

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

The Kinetic and Structural Investigation of Pilus Assembly and the Development of Sortase Inhibitors for Gram-Positive Bacteria

**Permalink**

<https://escholarship.org/uc/item/5k96k1r6>

**Author**

Sue, Christopher Kenji

**Publication Date**

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Kinetic and Structural Investigation of Pilus Assembly  
and the Development of Sortase Inhibitors for Gram-Positive Bacteria

A dissertation submitted in partial satisfaction of the  
requirements for the degree Doctor of Philosophy  
in Chemistry

by

Christopher Kenji Sue

2022

© Copyright by

Christopher Kenji Sue

2022

## ABSTRACT OF THE DISSERTATION

The Kinetic and Structural Investigation of Pilus Assembly  
and the Development of Sortase Inhibitors for Gram-Positive Bacteria

by

Christopher Kenji Sue

Doctor of Philosophy in Chemistry

University of California, Los Angeles, 2022

Professor Robert Thompson Clubb, Chair

Pathogenic multidrug resistant bacteria cause a range of serious infections in humans. These bacteria have developed mechanisms to counteract the lethal effects of currently used antibiotics, creating a need for novel therapeutics. Gram-positive bacteria display a wide assortment of cell surface proteins that are important for bacterial survival and host-pathogen interactions. These key surface structures included pili, proteinaceous fibers that assist in microbial survival by mediating adhesion to host tissues and aid in the formation of biofilm. A large number of gram-positive bacterial species assemble pili and append surface proteins to the cell wall using sortase cysteine transpeptidase enzymes. These enzymes link the components of the pilus together via covalent lysine isopeptide bonds which confer enormous tensile strength. This dissertation describes my investigation of the assembly mechanism of the

archetypal SpaA-pilus from *Corynebacterium diphtheriae*. It also describes my contributions to develop small molecule sortase inhibitors that could function as anti-infective agents and my work towards exploiting the activity of sortase enzymes as a protein engineering tool.

This thesis focuses on sortase enzymes and can be divided into two major sections: studies to determine how sortases construct pili (Chapters 2-4) and work designed to discover a small molecule inhibitor for the *Staphylococcus aureus* Sortase A enzyme (Chapter 5). All of the studies have been published in peer reviewed papers, with the exception of work detailed in Chapter 4. Chapter 2 describes the NMR solution structure of the lysine isopeptide bond interface that connects the pilin components of the pilus. This structural information combined with biophysical and cellular analyses led to the formulation of the “latch” mechanism of pilus assembly. Chapter 3 describes an enzyme kinetic study of the sortase enzyme from *C. diphtheriae* (<sup>Cd</sup>SrtA) which catalyzes the formation of lysine isopeptide bonds between components of the SpaA pilus. In this study, the rate-limiting step of catalysis was determined and variants of <sup>Cd</sup>SrtA with improved activity were discovered. Chapter 4 describes research that employed biochemical and structural approaches to investigate how the incorporation of the SpaB pilin subunit terminates pilus assembly. Key differences were observed between the reaction that terminates assembly and the process of polymerization that builds the shaft of the pilus. Chapter 5 describes efforts to discover a small molecule inhibitor of the *Staphylococcus aureus* Sortase A enzyme that has the potential to be a therapeutically useful anti-infective agent. The first half of Chapter 5 describes the work done to improve the activity of previously discovered pyridazinone-based molecules using computational and synthetic chemistry methods. The second half describes the implementation of a novel cell-based screen to identify sortase inhibitors. This work leveraged the unique sortase-dependent growth phenotype of *Actinomyces oris* to screen for sortase inhibitors. Over 200,000 small molecules were screened for their ability to impair *A.oris*'s growth, which led to the identification of three molecular

scaffolds that inhibit sortase activity *in vitro*. In total, my thesis research has shed light on how sortase enzymes assemble pili in gram-positive bacteria, led to improved variants of <sup>Cd</sup>SrtA sortase that can be used in protein engineering, and helped identify new small molecule sortase inhibitors with potential therapeutic applications.

