Further down the pipeline, a bispecific antibody targeting the type III secretion system needle tip protein PcrV and exopolysaccharide Psl has recently completed phase II clinical trials [33, 34]. Although these compounds represent only a handful of AVTs in the drug development pipeline, they exemplify the great potential AVTs have in combating drug resistance.

The Type IV Pilus (T4P)

The bacterial type IV pilus is a filamentous adhesin that is 6-8 nm wide, extends >1 μm from the cell body, and can withstand retraction forces greater than 100 pN [35, 36]. The T4P machinery shows homology with the bacterial type 2 secretion system (T2SS) and archaeal archaellum [37-39]. T4P systems can be divided into type 4a and type 4b subtypes. T4aP systems are frequently associated with human pathogens, such as *Neisseria gonorrhoeae* and *Pseudomonas aeruginosa,* and environmental species, like *Myxococcus xanthus* and *Thermus thermophilum*. In contrast, T4bP are primarily characterized in enterics, such as *Salmonella typhi*, *V. cholerae*, and various pathogenic strains of *Escherichia coli*. These have been historically categorized by the differences in prepilin and mature pilin structure, where the length of leader peptide sequences of T4a prepilin are 6-7 amino acids and T4b prepilin are longer at 15-30 amino acids. The difference in mature pilin size typically follows a similar pattern; however, it is now more

reliable to identify systems by the presence of characteristic gene clusters such as the *pilMNOP* cluster for T4aP [40-42]. Although both T4P subtypes play roles in bacterial pathogenesis, the functions, dynamics, and structure of the T4aP subtype are better understood.

Functions Mediated by the T4P

T4P are remarkably multifunctional and facilitate cell motility, surface attachment, biofilm formation, and DNA transformation by pilus extension and retraction [43]. Through these diverse processes, T4P allow bacteria to become successful pathogens by promoting colonization and long-term persistence in hosts, which manifests as acute and chronic infection respectively.

Twitching motility, or S motility in *M. xanthus*, manifests from repeated cycles of T4P extension, surface attachment, and retraction [44]. In bacteria that express pili peritrichously, like *N. gonorrhoeae*, multiple T4P can contact a surface at different points simultaneously. Movement occurs in the direction of the pili that retract with enough force to dislodge other pili, resulting in a random walk with directional persistence that allows surface exploration [45]. In contrast, rod-shaped bacteria producing polar pili tend to move across a surface in the direction of the leading pole. Directional reversal is controlled by chemosensory systems in some species, like the Frz system in *M. xanthus* [44, 46]. To pull the cell body forward, force must somehow be generated. Accordingly, twitching motility is associated with the T4aP subtype, which expresses an additional cytoplasmic ATPase motor to drive pilus depolymerization [47]. This motor, PilT, can generate stall forces of ~100 pN in *N. gonorrhoeae* and ~150 pN in *M. xanthus* [36]. Organisms lacking PilT do not display twitching motility, supporting that a retraction

ATPase is necessary to generate the needed force for motility [48-50]. Despite the differences in pili coordination, the molecular mechanism of T4P-mediated motility appears to be well conserved across species [44].

Almost paradoxical to its role in motility, T4P also helps bacteria to attach to surfaces and other bacterial cells. Individual pili are able to withstand shear forces up to 250 pN, a characteristic that likely aids in the initial colonization stage of infection [51]. Stable attachment to surfaces and other bacterial cells can result in microcolony formation, biofilm formation, and eventually dispersal to the environment [43, 52]. Biofilms are bacterial communities encased in an extracellular matrix composed of exopolysaccharides (EPS), extracellular DNA, and protein [53]. Bacteria in biofilms are more resistant to environmental stressors, including antibiotics and the host immune system [13, 54]. The matrix is not easily permeable to antibiotics, providing a literal layer of protection to embedded communities. In addition, bacteria in biofilms can be present in a low metabolic state, decreasing target protein expression or existing as persister cells [55]. Biofilms bring cells into close contact in an environment abundant in exogeneous genetic material, allowing easier gene transfer of resistance genes [13]. The second messenger cyclic-di-GMP (cdG) promotes switching between motile and sessile lifestyles by modulating the production of EPS [56]. Some assembly ATPases contain an N-terminal $MshE_N$ domain and reportedly regulate EPS synthesis by binding cdG [57-62]. Biofilms contribute to nosocomial infections by allowing bacteria to colonize medical implements, such as catheters and implants [52]. Acute infections can progress into chronic ones by biofilm formation within organs, as is the case of *P. aeruginosa* lung infection of cystic fibrosis patients [55, 63]. As an adhesin, the T4P plays a critical role in

the bacterial pathogenesis by mediating the initial attachment and long-term colonization stages of infection.

T4P assists in the process of natural transformation by binding extracellular DNA and importing it into the cell by pilus retraction [43]. Organisms such as *Acinetobacter*, *Neisseria*, and *Streptococcus* species are capable of natural transformation, taking up and integrating extracellular DNA into their genome [64-66]. Overexpression of specific pilin subtypes, such as ComP in *N. gonorrhoeae*, increases DNA binding [67]. Though the precise mechanism is not yet clear, the current model for Gram-negative bacteria suggests that T4aP-bound DNA is partially pulled into the periplasmic space by T4P retraction, where the periplasmic competence protein ComE binds the DNA and transports it through the remainder of the periplasm by a ratcheting mechanism [43, 67]. Regardless of mechanism, natural transformation appears to require the T4P, as transformation efficiency is reduced or completely lost if the assembly or retraction ATPases are mutated or absent [49, 65, 68, 69].

Structure and Assembly of T4P

Though forms of T4P systems can be found in Gram-negative and Gram-positive bacteria alike, its role and structure are best understood in Gram-negative bacteria [70]. This section will primarily focus on T4aP (henceforth T4P), using nomenclature from *Neisseria* spp. for simplicity.

The primary extracellular structure of the T4P is the pilus filament, a helical polymer of thousands of pilin subunits. The pilus is largely composed of a single major pilin species (PilE), but other minor pilin can be found in lower quantities (PilH, PilI, PilJ, PilK, PilV, PilL, ComP) for pilin recruitment or modulation of function [71-74].

Pilins that are not yet integrated into the pilus are maintained in an inner membrane pool in a monomeric form. Prepilin peptidase (PilD) must cleave a leader peptide sequence 6-7 amino acids in length from the prepilin for pilin production and its incorporation [41, 70]. Polymerization occurs at the base of the growing filament, extruding the fiber out into extracellular space during an extension event.

The rest of the T4P machine can be viewed as three interconnected complexes: the secretin channel, the alignment complex, and the motor complex (FIG 1-1). Broadly, the secretin (PilQ + PilW) and core alignment complex (PilM, PilN, PilO, PilP) form a pore through which the pilus may travel through the periplasm and outer membrane. This is capped at the base in the inner membrane by the motor complex. The PilQ secretin sits in the outer membrane and is stabilized by a pilotin (PilW). The lipoprotein PilP and heterodimers of the transmembrane proteins PilN and PilO form a ring complex in the periplasm, with PilP interfacing with the PilQ secretin through its C-terminus. PilM is additionally recruited to the PilNOP complex where it interfaces with PilN/O at the cytoplasmic face of the inner membrane. PilM subunits may then bind either the extension (PilF) or retraction (PilT) ATPases. The inner membrane platform protein (PilG) is localized to the central lumen of the motor protein. The platform protein is believed to be involved in pilin polymerization and depolymerization processes, potentially responsible for moving pilin in and out of the filament using energy generated by motor ATPase activity [43, 75-81]. The general arrangement of Gram-positive T4P machinery is similar to Gram-negative systems, with the key difference being the absence of a secretin since Gram-positive bacteria lack an outer membrane. Homologs of the alignment complex proteins are thought to interact with Gram-positive specific T4P-

associated proteins to form a channel to extrude the pilus through the thick peptidoglycan layer [70]. In all, a functional T4P machine is comprised of at least 10 conserved proteins.

The Assembly ATPase PilF/PilB

Central to T4P function is the motor protein PilF that drives assembly, or extension, of the pilus by coupling ATP hydrolysis to a conformational change that is transduced to the platform protein. This protein is known as PilB in many other bacteria, including *P. aeruginosa* and *M. xanthus*. In the remainder of the thesis, PilB and PilF will be used interchangeably in the discussions. While the retraction ATPase PilT is absent in some T4P systems, instead retracting their pili by stochastic disassembly events, the extension ATPase is necessary for all functional T4P [43].

The assembly ATPase PilF is one of the ATPases associated with diverse cellular activities (AAA+) family. Proteins in this family are involved in a variety of processes related to DNA replication and recombination, protein stability, cell motility, and secretion [82, 83]. The PilF ATPase has conserved Walker A and Walker B boxes that are involved in ATP hydrolysis and Mg^{2+} coordination [82-86]. PilF also contains arginine finger, tetra-cysteine, Asp-box, and His-box motifs. The arginine fingers are believed to have a role in hydrolysis and intersubunit communication, similar to arginine fingers in GTPases [82, 87]. The tetra-cysteine motifs bind Zn^{2+} to stabilize the hexameric structure [88, 89]. Mutations at the conserved acidic residues in the Asp-box motif and histidines in the His-box motif disrupt ATPase function [90, 91]. The specific roles of these Asp- and His-box motifs are still underexplored.

AAA+ ATPases are known to form oligomeric, usually hexameric, rings. Crystal structures of both the T4P extension and retraction ATPases reveal a hexameric structure. Specifically, PilF forms dimers of trimers with C_2 symmetry [90, 92]. A PilF monomer has an N-terminal domain (NTD) and a C-terminal domain (CTD) connected by a flexible linker. ATP binds at the Walker A and B motifs within the CTD ASCE core. During pilus assembly, PilF subunits has three distinct conformations that can bind ATP (open), release ADP (closed), and hold ADP in a post-hydrolysis state (open') [89, 90]. Conformational changes at one binding interface are propagated across the adjoining interfaces, producing open-open', closed-open, and open'-closed transitions across the hexamer. This causes a rotary motion that is transduced to the platform protein PilF. The platform protein is proposed to incorporate pilin into the filament by either scooping or thrusting motions. Although several studies have endeavored to characterize PilF dynamics, the specific direction of rotation may differ across T4P systems, with both clockwise and counterclockwise models proposed in the literature [81, 89, 90, 92].

Targeting the T4P by AVTs

Multiple studies have investigated the T4P as a target for AVTs. Two types of compounds were identified to disrupt T4P-related functions in *Neisseria* species. The first report indicated that a family of phenothiazines, particularly trifluoperazine and thioridazine, are able to reduce T4P levels and disperse *Neisseria meningitidis* microcolonies [93]. The target is believed to be the Na^+ pumping NADH-ubiquinone oxidoreductase complex, which has not been associated with T4P until this study [93]. The findings by Denis et al. have recently been supported by a more recent study showing phenothiazine-mediated inhibition of T4P-related functions in *A. baumannii* and

Acinetobacter nosocomialis [94]. A second study identified 1-[(piperidin-4-yl)methyl]piperidin4-ol (P4MP4) by phenotypic screening for inhibitors against *N. meningitidis* cell adhesion to epithelial cells [95]. P4MP4 is believed to target pilus biogenesis by inhibiting PilF. The compounds quercetin, benserazide, and levodopa were found in two high-throughput screens for inhibitors against the PilB assembly ATPase in *Chloracidobacterium thermophilum* [96, 97]. All three compounds were able to inhibit the T4P-powered S motility in *M. xanthus* [96, 97]. Benserazide and levodopa were additionally able to inhibit T4P-mediated biofilm formation by *A. nosocomialis* [96]. These findings indicate that the T4P is able to be targeted for drug discovery.

The assembly ATPase is particularly attractive as an AVT target due to its necessity in T4P biogenesis and functionality. Absence of PilF results in a loss of piliation and an apparent attenuation of pathogenesis [57, 91, 98-100]. Although the pilus filament is also a potential target, since it is the extracellular portion of T4P, the practicality is somewhat diminished by general variance in the pilin protein. This is especially true for organisms that are capable of antigenic variation, such as *Neisseria* species, that enables bacteria to alter their pilins through programmed genetic recombination events [101, 102]. Due to its ubiquity in T4P systems, the assembly ATPase is an ideal target for AVT development.

Neisseria gonorrhoeae **Overview**

N. gonorrhoeae, known also as gonococcus, is a Gram-negative obligate human pathogen that uses T4P in the initial stages of its infection [103]. Gonococci are the etiological agent of the sexually transmitted infection (STI) gonorrhea, the second most common STI behind chlamydia in the United States [104]. *N. gonorrhoeae* is classified

as an urgent ABR threat by the CDC, responsible for ~550,000 infections in the United States annually [1]. On a global scale, the WHO estimates gonococci were responsible for 87 million of the 376 million cases of treatable STIs in 2016 [105]. *N. gonorrhoeae* primarily colonizes the urogenital tract but can also infect the nasopharyngeal, ocular, or anal mucosa [18, 102]. Urogenital infections can manifest as urethritis, cervicitis, pelvic inflammatory disease, and increased HIV susceptibility or cause ectopic pregnancy or adverse fetal effects [18, 101].

N. gonorrhoeae is a major public health concern because resistance to all major classes of antibiotics has been identified in this pathogen [101]. A combinatorial therapy of ceftriaxone and azithromycin was previously recommended by the CDC and is commonly reported in the literature; however, due to increasing azithromycin resistance, the present recommended treatment for uncomplicated gonococcal infection is now a single dose of ceftriaxone [106, 107]. There are growing concerns of gonorrhea becoming untreatable by current antibiotics as multiple ceftriaxone-resistant strains of *N. gonorrhoeae* have been recently identified [19, 108-113]. Further, asymptomatic infections, which tends to be more common in women, may serve as a hidden reservoir and allow silent transmission of ABR *N. gonorrhoeae* strains within communities [101]. It is possible that *N. gonorrhoeae* will become an untreatable superbug without the development of new therapeutics.

The Role of T4P in *N. gonorrhoeae* Infection

During the initial stages of infection, the T4P is involved in microcolony formation at the surface of epithelial cells. Additionally, direct interaction between T4P and CR3 receptors expressed at the surface of cervical epithelial cells may be necessary

to initiate gonococcal infection [102, 114, 115]. T4P are additionally known to undergo phase and antigenic variation in *N. gonorrhoeae*. Phase variation is the reversible switching on or off of gene expression. T4P levels can be controlled by phase variation of the *pilC* genes, where piliation occurs when *pilC* is phased on and pili are underexpressed when *pilC* is off [116]. The *N. gonorrhoeae* genome contains multiple silent copies of *pilS* and one copy of *pilE*. *pilS*, which codes polypeptides for pilin, lacks the N-terminal coding region, ribosome binding site, and transcriptional promoter, all of which are present at the *pilE* locus. Homologous recombination events between any overlapping segment of *pilS* and *pilE* cause antigenic variation at the C-terminus of the major pilin PilE [117, 118]. Both methods allow *N. gonorrhoeae* to evade immune detection by the host immune system. Together, these indicate that the T4P is critical to *N. gonorrhoeae* pathogenesis.

Clostridioides difficile **Overview**

Clostridioides difficile, previously *Clostridium difficile*, is a Gram-positive spore-forming obligate anaerobe believed to use T4P for cell attachment and biofilm formation [57, 119]. *C. difficile* infections (CDI) are the most common healthcare-associated infection in the United States, reportedly responsible for ~225,000 nosocomial infections and ~13,000 deaths annually [1]. CDI can manifest as mild diarrhea to severe pseudomembranous colitis and occurs most often in elderly patients [1, 120]. *C. difficile* spores are resistant to changes in temperature and oxygen as well as exposure to common disinfectants including those used in healthcare settings. This allows *C. difficile* to persistent and be transmitted easily in healthcare and community settings through human contact, particularly by the fecal-oral route [30, 120, 121]. Clinical strains remain

susceptible to the antibiotics vancomycin, fidaxomicin, and metronidazole, the presently recommended drugs for CDI treatment [122, 123]. *C. difficile* is not currently considered ABR by the CDC but is particularly noteworthy due to the high incidence of CDI associated with antibiotic use. Despite this, ABR and hypervirulent strains have been isolated with increasing frequency across the globe [1, 121, 123].

There is a strong correlation between antibiotic use and CDI. This is because a balanced microbiome confers host resistance to *C. difficile* infection by competing for carbon sources and producing compounds that prevent spore germination [120]. Antibiotics weaken the healthy microbiome, exposing niches for *C. difficile* to colonize. Energy sources previously metabolized by others, such as sialic acid and succinate, become available to *C. difficile* in a dysbiotic microbiome. Additionally, species that would normally convert primary bile acids into secondary bile acids that normally inhibit vegetative *C. difficile* growth can be eliminated [30, 124]. Accumulation of primary bile acids like taurocholate promotes spore germination, particularly at the anoxic lower gastrointestinal tract [120]. Thus, susceptibility to CDI is strongly modulated by antibiotic use, mostly due to the disruption of a healthy microbiome.

The Role of T4P in *C. difficile* Infection

The *C. difficile* genome contains T4P genes in a primary T4P operon and three secondary gene clusters [70, 125]. Increased levels of cdG promotes host cell attachment, likely through the T4P [56, 57]. The primary T4P operon is upregulated in the presence of cdG and controls T4P expression by riboswitches [125]. Notably, *pilB1* mutants are unable to efficiently bind epithelial cells and cannot persist in the host as long as the wild type [57]. Minor pilins PilJ and PilW can bind DNA, a major component of *C. difficile*

biofilms [126, 127]. Mutating these proteins results in reduced biofilm biomass, suggesting these accessory pilins are needed for robust biofilm formation [127]. *C. difficile* has been shown to produce cellular mats and aggregates in animal models *in vivo*, suggesting biofilm formation may play a role in its pathogenesis [128, 129]. Overall, work on the *C. difficile* T4P supports its roles in epithelial cell attachment, long-term persistence, and biofilm formation [57, 99, 119], but the scope of its involvement in CDI remains unclear.

Small Compound Drug Discovery - Methods Overview

The drug discovery and development pipeline can take from 10 to 15 years on average, and it is estimated to cost upwards of $2 billion from target identification to market release [130-132]. Pre-clinical drug development is an iterative multi-step process that involves target identification and validation, assay development, compound screening, hit identification and verification, and lead generation and optimization [130, 132, 133]. Targets must be accessible to drugs and the impact of the drug must be measurable both *in vitro* and *in vivo*. Assays measuring this effect must also be reproducible, robust, and able to be scaled up for screening methods [130]. Screening involves challenging targets with small compounds and may involve *in vitro* high-throughput screening (HTS) or virtual screening (VS), a form of computer-aided drug discovery. From these screens, compounds that fulfill desired criteria (hits) are isolated and tested *in vitro* and *in vivo* to validate screening results. From these hits, lead compounds are generated and optimized for target selectivity and improved absorption, distribution, metabolism, excretion, and toxicity (ADMET) characteristics [133].

High-Throughput Screening

HTS allows systematic testing of hundreds of thousands of different compounds against the target of interest simultaneously in a large-scale assay, often leveraging robotics for automation. Targets in HTS can range from purified enzymes to whole cell systems [131]. This method is used to identify compounds with drug potential from pre-defined compound libraries [134]. Numerous assay types can be adapted to a HTS format, including those that measure competitive ligand binding, changes in enzymatic activity, and changes in fluorescence [134]. As described previously, potential inhibitors for *C. thermophilum* PilB were identified in two separate HTS. In the first HTS, a competitive binding assay based on displacement of the fluorescent ATP analog MANT-ATP was used [97]. Decreases in fluorescence indicated the compound competes with MANT-ATP in occupying the ATP binding site. The second HTS utilized a malachite green-based approach that measures ATPase activity of PilB [96]. Compounds that inhibit ATPase activity decrease the concentration of liberated phosphates from ATP hydrolysis, resulting in decreased quantifiable absorbance [135]. This method theoretically expands the screen's scope to include both orthosteric and allosteric inhibitors.

Structure-Based Virtual Screening

Structure-based virtual screening (SBVS) is a type of VS that predicts compound binding based on the known structure of a target [132]. SBVS can be viewed as an *in silico* HTS, where compound databases are screened against a computational model of the target protein through compound docking. Models may be derived from experimental data, like X-ray crystallography, or created computationally via homology modeling or

de novo computational methods [132, 136]. These structures represent one conformational state and can be influenced by ligand presence and crystallization conditions. Ensemble docking involves screening a compound database to multiple rigid structures of the same protein in different conformations, thus emulating docking to a dynamic protein rather than a singular state [137]. In SBVS, hit compounds are determined based on their predicted binding affinity. These compounds are evaluated by scoring functions that rank them based on the nature and number of protein-ligand interactions [132, 136]. Molecular dynamics simulations of protein-ligand interactions offer a more robust way to determine binding strength [137, 138]. Such a method was used to evaluate VS hits for monomeric *N. gonorrhoeae* PilF and *P. aeruginosa* PilB, where the three hits remained in stable complex for 100 ns [139]. Simulation of hits provides a more stringent metric before lead optimization but is generally more computationally demanding. SBVS provides a convenient and accessible method to screen for drug-like compounds; however, the hits identified in these screens require *in vitro* or *in vivo* validation.

Molecular Dynamics Simulations

Broadly speaking, molecular dynamics (MD) simulations are computational estimations of changes in a system over time. These estimations are determined by applying Newton's law of motion to every atom in the system, ultimately providing a view of structural and thermodynamic changes at a scale not normally achievable in a typical laboratory setting [140, 141]. Biological systems can be simulated to study biochemical processes, such as protein dynamics, ligand binding, and protein complex stability. System parameters, such as ion content, protonation state, and ligand presence,

can be controlled to simulate dynamics under particular conditions [140]. The energy of atoms is calculated by the force field model employed, which determines how energy functions are applied based on system-defined parameters [141]. MD simulation can be performed using a number of software and accompanying force field models, each best suited for different purposes (i.e., protein versus lipid simulation) [140-142].

The general workflow for MD simulation involves system preparation, energy minimization, equilibration, and production [141]. The first two steps initialize the system for MD, defining solvent and ion parameters and relaxing the system to a lower energy state by resolving conformational clashes [143]. Equilibration brings the system to a steady state, so the final trajectory is relevant to the desired conditions. This occurs in two phases: equilibration at constant temperature and volume (NVT) and equilibration at constant temperature and pressure (NPT) [141, 144]. The final stage, production, is the simulation from which the final trajectory data is derived and analyzed. Production must run for an adequate length of time to sufficiently sample a protein's conformational space, or to achieve trajectory convergence. Convergence can be determined by a number of methods, including measuring root-mean-square deviation (RMSD) over time, block averaging RMSD, and cluster analysis [141, 144, 145].

The use of MD simulation in conjunction with VS methods has increased steadily in the past 10 years, with approximately 39% of VS studies using MD simulation between 2015 and 2020 [146]. MD simulation can be used to analyze protein-ligand interactions, as discussed previously, to rationalize docking results and potentially reveal conformational changes [139, 146, 147]. Structures can also be extracted from MD trajectories and used for ensemble docking via clustering methods. Clustering algorithms

identify the most common protein conformations taken during the simulation based on structural similarities, or within a defined RMSD cut-off [142, 146, 148]. More recently, dynamic docking with MD simulations has been explored, wherein the process of ligand binding and unbinding is simulated. Although dynamic docking tends to be computationally expensive for routine use in VS when compared to other long-standing applications, it can provide greater insights on mechanisms of action and binding pathways [149, 150]. MD simulations may begin to play a greater role in the VS process as computing resources continue to become more accessible.

Tables and Figures for Chapter 1

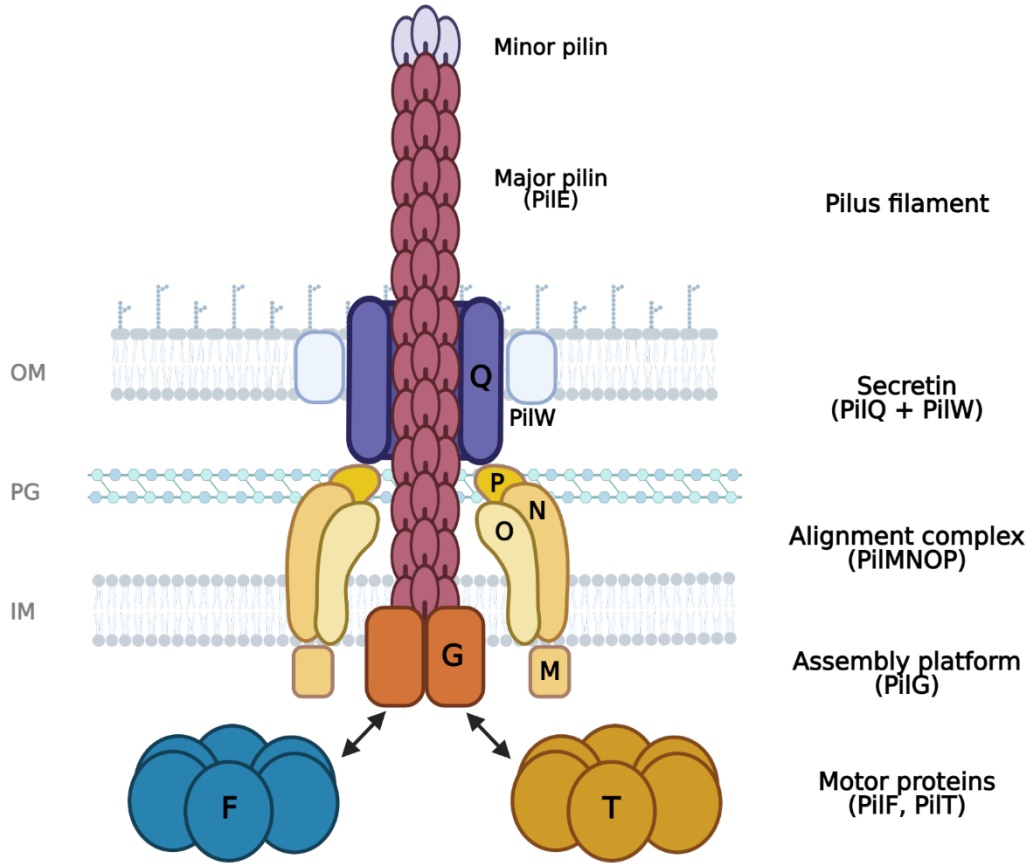

Figure 1-1. Schematic of the T4P machine in *N. gonorrhoeae.*
The T4P machine in Gram-negative bacteria is comprised of a pilus filament that travels through a pore created by adjoined secretin and alignment complexes that spans from the inner membrane to the outer membrane. Pilus assembly and disassembly is driven by two antagonistic ATPases that interface with the T4P machine at the cytosolic face of the inner membrane. **OM** = outer membrane, **PG** = peptidoglycan, **IM** = inner membrane. Image generated using BioRender.

Chapter 2 - Evaluating the use of *Chloracidobacterium thermophilum* PilB in virtual screening for inhibitory compounds

Abstract

The antivirulence strategy has been explored as an alternative to traditional antibiotic development. The bacterial type IV pilus is a virulence factor that aids host invasion and colonization in multiple antibiotic resistant bacterial pathogens. The PilB ATPase is critical because it hydrolyzes ATP to drive the assembly of the type IV pilus filament from pilin subunits. Recent studies have identified small-molecule inhibitors of the PilB ATPase. Here, we evaluate *Chloracidobacterium thermophilum* PilB (*Ct*PilB) as a model for structure-based virtual screening by molecular docking and molecular dynamics. A hexameric structure of *Ct*PilB was generated through homology modeling based on an existing crystal structure of PilB (5TSH). Using the *Ct*PilB homology model, molecular docking was performed to assess the binding of six PilB inhibitors described in the literature. From molecular dynamics simulations of *Ct*PilB in four different liganded states, we obtained four representative structures to examine the conformational plasticity of PilB and to improve docking analyses of the inhibitors by ensemble docking. Binding analyses revealed key residues for each of the six inhibitors investigated. Three known competitive inhibitors are found to consistently interact with conserved residues in the ATP binding pocket. One inhibitor from another study exhibits the greatest binding affinity but occupies the ATP binding pocket primarily through interactions with unconserved residues. The work in this chapter also gauges the use of a PilB monomeric model in virtual screening by evaluating the two top hits from a previously published virtual screen by molecular docking. The use of multiple conformations in molecular

screening can provide greater insight into compound flexibility within receptor sites and better inform future drug development.

Introduction

The bacterial type IV pilus (T4P) is a filamentous surface adhesin involved in an array of bacterial functions including surface attachment, biofilm formation, cell motility, and DNA transformation [43]. Reminiscent of a grappling hook, the pilus can extend several μm from the cell body, adhere to a surface, and retract, pulling the cell body towards the point of attachment [44]. T4P is one of the strongest known molecular motors, generating upwards of 150 pN during retraction [36]. Of the two T4P subtypes, the T4aP is most commonly associated with bacterial pathogens [41]. As a virulence factor, T4P contributes to the initial stages of infection by mediating surface attachment to clinical implements or host epithelial cells [52, 99]. Cells attached by T4P may stably aggregate into a biofilm, progressing infections into chronic disease [55, 63]. In some cases, infection is initiated through interactions between pili and host cell surface receptors [114, 151]. It is clear the T4P is a critical virulence factor in a number of pathogenic bacteria, including *Acinetobacter baumannii*, *Clostridioides difficile*, *Neisseria gonorrhoeae*, and *Pseudomonas aeruginosa*. The virulence of these organisms is notably attenuated when T4P is absent or abrogated [57, 64, 100, 151]. Due to its roles in pathogenesis, especially in priority bacteria pathogens, there is great interest in developing antivirulence therapeutics that target the T4P [1, 6, 29].

Pilus biogenesis requires the T4P supramolecular machinery composed of at least 10 conserved proteins [81]. Cytoplasmic ATPases at the base of the supramolecular machine provide the necessary energy for pilus extension and retraction through ATP

hydrolysis. Pilin polymerization, or extension, is driven by PilB, the assembly ATPase that allows pilin to be incorporated into the filament from pools of monomeric pilins in cell membrane. Depolymerization is conversely driven by a PilT retraction ATPase in T4aP systems [43, 81]. Repeated extension and retraction events by these ATPases are responsible for processes like twitching motility and natural transformation [44, 45, 50, 64]. Importantly, mutant bacteria without PilB have no pili and are diminished in pathogenesis, identifying PilB as a promising target for antivirulence drug development [57, 91, 98-100].

The PilB is a member of the ATPases associated with diverse cellular activities (AAA+) family, which belongs to the superfamily of additional strand catalytic E (ASCE) ATPases [82, 83]. PilB was previously characterized as a hexamer with C_2 symmetry where conformational states are reflected across the protein in paired protomers [89, 90, 92, 152]. Of the two primary annotation styles used to refer to PilB and PilT functional units [89, 90, 92, 152, 153], the packing unit annotation style presented by [90, 92, 153] best demonstrates how conformational changes may be propagated across the protein. Thus, the functional catalytic structure of PilB may be viewed as units comprised of the CTD of one chain and the N2D of the adjacent chain, or CTD_n:$N2D_{n+1}$ (FIG 2-1A, 2-1B) [90, 153]. The ATP binding interface can be found within the catalytic cavity of CTD_n formed at the surface between two packing units. These interfaces take on one of three conformational states: open (unliganded), closed (releases ADP), and open' (post-hydrolysis resting) (FIG 2-1C) [89, 90]. The ATP binding cavity is occluded in both the closed and open' interfaces by the adjacent packing unit (FIG 2-1C). In contrast, open interfaces are created by two parallel packing units,

creating an open channel for ATP binding (FIG 2-1C). Simultaneous ATP binding and hydrolysis followed by ADP release is thought to produce conformational changes within packing units that is propagated across the hexamer. This results in a simultaneous open-open', closed-open, and open'-closed transition that manifests as a rotary motion and is further propagated to the platform protein that builds pilin into the filament [81, 89, 90, 92].

At the monomeric level, PilB is comprised of two N-terminal domains (N1D and N2D) connected by a flexible linker to a C-terminal domain (CTD) containing the catalytic ASCE core [90, 92]. Within the ASCE core are conserved Walker A and Walker B structural motifs that are essential for ATP hydrolysis. The Walker A motif, also known as the P-loop, has the consensus GXXGXGK(S/T) that binds the phosphate groups and coordinates Mg^{2+} within the binding pocket. The conserved lysine is critical for interaction with the γ-phosphate, and mutation of this residue results in a significant decrease in ATP binding [82-86]. The PilB Walker B motif is atypical for AAA+ ATPases, with the sequence hhhhGE (where h is a hydrophobic residue). The catalytic glutamate promotes hydrolysis of the γ-phosphate by activating a water molecule for nucleophilic attack [83, 84, 86]. PilB also displays the Arg finger, Asp box, His box, and tetra-cysteine motifs defined for AAA+ extension ATPases [89, 90]. The arginine fingers lay N-terminal to the Walker A P-loop and orient toward the γ-phosphate of a bound ATP, believed to function as a phosphate sensor and mediator of intersubunit cooperativity [82, 85]. A glutamate within the Asp-box is thought to coordinate Mg^{2+} while the His-box histidine coordinates the γ-phosphate and may additionally contribute to intersubunit communication. Piliation levels are decreased by mutations of the

conserved residues in both the Asp-box and His-box motifs [90, 91]. Zn^{2+} binding at the tetra-cysteine motifs primarily functions to maintain structural stability of the hexameric ring [88, 89].

Virtual screening can serve as a complementary method to *in vitro* high-throughput screening. In particular, structure-based virtual screening (SBVS) is a target-based *in silico* screening method where library compounds are systematically docked into a three-dimensional model of the target protein [132, 136]. Structures are determined by experimental methods such as NMR spectroscopy or X-ray crystallography, or computationally generated such as homology modeling [132]. Docking, in essence, imitates the protein-ligand binding process and produces multiple poses of the ligand within a given cavity. These poses are then ranked based on predicted binding affinity [146]. By nature, experimentally derived structures represent a static conformation of a protein at a given condition, and they do not represent the dynamic nature of a protein under native conditions. Ensembles of different structures can be used in SBVS to compensate for the lack of protein flexibility in these structural models [137]. These structures may be determined experimentally or through molecular dynamics (MD) simulations. Ensemble docking has previously been successful in validating the binding of compounds not able to be understood by a single structural model, proving to be a useful approach for virtual screening [137, 154].

