Molecular Insights into Microbial Adhesion

Roger Davies Klein
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Atomic, Molecular and Optical Physics Commons, Biochemistry Commons, and the Microbiology Commons

Recommended Citation
https://openscholarship.wustl.edu/art_sci_etds/2208

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Molecular Insights into Microbial Adhesion
by
Roger Klein

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2020
St. Louis, Missouri
# Table of Contents

List of Figures .......................................................................................................................... vi
List of Tables ............................................................................................................................. viii
Acknowledgments .................................................................................................................... ix
Abstract ....................................................................................................................................... xii

Chapter 1: Introduction .............................................................................................................. 1
  1.1 Scope of Work and Attribution ......................................................................................... 1
  1.2 The Emerging Threat of Antibiotic Resistance ............................................................. 1
  1.3 Determinants of Bacterial Virulence ............................................................................... 3
  1.4 Chaperone-Usher Pathway Pili ..................................................................................... 4
     1.4.1 CUP Operon Structure and Classification ............................................................. 4
     1.4.2 CUP Pilus Biogenesis ......................................................................................... 6
     1.4.3 CUP Adhesins as Antivirulence Targets ............................................................... 8
  1.5 Curli Fibers ...................................................................................................................... 8
     1.5.1 The Role of Biofilms in Bacterial Adhesion .......................................................... 8
     1.5.2 Curli Fiber Structure .......................................................................................... 10
     1.5.3 Curli Fiber Assembly ......................................................................................... 11
  1.6 Figures ........................................................................................................................... 13

Chapter 2: Structure-Function Analysis of the Curli Accessory Protein CsgE .................. 15
  2.1 Scope of Work .................................................................................................................. 15
  2.2 Introduction ..................................................................................................................... 15
     2.2.1 CsgE Function .................................................................................................... 15
     2.2.2 CsgE Structure .................................................................................................. 16
  2.3 Structure-Function Analysis of CsgE ........................................................................... 18
     2.3.1 Functional Validation of Charged Residues by Site-Directed Mutagenesis .......... 18
     2.3.2 The CsgE-CsgA Interaction is Necessary for \emph{in vivo} Curli Fiber Formation ....... 19
     2.3.3 The CsgE-CsgG Interaction is Robust ................................................................. 20
     2.3.4 CsgE Oligomerization is Time-and Temperature-Dependent ............................ 21
     2.3.5 Key Residues for CsgE Function are Subject to Evolutionary Selection .......... 22
  2.4 Discussion ....................................................................................................................... 24
2.5 Future Directions.................................................................................................................26
  2.5.1 CsgE Oligomerization.................................................................................................26
  2.5.2 The CsgE-CsgG Interaction.......................................................................................29
  2.5.3 Structural Studies of CsgF.........................................................................................31
  2.5.4 Localization of Csg Proteins ....................................................................................33
  2.5.5 Characterizing the Inhibition of α-Synuclein by CsgE ..............................................34
2.6 Materials and Methods.......................................................................................................35
  2.6.1 Strains and Plasmids..................................................................................................35
  2.6.2 NMR Spectroscopy....................................................................................................36
  2.6.3 Thioflavin T Assays....................................................................................................36
  2.6.4 Quantification of CsgA in Curli-Expressing Cells ......................................................37
  2.6.5 Differential Scanning Fluorimetry.............................................................................38
  2.6.6 Bile Salt Sensitivity Assays........................................................................................38
  2.6.7 Native-PAGE and Size Exclusion Chromatography ................................................38
  2.6.8 SEC-MALS...............................................................................................................39
  2.6.9 Multisequence Alignment and Selective Pressure Analysis ......................................39
  2.6.10 Dynamic Light Scattering.......................................................................................40
2.7 Figures.................................................................................................................................41

Chapter 3: Optimizing Glycomimetic Antiadhesives for Treatment of UTIs .......................58
3.1 Scope of Work......................................................................................................................58
3.2 Introduction.........................................................................................................................58
  3.2.1 Clinical Characteristics of Urinary Tract Infections..................................................58
  3.2.2 Antibiotic Resistance and UTIs.................................................................................60
  3.2.3 Determinants of UTI Pathogenesis .........................................................................62
  3.2.4 The Structural Basis for Mannose Recognition by FimH .........................................64
  3.2.5 The Structural Basis for Galactose Recognition by FmlH ........................................65
3.3 Elucidating the Structure-Activity Relationship of Novel Glycomimetics.......................67
  3.3.1 X-ray Crystal Structure of ZFH269 in Complex with FimHLD ..................................67
  3.3.2 X-ray Crystal Structure Three Biphenyl Galactosides in Complex with FmlHLD ....68
3.4 Discussion and Future Directions .....................................................................................70
3.5 Materials and Methods.......................................................................................................73
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5.1</td>
<td>Protein Expression and Purification</td>
<td>73</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Enzyme-Linked Immunosorbent Assays</td>
<td>73</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Structure Determination</td>
<td>74</td>
</tr>
<tr>
<td>3.6</td>
<td>Figures</td>
<td>76</td>
</tr>
<tr>
<td>4.1</td>
<td>Scope of Work</td>
<td>85</td>
</tr>
<tr>
<td>4.2</td>
<td>Introduction</td>
<td>85</td>
</tr>
<tr>
<td>4.3</td>
<td>Structural and Functional Analysis</td>
<td>88</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Determination of the UclD&lt;sup&gt;LD&lt;/sup&gt; Structure</td>
<td>88</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Characterization of the Putative UclD&lt;sup&gt;LD&lt;/sup&gt; Binding Pocket</td>
<td>89</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Identification of a Molecular Ligand</td>
<td>90</td>
</tr>
<tr>
<td>4.4</td>
<td>Discussion and Future Directions</td>
<td>90</td>
</tr>
<tr>
<td>4.5</td>
<td>Materials and Methods</td>
<td>94</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Protein Purification</td>
<td>94</td>
</tr>
<tr>
<td>4.5.2</td>
<td>Crystallization and Structure</td>
<td>95</td>
</tr>
<tr>
<td>4.5.3</td>
<td>FimH and UclD Binding Studies</td>
<td>96</td>
</tr>
<tr>
<td>4.5.4</td>
<td>Differential Scanning Fluorimetry</td>
<td>97</td>
</tr>
<tr>
<td>4.5.5</td>
<td>Glycan Array Screening</td>
<td>97</td>
</tr>
<tr>
<td>4.6</td>
<td>Figures</td>
<td>98</td>
</tr>
<tr>
<td>5.1</td>
<td>Scope of Work</td>
<td>106</td>
</tr>
<tr>
<td>5.2</td>
<td>Introduction</td>
<td>106</td>
</tr>
<tr>
<td>5.3</td>
<td>Results</td>
<td>107</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Yeh and Yhl Pili are Common, Related, and Distinct</td>
<td>107</td>
</tr>
<tr>
<td>5.3.2</td>
<td>The YehD Lectin Domain Contains a Mobile Helix-Loop-Helix Domain</td>
<td>108</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Evolutionary Adaption of the Pilus Rod</td>
<td>109</td>
</tr>
<tr>
<td>5.3.4</td>
<td><em>Ex vivo and in vitro</em> YehD&lt;sup&gt;LD&lt;/sup&gt; and YhlD&lt;sup&gt;LD&lt;/sup&gt; Binding Studies</td>
<td>110</td>
</tr>
<tr>
<td>5.3.5</td>
<td>Mouse Studies</td>
<td>111</td>
</tr>
<tr>
<td>5.4</td>
<td>Discussion and Future Directions</td>
<td>111</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Exploration of the YehD&lt;sup&gt;LD&lt;/sup&gt; and YhlD&lt;sup&gt;LD&lt;/sup&gt; Structures</td>
<td>112</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Structural and Biomechanical Studies of the Yeh Rod</td>
<td>113</td>
</tr>
</tbody>
</table>
5.4.3 The YehB Chaperone and Donor Strand Complementation .................................................. 115
5.4.4 Structural Characterization of YhIDLD ............................................................................. 116
5.4.5 YehC Usher Activation ...................................................................................................... 117
5.4.6 Yeh Pili as Therapeutic Targets ...................................................................................... 118

5.5 Materials and Methods ........................................................................................................ 119
5.5.1 Evolutionary Adaption of the Pilus Rod ........................................................................... 119
5.5.2 Protein Purification ............................................................................................................ 119
5.5.3 Pilus Purification and Imaging .......................................................................................... 120
5.5.4 YehD Crystallography ....................................................................................................... 120
5.5.5 YehA Conservation Analysis ............................................................................................ 120
5.5.6 Glycan Array Screening .................................................................................................... 121

5.6 Figures and Tables .................................................................................................................. 122

Chapter 6: Conclusions ............................................................................................................ 134

Works Cited .................................................................................................................................. 137
List of Figures

Figure 1.1 Overview of CUP pilus structure and biogenesis ................................................................. 13
Figure 1.2. Working model for curli biogenesis ....................................................................................... 14
Figure 2.1. Charged residues on CsgE cluster into three groups ............................................................... 41
Figure 2.2. Chemical shift perturbations of 15N-labeled CsgE W48A/F79A by CsgA .................. 42
Figure 2.3. The presence of CsgA has little effect on the backbone amide residues of CsgE ........ 43
Figure 2.4. HSQC spectra of WT CsgE in the presence and absence of CsgA ...................... 44
Figure 2.5. Mutation of charged residues differentially affects the CsgE-CsgA interaction ...... 45
Figure 2.6. Thermal stability of select CsgE mutants ......................................................................... 47
Figure 2.7. Quantification of curli fiber formation on the bacterial surface ........................................ 48
Figure 2.8. Introduction of select point mutations has minimal effects on CsgE stability in vivo 49
Figure 2.9. Mutation of tail residues does not abrogate the CsgE-CsgG interaction................ 50
Figure 2.10. Size exclusion chromatograms of CsgE oligomerization ................................................. 51
Figure 2.11. CsgE W48A/F79A exists as a monodisperse monomer ............................................... 52
Figure 2.12. Mutations to aromatic head residues alter CsgE oligomerization kinetics.......... 52
Figure 2.13. Residues throughout the CsgE molecule are subject to selective pressures ..... 54
Figure 2.14. Dynamic light scattering of WT CsgE ........................................................................ 55
Figure 2.15. UV tracings from size-exclusion elution profiles ........................................................... 56
Figure 2.16. CsgE point mutants differentially inhibit α-synuclein polymerization ...................... 57
Figure 3.1. Overview of the Type 1 pilus structure .............................................................................. 76
Figure 3.2. Comparison of the FimH and FmlH lectin domains from E. coli UTI89 ...................... 77
Figure 3.3. X-ray crystal structures of ZFH269 and ZFH284 bound to FimHLD ............................. 78
Figure 3.4. Crystal structures of high-affinity galactosides in complex with FmlHLD ................. 79
Figure 3.5. Pharmacokinetic properties of AM2134 and AM2151 ...................................................... 80
Figure 3.6. C-linkages and acetyl groups improve the pharmacokinetics of ZFH269 ............... 81
Figure 4.1. Phylogenetic analysis reveals close homologues to the F17-like pilus ....................... 98
Figure 4.2. Type 1 and F17-like pili facilitate colonization of the murine gastrointestinal tract ... 99
Figure 4.3. Structural comparison of UclDLD and F17GLD ............................................................... 100
Figure 4.4. Comparison of residues in the putative binding site of UclDLD ................................. 101
Figure 4.5. UclDLD does not bind common mammalian glycan epitopes ........................................ 102
Figure 4.6. UclD does not bind to Caco-2 cells or common monosaccharides ............................. 103
Figure 4.7. Neither guanosine nor N1-methylguanosine affect the $T_m$ of UclD$^{LD}$ and FimH$^{LD}$ .. 104
Figure 5.1. Content, Conservation, and Phylogeny of the Yeh Operon .................................. 122
Figure 5.2. X-ray crystal structure of the YehD lectin domain ................................................. 123
Figure 5.3. Homology Models of the Yeh Pili Rod and DSE Reaction ...................................... 124
Figure 5.4 Conservation and selection among YehA Residues ............................................. 125
Figure 5.5. YehD$^{LD}$ and YhlD$^{LD}$ binds murine colonic tissue $ex vivo$ .................................. 126
Figure 5.6. YehD$^{LD}$ and YhlD$^{LD}$ bind colon-derived epithelial cells ..................................... 127
Figure 5.7. Deletion of $yeh$ and $yhl$ in murine models of gastrointestinal colonization .......... 128
Figure 5.8. Glycan array screening of YehD$^{LD}$ ...................................................................... 129
Figure 5.9. Metastructural analysis of Yeh pili ......................................................................... 130
Figure 5.10. Biomechanical properties of the Yeh pili ............................................................. 131
Figure 5.11. Homology modelling reveals a probable mechanism for DSC .............................. 132
List of Tables

Table 3.1. Structures and IC50s of Select Biphenyl Galactosides ..............................................82
Table 3.2. Data collection and refinement statistics for FimH^{LD}-ZFH269 ........................................83
Table 3.3. X-Ray data collection and refinement statistics for FmlH^{LD}-galactoside complexes ..84
Table 4.1. Collection and refinement statistics for UclD^{LD} ..........................................................105
Table 5.1. Collection and refinement statistics for YehD^{LD} ..........................................................133
Acknowledgments

The work presented herein was made possible by years of inspiration and support from a multitude of sources. Through decades of hard work, passion, and an unwavering commitment to collaboration, Scott Hultgren has built an amazing research program that will profoundly improve the health and well-being of millions of women worldwide. By providing me with opportunities to engage in meaningful research in a rigorous and interdisciplinary setting, Scott's support and mentorship has facilitated my development into a thoughtful and passionate scientist. I've also been fortunate to work with another outstanding mentor, Matt Chapman, during the course of my graduate training. I am very thankful for the warm welcome I would always receive from Matt and his lab during my visits to Ann Arbor. Matt’s dedication to his science, his collaborators, and his trainees exemplifies what it means to be a scientific mentor and role model.

I am also very thankful for the members of the Hultgren lab, past and present, who have helped me along my scientific journey. I’d especially like to acknowledge Karen Dodson for reviewing countless drafts, experiments, and ideas, and Jerry Pinkner for supplying me with protein, compounds, and the occasional reality check. I’d also like to thank fellow Benjamin Bay members Vasilios Kalas, Caitlin Spaulding, and Natalie Omattage for all the fun times, productive conversations, and questionable Google searches we shared over the years. They’ve been instrumental to my happiness in lab and development as a scientist. I’m also very grateful to Zac Cusumano, Matt Conover, Henry Schreiber, and Allie Paharik for providing me with instruction and data for a variety of projects, and Kevin Tamadonfar, whose generosity and humor have made the home stretch of my graduate career much more enjoyable.

I would also like to thank all the members of my thesis committee, past and present, who provided valuable feedback throughout my project. Daved Fremont, Carl Frieden and Jim Janetka
were each instrumental to the success of F17-like, curli, and glycomimetic projects, respectively. I would also like to thank all the past teachers, advisors, and mentors whose influence over my professional life cannot be understated. I’m particularly thankful for James Adams and Matt McGuire, who sparked my interest in science in high school, and Virginia Pett and Mark Snider, who introduced me to the core principles of scientific research during my undergraduate training.

I would like to thank anyone who has ever advocated for public support of the sciences; your efforts made this work possible. I would like to acknowledge NIH R01 grants AI099099, AI048689, DK108840, and MSTP Training Grant T32GM07200. I would also like to thank the Department of Molecular Microbiology, the Program in Computational and Molecular Biophysics, the Schlesinger and Olivo Student Travel Award, and the Sondra Schlesinger Graduate Student Fellowship for their support.

Finally, I would like to express my gratitude for the love and support of my family, without whom I would not be where I am today. My parents, Alicia and Martin Klein, have been role models for hard work and generosity every day for the past 29 years. By exposing me to new places, people, and ideas, they continue to kindle my lifelong love of learning and exploration. My sister, Charlotte, is a constant source of inspiration in my life for her intelligence, wit, and boundless compassion. And finally, I would like to thank my wife, Emma DeLooze-Klein, for all of the wonderful experiences we have shared together over the last eleven years. Words cannot begin to describe how thankful I am to come home to such a talented, engaging, and good-natured woman every day. There is no-one with whom I’d rather face life’s great challenges and celebrate life’s greatest achievements.

Roger Klein

*Washington University in St. Louis*

*May 2020*
Dedicated to Charlotte and Emma.
ABSTRACT OF THE DISSERTATION
Molecular Insights into Microbial Adhesion
by
Roger Klein
Doctor of Philosophy in Biology and Biomedical Sciences
Computation and Molecular Biophysics
Washington University in St. Louis, 2020
Professor Scott J. Hultgren, Chair

Antibiotic-resistant bacterial infections are a serious and immediate threat to global public health. In the United States alone, over 2 million individuals develop antibiotic-resistant infections annually, resulting in 23,000 deaths and $20 billion in excess health care costs. Virulence factors that allow bacteria to invade and persist within the host are promising targets for novel antimicrobial agents that could be used to curb the spread of antibiotic resistance. Development of therapeutics that can selectively eliminate pathogenic bacteria while sparing the beneficial host microbiota requires a detailed molecular understanding of critical virulence factors that facilitate interactions between pathogens and their environments. Two such factors are Chaperone-Usher Pathway (CUP) pili, which mediate bacterial adhesion, and curli, which constitute the major extracellular matrix component of many biofilms. In this dissertation, I have elucidated the structure of two pilus adhesins thought to mediate colonization of the gastrointestinal tract and determined the mechanism of CsgE assembly factor function in curli biogenesis.

CUP pili comprise hundreds of repeating structural subunits capped by an adhesin whose N-terminal lectin domain binds host receptors with stereochemical specificity, thus facilitating the colonization of specific host niches by Gram-negative bacteria. Using X-ray crystallography, I determined that the UclD adhesin on the F17-like pilus contains a transverse binding groove that
binds O-glycans in the colonic crypts. Conversely, I found that the YehD adhesin contains a unique helix-loop-helix motif that can pivot open to reveal a deep hydrophobic pocket. Additional structural and genetic studies of the Yeh pilus rod reveal a tightly-wound right-handed helix that has maintained its biomechanical response to external force through purifying evolutionary selection. I have also utilized X-ray crystallography to elucidate the structure-activity relationship of high-affinity compounds that inhibit the binding of two additional pilus adhesins, FimH and FmlH, to the urinary tract during acute and chronic urinary tract infection. Finally, I have performed genetic and biochemical studies to elucidate the role of CsgE in bacterial biofilm formation. During curli biogenesis, I found that CsgE uses a series of specific and nonspecific electrostatic interactions to bind the major amyloid subunit CsgA and deliver it to the CsgG outer membrane pore, thus preventing premature aggregation and cellular toxicity. In aggregate, these findings have advanced our molecular understanding of bacterial adhesion during pathogenesis and illuminated novel pathways for the treatment of antibiotic-resistant infections.
Chapter 1: Introduction

1.1 Scope of Work and Attribution

Portions of this chapter are reproduced from:

© 2015, Elsevier Inc.

1.2 The Emerging Threat of Antibiotic Resistance

Bacteria constitute a diverse domain of unicellular prokaryotic organisms whose presence is nearly ubiquitous in the environment and within eukaryotic hosts. Although they are best known for their ability to cause a diverse array of symptomatic infections in humans, the vast majority of human-associated bacteria asymptotically colonize a myriad of niches within the host. The relationship between these microbial communities and their host is typically commensalistic or even mutualistic (1). For example, bacteria within the gastrointestinal microbiota can liberate macro- and micronutrients from the host diet while modulating the host immune environment (2). Commensal bacteria are known to occupy several other niches on and within the human body, including the skin, respiratory tract, and genitourinary tract (3). In all these habitats, interbacterial competition between commensal and pathogenic bacteria can protect the host against colonization with pathogenic organisms in a process known as colonization resistance.

While the vast majority do not cause clinically-relevant human infections, bacteria can cause infections in a variety of organ systems and tissue types. While the habitats of some bacteria are restricted to specific niches within the host, other bacteria, such as Escherichia coli
(E. coli), can colonize and cause infection in a variety of sites, including the gastrointestinal tract, urinary tract, respiratory system, skin, and soft tissue (4). These infections contribute significantly to global morbidity and mortality, resulting in millions of deaths and tens of billions of dollars in direct and indirect healthcare costs annually (5). The use of broad-spectrum antimicrobial agents beginning in the 1940s led to a revolutionary paradigm shift in the treatment of bacterial infections. Since the discovery of penicillin, a multitude of antibiotic classes have been developed that target conserved vital functions within bacteria, including cell wall synthesis, protein synthesis, and bacterial metabolism (6).

Treatment with broad-spectrum antimicrobial agents, while effective, is not without side effects. Because most common antimicrobial agents possess activity against a variety of pathogenic and commensal bacterial species, antibiotic treatment can disrupt the beneficial host microbiota to the detriment of the host (7). Exposure to antimicrobial agents has been associated with an increased risk for a variety of negative sequelae, including an increased risk of diarrheal illness, vaginal yeast infections, urinary tract infections (UTIs), and certain metabolic disorders (8). Additionally, the widespread use, and often misuse, of antibiotics around the world has led to the rise of antibiotic resistant bacterial strains that are recalcitrant to previous standards of therapy (6). Mechanisms of resistance include augmented antibiotic efflux, modification of the antibiotic target through adaptive selection, and acquisition of antibiotic-degrading enzymes (9). These resistance mechanisms can be developed de novo or acquired through horizontal gene transfer between strains and species. The global rise in infections attributable to multidrug resist bacteria represents a serious and immediate public health crisis. Globally, 700,000 people die each year from infections caused by antibiotic resistant organisms, and by 2050, these infections could kill as many as 10 million people annually with an economic burden of approximately $100 trillion (10).
As we become increasingly cognizant of the threat posed by antibiotic resistance and the negative effects of antibiotic therapy on host health, the paradigm for the development of novel antimicrobial agents is shifting away from targeting core bacterial processes and towards targeting key determinants of virulence. Agents targeting these virulence factors could selectively eliminate pathogenic bacteria from the site of infection without affecting the structure of the beneficial native microbiota. By developing a detailed molecular understanding of virulence factor function, antimicrobial vaccines and small molecules can be rationally designed to accomplish these goals.

1.3 Determinants of Bacterial Virulence

Bacteria are versatile organisms that rely on a multitude of colonization factors to facilitate their growth and persistence in hosts and the environment. The virulence potential of any given bacterial strain is determined by the complex interactions that occur between bacteria and the host niches in which they reside. While the variation between and within host factors that contribute to this interaction makes broad characterization of a given strain’s pathogenic potential is extremely challenging, the role of many individual bacterial virulence factors during different stages of infection is well-known (11). Well-characterized virulence factors include nutrient sequestration systems, interbacterial defense systems, secreted cytotoxins, and motility factors (12-14). Bacteria also employ a variety of extracellular defense systems to protect against environmental and immune stressors. For example, Gram-negative bacteria incorporate lipopolysaccharides into their outer membrane to augment their membrane integrity to protect against chemical stressors (15). Both Gram-positive and Gram-negative bacteria express capsule, a polysaccharide matrix that protects the bacteria within from phagocytosis and recognition by the innate and adaptive immune systems (16, 17).
Another important virulence factor for many bacterial infections are structures that mediate colonization by promoting bacterial adhesion to host surfaces. To persist within a host, bacteria must overcome a variety of innate mechanical and immunological clearance mechanisms. For example, bacteria residing in the urinary tract are exposed to varying degrees of shear force arising from the accumulation and expulsion of urine (18). Similarly, bacteria in the gastrointestinal tract must withstand constant contractile events intended to promote the movement and expulsion of intestinal contents (19). To this end, many bacteria express proteinaceous extracellular fibers that mediate adhesion and promote persistence. This thesis details my studies of two proteinaceous extracellular fibers that participate in Gram-negative bacterial adhesion. The first, chaperone-usher pathway (CUP) pili, are fibrous extracellular appendages that mediate bacterial adhesion and biofilm formation by binding to specific host and environmental receptors (20). The second, curli, are assembled as the primary structural component of many bacterial biofilms that encapsulate bacteria and protect them from environmental and immune stressors, such as phagocytosis by macrophages and neutrophils (21). The structure, assembly, and function of these fiber types are discussed in more detail below.

1.4 Chaperone-Usher Pathway Pili

1.4.1 CUP Operon Structure and Classification

To facilitate colonization of environmental and host niches, Gram-negative bacteria utilize proteinaceous extracellular appendages known as chaperone-usher pathway (CUP) pili to adhere to biotic and abiotic surfaces and facilitate the formation of protective extracellular biofilms. *E. coli* encode 38 distinct CUP pilus operons in their pangenome, many of which are known to play a crucial role in a variety of clinically important infections (20). The specificity of a pilus for its
cognate receptor is conferred by a distal tip adhesin that binds its molecular ligand with stereochemical specificity (22). *E. coli* isolates encode an average of 12 CUP pilus operons, reflecting their ability to colonize multiple niches within the host (20). For example, the type 1 (fim) pilus mediates binding to mannosylated glycoproteins in the bladder during urinary tract infection (UTI) pathogenesis in uropathogenic *E. coli* (UPEC) isolates (23). Similarly, F17 and CS20 pili mediate adhesion to the mammalian gastrointestinal tract, facilitating host colonization by enterotoxigenic *E. coli* (ETEC) isolates (19, 24). Genomic analysis of the so-called *E. coli* 'piliome' reveals that the carriage rates of individual pilus types vary significantly (20). Some pilus types, including type 1, are found in greater than 75% of strains. Others, such as P pili, are found in fewer than 30% of strains. By correlating CUP pilus carriage with specific bacterial pathotypes, it is sometimes possible to infer a function for those pili during *E. coli* pathogenesis. For example, carriage of the P pilus is primarily restricted to extraintestinal pathogenic *E. coli* (ExPEC) isolates and correlates well with its established role in mediating adhesion to the kidney epithelium during pyelonephritis (25).