The dissertation of Christopher Kenji Sue is approved.

Steven G. Clarke

Joseph Ambrose Loo

Jose Alfonso Rodriguez

Robert Thompson Clubb, Committee Chair

University of California, Los Angeles

2022

## DEDICATION

To my mother, who encouraged me to dream,  
my father, who always took me to the best science museums,  
my brothers and my sister who showed me the path forward  
my lab mates who pushed me forward, and  
to Jag for always being there for me.

“The most important step a man can take. It’s not the first one, is it?  
It’s the next one. Always the next.” – Brandon Sanderson (Oathbringer)



# Table of Contents

|   |           |
|---|-----------|
| <b>Introduction: The Roles of Sortase Enzymes in Bacterial Pilus Assembly .....</b>   | <b>1</b>  |
| 1.1 Overview .....  | 2         |
| 1.2 Introduction to Sortase Transpeptidase Enzymes .....  | 2         |
| 1.3 Class A Sortase Enzymes Append Proteins to the Cell Wall .....  | 4         |
| 1.4 Pilus Assembly by Class C Sortases.....   | 6         |
| 1.5 Sortase Inhibitors .....  | 11        |
| 1.6 Figures.....  | 15        |
| 1.7 References.....   | 20        |
| <br><b>Sortase-assembled Pili in <i>Corynebacterium diphtheriae</i> are Built Using a Latch</b>   |           |
| <b>Mechanism .....</b>  | <b>31</b> |
| 2.1 Overview .....  | 32        |
| 2.2 Sortase-assembled pili in <i>Corynebacterium diphtheriae</i> are Built Using a Latch Mechanism<br>.....                                       | 33        |
| 2.3 Supplemental Information .....  | 44        |
| <br><b>Kinetics and Optimization of the Lysine-Isopeptide Bond-Forming Activity of the Pilin</b>  |           |
| <b>Sortase from <i>Corynebacterium diphtheriae</i>.....</b>   | <b>67</b> |
| 3.1 Overview .....  | 68        |
| 3.2 Kinetics and Optimization of the Lysine-Isopeptide Bond Forming Activity of the Pilin Sortase<br>from <i>Corynebacterium diphtheria</i> ..... | 69        |

|  |            |
|--|------------|
| <b>Unique Structural and Kinetic Features of the Terminating Pilin Subunit from <i>Corynebacterium diphtheriae</i></b> ..... | <b>81</b>  |
| 4.1 Overview .....   | 82         |
| 4.2 Introduction .....   | 83         |
| 4.3 Results and Discussion .....   | 84         |
| 4.4 Materials and Methods .....  | 95         |
| 4.5 Figures .....  | 100        |
| 4.6 Tables .....   | 110        |
| 4.7 References .....   | 114        |
| <b>The Search for Sortase Inhibitors</b> .....   | <b>118</b> |
| 5.1 Overview .....   | 119        |
| 5.2.1 NMR structure-based optimization of <i>Staphylococcus aureus</i> sortase A pyridazinone inhibitors .....               | 121        |
| 5.2.2 Supplementary Information .....  | 140        |
| 5.3.1 A Cell-based Screen in <i>Actinomyces oris</i> to Identify Sortase Inhibitors .....                                    | 149        |
| 5.3.2 Supplementary Information .....  | 161        |

# List of Figures

## Chapter 1: Introduction: The Roles of Sortase Enzymes in Bacterial Pilus Assembly

|  |    |
|--|----|
| <b>Figure 1.1</b> Overview of class A and class C sortases   | 15 |
| <b>Figure 1.2</b> The proposed molecular mechanism of class A Sortase enzymes from <i>S. aureus</i> .          | 16 |
| <b>Figure 1.3</b> Representative class A and C sortase structures  | 17 |
| <b>Figure 1.4</b> Pilus polymerization is catalyzed by sortase enzymes in <i>Corynebacterium diphtheriae</i> . | 19 |