There are two studies that have used computational methods to identify or evaluate inhibitors for the T4P assembly ATPase [95, 139]. Phenotypic screening for compounds that decrease cell adhesion and aggregation in *Neisseria meningitidis* identified “Compound B”, often referred to as P4MP4, as a potential inhibitor of the PilF

assembly ATPase [95]. Potential binding sites were investigated via molecular docking, revealing that P4MP4 may bind to the ATP binding pocket PilF. A second study performed virtual screening of a library of natural compounds containing piperidine moieties to identify inhibitors for *N. meningitidis* PilF and *P. aeruginosa* PilB [139]. This VS was performed against monomeric models of the respective proteins, and the binding stability of the three best hits were evaluated by MD simulation. After 100 ns of simulation, all three compounds remained within 2 Å of ATP binding pocket, with two compounds (CN0051517 and CN0030078) exhibiting binding free energies of > -10 kcal/mol as evaluated by AutoDock Vina.

Previous work using *Chloracidobacterium thermophilum* PilB has shown that the assembly ATPase can be used in high-throughput screening for PilB inhibitor with broad inhibitory activities in different bacteria [96, 97]. The hits from these studies include quercetin, benserazide, and levodopa. These compounds were found to inhibit *C. thermophilum* PilB *in vitro* and T4P-associated functions in *Myxococcus xanthus in vivo* [96, 97] Benserazide and levodopa were both additionally found to inhibit T4P-associated motility and biofilm formation in *Acinetobacter nosocomialis*, an opportunistic pathogen that is part of the *Acinetobacter calcoaceticus-Acinetobacter baumannii* complex [96, 155]. Further, the work in chapter 3 shows that quercetin is able to inhibit *Neisseria gonorrhoeae* PilF in a dose-dependent manner. These data validate *C. thermophilum* PilB as an experimental model for antivirulence research, supporting its application as a computational model for virtual screening.

This chapter seeks to validate *C. thermophilum* PilB (*Ct*PilB) as a model for anti-PilB virtual screening by molecular docking of the inhibitory compounds previously

reported as discussed above. By extension, this chapter investigates if virtual screening with a single monomeric unit, rather than a hexameric complex, is sufficient for structure-based virtual screening by comparison to published computational data [139]. Toward this end, we generated a structure of the *Ct*PilB ATPase through homology modeling based on the resolved structure of *Geobacter metallireducens* PilB (*Gm*PilB). This *Ct*PilB structure was used for molecular docking of ATP and ADP as well as the inhibitory compounds noted above. Four models of *Ct*PilB in varying liganded states were additionally generated and MD simulations were conducted to investigate morphological changes due to nucleotide binding. Representative structures were extracted, and molecular docking was repeated to assess changes in inhibitor binding caused by conformational plasticity after MD simulation. The results herein provide insight into the structural aspects of successful PilB inhibitors and can inform the design of future anti-PilB compounds.

Methods

Generation of the *Ct*PilB Homology Model

The sequence for *C. thermophilum* PilB was obtained from UniProt (Accession ID: G2LDI3) [156]. A homology model of hexameric *C. thermophilum* PilB was generated through the Advanced Homology Modeling module of the Biologics software in Schrodinger Maestro 12.5 using the crystal structure of *G. metallireducens* PilB (RCSB PDB ID: 5TSH) as a template [90]. Due to the limitations of available crystal structures, the N1D region of *Ct*PilB was unable to be modeled, resulting in a structure that begins at residue 218. Zn^{2+} and Mg^{2+} ions were modeled at the corresponding positions as the model template. The homology model was energy minimized in

Schrödinger's Prime software and validated with Ramachandran plots and QMEANDisco through the SwissModel tool suite and the ProSA web server [157-159].

Ligand and Structure Preparation for Molecular Docking

Structure preparation, molecular docking, and energy calculations were performed using the SiteMap, Prime, and Glide programs of the Schrödinger Maestro 12.5 software suite. Non-nucleotide inhibitory ligands were selected based on reports in literature [95-97, 139]. Ligand structure data files were downloaded from PubChem, the COlleCtion of Open Natural prodUcTs (COCONUT) online database [160], or recreated with the Schrödinger Maestro 2D Sketcher. All ligands were minimized using the Schrödinger LigPrep under the OPLS4 force field. The *Ct*PilB homology model was used for docking prior to MD simulation. After MD simulation, the structures that occupied the greatest percentage of the sampling period after structural clustering (see Trajectory Analysis) across all three replicates for a given condition were selected for docking. All protein structures were prepared using the Schrödinger protein preparation wizard to assign hydrogens and bond orders then energy minimized using Prime with the OPLS4 force field and VSGB solvation model. Binding sites were validated using the SiteMap software. Because all top predicted binding pockets overlapped with the ATP binding domain, the docking grid was defined by centering on the Walker A box residues in each protein chain and restricting box size to 20 Å which was sufficient to also capture the Walker B and other AAA+ ATPase-associated motifs. The Glide [161] program was used to generate a maximum of nine poses using SP mode and default parameters. The free energy of binding (ΔG_{Bind}) of each pose was calculated by the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) [162] module of Prime using the OPLS4

force field and VSGB solvation model. Ligand-protein fingerprint analysis was performed on the lowest energy poses of each ligand tested using the Schrödinger Interaction Fingerprints module [163, 164]. Cut-off distances for heavy atom and hydrogen atom interactions were set to 4.0 Å and 2.5 Å, respectively.

System Construction and Molecular Dynamics Simulations

All molecular dynamics simulations of the PilB hexamer were performed using the 2020.4 build of the GROMACS software suite and built with the July 2021 release of the CHARMM36m forcefield [165, 166]. Four systems were simulated with three replicates each for a total of 12 simulations. A TIP3P water model [167] in a cubic box was used with heavy atoms restrained and a minimum solute-box distance of 1.0 nm. Systems contained 150 mM NaCl and were neutralized with Na^+ and Cl^- counter ions. For systems containing nucleotides, ATP or ADP were pre-docked into the appropriate subunit interfaces using the Glide software in the Schrödinger software suite (see above). Energy minimization was performed with the steepest descent algorithm, with a maximum force constraint of 1000 kJ/mol/nm. Systems were equilibrated with a constant volume and temperature ensemble (NVT) at 298 K for 100 ps using the Berendsen weak-coupling algorithm [168]. Random starting velocities were assigned to each replicate from a Maxwellian distribution. An isothermal-isobaric ensemble (NPT) was then applied for 100 ps with a modified Berendsen thermostat [168] held at 298 K and Berendsen pressure coupling held at 1 bar. All production simulations were performed using the CHARMM36m forcefield [166]. Short-range cutoff of 1.2 nm was applied to all van der Waals interactions. Long-range electrostatic interactions were calculated using the particle-mesh Ewald method [169] with cubic interpolation, 0.16 nm Fourier grid

spacing, and Verlet cut-off scheme. Bonds were constrained by a fourth order P-LINCS algorithm [170], with an integration timestep of 2 fs. MD simulations were performed for 1 μs for all twelve replicates, totaling 12 μs of simulation time.

MD Simulation Trajectory Analysis

RMSD and RMSF analyses were performed using the initial homology structure as the reference structure. RMSD and RMSF analyses were done to determine the structural stability of each replicate system. Conformational convergence was observed over the final 500 ns of simulation, providing a total of 6 μs sampling time. Structural clustering was performed based on the protein backbone with the gromos algorithm and a RMSD cut-off of 0.275 nm. All analyses presented herein were performed over this 500 ns sampling period using predominant conformations determined by clustering. RMSD, RMSF, and clustering were generated and analyzed using GROMACS 2020.4 and in-house Python 3.9 scripts [165]. Visualization analysis was performed with PyMOL.

Results and Discussion

Evaluation of the *Ct*PilB homology model. Crystallization efforts have not yet been able to visualize the entire PilB protein such that only structures of the N2D and CTD regions together or N1D separately have been determined experimentally [89, 90, 92, 152]. Due to this, our homology model for *Ct*PilB (*Ct*PilB$_{HM}$) lacks the first 217 amino acids corresponding to the N1D region. Notably, the absence of the N1D domain does not appear to severely impact ATPase activity *in vitro* [58, 95]. Evaluation of the homology model via Ramachandran plots show that 96.84% of dihedral bond angles are favored (FIG 2-2A). Outliers were located within four flexible loop regions located at the external face of PilB. This is supported by a QMEANDisCo score of 0.77 ± 0.05 with the

lowest structural confidence at, again, the same flexible loop regions (FIG 2-2B). The structure quality of *Ct*PilB$_{HM}$ is within expected quality of similarly sized proteins determined by X-ray crystallography with a ProSA *Z-score* of -9.36 (FIG 2-2C) [159]. ProSA additionally determined that local model quality is overall favorable except at a flexible loop spanning from I296 to L307 (FIG 2-2D). Though this specific loop region appears to be extended in *Ct*PilB$_{HM}$ compared to the *Gm*PilB template, all other regions are strongly consistent with the template structure.

To further validate our structure, we performed molecular docking into the *Ct*PilB structure with ATP and ADP and assessed the ligand-protein interactions with the key ASCE ATPase residues, particularly the Walker A and B motifs (FIG 2-3A). As expected, the top binding pockets were found to overlap with the ATP binding pocket, and ATP was docked into the two open interfaces and ADP into the remaining closed and open' interfaces (FIG 2-3B). Both ADP and ATP generated poses where the phosphates are able to interact with the Walker A P-loop; however, some interfaces produced poses that are more U-shaped rather than the typical linear conformation found in crystal structures (FIG 2-4). This may be due to the size of the binding surfaces identified by the SiteMap program, which generates a surface based on receptor properties that are favorable for generalized ligand binding (FIG 2-3B) [171]. Despite this, ATP in *Ct*PilB still exhibited most expected interactions with the binding cavity residues, primarily via the β- and γ-phosphates (Table 2-1). Within the Walker A motif, ATP primarily interacts with T366 to T371, where the phosphate groups of ATP are surrounded by the P-loop backbone amides. No interactions with the first Walker A glycine G364 were detected, and P365 only participates in backbone interactions. All residues spanning from T366 to

T371 interact with the β- and γ-phosphates by the backbone amine group to some degree. T371 additionally makes hydrogen bonds with the β- or γ-phosphate. Interestingly, G367 to G369 seem to participate in hydrogen bonding with ATP when they are generally expected to interact via backbone amides only (Table 2-1) [85]. The arginine finger residue R310, Walker B catalytic glutamate E434, and Asp-box glutamate E396 contact ATP via sidechain interactions, orienting toward the triphosphate tail. The last three of these residues are expected to coordinate Mg^{2+} under biological conditions [83]. Interestingly, ATP makes an additional contact with R436, adjacent to the Walker B element. This arginine was previously noted to be critical for ATPase function, where mutation of this residue resulted in an 18-fold decrease in activity [156]. This contact was only seen for ATP, and not ADP.

Docking ADP into *Ct*PilB produced similar interactions, though ADP makes fewer overall contacts with the closed interface than the open' interface in the generated poses (Table 2-1). Contacts with R310 and E392 are lost in both closed and open' interfaces, though R325 still remains poised toward the nucleotide. Interactions with the Walker A residues are maintained, with new hydrogen bonds with K370 emerging through its side-chain group. Interestingly, His-box H459 creates a new contact by its side chain to the α-phosphate of ADP only in the closed binding interface. This histidine was found to coordinate the γ-phosphate group of AMP-PNP and not ADP in the open' interface of the *Gm*PilB structure [90]. ATP and ADP also showed a notable number of interactions with other arginine and aspartate residues either within CTD_n or at the adjacent packing unit. Hydrophobic interactions occurred primarily with leucine residues, especially L327, which neighbors the arginine finger R325. In all, both ATP and ADP

exhibit the key ligand-protein interactions within the ATPase catalytic pocket of the *Ct*PilB homology model, despite the variety of poses generated.

Evaluation of docked inhibitors to the *Ct*PilB homology model. Six compounds identified as potential or validated PilB inhibitors were docked into the *Ct*PilB model, and their binding was evaluated (FIG 2-5, FIG 2-6) [95-97, 139]. All six were predicted to bind at the ATP binding site in their respective studies. Free energy of binding (ΔG_{Bind}) was individually evaluated within the three interface conformations, where a more negative value suggests a greater binding affinity (Table 2-2). Table 2-3 contains a summary of ligand-complex residue interactions as determined by interaction fingerprinting of the best binding poses. All compounds make at least three contacts with a key ASCE residue, with five of six compounds exhibiting polar side-chain interactions with T371 in particular.

Quercetin exhibited the most consistent ΔG_{Bind} across the three conformations with an overall average of -27.67 ± 5.32 kcal/mol, indicating a fairly equal affinity for any of the three conformational states (Table 2-2). Quercetin docks just outside of the P-loop, contacting R325, T371, and E396 via hydroxyl groups (FIG 2-6C). The structurally similar compounds Benserazide and levodopa exhibit very similar overall average ΔG_{Bind} values of -22.76 ± 8.82 and -21.86 ± 7.55 kcal/mol, respectively (Table 2-2). Both of these compounds interact with the majority of Walker A motifs via their 'tail', especially levodopa which sits within the P-loop at the carboxyl end (FIG 2-6A). The favorable ΔG_{Bind} values and best binding poses for these three compounds are in line with existing experimental data where Benserazide, levodopa, and quercetin were all demonstrated to inhibit *Ct*PilB ATPase activity *in vitro* as competitive inhibitors [96, 97].

The natural compounds CNP0030078 and CNP0051517 also produce similar ΔG_{Bind} as each other with -27.09 ± 11.71 kcal/mol and -27.20 ± 12.10 kcal/mol, respectively. Structurally, these compounds differ only by a single fluorine group. CNP0030078 and CNP0051517 appear to have greater affinity for the open and open' interfaces as indicated by the two-fold difference in the ΔG_{Bind} value for the closed interface (Table 2-2). Notably, CNP0030078 made the least number of key residue contacts of all docked compounds, interacting with only three key residues (R325, G367, G369) CNP0051517, in comparison, makes six additional contacts with key residues (Table 2-3). Comparing the best binding poses of these two compounds reveal that the benzoyl group of CNP0051517 docks within the Walker A P-loop, but the same substituted ring in CNP0030078 does not. This suggests that CNP0051517 has potential as a competitive inhibitor at the ATP binding pocket, but CNP0030078 may not be able to fully enter the ATP binding pocket due to the additional fluorine substitution, despite its favorable ΔG_{Bind} value (FIG 2-6B).

P4MP4 displays the most favorable ΔG_{Bind} of all compounds (-34.48 ± 12.01 kcal/mol); however, there were also fewer poses generated during molecular docking (Table 2-2). Specifically, only one pose was generated when docked into the open' interface. P4MP4 is a bulky compound (564.73 g/mol) and occupies a large portion of the binding pocket, which is already partially occluded by the adjacent packing unit in the closed and open' interfaces (FIG 2-6C). The size of P4MP4 may allow it to interact with a greater number of residues, making contacts with all arginine fingers, the Walker B glutamate, an Asp-box glutamate, and six of eight Walker A residues (Table 2-3). Flexibility of this compound is likely restricted within the binding interface, accounting

for the lack of additional poses and reflected in the comparatively low standard deviations of the ΔG_{Bind} values.

***Ct*PilB can be simulated in different liganded states.** The molecular docking was performed using a single homology model of *Ct*PilB, which only represents a snapshot of a typically dynamic protein. Ensemble docking has previously been used to compensate for the lack of dynamic information in these static structures by using additional structures in different conformational states [137]. The inclusion of varying conformations may also reveal different ligand poses within the binding pocket that would not normally be seen with just a single crystal structure. To generate additional structures for molecular docking and to investigate major morphological changes due to nucleotide binding, we performed MD simulations of *Ct*PilB in the apoprotein ($CtPilB_{Apo}$), partially liganded ($CtPilB_{ATP}$ and $CtPilB_{ADP}$), and fully occupied ($CtPilB_{ATP+ADP}$) states. In the $CtPilB_{ATP}$ and $CtPilB_{ADP}$ simulations, ATP or ADP was docked into the open or closed interfaces, respectively. ATP and ADP were docked in the open and closed/open' states as described in the previous section for the $CtPilB_{ATP+ADP}$ simulations.

Simulation convergence was evaluated by visualization of change in root-mean-square deviation (RMSD) over time (FIG 2-7) and cluster analysis (FIG 2-8). The appearance of a plateau in average RMSD of all four systems indicates convergence is achieved in the final 500 ns of the simulations (FIG 2-7). This is supported by the predominant structures identified by clustering analysis with a 0.275 nm RMSD cutoff, where the major structure extracted from each replicate occupies at least 50% of the final 500 ns sampling period (FIG 2-8). Root-mean-square fluctuation (RMSF) indicates four

regions of high flexibility, which correspond to the four disordered loops identified previously by our validation methods (FIG 2-9). The fluctuation in these regions appears consistent across all replicates within each system, further supporting that convergence was achieved.

Morphology of *Ct*PilB structures changes after MD simulation with ATP/ADP. Examination of the predominant clustering structures showed that nucleotide binding at one interface appears to have propagative effects on the overall protein structure (FIG 2-8). Across all three replicates, the *Ct*PilB$_{Apo}$ and *Ct*PilB$_{ATP+ADP}$ systems generated predominant structures that either appear more symmetrical overall or closely identical to the homology model, respectively. The *Ct*PilB$_{ATP}$ and *C*tPilB$_{ADP}$ simulations showed greater heterogeneity, both producing at least one predominant structure that has a visible skew or deformation when the entire structure is viewed (FIG 2-8). Due to the limitations and resolution of MD simulation, the process of ATP hydrolysis cannot feasibly be simulated [172]. However, conformational deformation in the *Ct*PilB$_{ATP}$ and *C*tPilB$_{ADP}$ structures occurs at unoccupied interfaces adjacent to a bound interface, suggesting the presence of ATP or ADP is influencing the greater morphological character. These skewed structures occupy the least amount of sampling space compared to the other predominant structures in other replicates, which may be indicative of a less frequently visited conformation.

Key residue interactions with ATP/ADP are maintained after MD simulation. The ligand-protein interactions between *Ct*PilB and ATP or ADP after MD simulation were reevaluated by interaction fingerprinting analyses (Table 2-4). In at least two of three replicates, most interactions of ATP and ADP that were retained after

simulation occur with conserved residues. Non-conserved residues that were also retained neighbor a key motif, usually an arginine finger (R293) or the P-loop (N372). Notably, the structurally equivalent arginine of R293 in *Thermus thermophilum* PilB was previously described as an arginine finger [92]. R293 may not function as an arginine finger in *Ct*PilB, as interactions with R325 were more prevalent prior to and after MD simulation (Table 2-1, Table 2-4). Hydrogen bonding with P-loop residues is lost at most interfaces, with all previously recorded contacts preserved at just one open' interface. Backbone interactions are variably maintained, with G367 or G369 shared across most systems as expected [85]. Interactions with the Walker B glutamate E434, which assists in coordinating ATP hydrolysis, were only detected in ATP-bound interfaces. In contrast, ADP-bound interfaces experienced more interactions with the Asp-box glutamate E396. Overall, ATP and ADP maintained most expected interactions with the ATP binding pocket, particularly with the Walker A and B motifs.

Surprisingly, two binding interfaces in the *Ct*PilB$_{ATP+ADP}$ systems retained only arginine contacts (Table 2-4). This is potentially due to non-linear initial docking poses described previously, allowing ATP or ADP to move within the binding pocket. Previous work has shown that mutations of AAA+ ATPase arginine residues result in diminished or complete abolishment of ATPase activity, though they are also broadly known to coordinate ATP via the γ-phosphate [85, 173]. Continued contact with arginine residues suggests that these residues are critical to holding ATP or ADP in the binding pocket of *Ct*PilB.

Inhibitor docking after MD simulation reveals key interactions. To perform inhibitor docking with *Ct*PilB after MD simulation, the predominant structure that

occupied the greatest proportion of the 500 ns sampling period in each system was selected as a representative structure for ensemble docking (FIG 2-8). All four representative conformations occupy greater than 80% of the sampling period and appear more morphologically symmetric or similar to the *Ct*PilB homology model when the protein structures are overlayed (average RMSD of 5.16 ± 0.51 Å). Reevaluation of potential binding pockets by SiteMap again identifies the ATP binding domain as the most probable ligand pocket. The same six compounds were docked into each representative structure and their ΔG_{Bind} was reassessed. All compounds exhibit favorable ΔG_{Bind} values in the four additional structures, showing generally similar binding affinity in all docked structures (Table 2-5). The residue contacts made by the best binding compounds in the *Ct*PilB$_{Apo}$, *Ct*PilB$_{ATP}$, *C*tPilB$_{ADP}$, and *Ct*PilB$_{ATP+ADP}$ structures are summarized in Table 2-6, Table 2-7, Table 2-8, and Table 2-9 respectively. The best binding poses were mainly generated in the open or open' binding interfaces, except in the *Ct*PilB$_{ATP}$ structure where four of six best poses were in a closed interface. Comparison of pre-MD and post-MD docking reveals that certain interactions are important for inhibitor binding. All compounds contact at least one key catalytic residue, usually an arginine finger or Walker A P-loop residue.

Binding analyses of the three known PilB inhibitors reveals that these compounds bind the most key catalytic residues of the six examined here (Table 2-10). Quercetin interacts with five of the eight Walker A residues, but specifically maintained contact with G367, T371, and at least one arginine finger in three of the four representative structures. Quercetin interacts with the Walker B glutamate and His-box H459 in one structure and maintains the E396 contact in two. When all post-MD poses are aligned to

the best pose from *Ct*PilB$_{HM}$, quercetin has an average RMSD of 0.19 ± 0.08 Å (FIG 2-10, Table 2-11). The flavone backbone has limited flexibility and is able to rotate primarily across one bond, which is reflected in quercetin's best binding poses (FIG 2-10). Interactions with the majority of the P-loop region and at least one arginine finger are preserved with Benserazide in all structures, but Asp-box contacts occurred in only one structure (Table 2-10). Levodopa has a similar profile, retaining interactions with the P-loop in all structures but losing Asp-box and Walker B contacts. This compound appears to be able to contact arginine fingers in these structures, which was not observed in the original homology model (Table 2-3, Table 2-10). When compared to the respective *Ct*PilB$_{HM}$ poses, Benserazide has an average RMSD of 1.84 ± 0.57 Å while levodopa has an average of 0.89 ± 0.68 Å (Table 2-11). The overlay of best poses shows that the 'tails' of Benserazide and levodopa are able to rotate and contact the P-loop freely and have limited rotation at the trihydroxyphenyl- and dihydroxyphenyl- groups respectively (FIG 2-10). The homogeneity of levodopa poses are likely due to its relatively small size. These compounds are likely able to occupy most key residues involved in ATP catalysis, supporting the existing experimental knowledge that these three compounds function as competitive inhibitors of ATP.

The natural compounds CNP0030078 and CNP0051517 do not retain many interactions between the pre-MD and post-MD structures. CNP0030078 interacts with R325 in two of four post-MD structures. It is able to bind some Walker A residues in three of four structures but only the G367 interaction is preserved in two of the four (Table 2-10). Similarly, CNP0051517 contacts the P-loop via its benzoyl group in three structures but other interactions are generally lost, with no key interactions identified in

the *Ct*PilB$_{ATP}$ structure (Table 2-10). CNP0030078 and CNP0051517 have average RMSD values of 1.98 ± 1.25 Å and 2.10 ± 1.01 Å, with the greatest flexibility at the terminal benzoyl and trifluoromethylphenyl group (FIG 2-10, Table 2-11). Fluorination of compounds is often used in drug development to improve compound bioavailability, stability and ligand-protein interactions due to its high electronegativity and relatively small van der Waals radius (1.47 Å) [174]. The fluorine groups within the molecules likely account for most of the miscellaneous interactions observed, which are more prevalent than those in key motifs. Notably, the best poses identified in this chapter are similar to those found in the original study (FIG 2-10). Both compounds are positioned with trifluoromethylphenyl groups more interior than the benzoyl group in both this study and the study by Ozcan et al. [139] (FIG 2-6). Poses found by Ozcan et al. depict very few interactions with key motifs, specifically only a single threonine or glycine within the Walker A loop per compound [139]. The presence of adjacent packing units may influence how these compounds bind into PilB by introducing additional residues outside of the binding cavity that prevent trifluoromethylphenyl from entering the pocket. Nonetheless, both the original study and this chapter indicate that CNP0030078 and CNP0051517 are able to bind PilB to varying degrees, though their effectiveness *in vitro* has yet to be explored.

Similar to the natural compounds above, P4MP4 predominantly interacts with amino acids outside of the key motifs. P4MP4 makes variable contact with the P-loop residues, particularly P365 in three structures and G367 and T371 in two structures. At least one arginine finger interaction was identified in two of the four structures. Again, Asp-box contacts are lost, but at least one structure generated a new interaction with His-

box H459 (Table 2-8). Overall, the only contacts P4MP4 displays across multiple structures are with three Walker A residues P365, G367, and T371 (Table 2-10). P4MP4 also showed the greatest average RMSD at 3.65 ± 0.25 Å, likely owing to its greater molecular size and multiple flexible bonds to large groups (Table 2-11). Despite its average ΔG_{Bind} value being the most favorable of the compounds across all structures, P4MP4 has a reported IC_{50} of 175.10 ± 13.08 μM and only inhibits *N. meningitidis* PilF by ~35% when incubated at 200 μM [95]. Notably, the additional structures allowed more than one pose to be generated for all three interface states; however, not all binding pockets produced poses for P4MP4 in the post-MD structures. Collectively, it is possible that P4MP4 is unable to efficiently bind the ATP binding interface due to its size, despite its high binding favorability (Table 2-2).

The binding data have implications on the quality of these inhibitors as lead candidates, specifically considering compound size and structure. P4MP4 is particularly interesting as its molecular weight is 564.73 g/mol, which violates Lipinski's rule of 5 (RO5). The RO5 states a compound with a molecular weight greater than 500 may have poor absorption in the body when administered orally (FIG 2-5) [175]. The RO5 is often used to filter compounds that are less likely to be drug-like in virtual screening, as seen in the study by Ozcan et al. [139]. By this metric, P4MP4 may not be an ideal candidate for drug optimization, but the structure activity relationship analyses provided by Aubey et al. may inform future endeavors [95]. In their original study, Aubey et al. indicate that two major groups contributing to P4MP4 activity are the 2,4-dimethoxybenzene and piperidine moieties [95]. The three inhibitors known to competitively bind the ATP binding pocket in PilB (benserazide, levodopa, and quercetin) contain a di- or tri-

substituted rings with hydroxyl groups and resemble 2,4-dimethoxybenzene structurally (FIG 2-5). Similarly, the two natural compounds CNP0030078 and CNP0051517 contain highly electronegative fluorine groups (FIG 2-5). These likely mimic the negatively charged triphosphate tail of ATP and indicates that leads are more likely to be successful if they exhibit this characteristic, although the inclusion of highly electronegative groups such as fluorine may not be ideal.

In order to evaluate PilB as a model for virtual screening, this chapter utilized homology modeling to generate a hexameric computational model for the *C. thermophilum* PilB. Molecular docking was performed to investigate the binding characteristics of compounds that were found in recent published literature that may have inhibitory activity against PilB, which was supplemented by additional structures that were generated through molecular dynamics. Docking metrics retrieved using the hexameric *Ct*PilB complex align with experimentally derived results, where three compounds previously confirmed as orthosteric inhibitors were found to occupy the ATP binding pocket at multiple key residues, primarily the Walker A motif. The three other compounds investigated in this chapter may also occupy the ATP binding pocket, but they do not contact many key amino acids. Additionally, virtual screening using the monomeric subunit versus the hexameric complex potentially produces similar results, but larger scale screening will need to be performed for verification. The analyses in this chapter can be used to guide future anti-PilB drug optimization and support the use of *Ct*PilB as a model for structure-based virtual screening for inhibitory compounds.

Tables and Figures for Chapter 2

Table 2-1. Key interacting residues of ATP and ADP in the *Ct*PilB homology model. Fingerprinting summary of key and recurrent residues participating in protein-ligand interactions between ATP or ADP and the *Ct*PilB homology model prior to MD simulation. Walker A, Walker B, arginine finger, Asp-box, and His-box residues are bolded.

Interface	Ligand	H-Bond	Charged	Backbone	Sidechain	Polar	Hydrophobic
Open	ATP	R293, **R310**, **R325**, E330, **G367**, **S368**, **G369**, **T371**, **E392**, **E396**, R436, R552	R293, R305, **R310**, **R325**, E330, **K370**, **E392**, D393, **E396**, R424, **E434**, R436, R552	D328, K329, E330, **P365**, **T366**, **G367**, **S368**, **G369**, **K370**, **T371**, N372, D393, P394, **E396**, G433	R293, R305, **R310**, **R325**, L327, E330, **T366**, **K370**, **T371**, N372, L374, Y375, **E392**, D393, **E396**, F397, R424, L431, **E434**, R436, T457, R552	R293, R305, **R310**, **R325**, E330, **T366**, **K370**, **T371**, N372, **E392**, D393, **E396**, R424, **E434**, R436, T457, R552	L327, L374, Y375, F397, L431
Closed	ADP	**R325**, **G367**, **G369**, **K370**, **T371**, **H459**	**R325**, **K370**, **E396**, **E434**	L326, **P365**, **T366**, **G367**, **S368**, **G369**, **K370**, **T371**, G433, **E434**	**R325**, L327, **T366**, **K370**, **T371**, N372, **E396**, **E434**, T457, **H459**	**R325**, **T366**, **K370**, **T371**, N372, **E396**, **E434**, T457, **H459**	L327
Open'	ADP	A235, **R325**, L326, **T366**, **G367**, **S368**, **G369**, **K370**, **T371**, R552	D237, **R325**, D328, K331, **K370**, **E396**, R552	G234, A235, S236, L326, L327, D328, **P365**, **T366**, **G367**, **S368**, **G369**, **K370**, **T371**, N372, G551, R552	S236, D237, **R325**, L326, L327, D328, K331, L332, **T366**, **S368**, **K370**, **T371**, N372, **E396**, F397, R552	S236, D237, R325, D328, K331, **T366**, **S368**, **K370**, **T371**, N372, **E396**, R552	L326, L327, L332, F397

Table 2-2. Average free energy of binding of selected inhibitors.
The average free energy of binding (ΔG_{Bind}) of the six selected inhibitors docked into the three conformational states in the *Ct*PilB homology model. All energies are reported in kcal/mol. *No standard deviation available for P4MP4 in the Open' conformation.

Compound	Average ΔG_{Bind} Closed	Average ΔG_{Bind} Open	Average ΔG_{Bind} Open'	Overall Average ΔG_{Bind}
Benserazide	-16.98 ± 6.51	-22.79 ± 9.13	-28.51 ± 7.29	-22.76 ± 8.82
CNP0030078	-15.16 ± 8.76	-34.42 ± 6.47	-34.56 ± 4.49	-27.09 ± 11.71
CNP0051517	-14.86 ± 8.97	-32.57 ± 7.93	-36.19 ± 4.47	-27.2 ± 12.1
Levodopa	-21.09 ± 6.65	-26.55 ± 5.97	-17.95 ± 7.94	-21.86 ± 7.55
P4MP4	-24.85 ± 3.78	-48.04 ± 1.23	-28.41*	-34.48 ± 12.01
Quercetin	-25.59 ± 6.35	-27.97 ± 2.98	-29.44 ± 5.85	-27.67 ± 5.32

Table 2-3. Summary of interacting residues in the *Ct*PilB homology model.
Summary of residues participating in protein-ligand interactions of the best binding poses in the *Ct*PilB homology model as determined by fingerprint analysis. Walker A, Walker B, arginine finger, Asp-box, and His-box residues are bolded.

Structure	Compound	Interface	H-Bond	Charged	Backbone	Sidechain	Polar	Hydrophobic
*Ct*PilB$_{HM}$	Benserazide	Open'	K331, **T366**, **G367**, **K370**, **E396**	**R325**, **K370**, **E396**, R552	K331, **T366**, **G367**, **S368**, **G369**, **K370**	**R325**, **T366**, **K370**, **T371**, N372, **E396**, F397, R552	**R325**, **T366**, **K370**, **T371**, N372, **E396**, R552	F397
	CNP0030078	Open'	R552	**R325**, K331, R552	G234, A235, S236, L326, K331, **G367**, **G369**, G551, R552	S236, **R325**, K331, L339, L492, R552	S236, R325, K331, R552	L339, L492
	CNP0051517	Open	**R310**, **G369**	**R310**, **R325**, E330, **K370**, **E396**	**G367**, **S368**, **G369**	**R310**, **R325**, L327, E330, **T366**, **K370**, **T371**, **E396**	**R310**, **R325**, E330, **T366**, **K370**, **T371**, **E396**	L327
	Levodopa	Open	**G367**, **E434**, C436	**K370**, **E396**, **E434**, C436	**G367**, **S368**, **G369**, D393, G433	**K370**, **T371**, **E396**, **E434**, C436	**K370**, **T371**, **E396**, **E434**, C436	-
	P4MP4	Open/Open'	-	**R310**, **R325**, E330, **K370**, **E396**, (R424)	D328, K329, **G367**, D393, P394, **E396**, G433	**R310**, **R325**, L327, E330, **T366**, **K370**, **T371**, **E396**, (R424), T457	**R310**, **R325**, E330, **T366**, **K370**, **T371**, **E396**, (R424), T457	L327
	Quercetin	Open'	**E396**, R552	**R325**, K331, **E396**, R495	L326, **G369**, R552, V553	**R325**, L327, K331, L339, **T371**, **E396**, F397, L492, R495	**R325**, K331, **T371**, **E396**, R495	L327, L339, F397, L492

Table 2-4. Summary of interacting residues of ATP/ADP after MD simulation.
Summary of residues participating in protein-ligand interactions of the best binding poses of nucleotide-bound conditions that were maintained after 1 μs of MD simulation. All residues reported are present in at least two of three replicates.