CUP pili are composed of hundreds of repeating pilin subunits that are often arranged into a tightly-wound right-handed helical rod anchored to an outer membrane secretion channel known as the 'usher' (Figure 1.1) (18, 26). Each structural subunit within the pilus adopts a β-rich immunoglobulin (Ig)-like fold. To be considered a full, intact pilus operon, strains must encode the four minimum genes required for pilus structure and assembly: i) a major structural subunit, comprising the majority of the CUP rod; ii) an adhesin, typically two-domain; iii) a periplasmic chaperone and iv) an outer-membrane usher that coordinates subunit translocation and assembly. In addition, many CUP pilus operons encode between one and three distinct adaptor subunits that are incorporated between the helical pilus rod and the distal adhesin (26). The stoichiometric ratio of these adaptor proteins varies by pilus system. Because it encodes multiple
domains whose functions must remain intact for pilus assembly to occur, genes encoding the usher contain the highest degree of sequence conservation across CUP pilus operons (20). When a phylogenetic tree is constructed from all 38 CUP pilus ushers within the *E. coli* pangenome, six distinct evolutionary clades (α, β, κ, π, γ, σ) are observed.

### 1.4.2 CUP Pilus Biogenesis

CUP pilus assembly occurs via a tightly coordinated and highly conserved process. Structural subunits of the pilus fiber containing an N-terminal signal sequence are transported to the bacterial periplasm by the SecYEG apparatus (Figure 1.1). Following translocation, CUP subunits are recognized by a dedicated two-domain periplasmic chaperone (26). The primary sequences of all structural subunits and the C-terminal pilin domain of the tip adhesin encode a canonical Ig-like fold lacking its C-terminal β-strand, also known as the G strand. Following translocation, hydrophobic residues on the G1 β-strand of the pilus-specific periplasmic chaperone transiently occupy a series of hydrophobic pockets within the structural subunits’ incomplete immunoglobulin-like fold. This process, termed donor strand complementation (DSC), results in an atypical Ig-like fold in which the G1 donor strand runs parallel to the F strand of the structural subunit, rather than the more common antiparallel configuration (27). Failure of this process to occur results in misfolded periplasmic CUP subunits that are targeted for degradation by DegP. Following successful DSC, the chaperone-adhesin complex is recruited to the outer membrane usher by the usher’s periplasmic N-terminal domain. The usher catalyzes a second process known as donor strand exchange (DSE), wherein the free N-terminal extension of the incoming subunit replaces the chaperone’s G1 β-strand to permanently link the two subunits. This process, which occurs via a zip-in, zip-out mechanism, results in a more typical parallel arrangement between the F stand and NTE (28). The DSE-linked subunits are then translocated.
through the usher to the extracellular space, where the terminal adhesin can recognize ligands and mediate adhesion.

Because interactions formed between neighboring pilin subunits within the helical rod dictates the strength of this unwinding force, individual pili have adapted their biomechanical properties to optimize adhesion within different niches (19). For example, P pili require a greater force to unwind to prevent rapid elongation during constant urine flow in the kidney, while the CS20 fimbriae that mediate intestinal attachment in ETEC strains are more easily unwound to accommodate the short, strong forces bacteria experience in the intestines (19). Structural and biophysical studies of the Type 1 pilus rod reveal that residues responsible for the interaction between adjacent pilus subunits are subject to purifying selection, and that mutation of these residues leads to significant functional attenuation during in vivo infection in a mouse model of bladder infection (cystitis) (18). Conversely, a high rate of polymorphism among outward-facing residues reflects high rates of positive, diversifying selection, to promote antigenic variation and immune evasion (18).

In most cases, it is the N-terminal lectin domain of the pilus adhesin that confers the binding specificity of a particular pilus by recognizing the cognate molecular receptors with stereochemical specificity (26). X-ray crystallographic studies have provided tremendous insight into the structural basis of receptor recognition by providing molecular snapshots of pilus adhesion. However, recent studies have demonstrated that pili are dynamic structures. Recent genetic and biophysical analyses of the type 1 pilus adhesin, FimH, have identified pathoadaptive alleles that modulate ligand binding activity by altering the conformational flexibility of the adhesin (29-31). Alterations to these conformational dynamics affect the ability of uropathogenic E. Coli (UPEC) to adhere to the bladder epithelium, form IBCs, and sustain infection in a competitive
mouse model of chronic cystitis (29). Studies are currently underway to understand the generalizability of these findings to other pilus systems.

1.4.3 CUP Adhesins as Antivirulence Targets

Vaccination with pilus components has proven to be effective in the treatment of UTIs and other common infections. Specifically, vaccination with the chaperone-adhesin complex of both type 1 and fml pili have been shown to decrease the bladder burdens of UPEC in murine models of acute and chronic urinary tract infections (UTIs), respectively (32, 33). More recently, phase 1 clinical trials in human subjects showed a drastic reduction in the rate of recurrent UTI following vaccination with the FimCH chaperone-adhesin complex, and have been allowed for compassionate use in women with recurrent UTI recalcitrant to traditional antimicrobial therapy (Sequoia Biosciences, unpublished). Similarly, small molecules targeting CUP pili have shown great promise as antibiotic-sparing therapeutics that selectively remove pathogens from the niche in which they cause infection while sparing the native structure of the host microbiota (34, 35).

These compounds are high-affinity derivatives of the natural CUP pilus ligand, and competitively inhibit binding during infection. While these compounds have shown great promise in the prevention and treatment of infection in murine models of cystitis, a detailed molecular understanding of the structural basis of pilus adhesion is necessary to further optimize their potency and pharmacokinetic parameters.

1.5 Curli Fibers

1.5.1 The Role of Biofilms in Bacterial Adhesion

To aid colonization of and persistence in host or environmental niches, many bacteria coalesce to form encapsulated communities embedded within a complex hydrated matrix of
proteins, nucleic acids, and polysaccharides. These communities, known as biofilms, protect the bacteria within from physical and chemical stresses, such as oxidative damage and desiccation (36, 37). Biofilm deposits on household, industrial, and medical equipment result in a potential reservoir of infectious agents recalcitrant to traditional cleaning techniques. Further, biofilm formation by pathogenic bacterial strains within hosts drastically reduces the ability of the host immune response and antimicrobial agents to combat the infection, and therefore constitutes a major health concern (38-41).

Biofilm formation comprises a series of related steps. Following the reversible cell-surface and/or cell-cell adherence of planktonic microorganisms, CUP pili mediate a robust surface attachment and inclusion in an extracellular matrix (42). In *E. coli* and *Salmonella enterica* biofilms, aggregative fibers known as curli (sometimes referred to by their now-obsolete name, tafi) constitute the major proteinaceous component of this extracellular matrix (43). These matrices promote the formation of floating biofilms (pellicles) at the air-liquid interface of static liquid cultures and can mediate the adhesion of solid cultures to biotic and abiotic surfaces, such as animal and plant tissue, stainless steel, and glass (39).

Curli belong to a class of stable, ordered protein aggregates known as amyloids (44-46). Although commonly associated with pathological protein misfolding in human diseases (47, 48), a significant body of research now suggests that amyloids are also intentionally produced by a variety of organisms to fulfill important physiological functions, such as regulation of hydrophobicity during fungal reproduction or transcriptional regulation (49-51). Curli are among the best-studied functional amyloids, and are an integral part of the biofilm extracellular matrix produced by *Salmonella, Enterobacteriaceae*, and *E. coli* (52). Solid-state NMR studies have found that curli account for 85% of the total biofilm extracellular matrix, with polysaccharides constituting a majority of the remaining 15% (53). In addition to their structural role in maintaining
the biofilm architecture, curli fibers regulate inflammatory and immune responses in the gut, in part by decreasing the permeability of tight junctions in the intestinal epithelium (54, 55). Additionally, curli fibers can be engineered to scaffold a variety of functional biomolecules, with applications ranging from the delivery of self-sustaining therapeutics to environmental bioremediation (56-58).

1.5.2 Curli Fiber Structure

Curli are non-covalent heteropolymeric filaments of CsgA and CsgB subunits, present at ratios of approximately 20:1 CsgA:CsgB in in vivo wild-type fibers (59). Like other amyloid filaments, curli fibers are characterized structurally by their ‘cross-β spine’ architecture, in which repeating β-strand units are oriented perpendicular to the fiber axis (60, 61). This repetitive, tightly-ordered packing of β-strands confers high stability and physical robustness to the filaments. Indeed, curli are resistant to proteolytic degradation or dissolution by sodium dodecyl sulphate (SDS), and instead must be subjected to harsh treatment with formic acid or hexafluoroisopropanol (HFIP) in order to depolymerize fibers into individual subunits (62). Extracellular, in vivo fiber formation can be macroscopically monitored by the staining of bacterial cultures with Congo red, a dye that undergoes a spectral shift upon binding β-rich polymers such as curli and other amyloid species. Though a convenient research tool, precaution is needed due to the nonspecific binding of Congo red to other polymers including biofilm matrix components such as cellulose.

Curli-like fibers can also be formed in vitro, starting from the purified major structural subunit CsgA (46, 63, 64). This process can be monitored by the fluorescence increase from the Thioflavin T (ThT) dye as it binds the forming amyloid structure (62). Following an extended lag time that is associated with the formation of a metastable oligomeric nucleus, in vitro fiber formation occurs spontaneously and follows an exponential growth phase. Prior to nucleation,
concentrated CsgA samples frequently form amorphous aggregates that remain competent for incorporation into curli-fibers (64, 65). The lag time preceding fibrillation can be eliminated by the addition of pre-formed curli fragments or purified CsgB protein, which accelerates fiber formation in a process known as heteronucleation (63, 66, 67). This suggests that curli polymerization is a templated process, as is also seen for other amyloid structures. Deletion of csgB in vivo results in CsgA secretion away from the cell in an SDS-soluble, unpolymerized form [39]. However, when csgB− cells are placed in the vicinity of cells encoding csgB but lacking csgA, these soluble CsgA subunits can polymerize on the surface of CsgB+ cells in a process known as interbacterial complementation (46).

The exact structure of curli fibers has not yet been elucidated with molecular resolution. X-ray fiber diffraction, solid state NMR, and electron microscopy data on CsgA and CsgB fibers grown in vitro point to a cross-β architecture that is most consistent with stacked β-helical subunits rather than stacked parallel in-register β-sheet structures as often seen in other amyloids (68, 69). However, it should be noted that in vitro fibers are frequently more variable in width than their wild-type counterparts, and often show 3–4 nm protofilaments branching off from thicker filament bundles. These observations suggest that in vivo assembled curli may not be fully structurally isomorphous with in vitro grown curli-like fibers (68). This is consistent with the observation that the presence or absence of periplasmic, non-structural accessory proteins such as CsgE or CsgC in vivo can alter curli morphology (46, 70).

### 1.5.3 Curli Fiber Assembly

Upon visualization by electron microscopy, E. coli-associated curli usually appear as a tangled mass of linear, surface-associated fibers of 4–6 nm width and several microns length (46). Within biofilms, curli fibers form an interwoven mesh that supports the extracellular matrix and encapsulates individual cells. In these matrices, curli frequently appear detached from the
embedded cells. It is unclear if this observation represents a regulated process necessary for efficient biofilm matrix formation, or is an artefact of cell desiccation during sample preparation procedures.

The proteins involved in the structure and biogenesis of curli fibers are encoded by two divergently-transcribed operons; csgBAC and csgDEFG (71). Structurally, curli fibers are primarily composed of CsgA, a 13-kDa, β-rich protein (65, 72). These subunits are assembled into fibers in vivo by a mechanism known as nucleation-precipitation, sometimes referred to as Type VIII secretion, outlined in Figure 1.2 (21). Briefly, CsgA is targeted to the SecYEG translocon by an N-terminal signal peptide, which is proteolytically cleaved upon translocation into the periplasm and results in a 13.1 kDa mature periplasmic CsgA subunit (73). In the periplasm, CsgC maintains CsgA in an amorphous, β-strand deficient state to prevent toxicity to the host cell by inhibiting primary nucleation and/or elongation through a series of electrostatic interactions (44, 74, 75). Periplasmic CsgA is then targeted to the CsgG outer membrane pore by a 22-amino acid N-terminal signal sequence. The CsgG pore is composed of nine identical CsgG subunits, each of which contributes four β-strands to a 36-strand transmembrane β-barrel and three residues to a series of rings that constrict the secretion channel diameter to 9 Å, a size consistent with the expulsion of unfolded, soluble CsgA into the extracellular space (76-78). Once secreted, CsgA fiber formation and elongation are nucleated by CsgB on the cell surface in a CsgF-dependent manner (63, 67, 79, 80).
1.6 Figures

Figure 1.1 Overview of CUP pilus structure and biogenesis. Pilus subunits are transported to the periplasm, where they are recognized by dedicated chaperones that promote pilin folding by donating their G1 β-strand to transiently complete the subunits’ Ig-like fold. Chaperone-pilin complexes are then recruited to the outer membrane usher, where they are transported across the membrane following donor strand exchange with their neighboring subunit. Once assembled, the pilin rod forms a tight right-handed helix supporting a tip complex containing between 0-3 tip adaptor subunits and a two-domain adhesin. Sample structures shown represent components of the Type 1 pilus system, including FimDCH (PDB ID 3RFZ), FimCA (PDB ID 4DWH), FimA rod (PDB ID 6C53), and FimH in its bent and elongated conformation (PDB IDs 5JR4 and 5JQI, respectively).
Figure 1.2. Working model for curli biogenesis. Unstructured CsgA transported into the periplasmic space by the Sec translocon is maintained in a soluble, secretion-competent state by CsgC prior to delivery to the CsgG pore by the periplasmic accessory protein CsgE. Once secreted to the outer membrane, CsgA adopts an amyloid fold and is assembled into fibers by the action of CsgF and CsgB.
Chapter 2: Structure-Function Analysis of the Curli Accessory Protein CsgE

2.1 Scope of Work

Of the work described in this chapter, I was responsible for CsgE and CsgF crystallization attempts, design and performance of the Thioflavin T assays, design and performance of the bile salt sensitivity assays (with help from Dr. Kanna Nagamatsu), genetic conservation analyses, differential scanning fluorimetry assays, dynamic light scattering, and SEC-MALLS (with the help of Wenjie Wang and Chau Wu). Portions of this chapter are reproduced from:

© 2018, American Society for Microbiology

2.2 Introduction

2.2.1 CsgE Function

During curli fiber biogenesis, the accessory periplasmic protein CsgE is necessary for the transport of curli structural components through the CsgG pore (81). CsgE is thought to bind unfolded CsgA in the periplasm and deliver it to CsgG, where CsgE assembles into a nonameric ring that ‘caps’ the periplasmic vestibule of the outer membrane pore (76, 82). CsgE’s putative role as a gating and specificity factor for the CsgG pore is based on direct biochemical and
biological evidence. Co-immunoprecipitation studies have demonstrated a stable physical interaction between CsgE and CsgG (77). In addition, a complex consisting of 9 subunits each of CsgE and CsgG has been observed using cryo-electron microscopy (76). Single-channel current recordings and in vivo toxicity assays reveal that the presence of CsgE decreases the permeability of the CsgG channel to ions and small molecules in a manner dependent on CsgE concentration (76). CsgE can also differentially regulate protein translocation through the CsgG channel (81). Translocation of non-native client proteins is attenuated in the presence of CsgE, while translocation of curli proteins encoding the 20-21-residue N-terminal CsgG localization sequence is unaffected (81). Direct biochemical evidence for a CsgE-CsgA interaction has not been previously reported, although CsgE can delay CsgA amyloid formation in vitro, suggesting an interaction between CsgE and CsgA (81).

2.2.2 CsgE Structure

To determine the structural basis for CsgE function during curli biogenesis, I initially sought to generate an X-ray crystal structure of the CsgE protein. To this end, I assessed the ability of WT CsgE to form crystals in a variety of commercially-available sparse matrix crystallography screens using the hanging drop vapor diffusion method with a 200 nL total drop volume. The initial protein concentration used was 10 mg/mL, although this was varied throughout the screening and refinement process. Crystalline precipitate was observed in conditions containing calcium acetate, calcium chloride, and magnesium acetate, suggesting that divalent cations play an important role in stabilizing the protein during crystal packing. The presence of assorted PEG species and organic additives had a variable effect on CsgE crystallization, and a clear pattern did not emerge for the success of these additives. Unfortunately, further attempts to refine those initial hits using grid screens, additive screens, and streak seeding were unsuccessful. I attributed this to the heterogeneity of CsgE oligomeric species in solution. To circumvent this problem, I
also attempted to crystalize a mutant that was less prone to forming oligomers during purification; CsgE F67A. I screened 18 mg/mL of purified CsgE F67A in a variety of sparse matrix screens, revealing that magnesium chloride and ammonium sulfate at a pH between 7.5 and 8.5 yielded the most promising hits for this construct. However, attempts at further refinement were again unsuccessful. Following the successful generation of a solution-state NMR structure of the CsgE W48A/F79A double mutant by Qin Shu in the Frieden laboratory determined (Biological Magnetic Resonance Data Bank accession number: 25927) (Figure 2.1A), I discontinued my crystallization efforts to focus on functional assays.

CsgE W48A/F79A is functional in curli biogenesis \textit{in vivo} and behaves similar to WT CsgE in \textit{in vitro} amyloid polymerization assays. Thus, the structure of CsgE W48A/F79A provides insight into surface-exposed residues that mediate interactions with CsgA, CsgG and itself during homo-oligomerization (83). Examination of the electrostatic surfaces on CsgE reveal three regions of concentrated charge; i) a positively charged ‘head’, including R47, K70, and R71, ii) a negatively-charged ‘stem’, including E31, D33, and E85 and iii) a negatively-charged ‘tail’, including E3, E5, D102, D106, and E107 in the N- and C-terminal IDR (Figure 2.1B). When the CsgE NMR structure is fit into a low-resolution cryo-EM density map, two possible CsgE orientations were observed; one in which the disordered tails coalesce in the center of the disc, and one in which the positively-charged heads form the center (83).

To identify residues responsible for mediating the CsgE-CsgA interaction, the NMR chemical shift perturbation of all assigned CsgE W48A/F79A peaks was examined in the presence and absence of CsgA. Eleven residues (E31, S32, D33, Y34, S45, A46, R47, W48, R71, E85, and L88) were identified that exhibited a greater-than twofold higher chemical shift perturbation than average, suggesting that these residues are involved in the CsgE-CsgA interaction (Figure 2.2A-C). The four residues with the greatest chemical shifts cluster into two solvent-exposed and
oppositely-charged patches on CsgE (Figure 2.2D). Residues R47 and R71 are located in the $\beta_1$-$\beta_2$ and $\beta_3$-$\alpha_2$ loops at the ‘head’ of CsgE, coalescing to form a patch of positive charge (83). Conversely, residues E31 and E85 contribute to a negatively-charged patch at the ‘stem’ of CsgE near the intrinsically-disordered tails. A global comparison of the spectra in the presence and absence of CsgA demonstrates a high degree of overlap, suggesting that the backbone amides of $^{15}$N-labeled CsgE W48A/F79A are not perturbed upon addition of CsgA (Figure 2.3).

To verify that the observed chemical shifts were not an artefact unique to the CsgE W48A/F79A protein, we also conducted chemical shift perturbation studies with WT CsgE in the presence and absence of CsgA (Figure 2.4). The 2D $^1$H-$^{15}$N HSQC spectrum of WT CsgE overlapped with that of CsgE W48A/F79A, allowing for partial chemical shift assignment (83). As with the CsgE W48A/F79A spectrum, small perturbations in the chemical shifts were detected upon addition of CsgA that corresponded to residues E31, R71, and R47, indicating that these interactions are not unique to the monomeric double mutant.

2.3 Structure-Function Analysis of CsgE

2.3.1 Functional Validation of Charged Residues by Site-Directed Mutagenesis

CsgE can prevent purified CsgA from aggregating into an amyloid fiber (81). To investigate the role of the positively-charged head and negatively-charged ‘stem’ patches in the CsgE-CsgA interaction, we examined the effect of mutations in the head and stem residues of CsgE on polymerization of purified CsgA in vitro using the Thioflavin T (ThT) assay (65, 83). Addition of recombinant WT CsgE to freshly-purified CsgA at a substoichiometric ratio of 1:4 CsgE:CsgA increased the time necessary for CsgA to reach 50% of its maximal aggregation ($T_{50}$) (Figure
2.5A). CsgE R47A was incubated with CsgA at 25 °C for 48 hours and ThT fluorescence was measured in 20-minute intervals. No changes in the overall thermal stability of CsgE were observed for any of the mutants tested (Figure 2.6). We anticipated that disruption of the CsgE-CsgA interaction would shorten the observed T50, and indeed, CsgE R47A was less efficient than WT CsgE in extending the lag phase of CsgA (Figure 2.5C). To determine if this effect was charge- and residue-specific, nearby nonpolar residues were also mutated to alanine, and the resultant protein was assessed for its ability to delay CsgA aggregation. CsgE W51A and CsgE F67A performed similarly to WT CsgE, and CsgE P44A prolonged the lag phase relative to WT CsgE (Figure 2.5C). The E31 and E85 residues in the stem of CsgE were each mutated to both an alanine and a lysine to similarly investigate the role of the stem in mediating the CsgE-CsgA interaction. CsgE E85K, E31A and E31K all demonstrated an increased ability to extend the lag phase of CsgA (Figure 2.5D). Thus, the R47 head residue makes a critical charge-charge interaction with CsgA that is important for its inhibition of CsgA amyloidogenesis, while the E31 and E85 residues temper the CsgE-CsgA interaction by making the surface change of the CsgE molecule more negative.

2.3.2 The CsgE-CsgA Interaction is Necessary for in vivo Curli Fiber Formation

Because CsgE is required for curli biogenesis in vivo (81), we tested the ability of the CsgE head and stem mutants to mediate curli formation by ectopically expressing csgE mutants in MC4100 ΔcsgE. Curli expression was quantified from cells grown on solid YESCA agar at 30 °C for 40 hours. Colonies were isolated, treated with hexafluoroisopropanol (HFIP) to depolymerize curli fibrils, and subject to western blot analysis to measure total CsgA (62). The total level of CsgA detected in MC4100ΔcsgE mutants were decreased fourfold relative to WT MC4100 cells, demonstrating that CsgA is not accumulated at high levels in the periplasm (Figure 2.7). Mutation
of the positively-charged head residue R47 (CsgE R47E) or the hydrophobic residue W51 (CsgE W51A) reduced detectable CsgA levels to that of a \textit{csgE} null mutation, indicating a complete loss of CsgE function during curli biogenesis (Figure 2.7). CsgE P44A also reduced CsgA levels relative to WT CsgE, but not to the same degree as a \textit{\Delta csgE} mutant. Similarly, the ectopic expression of the CsgE E31A, CsgE E85A, CsgE E31K, and CsgE E85K mutants each facilitated less curli production than WT CsgE (Figure 2.7). The monodisperse double mutant, CsgE W48A/F79A, was able to fully complement curli biogenesis in a \textit{\Delta csgE} background (Figure 2.7).

### 2.3.3 The CsgE-CsgG Interaction is Robust

Overexpression of CsgG in \textit{E. coli} has been shown to be lead to toxicity in the presence of bile salts (76). Co-expression of CsgE rescues this toxicity, suggesting that CsgE is able to gate the CsgG channel (74, 76). Thus, we used a bile salt sensitivity assay to probe the CsgE-CsgG interaction, hypothesizing that CsgE mutants with a disrupted CsgE-CsgG interaction would be unable to rescue \textit{E. coli} overexpressing CsgG from bile-salt-induced toxicity. The negatively-charged tail residues in CsgE (E3, E5, D102, and D106) were mutated to alanine or lysine, and the mutants were expressed in \textit{E. coli} MC4100 \textit{\Delta csg} cells along with \textit{csgG} in the presence of bile salts. In addition, the mutations in the head and stem observed to affect the CsgE-CsgA interaction were also tested in this assay. All of the mutants were stably expressed (Figure 2.8). Serial dilutions of cells harboring these mutations were plated on McConkey agar supplemented with 2% w/v bile salts, and growth was quantified after 48 hours of incubation at 26 °C (Figure 2.9A). We found that all of the mutants tested were able to rescue cells from bile salt sensitivity, suggesting they did not interrupt the CsgE-CsgG interaction required for gating the channel (Figure 2.9A). However, in the curli biogenesis assay described above, we found that a D102K mutation resulted in a near-complete loss of curli assembly while the E5K mutation had no effect (Figure 2.9B).
2.3.4 CsgE Oligomerization is Time-and Temperature-Dependent

Previous reports have provided structural and biochemical evidence of CsgE’s propensity to adopt a variety of oligomeric forms (Figure 2.10A) (82, 83). To understand the influence of key residues on the kinetics of this process, we characterized the effect of point mutations on the oligomerization states of CsgE. Consistent with previous reports, we observed the partitioning of WT CsgE into two distinct peaks on a Superdex 200 10/300 column immediately following initial purification (Figure 2.10A) (83). The two primary species present corresponded to monomeric CsgE (>90% by peak area) and a small amount of nonamer. This distribution remained stable for three days, with a moderate shift towards the nonameric form occurring in the subsequent 17 days at 4 °C.