Structure	Ligand	Interface	H-Bond	Charged	Backbone	Sidechain	Polar	Hydrophobic
*Ct*PilB$_{ATP}$	ATP	Open	R310, R325, R552	R310, R325, R552, R424	D328, G367	R310, R325, L327, R552, R424	R310, R325, R552, R424	L327
	ATP	Open	R293, R310	R293, R310, K370	G367	R293, R310, T366, K370, T371	R293, R310, K370, T371	-
*Ct*PilB$_{ADP}$	ADP	Closed	G367, K370	K370, E396	G367, S368, G369, K370, T371	T366, K370, T371, E396	T366, K370, T371, E396	-
	ADP	Closed	K370	R325, K370, E396	G369	T366, K370, T371, E396	R325, T366, K370, T371, E396	-
*Ct*PilB$_{ATP+ADP}$	ADP	Closed	K370, T371	K370, E396	L326, G367, S368, G369, K370, T371	T366, K370, T371, E396	T366, K370, T371, E396	-
	ATP	Open	R310, G367, T371	R293, R310, K370, E392, E434	T366, G367, T371, D393, G433	R293, R310, T366, K370, T371, E434, E392	R293, R310, T366, K370, T371, E434, E392	-
	ADP	Open’	T366, G367, S368, G369, K370, T371	R325, K370	P365, T366, G367, S368, G369, K370, T371	R325, T366, S368, K370, T371, N372	R325, T366, S368, K370, T371, N372	-
	ADP	Closed	-	R325, K370, E396	T366, G367, S368, G369, K370, T371	K370, T371, R325, E396	K370, T371, R325, E396	-
	ATP	Open	R310, R325, R552	R310, R325, R424, R552	-	R310, R325, R424, R552	R310, R325, R424, R552	-
	ADP	Open’	-	R325, R552	-	R325, R552	R325, R552	-

Table 2-5. Average free energy of binding of selected inhibitors in post-MD structures.

The overall average free energy of binding (ΔG_{Bind}) of the six selected inhibitors docked into representative *Ct*PilB structures derived from cluster analysis. Overall average ΔG_{Bind} values derived from docking with the *Ct*PilB homology model are provided in the final column for comparison. All energy is reported in kcal/mol.

Compound	ΔG_{Bind} *Ct*PilB$_{Apo}$	ΔG_{Bind} *Ct*PilB$_{ATP}$	ΔG_{Bind} *Ct*PilB$_{ADP}$	ΔG_{Bind} *Ct*PilB$_{ATP+ADP}$	ΔG_{Bind} *Ct*PilB$_{HM}$
Benserazide	-20.65 ± 9.36	-23.47 ± 8.83	-25.82 ± 8.69	-20.52 ± 10.82	-22.76 ± 8.82
CNP0030078	-25.78 ± 10.55	-28.53 ± 8.51	-32.86 ± 8.59	-25.79 ± 18.77	-27.09 ± 11.71
CNP0051517	-25.91 ± 10.59	-27.31 ± 8.83	-31.65 ± 9.59	-24.15 ± 20.3	-27.2 ± 12.1
Levodopa	-19.41 ± 10.01	-20.94 ± 7.85	-23.85 ± 9.6	-26.15 ± 8.72	-21.86 ± 7.55
P4MP4	-28.08 ± 14.97	-40.34 ± 10.61	-38.47 ± 11.37	-40.07 ± 11.92	-34.48 ± 12.01
Quercetin	-26.72 ± 7.43	-26.23 ± 6.63	-30.06 ± 5.91	-31.16 ± 7.79	-27.67 ± 5.32

Table 2-6. Residue interactions in the *Ct*PilB$_{Apo}$ structure.
Summary of residues participating in protein-ligand interactions of the best binders docked into the *Ct*PilB$_{Apo}$ representative structure. Walker A, Walker B, arginine finger, Asp-box, and His-box residues are bolded. Residue interactions occurring with adjacent packing units are indicated by parentheses.

Structure	Compound	Interface	H-Bond	Charged	Backbone	Sidechain	Polar	Hydrophobic
*Ct*PilB$_{Apo}$	Benserazide	Open	D291, **T366**, **G367**, **G369**, **T371**	D291, R293, **R310**, **K370**	G292, **T366**, **G367**, **S368**, **G369**	D291, R293, **R310**, L327, **T366**, **K370**, **T371**, N372	D291, R293, **R310**, **T366**, **K370**, **T371**, N372	L327
	CNP0030078	Open'/Closed	-	(K355), R491, R552, E611	(A354), L492	L256, (K355), (P356), (Y357), **P365**, **S368**, N461, R491, L492, R552, E611	**S368**, N461, R491, R552, E611	L256, (P356), (Y357), **P365**, L492
	CNP0051517	Open/Open'	-	**R325**, (R424), R494, R552, R610	L326, **T366**, **G367**, **S368**, **G369**, (L423)	**R325**, L327, **T366**, (L423), (R424), L492, R494, R552, R610	**R325**, **T366**, (R424), R494, R552, R610	L327, (L423), L492
	Levodopa	Open	D291, **R325**, **K370**, **T371**	D291, **R310**, **R325**, **K370**	**P365**, **G367**, **G369**, **K370**, **T371**	D291, **R310**, **R325**, L327, **T366**, **S368**, **K370**, **T371**	D291, **R310**, **R325**, **T366**, **S368**, **K370**, **T371**	L327
	P4MP4	Open/Open'	P289	D291, **R310**, (R420), (R424), D462, R491, E611	P289, Q290, D291, (A417), (R420), L492	Q290, D291, **R310**, **P365**, (A417), (R420), (L423), (R424), N461, D462, R491, E611, V613	Q290, D291, **R310**, (R420), (R424), N461, D462, R491, E611	**P365**, (A417), (L423), V613
	Quercetin	Open/Open'	D291, L492	D291, **R310**, R491, E611	**P365**, **G367**, **S368**, R491, L492, R610, E611	D291, **R310**, **P365**, **S368**, N461, (L423), R491, E611	D291, **R310**, **S368**, N461, R491, E611	**P365**, (L423)

Table 2-7. Residue interactions in the *Ct*PilB$_{ATP}$ structure.
Summary of residues participating in protein-ligand interactions of the best binders docked into the *Ct*PilB$_{ATP}$ representative structure. Walker A, Walker B, arginine finger, Asp-box, and His-box residues are bolded. Residue interactions occurring with adjacent packing units are indicated by parentheses.

Structure	Compound	Interface	H-Bond	Charged	Backbone	Sidechain	Polar	Hydrophobic
*Ct*PilB$_{ATP}$	Benserazide	Open'	**T371**, P394, **E396**	**R325**, **E392**, **E396**	D393, P394, **E396**, **E434**	**R325**, L332, **T366**, **T371**, N372, Y375, **E392**, P394, **E396**	**R325**, **T366**, **T371**, N372, **E392**, **E396**	L332, Y375, P394
	CNP0030078	Open/Open'	**R310**	D291, R293, **R310**, (R424), R552	**T366**, **G367**, **S368**, L492	D291, R293, **R310**, L327, **P365**, **T366**, **T371**, (R424), L492, R552	D291, R293, **R310**, **T366**, **T371**, (R424) R552	L327, **P365**, L492
	CNP0051517	Closed	R552	E330, R495, K550, R552	G234, D254, G255, K338, L339, R552, V553, G554	S236, V253, E330, L492, R495, K550, R552	S236, E330, R495, K550, R552	V253, L492
	Levodopa	Closed	S236, **T371**, N372	D237, R252, **K370**, R552	**G367**, **G369**, **K370**, **T371**, **N372**	S236, D237, R252, **T366**, **K370**, **T371**, N372, R552	S236, D237, R252, **T366**, **K370**, **T371**, N372, R552	-
	P4MP4	Closed	S236, D254	D328, E330, K550, R552	G234, V253, D254, K331, **G367**, **G369**, **K370**, **T371**, N372, T373, R552	S236, V253, D328, E330, L339, **T371**, N372, T373, K550, R552	S236, D328, E330, **T371**, N372, T373, K550, R552	V253, L339
	Quercetin	Closed	P394, **E434**	**K370**, **E392**, **E396**, **E434**	**T366**, **G367**, **S368**, **G369**, P394, V395, **E396**	**T366**, **K370**, **T371**, N372, Y375, **E392**, **E396**, F397, **E434**, **H459**	**T366**, **K370**, **T371**, N372, **E392**, **E396**, **E434**, **H459**	Y375, F397

Table 2-8. Residue interactions in the *Ct*PilB$_{ADP}$ structure.
Summary of residues participating in protein-ligand interactions of the best binders docked into the *Ct*PilB$_{ADP}$ representative structure. Walker A, Walker B, arginine finger, Asp-box, and His-box residues are bolded. Residue interactions occurring with adjacent packing units are indicated by parentheses.

Structure	Compound	Interface	H-Bond	Charged	Backbone	Sidechain	Polar	Hydrophobic
*Ct*PilB$_{ADP}$	Benserazide	Open	**T366**, **S368**, R491, L492, E611	**R310**, R491, R552, E611	**T366**, **G367**, **S368**, R491, L492	**R310**, L327, **P365**, **S368**, R491, L492, R552, E611	**R310**, **S368**, R491, R552, E611	L327, **P365**, L492
	CNP0030078	Open'	R552	**R325**, K331, R552	L231, G234, S236, K331, L332, G551, R552	L231, S236, **R325**, L326, K331, M333, N372, L492, R552	S236, R325, K331, N372, R552	L231, L326, M333, L492
	CNP0051517	Open	D328	D328, E330, R552	L327, D328, K329, E330, **T366**, **G367**, L492	L327, D328, E330, **P365**, **S368**, L492, R552	D328, E330, **S368**, R552	L327, **P365**, L492
	Levodopa	Open	D291, **T366**, **S368**, R491, E611	D291, **R310**, R491, R552, E611	**P365**, **T366**, R491, L492	D291, **R310**, **P365**, **T366**, **S368**, R491, L492, R552, E611	D291, **R310**, **T366**, **S368**, R491, R552, E611	**P365**, L492
	P4MP4	Open/Open'	L492, R610	**R325**, (R424), R552, R610	**R310**, V311, S312, V323, (R424), **H459**, N461, L492	V323, **R325**, **P365**, (R424), **H459**, N461, L492, R552, R610	**R325**, (R424), **H459**, N461, R552, R610	V323, **P365**, L492
	Quercetin	Open'/Closed	S236, D254, N372, (T450)	R252, D254, **R325**, **E396**, R552	G234, S236, R252, V253, D254, G255, (T450), (G451)	S236, R252, D254, **R325**, **T371**, N372, **E396**, R552	S236, R252, D254, **R325**, **T371**, N372, **E396**, R552	-

Table 2-9. Residue interactions in the *Ct*PilB$_{ATP+ADP}$ structure.
Summary of residues participating in protein-ligand interactions of the best binders docked into the *Ct*PilB$_{ATP+ADP}$ representative structure. Walker A, Walker B, arginine finger, Asp-box, and His-box residues are bolded. Residue interactions occurring with adjacent packing units are indicated by parentheses.

Structure	Compound	Interface	H-Bond	Charged	Backbone	Sidechain	Polar	Hydrophobic
*Ct*PilB$_{ATP+ADP}$	Benserazide	Open'/Closed	R252, **G369**, T371	R252, **R325**, **K370**, (D426), R552	G234, A235, S236, R252, D254, G255, **G367**, **G369**, **K370**, **T371**	S236, R252, **R325**, M333, **T366**, **K370**, **T371**, (D426), R552	S236, R252, **R325**, **T366**, **K370**, **T371**, (D426), R552	M333
	CNP0030078	Open'/Closed	N428	D237, R252, **R325**, (D426), R552, E611	D254, G255, **T366**, **G367**, R491, L492	S236, D237, R252, **R325**, **P365**, **S368**, (D426), (N428), L492, R552, E611	S236, D237, R252, **R325**, **S368**, (D426), (N428), R552, E611	**P365**, L492
	CNP0051517	Open	**T371**	R293, R305, D308, D328, K329, E330, **E396**	E330, **G367**	R293, R305, D308, D328, K329, E330, L332, **T371**, Y375, **E396**, L492	R293, R305, D308, D328, K329, E330, **T371**, **E396**	L332, Y375, L492
	Levodopa	Open'	S236, **R325**, **T366**, **T371**	**R325**, **K370**, R552	K331, **T366**, **G367**, **G369**, **K370**, **T371**	S236, **R325**, M333, **T366**, **K370**, **T371**, N372, L492, R552	S236, **R325**, **T366**, **K370**, **T371**, N372, R552	M333, L492
	P4MP4	Open/Open'	**G367**, **T371**, R552	K329, R424, R552	E330, **G367**	K329, L332, **P365**, **T371**, L492, R552, L614	K329, **T371**, (R424), R552	L332, **P365**, L492, L614
	Quercetin	Open'/Closed	G234, (D426)	**R325**, (D426), R552	G234, R252, D254, G255, **G367**, **G369**, **K370**, **T371**, N372	S236, **R325**, M333, **T366**, **T371**, N372, R552	S236, **R325**, **T366**, **T371**, N372, R552	M333

Table 2-10. Summary of interacting residues of PilB inhibitors in post-MD representative structures.

Summary of residues participating in protein-ligand interactions of the best binding poses of PilB inhibitors across all four representative *Ct*PilB structures selected from clustering analysis. All residues reported are present in at least two of four representative structures.

Compound	H-Bond	Charged	Backbone	Sidechain	Polar	Hydrophobic
Benserazide	T366, G369, T371	R310, R325, K370	T366, G367, S368, G369	R310, R325, T366, K370, T371	R310, R325, T366, K370, T371	-
CNP0030078	-	R325	T366, G367	R325, P365, S368	S368	P365
CNP0051517	-	-	T366, G367	-	-	-
Levodopa	R325, T366, T371	R310, R325, K370	P365, T366, G367, G369, K370, T371	R310, R325, T366, S368, K370, T371	R310, R325, T366, S368, K370, T371	-
P4MP4	-	-	G367	P365, T371	T371	P365
Quercetin	-	R325, E396	G367, S368, G369	R325, T366, T371, E396	R325, T366, T371, E396	-

Table 2-11. Average RMSD of inhibitor best poses.
The root-mean-square deviation (RMSD) of the best binding poses of selected inhibitors in the post-MD representative structures. Poses are aligned to the corresponding best pose in the *Ct*PilB homology model. RMSD is reported in angstroms (Å).

Compound	RMSD $Ct\text{PilB}_{Apo}$	RMSD $Ct\text{PilB}_{ATP}$	RMSD $Ct\text{PilB}_{ADP}$	RMSD $Ct\text{PilB}_{ATP+ADP}$	Average RMSD (Å)
Benserazide	2.73	1.878	1.57	1.18	1.84 ± 0.57
CNP0030078	3.178	0.38	1.143	3.21	1.98 ± 1.25
CNP0051517	0.465	2.454	2.252	3.222	2.10 ± 1.01
Levodopa	0.083	0.862	1.965	0.661	0.89 ± 0.68
P4MP4	3.314	3.675	3.598	4.009	3.65 ± 0.25
Quercetin	0.179	0.159	0.115	0.317	0.19 ± 0.08

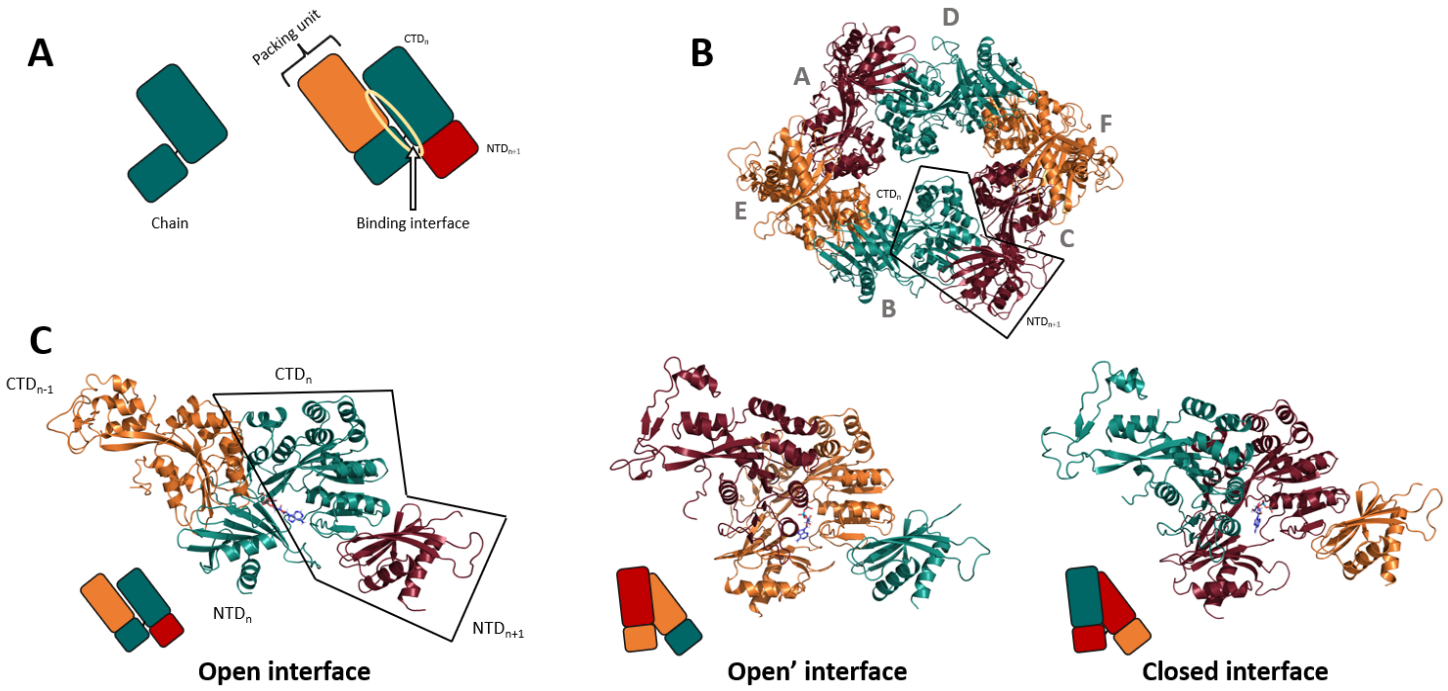

Figure 2-1. Schematic of PilB functional structure.
(A) Schematic depicting the functional units of the PilB ATPase. The CTD of one chain and the NTD of the adjacent chain comprise a functional packing unit of PilB. The ATP binding interface is located between two packing units. (B) The different chain designations in the *Ct*PilB homology model are indicated in gray. Chains conformations are designated by color: green = open, orange = open', and red = closed. A single packing unit is outlined in black. (C) Images of two packing units creating the three binding states. An open interface is composed of two parallel packing units, while packing units in the open' and closed interfaces are partially overlapping. Annotation style adapted from McCallum et al., 2017.

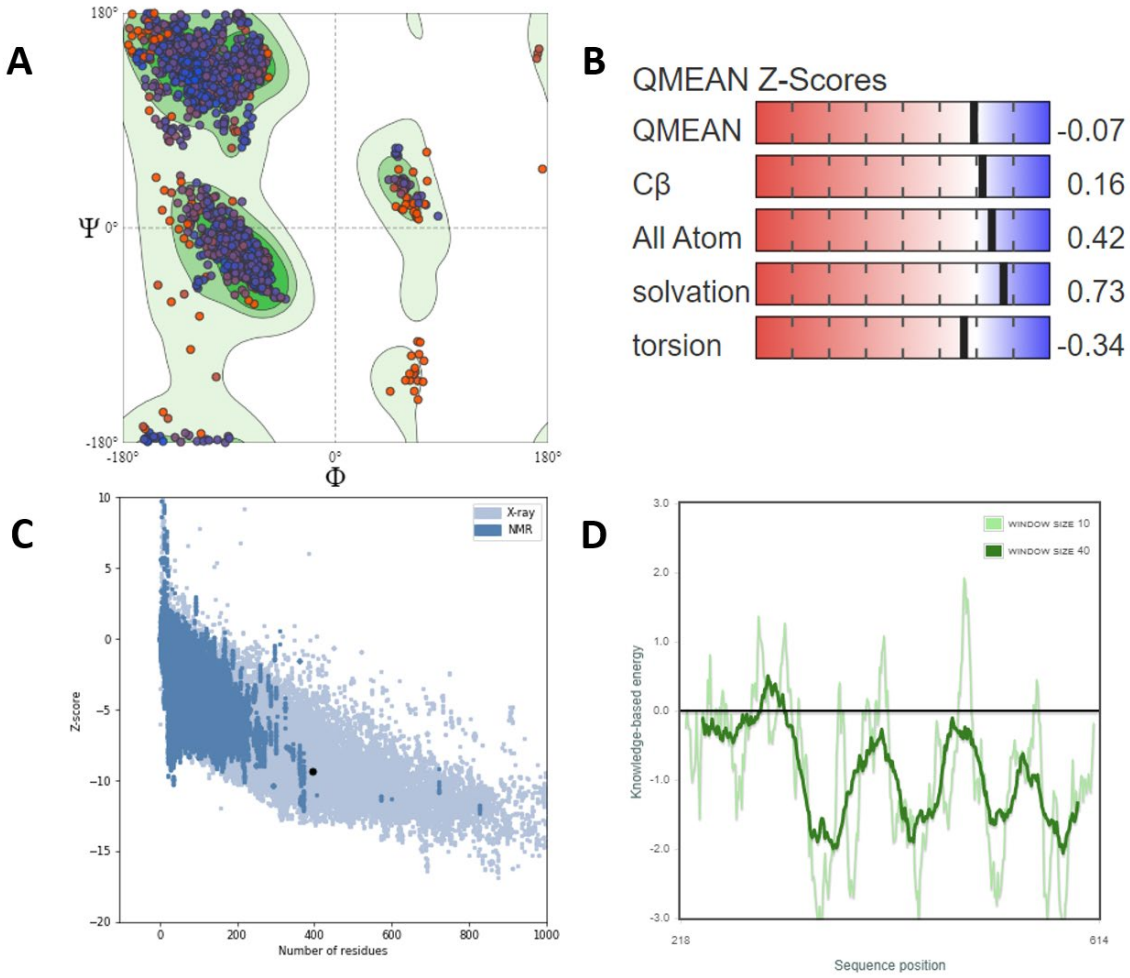

Figure 2-2. Validation of the *Ct*PilB homology model.
Structure validation of the energy minimized homology model of *Ct*PilB. (A) Ramachandran plot showed 96.84% of residues are favorable, with 3.43% allowed and 0.59% outliers. (B) QMEAN analysis indicates values are near 0. QMEANDisCo is 0.77 ± 0.05 according to the SwissModel tool suite. (C) ProSA produces a Z-score of -9.36, indicating an acceptable structure when compared to resolved structures of similar size (black dot). (D) Local model quality is acceptable across the model except within a flexible loop region (dark green line). Residues below the black line indicate favorable residues.

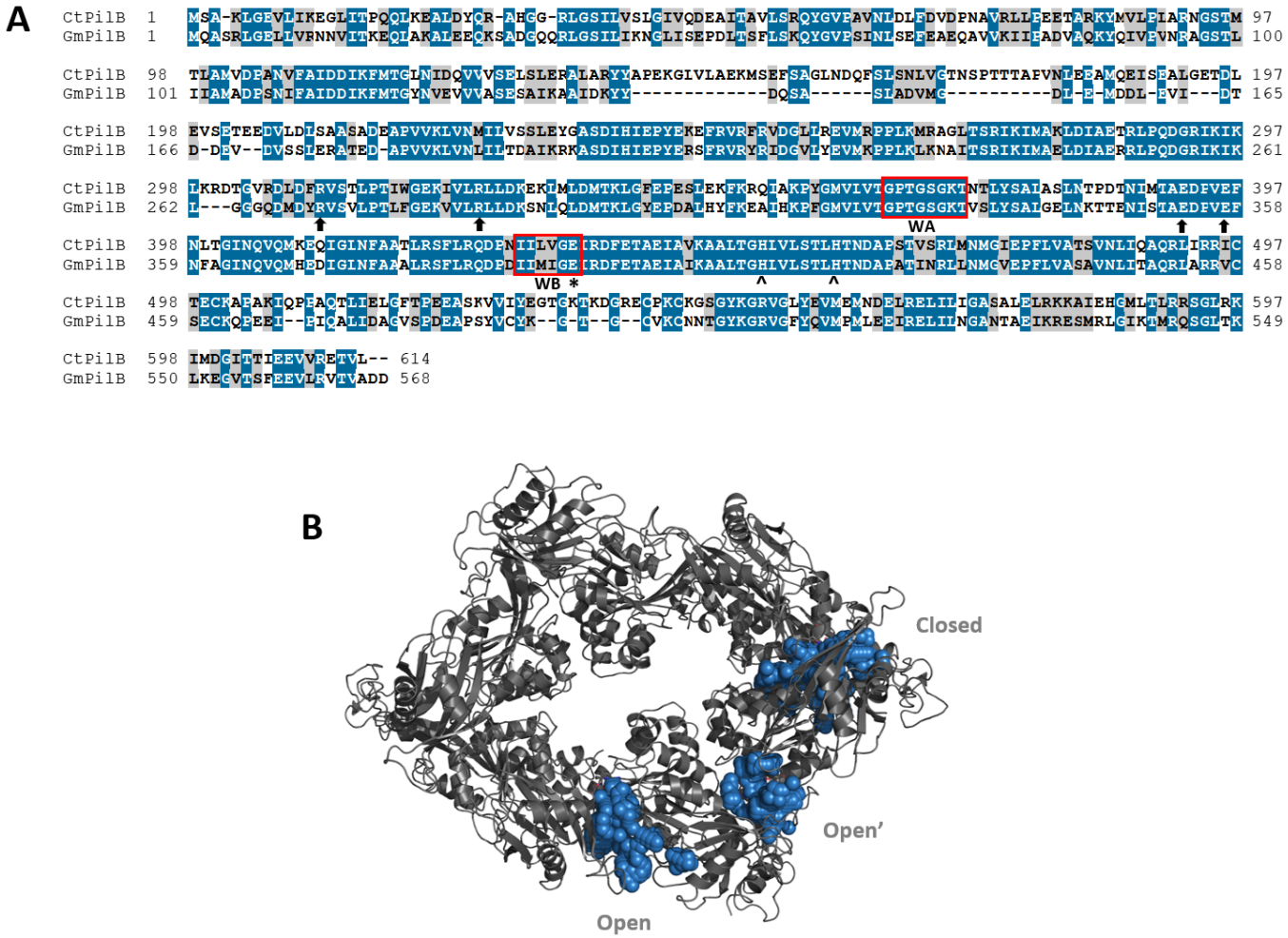

Figure 2-3. ATP binding motifs are conserved in PilB.
(A) Alignment of the amino acid sequences of PilB from *C. thermophilum* and the template PilB from *Geobacter metallireducens*. Walker A (WA) and Walker B (WB) residues are boxed in red. The catalytic glutamate in the WB motif is indicated by an asterisk (*). Arginine fingers and Asp-box motifs are indicated by black arrows. His-box histidines are indicated by carets (^). (B) The most probable ligand binding domains are at the ATP binding cavity, as determined by SiteMap (blue spheres).

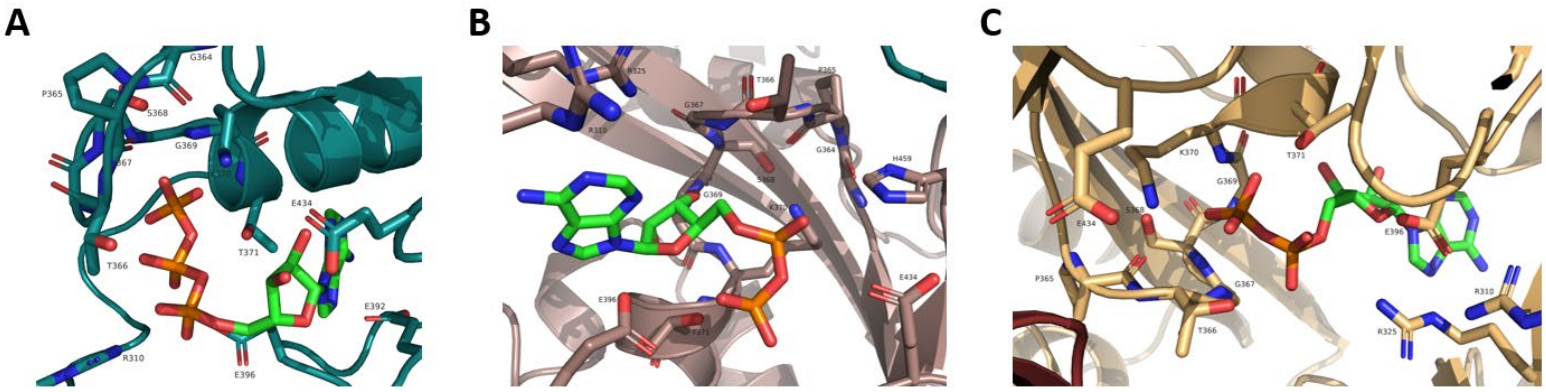

Figure 2-4. Representative images of ATP and ADP docked into *Ct*PilB.
Representative images of ATP or ADP docking to the ATP binding pocket of the *Ct*PilB homology model. Key residues are visualized as sticks and labeled by residue. (A) ATP docked into the open binding interface. (B) ADP docked into the closed interface. (C) ADP docked into the open' interface.

Compound	Molecular Formula	Molecular Weight (g/mol)	Reference
Benserazide	$C_{10}H_{15}N_3O_5$	257.24	(Dye et al., 2022)
CNP0030078	$C_{25}H_{20}F_4N_2O_2$	456.43	(Ozcan et al., 2023)
CNP0051517	$C_{25}H_{21}F_3N_2O_2$	438.44	(Ozcan et al., 2023)
Levodopa	$C_9H_{11}NO_4$	197.19	(Dye et al., 2022)
P4MP4	$C_{36}H_{40}N_2O_4$	564.73	(Aubey et al., 2019)
Quercetin	$C_{15}H_{10}O_7$	303.23	(Dye et al., 2022)

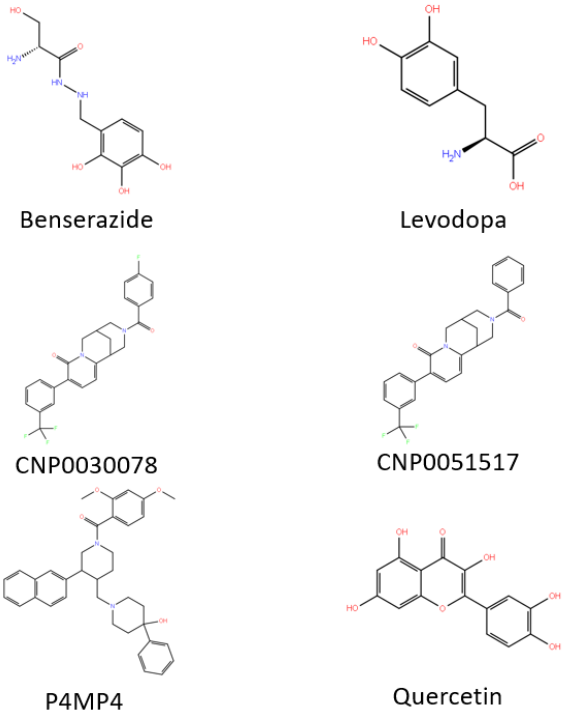

Figure 2-5. Characteristics of selected inhibitors.
The molecular formula and weight of the selected inhibitors identified in recent literature. The molecular structure of benserazide, levodopa, CNP0030078, CNP0051517, P4MP4, and quercetin are provided to the right.