To understand the variables affecting the dynamic process of CsgE assembly, we conducted additional size-exclusion chromatography (SEC) experiments on WT CsgE at different incubation times and temperatures. When incubated at 25 °C, the shift to the nonameric species occurred over a much shorter time scale than at 4 °C. After 5 hours at 25 °C, the nonamer peak area is approximately 20% of the monomer peak area, and after 28 hours, the nonamer peak area is approximately 40% of the monomer peak area (Figure 2.10B). At 37 °C, the monomeric peak loses >65% of its area within 1 h of incubation. The oligomeric species formed during this time include both a nonameric species and other larger oligomers (Figure 7C, D).

SEC-MALS tracings of WT CsgE reveal four distinct species; two forms of monomer, a small oligomer, and high-molecular weight oligomers and aggregates (Figure 2.11A). Partitioning of the monomer peak into monomeric and nonameric species is observed when it is re-run over a size exclusion column (Figure 2.10A.) In contrast, CsgE W48A/F79A appeared monodisperse on SEC and ran as a single monomeric band on a native-PAGE. The SEC-MALS tracing reveals a single dominant peak with a calculated molecular weight of 14.5 kDa (Figure 2.11B). To parse
out the individual contributions of W48 and F79 to the temperature-dependent oligomerization of CsgE, the CsgE W48A and CsgE F79A mutants were analyzed via SEC and native-PAGE. Both CsgE W48A and CsgE F79A appeared monomeric by SEC, although they were differentially retained on the column relative to both WT CsgE and the molecular weight standards (83) (Figure 2.12A). On native PAGE, purified WT CsgE appeared primarily nonameric at 4 °C and 25 °C with a shift to high-molecular-weight oligomers at 37 °C (Figure 2.12B). CsgE W48A/F79A was monomeric at all temperatures and time points tested (Figure 2.12C). Following short incubation times at 4 °C and 25 °C, CsgE W48A segregated into two bands corresponding to the monomer and nonamer, with the monomeric species disappearing after 24 hours at 25 °C (Figure 2.12D). At 37 °C, higher-molecular weight species began to appear after 5 hours and dominated the distribution (Figure 2.12D). In contrast, CsgE F79A was comparatively more heterogeneous at 4 °C and 25 °C and demonstrated a pronounced shift to high molecular weight aggregates after only an hour of incubation at 37 °C (Figure 2.12E). Thus, both W48 and F79 contribute to the oligomerization of CsgE at 4 °C and 25 °C, but W48 drives the formation of the high-molecular-weight CsgE species at 37 °C.

2.3.5 Key Residues for CsgE Function are Subject to Evolutionary Selection

To better understand the evolution of the csgE gene throughout all bacterial phyla, we examined the sequences of CsgE homologues in the EMBL-EBI database with an E-value of <10^-10. A total of 397 CsgE amino acid sequences representing 3,388 nucleotide sequences were aligned using the MAFFT FFT-NS-I x2 algorithm (84). Overall pairwise identity for this sequence set was 58.7%, with an overall pairwise similarity of 72.2%. Pairwise identity by residue is indicated by the height of the bars in Figure 2.13A and is mapped onto the CsgE structure in Figure 2.13B. These findings reveal a cluster of conservation near the head of CsgE, as well as
select residues in the negatively-charged, intrinsically-disordered tails. Of note, R43 and R47 have a pairwise identity of 96% and 100%, respectively, while the residues E3, E5, D102, D106, and D107 within the intrinsically-disordered tails each have a pairwise identity of >85% (Figure 2.13A). This supports our data that the concentration of positive charge at the head of CsgE and the negative charge in the tail regions of CsgE contribute to CsgE function. We also find that the E31 and E85 stem residues each have a pairwise identity of <35%, suggesting that the CsgE protein is more tolerant to variation at these residues.

To examine the selective pressures acting upon the csgE gene within the E. coli species, we obtained the nucleotide sequences for 1,985 csgE homologues across all E. coli genomes in the EMBL-EBI Bacterial Genomes database. A total of 41 unique amino acid sequences remained following removal of identical sequences. The mature amino acid sequences demonstrated an overall pairwise identity of 91.3% and an overall pairwise similarity of 93.3%. The ratio of synonymous to non-synonymous mutations (dS/dN) was examined to better identify the selection pressures acting on each residue within the csgE gene. Negative/purifying selection was found at 27 sites with a Bonferroni-corrected P-value less than 0.05 (indicated by dark blue in Figure 2.13A). Nine additional sites were identified with a Bonferroni-corrected P-value between 0.05 and 0.1 (indicated by light blue in Figure 2.13A). No sites of positive selection were identified. While the majority of the negatively-selected residues face inward to contribute to the core fold of CsgE, R47 and R43 are surface-exposed and subject to strong purifying selection (Figure 2.13B). Conversely, E31 and E85 are not subject to selective pressure, again indicating that some variability in these residues is tolerated.
2.4 Discussion

CsgE is a periplasmic curli assembly protein required for the transport of amyloidogenic CsgA to the CsgG pore (82). The NMR structure of a stable CsgE W48A/F79A mutant reveals a three-stranded anti-parallel β-sheet core flanked by three α-helices (83). In this study, we used NMR chemical shift perturbation to identify residues that participate in the CsgE-CsgA interaction. The roles of these residues were validated using *in vitro* and *in vivo* assays of CsgE function. We identified a region of positive charge formed by the confluence of the β₁-β₂ and β₃-α₂ loops at the head of CsgE that plays a key role in the CsgE-CsgA interaction and is required for curli assembly (Figure 2.5C, 2.7). The identity of residue E31 in the α₁-β₁ loop and E85 in α₃ in the stem of CsgE modulates the ability of CsgE to prevent CsgA aggregation *in vitro* and assemble curli fibers *in vivo* (Figure 2.5D, 2.7). Residue D102 in the negatively-charged C-terminal tail of CsgE contributes to curli assembly without disrupting the CsgE-CsgG interaction (Figure 2.9A, B).

The finding that charge-charge interactions underlie interactions within amyloidogenic proteins is consistent with existing literature (74, 85). During curli biogenesis, the periplasmic chaperone CsgC is known to prevent inappropriate periplasmic CsgA polymerization through a series of electrostatic interactions, instead targeting CsgA for degradation in the event of disrupted subunit secretion (75). Outside the curli system, computational and *in vitro* analyses of charged amino acids in amyloid β reveal the importance of electrostatic interactions in oligomerization (85). We demonstrate that the head mutant CsgE R47A shows a marked decrease in its ability to protect CsgA from aggregation, and that disruption of the R47 residue abrogates the ability of CsgE to assemble curli fibers without affecting degradation of CsgA in the periplasm (Figure 2.5A, 2.7.) Coupled with the invariance of this residue across all Gram-negative bacterial species, these data suggest a specific charge-charge interaction occurs between the head of CsgE and CsgA. This finding supports the “head centric” model of complex formation in which the positively-
charged head groups of CsgE coalesce in the center of the nonameric cap to bind CsgA and facilitate its transport to the CsgG pore.

NMR perturbation of the CsgE chemical shifts in the presence of CsgA also revealed a novel charge-charge interaction between the stem of CsgE and CsgA (Figure 2.2). Introduction of the CsgE E31K and CsgE E85K mutations resulted in a 60% reduction in curli fiber assembly despite an augmented ability to delay the lag phase of CsgA polymerization (Figure 2.5D, 2.7). These findings suggest that the surface charge in the stem of CsgE controls the balance between CsgA association and dissociation, and the low degree of conservation at residues 31 and 85 suggest that CsgE can tolerate substitutions at these positions.

CsgE harboring mutations in the head, stem, and tail regions retained the ability to block passage of small molecules through the CsgG pore (Figure 2.9). The persistence of this interaction despite disruption of charged residues is consistent with a previous report that mutations disrupting the formation of the nonameric CsgE cap in vitro were still able to complement curli biogenesis in vivo (83). In this study, we demonstrate that CsgE W48A/F79A is unable to form nonameric and high-molecular-weight species in a temperature-dependent manner in vitro. We also demonstrate that the W48A and F79A mutations individually shift the CsgE population to the monomeric species. The persistence of in vivo curli assembly despite an observed attenuation of in vitro oligomerization with CsgE W48A/F79A suggests that curli biogenesis may be affected by the presence of other curli proteins or periplasmic factors that are not present in vitro.

In sum, these data demonstrate that positively-charged residues in the ‘head’ of CsgE mediate specific charge-charge interactions with CsgA, supporting the “head-centric” cryo-EM fitting model described previously (83). We also found that negatively-charged stem residues modulate CsgE binding to CsgA, possibly to maintain a finely-tuned equilibrium that promotes
CsgA transport and secretion. This balance of specific and nonspecific charge-charge interactions provides a stark contrast to known protein cages like the GroEL/GroES chaperonin ABC toxin complexes. In these systems, the oligomeric cages recognize hydrophobic residues within a broad array of client proteins prior to encapsulation to facilitate their sequestration and promote protein folding (86-88). These insights into the structural basis of CsgE function enhance our mechanistic understanding of bacterial amyloid assembly during biofilm formation and provide insight into a mechanism of protein recognition distinct from those currently known to occur in the bacterial periplasm.

2.5 Future Directions

2.5.1 CsgE Oligomerization

In the current working model of curli biogenesis, periplasmic CsgE assembles into nonameric cap that traps CsgA in the periplasmic vestibule of CsgG, facilitating translocation to the outer membrane. Evidence supporting this model includes a series of low-resolution cryo EM structures of CsgE assembling into a nine-membered ring in the presence and absence of CsgG and the separation of purified, monomeric CsgE into two peaks after passage through a size exclusion column. In this work, we demonstrate the time- and temperature-dependence of this process using size exclusion chromatography and native-PAGE (2.10 and 2.12). Individually, W48A and F79A each showed a decreased propensity to form oligomers following incubation at high temperatures, which is consistent with previous studies indicating that the W48A/F79A double mutant cannot form higher molecular-weight oligomers. We also utilized SEC-MALS to verify that CsgE freshly prepared from cellular lysate exists in multiple assembly states, while CsgE W48A/F79A purified directly from cells is almost exclusively monomeric (Figure 2.11).
Several questions about the structural basis and functional role CsgE oligomerization remain. Although it is apparent that W48 and F79 both contribute to hydrophobic interactions between subunits that promote oligomerization, it is possible that other hydrophobic and charged residues on the surface of CsgE may also modulate the equilibrium between CsgE oligomeric states through a series of attractive and repulsive interactions. To further explore the self-assembly behavior of mutants assessed elsewhere in this study, I performed a series of SEC-MALS experiments to explore the effects of mutations to charged residues on the oligomerization of CsgE. I used dynamic light scattering (DLS) experiments with wild-type CsgE at various concentrations to characterize the species present in solution (Figure 2.14). From these studies, I determined that a concentration of 2 mg/mL was ideal to observe distinct oligomeric species while maximizing the total protein signal following elution from the column.

All CsgE constructs were individually purified with two sequential cobalt affinity columns. The resultant protein was then injected into a Superdex™ 200 Increase 10/300 GL column coupled to a light scattering detector and refractometer. Scattering and refraction profiles were compared with elution times during analysis (Figure 2.15.) As described above, WT CsgE, as well as many of the mutants tested, partitioned into four distinct peaks. The first peak to elute typically had a molecular mass of 400-500 kDa with a variable polydispersity, likely reflecting a range of higher-molecular weight oligomers and aggregates (ie, the association of multiple nonamers). A second, much smaller peak eluting after approximately 50 minutes typically had a molecular weight of approximately 110 kDa. This peak typically represented between 0-2% of the total sample volume, a likely corresponds to the nonameric species. Finally, two late-eluting peaks or one wide, polydisperse peak eluting between 55 and 70 minutes was observed. The first of these peaks typically had a calculated molecular of 14 kDa; consistent with the predicted size of the monomer. When present (as in WT CsgE), the second peak had a highly variable calculated
molecular weight between 5-20 kDa. The identity of this peak remains unknown. To verify the purity of our protein samples, we collaborated with Nicole Wagner to determine the purity of our protein samples via mass spectrometry. She identified low levels of CsgEW48A/F79A contaminant in the WT CsgE sample. These could have arisen from accidental contamination of laboratory apparatus, or incomplete elution and cleaning of the cobalt column between purification of the two constructs. However, no protein corresponding to a possible degradation product or other contaminant was observed. To determine if the presence of NaCl was altering the oligomerization of CsgE, these experiments were repeated following dialysis into 50 mM KPi, pH 7.3 + 150 mM NaCl. We found that buffer did alter the elution profile somewhat, suggesting that salt does have an effect on CsgE oligomerization. Specifically, the presence of NaCl decreased the elution delay between the first and second nonameric species. It also led to the emergence of a shoulder on the W48A/F79A monomeric peak. Together, these findings could indicate that the presence of salt may alter the intermolecular interactions required for assembly of the oligomer.

As described above, the SEC chromatograms of the mutant CsgE displayed a degree of complexity not previously observed in experiments performed in the Frieden laboratory (83). The differences observed between our traces and those supporting the time- and temperature-dependent oligomerization findings were likely due to a difference in purification approaches. Samples used to generate the SEC traces shown in this report and previously by Shu et al. were first purified by both affinity and size exclusion chromatography to remove high-molecular-weight contaminants. Thus, the species eluting from the second SEC column in those studies represented products formed from monomeric CsgE following purification on the first column. For future studies, protein will be purified according to the methodology developed by Shu et al, wherein the sample is first passed over a SEC column prior to SEC-MALS analysis to ensure that
high-molecular weight do not convolute the observed results. Additionally, these SEC-MALS analyses will be expanded to include mutation to more hydrophobic residues, including W48 and F79. Developing a more detailed understanding of the effect of CsgE surface residues on oligomerization may elucidate a link between CsgE oligomerization and CsgA and/or CsgG binding, which could in turn inform the design of future classes of small molecules that disrupt this critical interaction to inhibit biofilm formation.

2.5.2 The CsgE-CsgG Interaction

Based on the orientation of CsgE monomers within our favored ‘head centric’ CsgE-CsgG complex fitting model, we hypothesized that the negatively-charged, intrinsically-disordered tails of CsgE mediate interactions with CsgG. To our surprise, mutation of negatively-charged residues within these tails, including E3, E5, D102, and D106, did not abrogate the CsgE-CsgG interaction. There are four possible explanations for this outcome; i) the fitting model does not accurately reflect the orientation of CsgE within the complex; ii) the structure of the complexes assembled in vitro differ from those assembled in vivo; iii) the interaction is robust enough to accommodate single point mutations and iv) the bile salt sensitivity assay lacks the sensitivity to measure minor perturbations to the kinetics of complex formation. These possibilities are not mutually exclusive, and I believe (iii) and (iv) most likely explain our observed findings.

An alternative “tail centric” placement of CsgE monomers that is consistent with the existing low-resolution EM reconstruction has been proposed. It is also possible that the CsgE-CsgG complexes formed in vitro differ from those formed under physiologic conditions in vivo, as the CsgE-CsgG complexes imaged using cryo-EM were generated in detergent-destabilized lysosomes in vitro using high concentrations of protein and surfactant (89). These conditions could affect the structure or position of the intrinsically-disordered tails within the complex, or could induce significant local or global structural changes within either protein. However, our
finding that charged residues in the head and stem of the molecule are responsible for modulating the CsgE-CsgA interaction provides experimental support for the head-centric model. Additionally, CsgE harboring mutations in charged and hydrophobic residues elsewhere on the molecule were able to fully complement in the bile salt sensitivity assay, suggesting that they are also not necessary for complex assembly.

Because the putative interface between CsgE and CsgG comprises several discrete charge-charge interactions, we believe that the persistence of complementation likely arises from the robust nature of this interaction. To further dissect this interaction, one would have to generate a construct in which combinations of the aforementioned residues are deleted and tested for their ability to restore resistance to bile salts. Alternatively, one or both intrinsically-disordered tails could be deleted in their entirety and the resulting constructs tested in the bile salt sensitivity assay. If either mutant was created, it would be essential to confirm robust expression of the mutated protein via Western blot analysis. We are also aware of efforts by Remaut, et al. to generate a crystal structure of the CsgE-CsgG complex using chemical crosslinkers to generate a stable complex. These studies, or analogous studies using high-resolution CryoEM of CsgE-CsgG complexes within their native outer membrane environment, would provide additional insight into residues at the CsgE-CsgG interface.

An additional point of interest arises from the observation that the W48A/F79A mutant, which is present entirely as a monomer in vitro, does not result in a defect in curli biogenesis when complemented in vivo (83). There are two possible explanations for this observation: i) oligomerization is not required for translocation of CsgA through the CsgG pore ii) in vitro measurements of oligomerization do not fully reflect in vivo oligomerization. While both of these options are viable possibilities, it is possible that the presence of CsgA could alter the oligomerization of CsgE in vivo by inducing structural changes into CsgE that could promote self-
assembly, even in the absence of the hydrophobic interactions required for \textit{in vitro} assembly. The absence of CsgA from the bile salt sensitivity assay is another variable that could be affecting our results. According to the current assembly model, CsgE first binds unfolded CsgA in the periplasm prior to interacting with CsgG. Although the bile salt sensitivity assay demonstrates that binding to CsgA is not an absolute prerequisite for CsgE binding to CsgG, it is possible that CsgA may induce local conformational changes within CsgE that alter the kinetics of complex formation and affect the structure of the final complex. To test this hypothesis, a plasmid containing CsgA could be introduced to the existing one-plasmid bile salt sensitivity assay to determine if pore capping is altered by the presence of CsgA. CsgE mutants that do not bind CsgE (ie, R47A) could be used as negative experimental controls. Additionally, protein footprinting techniques such as fast photochemical oxidation of proteins (FPOP) or hydrogen-deuterium exchange (HDX) could be used to assess the solvent accessibility of surfaces on CsgE and CsgG in the presence and absence of CsgA (82). To ensure that the CsgE-CsgG complex assessed \textit{in vitro} corresponds to the native structure of the complex \textit{in vivo}, CsgG could be incorporated in nanodiscs to better recapitulate membrane conditions (90).

\textbf{2.5.3 Structural Studies of CsgF}

As discussed above, CsgF is a 13 kDa curli accessory protein known to play a role in the anchoring of the curli fiber to the outer membrane (79, 91). Deletion of CsgF results in the secretion of polymerized CsgA and CsgB subunits away from the host cell. CsgF is known to interact directly with the CsgG outer membrane pore, and is predicted to interact directly with the outer leaflet of the outer membrane. Primary sequence analysis suggests that the N- and C-terminal tails of CsgF are likely intrinsically disordered (92). To attain structural insights into CsgF function, I attempted to crystallize the CsgF protein for structural determination. Although I was able to obtain small, needle-shaped crystals, I was unable to phase the resultant native data set.
Attempts to derivatize the protein via iodine vapor diffusion were unsuccessful. Ultimately, work on the project was halted after we received notification that a group in Germany had successfully determined a solution NMR structure of CsgF (91). The structure, which was publicly released in 2018 (PDB ID 5M1U), reveals a core four-stranded \( \beta \) sheet flanked by a 21-residue \( \alpha \)-helix. Large regions of the CsgF surface contain exposed hydrophobic residues, supporting the hypothesis that CsgF requires interactions with multiple surfaces to fold and perform its required function in curli biogenesis. In contrast to previous findings, this group found that CsgA can suppress the polymerization of purified CsgA in a similar manner to CsgE at equimolar ratios. As expected, both the N- and C-terminal tails of CsgF are intrinsically disordered and highly mobile. The authors also discovered that CsgF was prone to oligomerization, requiring detergents to stabilize the monomeric state. These findings mirrored our own observations made during purification of the CsgF. Additional ThT studies suggested that CsgF may be aiding in the folding of CsgB into its mature amyloid form.

To better understand the interactions CsgF must make with CsgG and CsgB to facilitate curli biogenesis, future efforts would best be geared towards solving the structure of the outer membrane curli assembly complex, including CsgG, CsgE, and CsgF, in the absence of CsgA. Such a structure would inform stoichiometry and reveal intermolecular contacts, elucidating possible mechanisms for inhibition of curli assembly and biofilm formation. Due to the complexity of this structure and the high degree of variability that may be present in the stoichiometry, cryo-EM would likely be the most appropriate method for this level of structural analysis. Based on the known CsgF structure and sequence properties, it is likely that the \( \beta \) sheet templates CsgB folding and forms stable intramolecular contacts that anchor the fiber to the cell surface. CsgF also likely interacts with both the membrane and the extracellular loops of CsgG. Because these interactions occur on the outer membranes of cells, they may represent prime targets for the development of
small molecular inhibitors of curli secretion complex assembly. Furthermore, studying the effect of curlicides on the formation of these processes may elucidate their mechanism of action, allowing for the structure-informed optimization of curlicide binding to their targets (93).

2.5.4 Localization of Csg Proteins

Although the individual components of the curli system have been well studied, comparatively little is known about the localization of individual subunits within bacterial cells synthesizing E. coli. Previous studies have observed a nonrandom distribution of CsgG on the surface of curli-expressing E. coli. Interestingly, deletion of other curli subunits, such as CsgA, disrupts this polar localization. To better understand the factors governing the distribution of curli subunits within a cell, I proposed using superresolution microscopy to track the localization of single curli subunits real time in live bacterial cells (94). Superresolution microscopy utilizes standard optical imaging systems coupled with advanced image analysis software to determine the localization of single fluorescent particles with 20 nm resolution. By fusing common fluorescence proteins like photoactivatable mCherry (PA mCherry) or mCitrine to target proteins of interest, it is possible to track the motion of single protein molecules in live cells (95).

To this end, I created an mCherry-CsgE fusion construct to monitor the localization of CsgE in live cells. These recombinant proteins were initially expressed from the pTRC99a plasmid into which they were cloned. This fusion protein is fully able to complement curli production in E. coli MC4100 cells lacking the WT csgE gene, thus verifying that the fluorescent tag does not impact CsgE function. Preliminary experiments with the Biteen lab at the University of Michigan verified the successful expression and folding of PA-mCherry, but were unable to determine subcellular localization of single particles due to the large quantity of protein being expressed. To mitigate this problem, CsgE will need to be transferred to a low-copy number plasmid or into its
native locus on the bacterial chromosome to decrease the total quantity of protein present and allow for single-molecule tracking.

I also designed two mCitrine-conjugated CsgG constructs to allow for simultaneous localization of both CsgE and CsgG. To optimize the efficiency of curli assembly, one would expect these proteins to co-localize in the bacterial cell. These constructs were ultimately created by Kanna Nagamatsu in the Chapman lab, and will soon be tested for adequate folding and complementation of MC4100ΔcsgG cells. If CsgG function is unaffected, CsgG and CsgE could be visualized simultaneously to examine the relative localization of the two proteins in the cell. This would also provide stoichiometric information about the CsgG-CsgE interaction. Further, the localization of CsgG and CsgE could be monitored in the absence of various curli subunits to determine the effect of other subunits on CsgG localization. Curlicides that disrupt surface fiber formation could be introduced to see if loss of fiber assembly, even in the presence of expression of all curli subunits, disrupts CsgG localization. Mutations to CsgE and CsgG could also be introduced and the effect on formation of the CsgG-CsgE complex observed directly by assessing the changes in colocalization patterns.

2.5.5 Characterizing the Inhibition of α-Synuclein by CsgE

Previous studies have demonstrated that the potent inhibitor of curli assembly, CsgC, can also decrease the rate of α-synuclein by interacting with common sequence element found in both CsgA and α-synuclein (73). Conversely, small molecular inhibitors of curli synthesis have a variable effect on the formation of α-synuclein fibers; some accelerate fiber formation, while some abrogate it (93). Previous reports have also attempted to characterize the effect of CsgE on CsgC aggregation with mixed results (96). I used the ThT assay to further explore the effect of CsgE on α-synuclein fiber formation. I found that CsgE was able to delay α-synuclein fiber formation at a 4:1 ratio (Figure 2.16). Additionally, I found that certain CsgE point mutants (W51A and, to a
lesser degree, P44A) were able to further prolong the lag phase of α-synuclein polymerization relative to WT CsgE. Interestingly, the mutants that most effectively prolonged the lag phase of α-synuclein polymerization were not the most effective inhibitors of CsgA polymerization, suggesting that CsgE may interact with α-synuclein by a different mechanism.

To follow up on these studies, the ThT assay should be repeated with a larger panel of mutants. Additionally, monitoring the reaction via atomic force microscopy as described by would likely shed light on the mechanisms of CsgA and α-synuclein inhibition, allowing us to better dissect the mechanisms by which proteinaceous inhibitors of amyloid assembly exert their effects (44). These studies could also be expanded to include tau and amyloid β, which are not inhibited by CsgC but can cross-seed CsgA amyloid formation.