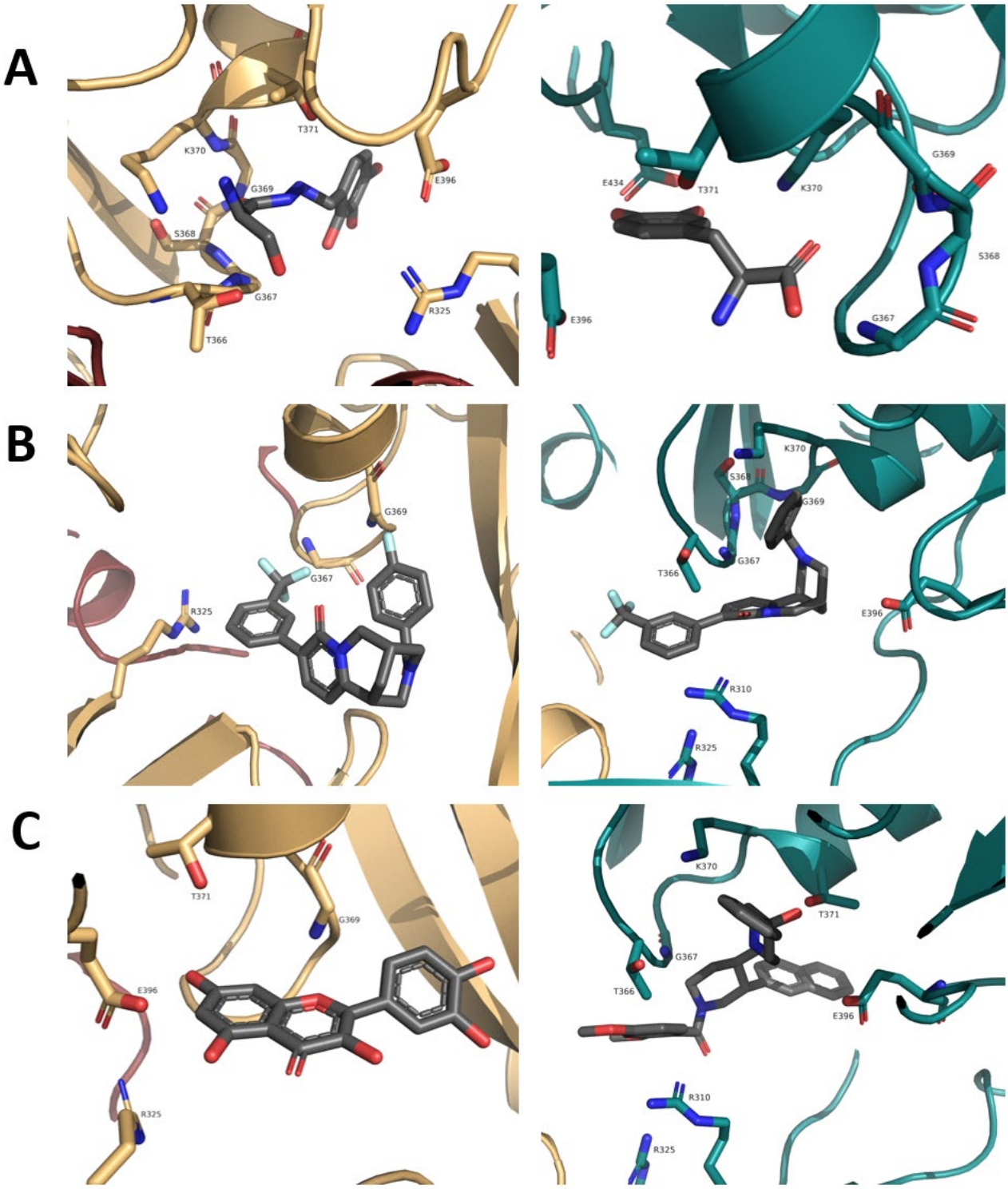

Figure 2-6. Best poses of selected inhibitors in the *Ct*PilB homology model.
The best binding poses of selected inhibitors following docking into the *Ct*PilB homology model. Best binding poses primarily occurred in the open' (yellow) or closed (teal) interfaces. Key residues are labeled by residue and represented by sticks. (A) Benserazide (left) and levodopa (right). (B) CNP0030078 (left) and CNP0051517 (right). (C) Quercetin (left) and P4MP4 (right).

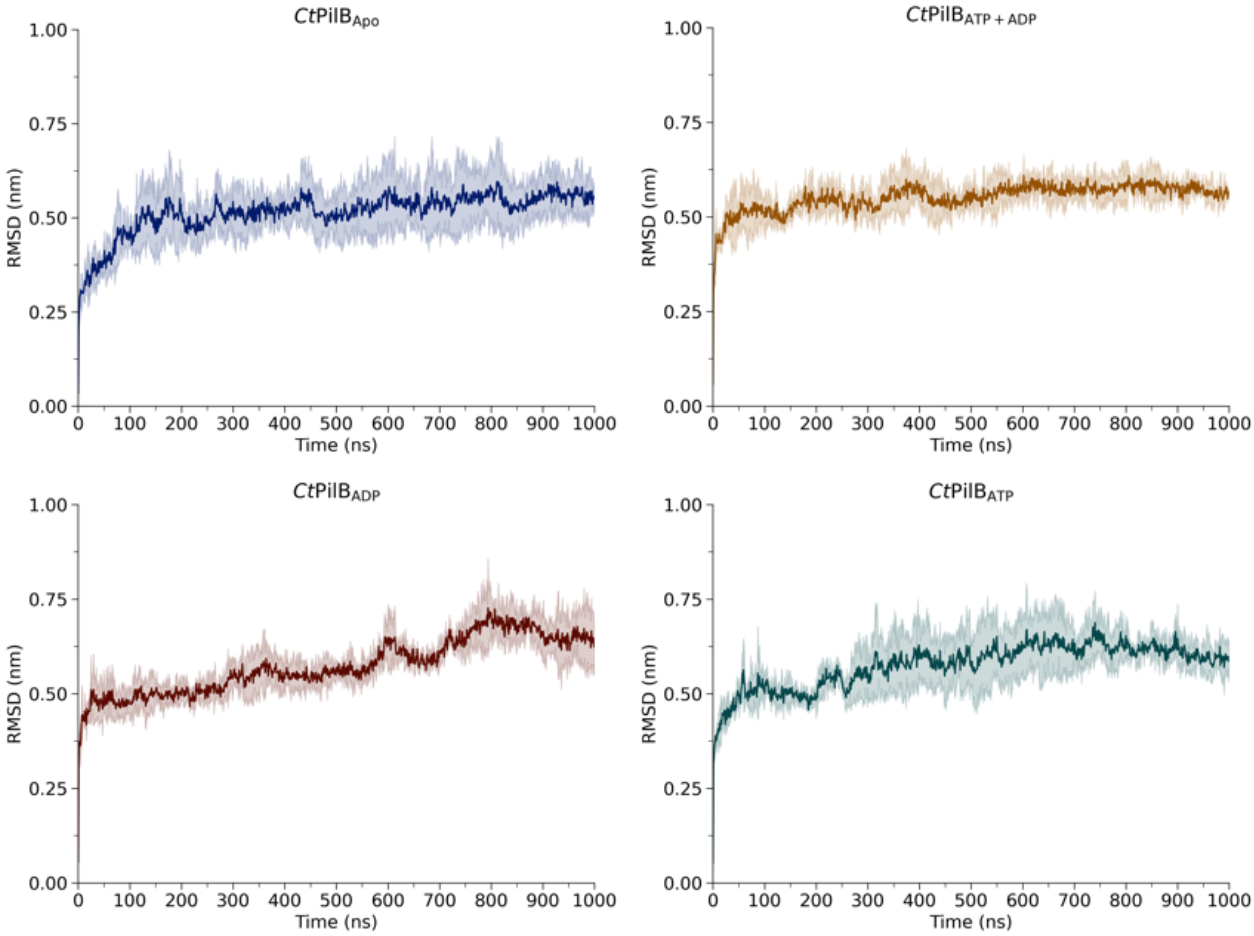

Figure 2-7. Average RMSD of *Ct*PilB in MD simulation.
Average root-mean-square deviation (RMSD) of the protein backbone for the *Ct*PilB$_{Apo}$, *Ct*PilB$_{ATP+ADP}$, *C*tPilB$_{ADP}$, and *Ct*PilB$_{ATP}$ systems after 1 μs of simulation time (CHARMM36m), with three replicates for each system. Shaded area represents the standard deviation of the three replicates.

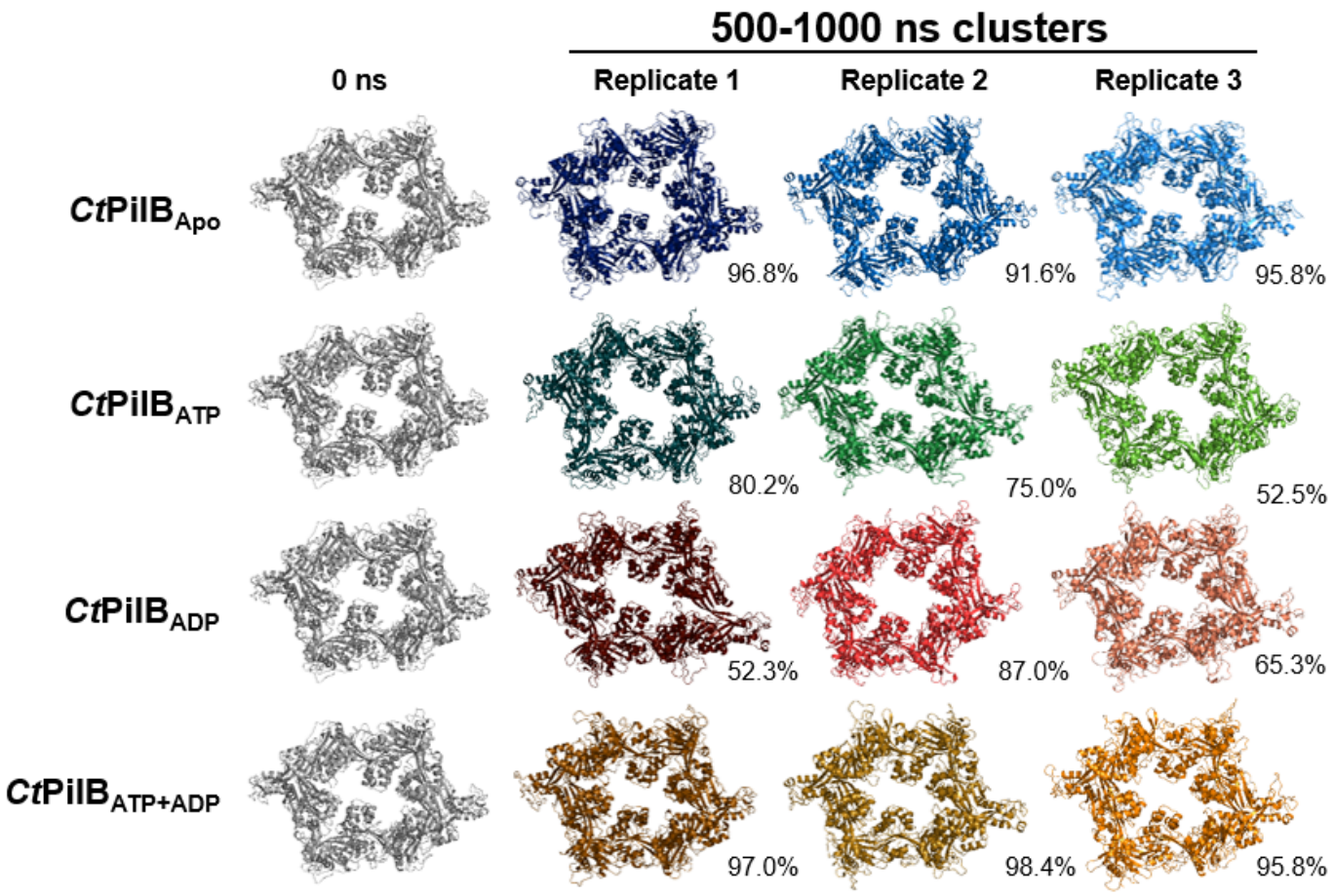

Figure 2-8. Predominant structures obtained by clustering analysis.
Cartoon representation of each predominant structure of all clusters prior to and after MD simulation. Clustering was performed over the final 500 ns of each simulation using the gromos algorithm with a 0.275 nm RMSD cut-off. '0 ns' (gray) represents the initial *Ct*PilB structure prior to simulation. Percentages to the right of each structure indicate the portion of the sampling period that structure occupies. Structure colors are assigned to system types: *Ct*PilB$_{Apo}$ = blue, *Ct*PilB$_{ATP}$ = green, *C*tPilB$_{ADP}$ = red, *Ct*PilB$_{ATP+ADP}$ = orange.

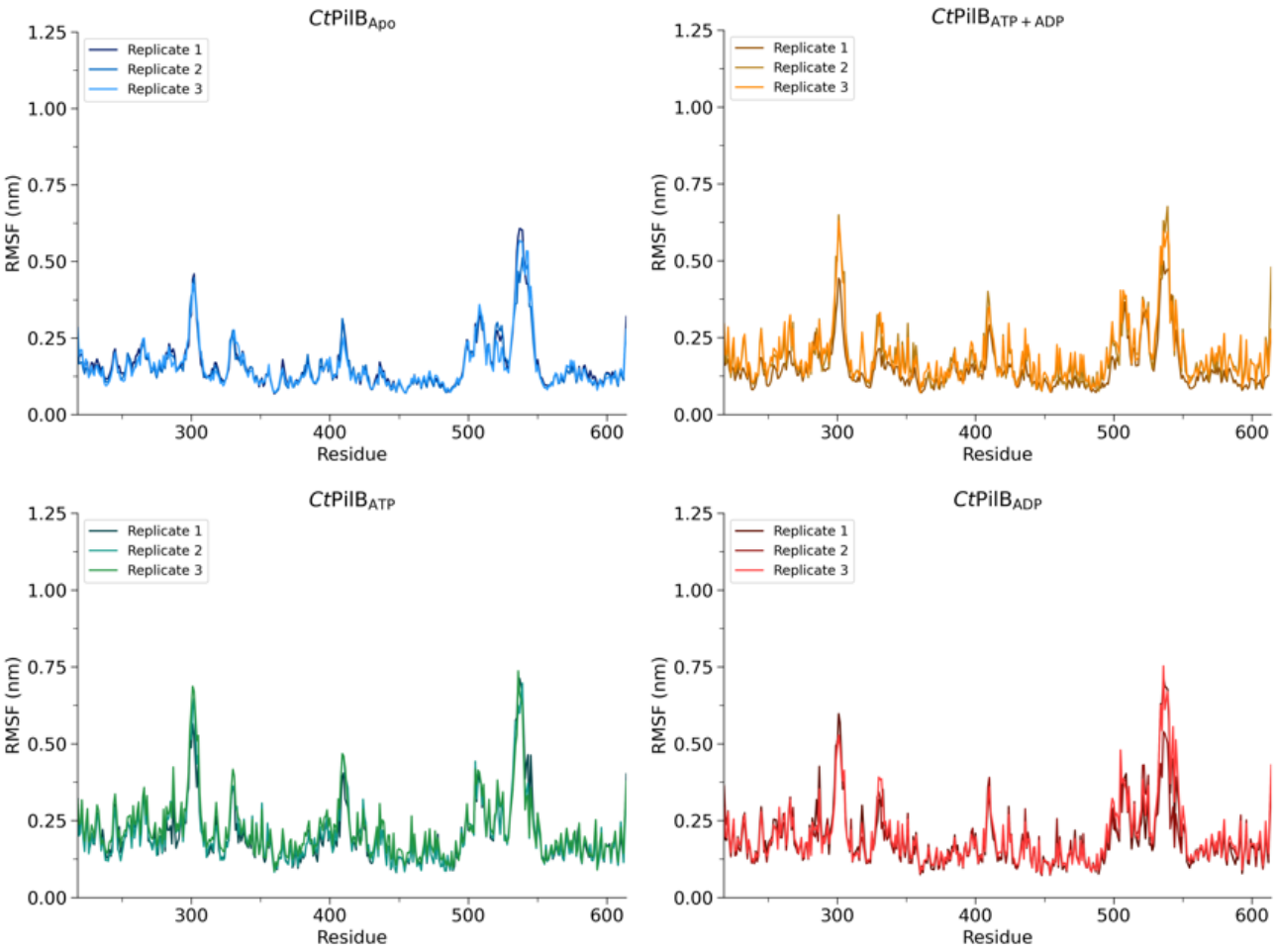

Figure 2-9. Average RMSF of *Ct*PilB in MD simulation.
Average root-mean-square fluctuation (RMSF) of the protein backbone for the *Ct*PilB$_{Apo}$, *Ct*PilB$_{ATP+ADP}$, *C*tPilB$_{ADP}$, and *Ct*PilB$_{ATP}$ systems after 1 μs of simulation time (CHARMM36m), with three replicates for each system. Each replicate is represented by a differently shaded line based on system conditions: *Ct*PilB$_{Apo}$ = blue, *Ct*PilB$_{ATP}$ = green, *C*tPilB$_{ADP}$ = red, *Ct*PilB$_{ATP+ADP}$ = orange.

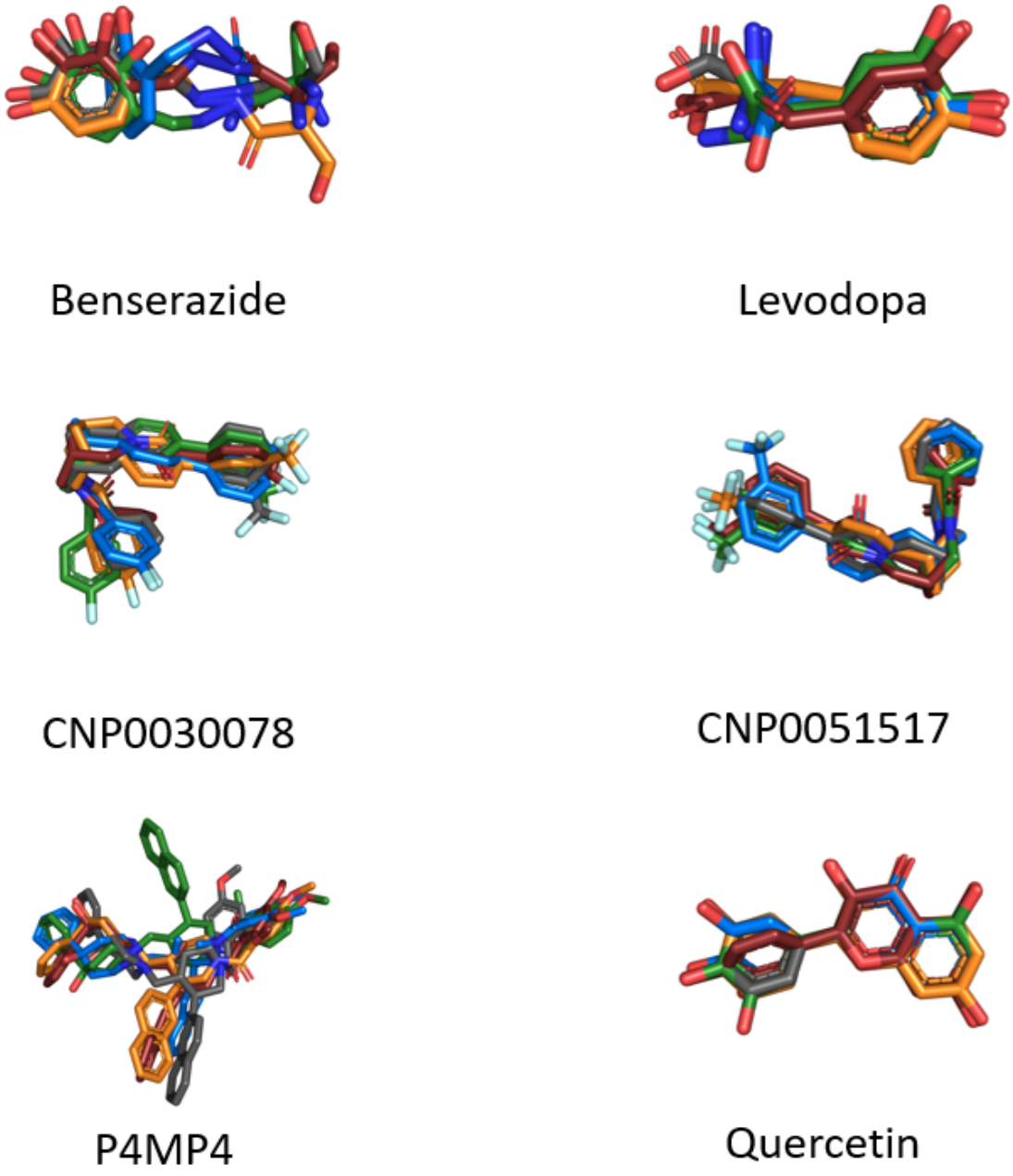

Figure 2-10. Overlay of best inhibitor poses.
Overlay of the best compound poses obtained from molecular docking. The poses that generated the most favorable ΔG_{Bind} values for a compound were extracted and aligned with the corresponding best binding pose in the *Ct*PilB homology model. The carbon backbone of each pose is shaded based on system conditions: *Ct*PilB$_{HM}$ = gray, *Ct*PilB$_{Apo}$ = blue, *Ct*PilB$_{ATP}$ = green, *C*tPilB$_{ADP}$ = red, *Ct*PilB$_{ATP+ADP}$ = orange.

Chapter 3 - Characterization of *Neisseria gonorrhoeae* and *Clostridioides difficile* T4P assembly ATPases for antivirulence drug discovery

Abstract

The bacterial type IV pilus (T4P) is a cell surface appendage with a variety of functions in cell motility, biofilm formation, horizontal gene transfer, and pathogenesis. Several noteworthy human pathogens including *Neisseria gonorrhoeae* and *Clostridioides difficile* utilize T4P as a virulence factor to colonize their hosts and establish disease. Critical to T4P biogenesis is the PilF/B ATPase that drives the assembly of the pilus filament. Disruption of this assembly motor ATPase has been shown to abrogate T4P production in multiple organisms, suggesting PilF/B may be exploited for antivirulence drug discovery and development. Here, we report the heterologous expression and purification of an N-terminally truncated *N. gonorrhoeae* PilF by heterologous expression and chromatography. Conditions for its ATPase activity were optimized through malachite green-based endpoint assays. We show here that the successful purification of this truncated PilF (*Ng*PilF$_{166}$) as an active ATPase and our efforts to formulate optimized buffer and reaction conditions for its enzymatic activity *in vitro*. Additionally, this chapter documents the various attempts to purify *C. difficile* PilB in an active form. Seven *Cd*PilB expression plasmid constructs were created, and they were used for their expression and purification by induction and auto-induction methods. Of the seven created, only two produced soluble protein in the conditions tested. One construct remains untested but may potentially be useful for future experimentation. Full-length *Cd*PilB eluted in dual peaks in FPLC and only the peak fractions eluted early in size exclusion had detectable ATPase activity While the overall attempts to produce a

greater yield of active *Cd*PilB were unsuccessful, these efforts may prove informative for future designs of constructs and purification schemes.

Introduction

Accelerated by antibiotic misuse, antibiotic resistance (ABR) has become increasingly commonplace among bacterial pathogens, threatening human health on a global scale. Infections related to ABR are estimated to be responsible for approximately 4.9 million deaths globally in 2019, and 1.27 million among these were directly attributed to ABR pathogens. In the United States alone, ABR infections are believed to cause 2.8 million infections, resulting in 35,000 deaths annually [1, 5]. It is projected that ABR-related disease could become responsible for up to 10 million deaths by 2050 if no new therapeutics are developed to replace our near-depleted repertoire of antibiotics [7]. 90% of highly resistant pathogens identified by the Centers for Disease Control and Prevention (CDC) as antimicrobial resistant are bacterial in origin. Taken together with the considerable overlap with the ABR priority pathogens indicated by the World Health Organization (WHO), it is clear there is a critical need for development of novel antibiotics [1, 6].

Ultimately, the emergence of ABR is tied to the core functions of antibiotics. That is, antibiotics target cellular processes essential for sustaining life, such as DNA, RNA, and protein synthesis as well as cell wall biosynthesis and maintenance [8]. Though this targeting of vital functions was once touted as the main strength of antibiotics, it is also paradoxically its greatest weakness. Because antibiotics target vital functions, they are intrinsically broad-spectrum and can thus eliminate both commensal and pathogenic species alike. This exerts a strong evolutionary pressure to select for organisms with

resistance mechanisms that are able to overcome antibiotic effects [10]. Major mechanisms of resistance include target modification, increased antibiotic efflux, decreased antibiotic uptake, and antibiotic inactivation. These mechanisms can rapidly spread within bacterial communities through horizontal gene transfer; thus, it is likely resistance will quickly emerge soon after new antibiotics are introduced [8, 14].

One alternative to antibiotics is to target bacterial virulence rather than critical life processes. The antivirulence approach aims to disrupt the production or function of components or products known also as virulence factors that are tied to the capability of bacterial pathogens to cause disease. Common bacterial virulence factors include adhesins, toxins and host interfering factors, as well as protein secretion and virulence regulatory systems [27, 28]. Virulence factors are typically most relevant in the scope of pathogenesis, and they are non-essential for growth and survival of the pathogen aside from the infection process. These antivirulence therapeutics are expected to exert little, if any, selective pressures on the growth and survival of pathogens or non-pathogens because they are not bactericidal or bacteriostatic *per se* .[176] Recent research is continuing to explore the realm of antivirulence therapeutics as an alternative to overcome the ABR pandemic.

Neisseria gonorrhoeae and *Clostridioides difficile* are two bacterial species noted by the CDC as urgent threats due to the ABR pandemic [1]. *N. gonorrhoeae* is a Gram-negative obligate human pathogen responsible for the sexually transmitted infection (STI) gonorrhea. ABR strains of *N. gonorrhoeae* have been found to have all major mechanisms of antibiotic resistance and are unable to be treated by most major classes of antibiotics with few exceptions [102, 104]. *N. gonorrhoeae* generally colonizes the

human genital mucosa, where infection can result in cervicitis and pelvic inflammatory disease, adverse fetal defects in pregnant women, and an increased risk of HIV [101, 104]. Gonococcal infection is largely asymptomatic in women, serving as hidden reservoirs for potentially ABR strains of *N. gonorrhoeae* that can be transmitted silently amongst a population. Recently, the recommended treatment has been changed to a single dose of ceftriaxone due to the emergence of azithromycin-resistant strains[101, 106, 107].

C. difficile is Gram-positive, spore-forming anaerobe responsible for gastrointestinal disease, most notably pseudomembranous colitis. The conditions that allow *C. difficile* infections to arise mirror that of opportunistic ABR pathogens after antibiotic use. Its pathogenicity is closely tied to its ability to thrive in a dysbiotic microbiome, with commensal populations keeping *C. difficile* levels in check by converting primary bile acids into secondary bile acids that inhibit *C. difficile* vegetative growth. Antibiotic use weakens normal microbiota and promotes *C. difficile* proliferation by increasing the availability of energy sources, such as succinate and sialic acid, and primary bile acids that promote spore germination [120]. Sporulation allows *C. difficile* to survive oxic conditions, heat stress, and exposure to disinfectants for long periods of time, allowing transmission in healthcare and community settings [30, 120]. Although it is not presently considered antibiotic resistant, there have been increasing reports of ABR *C. difficile* strains worldwide. Currently, the antibiotics vancomycin, fidaxomicin, and metronidazole are recommended to treat *C. difficile* infections [1, 30, 121, 122].

Common to these two pathogens is their use of the bacterial type IV pilus (T4P) as a virulence factor for bacterial adhesin. The T4P is an extracellular filamentous appendage associated with surface motility, surface attachment, and DNA transformation [43].

Stable surface attachment can eventually result in biofilm formation, manifesting pathologically as chronic disease [177]. There are two subtypes of T4P, with the type 4a subtype (T4aP) most commonly identified with pathogenesis as well as bacterial motility [40]. A supramolecular complex of at least 10 highly conserved proteins, collectively called the T4P machine, is required for T4P biogenesis. This T4P machine anchors the pilus to the cell body and provides a channel through which the pilus extends and retracts. The pilus filament is composed of thousands of polymerized pilin subunits, whose assembly is driven by ATP hydrolysis by a cytoplasmic ATPase known as PilB in *C. difficile* and PilF in *N. gonorrhoeae*. This assembly ATPase allows pilin to be incorporated into the growing pilus at its base, extruding over a few μm from the cell surface. A retraction ATPase is present in T4aP system, driving the pilus depolymerization [43, 70, 75, 81]. Repeated cycles of ATPase-mediated extension and retraction can pull cells forward as a form of motility in key Gram-negative pathogens like *N. gonorrhoeae, Acinetobacter baumannii* and *Pseudomonas aeruginosa*. T4P are also associated with DNA transformation, where extension and retraction may bring DNA to the cell body [50, 64]. Specifically in *N. gonorrhoeae*, the T4P is critical in the early establishment of disease by allowing *N. gonorrhoeae* to form microcolonies at the epithelial cell surface. Initiation of infection may also require interactions between the T4P of the bacterium and the human epithelial surface receptor CR3 [102, 114, 115]. The role of T4P in *C. difficile* is less clear; however, it has been demonstrated that T4P may contribute to the early stages of biofilm formation and long-term persistence at the intestinal epithelium [57, 119].

Since the main extracellular appendage of the T4P is the pilus, it follows that most T4P-associated functions are mediated by the pilus itself. Pilin, and by extension the pilus filament, are not ideal antivirulence targets due to antigenic variation in some species, including *Neisseria* spp. [102]. Notably, mutation or loss of the assembly ATPase PilB/F results in the absence of pili at the cell surface. These mutants often exhibit attenuated virulence [57, 86, 95, 100]. This supports that the assembly ATPase is critical for T4P assembly and its functionality. Previous work in this laboratory has shown that the *Chloracidobacterium thermophilum* PilB can be used in high-throughput screens for inhibitory compounds [96, 97]. The inhibitors found in these studies were shown to additionally be effective on PilB from environmental species *Myxococcus xanthus* and opportunistic nosocomial pathogen *Acinetobacter nosocomialis*. Other work has also reported another compound that impacts T4P-associated cell aggregation and reduces T4P levels in *Neisseria meningitidis*. This compound, P4MP4, is believed to target the assembly ATPase [95]. These studies lend credence to the assembly ATPase as a target for antivirulence therapeutics.

With the aim of establishing *N. gonorrhoeae* PilF as a model for anti-T4P antivirulence drug discovery, this chapter reports on the expression and purification of the assembly ATPase PilF from *N. gonorrhoeae* in a N-terminally truncated form (NgPilF$_{166}$). Malachite green-based endpoint ATPase assays were used to assess and optimize the conditions for NgPilF$_{166}$ ATPase activity *in vitro*. Here, we show successful construction and purification of NgPilF$_{166}$ and provide a basic kinetic profile of its ATPase activity. Additionally, this chapter reports an optimized buffer composition and reaction conditions for NgPilF$_{166}$ ATPase activity at a physiologically relevant

temperature. This chapter also discusses the approaches taken to purify an active PilB from *C. difficile* (*Cd*PilB). Seven expression constructs for *Cd*PilB were created and used for induction, auto-induction and purification. Two of these constructs were found to be more amenable to purification, although most of the purified proteins were inactive as ATPases.

Experimental Procedures

Growth Conditions and Bacterial Strains

The *Escherichia coli* strain XL1-Blue was used for plasmid construction and maintenance and BL21(DE3) was used for protein expression. All *E. coli* strains were grown in Luria-Bertani broth (LB) at 37°C, unless otherwise stated. 100 µg/mL of ampicillin was included in *E. coli* XL1-Blue cultures and 100 µg/mL of ampicillin and chloramphenicol in *E. coli* BL21(DE3) cultures for the maintenance of plasmids, when necessary.

Plasmid Construction of *Ng*PilF and *Cd*PilB

A complete list of primers used for the work in this chapter can be found in Table 3-1, with primers for pilot constructs included here. The coding region of *Ng*PilF$_{166}$ was amplified from a *Neisseria gonorrhoeae* FA 1090 chromosomal DNA template by PCR using Phusion DNA polymerase (New England Biolabs) with the primers *Ng*PilF_F (TAGCATATGGAGGCAGAAGACGGCCCTGTTCC) and *Ng*PilF_R (CCGAAGCTTATCGTTGGTATTTGCCGTTACCTCTTCCAATGAAG). The DNA fragment was then digested with NdeI and HindIII and cloned into a pET-22b (Novagen) backbone, producing pJR111. This plasmid contains a truncated form of *Ng*PilF, where

the first 165 amino acids comprising the N1D are absent from the N-terminus, with a 6 x His tag at its C-terminus.

The coding region of *Cd*PilB (*pilB1*) was similarly amplified from a *Clostridioides difficile* 630Δ*erm* chromosomal DNA template using the primers *Cd*PilBpET22_F (TGGCTAAAAAAGTTAGAATAGGAGA) and *Cd*PilBpET22_R (CATAAGCTTATCAGTCATCAAAATTATTCTAATCATTT). The DNA fragment was digested with HindIII. The pET-22b backbone was digested with NdeI, treated with T4 DNA polymerase (new England Biolabs), then digested with HindIII. The digested DNA fragment was cloned into the pET-22b backbone to create pJR101. Site-directed mutagenesis by the Q5 High Fidelity DNA polymerase mutagenesis kit (New England Biolabs) was used to remove an internal NdeI restriction enzyme site. The plasmid was recircularized using KLD Enzyme Mix (New England Biolabs), creating pJR102. Six plasmids (pJR105, pJR106, pJR107, pJR109, pJR110, and pJR112) were subsequently generated using pJR102. Three of these plasmids (pJR109, pJR1110, pJR112) are various N-terminal truncations of *Cd*PilB, removing 127, 137, and 169 amino acids, respectively. The other three plasmids (pJR105, full-length; pJR106, 127 amino acid truncation, pJR107; 137 amino acid truncation) are *Cd*PilB-Hcp1 fusions. The hexameric *P. aeruginosa* protein Hcp1 is added as a stabilizing tag with the intention to improve stability of the hexameric PilB complex [178].

To produce the Hcp1 stabilizing tag, the coding region of Hcp1 was amplified from a *P. aeruginosa* PAO1 chromosomal DNA template with the primers Hcp1_F (GAACATATGCATGGCCATGGTACCCGGGATCCATGGCTGTTGATATGTTCAT CAAGATCGGC) and Hcp1_R (AATGAATTCAAGGCCTGCACGTTCTGGCGGAT).

This DNA fragment was digested with NdeI and BamHI, ligated to a pET-22b backbone, then digested with NdeI and EcoRI for ligation with *Cd*PilB. This fusion contains a 6-aa flexible linker (GSGSGS) between the C-terminus of *Cd*PilB and N-terminus of Hcp1, as well as a 6 x His tag at the C-terminus of Hcp1.

pJR101, pJR102, pJR105, pJR106, pJR107, pJR109, pJR110, pJR111, and pJR112 were screened by transformation of *E. coli* XL1-Blue, then sequenced for confirmation via Sanger sequencing by the Genomics Sequencing Center at the Fralin Life Sciences Institute of Virginia Tech. All plasmids except pJR102 were subsequently transformed into the expression strain *E. coli* BL21(DE3).