2.6 Materials and Methods

2.6.1 Strains and Plasmids

6x-his-tagged CsgE expression constructs containing single amino acid substitutions were generated in the pNH27 plasmid (MC4100 CsgE in pTRC99a) using the QuikChange protocol (30, 40) and transformed into *E. coli* C600. For *in vivo* curli biogenesis studies, WT CsgE from *E. coli* MC4100 was cloned into pTRC99a encoding the pBR322 origin of replication. Mutations were generated using the QuikChange system. For bile salt sensitivity studies, mutations in CsgE were made in pTRC99a encoding CsgE and CsgG (pLR12) using the QuikChange protocol. Plasmids harboring mutations were transformed into MC4100ΔcsgE (LSR11) cells (24). 6x-his tagged CsgA lacking the 20-residue N-terminal signal sequence was expressed in the cytoplasm of NEBC2566 cells containing the pET11d-derived pNH11 plasmid (34).
2.6.2 NMR Spectroscopy

15N-labeled CsgE for NMR spectroscopy was generated as previously described using M9 minimal medium containing 1 g/L of 15N-ammonium sulfate (33). Deuterium oxide (D2O) and 15N-ammonium sulfate were obtained from Cambridge Isotope Laboratories. 80 µM CsgE (WT or W48A/F79A) were incubated in the presence or absence of 80 µM unlabeled CsgA for 24 hours. NMR spectra were collected on Bruker AVANCE III 600 MHz spectrometer equipped with a cryogenic triple resonance probe. The proton chemical shifts were internally referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS), and the chemical shifts of 15N were referenced indirectly to DSS using the absolute frequency ratios. All NMR data were processed using Bruker TopSpinTM 3.2. The chemical shift perturbation was analyzed using the program NMRFAM-SPARKY(32). The average Euclidean distance change was calculated as 

$$
\delta = \sqrt{\frac{1}{2} [\delta_H^2 + (\alpha \cdot \delta_N^2)]}
$$

where $\delta_H$ and $\delta_N$ are the absolute chemical shift of 1H and 15N, respectively; the scale factor $\alpha$ is 0.14.

2.6.3 Thioflavin T Assays

ThT CsgA polymerization assays were conducted as previously reported (8, 34). WT and mutant CsgE were purified from the lysate of E. coli C600 cells containing pNH27 by cobalt affinity chromatography (resin: Goldbio Product #H310). Following dialysis into 20 mM Tris pH 8.0, samples were run over an anion exchange column (Mono Q, GE healthcare) and eluted with an NaCl gradient. Samples were dialyzed into 50 mM KHPO4, pH 7.3 and stored at 4 °C. CsgA was purified from NEBC2566/pNH11 cells lysed in 50 mM KPi, pH 7.3 supplemented with 8 M guanidine hydrochloride for 24 hours. Cellular debris was cleared by centrifugation and the resulting supernatant was loaded onto an Ni-NTA column. CsgA was eluted in 50 mM K2HPO4, pH 7.3 + 125 mM imidazole and filtered through a 4-mL Ultracel 30-kD MWCO Amicon® Ultra
spin column (Millipore Ref# YFC803024) to remove aggregates and oligomers. Monomeric CsgA was buffer-exchanged into 50 mM K₂HPO₄, pH 7.3 using a 5-mL, 7K MWCO Zeba™ Spin Desalting Column (Thermo Scientific #89892). CsgA concentration was determined using the A₂₈₀ with a molar extinction coefficient of 11460 M⁻¹ cm⁻¹.

Purified CsgA protein was immediately added to the wells of 96-well black polystyrene nonbinding plates (Corning Ref#3650) to final concentration of 20 μM. CsgE was added to a final concentration of 5 μM, unless otherwise noted. Plates were incubated in a Tecan Infinite 200Pro plate reader for 48 hours at 25 °C. Reads were taking following 3 s of linear shaking every 20 minutes. The excitation and emission wavelengths were set to 438 nm and 495 nm, respectively.

### 2.6.4 Quantification of CsgA in Curli-Expressing Cells

MC4100ΔcsgE cells complemented with WT or mutant csgE on a pTRC99a plasmid were grown overnight in LB supplemented with 100 μg/mL ampicillin. 5 μL of each sample were spotted into a YESCA agar plate, which was incubated at 30 °C for 40 hours. Cells were scraped and resuspended in 1 mL of PBS. Cell density was normalized using optical density at 600 nm. 200 μL cells at an OD₆₀₀ of 1.0 were pelleted. Cells were then resuspended in 100 μL hexafluoro-2-pronanol (HFIP) to depolymerize preformed curli fibers. Samples were then dried in a rotary evaporator and resuspended in 100 μL SDS sample buffer (60 mM Tris pH 6.8, 5% β-mercaptoethanol, 10% glycerol, 3% SDS). Samples were run on a 15% SDS-PAGE gel, transferred to a Immobilon-PSQ PVDF membrane (Millipore Cat #ISEQ00010), and developed using mouse α-CsgA antibody. Quantification was performed by densitometry using the Quantity One image analysis software. Significance was computed relative to MC4100ΔcsgE complemented with WT pTRC99a-csgE on the appropriate control plasmid using the Mann-Whitney U-test.
2.6.5 Differential Scanning Fluorimetry

The thermal stability of CsgE was assessed using differential scanning fluorimetry. Purified CsgE was dialyzed into 50 mM KHPO$_4$, pH 7.3. In addition to Sypro® Orange Protein Gel Stain (Sigma #5692), each well contained 2 µg of WT or mutant CsgE and buffer with and without 150 mM NaCl to a final volume of 70 µL. Samples were heated from 20 °C to 100 °C in 30 s/0.5 °C increments using a Bio-Rad C1000 thermocycler with CFX96 RT–PCR attachment. The melting temperature was determined by fitting the melt curves to the Boltzman equation \[ y = A2 + (A1 - A2)/(1 + \exp(x - Tm)/dx) \] (41).

2.6.6 Bile Salt Sensitivity Assays

MC4100Δcsg cells were transformed with pTRC99a containing both WT csgG and WT or mutant csgE in the multiple cloning site under control of the lac operon (pLR12). Cells were grown to an OD of 0.1 in LB broth at 37 °C shaking and pelleted by centrifugation at 6500 rpm for 10 minutes. Cells were resuspended in 1 x PBS to an OD of 1, and 3 µL each of eight serial dilutions were plated on McConkey agar plates supplemented with 2% bile salts, 100 µg/mL ampicillin, and 0.1 mM IPTG.

2.6.7 Native-PAGE and Size Exclusion Chromatography

CsgE used in SEC and Native-PAGE analysis was expressed and purified as previously described (33). Briefly, WT and mutant CsgE were expressed from the pNH27 plasmid and subject to an initial cobalt affinity column. These samples were run over a HiPrep 26/60 Sephacryl S-100 HR column (GE Healthcare) and eluted with 50 mM potassium phosphate, 150 mM NaCl, pH 7.4. The peak corresponding to monomeric CsgE was isolated and used for subsequent SEC and Native-PAGE experiments. SEC analysis of CsgE oligomerization was performed using a Superdex 200 10/300 column (GE Healthcare) and a flow rate of 0.5 ml/min. The elution buffer
was 50 mM potassium phosphate, 150 mM NaCl, pH 7.4. SEC analysis performed at different
temperatures was slightly shifted in profile and peak intensity. For comparison, we only show SEC
analysis carried at room temperature (about 25°C) for CsgE pre-incubated at different
temperatures (4, 25 or 37 °C). Some degradation of CsgE was seen after 28 hours of incubation
at 25 °C.

2.6.8 SEC-MALS

Oligomerization of CsgE was analyzed by size-exclusion chromatography coupled with
multi-angle light scattering. WT and mutant CsgE were purified with two sequential cobalt affinity
columns and dialyzed into 50 mM KHPO₄, pH 7.3 with and without 150 mM NaCl and concentrated
to 2 mg/mL. Protein was injected into a Superdex™ 200 Increase 10/300 GL column (GE
Healthcare #28-9909-44) and run at a flow rate of 0.3 mL/min using an Agilent Technologies 1260
Infinity HPLC system. Light scattering was measured using the DAWN® HELEOS® II multi-angle
static light scattering detector (Wyatt technology) coupled to an Optilab T-rEX refractometer
(Wyatt Technology). Data were collected and analyzed using the ASTRA software.

2.6.9 Multisequence Alignment and Selective Pressure Analysis

To determine the evolutionary pressures acting on codons encoding individual amino acid
residues within the E. coli species and across all taxa of bacteria, a PHMMR (EMBL-EBI) search
was conducted against the “Ensembl Genomes Bacteria” database using an E-value cutoff of
1.0e-10. Nucleotide sequences were downloaded into two separate bins: one containing all CsgE
homologues, and one containing homologues found exclusively in E. coli.

For the E. coli-only search, 2002/2003 hits were downloaded from the EMBL-EBI
database and loaded into Geneious R10 (42). Duplicate sequences were removed to generate a
list of 93 unique E. coli sequences. Sequences with ORFs less than 300 nucleotides in length
were then removed to generate a list of 80 sequences. These sequences were then translated
into amino acid sequences and aligned using the MAFFT FFT-NS-I x2 algorithm with the BLOSUM62 scoring matrix (35, 43, 44). These amino acid alignments were then translated into codon-based nucleotide alignments using the Pal2Nal webserver (45). Recombination analysis was performed using the GARD method on the DataMonkey webserver (46-48). Positive/negative selection analysis was performed using the FEL algorithm using the GARD-generated phylogenetic tree and the HKY85 nucleotide evolution model (49).

2.6.10 Dynamic Light Scattering

WT CsgE was purified as described above, dialyzed into 50 mM KPi, pH 7.3, and concentrated to 5 mg/mL. Protein dilutions were made in 50 mM KPi, pH 7.3, immediately prior to analysis. 35 µL of protein were loaded into a 384-well plate. All concentrations were performed in triplicate. Protein was maintained at 4 °C during preparation. Readings were taken after 0, 2, 8 and 24 hour of incubation at 25 °C using a DynaPro® Plate Reader™ II (Wyatt). Data were analyzed using the DYNAMICS software.
2.7 Figures

**Figure 2.1.** Charged residues on CsgE cluster into three groups. (A) A positively-charged cluster of residues (R43, R47, K70, and R71) define the ‘head’ of the molecule, while negatively charged residues (E3, E5, D102, D106) are enriched at the disordered ‘tail’ of the molecule. (B) Primary sequence of CsgE, with positive, negative, and hydrophobic residues highlighted in blue, red, and green, respectively.
Figure 2.2. Chemical shift perturbations of $^{15}$N-labeled CsgE W48A/F79A by CsgA. Chemical shift changes of A) $^{15}$N, B) $^1$H, and C) the average change of Euclidean distance in ppm units. B) Structural mapping of chemical shift perturbations (CSP). Residues E31, S32, D33, Y34, S45, A46, R47, A48, R71, E85, and L88 have relatively high CSP, about 2-fold higher than the average shifts. These residues form two oppositely charged clusters in the structure: a negatively-charged region in the stem of the protein containing residues E31, D33 and E85, and a positively-charged region at the head of the protein containing R47 and R71. Positively-charged surface regions are colored blue, while negatively-charged regions are colored red.
Figure 2.3. The presence of CsgA has little effect on the backbone amide residues of CsgE.

Overlay of $^1$H-$^{15}$N HSQC spectra of $^{15}$N-labeled CsgE W48A/F79A (80 μM) in the presence (red) and absence (blue) of unlabeled CsgA (80 μM) for 24 hours. The green boxes illustrate a few residues (such as E31, D33, R71, A46 and E85) with relatively higher chemical shift perturbations than others upon the addition of CsgA.
Figure 2.4. HSQC spectra of WT CsgE in the presence and absence of CsgA. Overlay of $^1$H-$^{15}$N HSQC spectra of $^{15}$N-labeled CsgE WT (80 µM/0.4 M arginine) in the presence (red) and absence (blue) of unlabeled CsgA (80 µM). Residues E31, A46, R47 and R71 were assigned based on the chemical shift of the double mutant CsgE W48A/F79A.
Figure 2.5. Mutation of charged residues differentially affects the CsgE-CsgA interaction.

(A) 20 μM CsgA were incubated in the presence (solid line) and absence (dashed line) of CsgE at 25 °C. ThT fluorescence was measured every 20 minutes for 48 hours to determine the extent of CsgA fiber polymerization. Consistent with previous reports, the presence of CsgE disrupts the aggregation of CsgA into fibers. (B) Mutations were made to positively-charged and nonpolar residues at the head of CsgE and the negatively-charged E31 and E85 residues in the stem. Positively-charged residues are colored blue, negatively-charged residues are colored red, and uncharged residues are colored green. (C) Time necessary for CsgA to reach 50% of maximal aggregation (T_{50}) when incubated with each purified mutant, as measured by Thioflavin T fluorescence. The T_{50} of CsgA autopolymerization is decreased by over 50% in the CsgE^{R47A}
variant relative to WT CsgE, while mutation of nearby uncharged residues has a much more subdued effect. \textbf{(D)} Introduction of positively-charged residues near the tail of the CsgA molecule increases the $T_{50}$ of CsgA aggregation.
Figure 2.6. Thermal stability of select CsgE mutants. Thermal stability was assessed via differential scanning fluorimetry using the Sypro Orange hydrophobic dye. $T_m$ was determined by identifying the inflection point from the sigmoidal region of the melting curve.
Figure 2.7. Quantification of curli fiber formation on the bacterial surface. Curli fiber formation was measured by Western blot densitometry following treatment of cells with HFIP to depolymerize curli fibrils.
Figure 2.8. Introduction of select point mutations has minimal effects on CsgE stability in vivo. Cells harboring the pLR12 plasmid were induced with IPTG for 30 minutes. 100 µL of cells normalized to an OD600 of 1.0 were pelleted and resuspended in 2x SDS sample buffer. α-HA and α-CsgG antibodies were used to detect the CsgE and CsgG proteins, respectively.
Figure 2.9. Mutation of tail residues does not abrogate the CsgE-CsgG interaction. (A) MC4100Δcsg cells overexpressing CsgG from pMC1 were complemented with pTRC99a encoding WT and mutant CsgE. Each variant tested, including negatively-charged residues in the N- and C-terminal IDRs, was able to rescue cells from CsgG-mediated toxicity. (B) CsgE D102K, but not CsgE E3K, blocks curli assembly on the outer membrane of the cell.
Figure 2.10. Size exclusion chromatograms of CsgE oligomerization. Representative results of purified CsgE (20 µM) incubated as a function of time are shown for three temperatures; 4 °C (A), 25 °C (B), and 37 °C (C). (D) Relative distribution of each species at both 25 °C and 37 °C as a function of time.
Figure 2.11. CsgE W48A/F79A exists as a monodisperse monomer. Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) elution profiles of WT CsgE (solid black) and CsgE W48A/F79A (solid blue) from a Superdex™ 200 Increase 10/300 GL column. Calculated molecular masses from each peak of each trace are shown as dotted lines. WT CsgE elutes in four peaks; a large, high-molecular weight aggregate representing 32% of the loaded mass, a small intermediate species corresponding to a small oligomer (2% of the mass), and two lower-molecular weight species with calculated masses of 13.2 and 19.6 kDa, respectively. By contrast, the CsgE W48A/F79A is present almost exclusively as a monomer, with a calculated molecular weight of 14.5 kDa.
Figure 2.12. Mutations to aromatic head residues alter CsgE oligomerization kinetics (A)

Despite the observed shifts on the Native-PAGE gel, both the single and double mutants appear as monomers on a Superdex 200 10/300 column. B) Native-PAGE analysis demonstrates that WT CsgE is primarily present as a nonamer at 4°C and 25°C, but begins to form higher-molecular-weight oligomers following incubation at 37°C. (C) CsgE W48A/W79A remains exclusively monomeric at all temperatures and incubation times tested. (D) CsgE F79A is present as both a monomer and nonamer at 25°C, with a pronounced shift to higher oligomers at 37°C. (E) W48A remains a mixture of monomers and nonamers during short incubation times at 25°C and 37°C, with the emergence of the higher molecular weight species only occurring after prolonged incubation at 37°C.
Figure 2.13. Residues throughout the CsgE molecule are subject to selective pressures.

(A) Residues subject to negative selection occur preferentially on the core and at the head of the CsgE protein. Bars shown in dark blue correspond to strong purifying selection ($P<0.05$), while bars shown in light blue correspond to $0.10 > P > 0.05$. (B) Residues that are more conserved across Gram-negative bacterial species are shown in red, while residues that are less conserved are shown in blue.
Figure 2.14. Dynamic light scattering of WT CsgE. Histograms of molecular radii of WT CsgE at different protein concentrations in 50 mM KPi, pH 7.3 immediately after dilution. All experiments were performed in triplicate; representative distributions are displayed. Readings were also performed following incubation at 25 °C for 2, 4, 8, and 24 hours (not shown).
Figure 2.15. UV tracings from size-exclusion elution profiles. CsgE point mutants elute at different times in the presence and absence of NaCl.
Figure 2.16. CsgE point mutants differentially inhibit \(\alpha\)-synuclein polymerization. CsgE were incubated at a 1:4 molar ratio with purified \(\alpha\)-synuclein for 70 hours at 37 ° in the presence of a glass bead and constant aggregation.
Chapter 3: Optimizing Glycomimetic Antiadhesives for Treatment of UTIs

3.1 Scope of Work

Of the work described in this chapter, I was responsible for the crystallization of FimHLD with ZFH269 and ZFH284, and for solving the FimHLD-ZFH269 crystal structure. I was also responsible for generating co-crystals FmlHLD in complex with AM2050 and AM2119, and for solving the FmlHLD-AM2050, -AM2119, and -AM2134 crystal structures and the subsequent data analysis. Portions of this chapter are reproduced from:


©2016, American Chemical Society

3.2 Introduction

3.2.1 Clinical Characteristics of Urinary Tract Infections

Urinary tract infections (UTIs) are common, recurrent, and costly bacterial infections that have a profound impact on public health. In the United States alone, UTIs result in over 10 million office visits and 2-3 million emergency department visits annually and contribute nearly $2.5 billion to direct and indirect healthcare costs (97, 98). Over 60% of females with be diagnosed with a UTI in their lifetime (99). Although the annual incidence of physician-diagnosed UTI is greater than 10% for females and 3% for males (100), individual risk depends on a variety of factors,
including age, sexual activity, family history, medical comorbidities, and an individual's history of UTI (101). The propensity for infection to recur despite appropriate antibiotic treatment also contributes significantly to the incidence and morbidity of UTI. Depending on the presence of behavioral and physiological risk factors, more than 30% of women will suffer one or more additional infections within 12 months of an initial UTI despite appropriate antibiotic therapy to clear the initial infection (102). In one study of postmenopausal women, almost 70% of individuals receiving appropriate antibiotic therapy for an initial UTI experienced at least 1 recurrence within 12 months, with a mean number of recurrences of 2.9 (103).

Clinical classification of UTIs is made based on the affected organ systems, duration of symptoms, and patient characteristics. In the overwhelming majority of UTIs, pathogens enter the urinary tract through the urethral meatus before ascending up the urethra and into the bladder lumen. Isolated infections of the bladder are referred to as cystitis. Acute cystitis in nonpregnant, premenopausal patients without functional urinary tract abnormalities is classified as “uncomplicated”, while cystitis in pregnant patients, those with comorbidities (functional urinary tract abnormalities, history of renal transplant, being immunocompromised), or those with involvement of other organ systems is classified as “complicated” (104). Common symptoms of cystitis include painful or difficult urination (dysuria), elevated frequency of urination, urinary urgency, and suprapubic pain (105). In approximately 0.34% of cases, pathogens causing cystitis ascend further through the ureters into the kidney, where they cause an infection termed pyelonephritis (106). These are always classified as “complicated” UTIs. Clinical signs and symptoms suggestive of pyelonephritis include: i) flank pain; ii) costovertebral angle tenderness; iii) fever > 99.9 °F; and iv) other signs or symptoms of systemic illness, including chills, fatigue, or nausea/vomiting. Pyelonephritis, if left untreated, may progress to bacteremia and, in rare cases, septicemia (107).
Although the most common causative agent for community acquired UTIs is uropathogenic *Escherichia coli* (UPEC), UTIs can be caused by a variety of Gram-positive and Gram-negative bacterial pathogens, with rates dependent on the site of acquisition and the presence of medical comorbidities. Approximately 75% of uncomplicated UTIs are caused by uropathogenic *Escherichia coli* (UPEC), 6% are caused by *Klebsiella pneumoniae*, 6% by *Staphylococcus saprophyticus*, 5% by *Enterococcus* spp, and 6% others (108). In contrast, only 65% of complicated UTIs are caused by UPEC, while 11% are caused by *Enterococcus* spp, 8% by *K. pneumonia*, and 7% by *Candida* spp (108). Healthcare-associated UTIs, 80% of which are catheter-associated UTIs (CAUTIs), are more likely to progress to pyelonephritis and urosepsis than their community-associated counterparts (109). These sequelae are a major source of morbidity and mortality. In one study from the UK’s National Health Service, pyelonephritis and CAUTI combined to account for over 66% of all reported Gram-negative bacteremia episodes in the hospital (107). Further, mortality within patients suffering from *E. coli* bacteremia ranges between 5% and 30% over a 30-day period (110) and in the United States, urosepsis accounts for approximately 25% of all sepsis cases, and claims in excess of 55,000 lives annually (111).

### 3.2.2 Antibiotic Resistance and UTIs

The rapid emergence of antibiotic resistance among UTI pathogens mirrors that of many other bacterial infections, and is predicted to have dire economic consequences worldwide (10). Further, antibiotic use for the treatment of UTI contributes significantly to resistance, as 9% of all antibiotics prescribed in an ambulatory setting are prescribed to treat UTIs (112). Presently, over 6% of *Enterobacteriaceae* clinical UTI isolates are multidrug resistant, and 1% of all isolates are extensively drug-resistant (113). Pathogens that cause healthcare-associated UTI are more diverse, and are associated with higher rates of antibiotic resistance (114). Extended-spectrum beta lactamase production and fluoroquinolone resistance continue to emerge among
uropathogens and have begun to severely limit options for empiric therapy (115, 116). More recently, the worldwide dissemination of plasmids encoding colistin resistance has underscored the urgency of the antibiotic resistance crisis (117, 118).

Diagnosis of an acute, uncomplicated UTI is often made based on clinical suspicion alone, although a thorough history and physical are helpful to exclude complicated UTI or alternative diagnoses. While urine cultures are preferred, empiric therapy in the absence of culture results is often prescribed based on risk of infection with a multidrug-resistant (MDR) Gram-negative pathogen. When MDR risk is low, nitrofurantoin (100 mg BID PO x 5 days), trimethoprim-sulfamethoxazole (TMP/SMX, 160/800 mg BID PO x 3 days), and fosfomycin (3g 1x PO) are the preferred first-line agents (98, 104, 105). Urine culture and antibiotic susceptibility testing are indicated in patients with one or more risk factors for infection with an MDR organism. Risk factors include: i) extended stays in healthcare facilities; ii) recent travel to areas with high MDR UTI prevalence, including India, Israel, Spain, and Mexico; and iii) recent history of antibiotic use, including fluoroquinolones, TMP-SMX, or third+ generation cephalosporins. For patients with an elevated risk of MDR UTI, oral beta-lactams and fluoroquinolones are recommended as empiric treatment, with targeted therapy based on susceptibility testing being the ultimate goal.

Diagnosis of a complicated UTI requires a thorough physical examination, urinalysis, and culture with susceptibility testing. Outpatients with no risk factors for infection with MDRs (especially ESBL isolates) can be treated empirically with ciprofloxacin (500 mg BID PO for 6-7 days) or levofloxacin (750 mg QD PO 5-7 days) (119). In patients with allergies or at risk for a fluoroquinolone-resistant UTI, ceftriaxone or ertapenem (each 1g IV) followed by TMP-SMX, amoxicillin-clavulanate, cefodoxime, cefdinir, or cefadroxil are preferred. In patients with a high risk for an MDR infection, an initial dose of ertapenem (1g IV) followed by fluoroquinolones (ciprofloxacin 500 mg BID 5-7 days) is preferred. Hospitalization is indicated when patients
experience a high fever (>101 °F), pain, inability to maintain oral hydration, or show signs of systemic illness.

Although they are a vital component of our therapeutic arsenal, treatment with broad-spectrum antibiotics is not without risks. In addition to driving resistance, antibiotic treatment can lead to disruption of the beneficial host microbiota, increasing the risk for opportunistic organisms to expand within a given niche and cause infection. In the gastrointestinal tract (GI) tract, antibiotic consumption increases inflammation, weakens the host immune response, and encourages the preferential proliferation of *E. coli* by increasing nitrate availability (120). Because the GI tract is the ultimate source for UPEC that cause cystitis, expansion of *E. coli* populations within the gastrointestinal niche is also associated with increased risk of subsequent UTI recurrence (121, 122). Thus, to counter the growing threat of antimicrobial resistance and spare the beneficial microbiota from the detrimental effects of broad spectrum antibiotics, a concerted effort is necessary to optimize the usage of existing antimicrobial agents while developing new, targeted alternative agents. A detailed mechanistic understanding of UTI pathogenesis is crucial for our ability to develop such novel agents to effectively treat UTI in the future.

### 3.2.3 Determinants of UTI Pathogenesis

Adhesion of UPEC to the bladder epithelium is a critical step in the pathogenesis of UTI, allowing UPEC to resist the flow of fluids in the urinary tract to maintain colonization of urinary mucosal tissues. During acute cystitis, this adhesion is mediated by the type 1 chaperone-usher pathway (CUP) pilus adhesin, FimH (26, 33). CUP pili are proteinaceous extracellular appendages expressed by Gram-negative bacteria to facilitate colonization and persistence in a variety of host and environmental habitats. Many CUP pili, including type 1, are composed of a helical rod containing hundreds of copies of the major pilus subunit joined to an adhesive tip structure, including a distal two-domain adhesin comprising an N-terminal ligand binding domain
and a C-terminal pilin domain (Figure 3.1A) (18). The N-terminal domain of the adhesin typically contains a binding pocket that recognizes its molecular receptor(s) with stereochemical specificity, thus determining the tissue tropism of the pilus (26). The FimH adhesin on type 1 pili binds mannosylated uroplakins on the bladder epithelium (123). Following adhesion, UPEC are internalized into the superficial umbrella cells of the bladder, where they are protected from the host immune response and can replicate to form intracellular bacterial communities (IBC). The bacteria within these communities then flux out of their host cells, re-entering the bladder lumen and beginning the IBC cycle anew (22). UPEC isolates lacking the type 1 pilus operon are unable to cause UTI in murine models of infection.