Expression and purification of *Ng*PilF$_{166}$ by IPTG induction

Cultures of *E. coli* BL21(DE3) with pJR111 were grown in 75 mL of LB overnight at 37°C using a rotary shaker set to 250 rpm. Four flasks with 500 mL LB were inoculated to an OD_{600} of 0.1 with the overnight cultures and grown at 37°C and 180 rpm until an OD_{600} of 0.6 was reached. The cultures were then transferred to 18°C and isopropyl ß-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.2 mM to induce PilF expression. Cells were then allowed to grow at 18°C for 22-24 hours. Cells were harvested by centrifugation at 4°C, weighed, then resuspended at 5 mL/g in cold Buffer A (50 mM TAPS: Tris pH 8.7, 75 mM KCl, 50 mM NaOAc, 10 mM $MgCl_2$, 5 mM imidazole, 5% glycerol, and 5 mM β-mercaptoethanol). The mixture was supplemented with 0.3 mM PMSF prior to lysis. The cell suspension was then sonicated on ice for 20 cycles of 15 seconds in 45 second intervals using the Sonic Dismembrator Model 500 (Fisher Scientific), at 50% amplitude. Lysis was visualized by phase contrast microscopy, then centrifuged at 40,000 × g for 1 hour at 4°C to isolate the supernatant.

The whole cell lysate, insoluble fraction, and soluble fractions were adjusted to OD_{600} ~15 and visualized by SDS-PAGE to confirm the presence of target protein.

The supernatant was collected and filtered through a 0.45 µm syringe membrane filter before loading onto a 20 mL Ni-NTA column equilibrated with Buffer A. 6 x His tagged *Ng*PilF was eluted using a gradient of 5 mM to 500 mM imidazole with a flow rate of 0.5 mL/min collected in 5 mL fractions. Fractions containing PilF protein were identified by SDS-PAGE, then pooled, concentrated, and filtered with a 0.45 µm syringe membrane filter before loading onto a Sephacryl S-300 26/60 column equilibrated with gel filtration buffer (50 mM TAPS:Tris pH 8.7, 75 mM KCl, 50 mM NaOAc, 10 mM $MgCl_2$, 5% glycerol, and 2 mM TCEP). Protein was subsequently eluted using gel filtration buffer at a flow rate of 1.0 mL/min in 10 mL fractions. Fractions containing *Ng*$PilF_{166}$ were identified by SDS-PAGE, then pooled, concentrated, and filtered. Protein concentrations were determined via Bradford assay against bovine albumin serum as a standard (BioRad). Protein was aliquoted, flash frozen with liquid nitrogen, then stored at -80°C.

Expression of *Cd*PilB variant proteins by IPTG induction and auto-induction

For small-scale IPTG experiments at 37°C, cultures of *E. coli* BL21(DE3) containing pJR101 were grown in 5 mL LB overnight at 37°C using a rotary shaker set to 250 rpm. Six flasks with 25 mL LB each were inoculated to an OD_{600} of 0.1 using the overnight cultures and grown at 37°C and 250 rpm until an OD_{600} of 0.6 was achieved. IPTG was added to the flasks to the final concentrations of 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 mM IPTG to induce PilB expression. Cultures were grown for 3 hours at 37°C. 1.5 mL

aliquots were collected and pelleted by centrifugation at 16,000 × g for 1 minute at room temperature.

For IPTG experiments performed at 16°C, 10 mL of overnight inoculum was prepared in the same manner as the 37°C experiments. Eight flasks with 25 mL LB each were inoculated to OD_{600} of 0.1 and grown at 37°C and 250 rpm to an OD_{600} of 0.6. IPTG was added to the final concentrations of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, and 0.8 mM. Cultures were transferred to 16°C to grow for 24 hours. 1.5 mL aliquots were collected and harvested by centrifugation at 16,000 × g for 1 minute at 4°C. Larger scale IPTG induction using 0.1 mM IPTG was prepared similarly, except that two flasks with 500 mL of LB were inoculated to OD_{600} of 0.1. These were grown to OD_{600} of 0.6, then IPTG was added to 0.1 mM and cells were allowed to grow for 24 hours at 16°C at 250 rpm. Cultures were combined and cells were harvested by centrifugation at 5,000 × g for 30 minutes.

For small-scale experiments using the auto-induction method, 5 mL of pJR101 was cultured overnight for inoculum as described above. Two flasks with 25 mL auto-induction media (1% tryptone, 0.5% yeast extract, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 85 mM NaCl, 2 mM $MgSO_4$) each were inoculated to OD_{600} of 0.1. Inducing sugar mixtures containing lactose (‘lac’; 0.5% glycerol, 0.05% glucose, 0.2% lactose) or galactose (‘gal’; 0.5% glycerol, 0.05% glucose, 0.1% galactose) were also added to the auto-induction media at this point, replacing the need to add IPTG later [179]. Cultures were grown to an OD_{600} of 0.6 at 37°C and 250 rpm, then transferred to 18°C to grow for 36 hours. A 1.5 mL sample of each culture was taken at 24 hours and 36 hours.

For large-scale auto-induction experiments, cultures of *E. coli* BL21(DE3) with pJR101, pJR105, pJR106, pJR107, pJR109, pJR110, or pJR112 were grown in an overnight culture of 75 mL LB using the same growth conditions as above. Two flasks per construct containing 500 mL of auto-induction media and galactose or lactose inducing sugar mixture were inoculated to an OD_{600} of 0.1 with the respective overnight cultures. Cultures were grown at 37°C and 250 rpm until an OD_{600} of 0.4 was achieved. The OD_{600} prior to the temperature shift was changed from 0.6 to 0.4 to account for the extra time needed to cool a larger culture volume. The cultures were then transferred to 18°C and allowed to grow for 24 hours. Cultures from the two flasks were combined and 1.5 mL aliquots were taken to observe protein expression and solubility. Cells were harvested by centrifugation at 5,000 × g for 30 minutes. Pellets were weighed then stored at -80°C for later purification.

Protein expression was analyzed by resuspending cell pellets using 500 µL Buffer A and visualizing by SDS-PAGE. To analyze the protein solubility in the small-scale inductions, pellets were resuspended in the same manner then solubilized by sonication on ice for 20 cycles of 5 seconds in 30 second intervals using the Sonic Dismembrator Model 500 (Fisher Scientific). The cell suspensions were centrifuged at 20,000 × g for 10 minutes at 4°C to separate the soluble and insoluble fractions. Pellets from large scale experiments were resuspended using 5 mL/g of cold Buffer A (50 mM TAPS: Tris pH 8.7, 75 mM KCl, 50 mM NaOAc, 10 mM $MgCl_2$, 5 mM imidazole, 5% glycerol, and 5 mM β-mercaptoethanol) and sonicated on ice for 20 cycles of 15 seconds in 45 second intervals at 50% amplification. Cell suspensions were centrifuged at 40,000 × g for 1

hour at 4°C. Samples of the whole cell lysate, insoluble fraction, and soluble fractions were adjusted to OD_{600} ~10 and visualized by SDS-PAGE.

Purification of *Cd*PilB proteins

Cell pellets were resuspended in 5 mL/g cold Buffer A and 0.3 mM PMSF. To separate insoluble and soluble fractions, the cells were lysed by sonication then centrifuged as described in the previous section. To observe protein solubility, samples of the cell lysate, insoluble fraction, and soluble fractions were adjusted to OD_{600} ~15 and processed through SDS-PAGE. The supernatant containing the soluble proteins was filtered using a 0.45 µm syringe filter. A 20 mL Ni-NTA affinity column was pre-equilibrated with Buffer A, then the filtered supernatant was loaded and eluted using a gradient of 5 mM (Buffer A) to 500 mM imidazole (Buffer B). A flow rate of 0.5 mL/min was used and protein was collected in 5 mL fractions. SDS-PAGE was used to identify peak fractions containing *Cd*PilB. These fractions were pooled, concentrated, and filtered using a 0.45 µm syringe filter. This was loaded onto a Sephacryl S-300 26/60 column was pre-equilibrated with gel filtration buffer (50 mM TAPS:Tris pH 8.7, 75 mM KCl, 50 mM NaOAc, 10 mM $MgCl_2$, 5% glycerol, and 2 mM TCEP). A flow rate of 0.5 mL/min and fraction size of 10 mL was used. *Cd*PilB was eluted using gel filtration buffer. Peak fractions with *Cd*PilB were identified using SDS-PAGE. These were concentrated and filtered with a 0.45 µm syringe filter. Concentrations were determined through a Bradford assay (BioRad). Fractions were aliquoted then flash frozen with liquid nitrogen. Protein samples were subsequently stored at -80°C.

ATPase assays for condition optimization

ATPase activity of *Ng*PilF$_{166}$ was assessed by malachite green-based colorimetric endpoint assays. Reactions containing 2.7 µg/mL *Ng*PilF$_{166}$, 0.5 mM ATP, and selected buffering agents and salts were prepared in PCR tube strips. The initial assessment of activity used *Ct*PilB buffer (50 mM TAPS:Tris pH 8.7, 75 mM KCl, 50 mM NaOAc, 10 mM $MgCl_2$, and 2 mM TCEP) [156] and *Nm*PilF buffer (20 mM HEPES pH 8.0, 150 mM NaCl, and 5 mM $MgCl_2$) [95]. Reaction mixtures containing buffer and ATP but lacking protein were included as negative controls. These reactions were allowed to proceed at 37°C in an Applied Biosystems ProFlex PCR system for 30 minutes. Reactions were terminated by transferring the reaction to 4°C and adding 4% TCA for thermal and chemical stoppage. These were then transferred to a 96-well plate containing a series of KH_2PO_4 phosphate standards. 100 µL of malachite green reagent (0.034% malachite green oxalate, 10 mM ammonium molybdate, 1 N HCl, 3.4% EtOH, and 0.01% Tween-20) were added to each well and incubated at room temperature for 10 minutes. For optimization of KCl concentrations, reaction mixtures and controls were prepared directly in a 96-well plate and incubated at 37°C for 30 minutes. The plate was then transferred directly onto ice and 4% TCA was added to end the reaction. The KH_2PO_4 phosphate standard was added to the plate, then malachite green reagent was added to each well as described previously. The absorbance at 620 nm was measured using a Tecan Infinite F200 Pro microplate reader. Liberated phosphate in each sample was determined against the KH_2PO_4 phosphate standard curve. Activity of *Ng*PilF$_{166}$ was calculated by subtracting the background phosphate from the negative control, then dividing the concentration of liberated phosphate without background in nanomoles by protein concentration in milligrams and reaction time in minutes (nmol Pi/mg PilF/min).

For condition optimization, activity was normalized by subtracting the minimum value of the dataset then dividing by the difference between the maximum and minimum values, such that the absence or lowest value of a given variable has a value of 0 and the maximum normalized activity has a value of 1. Data normalization was performed to visualize the optima of each variable with greater clarity and remove differences between the assaying methods used. All assays were performed in triplicate.

Inhibitor assays

The ATPase assay described above was modified to observe the effects of selected compounds on the activity of *Ng*PilF$_{166}$. Compounds were added at the concentrations indicated to the initial reaction mixture containing *Ng*PilF$_{166}$, ATP, and the optimized activity buffer described herein. Mixtures containing buffer, ATP, and either quercetin or Benserazide were included as negative controls. All other aspects of the ATPase assay procedure remained the same. The half maximal inhibitory concentration (IC_{50}) of quercetin was determined using a 4-parameter logistic model [180].

Results

Expression and purification of *Ng*PilF$_{166}$. Recent work with *Neisseria meningitidis* showed that *Nm*PilF can be stably purified when truncated at the N-terminus by 165 amino acids, removing the first N-terminal domain and adjoining linker region. This protein appeared to be highly active and amenable to *in vitro* experimentation [95]. Since *N. gonorrhoeae* and *N. meningitidis* PilF are 99% identical in amino acid sequences, a similar truncated *N. gonorrhoeae* PilF may be purified and serve as the basis for high-throughput screening for inhibitory compounds. Briefly, the coding sequence for

a N1D-truncated *N. gonorrhoeae* PilF, hereafter referred to as *Ng*PilF$_{166}$, was PCR amplified from FA 1090 genomic DNA and cloned into a plasmid to create pJR111 with *pilF* under the control of an inducible T7 promoter. Expression of *Ng*PilF$_{166}$ was induced by addition of IPTG, after which the cells were harvested, lysed, and fractionated to isolate the soluble fraction containing the target protein. This protein contains a 6×His tag at its C-terminal, allowing purification of the supernatant by fast protein liquid chromatography (FPLC). The supernatant containing *Ng*PilF$_{166}$ was first processed through Ni-NTA affinity chromatography and began elution at approximately 150 mM imidazole. The collected fractions were subsequently purified through size exclusion chromatography (SEC).

Interestingly, the elution profiles from both steps of chromatography showed two distinct absorbance peaks (FIG 3-2). The first peak of the elution profile resulting from affinity chromatography shows no notable protein banding when visualized via SDS-PAGE. Due to the absence of banding but notable absorbance signal, the proteins in these fractions likely simply exhibit high absorbance. When analyzed by SDS-PAGE, the second peak does appear to contain the target protein, producing the expected signal at ~45 kDa and less prominent banding at lower molecular weights. These additional bands are possibly a result of contamination or protein degradation during the elution process. When analyzing the SEC fractions by SDS-PAGE, both peaks in the elution profile show near identical protein banding, with the most prominent band at ~45 kDa. Notably, despite producing a similar banding pattern, the first SEC peak produced very little protein, with overall yield drastically greater in the fractions within the second peak. When both fractions were allowed to incubate at room temperature for 90 minutes,

protein from the first peak showed precipitation, while protein from the second peak displayed no precipitation or decrease in protein concentration (data not shown). This indicates that the proteins of the second peak are of a much stabler form.

***Ng*PilF$_{166}$ is an active ATPase.** The activity of *Ng*PilF$_{166}$ was assessed via a malachite green-based endpoint assay, a colorimetric assay that measures the concentration of phosphate liberated during the ATP hydrolysis process [135]. Activity was tested in both our in-house activity buffer used for *Chloracidobacterium thermophilum* PilB (*Ct*PilB) (50 mM TAPS:Tris pH 8.7, 75 mM KCl, 50 mM NaOAc, 10 mM $MgCl_2$, and 2 mM TCEP) [156] and literature-derived buffer used for *N. meningitidis* PilF (*Nm*PilF) (20 mM HEPES pH 8.0, 150 mM NaCl, and 5 mM $MgCl_2$) [95]. The reaction was performed using 2.7 µg/mL *Ng*PilF$_{166}$ at 37°C for 30 minutes using 0.5 mM ATP. *Ng*PilF$_{166}$ appears to be active in both buffers (FIG 3-3). However, activity is significantly elevated in the in-house buffer. *Ng*PilF$_{166}$ activity in the *Ct*PilB buffer was on average 352 ± 19.7 nmol Pi/mg/min while activity in the *Nm*PilF buffer was 92.5 ± 11.2 nmol Pi/mg/min. Given the major differences in buffer composition and possible degradation (low molecular weight banding present in SDS-PAGE, FIG 3-1), we sought to optimize the assay conditions to improve protein stability using our in-house activity buffer as the initial base recipe.

***Ng*PilF$_{166}$ activity is optimal in Tris buffer.** To determine the optimal buffering agent for *Ng*PilF$_{166}$ activity, 7 different buffering agents or agent combinations were tested based on various literature focused on PilB or PilB-like proteins [88, 156, 181, 182]. All buffering agents were pH-adjusted to 8.7, which is the pH of the *Ct*PilB buffer used by our laboratory. All other reaction conditions remained the same as described in

the previous section. As shown in FIG 3-4, *Ng*PilF$_{166}$ with Tris permits the greatest activity, with the second-best buffer agent combination being TAPS:Tris. TAPS, HEPES, MOPS, and Tris:MOPS:CHES showed comparable activity levels, while CHES alone showed very little activity. The combination of Tris:MOPS:CHES resulted in the lowest activity of the three Tris-containing buffers tested here. Since CHES showed little ATP hydrolysis when used as the sole buffering agent, it is possible that CHES may have an inhibitory effect on *Ng*PilF$_{166}$ that is partially resolved by the inclusion of Tris or MOPS. Based on these results, Tris was chosen for further optimization of the buffer and reaction conditions for analyzing *Ng*PilF$_{166}$ activity.

***Ng*PilF$_{166}$ is active at physiologically relevant temperatures.** To investigate the impact of reaction temperature on activity, endpoint assays were performed at various temperatures within the permissible temperature range of mesophiles. As expected, ATP hydrolysis steadily increases, then decreases with increasing temperature. Optimum activity is reached between 43°C and 46°C, after which hydrolysis begins to drop off, likely due to protein denaturation (FIG 3-5A). This is in line with the expected temperature tolerance of mesophilic organisms, which is approximately between 20°C and 45°C [183]. Because *N. gonorrhoeae* is an obligate pathogen of humans with atypical body temperature of 37°C, this temperature is chosen for the remainder of this study .

We next determined the optimum pH for *Ng*PilF$_{166}$ activity at 37°C using Tris as a buffering agent. The buffer thus adjusted to different pH within the working range of Tris from 7.0 to 9.2. From the results shown in FIG 3-5B, pH of 8.7 was determined to be optimal for *Ng*PilF$_{166}$ ATPase activity.

The in-house activity buffer for PilB contained two primary salts, NaOAc and KCl. In comparison, PilB/PilF buffers in the literature frequently contained only NaCl [95]. Changes or complete removal of NaOAc or inclusion of NaCl to the buffer composition appeared to have no significant influence on *Ng*PilF$_{166}$ activity (data not shown). In contrast, it appears that KCl is required for robust *Ng*PilF$_{166}$ activity (FIG 3-5C). Activity reaches a plateau at 250 mM KCl, after which there is no significant difference in ATP hydrolysis with increasing KCl concentration. From this, 300 mM KCl was selected for use in the final buffer composition, being within the lower range of the plateau that elicits similar activity as higher [KCl].

Since ATP requires Mg^{2+} for coordination in the ATP binding pocket of PilB and PilF ATPase, we sought to determine the influence of varying Mg^{2+} concentration on *Ng*PilF$_{166}$ activity (FIG3-5D). As expected, *Ng*PilF$_{166}$ displays low activity in the absence of supplemented Mg^{2+}. This low activity, rather than complete abrogation, may be attributed to potential carry-over of ions from the purification process. The rate of ATP hydrolysis reaches an optimum around 2 mM Mg^{2+} and does not significantly change as [Mg^{2+}] approaches 10 mM under our assay conditions with 0.5 mM ATP. Due to the dependence of ATP on [Mg^{2+}], 5 mM Mg^{2+} was selected to accommodate a wider range of [ATP] in kinetic analyses.

The final optimized buffer conditions were determined to be 50 mM Tris at pH 8.7, 300 mM KCl, and 5 mM $MgCl_2$. 37°C was chosen because it is the most physiologically relevant as *N. gonorrhoeae* is an obligate human pathogen. This temperature likely will result in less protein denaturation than at higher temperatures under the assay conditions.

***Ng*PilF$_{166}$ ATPase displays Michaelis-Menten kinetics.** To analyze the relationship between *Ng*PilF$_{166}$ activity and ATP concentration, a series of endpoint assays were performed with ATP concentrations ranging from 0 mM to 2.75 mM using the optimized buffer conditions described above. Figure 3-6 indicates that the rate of ATP hydrolysis by *Ng*PilF$_{166}$ increases with ATP concentration, resulting in a logarithmic curve. Activity appears to approach a plateau around 2 mM ATP. Plotting *Ng*PilF$_{166}$ activity against [ATP] closely resembles the classical Michaelis-Menten model of kinetics (FIG 3-6). From this, we can extract an approximate V_{max} of 625 nmol/mg/min and K_M of 0.4 mM ATP, where V_{max} is the maximum rate of hydrolysis and K_M is the concentration of ATP at ½ V_{max}. Given that the average intracellular concentration of ATP is between 1 to 5 mM, a K_M of 0.4 mM ATP is physiologically relevant [184]. Overall, this data suggests that *Ng*PilF$_{166}$ activity obeys Michaelis-Menten kinetics within this range of [ATP].

Quercetin inhibits *Ng*PilF$_{166}$. To analyze the effect of selected inhibitors on *Ng*PilF$_{166}$ activity, malachite green-based inhibition assays were performed using the flavonoid quercetin and Benserazide, a drug approved for the treatment of Parkinson's disease. As shown in FIG 3-7B, the inclusion of Benserazide caused no significant change in ATP hydrolysis by *Ng*PilF$_{166}$, with relative activity remaining at ~100% to Benserazide concentrations up to 162 μM. In contrast, FIG 3-7A indicates quercetin inhibits *Ng*PilF$_{166}$ activity in a concentration-dependent manner, with a half maximal inhibitory concentration (IC_{50}) of 24.9 ± 7.7 μM. Activity is completely abolished at concentrations ≥ 256 μM of quercetin. Although this IC_{50} value is higher than that

reported for *Ct*PilB [97], quercetin is nonetheless effective against *Ng*PilF$_{166}$, under our experiment conditions.

Expression constructs for *Cd*PilB variant proteins. *C. difficile* PilB, hereafter *Cd*PilB, was initially the intended target for optimization and assaying (as above), particularly in future high-throughput screening for inhibitory compounds. However, we continually encountered poor yield and low ATPase activity. Several approaches were attempted to overcome these obstacles. To improve both stability and yield, 8 recombinant plasmids were constructed using genomic DNA from *C. difficile* 630Δ*erm* to generate 7 variants of *Cd*PilB. The initial plasmid (pJR101), which contained the gene for full-length *Cd*PilB, was modified to remove an internal NdeI restriction enzyme site, and the resulting plasmid (pJR102) was used as the template for the subsequent variant proteins. Plasmids for the expression of the following proteins were created: full-length *Cd*PilB; N-terminal truncated *Cd*PilB lacking the first 127 (*Cd*PilB$_{128}$), 137 (*Cd*PilB$_{138}$), or 169 (*Cd*PilB$_{170}$) amino acids; and full-length *Cd*PilB, *Cd*PilB$_{128}$, and *Cd*PilB$_{138}$ C-terminally fused to *P. aeruginosa* Hcp1 protein as a stability tag. All *Cd*PilB constructs were tagged at the C-terminus with 6×His for purification via affinity chromatography, except the Hcp1 fusions which had the 6×His tag at the CTD of Hcp1. All N-terminal truncated variants lack the cyclic-di-GMP binding domain, with *Cd*PilB$_{128}$ and *Cd*PilB$_{138}$ closely reflecting the truncation described previously for *Ct*PilB [58] and *Cd*PilB$_{170}$ reflecting the truncation described for *Nm*PilF [95], respectively.

Expression of *Cd*PilB. Initial protein expression conditions were investigated using the full-length *Cd*PilB. As seen in FIG 3-8A, a small-scale traditional IPTG induction with 0.5 - 3 mM IPTG at 37°C produces a respectable yield of *Cd*PilB after just 3 hours

of induction; however, very little of this protein is soluble. Protein solubility may be improved by inducing at lower temperatures, effectively slowing the rate of expression and allowing more time for proper protein folding [185]. Decreasing the incubation temperature to 16°C and increasing the incubation length to 24 hours using 0.1 - 0.8 mM IPTG produces a comparable level of protein as the 37°C induction condition. At this lower temperature and incubation length, protein expression levels are not appreciably different after 0.1 mM IPTG (FIG 3-8B). Unfortunately, upscaling the induction to 1 L using 0.1 mM IPTG at 16°C still produced mostly insoluble protein (FIG 3-8C). An auto-induction method was explored in lieu of IPTG induction. In this method, glucose, glycerol, and an inducing sugar are directly included in the growth media rather than supplemented later as in IPTG induction (is this described in Materials and Methods? A record of exact conditions will be the most helpful for someone that may take over this project). This leverages the preference of glucose over lactose or galactose by *E. coli*, allowing a more gradual induction of protein expression compared to IPTG [186, 187]. The lowest temperature auto-induction is reportedly effective at is 18°C [187], so future auto-induction experiments were performed at 18°C and not 16°C. Separately, both lactose and galactose were used as inducing sugars to determine if either could improve the yield of *Cd*PilB. As shown in FIG 3-9A, a 25 mL auto-induction using both sugars resulted in similar levels of protein expression after 36 hours at 18°C. After upscaling the volume of auto-induction to two flasks of 500 mL culture, the differences between the inducing sugars become indistinguishable. FIG 3-9B depicts fractionated cell suspension after a 24 hour auto-induction with galactose in 1 L auto-induction media.

Approximately half of the proteins expressed were present in the soluble fraction, a considerable improvement over the IPTG induction.

Purification of *Cd*PilB. Purification of these auto-induced proteins by Ni-NTA affinity chromatography and SEC produces a perplexing elution profile (FIG 3-9C). Two absorbance peaks can be seen on both the Ni-NTA affinity and size exclusion chromatograms, suggesting *Cd*PilB may be eluting in multiple oligomeric conformations. The first peak had a stronger absorbance signal and appeared to produce active protein when tested in endpoint assays. However, this first peak also yielded very little protein, making downstream assaying impractical. Modifications were made to the auto-induction process in an effort to resolve the 'two peak problem'. Inclusion of a diguanylate cyclase into the expression strain to additionally produce cyclic-di-GMP (cdG), a second messenger that is known to bind PilB in other bacteria, did not change the FPLC profile and had no impact on activity. In contrast, the addition of Zn^{2+} into the auto-induction media resulted in a change in elution profile to a single peak. This, however, produced *Cd*PilB in an inactive form. Adding ATP to the auto-induction media also changes the elution profile, that is, the absorbance signal dramatically decreases to a point that becomes nearly undetectable.

Stabilization of *Cd*PilB by Hcp1 fusions. Work using a N-terminus truncated form of the *Vibrio cholerae* PilB-like protein GspE showed that fusion to the *P. aeruginosa* protein Hcp1 as a helper tag can induce hexamerization [178]. This method was additionally used in stabilizing *Myxococcus xanthus* PilB in hexameric form [188]. To determine whether this approach would be effective for *Cd*PilB as well, three *Cd*PilB constructs were fused to Hcp1: full-length *Cd*PilB and two N-terminal truncations,

*Cd*PilB$_{128}$ and *Cd*PilB$_{138}$. The expression profile for the three PilB-Hcp1 fusions is shown in FIG 3-10A. Expression levels are mostly identical between lactose and galactose, except that *Cd*PilB$_{138}$-Hcp1 does not appreciably induce under galactose. Analysis of protein solubility that *Cd*PilB$_{128}$-Hcp1 and *Cd*PilB$_{138}$-Hcp1 are produced in an insoluble form, while *Cd*PilB-Hcp1 has solubility comparable to the non-Hcp1 fusion (FIG 3-10B). Similar to the full-length non-Hcp1 fusion, *Cd*PilB-Hcp1 produced two absorbance peaks during FPLC, with the first peak yielding active protein in small quantities.

Additional work with *Cd*PilB variants. An additional N-terminal truncation, *Cd*PilB$_{170}$, was created in reflection of the truncation made by [95]. Preliminary auto-induction experiments indicate this protein is soluble under the described conditions (see Experimental Procedures); however, it has not yet been purified nor has its activity tested.

Overall, purification of these proteins by affinity and size exclusion chromatography yielded low yield of protein regardless of the variant, assuming the protein was found to be soluble and able to be purified. Some of these constructs, primarily the full-length variants, were found to be active via endpoint assay. Despite this, the yields of active *Cd*PilB variants were impractical to use for downstream applications.

Discussion

Here, we demonstrate that an N1D-truncated form of *Ng*PilF can be successfully constructed and purified. We additionally report optimized buffer conditions for *Ng*PilF$_{166}$ ATPase assay, and that this protein has robust ATPase activity that follows Michaelis-Menten kinetics. *Ng*PilF$_{166}$ is able to be inhibited by quercetin, thus showing potential for use as a model protein in anti-T4P antivirulence research.

The appearance of two distinct signals on the chromatograms in Ni-NTA affinity chromatography and SEC for *Ng*PilF$_{166}$ was unexpected. Effectively, of the two seen in affinity chromatography, only one peak contained *Ng*PilF$_{166}$. The two peaks from SEC, however, both contained *Ng*PilF$_{166}$ according to SDS-PAGE gels (FIG 3-1). It is likely that *Ng*PilF$_{166}$ is eluting as two different oligomeric species. This results in the larger oligomeric form eluting first, followed later by the second smaller, more abundant species. Thus, given that SDS-PAGE is a denaturing process, it is not surprising that protein from these two peaks produced identical banding patterns at the monomeric molecular weight of ~45 kDa. Since this second species showed apparent ATPase activity, it is likely eluting in a hexameric form [85, 188]. Another potential, but less likely, speculation is that there is some other protein or protein fragment that is co-purifying with PilF in that first peak and the complex is not stable enough to be seen by SDS-PAGE. It is unknown whether a similar pattern was observed for *Nm*PilF$_{166.}$

*Ng*PilF$_{166}$ was shown to have robust *in vitro* activity under the reported optimized conditions, whose composition is relatively minimal when compared to the *Ct*PilB buffer used previously. The activity of *Nm*PilF$_{166}$ in its respective buffer was reported to have a V_{max} of 25.1 ± 24.8 nmol/mg/min with a K_m of 0.45 ± 0.05 mM [95]. Although the K_m values of *Ng*PilF$_{166}$ and *Nm*PilF$_{166}$ are quite similar, there is a 25-fold difference between the V_{max} values (FIG 3-6). Part of this may be attributed to differences in the buffer compositions used for kinetics analyses as it was shown that *Ng*PilF$_{166}$ was dramatically more active in the in-house buffer (FIG 3-3). Yet, *Ng*PilF$_{166}$ activity in the original *Nm*PilF buffer was still nearly four times greater than the reported V_{max} for *Nm*PilF$_{166}$. Perplexingly, the kinetic analysis performed in [95] used a higher concentration of

protein than what is reported here. It is possible that other proteins or protein fragments may have also eluted with *Ng*PilF$_{166}$ and are contributing to this apparent ATPase activity, whose presence becomes visible when peak fractions are analyzed by SDS-PAGE (FIG 3-1). It is also likely, however, that these additional bands are products of *Ng*PilF$_{166}$ degradation, as buffer conditions were not yet optimized at the time of purification.

The compounds quercetin and Benserazide were chosen for the inhibition assays due to their previous identification as PilB inhibitors in high-throughput screens targeting *Ct*PilB [96, 97]. Most notably, *C. thermophilum* is an environmental, non-pathogenic species isolated from microbial mats from Yellowstone National Park, WY, USA [189]. This raises the concern of whether protein derived from a non-pathogenic species would be relevant for pathogens. Due to high conservation among the catalytic regions of PilB homologs, it is likely that inhibitors of this PilB would be effective on PilB homologs derived from other, pathogenic bacteria. Recent work by Dye et al. showed that Benserazide has inhibitory activity against T4P-related functions in *Acinetobacter nosocomialis*, an opportunistic nosocomial pathogen [96]. Our work has shown that Benserazide is not inhibitory toward *Ng*PilF$_{166}$. Quercetin, on the other hand, was effective against *Ng*PilF$_{166}$ (FIG 3-7). Alignment of *Ct*PilB and *Ng*PilF sequences shows a ~40% identity between the proteins. As expected for AAA+ proteins, the major catalytic domains, the Walker A and Walker B boxes, are both highly conserved. Although it was reported that Benserazide is an orthosteric inhibitor in *Ct*PilB [96], it is possible that there is some difference in *Ct*PilB and *Ng*PilF structure that prevents it from disrupting NgPilF activity at the C-terminal domain.

The recurrent problems of poor yield and poor ATPase activity resulted in the use of *Ng*PilF over *Cd*PilB. Particularly, during purification *Cd*PilB unexpectedly exhibited a double signal pattern on both FPLC chromatograms, similar to *Ng*PilF$_{166}$. Yet, the major species showed no apparent ATPase activity when tested through endpoint assay.

Using a T7 expression system, IPTG induces expression of target genes by binding the *lac* repressor, relieving inhibition on the *lac* operator and T7 promoter in a manner similar to allolactose. However, because IPTG is non-hydrolyzable it remains at constant levels in the cell, and 50% of total cell mass can become protein produced under the induced operon. This likely puts stress upon the cells, and such bulk production of protein can cause poor protein solubility and aggregation [187, 190]. Auto-induction methods are said to be able to produce similar, even higher, levels of protein expression when compared to a traditional IPTG-based approach. In this method, expression is achieved by the inclusion of glycerol, glucose, and lactose/galactose as carbon sources in the auto-induction media, allowing *E. coli* to gradually utilize lactose/galactose as an inducing sugar after glucose is depleted [179, 187, 190]. Thus, auto-induction is a more 'hands-off' method, requiring less monitoring of culture density since inducing sugars are added to the initial culture media. These aspects were attractive and appeared as a potential solution for our poor protein quality. This alternative method was more effective in producing more soluble protein relative to insoluble *Cd*PilB, but ultimately purified protein concentrations were still low. Longer incubation, from 36 hours up to 72 hours (data not shown), did not have appreciably higher yields. Yields also did not appreciably differ between use of lactose versus galactose for induction, although the auto-induction process is slower to initiate when galactose is used (FIG 3-9A). This is

expected as galactose is a weak inducer of the *lac* operon and must be present in high concentrations to initiate induction. *E. coli* BL21(DE3) lacks the *gal* operon required for galactose metabolism, thus galactose is able to accumulate to levels that eventually allow for similar levels of protein expression [191, 192].

Further, these issues were not able to be overcome by supplementation of the secondary messenger cdG or critical ions like Zn^{2+} or Mg^{2+}. Addition of a diguanylate cyclase only increased the viscosity of the bacterial culture after auto-induction. This much was expected since PilB and related proteins are known to bind cdG to modulate exopolysaccharide production [57, 58]. PilB contains a tetra-cysteine motif that binds Zn^{2+}, which was shown to improve PilB stability in *Thermus thermophilum* [88]. The addition of Zn^{2+} did cause only one peak to appear on chromatograms. This single peak suggests *Cd*PilB was stabilized in one form. Though this shift was unfavorable, since only inactive protein was produced, it does reveal that conditions are able to be altered to produce one primary form. ATP was later included in the auto-induction media because some assembly ATPases, such as T2SS XspE in *Xanthomonas campestris*, require ATP for hexamerization [193]. This was initially promising, however, the chromatogram signals completely vanished during purification and extremely low concentrations of *Cd*PilB were only able to be identified after processing fractions previously known to contain *Cd*PilB.

Past work has seen success in producing fusion proteins using *P. aeruginosa* Hcp1 as a tag [178, 188]. Hcp1 is thought to bring the hexameric subunits into close proximity as the tag itself oligomerizes into a hexamer. Our full-length *Cd*PilB-Hcp1 fusion was effectively the same as the non-Hcp1 fusion protein, exhibiting the same two

peak elution profile with low yield of active protein (FIG 3-9B & C). The two N-terminal truncation fusions, *Cd*PilB$_{128}$-Hcp1 and *Cd*PilB$_{138}$-Hcp1, were insoluble and unable to be purified. It is worth noting that *Cd*PilB$_{128}$ and *Cd*PilB$_{138}$ had poor solubility with and without the Hcp1 assistant tag, suggesting that some aspect of these truncations may be causing misfolding or aggregation. Overall, there may be some intrinsic property of *Cd*PilB that is remarkably unamenable to overexpression methods, or there are particular considerations for this type of protein that are not fulfilled with these common protocols. A final N-terminal truncation, *Cd*PilB$_{170}$, remains untested. Based on the success of *Ng*PilF$_{166}$, future work may find it worthwhile to investigate whether this particular form is better suited for biochemical experimentation.

Tables and Figures for Chapter 3

Table 3-1. Primers utilized in Chapter 3.

Primer Name	Sequence	Associated Plasmid(s)
CdPilBpET22_F	TGGCTAAAAAAGTTAGAATAGGAGA	pJR101
CdPilBpET22_R	CATAAGCTTATCAGTCATCAAAATTATTCTAATCATTT	pJR101, pJR109, pJR110, pJR112
CdPilB_SDM_F	GAAGCAGATATTGGAGAAAAAAAG	pJR102
CdPilB_SDM_R	ATACGAAACCTTACAATGC	pJR102
Hcp1_F	GAACATATGCATGGCCATGGTACCCGGGATCCATGGC TGTTGATATGTTCATCAAGATCGGC	-
Hcp1_R	AATGAATTCAAGGCCTGCACGTTCTGGCGGAT	-
CdPilB_N127_F	ACTCATATGAGTACAGAAGATATAAACAAAGC	pJR106, pJR109
CdPilB_N137_F	TTGCATATGGGTCATGCTTATGCACATTCA	pJR107, pJR110
CdPilB_hcp1_R	TAAGGTACCACTTCCACTTCCACTTCCATCAGTCATCA AAATTATTCTA	pJR105, pJR106, pJR107
CdPilB169_F	ACCCATATGGATGTAAATGCTGCTCCAATAGTA	pJR112
NgPilF166_F	TAGCATATGGAGGCAGAAGACGGCCCTGTT	pJR111
NgPilF166_R	CCGAAGCTTATCGTTGGTATTTGCCGTTACCTCTTCC	pJR111

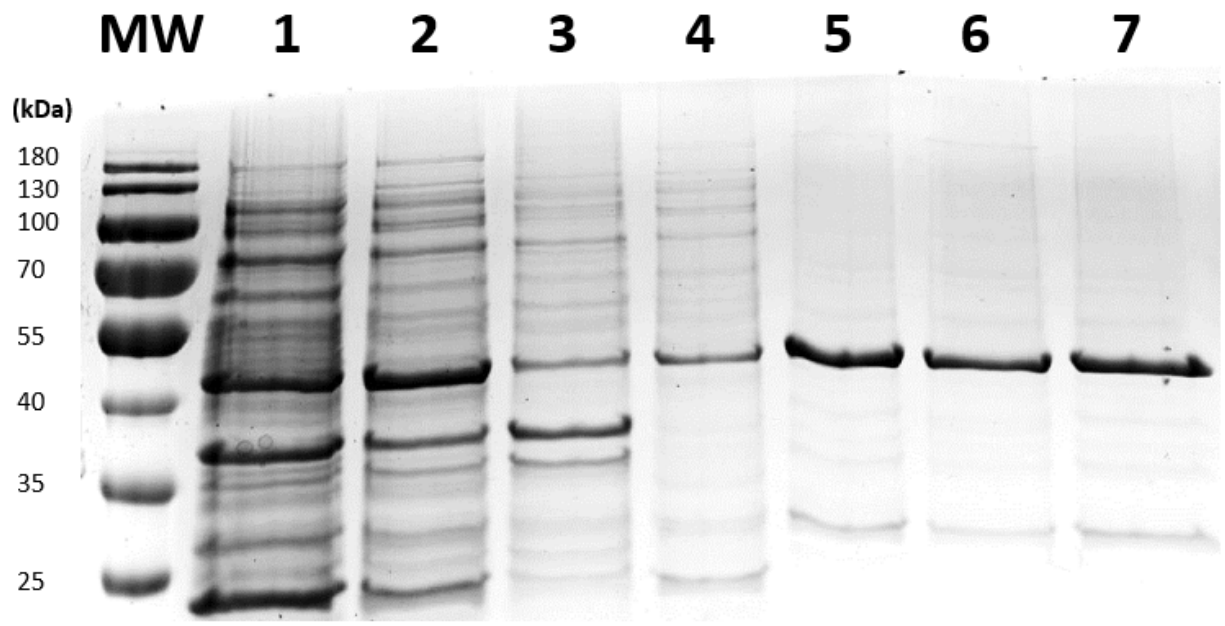

Figure 3-1. Expression and purification of *Ng*PilF$_{166}$.
Protein samples were run through SDS-PAGE and stained with Coomassie Blue. Molecular weight (MW) standards were loaded in the first lane with masses in kilodaltons (kDa) indicated to the left. From left to right, (1) whole cell lysate before induction, (2) whole cell lysate after induction, (3) insoluble fraction and (4) soluble fractions from (2), (5) pooled fractions from Ni-NTA affinity chromatography, and pooled fractions from peak 1 (6) and peak 2 (7) of size exclusion chromatography. *Ng*PilF$_{166}$ produces a signal at ~45 kDa.

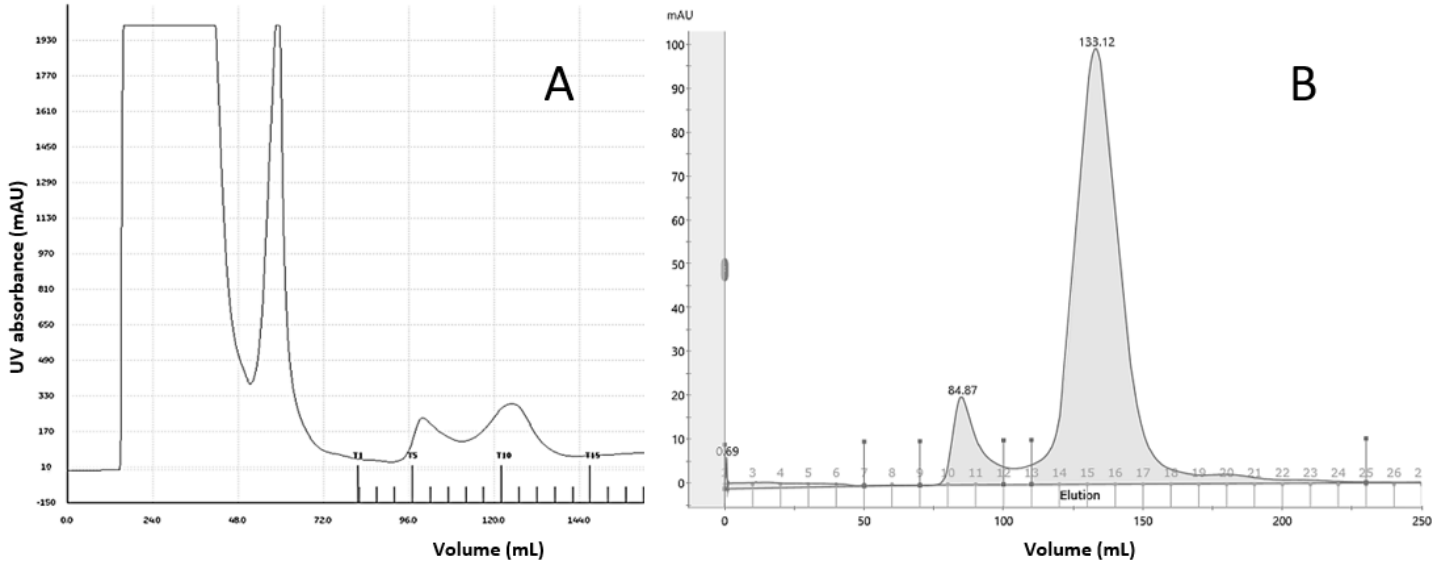

Figure 3-2. Purification of *Ng*PilF$_{166}$ by chromatography.
Chromatograms from the purification of *Ng*PilF$_{166}$. (**A**) The elution profile from Ni-NTA affinity chromatography depicts two high absorbance peaks from factions T4-T6 and T8-T12. *Ng*PilF$_{166}$ was only present in the T8-T12 peak. Flow-through proteins during the wash step create high absorbance signals, seen at the left of the chromatogram, which may be ignored. (**B**) The elution profile from size exclusion chromatography similarly shows two high absorbance peaks from fractions 10-11 and 13-17. *Ng*PilF$_{166}$ is present in both peaks, but only the second peak produces active protein with practical yield.

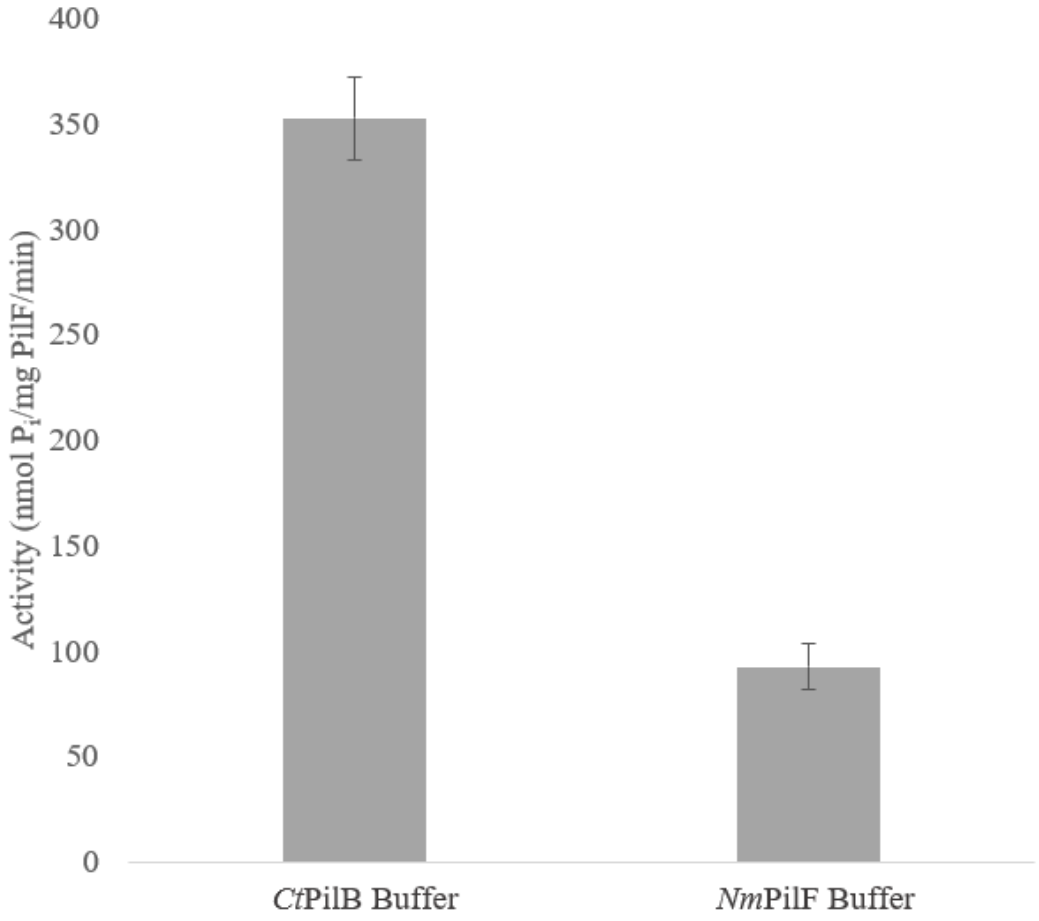

Figure 3-3. *Ng*PilF$_{166}$ is an active ATPase.
The activity of *Ng*PilF$_{166}$ was tested using *Ct*PilB ATPase buffer ("*Ct*PilB Buffer") or literature-derived buffer ("*Nm*PilF Buffer") by MLG endpoint assay. Error bars represent the standard deviation of three independent experiments performed in triplicate.

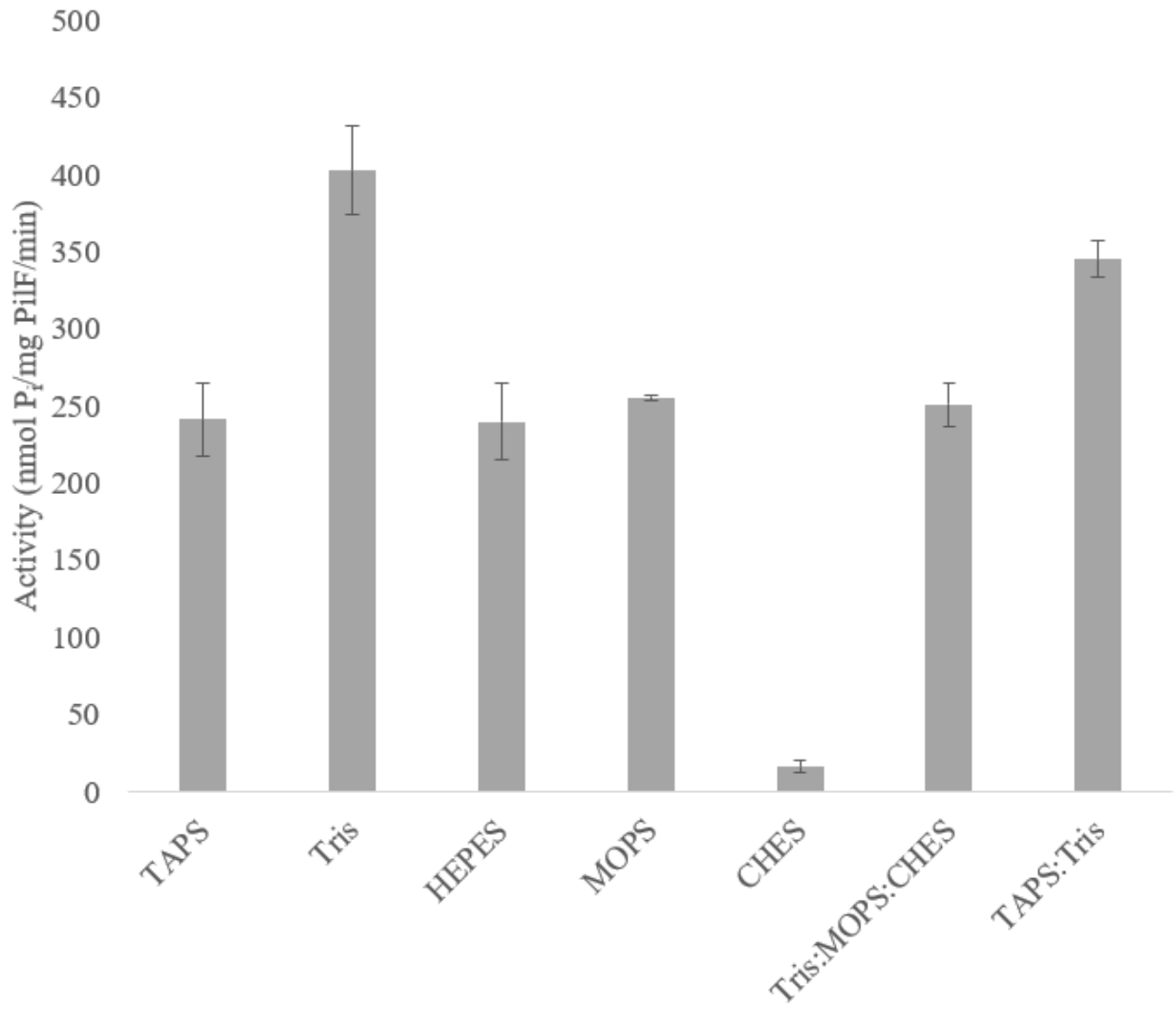

Figure 3-4. *Ng*PilF$_{166}$ activity is optimal in Tris buffer.
The activity of *Ng*PilF$_{166}$ was observed in the presence of 7 different buffering agents or agent combinations by endpoint assay. All buffering agents were added at 50 mM at pH 8.7; otherwise, all other buffer components were held constant with 75 mM KCl, 50 mM NaOAc, 10 mM $MgCl_2$ and 0.5 mM ATP. Error bars represent the standard deviation of three independent experiments performed in triplicate.

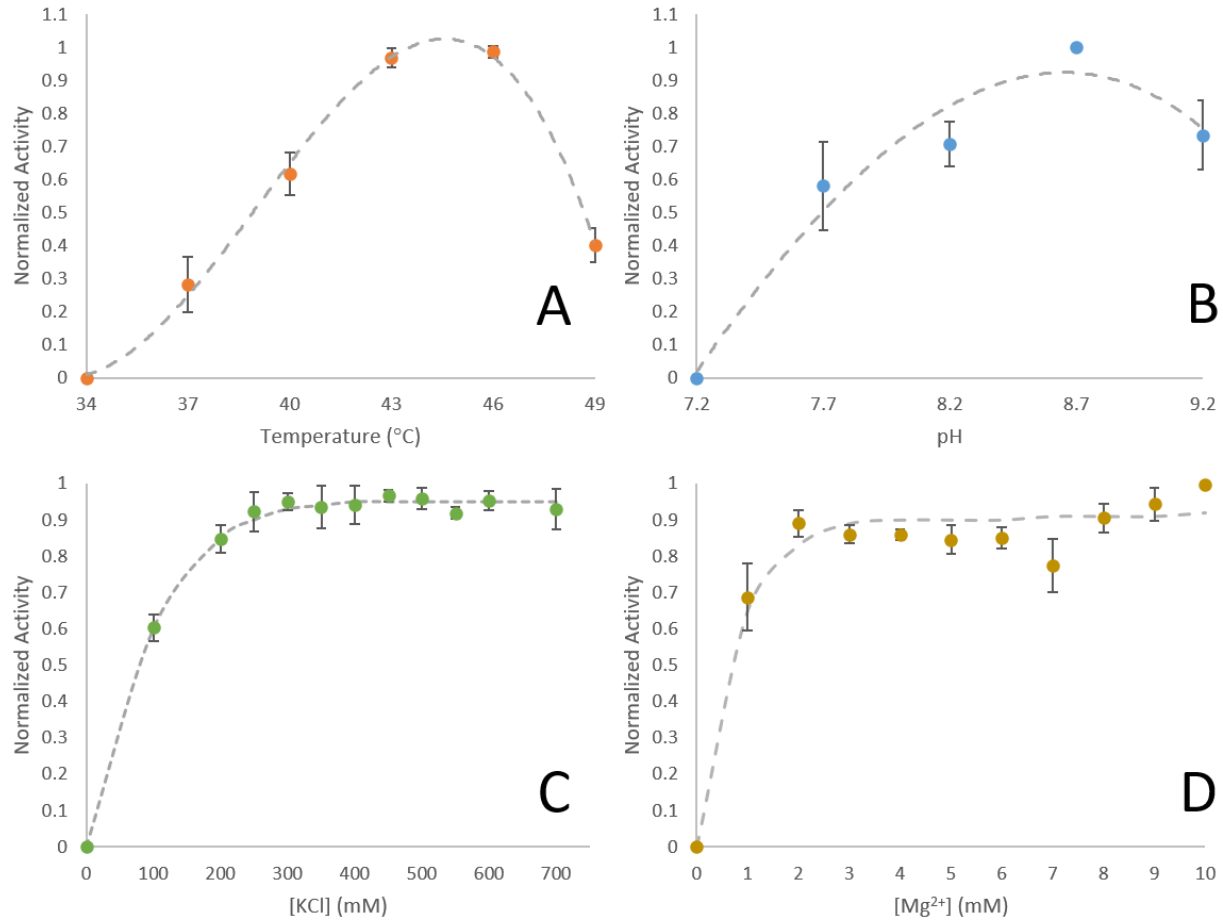

Figure 3-5. Assay condition optimization for *Ng*PilF$_{166}$ ATPase activity.
The impact on *Ng*PilF$_{166}$ ATPase activity by change in (A) temperature, (B) pH, (C) KCl, and (D) Mg^{2+} concentration are shown. Endpoint MLG assays were used for these experiments. Data was normalized to better visualize the point of optimal activity. Error bars represent the standard deviation of three independent experiments performed in triplicate.

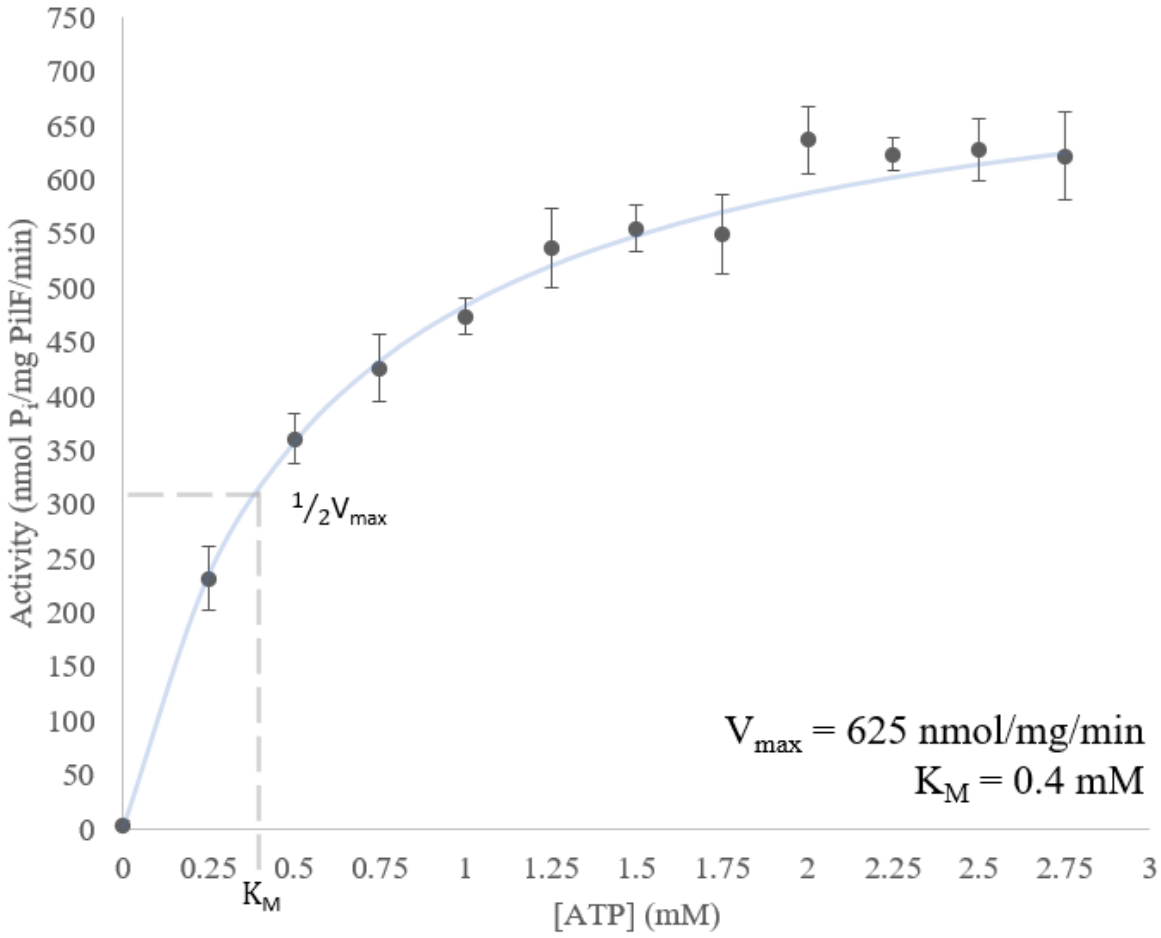

Figure 3-6. *Ng*PilF$_{166}$ ATPase displays Michaelis-Menten kinetics.
The reaction rate of *Ng*PilF$_{166}$ as a function of ATP concentration. The Michaelis constant K_M is calculated as the concentration of ATP at half the maximum reaction velocity, or $1/2V_{max}$. Endpoint MLG assays were used for this experiment. Error bars represent the standard deviation of three independent experiments performed in triplicate.

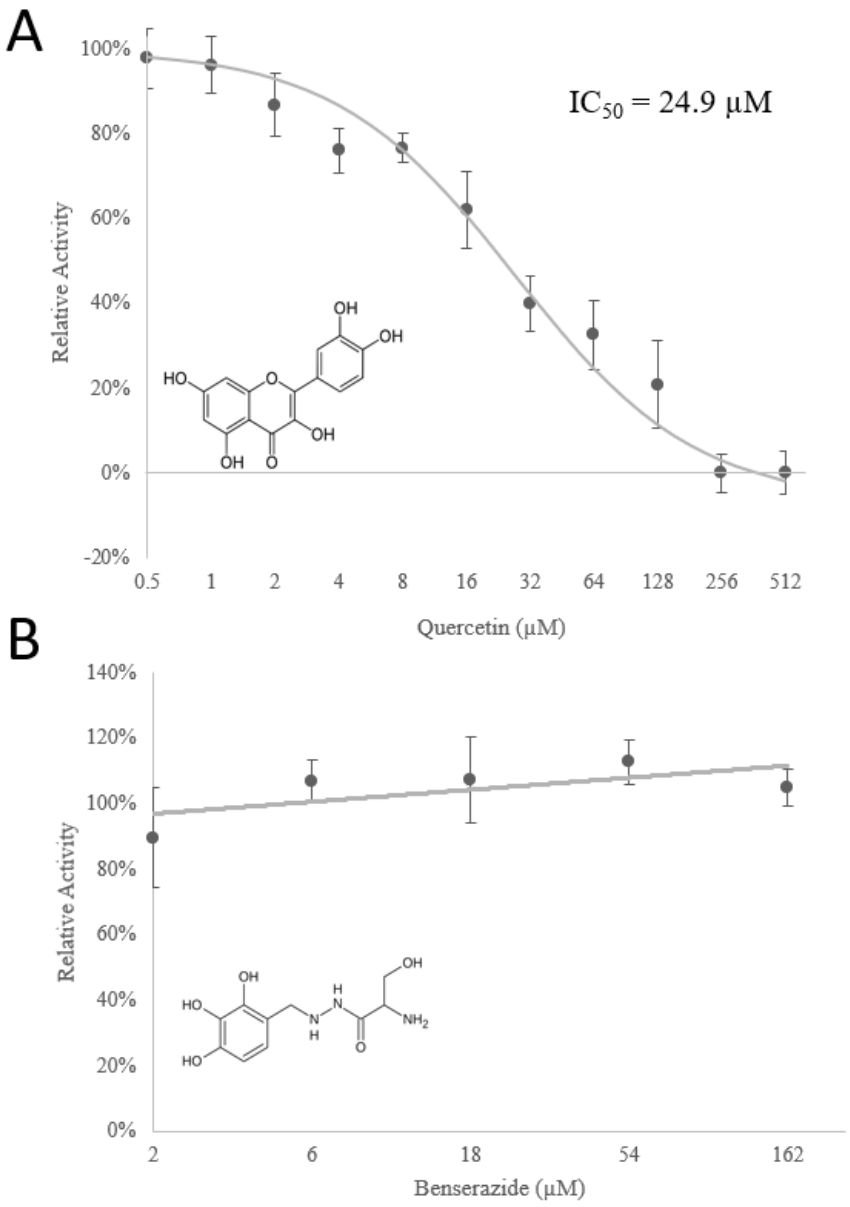

Figure 3-7. Quercetin inhibits *Ng*PilF$_{166}$ activity.
The activity of *Ng*PilF$_{166}$ was determined in the presence of (A) the flavonoid quercetin and (B) Benserazide at various concentrations. An IC_{50} of 24.9 µM was determined by curve fitting to a dose-response curve. Error bars represent the standard deviation of three independent experiments performed in triplicate. Inset: the molecular structures of quercetin and Benserazide.

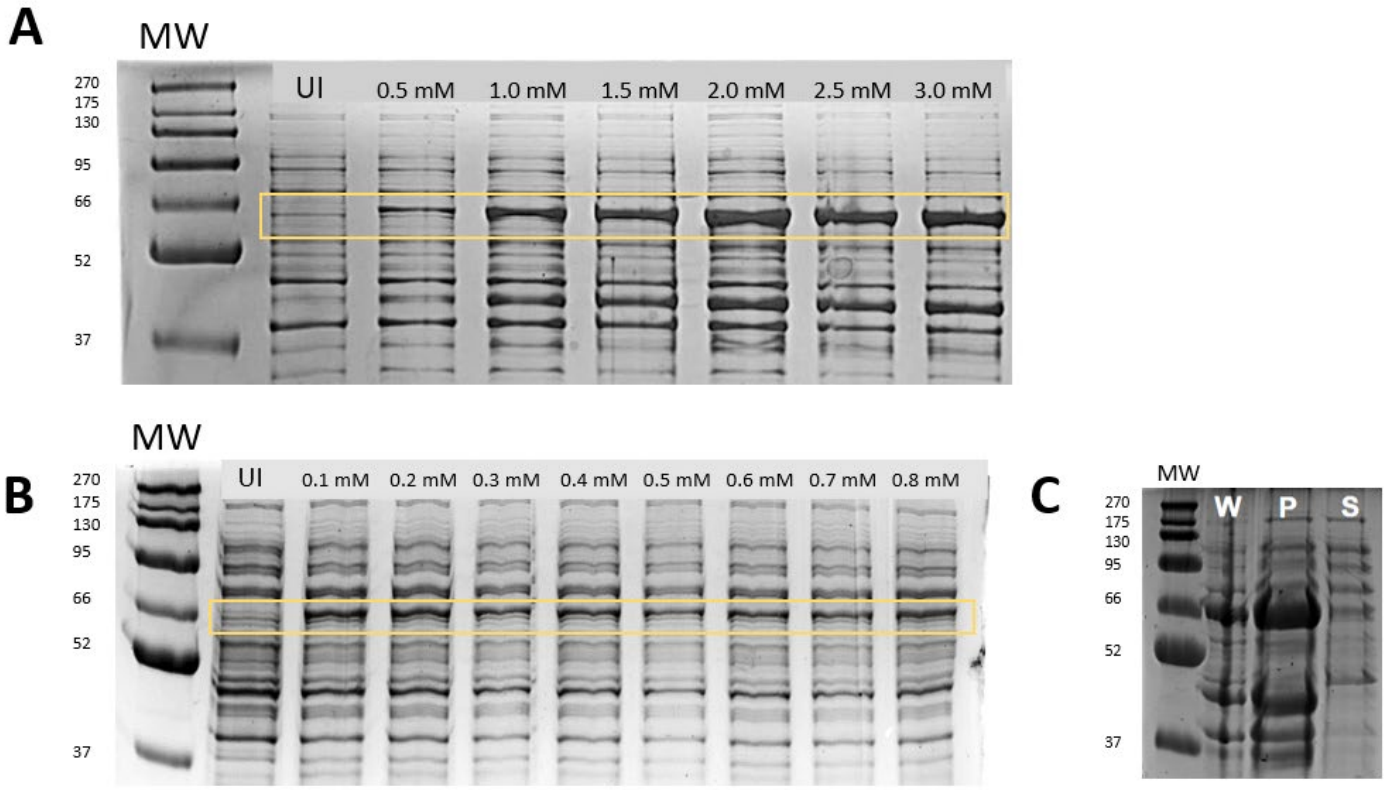

Figure 3-8. Induction of *Cd*PilB using IPTG.
SDS-PAGE gels depicting the expression of *Cd*PilB at (**A**) 37°C for 3 hours or (**B**) 16°C for 24 hours using various [IPTG]. *Cd*PilB produces a band at ~64 kDa (yellow box). (**C**) The solubility profile of *Cd*PilB expressed under the conditions in **B**. **UI** = uninduced, **W** = whole cell lysate, **P** = insoluble fraction, **S** = soluble fraction.