In a murine models of cystitis, infection with UTI clinical isolates results in two distinct outcomes: i) spontaneous symptomatic resolution and bacterial clearance; and ii) chronic, high-titer bacteriuria that does not self-resolve in the absence of antibiotics (124). Bladders of mice experiencing chronic cystitis show markers of chronic inflammation, necrosis, and urothelial hyperplasia. This inflammatory remodeling leads to the elaboration of glycan moieties that are not present in healthy bladders (32). This change is exploited by UPEC, which can express F9 pili, also known as fml (fim-like) pili, that are structurally similar to type 1 pili but recognize terminal galactose and N-acetylgalactosamine (GalNAc) residues found in core-1 and -2 O-glycans exposed by inflammation during chronic cystitis (32). Fml pili can also recognize Thomsen-Friedenreich (TF) antigen (Gal[β1-3GalNAc]) on the kidney epithelium to facilitate kidney infection. Deletion of fmlH attenuates clinical UPEC isolates in mouse models of chronic UTI, while complementation of fmlH restores urovirulence.

Because of the crucial role type 1 and fml pili play in UPEC pathogenesis, they have been a subject of intense therapeutic investigation. Vaccination with the FimCH chaperone-adhesin complex can effectively prevent and treat both acute and chronic UTI in murine models of cystitis.
This efficacy has also been demonstrated in humans during a recent Phase 1 clinical trial in which women with a recent history of UTI experienced a 72% reduction in total UTI recurrences following administration of the FimCH vaccine. Based on this data, the vaccine is allowed by the FDA for compassionate use in patients experiencing severe episodes of recurrent UTI recalcitrant to traditional antibiotic therapy (Sequoia Biosciences, unpublished). Vaccination against the FmlH pilus adhesin is also protective against establishment of chronic cystitis in murine models. A second therapeutic approach is the competitive inhibition of type 1 and fml pilus binding to their natural receptors on the bladder epithelium with small molecule glycomimetic compounds. These compounds can effectively reduce the bladder UPEC burden in murine models of acute and chronic cystitis (34, 125). However, further improvement of the biochemical and pharmacokinetic properties of glycomimetic compounds for use as a therapeutic agents requires a detailed understanding of the structural basis of their interaction with each adhesin. The interactions for FimH and FmlH are described below.

3.2.4 The Structural Basis for Mannose Recognition by FimH

The structural basis of mannose recognition by the FimH lectin domain (FimH\textsuperscript{LD}) has been known for some time. The mannose-binding site is located in a deep pocket formed by residues from three loops (N46, D47, D54, Q133, N135, and D140) at the distal end of FimH\textsuperscript{LD} and the N-terminal tail (F1) of the mature protein (Figure 3.1B) (123). FimH’s affinity for mannosylated ligands is regulated in part by the conformational state of the FimH lectin domain relative to the pilin domain (29). Full-length FimH can sample a range of conformational ensembles when placed in a tip-like setting. In the high-affinity “relaxed” state, the three mannose binding loops and core \(\beta\)-sandwich fold are tightly packed, allowing the hydroxyl groups on mannose to form more favorable interactions with the backbone amide of Phe1 and the sidechains of N46, D47, D54, Q133, N135, and D140 (Figure 3.1B). Conversely, when FimH adopts the tense conformation,
the binding loops are displaced due to widening of the core β-sandwich fold, decreasing their affinity for mannose (Figure 3.1C). The ability of FimH to sample these conformational states is influenced by the identity of key residues that are positively selected in clinical UPEC isolates (29, 31).

Despite its ability to inhibit FimH adhesion to mannosylated glycoproteins in vitro, α-D-mannose alone is insufficient to prevent clinical UTI in human or murine models of acute cystitis. As such, novel compounds that display an increased in vitro activity against FimH were developed to improve in vivo therapeutic outcomes (126, 127). To this end, a structure-based design approach was used to develop new, high-affinity mannosides that exploit interactions with hydrophobic residues at the edges of the binding pocket, especially Y48 (Figure 3.1D) (128, 129). These compounds included butyl mannose, alkyl-phenyl mannosides, and eventually, biphenyl mannosides. The ability of novel compounds to inhibit FimH-mediated agglutination of guinea pig red blood cells, which display mannosylated ligands on their surface, is used as an initial screening tool for potency. Thus, the in vitro potency of mannoside compounds is typically reported as the concentration of compound required to achieve 90% hemagglutination inhibition (HAI). The improved potency of biphenyl mannosides stems from the ability of the second phenyl ring to form key π-stacking interactions with the Y48 “gate” adjacent to the mannose-binding site. Addition of a methylester group at the meta position of the second phenyl ring created a strong hydrogen bond between the carbonyl group and R98, resulting in an HAI of 1 µM (approximately 18,000 x lower than α-D-mannose alone) (129).

3.2.5 The Structural Basis for Galactose Recognition by FmlH

Although global phylogenetic analysis of all pilus ushers reveals a close evolutionary relationship between fml and type 1 pili, the primary sequences of FmlHLD and FimHLD are only 44.9% identical (32). X-ray crystal structures of apo- and receptor-bound FmlH have illuminated
the molecular basis of receptor recognition by FmlH^{LD}. Despite the divergence of their primary sequences, the core β-sandwich folds of FimH^{LD} and FmlD^{LD} are highly conserved (Figure 3.2A). Crystallization of FmlH^{LD} in complex with TF reveals that hydroxyl groups on the galactose ring form an extensive direct and water-mediated hydrogen bonding network with the backbone of F1 and the side chains of D45, D53, K132, and N140 (Figure 3.2B) (32). The functions of the latter four residues are analogous to the role of D7, D54, Q133, D140 in FimH^{LD} binding to mannose. Additionally, R142 forms a hydrogen bond with the carbonyl oxygen on the N-acetyl group on GalNAc. Unlike FimH, residues in loop 1 do not interact with the sugar ring.

Studies by the Hultgren and Janetka labs have utilized iterative rounds of *in silico* docking analysis, X-ray crystallography, and directed organic synthesis to develop potent inhibitors of FmlH binding (125). These studies ultimately culminated with the synthesis of 29β, a galactose ring fused to a phenyl ring ortho-substituted with benzoic acid, and 29β-NAc, a fusion of GalNAc ring, a phenyl ring, and benzoic acid (Table 1). This ring configuration is different than that of the biphenyl mannoside compounds, in which the second phenyl ring is linked to the para position of the first ring relative to the mannose moiety. The potency of these compounds was assessed using a competition-based enzyme-linked immunosorbent assay (ELISA) in which FmlH binding to desialylated bovine submaxillary mucin (BSM) was quantified in the presence and absence of galactoside. 29β and 29β-NAc were able to achieve 50% inhibition of FmlH binding at a concentration (IC50) of approximately 8 µM and 650 nM, respectively (125). Co-crystallization of FmlH^{LD} with 29β-NAc revealed that phenyl ring one forms a face-to-edge π stacking interaction with Y46, while the meta-substituted carboxylic acid group forms hydrogen binding interacts with R142 and the backbone of F1. Transurethral administration of 29β-NAc compounds to mice experiencing chronic cystitis successfully reduced UPEC bacterial burdens in the kidney and bladder (125). Upon co-administration with mannoside 4Z269, further reductions in bladder and
kidney titers were observed, suggesting the possibility for synergistic combination treatment as an antibiotic-sparing therapeutic approach against UPEC UTI.

3.3 Elucidating the Structure-Activity Relationship of Novel Glycomimetics

3.3.1 X-ray Crystal Structure of ZFH269 in Complex with FimH^{LD}

To further optimize the binding affinity of mannosides to FimH, two novel compounds were synthetized by Han, et al. (34, 129). The first, ZFH269, contains an ortho methyl group on the first biphenyl ring and an amide at the meta position of the second phenyl ring. ZFH269 has HAI of 62 nM, which represents a 16-fold improvement over the methylester (Figure 3.3A). The second, ZFH284, contains a second meta-substituted amide on the second ring has an HAI of 16 nM (Figure 3.3B). To better understand the molecular interactions that underlie these increases in potency, I generated crystals of ZFH269 and ZFH284 bound to FimH^{LD}. I solved the structure of ZFH269 to 1.75 Å resolution (PDB ID: 5F3F, Figure 3.3C), while ZFH284 was solved by Vasilios Kalas to 1.67 Å resolution (PDB ID 5F2F, Figure 3.3D). As observed in previous structures, the mannose moiety forms an extensive hydrogen bonding network with residues in loops 1-3. Additionally, the hydroxyl group from the O-glycosidic bond forms water-mediated hydrogen bonds with Asn138 and Asp140, and the second phenyl ring forms face-to-face π stacking interactions with Tyr48.

I also observed that the ortho-methyl substituent on first phenyl ring of ZFH269 occupies a small hydrophobic pocket formed by I52, Y137, and N138. This results in a slight twist in the second phenyl ring that strengthens the π-π stacking interactions with Tyr48. The meta-substituted amide on the second phenyl ring occupies two distinct conformations with roughly
equal occupancy in which methyl group of the amide is flipped approximately 180 degrees relative to the nitrogen. In both conformations, the nitrogen forms a hydrogen bond with the side chain of R98, while the carbonyl oxygen forms a hydrogen bond with the hydroxyl group of Tyr48. In the ZFH284 structure, the meta-amide oriented upward forms a water-mediated hydrogen bond with the hydroxyl group of Tyr137, thus accounting for the higher affinity.

3.3.2 X-ray Crystal Structure Three Biphenyl Galactosides in Complex with FmlH^{LD}

Using the biphenyl scaffold as a basis for further functionalization, the Janetka lab generated a series of compounds with a variety of substituents on the first and second phenyl rings designed to maximize interactions with FmlH and increase potency (Table 1). They synthesized 29β and 29β-NAc derivatives containing a nitro at the meta position of the second phenyl ring (AM2050 and AM2119, respectively). They also replaced the carboxylic acid in 29β and 29β-NAc with a sulfonamide group, resulting in AM2043 and AM2038, respectively. The meta position of the first phenol ring on AM2038 was further functionalized with methyl- (AM2110), fluoro- (AM2109) and trifluoromethyl- (AM2134) substituents. Finally, the sulfonamide methyl group of AM2110 was replaced with a cyclopropyl group to create AM2151. All sulfonamide-substituted biphenyl GalNAc derivatives had IC50s of 272 nM or below, and substitution with any functional group at the meta position of the first phenyl ring further improved the IC50 to 151 nM or below. AM2134 exhibited the highest potency of the compounds measured, with an IC50 of 85 nM.

To determine the structural basis of galactoside binding, we selected three of the novel biaryl galactosides for further structure-activity analysis using X-ray crystallography. The first, AM2134, was selected for its superior potency relative to the other compounds. I solved the structure of this compound in complex with FmlH^{LD} to 1.75 Å resolution (Figure 3.4B). As
previously observed in the 29β-NAc-FmlH\textsuperscript{LD} co-crystal structure, the terminal galactose ring of each compound forms hydrogen bonds with the amide backbone of F1, the side chains of D45, Y46, and D53 in loop 2, and the side chains of K132 and N140 in Loop 3. The nitrogen in the N-acetylglucosamine group forms hydrogen bonds with K132 and a water present in the binding pocket. We also observed an additional water-mediated hydrogen bond between the N-acetylglucosamine carbonyl and R142 that had not been previously appreciated. The sulfonamide nitrogen atom forms a hydrogen bond with the backbone hydroxyl group of F1. Additionally, one of the sulfonamide oxygens interacts with the side chain of S2, the side chain of S10 side chain and backbone of I11 in loop 1. The addition of the meta-trifluoromethyl group to the first phenyl ring likely locks the position of the second phenyl ring at an 80 degree offset relative to the first ring, providing a favorable entropic contribution to binding. Additionally, one of the fluorine atoms interacts directly with D45 and indirectly with S2 through a water molecule.

AM2050 and AM2119 were chosen as crystallography candidates to better understand the influence of the N-acetyl group on potency. In most cases, removal of the N-acetyl group decreases the potency of the compound fivefold (29β-NAc vs 29β, AM2038 vs AM2045). However, removal of the N-acetyl group from AM2119 to create AM2050 increased the potency fourfold. To determine the structural basis for this finding, I solved crystal structures of FimH\textsuperscript{LD} in complex with AM2050 and AM2119 to 1.39 Å and 1.31 Å resolution, respectively (Figure 3.4C and D). In both structures, the nitro oxygens on the second phenyl ring form two interactions with R142, while the carboxyl oxygens interact with the side chain of S2 and the hydroxyl group on the backbone of S2. In AM2119, one nitro oxygen resides within 3.2 Å of the N-acetylglucosamine carbonyl, causing the second phenyl ring to tilt 45° relative to the plane of the first phenyl ring. In contrast, the angular offset between the plane of the two rings is 32.5° in the AM2050.
3.4 Discussion and Future Directions

High-affinity mannose and galactose analogues show great promise in the treatment and prevention of acute and chronic UTI through the inhibition of CUP pilus binding to glycoprotein ligands in the urinary tract. A detailed structural understanding of the adhesin binding pocket and surrounding residues is crucial to the development of high-affinity glycomimetic compounds that utilize multiple intermolecular contacts with the protein. For example, tyrosine residues near the active site of FimH and FmlH (Y48 and Y46, respectively) have been exploited to increase affinity via π-π stacking interactions with phenyl rings added to the core sugar moiety (34, 125, 129). In both cases, functionalization of a biphenyl scaffold was able to increase the potency and maximize the therapeutic potential of each compound. Mannosides ZFH269 and ZFH284 were designed to optimize the π-π stacking interactions of the second phenyl ring with Y48 while increasing the number of polar contacts with neighboring residues. Using X-ray crystallography, I found that the ortho-methyl substituent on first phenyl ring of ZFH269 occupies a small hydrophobic pocket formed by I52, Y137, and N138, optimizing the ring geometry to facilitate stronger π stacking interactions. I also found that the nitrogen and carbonyl oxygens on the amide group of the second ring form hydrogens bonds with R98 and Tyr48, respectively. These interactions improved the HAI titers of ZFH269 and ZFH284 to 62 nM and 16 nM, respectively. Subsequent studies demonstrated that these compounds can more effectively decrease the bacterial burden in murine models of cystitis and gastrointestinal models of UPEC colonization.

Similar principles were applied to the successful synthesis and crystallization of novel high-affinity biphenyl galactose analogues. The most potent of these compounds, AM2134, contains a sulfonamide moiety in the meta position on the second phenyl ring that engages in novel polar contacts with loop 1 of FmlH. Additionally, the trifluoromethyl group at the meta position of the first phenyl ring interacts with the side chain of D45, stabilizing the ring structure
and providing positive entropic contributions to binding. Together, these additional interactions confer a 5-fold improvement in potency relative to the former lead compound, 29β-NAc. Further improvements to galactoside potency could include the replacement of the trifluoromethyl group with a potent electron donor, such as an acetyl group, to exploit the interaction with D45.

Although the O-linked biphenyl mannosides and galactosides demonstrate strong efficacy in in vivo models of acute and chronic cystitis, their therapeutic potential is limited by their urine bioavailability and half-life. For example, galactoside AM2134 possesses poor pharmacokinetic characteristics, with a serum half-life of only 70 minutes and an oral bioavailability of only 0.29% (Figure 3.5A.) Modest improvement to these parameters was observed upon replacement of the sulfonamide methyl group with a cyclopropyl group and replacement of the trifluoromethyl group with a methyl group (Figure 3.5B). Efforts are currently underway to improve these pharmacokinetic parameters further using established principles of drug design. To this end, two modifications were made to mannoside ZFH269: i) replacement of the hydroxyl groups on the mannose ring with acetyl groups to create a prodrug with improved urine bioavailability and ii) elimination of the O-glycosidic bond, which can hydrolyze during passage through the gastrointestinal tract and lead to decreased oral bioavailability. Both of these modifications significantly improved plasma and urine bioavailability (Figure 3.6A), resulting in improved performance as a therapeutic in a murine model of acute UTI (Figure 3.6B). Successful synthesis of these compounds and testing in murine models of chronic cystitis are necessary predecessors for phase 1 clinical trials to evaluate the safety and optimize the dosing schedule in human subjects.

One outstanding question generated by this structural work is the seemingly contradictory deleterious effect of the N-acetyl group on AM2119 potency. Typically, comparison of compounds containing GalNAc to their non-N-acetylated counterparts results in a drastic loss of potency.
However, in the case of AM2050 and AM2119, this trend is reversed. Although the presence of the N-acetyl group on the galactose ring did alter the tilt of the second phenyl ring relative to the first phenyl ring by approximately $12^\circ$, no readily apparent cause for this profound loss of potency was observed. Interactions between the meta nitro group and R142 were unaffected by the differential ring tilt. Similarly, interactions between the carboxyl group and the side chain of S2 were also unaffected. The only moderate difference observed was a shortening of the distance between the second carboxyl oxygen and the backbone of S10 and I11 in loop one in AM2050 relative to AM2119. While it is possible that these effects result in a moderate decline in potency, further study will be necessary to elucidate the full mechanism of reduced potency. To this end, 29β and 29β-NAc derivatives containing only the meta nitro group will be synthesized and tested for potency in vitro. If the nitro group preferentially decreases potency in the presence of an N-acetyl group, that compound will be crystallized with FmlH$^{LD}$ to see if the intramolecular interaction between these two groups persists even in the absence of the carboxylic acid.

Taken together, structural studies of biphenyl mannose and galactose analogues have highlighted key principles in the synthesis of high-affinity inhibitors of protein binding. When possible, compounds should be designed to exploit contacts with nearby residues beyond the confines of the canonical active site. Further, polar and hydrophobic interactions can be used to optimize ligand geometry to maximize the strength of existing interactions. Finally, known methods of improving pharmacokinetic properties of small molecules can drastically improve the bioavailability and half-life of potent lead compounds, thereby maximizing the therapeutic potential of compounds with high in vivo potency.
3.5 Materials and Methods

3.5.1 Protein Expression and Purification

FimH lectin domain (FimH\textsuperscript{LD}; residues 1-158) from \textit{E. coli} strain J96 was expressed and purified as described previously and dialyzed into 10 mM MES pH 5.8, 50 mM NaCl (129). FmlH protein used in crystallographic studies was expressed and purified as previously described (125). Protein used in ELISA assays was biotinylated using an NHS-PEG4-Biotin and Biotinylation Kits (ThermoFisher).

3.5.2 Enzyme-Linked Immunosorbent Assays

Enzyme-linked immunosorbent assays (ELISAs) were used to quantify the IC\textsubscript{50} of different galactoside compounds as previously described (125). Briefly, 1 µg bovine submaxillary mucin (Sigma) in 100 µL PBS were incubated with Immulon 4HBX 96-well plates overnight prior to treatment with 1 mU \textit{Arthrobacter ureafaciens} sialidase for 1 hour at 37 °C to remove terminal sialic acid sugars. Wells were then blocked with 200 µL PBS + 1% BSA for 2 hours at room temperature. Biotinylated FmlH\textsuperscript{LD} was diluted to 20 µg/mL in blocking buffer and incubated in the presence or absence of galactoside compounds serially diluted 2x down eight rows for 1 hour at room temperature. Wells were washed three times with PBS 0.05% TWEEN-20 then incubated with 100 µL of streptavidin-HRP conjugate (BD Biosciences; 1:2,000 dilution in blocking buffer) for one hour. After three additional PBS +0.05% TWEEN washes, plates were developed with 100 µL of tetramethylbenzidine (BD Biosciences) substrate and quenched with 50 µL of 1 M \textit{H}_2\textit{SO}_4. Total bound portion concentration was measured by the absorbance at 450 nm. IC\textsubscript{50}s were determined using the Graphpad Prism software.
3.5.3 Structure Determination

To generate crystals of FimH\textsuperscript{LD} bound to ZFH259, FimH\textsuperscript{LD} was pre-incubated with ZFH269 at final concentrations of 1.27 μM (approx. 21 mg/mL) and 1.7 μM, respectively (a 1.4-fold molar ratio). 100 nL + 100 nL drops were set up using a TTP Labtech mosquito\textsuperscript{®} crystal robot using 70 μL mother liquor (0.2 M AmSO\textsubscript{4}, 20% PEG 3350.) Cuboidal crystals appeared after 72 hours of incubation at 18 °C. After approximately two weeks, crystals were transferred to mother liquor supplemented with 20% v/v glycerol as cryoprotectant and flash frozen in liquid nitrogen prior to storage and data collection. Diffraction data were collected at 100°K. X-rays were generated by a Rigaku MicroMax 007 generator coupled to a Rayonix Marmux X-ray source (Evanston, IL.) Images were collected on a Mar345 image plate detector. Data were indexed and integrated in iMosflm and scaled by Scala. The space group was then determined by POINTLESS. Phases were obtained by molecular replacement using PHASER and a .pdb of a previously solved FimH\textsuperscript{LD} structure (PDB ID 3MCY) stripped of all waters, additives, and ligands. The Fo-Fc map unambiguously indicated electron density corresponding to mannoside ZFH269 in two distinct binding conformations within the ligand-binding pocket. Chemical restraints on ZFH269 were generated by eLBOW and used in subsequent rounds of refinement with phenix.refine from the PHENIX package. Space group, collection, and refinement statistics for ZFH269 are listed in Table 2.

All FmlH\textsuperscript{LD} solutions were generated by adding 10 μL 50 mM galactoside dissolved in 100% DMSO to FmlH in 10 mM HEPES pH 7.5 + 50 mM NaCl immediately before setting up drops for a final concentration of 9 mg/mL FmlD\textsuperscript{LD}, 5 mM galactoside, and 10% DMSO. Co-crystals of FmlH\textsuperscript{LD}-AM2134 were grown by mixing 1 μL protein solution with 1 μL motherliquor containing 0.7 M LiSO\textsubscript{4} + 20% PEG 8000 on a glass coverslip over 1 mL motherliquor. Thin, needle-like crystals appeared after approximately 72 hours. Crystals were cryoprotected in 1 M
LiSO₄ + 20% PEG 8000 + 25% glycerol for 10 seconds before and flash-freezing in liquid nitrogen. Crystals of FmlH⁹⁻AM2119 were grown by mixing 1 μL protein solution (9 mg/mL FmlH⁹⁻, 5 mM galactoside AM2119, 9 mM HEPES pH 7.5, 45 mM NaCl) with 1 μL 0.1 M Tris 8.0 + 0.8 M AmSO₄ using the hanging drop vapor diffusion method. Square pyramidal crystals began appearing after approximately 24 hours and continued to grow for seven days. Crystals were harvested after 10 days, cryoprotected in a solution containing 0.1 M Tris 8.0, 0.8 M AmSO₄, and 30% glycerol for 10 seconds, and flash-frozen in liquid nitrogen.

Crystals of FmlH⁹⁻AM2050 were generated by soaking existing FmlH⁹⁻2119 crystals in drops containing fresh FmlH⁹⁻ protein, mother liquor, and AM2050 to replace the bound ligand. Drops containing 9 mg/mL FmlH⁹⁻, 2.5 mM AM2050, 10% DSMO, 0.1 M Tris 8.0 and 0.8 M AmSO₄ were allowed to equilibrate over 1 mL of well solution containing 0.1 M Tris 8.0 + 0.8 M AmSO₄ for two days. FmlH⁹⁻AM2119 co-crystals were then transferred the pre-equilibrated drops and allowed to soak for 48 hours before cryoprotection in 0.1 M Tris 8.0, 0.8 M AmSO₄, and 30% and flash-freezing in liquid nitrogen. All data were collected on ALS Beamline 4.2.2 at an X-ray wavelength of 1.00 Å. Raw data were processed using XDS, Aimless, and Pointless. The phase problem was solved using Phaser-MR in the Phenix suite using the apo FmlH⁹⁻ structure (PDBID: 6AOW) as a search model. Iterative rounds of Phenix. Refine and Coot were used to refine the final model. Guided ligand replacement was performed using Phenix. Space group, collection, and refinement statistics for all FmlH⁹⁻-galactoside structures are listed in Table 3.
Figure 3.1. Overview of the Type 1 pilus structure. A) Schematic representation of the Type 1 pilus bioassembly platform. B) The mannose residue forms several hydrogen binding interactions with the binding pocket of FimH. C) FimH can adopt two distinct conformational states; a high affinity “relaxed” state, and a low-affinity “tense” state. (D) Electrostatic surface representation of the FimH lectin domain. Residues Ile13, Try48, Ile52, and Tyr137 form a hydrophobic ridge around the deep binding pocket.
Figure 3.2. Comparison of the FimH and FmlH lectin domains from *E. coli* UTI89. A) Structural superposition of FimH$^{LD}$ and FmlH$^{LD}$ demonstrates a high degree of secondary structure overlap (RMSD=0.963) despite a relatively low sequence identity (44.9%). B) Superposition of the FmlH$^{LD}$ galactose binding residues (blue) with the FimH$^{LD}$ mannose binding residues (green) reveals a conserved direct- and water-mediated hydrogen bonding network responsible for coordination of the sugar ring.
Figure 3.3. X-ray crystal structures of ZFH269 and ZFH284 bound to FimH\textsuperscript{LD}. A) Structure and HAI of ZFH269. B) Structure and HAI of ZFH284. In the X-ray crystal structures of ZFH269 (C) and ZFH284 (D), both compounds form hydrogen bonds, nonpolar interactions, and π-π stacking interactions with FimH\textsuperscript{LD}. 
Figure 3.4. Crystal structures of high-affinity galactosides in complex with FmlH^{LD}. 

A) Crystal structure of 29β-NAc in complex with FmlH^{LD} (PDB 6AS8.) μ

Crystal structure of AM2134 in complex with FmlH^{LD} solved to 1.75 Å. 

C) Crystal structure of AM2119 in complex with FmlH^{LD} solved to 1.31 Å resolution. 

D) Crystal structure of AM2050 in complex with FmlH^{LD} solved to 1.39 Å resolution.
Figure 3.5. Pharmacokinetic properties of AM2134 and AM2151. A) *In vivo* pharmacokinetic analysis of AM2134 in rats reveals a serum half-life of approximately 70 minutes, with an oral bioavailability of 0.29%. B) Addition of a cyclopropyl group to the sulfonamine group and substitution of the trifluoromethyl group with a methyl group increases the serum half life to 88 minutes and increases the oral bioavailability to 0.728%.
Figure 3.6. C-linkages and acetyl groups improve the pharmacokinetics of ZFH269. A) Mannoside availability in murine urine is increased 15-fold after 8 hours upon replacement of the mannose hydroxyl groups with acetyl groups and 50-fold upon addition of a C-linkage in place of the O-glycosidic bond. B) Bladder bacterial burdens are decreased significantly upon addition of a C-linkage and acetyl groups to the ZFH269 backbone in a murine model of acute cystitis. C) Addition of a C-linkage or acetyl groups to the ZFH269 backbone does not affect bladder bacterial burdens in a murine model of chronic cystitis.
<table>
<thead>
<tr>
<th>Compound Name</th>
<th>Structure</th>
<th>IC50 (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>29β</td>
<td><img src="image1.png" alt="Structure" /></td>
<td>-</td>
</tr>
<tr>
<td>29β-NAc</td>
<td><img src="image2.png" alt="Structure" /></td>
<td>650</td>
</tr>
<tr>
<td>AM2050</td>
<td><img src="image3.png" alt="Structure" /></td>
<td>418</td>
</tr>
<tr>
<td>AM2119</td>
<td><img src="image4.png" alt="Structure" /></td>
<td>1717</td>
</tr>
<tr>
<td>AM2038</td>
<td><img src="image5.png" alt="Structure" /></td>
<td>272</td>
</tr>
<tr>
<td>AM2043</td>
<td><img src="image6.png" alt="Structure" /></td>
<td>1113</td>
</tr>
<tr>
<td>AM2098</td>
<td><img src="image7.png" alt="Structure" /></td>
<td>110</td>
</tr>
<tr>
<td>AM2109</td>
<td><img src="image8.png" alt="Structure" /></td>
<td>151</td>
</tr>
<tr>
<td>AM2134</td>
<td><img src="image9.png" alt="Structure" /></td>
<td>85</td>
</tr>
<tr>
<td>AM2151</td>
<td><img src="image10.png" alt="Structure" /></td>
<td>120</td>
</tr>
</tbody>
</table>

Table 3.1. Structures and IC50s of Select Biphenyl Galactosides
Table 3.2. Data collection and refinement statistics for FimH\textsuperscript{LD}-ZFH269

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>FimH-ZFH269 (5F3F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>P21 21 21</td>
</tr>
<tr>
<td>Cell Dimensions</td>
<td></td>
</tr>
<tr>
<td>$a, b, c$ (Å)</td>
<td>48.97, 55.66, 60.89</td>
</tr>
<tr>
<td>$\alpha, \beta, \gamma$ (°)</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>24.48-1.76</td>
</tr>
<tr>
<td>$R_{\text{merge}}$\textsuperscript{a}</td>
<td>0.05</td>
</tr>
<tr>
<td>$I/\sigma (I)$\textsuperscript{a}</td>
<td>4.23</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.7</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>No. Reflections (unique)</td>
<td>248304 (17107)</td>
</tr>
<tr>
<td>$R_{\text{work}}$</td>
<td>0.1352</td>
</tr>
<tr>
<td>$R_{\text{free}}$</td>
<td>0.1718</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>1209</td>
</tr>
<tr>
<td>Ligand</td>
<td>58</td>
</tr>
<tr>
<td>Solvent</td>
<td>237</td>
</tr>
<tr>
<td>Wilson B-Factor (Å$^2$)</td>
<td>15.2</td>
</tr>
<tr>
<td>R.m.s. Deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.009</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.41</td>
</tr>
</tbody>
</table>
### Table 3.3. X-Ray data collection and refinement statistics for FmlH\textsuperscript{LD}-galactoside complexes.

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>FmlH-AM2050</th>
<th>FmlH-AM2119</th>
<th>FmlH-AM2134</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space Group</strong></td>
<td>P43 21 2</td>
<td>P43 21 1</td>
<td>C 2 2 21</td>
</tr>
<tr>
<td><strong>Cell Dimensions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å)</td>
<td>57.1, 57.1, 121.3</td>
<td>57.2, 57.1, 120.6</td>
<td>39.4, 110.2, 190.2</td>
</tr>
<tr>
<td>$\alpha$, $\beta$, $\gamma$ (°)</td>
<td>90, 90, 90</td>
<td>90, 90, 90</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Resolution (Å)(^a)</td>
<td>38.32-1.39 (1.44-1.39)</td>
<td>41.49-1.31 (1.36-1.31)</td>
<td>45.1-1.75 (1.81-1.75)</td>
</tr>
<tr>
<td>$R_{\text{merge}}$(^a)</td>
<td>0.088 (0.526)</td>
<td>0.074 (0.886)</td>
<td>0.092 (0.398)</td>
</tr>
<tr>
<td>$R_{\text{meas}}$(^a)</td>
<td>0.1915 (0.545)</td>
<td>0.077 (0.922)</td>
<td>0.99 (0.429)</td>
</tr>
<tr>
<td>$I/\sigma(I)$(^a)</td>
<td>23.2 (5.18)</td>
<td>25.48 (2.95)</td>
<td>17.65 (5.07)</td>
</tr>
<tr>
<td>$CC_{1/2}$(^a)</td>
<td>0.999 (0.95)</td>
<td>1.000 (0.883)</td>
<td>0.998 (0.941)</td>
</tr>
<tr>
<td>Completeness (%)(^a)</td>
<td>99.9 (100.0)</td>
<td>99.91 (99.98)</td>
<td>99.9 (100.0)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>14.2 (14.2)</td>
<td>13.9 (13.2)</td>
<td>7.1 (7.3)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. Reflections</td>
<td>41074 (4031)</td>
<td>49124 (4810)</td>
<td>20228 (2088)</td>
</tr>
<tr>
<td>$R_{\text{work}}$(^a)</td>
<td>0.178 (0.200)</td>
<td>0.168 (0.225)</td>
<td>0.162 (0.180)</td>
</tr>
<tr>
<td>$R_{\text{free}}$(^a)</td>
<td>0.197 (0.223)</td>
<td>0.1794 (0.226)</td>
<td>0.182 (0.211)</td>
</tr>
<tr>
<td>No. atoms</td>
<td>1503</td>
<td>1420</td>
<td>1522</td>
</tr>
<tr>
<td>Protein</td>
<td>1188</td>
<td>1164</td>
<td>1212</td>
</tr>
<tr>
<td>Ligand</td>
<td>72</td>
<td>78</td>
<td>36</td>
</tr>
<tr>
<td>Solvent</td>
<td>243</td>
<td>178</td>
<td>274</td>
</tr>
<tr>
<td>B factors</td>
<td>12.96</td>
<td>14.79</td>
<td>14.1</td>
</tr>
<tr>
<td>Protein</td>
<td>10.29</td>
<td>12.45</td>
<td>11.1</td>
</tr>
<tr>
<td>Ligand</td>
<td>16.67</td>
<td>22.61</td>
<td>10.59</td>
</tr>
<tr>
<td>Solvent</td>
<td>24.89</td>
<td>26.69</td>
<td>27.85</td>
</tr>
<tr>
<td>R.m.s. Deviations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.009</td>
<td>0.009</td>
<td>0.007</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.47</td>
<td>1.35</td>
<td>1.11</td>
</tr>
<tr>
<td>Ramachandran favored (%)</td>
<td>97.3</td>
<td>97.97</td>
<td>97.45</td>
</tr>
<tr>
<td>Ramachandran allowed (%)</td>
<td>2.7</td>
<td>2.03</td>
<td>2.55</td>
</tr>
</tbody>
</table>

\(^a\)Values in parentheses reflect the highest resolution shell
Chapter 4: Structural and Functional Analysis of the F17-like Adhesin, UclD

4.1 Scope of Work

Of the work described in this chapter, I was responsible for the crystallization of UclD\textsuperscript{LD} and for solving the UclD\textsuperscript{LD} crystal structure with the help of Dr. Daved Fremont. I was also responsible for the design and performance of the differential scanning fluorimetry experiments, cell binding experiments (with the exception of hBMEC binding assays, which were performed by Dr. Matthew Conover), and for submission of samples to the Center for Functional Glycomics for glycan array screening. Portions of this chapter are reproduced from:


©2017, Springer Nature Limited

4.2 Introduction

As discussed above, Gram-negative bacteria express CUP pili to mediate adhesion to a variety of surfaces, facilitating their dissemination through the environment and colonization of the host (22). While the structure and function of some CUP pili, such as type 1 and fml from the \(\gamma\) clade and P from the \(\pi\) clade, have been well studied, the role of most pilus types remains unknown (20). To direct our structural and functional exploration of the \textit{E. coli} pilome, we sought
identify pili that may contribute to UPEC pathogenesis in patients suffering from UTI. To this end, a sequence database created from UTI strains isolated from women suffering from chronic, recurrent UTI (rUTI) was analyzed to look for enrichment of specific CUP operons relative to their overall carriage in *E. coli* (35). Of all pilus types included in this analysis, the F17-like operon was the most strongly enriched in these strains relative to the total *E. coli* population. Located in the \( \gamma \) clade, the F17-like (also known as Uca-like, or Ucl) pilus is encoded by a minimal four-gene operon encoding the major subunit, UclA, a periplasmic chaperone, UclB, an outer membrane usher, UclC, and a two-domain tip adhesin, UclD. Although the F17-like pilus is only present in 11% of all *E. coli* strains, genomic analysis demonstrated that 50% of strains belonging to *E. coli* clade B2, which includes most UPEC isolates from patients in the United States and Europe, encode the F17-like pilus (20). Further, 13 out of 14 strains isolated from patients with recurrent UTI encoded the F17-like pilus operon. This stark enrichment of UPEC encoding F17-like isolated from rUTI sufferers suggests that the F17-like pilus may allow UPEC to persist in the host despite appropriate antibiotic therapy for a UTI and reemerge to cause rUTI following cessation of treatment.

The F17-Like pilus derives its name from the primary sequence homology of UclC with the usher from the *E. coli* F17 pilus operon. F17 pili facilitate bacterial adhesion to N-acetylglucosamine (GlcNAC)-containing receptors on the intestinal microvilli of large ruminants, such as cattle and sheep (24). The colonization resulting from this interaction ultimately leads to diarrhea and septicemia in these animals (130). When phylogenetic analyses are expanded to include all Gram-negative bacterial species, the UclC usher shares the greatest degree of sequence identity with the Uca pilus from *Proteus mirabilis* (Figure 4.1). Homology is also seen between the UclC and the ECs1278 usher, which is enriched in Enterohemorrhagic *E. coli* (EHEC) isolates (35). In aggregate, the genetic similarities suggest that UPEC may have acquired the
F17-like pilus from *Proteus* via a horizontal gene transfer event, allowing UPEC to persist in the gastrointestinal tract despite appropriate treatment with broad-spectrum antibiotics.

To assess the role of F17-like pili in gastrointestinal colonization, C3H/HEN mice pretreated with streptomycin to eliminate colonization resistance conferred by the native microbiota were coinfected with differentially-marked WT UTI89 and UTI89\(Δucl\) (35). Measurement of fecal bacterial titers revealed that deletion of the F17-like pilus resulted in a \(~1000\)-fold competitive disadvantage in colonization of the gastrointestinal tract (Figure 4.2A). Interestingly, UTI89\(Δfim\) (UTI89 lacking the type 1 pilus operon) also exhibited a competitive defect (Figure 4.2B). Concomitant deletion of both operons resulted in a greater competitive defect than deletion of either pilus alone, suggesting that these pili play nonredundant roles in colonization of the gastrointestinal tract (Figure 4.2C). To confirm the tropism of these pili for the murine colon, FimH\(^{LD}\) and the lectin domain of the UclD adhesin (UclD\(^{LD}\)) were purified and conjugated to a fluorescent tag, and binding was assessed to murine colon tissue *ex vivo*. Binding of UclD\(^{LD}\) was observed in the lower portions of the colonic crypts, while binding of FimH\(^{LD}\) was observed in the upper portion of the colonic crypts (Figure 4.2D and E) (35). As expected, pre-treatment of colonic tissue with PNGase, a nonspecific glycosidase that removes all N-linked glycan moieties, including mannose, abrogated FimH\(^{LD}\) binding. Interestingly, treatment with O-glycosidase, but not PNGase, was able to block UclD\(^{LD}\) binding to the lower crypts, indicating that UclD may recognize an O-linked ligand.

A separate study using comparative proteomics to identify factors preferentially expressed by UPEC during growth in human urine identified F17-like pili as an adherence factor that facilitates bacterial binding to desquamated bladder epithelial cells (131). Additionally, F17-like pili were found to contribute to biofilm formation on abiotic surfaces in human urine. However, when UTI-susceptible mice were concurrently colonized with WT UTI89 and UTI89\(Δucl\), no
competitive defect was observed in bacteria lacking the operon (35). This suggests that the in vitro uroepithelial cell binding phenotype may only occur in the presence of pilus overexpression, and may not play a vital role in pathogenesis.

4.3 Structural and Functional Analysis of UclD<sub>LD</sub>

4.3.1 Determination of the UclD<sub>LD</sub> Structure

To better understand the mechanism of F17-like-mediated gastrointestinal colonization, I generated crystals of UclD<sub>LD</sub> following sequential rounds of sparse matrix screening and crystal refinement. Crystals were grown using the hanging-drop vapor diffusion method over 1 mL of mother liquor solution (0.1 M potassium phosphate (monobasic), 0.2 M potassium iodide and 20% PEG 3350.) Crystals were cryoprotected in mother liquor supplemented with 20% glycerol prior to flash-freezing in liquid nitrogen. Although crystals of the native protein diffracted to ~1.8 Å resolution, I was unable to phase the data using molecular replacement with a variety of search models, including F17G<sub>LD</sub>, FimH<sub>LD</sub>, PapG<sub>LD</sub>, and FmlH<sub>LD</sub>. Thus, I generated selenomethionine-labeled UclD<sub>LD</sub> and grew crystals under the same conditions used to generate native crystals. Data were collected on these crystals, which were determined to be in the P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> space group. Phasing was performing using single-wavelength anomalies dispersion (SAD) and the structure was solved to a resolution of 1.6 Å (PDB 5VQ5, Figure 4.3A). Collection and refinement statistics can be found in Table 4.1. This structure revealed that UclD<sub>LD</sub> adopts the canonical β-sandwich fold observed in the lectin domains of other CUP pilus adhesins, including both F17G and FimH. In UclD<sub>LD</sub>, this fold is composed of 11 individual β-strands and two short α-helices found in the interstrand loops. Although the amino acid sequences of the two lectin domains are less than 25%, identical, global structural alignment of UclD<sub>LD</sub> with the F17G adhesin lectin domain (F17G<sub>LD</sub>,...
Figure 4.3B) revealed that the two structures overlay with a Cα RMSD of only 3.7 Å (Figure 4.3C). A structure-based primary sequence alignment of the secondary sequence elements in UclD\textsuperscript{LD} and F17G\textsuperscript{LD} reveals the conservation of every β-strand contributing to the core fold (Figure 4.3D). The primary differences between these two proteins is the presence of two large insertions between β4/β5 and β8/β9 present in UclD but not F17G.

**4.3.2 Characterization of the Putative UclD\textsuperscript{LD} Binding Pocket**

The availability of a co-crystal structure of F17G\textsuperscript{LD} bound to its natural ligand, GlcNAc, allows for the inference of the location of the putative UclD\textsuperscript{LD} ligand-binding site. In contrast to the deep binding pockets located at the distal tips of FimH and FmID, F17G binds GlcNAc in a shallow pocket located on the side of the lectin domain (24, 132). This pocket primarily comprises residues in the β5-β7 loop and in β8 and β9 themselves. Specifically, the side chains of F17G-D88, -T89, -S116, -Q118 and the backbone of A43 form hydrogen bonding interactions with the hydroxyl groups on the glucose ring, while the side chain of F17G-T115 forms a hydrogen bond with the N-acetyl nitrogen. Additionally, the aromatic side chain of F17G-W109 likely forms π stacking interactions with the pyranose ring of the galactose moiety, a feature common among carbohydrate-binding lectin proteins.

Like F17G\textsuperscript{LD}, UclD\textsuperscript{LD} contains a putative shallow binding pocket located on the side of the domain dominated by UclD-W103. Other residues lining the putative active site include UclD-P104, -D119, -N131, -T134, and -S136. Interestingly, comparison of this structure to another UclD\textsuperscript{LD} structure generated concurrently in the laboratory of our collaborators reveals two different conformations of the UclD-W103 side chain (Figure 4.4). Because the binding pocket does not form close contacts with neighboring subunits in the crystal lattice that could explain this conformational variation, this could indicate physiologically-relevant conformational flexibility in the UclD-W103 residue that could modulate binding to a sugar ring.
4.3.3 Identification of a Molecular Ligand

To identify the molecular ligand recognized by UclD<sup>LD</sup>, I submitted biotinylated, purified protein to the Center for Functional Glycomics (CFG) for screening against their mammalian glycan array, which comprises over 600 unique \(N\)- and \(O\)-linked glycan structures (133). No significant binding to any of these ligands was observed (Figure 4.5A). To ensure that the process of immobilizing the glycans into the chip was not affecting their access to the active site of soluble UclD<sup>LD</sup>, I also used differential scanning fluorimetry (DSF) to assess binding to a series of monosaccharides, including the F17G receptor, GlcNAc. However, there was no thermal shift upon exposure to any of the monosaccharides tested (Figure 4.5B).

To aid in the identification of a molecular ligand for UclD<sup>LD</sup>, I sought to develop a high-throughput \textit{in vitro} assay that would allow us to screen small molecules for their ability to inhibit UclD binding to its native ligand. To this end, I tested the ability of UclD<sup>LD</sup> to bind a variety of immortalized cell lines, including Caco-2, a human epithelial colorectal adenocarcinoma cell line. We did not observe significant UclD<sup>LD</sup> binding to Caco-2 cells or any of the other cell lines tested, including bladder epithelial carcinoma (5637), bladder transitional carcinoma (T24), brain microendothelial (hBMEC) and kidney carcinoma (A498) cells (Figure 4.6A and data not shown). However we did observe robust FimH<sup>LD</sup> binding to Caco-2 cells that was inhibited by mannose, providing further verification that FimH binds mannosylated glycoproteins in the gastrointestinal tract to facilitate GI colonization (Figure 4.6B).

4.4 Discussion and Future Directions

Evidence from genomic analyses and \textit{in vivo} murine models of gastrointestinal colonization indicate the F17-like pilus adhesin, UclD, mediates binding to \(O\)-glycans in the colon
of mice and humans. In this study, we sought to determine the structural basis for this interaction and identify the molecular ligand recognized by UclD. In pursuit of this goal, we solved a high-resolution crystal structure of UclD$^{LD}$, revealing a putative binding pocket comprising residues UclD-W103, -P104, -D119, -N131, -T134, and -S136. The presence of the apparently-flexible and solvent-exposed UclD-W103 in the binding pocket suggested that UclD, like many other CUP adhesins, may recognize a glycan receptor expressed on the surface of the host’s gastrointestinal epithelium. However, screening against 609 common glycan structures on the CFG mammalian glycan array yielded no significant hits (Figure 4.5A). There are several possible explanations for this finding, including that the ligand recognized by UclD may not be a carbohydrate or that the immobilization of glycan ligands on the surface of the microarray chip is not conducive to UclD binding. However, the linker anchoring the glycans to the microarray chip should be of sufficient length to allow binding to the transverse binding pocket seen in the structure of UclD$^{LD}$.

Examination of the surface topology near the putative binding site reveals a large nonpolar groove to the left of the binding site that may bind a nonpolar ligand, such as a glycolipid or bile salt, rather than a traditional glycan (Figure 4.4). To test this hypothesis, one could use an ELISA-based binding assay to assess the ability of purified UclD$^{LD}$ to bind a series of commercially-available bile salts or glycolipids (ie, Total E. coli or Bovine Total Liver Extract) Additionally, one could perform a Folch extraction of homogenized murine tissue to generate a tissue-specific glycolipid mix and use an ELISA-based approach to quantify UclD$^{LD}$ binding. Any hits could be further characterized using thin layer chromatography and mass spectrometry. One other explanation is that the UclD ligand is not a glycan found on mammalian cells. Because UclD appears to function in the gastrointestinal environment, which is the home of a complex bacterial community, it is possible that UclD binds glycans elaborated by other members of the intestinal microflora. To determine if the UclD ligand is a bacterial glycan, we plan to submit the UclD$^{LD}$
protein to the CFG microbial glycan array. It is also possible that UclD binds components of the extracellular milieu to facilitate biofilm formation. *In silico* docking analyses performed by Jon Helander in the Janetka lab revealed that five-membered rings could form plausible hydrophobic and polar interactions with the putative UclD binding site (134). In particular, N1-methyl-guanosine was identified as a possible binding partner. Because DNA is a major constituent of part of the gastrointestinal extracellular matrix, we hypothesized that the Ucl pili may be important for biofilm formation. However, *in vitro* testing of binding using DSF demonstrated no shift in the melting temperature of UclD with or without guanosine or methyl-guanosine (Figure 4.7).

It is also possible that the ligand-binding binding pocket of the UclD lectin domain alone may not adopt a high-affinity ligand binding state or mirror the conformation it adopts when incorporated into the tip of F17-like pili. This possibility is not unprecedented. As noted above, the affinity of FimH for its natural ligand, mannose, depends on the conformation of its lectin and pilin domains (29). Thus, it is possible that the absence of a pilin domain or a DSE-paired major pilus subunit may be altering the position of the loops or side chains near the putative ligand-binding site, decreasing the affinity of the UclDLD construct for its native ligand. However, the observed ex vivo binding to murine colonic tissue suggests that UclDLD retains at least some binding activity. Towards assessing the possibility that the binding pocket residue orientations in the UclDLD construct used for structural studies do not represent the high-affinity binding form of the protein, I have cloned the full-length UclD adhesin and the his-tagged UclB chaperone into the pBAD plasmid and pTRC99a plasmids, respectively. Previous studies in other pilus systems have been able to express and purify a stable periplasmic chaperone-adhesin complex using this approach. Unfortunately, despite robust expression of the 6x-his-tagged UclB chaperone, we were unable to obtain sufficient quantities of the chaperone-adhesin complex for binding or structural analysis. Further optimization of this expression system using different plasmids and/or expression cells
may improve yields, allowing for studies to be conducted using the full length adhesin. Additionally, peptides corresponding to the N-terminal extension of the UclA major subunit were also designed and ordered to generate a stable peptide-bound full length adhesin for use in structural and functional studies following an in vitro DSE reaction. Finally, I have also cloned the whole F17-like pilus operon into the inducible pTRC99a expression system and verified the expression using gold-bead-conjugated Ucl\textsubscript{LD} antibodies. These pili can be purified and used directly to assess ex vivo binding.

I cloned and expressed three targeted point mutants to the putative binding pocket of the Ucl\textsubscript{LD} (W103A, D119A, and N131A.) However, we have been unable to replicate the in vitro WT binding observed by Remaut, et al, and thus, do not have an in vitro readout of binding to assess the activity of these mutants. If binding can be replicated, these mutants should be tested for binding activity to verify the identity of the active site. Remaut, et al. have also created a panel of mutants, and report that the W103A, T134Q and S136Q mutants each exhibit a loss of ex vivo binding to colonic tissue. Additionally, treatment of the tissue with glycosidases that eliminate specific glycan moieties may also elucidate the natural binding partner of the F17-like pilus. Binding to human colon tissue should also be assessed to confirm the presence of the native ligand in the human gastrointestinal tract. Future studies could work to integrate these mutants into their native locus in the UTI89 chromosome using the scarless positive-negative selection system developed by Chen, et al to test their effect on tropism in the murine gastrointestinal colonization model.

In addition to their potential as targets for antimicrobial therapy, a more thorough understanding of CUP pili that mediate gastrointestinal colonization may also allow for the rational design of probiotic strains that can stably colonize the gastrointestinal tract. By arming commensal strains with the same factors used by pathogenic bacteria to mediate binding, probiotics could be
engineered to occupy the same niche as enteropathogens and uropathogens (135-139). For this approach to succeed, a more detailed understanding of F17-like pilus expression from its native locus is required. Efforts have been made to characterize the expression of the F17-like pilus using Western blot RT-PCR in a variety of conditions, including aerobic, microaerophilic, and anaerobic growth in a variety of media. Future attempts will be geared towards assessing the role of common fecal contents, including bile salts and bacterial metabolites, in F17-like expression. If induction of the F17-like operon is observed, specific components of feces could be extracted to narrow down the inducing factors.