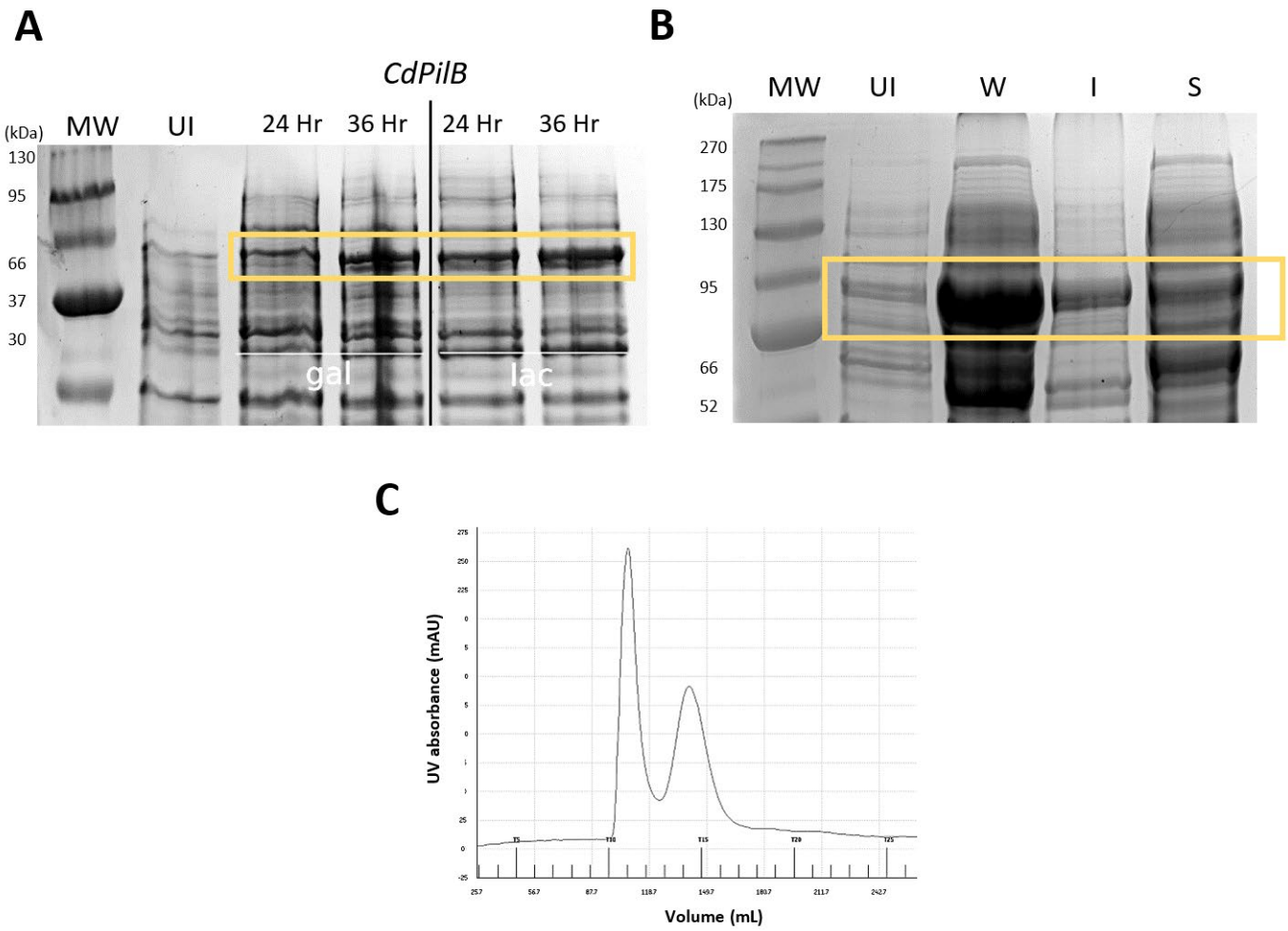

Figure 3-9. Auto-induction of *Cd*PilB.
(A) SDS-PAGE gel depicting the auto-induction of *Cd*PilB at 18°C for 24 or 36 hours using either galactose (gal) or lactose (lac) as the inducing sugar. (B) The solubility profile of *Cd*PilB after auto-induction using galactose. *Cd*PilB produces a band at ~64 kDa (yellow box). (C) The chromatogram from FPLC purification shows two high absorbance peaks. The shape and relative peak intensity are identical in both affinity chromatography and size exclusion chromatography. Only the first peak produces active protein, albeit with little yield. **UI** = uninduced, **W** = whole cell lysate, **I** = insoluble fraction, **S** = soluble fraction.

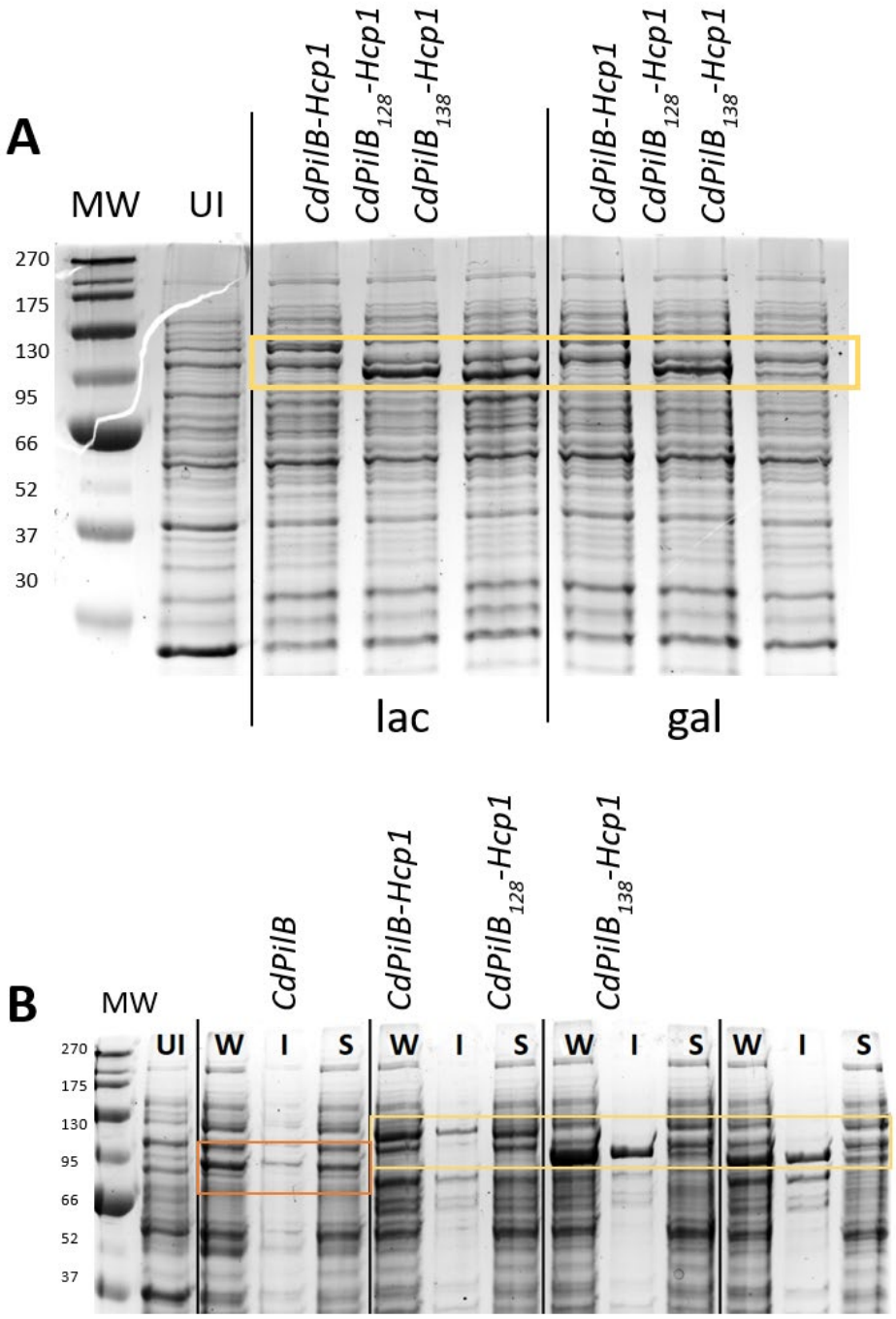

Figure 3-10. Expression of *Cd*PilB-Hcp1 fusion proteins.
(A) SDS-PAGE gel depicting the expression of *Cd*PilB-Hcp1 fusions by auto-induction methods at 18°C for 24 hours using either galactose (gal) or lactose (lac) as the inducing sugar. (B) The solubility profile of the *Cd*PilB-Hcp1 fusions after auto-induction using lactose. Non-Hcp1 tagged *Cd*PilB is included for comparison. *Cd*PilB-Hcp1 produces a band at ~83 kDa, *Cd*PilB$_{128}$-Hcp1 at ~69 kDa, and *Cd*PilB$_{138}$-Hcp1 at ~68 kDa (yellow box). *Cd*PilB produces a band at ~64 kDa (orange box). **UI** = uninduced, **W** = whole cell lysate, **I** = insoluble fraction, **S** = soluble fraction.

Chapter 4 - Concluding Remarks

The work in this thesis aimed to develop the foundation for antivirulence drug discovery targeting the bacterial type IV pilus by establishing the assembly ATPase PilB/PilF as a model for virtual screening and high-throughput screening. Chapter 2 evaluated the use of *Chloracidobacterium thermophilum* PilB as a model for *in silico* structure-based virtual screening to identify PilB inhibitors. Six PilB inhibitors reported in the literature were selected for molecular docking, using a *Ct*PilB model generated by homology modeling. Additional structural models or conformational ensembles were generated through molecular dynamics simulations of *Ct*PilB in differing liganded states. Molecular docking of inhibitors into the conformational ensembles revealed important structural characteristics of the PilB inhibitors. The data from the binding analyses align with existing experimental knowledge and support the use of *Ct*PilB in computational screening methods. Chapter 3 characterized an assembly ATPase of pathogenic origin by detailing the expression and purification of *Neisseria gonorrhoeae* PilF and the optimized conditions for its activity in malachite green-based endpoint assays. Here, it is shown to be a robust and active ATPase displaying Michaelis-Menten kinetics, with a V_{max} of ~625 nmol/mg/min. Chapter 3 also serves as a record of methods used toward expressing and purifying active *Clostridioides difficile* PilB with the hope that it will provide guidance in future endeavors with this protein.

Future work may consider the data presented here to further antivirulence efforts targeting the type IV pilus assembly ATPase. Structure-based virtual screening using compound libraries can be performed using the hexameric *Ct*PilB structure and the simulation-derived structural ensemble to identify PilB-specific inhibitors, particularly competitive inhibitors. It would be interesting to investigate inhibitory compound-induced morphological changes and the stability of PilB-inhibitor interactions through molecular dynamics simulations. While this could

be computationally expensive, it would supply an additional metric to judge the most promising hits, especially with improvement in computational capabilities. Future experimental work may also explore the development of high-throughput screens using *N. gonorrhoeae* PilF as a target. This is especially appealing as *N. gonorrhoeae* is a high-priority pathogen with widespread antibiotic resistance and T4P is indispensable for Neisserial pathogenesis. Purifying and characterizing an active *C. difficile* PilB could serve as an exciting challenge for future work as well. In all, the research presented here may inform future decisions in the process of antivirulence therapeutic discovery and drug optimization efforts targeting the type IV pilus and its core assembly ATPase.

References

1. CDC. Antibiotic resistance threats in the United States. Centers for Disease Control and Prevention. 2019:1-113.
2. Hutchings MI, Truman AW, Wilkinson B. Antibiotics: past, present and future. Curr Opin Microbiol. 2019;51:72-80. Epub 20191113. doi: 10.1016/j.mib.2019.10.008. PubMed PMID: 31733401.
3. Munita JM, Arias CA. Mechanisms of Antibiotic Resistance. Microbiol Spectr. 2016;4(2). doi: 10.1128/microbiolspec.VMBF-0016-2015. PubMed PMID: 27227291; PubMed Central PMCID: PMCPMC4888801.
4. Reygaert WC. An overview of the antimicrobial resistance mechanisms of bacteria. AIMS Microbiol. 2018;4(3):482-501. Epub 20180626. doi: 10.3934/microbiol.2018.3.482. PubMed PMID: 31294229; PubMed Central PMCID: PMCPMC6604941.
5. Antimicrobial Resistance C. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399(10325):629-55. Epub 20220119. doi: 10.1016/S0140-6736(21)02724-0. PubMed PMID: 35065702; PubMed Central PMCID: PMCPMC8841637.
6. World Health O. Prioritization of pathogens to guide discovery, research and development of new antibiotics for drug-resistant bacterial infections, including tuberculosis. Geneva: World Health Organization, 2017 2017. Report No.: 9789240026438 (electronic version) Contract No.: WHO/EMP/IAU/2017.12.
7. O'Neill J. Tackling drug-resistant infections globally: final report and recommendations. 2016.
8. Levy SB, Marshall B. Antibacterial resistance worldwide: causes, challenges and responses. Nature Medicine. 2004;10(S12):S122-S9. doi: 10.1038/nm1145.
9. Bush K. Antimicrobial agents targeting bacterial cell walls and cell membranes. Rev Sci Tech. 2012;31(1):43-56. doi: 10.20506/rst.31.1.2096. PubMed PMID: 22849267; PubMed Central PMCID: PMC22849267.
10. Davies J, Davies D. Origins and evolution of antibiotic resistance. Microbiol Mol Biol Rev. 2010;74(3):417-33. doi: 10.1128/MMBR.00016-10. PubMed PMID: 20805405; PubMed Central PMCID: PMCPMC2937522.
11. Darby EM, Trampari E, Siasat P, Gaya MS, Alav I, Webber MA, et al. Molecular mechanisms of antibiotic resistance revisited. Nat Rev Microbiol. 2023;21(5):280-95. Epub 20221121. doi: 10.1038/s41579-022-00820-y. PubMed PMID: 36411397.
12. Fournier PE, Richet H. The epidemiology and control of *Acinetobacter baumannii* in health care facilities. Clin Infect Dis. 2006;42(5):692-9. Epub 20060126. doi: 10.1086/500202. PubMed PMID: 16447117.
13. Sharma D, Misba L, Khan AU. Antibiotics versus biofilm: an emerging battleground in microbial communities. Antimicrob Resist Infect Control. 2019;8:76. Epub 20190516. doi: 10.1186/s13756-019-0533-3. PubMed PMID: 31131107; PubMed Central PMCID: PMCPMC6524306.
14. Santajit S, Indrawattana N. Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. Biomed Res Int. 2016;2016:2475067. Epub 20160505. doi: 10.1155/2016/2475067. PubMed PMID: 27274985; PubMed Central PMCID: PMCPMC4871955.
15. Tooke CL, Hinchliffe P, Bragginton EC, Colenso CK, Hirvonen VHA, Takebayashi Y, et al. beta-Lactamases and beta-Lactamase Inhibitors in the 21st Century. J Mol Biol. 2019;431(18):3472-500. Epub 20190405. doi: 10.1016/j.jmb.2019.04.002. PubMed PMID: 30959050; PubMed Central PMCID: PMCPMC6723624.

16. Fernandez L, Hancock RE. Adaptive and mutational resistance: role of porins and efflux pumps in drug resistance. Clin Microbiol Rev. 2012;25(4):661-81. doi: 10.1128/CMR.00043-12. PubMed PMID: 23034325; PubMed Central PMCID: PMCPMC3485749.
17. Fisher JF, Mobashery S. beta-Lactam Resistance Mechanisms: Gram-Positive Bacteria and *Mycobacterium tuberculosis*. Cold Spring Harb Perspect Med. 2016;6(5). Epub 20160502. doi: 10.1101/cshperspect.a025221. PubMed PMID: 27091943; PubMed Central PMCID: PMCPMC4852796.
18. Unemo M, Seifert HS, Hook EW, 3rd, Hawkes S, Ndowa F, Dillon JR. Gonorrhoea. Nat Rev Dis Primers. 2019;5(1):79. Epub 20191121. doi: 10.1038/s41572-019-0128-6. PubMed PMID: 31754194.
19. Tomberg J, Unemo M, Ohnishi M, Davies C, Nicholas RA. Identification of amino acids conferring high-level resistance to expanded-spectrum cephalosporins in the penA gene from *Neisseria gonorrhoeae* strain H041. Antimicrob Agents Chemother. 2013;57(7):3029-36. Epub 20130415. doi: 10.1128/AAC.00093-13. PubMed PMID: 23587946; PubMed Central PMCID: PMCPMC3697319.
20. Yoon EJ, Balloy V, Fiette L, Chignard M, Courvalin P, Grillot-Courvalin C. Contribution of the Ade Resistance-Nodulation-Cell Division-Type Efflux Pumps to Fitness and Pathogenesis of *Acinetobacter baumannii*. mBio. 2016;7(3). Epub 20160531. doi: 10.1128/mBio.00697-16. PubMed PMID: 27247231; PubMed Central PMCID: PMCPMC4895114.
21. Melnyk AH, Wong A, Kassen R. The fitness costs of antibiotic resistance mutations. Evol Appl. 2015;8(3):273-83. Epub 20140827. doi: 10.1111/eva.12196. PubMed PMID: 25861385; PubMed Central PMCID: PMCPMC4380921.
22. Rice PA, Shafer WM, Ram S, Jerse AE. *Neisseria gonorrhoeae*: Drug Resistance, Mouse Models, and Vaccine Development. Annu Rev Microbiol. 2017;71:665-86. doi: 10.1146/annurev-micro-090816-093530. PubMed PMID: 28886683.
23. Allen RC, Popat R, Diggle SP, Brown SP. Targeting virulence: can we make evolution-proof drugs? Nat Rev Microbiol. 2014;12(4):300-8. doi: 10.1038/nrmicro3232. PubMed PMID: 24625893.
24. Dunai A, Spohn R, Farkas Z, Lazar V, Gyorkei A, Apjok G, et al. Rapid decline of bacterial drug-resistance in an antibiotic-free environment through phenotypic reversion. Elife. 2019;8. Epub 20190816. doi: 10.7554/eLife.47088. PubMed PMID: 31418687; PubMed Central PMCID: PMCPMC6707769.
25. Praski Alzrigat L, Huseby DL, Brandis G, Hughes D. Fitness cost constrains the spectrum of marR mutations in ciprofloxacin-resistant *Escherichia coli*. J Antimicrob Chemother. 2017;72(11):3016-24. doi: 10.1093/jac/dkx270. PubMed PMID: 28962020; PubMed Central PMCID: PMCPMC5890708.
26. Basra P, Alsaadi A, Bernal-Astrain G, O'Sullivan ML, Hazlett B, Clarke LM, et al. Fitness Tradeoffs of Antibiotic Resistance in Extraintestinal Pathogenic *Escherichia coli*. Genome Biol Evol. 2018;10(2):667-79. doi: 10.1093/gbe/evy030. PubMed PMID: 29432584; PubMed Central PMCID: PMCPMC5817949.
27. Rasko DA, Sperandio V. Anti-virulence strategies to combat bacteria-mediated disease. Nat Rev Drug Discov. 2010;9(2):117-28. Epub 20100118. doi: 10.1038/nrd3013. PubMed PMID: 20081869.
28. Van Avondt K, van Sorge NM, Meyaard L. Bacterial immune evasion through manipulation of host inhibitory immune signaling. PLoS Pathog. 2015;11(3):e1004644. Epub

20150305. doi: 10.1371/journal.ppat.1004644. PubMed PMID: 25742647; PubMed Central PMCID: PMCPMC4351076.
29. Dickey SW, Cheung GYC, Otto M. Different drugs for bad bugs: antivirulence strategies in the age of antibiotic resistance. Nat Rev Drug Discov. 2017;16(7):457-71. Epub 20170324. doi: 10.1038/nrd.2017.23. PubMed PMID: 28337021.
30. Czepiel J, Drozdz M, Pituch H, Kuijper EJ, Perucki W, Mielimonka A, et al. *Clostridium difficile* infection: review. Eur J Clin Microbiol Infect Dis. 2019;38(7):1211-21. Epub 20190403. doi: 10.1007/s10096-019-03539-6. PubMed PMID: 30945014; PubMed Central PMCID: PMCPMC6570665.
31. Nait Chabane Y, Mlouka MB, Alexandre S, Nicol M, Marti S, Pestel-Caron M, et al. Virstatin inhibits biofilm formation and motility of *Acinetobacter baumannii*. BMC Microbiology. 2014;14(1):62. doi: 10.1186/1471-2180-14-62.
32. Hung DT, Shakhnovich EA, Pierson E, Mekalanos JJ. Small-Molecule Inhibitor of *Vibrio cholerae* Virulence and Intestinal Colonization. Science. 2005;310(5748):670-4. doi: 10.1126/science.1116739.
33. DiGiandomenico A, Keller AE, Gao C, Rainey GJ, Warrener P, Camara MM, et al. A multifunctional bispecific antibody protects against *Pseudomonas aeruginosa*. Science Translational Medicine. 2014;6(262):262ra155-262ra155. doi: 10.1126/scitranslmed.3009655.
34. Chastre J, François B, Bourgeois M, Komnos A, Ferrer R, Rahav G, et al. Safety, efficacy, and pharmacokinetics of gremubamab (MEDI3902), an anti-*Pseudomonas aeruginosa* bispecific human monoclonal antibody, in *P. aeruginosa*-colonised, mechanically ventilated intensive care unit patients: a randomised controlled trial. Critical Care. 2022;26(1):355. doi: 10.1186/s13054-022-04204-9.
35. Craig L, Volkmann N, Arvai AS, Pique ME, Yeager M, Egelman EH, et al. Type IV pilus structure by cryo-electron microscopy and crystallography: implications for pilus assembly and functions. Mol Cell. 2006;23(5):651-62. doi: 10.1016/j.molcel.2006.07.004. PubMed PMID: 16949362.
36. Clausen M, Jakovljevic V, Sogaard-Andersen L, Maier B. High-force generation is a conserved property of type IV pilus systems. J Bacteriol. 2009;191(14):4633-8. Epub 20090508. doi: 10.1128/JB.00396-09. PubMed PMID: 19429611; PubMed Central PMCID: PMCPMC2704717.
37. Craig L, Li J. Type IV pili: paradoxes in form and function. Curr Opin Struct Biol. 2008;18(2):267-77. Epub 20080204. doi: 10.1016/j.sbi.2007.12.009. PubMed PMID: 18249533; PubMed Central PMCID: PMCPMC2442734.
38. Reindl S, Ghosh A, Williams GJ, Lassak K, Neiner T, Henche AL, et al. Insights into FlaI functions in archaeal motor assembly and motility from structures, conformations, and genetics. Mol Cell. 2013;49(6):1069-82. Epub 20130214. doi: 10.1016/j.molcel.2013.01.014. PubMed PMID: 23416110; PubMed Central PMCID: PMCPMC3615136.
39. Albers SV, Jarrell KF. The archaellum: how Archaea swim. Front Microbiol. 2015;6:23. Epub 20150127. doi: 10.3389/fmicb.2015.00023. PubMed PMID: 25699024; PubMed Central PMCID: PMCPMC4307647.
40. Li J, Egelman EH, Craig L. Structure of the *Vibrio cholerae* Type IVb Pilus and stability comparison with the *Neisseria gonorrhoeae* type IVa pilus. J Mol Biol. 2012;418(1-2):47-64. Epub 20120221. doi: 10.1016/j.jmb.2012.02.017. PubMed PMID: 22361030; PubMed Central PMCID: PMCPMC3389824.

41. Giltner CL, Nguyen Y, Burrows LL. Type IV pilin proteins: versatile molecular modules. Microbiol Mol Biol Rev. 2012;76(4):740-72. doi: 10.1128/MMBR.00035-12. PubMed PMID: 23204365; PubMed Central PMCID: PMCPMC3510520.
42. Pelicic V. Type IV pili: e pluribus unum? Mol Microbiol. 2008;68(4):827-37. Epub 20080408. doi: 10.1111/j.1365-2958.2008.06197.x. PubMed PMID: 18399938.
43. Craig L, Forest K, Maier B. Type IV pili: dynamics, biophysics and functional consequences. Nature Reviews Microbiology. 2019;17(7):429-40. doi: 10.1038/s41579-019-0195-4.
44. Maier B, Wong GCL. How Bacteria Use Type IV Pili Machinery on Surfaces. Trends Microbiol. 2015;23(12):775-88. Epub 20151022. doi: 10.1016/j.tim.2015.09.002. PubMed PMID: 26497940.
45. Marathe R, Meel C, Schmidt NC, Dewenter L, Kurre R, Greune L, et al. Bacterial twitching motility is coordinated by a two-dimensional tug-of-war with directional memory. Nat Commun. 2014;5:3759. Epub 20140507. doi: 10.1038/ncomms4759. PubMed PMID: 24806757.
46. Sun H, Zusman DR, Shi W. Type IV pilus of *Myxococcus xanthus* is a motility apparatus controlled by the frz chemosensory system. Curr Biol. 2000;10(18):1143-6. doi: 10.1016/s0960-9822(00)00705-3. PubMed PMID: 10996798.
47. Burrows LL. *Pseudomonas aeruginosa* twitching motility: type IV pili in action. Annu Rev Microbiol. 2012;66:493-520. Epub 20120702. doi: 10.1146/annurev-micro-092611-150055. PubMed PMID: 22746331.
48. Skerker JM, Berg HC. Direct observation of extension and retraction of type IV pili. Proc Natl Acad Sci U S A. 2001;98(12):6901-4. Epub 20010529. doi: 10.1073/pnas.121171698. PubMed PMID: 11381130; PubMed Central PMCID: PMCPMC34450.
49. Wolfgang M, Lauer P, Park HS, Brossay L, Hebert J, Koomey M. PilT mutations lead to simultaneous defects in competence for natural transformation and twitching motility in piliated *Neisseria gonorrhoeae*. Mol Microbiol. 1998;29(1):321-30. doi: 10.1046/j.1365-2958.1998.00935.x. PubMed PMID: 9701824.
50. Tala L, Fineberg A, Kukura P, Persat A. *Pseudomonas aeruginosa* orchestrates twitching motility by sequential control of type IV pili movements. Nat Microbiol. 2019;4(5):774-80. Epub 20190225. doi: 10.1038/s41564-019-0378-9. PubMed PMID: 30804544; PubMed Central PMCID: PMCPMC6522360.
51. Beaussart A, Baker AE, Kuchma SL, El-Kirat-Chatel S, O'Toole GA, Dufrene YF. Nanoscale adhesion forces of *Pseudomonas aeruginosa* type IV Pili. ACS Nano. 2014;8(10):10723-33. Epub 20141006. doi: 10.1021/nn5044383. PubMed PMID: 25286300; PubMed Central PMCID: PMCPMC4212785.
52. Jamal M, Ahmad W, Andleeb S, Jalil F, Imran M, Nawaz MA, et al. Bacterial biofilm and associated infections. J Chin Med Assoc. 2018;81(1):7-11. Epub 20171015. doi: 10.1016/j.jcma.2017.07.012. PubMed PMID: 29042186.
53. Petrova OE, Sauer K. Sticky Situations: Key Components That Control Bacterial Surface Attachment. Journal of Bacteriology. 2012;194(10):2413-25. doi: 10.1128/JB.00003-12.
54. Schulze A, Mitterer F, Pombo JP, Schild S. Biofilms by bacterial human pathogens: Clinical relevance - Development, composition and regulation - Therapeutical strategies. Microbial Cell. 2021;8(2):28-56. doi: 10.15698/MIC2021.02.741.
55. Ciofu O, Tolker-Nielsen T, Jensen PO, Wang H, Hoiby N. Antimicrobial resistance, respiratory tract infections and role of biofilms in lung infections in cystic fibrosis patients. Adv

Drug Deliv Rev. 2015;85:7-23. Epub 20141202. doi: 10.1016/j.addr.2014.11.017. PubMed PMID: 25477303.
56. Jenal U, Reinders A, Lori C. Cyclic di-GMP: second messenger extraordinaire. Nat Rev Microbiol. 2017;15(5):271-84. Epub 20170206. doi: 10.1038/nrmicro.2016.190. PubMed PMID: 28163311.
57. McKee RW, Aleksanyan N, Garrett EM, Tamayo R. Type IV pili promote *Clostridium difficile* adherence and persistence in a mouse model of infection. Infection and Immunity. 2018;86(5):1-13. doi: 10.1128/IAI.00943-17.
58. Dye KJ, Yang Z. Cyclic-di-GMP and ADP bind to separate domains of PilB as mutual allosteric effectors. Biochem J. 2020;477(1):213-26. doi: 10.1042/BCJ20190809. PubMed PMID: 31868878; PubMed Central PMCID: PMCPMC6957770.
59. Black WP, Wang L, Jing X, Saldana RC, Li F, Scharf BE, et al. The type IV pilus assembly ATPase PilB functions as a signaling protein to regulate exopolysaccharide production in *Myxococcus xanthus*. Sci Rep. 2017;7(1):7263. Epub 20170804. doi: 10.1038/s41598-017-07594-x. PubMed PMID: 28779124; PubMed Central PMCID: PMCPMC5544727.
60. Roelofs KG, Jones CJ, Helman SR, Shang X, Orr MW, Goodson JR, et al. Systematic Identification of Cyclic-di-GMP Binding Proteins in *Vibrio cholerae* Reveals a Novel Class of Cyclic-di-GMP-Binding ATPases Associated with Type II Secretion Systems. PLOS Pathogens. 2015;11(10):e1005232. doi: 10.1371/journal.ppat.1005232.
61. Wu WL, Liao JH, Lin GH, Lin MH, Chang YC, Liang SY, et al. Phosphoproteomic analysis reveals the effects of PilF phosphorylation on type IV pilus and biofilm formation in *Thermus thermophilus* HB27. Mol Cell Proteomics. 2013;12(10):2701-13. Epub 20130704. doi: 10.1074/mcp.M113.029330. PubMed PMID: 23828892; PubMed Central PMCID: PMCPMC3790284.
62. Jones CJ, Utada A, Davis KR, Thongsomboon W, Zamorano Sanchez D, Banakar V, et al. C-di-GMP Regulates Motile to Sessile Transition by Modulating MshA Pili Biogenesis and Near-Surface Motility Behavior in *Vibrio cholerae*. PLoS Pathog. 2015;11(10):e1005068. Epub 20151027. doi: 10.1371/journal.ppat.1005068. PubMed PMID: 26505896; PubMed Central PMCID: PMCPMC4624765.
63. Hoiby N, Ciofu O, Bjarnsholt T. *Pseudomonas aeruginosa* biofilms in cystic fibrosis. Future Microbiol. 2010;5(11):1663-74. doi: 10.2217/fmb.10.125. PubMed PMID: 21133688.
64. Harding CM, Tracy EN, Carruthers MD, Rather PN, Actis LA, Munson RS, Jr. *Acinetobacter baumannii* strain M2 produces type IV pili which play a role in natural transformation and twitching motility but not surface-associated motility. mBio. 2013;4(4). Epub 20130806. doi: 10.1128/mBio.00360-13. PubMed PMID: 23919995; PubMed Central PMCID: PMCPMC3735195.
65. Lam T, Ellison CK, Eddington DT, Brun YV, Dalia AB, Morrison DA. Competence pili in *Streptococcus pneumoniae* are highly dynamic structures that retract to promote DNA uptake. Mol Microbiol. 2021;116(2):381-96. Epub 20210416. doi: 10.1111/mmi.14718. PubMed PMID: 33754381.
66. Seifert HS, Ajioka RS, Marchal C, Sparling PF, So M. DNA transformation leads to pilin antigenic variation in *Neisseria gonorrhoeae*. Nature. 1988;336(6197):392-5. doi: 10.1038/336392a0. PubMed PMID: 2904127.
67. Aas FE, Wolfgang M, Frye S, Dunham S, Lovold C, Koomey M. Competence for natural transformation in *Neisseria gonorrhoeae*: components of DNA binding and uptake linked to type

IV pilus expression. Mol Microbiol. 2002;46(3):749-60. doi: 10.1046/j.1365-2958.2002.03193.x. PubMed PMID: 12410832.
68. Averhoff B, Friedrich A. Type IV pili-related natural transformation systems: DNA transport in mesophilic and thermophilic bacteria. Arch Microbiol. 2003;180(6):385-93. Epub 20031031. doi: 10.1007/s00203-003-0616-6. PubMed PMID: 14593449.
69. Seitz P, Blokesch M. DNA-uptake machinery of naturally competent *Vibrio cholerae*. Proc Natl Acad Sci U S A. 2013;110(44):17987-92. Epub 20131014. doi: 10.1073/pnas.1315647110. PubMed PMID: 24127573; PubMed Central PMCID: PMCPMC3816411.
70. Melville S, Craig L. Type IV pili in Gram-positive bacteria. Microbiol Mol Biol Rev. 2013;77(3):323-41. doi: 10.1128/MMBR.00063-12. PubMed PMID: 24006467; PubMed Central PMCID: PMCPMC3811610.
71. Giltner CL, Habash M, Burrows LL. *Pseudomonas aeruginosa* minor pilins are incorporated into type IV pili. J Mol Biol. 2010;398(3):444-61. Epub 20100323. doi: 10.1016/j.jmb.2010.03.028. PubMed PMID: 20338182.
72. Helaine S, Dyer D, Nassif X, Pelicic V, Forest K. 3D structure/function analysis of PilX reveals how minor pilins can modulate the virulence properties of type IV pili. Proceedings of the National Academy of Sciences. 2007;104(40):15888-93. doi: 10.1073/pnas.0707581104.
73. Kolappan S, Coureuil M, Yu X, Nassif X, Egelman EH, Craig L. Structure of the *Neisseria meningitidis* Type IV pilus. Nature Communications. 2016;7(1):13015-. doi: 10.1038/ncomms13015.
74. Wang F, Coureuil M, Osinski T, Orlova A, Altindal T, Gesbert G, et al. Cryoelectron Microscopy Reconstructions of the *Pseudomonas aeruginosa* and *Neisseria gonorrhoeae* Type IV Pili at Sub-nanometer Resolution. Structure. 2017;25(9):1423-35 e4. doi: 10.1016/j.str.2017.07.016. PubMed PMID: 28877506; PubMed Central PMCID: PMCPMC8189185.
75. Carbonnelle E, Helaine S, Nassif X, Pelicic V. A systematic genetic analysis in *Neisseria meningitidis* defines the Pil proteins required for assembly, functionality, stabilization and export of type IV pili. Mol Microbiol. 2006;61(6):1510-22. doi: 10.1111/j.1365-2958.2006.05341.x. PubMed PMID: 16968224.
76. Szeto TH, Dessen A, Pelicic V. Structure/function analysis of *Neisseria meningitidis* PilW, a conserved protein that plays multiple roles in type IV pilus biology. Infect Immun. 2011;79(8):3028-35. Epub 20110606. doi: 10.1128/IAI.05313-11. PubMed PMID: 21646452; PubMed Central PMCID: PMCPMC3147589.
77. Tammam S, Sampaleanu LM, Koo J, Sundaram P, Ayers M, Chong PA, et al. Characterization of the PilN, PilO and PilP type IVa pilus subcomplex. Mol Microbiol. 2011;82(6):1496-514. Epub 20111118. doi: 10.1111/j.1365-2958.2011.07903.x. PubMed PMID: 22053789.
78. Ayers M, Sampaleanu LM, Tammam S, Koo J, Harvey H, Howell PL, et al. PilM/N/O/P proteins form an inner membrane complex that affects the stability of the *Pseudomonas aeruginosa* type IV pilus secretin. J Mol Biol. 2009;394(1):128-42. doi: 10.1016/j.jmb.2009.09.034. PubMed PMID: 19857645.
79. Berry JL, Phelan MM, Collins RF, Adomavicius T, Tonjum T, Frye SA, et al. Structure and assembly of a trans-periplasmic channel for type IV pili in *Neisseria meningitidis*. PLoS Pathog. 2012;8(9):e1002923. Epub 20120913. doi: 10.1371/journal.ppat.1002923. PubMed PMID: 23028322; PubMed Central PMCID: PMCPMC3441751.

80. McCallum M, Burrows LL, Howell PL. The Dynamic Structures of the Type IV Pilus. Microbiol Spectr. 2019;7(2). doi: 10.1128/microbiolspec.PSIB-0006-2018. PubMed PMID: 30825300.
81. Chang YW, Rettberg LA, Treuner-Lange A, Iwasa J, Sogaard-Andersen L, Jensen GJ. Architecture of the type IVa pilus machine. Science. 2016;351(6278):aad2001. Epub 20160310. doi: 10.1126/science.aad2001. PubMed PMID: 26965631; PubMed Central PMCID: PMCPMC5929464.
82. Snider J, Houry WA. AAA+ proteins: diversity in function, similarity in structure. Biochem Soc Trans. 2008;36(Pt 1):72-7. doi: 10.1042/BST0360072. PubMed PMID: 18208389.
83. Iyer LM, Leipe DD, Koonin EV, Aravind L. Evolutionary history and higher order classification of AAA+ ATPases. J Struct Biol. 2004;146(1-2):11-31. doi: 10.1016/j.jsb.2003.10.010. PubMed PMID: 15037234.
84. Erzberger JP, Berger JM. Evolutionary relationships and structural mechanisms of AAA+ proteins. Annu Rev Biophys Biomol Struct. 2006;35:93-114. doi: 10.1146/annurev.biophys.35.040405.101933. PubMed PMID: 16689629.
85. Thomsen ND, Berger JM. Structural frameworks for considering microbial protein- and nucleic acid-dependent motor ATPases. Mol Microbiol. 2008;69(5):1071-90. Epub 20080721. doi: 10.1111/j.1365-2958.2008.06364.x. PubMed PMID: 18647240; PubMed Central PMCID: PMCPMC2538554.
86. Jakovljevic V, Leonardy S, Hoppert M, Sogaard-Andersen L. PilB and PilT are ATPases acting antagonistically in type IV pilus function in *Myxococcus xanthus*. J Bacteriol. 2008;190(7):2411-21. Epub 20080125. doi: 10.1128/JB.01793-07. PubMed PMID: 18223089; PubMed Central PMCID: PMCPMC2293208.
87. Bos JL, Rehmann H, Wittinghofer A. GEFs and GAPs: critical elements in the control of small G proteins. Cell. 2007;129(5):865-77. doi: 10.1016/j.cell.2007.05.018. PubMed PMID: 17540168.
88. Salzer R, Herzberg M, Nies DH, Joos F, Rathmann B, Thielmann Y, et al. Zinc and ATP binding of the hexameric AAA-ATPase PilF from *Thermus thermophilus*: role in complex stability, piliation, adhesion, twitching motility, and natural transformation. J Biol Chem. 2014;289(44):30343-54. Epub 20140908. doi: 10.1074/jbc.M114.598656. PubMed PMID: 25202014; PubMed Central PMCID: PMCPMC4215219.
89. Solanki V, Kapoor S, Thakur KG. Structural insights into the mechanism of Type IVa pilus extension and retraction ATPase motors. FEBS J. 2018;285(18):3402-21. Epub 20180818. doi: 10.1111/febs.14619. PubMed PMID: 30066435.
90. McCallum M, Tammam S, Khan A, Burrows LL, Howell PL. The molecular mechanism of the type IVa pilus motors. Nat Commun. 2017;8:15091. Epub 20170505. doi: 10.1038/ncomms15091. PubMed PMID: 28474682; PubMed Central PMCID: PMCPMC5424180.
91. Chiang P, Sampaleanu LM, Ayers M, Pahuta M, Howell PL, Burrows LL. Functional role of conserved residues in the characteristic secretion NTPase motifs of the *Pseudomonas aeruginosa* type IV pilus motor proteins PilB, PilT and PilU. Microbiology (Reading). 2008;154(Pt 1):114-26. doi: 10.1099/mic.0.2007/011320-0. PubMed PMID: 18174131.
92. Mancl JM, Black WP, Robinson H, Yang Z, Schubot FD. Crystal Structure of a Type IV Pilus Assembly ATPase: Insights into the Molecular Mechanism of PilB from *Thermus thermophilus*. Structure. 2016;24(11):1886-97. Epub 20160922. doi: 10.1016/j.str.2016.08.010. PubMed PMID: 27667690.

93. Denis K, Le Bris M, Le Guennec L, Barnier JP, Faure C, Gouge A, et al. Targeting Type IV pili as an antivirulence strategy against invasive meningococcal disease. Nat Microbiol. 2019;4(6):972-84. Epub 20190325. doi: 10.1038/s41564-019-0395-8. PubMed PMID: 30911127.
94. Vo N, Sidner BS, Yu Y, Piepenbrink KH. Type IV Pilus-Mediated Inhibition of *Acinetobacter baumannii* Biofilm Formation by Phenothiazine Compounds. Microbiol Spectr. 2023:e0102323. Epub 20230621. doi: 10.1128/spectrum.01023-23. PubMed PMID: 37341603.
95. Aubey F, Corre JP, Kong Y, Xu X, Obino D, Goussard S, et al. Inhibitors of the *Neisseria meningitidis* PilF ATPase provoke type IV pilus disassembly. Proc Natl Acad Sci U S A. 2019;116(17):8481-6. Epub 20190404. doi: 10.1073/pnas.1817757116. PubMed PMID: 30948644; PubMed Central PMCID: PMCPMC6486710.
96. Dye KJ, Vogelaar NJ, O'Hara M, Sobrado P, Santos W, Carlier PR, et al. Discovery of Two Inhibitors of the Type IV Pilus Assembly ATPase PilB as Potential Antivirulence Compounds. Microbiology Spectrum. 2022;10(6):e03877-22. doi: 10.1128/spectrum.03877-22.
97. Dye KJ, Vogelaar NJ, Sobrado P, Yang Z. High-Throughput Screen for Inhibitors of the Type IV Pilus Assembly ATPase PilB. mSphere. 2021;6(2). Epub 20210303. doi: 10.1128/mSphere.00129-21. PubMed PMID: 33658276; PubMed Central PMCID: PMCPMC8546689.
98. Heiniger RW, Hanne, Winther-Larsen C, Pickles RJ, Koomey M, Wolfgang MC. Infection of human mucosal tissue by *Pseudomonas aeruginosa* requires sequential and mutually dependent virulence factors and a novel pilus-associated adhesinc mi_1461 1158..1173. 2010. doi: 10.1111/j.1462-5822.2010.01461.x.
99. Purcell EB, McKee RW, Bordeleau E, Burrus V, Tamayo R. Regulation of Type IV Pili Contributes to Surface Behaviors of Historical and Epidemic Strains of *Clostridium difficile*. J Bacteriol. 2016;198(3):565-77. Epub 20151123. doi: 10.1128/JB.00816-15. PubMed PMID: 26598364; PubMed Central PMCID: PMCPMC4719444.
100. Persat A, Inclan YF, Engel JN, Stone HA, Gitai Z. Type IV pili mechanochemically regulate virulence factors in *Pseudomonas aeruginosa*. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(24). doi: 10.1073/pnas.1502025112.
101. Lim KYL, Mullally CA, Haese EC, Kibble EA, McCluskey NR, Mikucki EC, et al. Anti-Virulence Therapeutic Approaches for *Neisseria gonorrhoeae*. Antibiotics (Basel). 2021;10(2). Epub 20210121. doi: 10.3390/antibiotics10020103. PubMed PMID: 33494538; PubMed Central PMCID: PMCPMC7911339.
102. Quillin SJ, Seifert HS. *Neisseria gonorrhoeae* host adaptation and pathogenesis. Nat Rev Microbiol. 2018;16(4):226-40. Epub 20180212. doi: 10.1038/nrmicro.2017.169. PubMed PMID: 29430011; PubMed Central PMCID: PMCPMC6329377.
103. Higashi DL, Lee SW, Snyder A, Weyand NJ, Bakke A, So M. Dynamics of *Neisseria gonorrhoeae* attachment: microcolony development, cortical plaque formation, and cytoprotection. Infect Immun. 2007;75(10):4743-53. Epub 20070806. doi: 10.1128/IAI.00687-07. PubMed PMID: 17682045; PubMed Central PMCID: PMCPMC2044525.
104. CDC. Sexually transmitted disease surveillance 2018. 2019.
105. Rowley J, Vander Hoorn S, Korenromp E, Low N, Unemo M, Abu-Raddad LJ, et al. Chlamydia, gonorrhoea, trichomoniasis and syphilis: global prevalence and incidence estimates, 2016. Bull World Health Organ. 2019;97(8):548-62P. Epub 20190606. doi: 10.2471/BLT.18.228486. PubMed PMID: 31384073; PubMed Central PMCID: PMCPMC6653813.

106. Derbie A, Mekonnen D, Woldeamanuel Y, Abebe T. Azithromycin resistant gonococci: a literature review. Antimicrob Resist Infect Control. 2020;9(1):138. Epub 20200818. doi: 10.1186/s13756-020-00805-7. PubMed PMID: 32811545; PubMed Central PMCID: PMCPMC7436955.
107. St Cyr S, Barbee L, Workowski KA, Bachmann LH, Pham C, Schlanger K, et al. Update to CDC's Treatment Guidelines for Gonococcal Infection, 2020. MMWR Morb Mortal Wkly Rep. 2020;69(50):1911-6. Epub 20201218. doi: 10.15585/mmwr.mm6950a6. PubMed PMID: 33332296; PubMed Central PMCID: PMCPMC7745960.
108. Ohnishi M, Golparian D, Shimuta K, Saika T, Hoshina S, Iwasaku K, et al. Is *Neisseria gonorrhoeae* initiating a future era of untreatable gonorrhea?: detailed characterization of the first strain with high-level resistance to ceftriaxone. Antimicrob Agents Chemother. 2011;55(7):3538-45. Epub 20110516. doi: 10.1128/AAC.00325-11. PubMed PMID: 21576437; PubMed Central PMCID: PMCPMC3122416.
109. Ohnishi M, Saika T, Hoshina S, Iwasaku K, Nakayama S-I, Watanabe H, et al. Ceftriaxone-Resistant *Neisseria gonorrhoeae*, Japan. Emerging Infectious Disease journal. 2011;17(1):148. doi: 10.3201/eid1701.100397.
110. Nakayama S, Shimuta K, Furubayashi K, Kawahata T, Unemo M, Ohnishi M. New Ceftriaxone- and Multidrug-Resistant *Neisseria gonorrhoeae* Strain with a Novel Mosaic penA Gene Isolated in Japan. Antimicrob Agents Chemother. 2016;60(7):4339-41. Epub 20160620. doi: 10.1128/AAC.00504-16. PubMed PMID: 27067334; PubMed Central PMCID: PMCPMC4914677.
111. Unemo M, Golparian D, Nicholas R, Ohnishi M, Gallay A, Sednaoui P. High-level cefixime- and ceftriaxone-resistant *Neisseria gonorrhoeae* in France: novel penA mosaic allele in a successful international clone causes treatment failure. Antimicrob Agents Chemother. 2012;56(3):1273-80. Epub 20111212. doi: 10.1128/AAC.05760-11. PubMed PMID: 22155830; PubMed Central PMCID: PMCPMC3294892.
112. Lahra MM, Ryder N, Whiley DM. A New Multidrug-Resistant Strain of *Neisseria gonorrhoeae* in Australia. New England Journal of Medicine. 2014;371(19):1850-1. doi: 10.1056/NEJMc1408109.
113. Cámara J, Serra J, Ayats J, Bastida T, Carnicer-Pont D, Andreu A, et al. Molecular characterization of two high-level ceftriaxone-resistant *Neisseria gonorrhoeae* isolates detected in Catalonia, Spain. Journal of Antimicrobial Chemotherapy. 2012;67(8):1858-60. doi: 10.1093/jac/dks162.
114. Edwards JL, Brown EJ, Ault KA, Apicella MA. The role of complement receptor 3 (CR3) in *Neisseria gonorrhoeae* infection of human cervical epithelia. Cell Microbiol. 2001;3(9):611-22. doi: 10.1046/j.1462-5822.2001.00140.x. PubMed PMID: 11553013.
115. Poole J, Day CJ, Haselhorst T, Jen FEC, Torres VJ, Edwards JL, et al. Repurposed drugs that block the gonococcus-complement receptor 3 interaction can prevent and cure gonococcal infection of primary human cervical epithelial cells. mBio. 2020;11(2). doi: 10.1128/mBio.03046-19.
116. Jonsson AB, Nyberg G, Normark S. Phase variation of gonococcal pili by frameshift mutation in pilC, a novel gene for pilus assembly. The EMBO Journal. 1991;10(2):477-88. doi: https://doi.org/10.1002/j.1460-2075.1991.tb07970.x.
117. Palmer GH, Bankhead T, Seifert HS. Antigenic Variation in Bacterial Pathogens. Microbiol Spectr. 2016;4(1). doi: 10.1128/microbiolspec.VMBF-0005-2015. PubMed PMID: 26999387; PubMed Central PMCID: PMCPMC4806564.

118. Vink C, Rudenko G, Seifert HS. Microbial antigenic variation mediated by homologous DNA recombination. FEMS Microbiol Rev. 2012;36(5):917-48. Epub 20120117. doi: 10.1111/j.1574-6976.2011.00321.x. PubMed PMID: 22212019; PubMed Central PMCID: PMCPMC3334452.
119. Maldarelli GA, Piepenbrink KH, Scott AJ, Freiberg JA, Song Y, Achermann Y, et al. Type IV pili promote early biofilm formation by *Clostridium difficile*. Pathog Dis. 2016;74(6). Epub 20160630. doi: 10.1093/femspd/ftw061. PubMed PMID: 27369898; PubMed Central PMCID: PMCPMC5985507.
120. Abt MC, McKenney PT, Pamer EG. *Clostridium difficile* colitis: pathogenesis and host defence. Nat Rev Microbiol. 2016;14(10):609-20. Epub 20160830. doi: 10.1038/nrmicro.2016.108. PubMed PMID: 27573580; PubMed Central PMCID: PMCPMC5109054.
121. Dilnessa T, Getaneh A, Hailu W, Moges F, Gelaw B. Prevalence and antimicrobial resistance pattern of *Clostridium difficile* among hospitalized diarrheal patients: A systematic review and meta-analysis. PLoS One. 2022;17(1):e0262597. Epub 20220113. doi: 10.1371/journal.pone.0262597. PubMed PMID: 35025959; PubMed Central PMCID: PMCPMC8758073.
122. Sholeh M, Krutova M, Forouzesh M, Mironov S, Sadeghifard N, Molaeipour L, et al. Antimicrobial resistance in *Clostridioides (Clostridium) difficile* derived from humans: a systematic review and meta-analysis. Antimicrob Resist Infect Control. 2020;9(1):158. Epub 20200925. doi: 10.1186/s13756-020-00815-5. PubMed PMID: 32977835; PubMed Central PMCID: PMCPMC7517813.
123. Wickramage I, Spigaglia P, Sun X. Mechanisms of antibiotic resistance of *Clostridioides difficile*. J Antimicrob Chemother. 2021;76(12):3077-90. doi: 10.1093/jac/dkab231. PubMed PMID: 34297842; PubMed Central PMCID: PMCPMC8598299.
124. Joyce SA, Shanahan F, Hill C, Gahan CG. Bacterial bile salt hydrolase in host metabolism: Potential for influencing gastrointestinal microbe-host crosstalk. Gut Microbes. 2014;5(5):669-74. doi: 10.4161/19490976.2014.969986. PubMed PMID: 25483337; PubMed Central PMCID: PMCPMC4615832.
125. Bordeleau E, Purcell EB, Lafontaine DA, Fortier LC, Tamayo R, Burrus V. Cyclic di-GMP riboswitch-regulated type IV pili contribute to aggregation of *Clostridium difficile*. J Bacteriol. 2015;197(5):819-32. Epub 20141215. doi: 10.1128/JB.02340-14. PubMed PMID: 25512308; PubMed Central PMCID: PMCPMC4325102.
126. Dawson LF, Peltier J, Hall CL, Harrison MA, Derakhshan M, Shaw HA, et al. Extracellular DNA, cell surface proteins and c-di-GMP promote biofilm formation in *Clostridioides difficile*. Sci Rep. 2021;11(1):3244. Epub 20210205. doi: 10.1038/s41598-020-78437-5. PubMed PMID: 33547340; PubMed Central PMCID: PMCPMC7865049.
127. Ronish LA, Sidner B, Yu Y, Piepenbrink KH. Recognition of extracellular DNA by type IV pili promotes biofilm formation by *Clostridioides difficile*. J Biol Chem. 2022;298(10):102449. Epub 20220903. doi: 10.1016/j.jbc.2022.102449. PubMed PMID: 36064001; PubMed Central PMCID: PMCPMC9556784.
128. Lawley TD, Clare S, Walker AW, Goulding D, Stabler RA, Croucher N, et al. Antibiotic treatment of *Clostridium difficile* carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts. Infect Immun. 2009;77(9):3661-9. Epub 20090629. doi: 10.1128/IAI.00558-09. PubMed PMID: 19564382; PubMed Central PMCID: PMCPMC2737984.

129. Soavelomandroso AP, Gaudin F, Hoys S, Nicolas V, Vedantam G, Janoir C, et al. Biofilm Structures in a Mono-Associated Mouse Model of *Clostridium difficile* Infection. Front Microbiol. 2017;8:2086. Epub 20171025. doi: 10.3389/fmicb.2017.02086. PubMed PMID: 29118745; PubMed Central PMCID: PMCPMC5661025.
130. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. Br J Pharmacol. 2011;162(6):1239-49. doi: 10.1111/j.1476-5381.2010.01127.x. PubMed PMID: 21091654; PubMed Central PMCID: PMCPMC3058157.
131. Thrithamarassery Gangadharan N, Venkatachalam AB, Sugathan S. High-Throughput and In Silico Screening in Drug Discovery. In: Abdulhameed S, Pradeep NS, Sugathan S, editors. Bioresources and Bioprocess in Biotechnology: Volume 1: Status and Strategies for Exploration. Singapore: Springer Singapore; 2017. p. 247-73.
132. Leelananda SP, Lindert S. Computational methods in drug discovery. Beilstein J Org Chem. 2016;12:2694-718. Epub 20161212. doi: 10.3762/bjoc.12.267. PubMed PMID: 28144341; PubMed Central PMCID: PMCPMC5238551.
133. Sinha S, Vohora D. Drug Discovery and Development. Pharmaceutical Medicine and Translational Clinical Research2018. p. 19-32.
134. Wildey MJ, Haunso A, Tudor M, Webb M, Connick JH. High-Throughput Screening. Platform Technologies in Drug Discovery and Validation. Annual Reports in Medicinal Chemistry2017. p. 149-95.
135. Henkel RD, VandeBerg JL, Walsh RA. A microassay for ATPase. Analytical Biochemistry. 1988;169(2):312-8. doi: https://doi.org/10.1016/0003-2697(88)90290-4.
136. Lionta E, Spyrou G, Vassilatis DK, Cournia Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. Curr Top Med Chem. 2014;14(16):1923-38. doi: 10.2174/1568026614666140929124445. PubMed PMID: 25262799; PubMed Central PMCID: PMCPMC4443793.
137. Amaro RE, Baudry J, Chodera J, Demir O, McCammon JA, Miao Y, et al. Ensemble Docking in Drug Discovery. Biophys J. 2018;114(10):2271-8. Epub 20180330. doi: 10.1016/j.bpj.2018.02.038. PubMed PMID: 29606412; PubMed Central PMCID: PMCPMC6129458.
138. Morris CJ, Corte DD. Using molecular docking and molecular dynamics to investigate protein-ligand interactions. Modern Physics Letters B. 2021;35(08). doi: 10.1142/s0217984921300027.
139. Ozcan A, Keskin O, Sariyar Akbulut B, Ozbek P. Piperidine-based natural products targeting Type IV pili antivirulence: A computational approach. J Mol Graph Model. 2023;119:108382. Epub 20221126. doi: 10.1016/j.jmgm.2022.108382. PubMed PMID: 36463631.
140. Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. Neuron. 2018;99(6):1129-43. doi: 10.1016/j.neuron.2018.08.011. PubMed PMID: 30236283; PubMed Central PMCID: PMCPMC6209097.
141. Braun E, Gilmer J, Mayes HB, Mobley DL, Monroe JI, Prasad S, et al. Best Practices for Foundations in Molecular Simulations [Article v1.0]. Living J Comput Mol Sci. 2019;1(1). Epub 20181129. doi: 10.33011/livecoms.1.1.5957. PubMed PMID: 31788666; PubMed Central PMCID: PMCPMC6884151.
142. Salsbury FR, Jr. Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. Curr Opin Pharmacol. 2010;10(6):738-44. doi: 10.1016/j.coph.2010.09.016. PubMed PMID: 20971684; PubMed Central PMCID: PMCPMC2981647.

143. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. Journal of Computational Chemistry. 1983;4(2):187-217. doi: 10.1002/jcc.540040211.
144. Grossfield A, Patrone PN, Roe DR, Schultz AJ, Siderius DW, Zuckerman DM. Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations [Article v1.0]. Living J Comput Mol Sci. 2018;1(1). Epub 20181027. doi: 10.33011/livecoms.1.1.5067. PubMed PMID: 30533602; PubMed Central PMCID: PMCPMC6286151.
145. Romo TD, Grossfield A. Block Covariance Overlap Method and Convergence in Molecular Dynamics Simulation. J Chem Theory Comput. 2011;7(8):2464-72. Epub 20110701. doi: 10.1021/ct2002754. PubMed PMID: 26606620.
146. Sabe VT, Ntombela T, Jhamba LA, Maguire GEM, Govender T, Naicker T, et al. Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. Eur J Med Chem. 2021;224:113705. Epub 20210715. doi: 10.1016/j.ejmech.2021.113705. PubMed PMID: 34303871.
147. Rastelli G. Emerging topics in structure-based virtual screening. Pharm Res. 2013;30(5):1458-63. Epub 20130307. doi: 10.1007/s11095-013-1012-9. PubMed PMID: 23468050.
148. Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. Med Res Rev. 2006;26(5):531-68. doi: 10.1002/med.20067. PubMed PMID: 16758486.
149. De Vivo M, Cavalli A. Recent advances in dynamic docking for drug discovery. WIREs Computational Molecular Science. 2017;7(6). doi: 10.1002/wcms.1320.
150. Gioia D, Bertazzo M, Recanatini M, Masetti M, Cavalli A. Dynamic Docking: A Paradigm Shift in Computational Drug Discovery. Molecules. 2017;22(11). Epub 20171122. doi: 10.3390/molecules22112029. PubMed PMID: 29165360; PubMed Central PMCID: PMCPMC6150405.
151. Jennings MP, Jen FEC, Roddam LF, Apicella MA, Edwards JL. *Neisseria gonorrhoeae* pilin glycan contributes to CR3 activation during challenge of primary cervical epithelial cells. Cellular Microbiology. 2011;13(6):885-96. doi: 10.1111/j.1462-5822.2011.01586.x.
152. Collins R, Karuppiah V, Siebert CA, Dajani R, Thistlethwaite A, Derrick JP. Structural cycle of the *Thermus thermophilus* PilF ATPase: the powering of type IVa pilus assembly. Sci Rep. 2018;8(1):14022. Epub 20180919. doi: 10.1038/s41598-018-32218-3. PubMed PMID: 30232337; PubMed Central PMCID: PMCPMC6145873.
153. Satyshur KA, Worzalla GA, Meyer LS, Heiniger EK, Aukema KG, Misic AM, et al. Crystal structures of the pilus retraction motor PilT suggest large domain movements and subunit cooperation drive motility. Structure. 2007;15(3):363-76. doi: 10.1016/j.str.2007.01.018. PubMed PMID: 17355871; PubMed Central PMCID: PMCPMC1978094.
154. Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgensen WL, Lins RD, et al. Developing a dynamic pharmacophore model for HIV-1 integrase. J Med Chem. 2000;43(11):2100-14. doi: 10.1021/jm990322h. PubMed PMID: 10841789.
155. Vijayakumar S, Biswas I, Veeraraghavan B. Accurate identification of clinically important *Acinetobacter* spp.: an update. Future Sci OA. 2019;5(6):FSO395. Epub 20190627. doi: 10.2144/fsoa-2018-0127. PubMed PMID: 31285840; PubMed Central PMCID: PMCPMC6609899.

156. Sukmana A, Yang Z. The type IV pilus assembly motor PilB is a robust hexameric ATPase with complex kinetics. Biochem J. 2018;475(11):1979-93. Epub 20180615. doi: 10.1042/BCJ20180167. PubMed PMID: 29717025.
157. Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. QMEANDisCo—distance constraints applied on model quality estimation. Bioinformatics. 2020;36(6):1765-71. doi: 10.1093/bioinformatics/btz828.
158. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research. 2018;46(W1):W296-W303. doi: 10.1093/nar/gky427.
159. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007;35(Web Server issue):W407-10. Epub 20070521. doi: 10.1093/nar/gkm290. PubMed PMID: 17517781; PubMed Central PMCID: PMCPMC1933241.
160. Sorokina M, Merseburger P, Rajan K, Yirik MA, Steinbeck C. COCONUT online: Collection of Open Natural Products database. J Cheminform. 2021;13(1):2. Epub 20210110. doi: 10.1186/s13321-020-00478-9. PubMed PMID: 33423696; PubMed Central PMCID: PMCPMC7798278.
161. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem. 2004;47(7):1739-49. doi: 10.1021/jm0306430. PubMed PMID: 15027865.
162. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov. 2015;10(5):449-61. Epub 20150402. doi: 10.1517/17460441.2015.1032936. PubMed PMID: 25835573; PubMed Central PMCID: PMCPMC4487606.
163. Deng Z, Chuaqui C, Singh J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. J Med Chem. 2004;47(2):337-44. doi: 10.1021/jm030331x. PubMed PMID: 14711306; PubMed Central PMCID: PMC14711306.
164. Singh J, Deng Z, Narale G, Chuaqui C. Structural interaction fingerprints: a new approach to organizing, mining, analyzing, and designing protein-small molecule complexes. Chem Biol Drug Des. 2006;67(1):5-12. doi: 10.1111/j.1747-0285.2005.00323.x. PubMed PMID: 16492144.
165. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1-2:19-25. doi: 10.1016/j.softx.2015.06.001.
166. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017;14(1):71-3. Epub 20161107. doi: 10.1038/nmeth.4067. PubMed PMID: 27819658; PubMed Central PMCID: PMCPMC5199616.
167. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics. 1983;79(2):926-35. doi: 10.1063/1.445869.
168. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics. 1984;81(8):3684-90. doi: 10.1063/1.448118.

169. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. The Journal of Chemical Physics. 1995;103(19):8577-93. doi: 10.1063/1.470117.
170. Hess B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. Journal of Chemical Theory and Computation. 2008;4(1):116-22. doi: 10.1021/ct700200b.
171. Halgren T. New method for fast and accurate binding-site identification and analysis. Chem Biol Drug Des. 2007;69(2):146-8. doi: 10.1111/j.1747-0285.2007.00483.x. PubMed PMID: 17381729.
172. Sun R, Sode O, Dama JF, Voth GA. Simulating Protein Mediated Hydrolysis of ATP and Other Nucleoside Triphosphates by Combining QM/MM Molecular Dynamics with Advances in Metadynamics. J Chem Theory Comput. 2017;13(5):2332-41. Epub 20170404. doi: 10.1021/acs.jctc.7b00077. PubMed PMID: 28345907; PubMed Central PMCID: PMCPMC5425946.
173. Ogura T, Whiteheart SW, Wilkinson AJ. Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases. J Struct Biol. 2004;146(1-2):106-12. doi: 10.1016/j.jsb.2003.11.008. PubMed PMID: 15095758.
174. Purser S, Moore PR, Swallow S, Gouverneur V. Fluorine in medicinal chemistry. Chem Soc Rev. 2008;37(2):320-30. Epub 20071213. doi: 10.1039/b610213c. PubMed PMID: 18197348.
175. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. 2001;46(1):3-26. doi: https://doi.org/10.1016/S0169-409X(00)00129-0.
176. Hotinger JA, Morris ST, May AE. The Case against Antibiotics and for Anti-Virulence Therapeutics. Microorganisms. 2021;9(10). Epub 20210928. doi: 10.3390/microorganisms9102049. PubMed PMID: 34683370; PubMed Central PMCID: PMCPMC8537500.
177. O'Toole GA, Kolter R. Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. Mol Microbiol. 1998;30(2):295-304. doi: 10.1046/j.1365-2958.1998.01062.x. PubMed PMID: 9791175.
178. Lu C, Turley S, Marionni ST, Park YJ, Lee KK, Patrick M, et al. Hexamers of the type II secretion ATPase GspE from *Vibrio cholerae* with increased ATPase activity. Structure. 2013;21(9):1707-17. Epub 20130815. doi: 10.1016/j.str.2013.06.027. PubMed PMID: 23954505; PubMed Central PMCID: PMCPMC3775503.
179. Studier FW. Stable Expression Clones and Auto-Induction for Protein Production in *E. coli*. In: Chen YW, editor. Structural Genomics: General Applications. Totowa, NJ: Humana Press; 2014. p. 17-32.
180. Sebaugh JL. Guidelines for accurate EC50/IC50 estimation. Pharm Stat. 2011;10(2):128-34. doi: 10.1002/pst.426. PubMed PMID: 22328315.
181. Hockenberry AM, Hutchens DM, Agellon A, So M. Attenuation of the Type IV Pilus Retraction Motor Influences *Neisseria gonorrhoeae* Social and Infection Behavior. mBio. 2016;7(6). Epub 20161206. doi: 10.1128/mBio.01994-16. PubMed PMID: 27923924; PubMed Central PMCID: PMCPMC5142622.
182. Li Y, Huang S, Zhang X, Huang T, Li H. Cloning, expression, and functional analysis of molecular motor pilT and pilU genes of type IV pili in Acidithiobacillus ferrooxidans. Appl

Microbiol Biotechnol. 2013;97(3):1251-7. Epub 20120718. doi: 10.1007/s00253-012-4271-1. PubMed PMID: 22805785.
183. Schiraldi C, De Rosa M. Mesophilic Organisms. Encyclopedia of Membranes2014. p. 1-2.
184. Mempin R, Tran H, Chen C, Gong H, Kim Ho K, Lu S. Release of extracellular ATP by bacteria during growth. BMC Microbiology. 2013;13(1):301. doi: 10.1186/1471-2180-13-301.
185. Arya R, Sabir JSM, Bora RS, Saini KS. Optimization of Culture Parameters and Novel Strategies to Improve Protein Solubility. In: García-Fruitós E, editor. Insoluble Proteins: Methods and Protocols. New York, NY: Springer New York; 2015. p. 45-63.
186. Studier FW. Stable Expression Clones and Auto-Induction for Protein Production in E. coli. In: Chen Y, editor. Structural Genomics. Methods in Molecular Biology. 1091. Totowa, NJ: Humana Press; 2014. p. 17–32.
187. Studier FW. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif. 2005;41(1):207-34. doi: 10.1016/j.pep.2005.01.016. PubMed PMID: 15915565.
188. Bischof LF, Friedrich C, Harms A, Sogaard-Andersen L, van der Does C. The Type IV Pilus Assembly ATPase PilB of *Myxococcus xanthus* Interacts with the Inner Membrane Platform Protein PilC and the Nucleotide-binding Protein PilM. J Biol Chem. 2016;291(13):6946-57. Epub 20160205. doi: 10.1074/jbc.M115.701284. PubMed PMID: 26851283; PubMed Central PMCID: PMCPMC4807279.
189. Tank M, Bryant DA. Nutrient requirements and growth physiology of the photoheterotrophic Acidobacterium, *Chloracidobacterium thermophilum*. Front Microbiol. 2015;6:226. Epub 20150327. doi: 10.3389/fmicb.2015.00226. PubMed PMID: 25870589; PubMed Central PMCID: PMCPMC4376005.
190. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol. 2014;5:172. Epub 20140417. doi: 10.3389/fmicb.2014.00172. PubMed PMID: 24860555; PubMed Central PMCID: PMCPMC4029002.
191. Studier FW, Daegelen P, Lenski RE, Maslov S, Kim JF. Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes. J Mol Biol. 2009;394(4):653-80. Epub 20090915. doi: 10.1016/j.jmb.2009.09.021. PubMed PMID: 19765592.
192. Xu J, Banerjee A, Pan SH, Li ZJ. Galactose can be an inducer for production of therapeutic proteins by auto-induction using *E. coli* BL21 strains. Protein Expr Purif. 2012;83(1):30-6. Epub 20120308. doi: 10.1016/j.pep.2012.02.014. PubMed PMID: 22425658.
193. Shiue SJ, Kao KM, Leu WM, Chen LY, Chan NL, Hu NT. XpsE oligomerization triggered by ATP binding, not hydrolysis, leads to its association with XpsL. EMBO J. 2006;25(7):1426-35. Epub 20060309. doi: 10.1038/sj.emboj.7601036. PubMed PMID: 16525507; PubMed Central PMCID: PMCPMC1440322.