4.5 Materials and Methods

4.5.1 Protein Purification

To generate purified UclD^{LD} for the P2_{1}2_{1}2_{1} space group, DNA from the UTI89 uclD gene encoding the N-terminal 217 amino acids of the protein were cloned into pTRC99a with a C-terminal 6-His tag. After cleavage of the signal sequence, the mature form of UclD^{LD}_{6}×His contained 203 amino acids. This construct was expressed in the periplasm of *E. coli* DL41(DE3) cells (a methionine auxotroph strain suitable for expression of native or selenomethionine-labelled protein) by induction for 1 h with IPTG after growth to OD_{600} 4.0 in LB + Superbroth (Fisher). Periplasmic extracts were performed by treating cells with 20 mM Tris 8.0, 20% sucrose, 1 mM EDTA, and 10 ug/mL lysozyme. Following quenching with MgCl, extracts were first dialyzed against PBS supplemented with 250 mM NaCl, then bound to a cobalt (Goldbio) column; bound proteins were eluted with PBS containing 250 mM NaCl and a 0-300 mM imidazole gradient. Pooled fractions containing UclD^{LD} were dialyzed into 20 mM
MES, pH 5.8, bound to an HR16/10 Mono S cation exchange column (GE Healthcare), and eluted with a 0-500 mM NaCl gradient.

Selenomethionine-labelled protein was grown as previously described (35). All purification buffers were supplemented with 2 mM β-mercaptoethanol and 1 mM EDTA to prevent oxidation. EDTA was omitted from the periplasmic dialysis buffer to prevent chelation of immobilized cobalt.

4.5.2 Crystallization and Structure Determination

For the UclD<sup>LD</sup> structure solved in the P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> space group, native- and selenomethionine-labelled UclD<sup>LD</sup> (10 mg ml<sup>−1</sup> 10 mM MES pH 5.8) was crystallized by the hanging drop vapor diffusion method against a well solution containing 0.1 M potassium phosphate (monobasic), 0.2 M potassium iodide, and 20% PEG 3350. One microliter of the protein solution was mixed with 1 μl well solution and incubated at 18 °C. Crystals were harvested and transferred to a solution containing 0.1 M potassium phosphate (monobasic), 0.2 M potassium iodide and 20% PEG 3350 supplemented to 20% glycerol before being flash-frozen in a bath of liquid nitrogen.

Data were collected at beamline 4.2.2 (ALS Berkeley) to 1.6 Å resolution. Data were indexed and processed with XDS (140), scaled and merged using AIMLESS in the CCP4 suite (141), phased with the single anomalous dispersion (SAD) method using phenix.autosol, and refined with phenix.refine (142). Data and refinement statistics can be found in Table 4.1. r.m.s.d. values were calculating using the DALI server (143). Structural alignments were performed in PROMALS3D using the default settings. Secondary structure assignments for UclD<sup>LD</sup> were completed using DSSP. The structure was deposited into the RCSB Protein Data Bank (PDB deposition ID# 5VQ5.)
4.5.3 FimH and UclD Binding Studies

Caco-2 cells (ATCC number HTB-37) were cultured in minimum essential medium (MEM) supplemented with 20% FBS. Cell cultures tested negative for mycoplasma. Cells were split into 48-well plates, grown to 100% confluence and then fixed with paraformaldehyde for 15 min followed by treatment with blocking buffer (PBS containing 2% BSA) for 2 h. A truncated FimH, corresponding to residues 1–178 of the mature FimH adhesin (FimH LD), expressed in *E. coli* and purified as described previously 37, was serially diluted in blocking buffer and incubated with the fixed Caco-2 cells for 1 h at room temperature. To test the effect of D-mannose or M4284 on FimH binding, 0.2 mg ml⁻¹ FimH LD was pre-incubated for 5 min in the presence or absence of 1 mM D-mannose (Sigma-Aldrich) or 1 mM M4828 (in 20 mM Tris, pH 8.0, or 20 mM Tris or 10% cyclodextrin, respectively) before serial dilution and incubation. Wells were washed four times with PBS and 0.05% Tween 20 (PBST) before incubation with a polyclonal rabbit anti-T3 antibody against FimH LD (generated against FimH residues 1–165; (29)) for 1 h at room temperature. After another series of four washes, secondary antibody (goat anti-rabbit Ig conjugated to horseradish peroxidase; ThermoFisher, 32460) was incubated with the cells for 1 h at room temperature (24 °C) before washing in PBST. Plates were developed with the BD OptEIA TMB substrate reagent kit for 5 min at room temperature (24 °C) before quenching with 1 M H₂SO₄. Binding was assessed by measuring the absorbance at 450 nM on a TECAN infinite 2 PRO plate reader. Wells lacking protein were used as control. All conditions were examined in quadruplicate. Experiments assessing UclD LD binding to 5637 cells (ATCC HTB-9), T24 (ATCC HTB4), and A-498 (ATCC HTB-44) were repeated using the same protocol. Binding was detected using a polyclonal rabbit antibody generated against the UclD LD construct used for crystallographic and binding studies.
4.5.4 Differential Scanning Fluorimetry

Purified UclD\textsuperscript{LD} (1.4 μg per well) was incubated with 5x Sypro orange fluorescent dye (Sigma-Aldrich) in 20 mM Tris, pH 8.0, with or without 10 mM monosaccharide or other compounds in a total volume of 70 μl. Samples were heated from 20 °C to 100 °C in 30 s, 0.5 °C increments using a Bio-Rad C1000 thermocycler with CFX96 RT–PCR attachment. The reported melting temperatures were determined by the inflection point of the sigmoidal graph.

4.5.5 Glycan Array Screening

6x-his tagged UclD\textsuperscript{LD} was purified as described above and biotinylated using the NHS-PEG4-Biotin and Biotinylation Kits (ThermoFisher) as previously described. Biotinylated protein was dialyzed into TSM binding buffer (20 mM Tris–HCl, pH 7.4, 150 mM NaCl, 2 mM CaCl\textsubscript{2}, 2 mM MgCl\textsubscript{2}, 0.05% Tween 20, and 1% BSA). Protein was concentrated to 0.2 mg/mL, and sent to the CFG glycan array (133). Protein was detected using fluorescently-labelleld streptavidin.
Figure 4.1. Phylogenetic analysis reveals close homologues to the F17-like pilus. *E. coli* F17-like pili are most closely related to Uca pili from Proteus species, suggesting that UPEC acquired F17-like pili through a horizontal gene transfer event to facilitate persistence in the gastrointestinal niche.
Figure 4.2. Type 1 and F17-like pili facilitate colonization of the murine gastrointestinal tract. A) Deletion of the \textit{fim} operon results in a \textasciitilde100 fold competitive defect in a murine model of gastrointestinal colonization. B) Deletion of the \textit{ucl} operon results in a \textasciitilde1000 fold competitive defect in a murine model of gastrointestinal colonization. C) Concomitant deletion of the \textit{fim} and \textit{ucl} operons results in a more severe competitive defect than either operon alone, suggesting that these pili play nonredundant roles in gastrointestinal colonization. D) FimH\textsuperscript{LD} binds to \textit{N}-glycans in the upper crypts of the murine colon. E) UclD\textsuperscript{LD} binds to \textit{O}-glycans in the lower crypts of the murine colon.
Figure 4.3. Structural comparison of UclD\textsuperscript{LD} and F17G\textsuperscript{LD}. 

A) Cartoon representation of the structure of UclD\textsuperscript{LD}. 

B) Cartoon representation of the structure of F17G\textsuperscript{LD} bound to its natural ligand, \textit{N}-acetyl glucosamine (GlcNAc). 

C) Surface representation of UclD. The putative binding pocket is highlighted in magenta. Two short insertions in UclD not present in F17G are highlighted in yellow and orange. 

D) Structural alignment of UclD\textsuperscript{LD} and F17G\textsuperscript{LD}. Stars represent residues believed to participate in ligand binding.
Figure 4.4. Comparison of residues in the putative binding site of UclD<sup>LD</sup>. The W103 residue adopts two different conformations in the active site of the two crystallized forms of UclD, suggesting possible mobility. The analogous residues are shown in the F17G structure bound to GlcNAc.
Figure 4.5. UclD<sup>LD</sup> does not bind common mammalian glycan epitopes.  A) No significant hits were observed when UclD<sup>LD</sup> was screened against the 609-residue Center for Functional Glycomics (CFG) mammalian glycan array. Successful binding events are expected to reach intensities of 5-10,000 relative fluorescence units or more. B) A shift in the melting temperature of UclD<sup>LD</sup> is not observed in the presence of soluble galactose, GalNAc, glucose, GlcNAc, or mannose, suggesting a lack of binding.
Figure 4.6. **UclD^{LD} does not bind to Caco-2 cells or common monosaccharides.** A) FimH^{LD}, but not UclD^{LD}, binds to Caco-2 human epithelial colorectal adenocarcinoma cells. B) FimH binding to Caco-2 cells can be inhibited by mannose and high-affinity mannoside M4284.
Figure 4.7. Neither guanosine nor N1-methylguanosine affect the $T_m$ of UclD$^{LD}$ and FimH$^{LD}$.

7 µL of 50 mM guanosine or N1-methylguanosine solubilized in 100% DMSO were added to 2.5 µG UclD$^{LD}$ or FimH$^{LD}$ to a final concentration of 5 mM N1-methylguanosine, 10% DMSO. The expected melting temperature shift of 7 °C was observed for FimH$^{LD}$ in the presence of 5 mM mannose and 10% DMSO, confirming that DMSO did not affect binding.
Table 4.1. Collection and refinement statistics for UcID\textsuperscript{LD}.

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>UcID\textsuperscript{LD} (5VQ5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>P2, 2, 2, 2, 2</td>
</tr>
<tr>
<td>Cell Dimensions</td>
<td></td>
</tr>
<tr>
<td>(a, b, c) (Å)</td>
<td>39.0, 58.6, 175.1</td>
</tr>
<tr>
<td>(\alpha, \beta, \gamma) (°)</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>55.6-1.6 (1.66-1.60)</td>
</tr>
<tr>
<td>(R\text{merge})\textsuperscript{a}</td>
<td>0.05 (0.43)</td>
</tr>
<tr>
<td>(I/\sigma (I))\textsuperscript{a}</td>
<td>17.68 (1.43)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (99.9)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>No. Reflections</td>
<td>108142</td>
</tr>
<tr>
<td>(R\text{work})</td>
<td>0.172</td>
</tr>
<tr>
<td>(R\text{free})</td>
<td>0.218</td>
</tr>
<tr>
<td>No. atoms</td>
<td>3289</td>
</tr>
<tr>
<td>Protein</td>
<td>3804</td>
</tr>
<tr>
<td>Iodide</td>
<td>6</td>
</tr>
<tr>
<td>Solvent</td>
<td>479</td>
</tr>
<tr>
<td>B Factors (Å\textsuperscript{2})</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>22.4</td>
</tr>
<tr>
<td>Iodide</td>
<td>62.6</td>
</tr>
<tr>
<td>Water</td>
<td>36.2</td>
</tr>
<tr>
<td>R.m.s. Deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.002</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.58</td>
</tr>
</tbody>
</table>
Chapter 5: Structural Characterization of the Yeh Pilus

5.1 Scope of Work

Of the work described in this chapter, I was responsible for the crystallization of YehD\textsuperscript{LD} and for solving the YehD\textsuperscript{LD} crystal structure. I have crystallized and generated a native data set for the YhlD\textsuperscript{LD} structure. I was also responsible all genetic/genomic analyses, and for the design and generation of the Yeh-full-length operon constructs. I designed the Yhl deletion knockouts used in the mouse studies, and performed preliminary \textit{ex vivo} immunohistochemistry studies of YehD\textsuperscript{LD} and YhlD\textsuperscript{LD} binding to murine colonic tissue and a preliminary competitive infection of streptomycin-treated C3H/HeN mice with \textit{E. coli} Sakai and Sakai::yhl.

5.2 Introduction

Yeh- and Yeh-like pili are two related pilus types found within the $\gamma$ pilus clade (20). The Yeh pilus is present several \textit{E. coli} pathotypes found in clades A, B1, B2, and D, including the vast majority of UPEC and ETEC isolates (20). Distribution of Yeh-like is limited primarily to clade E strains, which include enterohemorrhagic \textit{E. coli} (EHEC) isolates. Previous reports have unsuccessfully attempted to elucidate a function for the Yeh pilus. \textit{E. coli} isolate UTI89::yeh did not exhibit a competitive disadvantage during infection of the murine gastrointestinal tract relative to WT UTI89 (35). However, based on the phylogenetic relationship of the Yeh Yhl pili to known colonizers of the gastrointestinal tract, we sought to further characterize these common pilus types. In this study, we present a structure of the lectin domain of the Yeh pilus adhesin YehD.
revealing a novel helix-loop-helix morphology in addition to the canonical beta sandwich motif seen in most pilus adhesins. I have also used genetic analyses to identify negatively-selected residues within the YehA major subunit that may mediate adhesion to neighboring pilus subunits, and mapped those residues on a homology model of the Yeh pilus rod. Finally, we conducted *in vitro* and *ex vivo* binding studies to suggest that YehD may bind murine colonic tissue. In aggregate, these data provide structural insight into a common CUP pilus type that could serve as a viable vaccination candidate against *E. coli* infections.

5.3 Results

5.3.1 Yeh and Yhl Pili are Common, Related, and Distinct

The Yeh pilus operon encodes four CUP pilus genes, reflecting the minimum requirements necessary to assemble a CUP pilus (Figure 5.1A). YehB is a putative periplasmic chaperone that facilitates the transport of the YehD tip adhesin and the YehA major structural subunit through the periplasm to the YehC usher (Figure 5.2B). The cognate genes in the Yeh-like operon, termed YhlA-YhlD, likely perform similar functions. Comparison of the primary amino acid sequences of the mature Yeh and Yhl pilus subunits reveals that the major subunits are 87.3% identical, the chaperones are 89.3% identical, and the ushers are 91.4% identical. The high degree of usher and chaperone conservation observed here is consistent with previous analyses from other systems. Due to the highly coordinated and conserved nature of their function, CUP pilus ushers are typically more conserved than other genes within the operon. By contrast, the adhesins were only 34.6% identical, and the lectin domains were only 20.1% identical, suggesting a possible functional divergence of the two pili. Construction of a phylogenetic tree with all YehD and YhlD homologues in Gram-negative bacteria reveals that Salmonella and Citrobacter species also
encode two distinct forms of the Yeh pilus. Providencia and Edwardsiella, two opportunistic human pathogens that sometimes cause catheter-associated UTIs and other extraintestinal infections, also encode Yeh homologues.

In an attempt to glean functional insights from the distribution of Yeh and Yhl species across E. coli pathotypes, we examined the carriage pattern of yehC and yhlC in the E. coli pangenome. We found that 82% of 240 publically-available E. coli reference genomes contained yehC, and an additional 13% contained yhlC. YhlC was commonly found in E. coli Clade E, which is known to contain primary EHEC isolates (11). Interestingly, 50% of strains encoding the Yeh-like operon also encode FimH allelic variant N135K. Residue N135 is known to participate in mannose binding in the intestine and exhibits function defects when mutated to alanine and tested in vitro.

5.3.2 The YehD Lectin Domain Contains a Mobile Helix-Loop-Helix Domain

Pilus adhesion to biotic or abiotic surfaces is typically mediated by the N-terminal lectin domain of a two-domain adhesin located at the distal tip of the pilus. To gain functional insights into the mechanism of adhesion for the Yeh pilus, we obtained an X-ray crystal structure of the YehD lectin domain (YehDLD) to 1.6 Å resolution (Figure 5.2A). Selenomethionine-labelled Yeh (12.7 mg/mL in 20 mM MES pH 5.8) was crystallized in 0.2 M AmH_2PO_4 + 20% w/v PEG 3350. Data were collected at the selenium peak and phased using single-wavelength anomalous diffraction. In this structure, 9 β-strands coalesce to form a core β-sandwich fold similar to those observed in the lectin domains of other CUP pilus adhesins, such as FimH, PapG, FmlD, and UclD (25, 32, 35, 123). However, this structure also revealed a noncanonical helix-loop-helix motif that has not, to our knowledge, been observed in any other pilus adhesin. The two helices within
this motif are linked by a disulfide bond near the end of the loop that extends approximately two thirds of the way down the side of the lectin domain.

To determine if the helix-loop-helix domain can flip away from the core β-sandwich fold, I performed molecular dynamics simulations using the lectin domain crystal structure as the starting coordinates. Significant movement in the helices relative to the body of the protein was not seen in a 2 ns conventional simulation using the GROMACS software. Subsequently, 12 independent runs of 500 ns each performed by Justin Porter in the Bowman lab were only able to observe modest movement of the helices. However, utilization of the Fluctuation Amplification of Specific Traits (FAST) method to reveal states in a guided, unbiased manner allowed us to observe domain movement (Figure 2B, C, E, and F) (144). These simulations revealed that the helices are capable of flipping upwards to expose a series of buried residues.

### 5.3.3 Evolutionary Adaption of the Pilus Rod

To explore the structure of the Yeh pilus rod in the absence of a known structure, I generated a homology model of Yeh using Phyre 2's one-to-one threading function (145). The resultant model reveals a high degree of predicted overlap between the FimH and YehA structures despite a sequence homology of only 25%. When these homology models of YehA are aligned with the established structure of the type 1 pilus rod (Figure 5.3A), the conserved mechanism of donor strand exchange is observed (18). Specifically, residues L8, I10, G12, and I14 on the N-terminal extension of YehA are predicted to occupy the P2-P5 sites, respectively, of the following subunit (Figure 5.3B). Previous studies have noted that residues responsible for mediating key intersubunit interactions within the pilus rod are subject to purifying selective pressure. To assess the evolutionary pressures acting upon major subunits within the Yeh pilus rod, I performed dN/dS analysis on all *E. coli* YehA major pilin sequences. Of the 158 residues in the mature YehA protein, we found 36 residues subject to negative (purifying) selection with a P-
value of 0.1 (Figure 5.4A). No residues subject to positive (adaptive) selection. An additional 42 codons were too highly conserved for quantitative selection analysis (Figure 5.4A). Mapping of these residues to the Yeh pilus homology model reveals that the majority of the negatively-selected residues are located within the core of the Ig-like fold. However, several of the negatively-selected surface residues are located at the predicted interface between adjacent subunits (Figure 5.4B-D). A cluster of residues including I76, I81, D85, and I88 forms polar and hydrophobic contacts with the T48, Y49, P129, K131, and K133 residues of the N+3 subunit located vertically above it. A second cluster containing negatively-selected/conserved residues Y119, K120, A127, and P129 interact with L75, I76, N78, K80, D109, and N115 on the N+3 subunit. Additionally, 8 residues located in the NTE of YehA are also subject to negative selection. These residues include the hydrophobic residues at the 8 and 10 positions plus a glycine at position 12, underscoring the specificity of the donor strand exchange process.

5.3.4 **Ex vivo and in vitro YehD^{LD} and YhlD^{LD} Binding Studies**

Because Yeh and Yhl pili are located in the same sub-clade as known mediators of gastrointestinal colonization, including F17 and F17-like pili, we sought to assess the ability of purified YehD^{LD} and YhlD^{LD} to bind murine colonic tissue (20). To this end, colonic tissue was harvested from 8-week-old C3H/HeN mice, fixed in methacarn, and embedded in paraffin for storage. Prior to incubation with YehD^{LD}, YhlD^{LD}, or FimH^{LD} (positive control) labelled with Alexa Fluor 647, deparaffinized slides were boiled in sodium citrate buffer and blocked with 2% BSA and an endogenous biotin blocking solution. Slides were then stained with Hoechst dye to label the cellular nuclei and visualized via confocal microscopy. All constructs tested bound to the murine gastrointestinal tract (Figure 5.5A-D).

To develop a more simple assay to assess binding, primary colonic epithelial cells were cultured into spheroids *in vitro* and grown as a monolayer on transwell plates. Membranes were
fixed, blocked with BSA and an endogenous biotin blocking solution, and incubated with labelled lectin domains. Like the colonic tissue, we observed that YehD<sup>LD</sup> and YhlD<sup>LD</sup>, along with the FimH<sup>LD</sup> positive control, all bound these primary-culture derived monolayers (Figure 5.6A-P). While FimH appeared to bind throughout the cells, YehD appeared to bind the edges of cells, and YhlD<sup>LD</sup> bound in clumps, often colocalizing with UEA-1 lectin binding, suggesting it may bind a component of secreted or cell-associated mucins.

5.3.5 Mouse Studies

Despite sharing common ancestry with known colonization factors of the gastrointestinal tract, deletion of the <i>yeh</i> operon from urinary tract isolate UTI89 did not affect colonization in a murine model gastrointestinal infection (Figure 5.7A). To determine if Yhl pili were necessary for gastrointestinal colonization, the <i>yhl</i> operon was knocked out of EHEC isolate O157:H7:Sakai, and colonization efficiency was compared in C3H/HEN mice using a combined DSS/Streptomycin treatment model of gastrointestinal infection and colitis. Ultimately, no significant competitive defect in colonization was observed based on fecal titers collected between one and eight days post-infection (Figure 5.7B). Single infection in similar models also showed no significant colonization defect up to 12 days post-infection (Figure 5.7C,D).

5.4 Discussion and Future Directions

CUP pili are key virulence factors expressed by a variety of Gram-negative bacterial pathogens to mediate adhesion to host tissues (22). Despite being found in 80-90% of all <i>E. coli</i> genomes, previous analyses have revealed little about the structure or function of the Yeh pilus. In this work, we present a 1.6 Å resolution structure of the YehD lectin domain, revealing a noncanonical helix-loop-helix motif protruding from an interstrand loop at the distal end of the
adhesin. In vitro and ex vivo binding studies reveal binding of both YehD^{LD} and YhlD^{LD} to murine colons, suggesting that these pili may play a role in gastrointestinal colonization. However, identification of a molecular ligand has been hitherto unsuccessful. Further analysis of these findings, along with possible paths to identify specific targets that could demonstrate utility in treating E. coli infections, are outlined below.

5.4.1 Exploration of the YehD^{LD} and YhlD^{LD} Structures

Helix-loop-helix and helix-turn-helix domains have been implicated in a variety of biological functions across all kingdoms of life. Helix-loop-helix domains mediate the dimerization of transcription factors (146), while helix-turn-helix domains are common structural motifs that typically mediate DNA binding (147). Similar constructs, such as the two-helix finger domain of the SecA inner membrane transportation complex, have also been described (148). However, the primary sequence of the YehA helix-loop-helix motif shares no sequence similarity with any of these established functional domains. Because it overlays the side of the core β-sandwich fold responsible for mediating ligand in F17G and UclD, I hypothesized that this helix may be a mobile gate that controls access to a hidden ligand-binding site. Although I did not observe helix movement using traditional molecular dynamics simulations, guided molecular dynamics performed by Max Zimmerman in the Bowman lab revealed that this loop can extend away from the core fold of the lectin domain (Figure 5.2C). The resulting pocket, which contains a mix of polar and nonpolar residues, could feasibly participate in ligand binding.

To identify the native ligand for the Yeh pilus following unsuccessful glycan array screening of YehD^{LD} (Figure 5.8), Jon Helander in the Janetka lab performed in silico docking experiments using both the crystallized and helix-opened state of YehD^{LD} (134). Screening against a deep, narrow pocket formed by the tip of the helices and the nearby β-sheets suggests that rhamnose, nigirin, fucose, and guanosine may be candidates for Yeh binding. However,
neither rhamnose of fucose resulted in a thermal shift when tested with differential scanning fluorimetry (data not shown). Docking analyses performed with the ‘open’ helical configuration identified substituted 5- and 6-membered rings (2-(Acetylamino)-2-Deoxy-a-D-glucopyranose, D-altro-D-manno-Heptose, and 4-O-hexopyranosylhex-2-ulofuranose) that hydrogen bond to R151 in the mobile helix. However, before further time and effort are spent analyzing the in vitro binding of these compounds to YehD, additional probabilistic analyses of the states identified via molecular dynamics are needed. Rough estimates suggest that the probability of the helix-loop-helix domain residing in any of the open states observed thus far is low, suggesting that these states are only transiently occupied. Longer simulations are currently underway to see if stable or metastable states are identified that would be more appropriate screening targets.

5.4.2 Structural and Biomechanical Studies of the Yeh Rod

To assess the overall morphology of the Yeh pilus rod and tip complex, Yeh pili were expressed from an inducible pTRC99a expression plasmid and subject to negative-stain electron microscopy for further analysis. Initial staining of cells with uranyl acetate revealed a series of short, rigid pili expressed from the surface of cells (Figure 5.9A). Cells expressing the empty vector without the Yeh operon appeared bald via EM (data not shown.) When piliated cells were stained with uranyl formate to decrease grain size and increase resolution, the majority of cell-associated pili appeared broken, suggesting that the pili may be fragile under traditional staining conditions. However, tip complexes approximately 17 nm in length could be observed on some pili (Figure 5.9B). This tip length would be consistent with a linear complex containing a single two-domain adhesin, which is predicted to be approximately 10.5 nm in its elongated conformation, and a single major subunit, which is predicted into be approximately 5 nm long. Additionally, a central clearing can be observed in the center of the pilus rods, indicating the
possible presence of a central clearing. These central clearings have been observed in the center of other pilus rods, including type 1 and P pili.

To generate a high-resolution structure of the Yeh pilus rod and better understand the intramolecular contacts present upon helical pilus assembly, we sent purified pilus samples to Dr. Ed Egelman and Dr. Weili Zheng at the University of Virginia for Cryo-EM analysis. Preliminary reconstructions of the Yeh pilus rod reveal a right-handed helix similar in structure to the type 1 helical rod previously-described (Figure 5.9C). Unfortunately, further refinement of the structure revealed that the pili that were imaged actually corresponded to contamination with type 1 pili. This finding was surprising, given that N-terminal sequencing and mass spectrometry of the pilus sample both revealed a large excess of Yeh pili over type 1 pili, and empty vector controls exhibited no pilus formation when visualized via negative-stain EM. Thus, expression of the Yeh pilus must be activating transcription of the type 1 pilus from its native operon, even in shaking conditions not normally conducive to type 1 pilus expression.

To compare the biomechanical properties of the Yeh pilus to other CUP pilus types, we collaborated with Dr. Magnus Andersson at the University of Umeå, Sweden to measure the force required to unwind the Yeh pilus (18, 19). Beads associated with the tip adhesin of cell-associated pili were manipulated using optical tweezers to apply lateral force away from the cell. Measurement of the biomechanical properties of the pilus rod have revealed that the unwinding force response shows three distinct phases; initial elastic stretching of the helix, helix unwinding, and elastic stretching of the tertiary structure of the linear fiber (Figure 5.10). However, because these pili may also correspond to type 1 pili contaminants, further study will be necessary using Yeh pili expressed in a UTI89Δfim background.

Once the correct experimental conditions required to provide quantitative measurement of the unwinding force are determined, a number of experiments could be performed to further
characterize the Yeh pilus. For example, previous studies have revealed that, at low velocities, the velocity of unwinding is independent of the force applied, but at a certain velocity, the force required to unwind the pilus increases logarithmically with velocity (19). The point at which these two unwinding regimes meet, called the corner velocity, often yields clues about the native environment in which Yeh pili are designed to function. Pili that must remain bound in the presence of strong, sustained external shear forces, such as type 1 pili in the bladder during urine expulsion, typically have fairly low corner velocities. Conversely, pili that must absorb short periods of strong force, such as those encoded by ETEC in the intestines, typically have higher corner velocities (19). Measurement of these parameters may provide a clue to Yeh function. Additionally, the effect of mutations to conserved residues located at the interface between laterally-associated subunits (ie, N and N+3) on the unwinding force could also be assessed to validate the importance of these interactions.

5.4.3 The YehB Chaperone and Donor Strand Complementation

Close examination of the subunit-subunit interactions mediated by donor strand exchange reveals a similar pattern to those previously observed in other pilus systems (27). Residues L8, I10, G12, and I14 of the YehA NTE likely occupy the P2-P5 pockets of the N+1 subunit, completing its hydrophobic core of the Ig-like fold through DSE. Sequence analysis of the YehB usher indicates an F-G loop length of 14 residues with a conserved asparagine at residue 89 and arginine at residue 113, suggesting that this chaperone belongs to the FGS family. Based on sequence analysis and homology modelling of the YehB chaperone, we would expect the M110, F108, L106, and N104 residues on the G1 strand of the chaperone to transiently occupy the P1, P2, P3, and P4 pockets, respectively, of the incoming pilin subunits, forming form an atypical Ig-like fold in which the G1 strand runs anti parallel to the A strand and parallel to the F strand (Figure 511). I have cloned the YehB chaperone into pTRC99a under the lactose promoter and the full-
length YehD adhesin into pBAD under the arabinose promoted to generate complexes to verify these predictions. This construct can also be used for the identification of possible binding partners, including glycan array screening, \textit{in vitro} pulldowns of homogenized colonic tissue, and differential scanning fluorimetry.

5.4.4 Structural Characterization of YhlD\textsuperscript{LD}

\textit{In vitro} and \textit{ex vivo} binding studies using YehD\textsuperscript{LD} and YhlD\textsuperscript{LD} revealed that YhlD\textsuperscript{LD} binds murine colon more intensely than YehD\textsuperscript{LD}. Based on the prevalence of the Yhl operon in EHEC strains and the observation that Yhl often co-exists with strains encoding a binding-attenuated FimH allele, I have begun efforts to generate a structure of the YhlD\textsuperscript{LD} to determine the molecular basis of intestinal recognition. Secondary structure prediction using the primary sequence of YhlD\textsuperscript{LD} suggests a similar overall topology to YehD\textsuperscript{LD}, including the presence of the helix-loop-helix domain. I have been able to grow thin needle-like crystals of YhlD\textsuperscript{LD} using 95 mM sodium citrate, pH 5.6, 5% glycerol, 19% PEG 4000, and 19% isopropanol. Although the diffraction pattern of these crystals was poor to the naked eye, the images processed in the I222 space group to a 2.1 Å. Unfortunately, despite the high degree of predicted structural homology, I was unable to obtain a valid molecular replacement solution using Phenix’s Phaser program using YehD as a search model.

To generate experimental phases for the YhlD\textsuperscript{LD} structure, I attempted to derivatize YhlD\textsuperscript{LD} crystals with a variety of heavy atoms, including iodine and platinum. For iodine, I attempted both iodine crystal vapor diffusion and potassium iodide soaks. Unfortunately, despite obtaining reasonable diffraction for some of the crystals tested, we were unable to locate an anomalous signal, suggesting that the iodine atoms were not consistently incorporated into the protein. I also placed small crystals of potassium tetrachloroplatinate transiently into the native crystal drops and allowed them to partially dissolve. However, a technical problem with the monochromator at the
beamline rendered us unable to collect data on those crystals to look for anomalous diffraction. Because the mature YhID<sup>LD</sup> sequence does not encode any methionine residues, I have begun mutating three sites to methionines. These positions were chosen because they correspond to the analogous positions in the Yeh protein where methionines reside. If cloning is successful and protein can be stable expressed, we will grow selenomethionine protein, crystallize it, and solve the structure with MAD.

YhID<sup>LD</sup> will also be tested in the CFG mammalian and bacterial glycan arrays to assess binding to common sugar moieties.

**5.4.5 YehC Usher Activation**

Because the structure of the YehD lectin domain is unique among crystallized lectin domains, structural study of other pilus components could elucidate novel mechanisms of pilus assembly, including usher activation and pore transit (149). Studies in other pilus systems have revealed that the chaperone-bound pilus adhesin can also serve as the initiator of pilus assembly by forming crucial interactions with the periplasmic N- and C-terminal domains of their cognate usher. Quantifying the formation of the Yeh pilus in the presence and absence of the YehD gene in the operon would allow us to determine if the adhesin is required for usher activation and initiation of assembly. If YehD is found to be necessary, directed mutagenesis (i.e., removing the helix-loop-helix domain or mutating residues in the distal β-β and α-β-loops) could be used to identify regions that interact with the usher to initiate assembly.

In addition to the glycomimetic antiadhesives discussed in Chapter 3, a second class of small molecular inhibits of pilus function known as pilicides are known to inhibit pilus biogenesis by targeting the interaction between the chaperone-adhesin complex and the usher during pilus activation (Omatage *et al*, under review). These compounds display activity against the biogenesis of both type 1 and P pili. To determine if the structural basis of usher activation and
pilus assembly is also conserved with the Yeh system, the activity of these compounds on usher biogenesis should be assessed by growing cells in the presence of inhibitory concentrations of pilicides and assessing piliation by negative stain EM in a Δfim background.

5.4.6 Yeh Pili as Therapeutic Targets

Because they are encoded by over 80% of all *E. coli* isolates, the Yeh pili may make attractive vaccine candidates to prevent infection in several niches by a variety of *E. coli* pathotypes. To assess this, one could immunize mice with the YehD lectin domain or the YehBD chaperone-adhesin complex and assess the susceptibility of these mice to colonization or infection in a variety of disease models, including acute and chronic cystitis, gastrointestinal colonization, sinusitis, and pneumonia. Because the expression of Yeh pili has not been spatially or temporally characterized, it is possible that vaccination could perturb the structure of the gastrointestinal microbiota and render the host susceptible to unwanted colonization. To assess this possibility, 16S sequencing should be performed before and after treatment, using mock-vaccinated mice as controls. If vaccination with a Yeh pilus component is able to decrease infection rates in any of the models tested, it could be a viable clinical candidate alone or as part of a multivalent vaccine against a variety of pilus types. Additionally, if conservation of usher activation and pilus assembly are observed, pilicides could also prove effective in inhibiting Yeh-mediated gastrointestinal colonization.
5.5 Materials and Methods

5.5.1 Evolutionary Adaption of the Pilus Rod

To create a his-tagged YehD lectin domain for structural studies, the DNA sequence corresponding to the C-terminal 195 amino acids representing the signal sequence and lectin domain of the UTI89 YehD adhesin were cloned between the EcoR1 and BamH1 sites of the expression vector pTRC99a with an C-terminal 6-His tag. The Yeh pilus operon from UTI89 was cloned into the pTRC99a plasmid under control of the lac operon at the EcoR1 restriction site using the InFusion protocol (Clontech Takara). Resulting plasmids were sequence-verified and transformed into C600 cells.

The his-tagged YhlD lectin domain for structural studies was generated by cloning the N-terminal 190 amino acids of the Sakai YhlD adhesin, representing the signal sequence and lectin domain of the Sakai YhlD adhesin, into pTRC99a using the InFusion protocol (Clontech Takara).

5.5.2 Protein Purification

The YehDLD and YhlDLD constructs were transformed into E. coli DL41(BE3), a methionine auxotroph, for expression. Cells were grown to an OD of 4.0 in LB broth + Superbroth (Fisher) in a New Brunswick fermenter, induced with 0.1 mM IPTG for 1 hour, then harvested by centrifugation at 3,500 x g for 20 minutes. Contents of the bacterial periplasm were isolated as previous described, dialyzed against 1 x PBS + 250 mM NaCl, and bound to a cobalt affinity column. Protein was eluted with 20 mM Tris 8.0 + 250 mM imidazole, fractions containing protein at the expected molecular weight diluted 4x, and the samples were bound to a cation exchange column (Mono S, GE Healthcare) and eluted in MES pH 6.5 + 500 mM NaCl.
5.5.3 Pilus Purification and Imaging

During expression, cells were grown to an OD600 of 0.6 under shaking conditions in liquid LB media. Pilus production was induced by 0.05 mM IPTG for one hour. Cells were washed 3x via centrifugation in 1 mM Tris, pH 8.0. Pili were removed from cells by shearing as previously described. Pilus identity following shear purification was confirmed with N-terminal sequencing (Midwest Analytical, Inc).

For ultrastructural analysis of whole bacteria by negative staining, bacteria were fixed with 1 % glutaradehyde (final concentration). Fixed bacteria were allowed to absorb onto freshly glow discharged, formvar/carbon-coated copper grids for 2 min. Grids were washed in dH2O (2X 30 sec) and stained with 1% aqueous uranyl acetate (Ted Pella Inc., Redding CA) for 1 min. Excess liquid was gently wicked off and grids were allowed to air dry. For subsequent analyses, uranyl acetate was replaced with uranyl formate to decrease grain size and increase contract. Staining and imaging were performed by Dr. Wandy Beatty and Greg Strout.

5.5.4 YehD Crystallography

Selenomethionine-labelled Yeh (12.7 mg/mL in 20 mM MES pH 5.8) was crystallized in 0.2 M AmH2PO4 + 20% w/v PEG 3350. Crystals were cryoprotected in a solution of motherliquor + 20% glycerol. Data were collected at the selenium peak at the ALS MBC beamline 4.2.2. Data were processed using XDS, aimless, and pointless, and the structure was phased single-wavelength anomalous diffraction in Phenix's Phaser program. Data were subject to iterative rounds of refinement using phenix.refine and Coot. Data and collection statistics for the YehDLD structure can be found in Table 5.1.

5.5.5 YehA Conservation Analysis

Conservation and evolution were analyzed as described previously (Spaulding, et al.) Briefly, YehA and YehD homologues were identified from the Ensembl Genomes Bacteria
database via the phmmer webserver using the UTI89 nucleotide sequences. Sequences with an E-value of $<10^{-10}$ were downloaded and aligned using Geneious R10. Duplicate sequences were removed, and sequence identities were calculated using Geneious R10.

To determine the selective pressure acting upon the YehA gene, a codon-based alignment was performed using the MAFFT FFT-NS-I x2 algorithm with the BLOSUM62 scoring matrix following by Pal2Nal webserver. Model selection and GARD recombination analyses were performed using the DataMonkey webserver. Positive/negative selection analysis was performed using the Fixed Effects Likelihood (FEL) model with a significance cutoff of $P < 0.1$.

### 5.5.6 Glycan Array Screening

6x-his tagged UclD<sup>LD</sup> was purified as described above and biotinylated using the NHS-PEG4-Biotin and Biotinylation Kits (ThermoFisher) as previously described. Biotinylated protein was dialyzed into TSM binding buffer (20 mM Tris–HCl, pH 7.4, 150 mM NaCl, 2 mM CaCl<sub>2</sub>, 2 mM MgCl<sub>2</sub>, 0.05% Tween 20, and 1% BSA). Protein was concentrated to 0.2 mg/mL, and sent to the CFG glycan array. Protein was detected using fluorescently-labelled streptavidin.
5.6 Figures and Tables

**Figure 5.1. Content, Conservation, and Phylogeny of the Yeh Operon.** A) The Yeh and Yhl operons each contain four genes, labelled YehA-YehD, respectively. B) Representation of the likely structure of the Yeh pilus. C) YehD homologues across all Gram-negative species were aligned and mapped to a maximum-likelihood phylogenetic tree.
Figure 5.2. X-ray crystal structure of the YehD lectin domain. A) X-ray crystal structure of the YehD lectin domain solved to 1.6 Å resolution. In addition to the canonical elongated β-barrel, YehD contains a helix-loop-helix domain not found in other CUP pilus lectin domains. B-C) Molecular dynamics simulations reveal motion of the helix-loop-helix domain. D) Electrostatic surface representation of the YehD crystal structure. E-F) Electrostatic surface representation of YehD with the helices flipped away from the body of the protein.
Figure 5.3. Homology Models of the Yeh Pilus Rod and DSE Reaction. A) Surface representation of a helical rod composed of YehA homology models based on the FimA structure within the helix. B) Homology modelling predicts that the residues L8, I10, G12, and I14 occupy the P2-P5 sites, respectively, of the following subunit.
Figure 5.4 Conservation and selection among YehA Residues. A) Sequence identity (bar height) and selective pressures (bar color) on YehA residues. Residues colored red are subject to negative (purifying) selection ($P < 0.1$). Residues colored blue are too conserved for selection to be measured. Residues colored green have no significant section. When these residues are mapped to a homology model of the Yeh pilus and viewed from the side (B), top (C), and bottom (D), one can observe that residues making interactions with adjacent pilin subunits are subject to negative, purifying selection.
Figure 5.5. YehD\textsuperscript{LD} and YhlD\textsuperscript{LD} binds murine colonic tissue \textit{ex vivo}. Murine colonic tissue was harvested from 8-week-old C3H/HeN mice, fixed in methacarn, boiled in sodium citrate to retrieve antigens, and stained with Hoechst and Alexa-Fluor-647-conjugated lectin domains or a PBS control. 

A) PBS Control.  
B) FimH\textsuperscript{LD} binding to murine colon tissue sections.  
C) YehD\textsuperscript{LD} binding to murine colon tissue sections.  
D) YhlD\textsuperscript{LD} binding to murine colon tissue sections.
Figure 5.6. YehD<sup>LD</sup> and YhlD<sup>LD</sup> bind colon-derived epithelial cells. Primary colon epithelial cells isolated from murine intestines were grown on transwells, blocked for endogenous biotin incubated with labelled adhesins, and fixed with methacarn. Following methacarn fixation, membranes were incubated with TRITC-conjugated UAE Lectin and Hoechst stain. A-D) PBS (no-adhesin LD) control. E-H) FimH<sup>LD</sup> positive control. I-L) YehD<sup>LD</sup>. M-P) YhlD<sup>LD</sup>.
Figure 5.7. Deletion of \textit{yeh} and \textit{yhl} in murine models of gastrointestinal colonization. (A) Concurrent infection of streptomycin-treated C3H/HeN mice with \textit{E. coli} UTI89 and UTI89\textit{Δyeh} reveals no competitive phenotype for the Yeh operon. (B) Concurrent infection of streptomycin- and DSS-treated C3H/HeN mice with \textit{E. coli} Sakai and Sakai\textit{Δyhl} reveals no competitive phenotype from the Yhl operon. (C) Single infection of 7-week of streptomycin- and 2\% wt/vol DSS-treated C3H/HeN mice with Sakai or Sakai\textit{Δyhl} reveals no colonization defect upon deletion of the yhl operon. Single infection of 7-week of streptomycin- and DSS-treated C3H/HeN mice with Sakai or Sakai\textit{Δyhl} reveals no colonization defect upon deletion of the Yhl operon. (C) Single infection of 7-week of streptomycin- and 3\% wt/vol DSS-treated C3H/HeN mice with Sakai or Sakai\textit{Δyhl} reveals no colonization defect upon deletion of the Yhl operon.
Figure 5.8. Glycan array screening of YehD^{LD}. Glycan array screening of YehD^{LD} against the Center for Functional Glycomics mammalian glycan array revealed no significant hits against and of the 609 glycan structures printed on the chip. The expected fluorescence of physiologically-relevant binders would be 5-10,000 relative fluorescence units or greater.
Figure 5.9. Metastructural analysis of Yeh pili. A) Cells induced to express Yeh pili from pTRC99a-Yeh were stained with uranyl acetate and visualized with negative-strain EM. B) High-resolution image of the a pilus tip found on Yeh-expressing cells stained with uranyl formate to improve grain size. C) Cryo-EM density of the Fim pilus rod contaminant, which reveals FimA subunits tightly wound into a right-handed helical rod with a twist of 115° per residue and a rise of 8 Å per residue. It is not yet known of all surface-associated pili observed via negative-stain EM correspond to the type 1 pilus contaminant.
Figure 5.10. Biomechanical properties of the Yeh pilus. The force-extension curve of the Yeh pilus was generated by applying increasing force to the tip of cell-associated Yeh pili using optical tweezers. As observed in other systems, pilus unwinding occurs in three distinct phases; an initial slope corresponding to elastic stretching of the helix, a plateau corresponding to helix unwinding, and a second upward slope corresponding to elastic stretching within the tertiary structure of the individual subunits.
Figure 5.11. Homology modelling reveals a probable mechanism for DSC. M110, F108, L106, and N104 residues on the G1 strand of the chaperone (cyan, spheres) to transiently occupy the P1, P2, P3, and P4 pockets of YehA (green).
Table 1. Collection and Refinement Statistics for the YehD lectin domain.

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>YehD&lt;sub&gt;LD&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>P2&lt;sub&gt;1&lt;/sub&gt;, 2, 2</td>
</tr>
<tr>
<td>Cell Dimensions</td>
<td></td>
</tr>
<tr>
<td>( a, b, c ) (Å)</td>
<td>50.3, 71.8, 102.6</td>
</tr>
<tr>
<td>( \alpha, \beta, \gamma ) (°)</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>58.8-1.6</td>
</tr>
<tr>
<td>( R_{\text{merge}} ) &lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.05 (0.43)</td>
</tr>
<tr>
<td>( I/\sigma(I) ) &lt;sup&gt;a&lt;/sup&gt;</td>
<td>21.7</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.2</td>
</tr>
</tbody>
</table>

**Refinement**

| No. Reflections                     | 96294             |
| \( R_{\text{work}} \)               | 0.212             |
| \( R_{\text{free}} \)               | 0.242             |
| No. atoms                            | 2908              |
| Protein                              | 2543              |
| Solvent                              | 365               |
| B Factors (Å<sup>2</sup>)            | 23.7              |

**R.m.s. Deviations**

| Bond lengths (Å)                     | 0.009             |
| Bond angles (°)                      | 1.15              |
Chapter 6: Conclusions

Bacteria express fibrous extracellular appendages to promote adhesion to and persistence within multiple environmental and host habitats. Because these processes are crucial determinants of bacterial virulence, inhibition of fiber formation or function represents a promising therapeutic approach to curb the proliferation of antibiotic resistance while circumventing the detrimental effects of broad-spectrum antibiotic use on the beneficial host microbiota. The preceding chapters describe novel insights into the structure, function, and assembly of two classes of extracellular fibers.

The first fiber class, curli, is a major structural component of many bacterial biofilms, which protect bacteria from mechanical, chemical, and immunological stressors. Curli fibers are composed of repeating amyloidogenic subunits that have structural and biophysical similarities to pathologic amyloids such as \( \alpha \)-synuclein, amyloid \( \beta \), and tau. The assembly of these amyloids on the outer membrane occurs via the nucleation precipitation pathway, in which unfolded CsgA structural subunits are shuttled through the bacterial periplasm by the accessory assembly protein CsgE prior to expulsion and polymerization. In this study, I sought to characterize the structural basis of the interaction between CsgE and CsgA. I found that CsgE binds CsgA and prevents premature polymerization using a series of charged surface residues. I also found that potentiation of the CsgE-CsgA interaction has detrimental effects on curli assembly, suggesting that release of CsgA through the CsgG pore requires a fine balance between association and dissociation. Investigation of the CsgE-CsgG interaction revealed that this interface is impervious to disruption by a single point mutant. I also conducted a series of preliminary studies examining the effects of surface mutants on CsgE assembly. As described above, future studies characterizing the interplay between CsgE oligomerization and the formation of the CsgE-CsgG
and CsgE-CsgA complexes could provide further insight into the multifaceted role of CsgE. The translational relevance of these studies is twofold. A better understanding of the mechanisms by which CsgE inhibits CsgA amyloid polymerization could be extended to pathogenic amyloid systems. In contrast to previous reports, I found that CsgE can inhibit the polymerization of α-synuclein \textit{in vitro}, and that mutations to CsgE surface residues modulate this inhibitory behavior. Expanding these studies to include amyloid β and tau could reveal novel mechanisms of amyloid inhibition that could inform the design of novel inhibitors of amyloid assembly. Additionally, a more thorough understanding of \textit{in vivo} biofilm assembly could inform the development of compounds that inhibit biofilm formation, thus rendering Gram-negative bacteria more susceptible to clearance.

I have also studied the structure and function of CUP pili. Like curli, CUP pili are proteinaceous extracellular fibers expressed by Gram-negative bacteria to promote surface adhesion and biofilm formation. Unlike curli, CUP pili possess adhesins at the tip of their fibers that recognize specific host ligands with stereochemical specificity. These ligands are typically glycans elaborated on host surfaces in a tissue- and niche-specific manner. Two CUP pilus types expressed by uropathogenic \textit{E. coli}, type 1 and fml, bind mannose and galactose-containing glycoproteins present in the bladder during acute and chronic cystitis, respectively. Early in my graduate career, I crystallized a potent competitive inhibitor of mannose binding, ZFH269, and determined its X-ray crystal structure. This compound was shown to be efficacious in the prevention and treatment of acute cystitis in mouse models of infection. Information gleaned about the structure-activity relationship of mannose analogues from this study informed the development of more potent mannosides with more favorable pharmacokinetic profiles. I also solved the X-ray crystal structure of three potent competitive inhibitors of galactose binding; AM2050, AM2119, and AM2134. The information gleaned from these studies will also be used to
develop more potent galactosides to treat and prevent chronic cystitis and pyelonephritis. Future efforts will be directed towards optimizing the interactions between residues near and adjacent to the FmlH adhesin’s active site and the galactoside molecule.

My work also included the study of two hitherto undercharacterized CUP pilus types; F17-like (Ucl) and Yeh. Both of these pilus types are located in the same evolutionary clade as known colonizers of the gastrointestinal tract, and thus, were hypothesized to mediate gastrointestinal colonization by \textit{E. coli}. I solved the X-ray crystal structure of the UclD and YehD adhesin lectin domains to determine the structural basis of ligand recognition. Although the F17-like pilus was indeed found to play a role in gastrointestinal colonization, I was unable to identify the molecular ligand recognized by UclD despite extensive \textit{in vitro} and \textit{in silico} screening. Thus, future studies should be geared towards identifying the molecular receptor for UclD to serve as the basis for the design of small molecules that inhibit F17-like pilus binding. While the structure of UclD\textsubscript{LD} revealed a canonical \(\beta\)-sandwich with a putative binding pocket located along the side of the core fold, the structure of YehD\textsubscript{LD} revealed a noncanonical helix-loop-helix domain extending down the side of the core fold with no readily-apparent active site. \textit{In vitro} and \textit{ex vivo} binding studies using YehD\textsubscript{LD} and its close phylogenetic relative, YhlD\textsubscript{LD}, reveal that these two pilus types are capable of binding murine colonic tissue. Future studies in the Yeh/Yhl pilus system should include phasing of an existing native YhlD\textsubscript{LD} data set, identification of a molecular ligand for both YehD and YhlD, assessment of the ability of pilus biogenesis inhibitors known as pilicides to prevent Yeh pilus assembly, and investigation of the YehBD chaperone-adhesin complex as a possible vaccine candidate. Together with the advances gained from studying curli biofilm biogenesis, these studies of CUP pilus structure and function have identified targets for antibiotic-sparing therapeutics by revealing novel molecular insights into microbial adhesion.
Works Cited