## Chapter 2: Sortase-assembled pili in *Corynebacterium diphtheriae* are built using a latch mechanism

|  |    |
|--|----|
| <b>Figure 2.1</b> Structure of the <sup>N</sup> SpaA-signal peptide complex                                    | 35 |
| <b>Figure 2.2</b> The AB loop undergoes a disordered-to-ordered transition upon crosslinking                   | 37 |
| <b>Figure 2.3</b> SAXS model of the crosslinked interpilin interface   | 38 |
| <b>Figure 2.4</b> <i>In vitro</i> and <i>in vivo</i> validation of key residues on the SpaA acceptor domain    | 39 |
| <b>Figure 2.5</b> Latch mechanism of sortase-catalyzed pilus biogenesis  | 41 |
| <b>Figure S2.1</b> Enzymatic production of the <sup>N</sup> SpaA-signal complexes                              | 44 |
| <b>Figure S2.2</b> NMR relaxation data for apo- <sup>N</sup> SpaA and <sup>N</sup> SpaA-signal peptide complex | 46 |
| <b>Figure S2.3</b> Representative intermolecular NOE data  | 48 |

|  |    |
|--|----|
| <b>Figure S2.4</b> Multiple sequence alignment of pilin proteins   | 50 |
| <b>Figure S2.5</b> Contact map summarizing protein-peptide interactions in the NMR structure of the <sup>N</sup> SpaA-signal complex                                       | 52 |
| <b>Figure S2.6</b> <sup>N</sup> SpaA thermostability and proteolytic stability   | 53 |
| <b>Figure S2.7</b> NMR relaxation data indicates that the <sup>C</sup> SpaA- <sup>N</sup> SpaA dimer is rigid  | 55 |
| <b>Figure S2.8</b> Small angle X-ray scattering data of the <sup>C</sup> SpaA- <sup>N</sup> SpaA complex   | 57 |
| <b>Figure S2.9</b> The structure of the <sup>C</sup> SpaA- <sup>N</sup> SpaA dimer is compatible with NMR data and distinct from packing interactions observed in crystals | 59 |
| <b>Figure S2.10</b> <i>In vivo</i> biochemical and electron microscopic analysis of <sup>C</sup> SpaA- <sup>N</sup> SpaA interfacial mutants                               | 62 |
| <b>Figure S2.11</b> <i>In vitro</i> analysis of transpeptidation rates and proteolytic stability of <sup>C</sup> SpaA- <sup>N</sup> SpaA interfacial mutants               | 63 |
| <b>Figure S2.12</b> Expanded HPLC kinetics data  | 65 |

### **Chapter 3: Kinetics and Optimization of the Lysine–Isopeptide Bond Forming Sortase Enzyme from *Corynebacterium diphtheriae***

|   |    |
|---|----|
| <b>Figure 3.1</b> <i>C. diphtheriae</i> <sup>Cd</sup> SrtA pilin sortase catalyzes lysine isopeptide bond formation | 71 |
| <b>Figure 3.2</b> <sup>Cd</sup> SrtA transpeptidation assay   | 72 |
| <b>Figure 3.3</b> Characterization of <sup>Cd</sup> SrtA <sup>3M</sup> and <sup>Cd</sup> SrtA <sup>Δ</sup>          | 73 |
| <b>Figure 3.4</b> Role of the X residue in transpeptidation   | 74 |
| <b>Figure 3.5</b> Labeling activity of <sup>Cd</sup> SrtA <sup>3M</sup> and <sup>Cd</sup> SrtA <sup>Δ</sup>         | 75 |

## Chapter 4: Unique Structural and Kinetic Features of the Terminating Pilin Subunit from *Corynebacterium diphtheriae*

|  |     |
|--|-----|
| <b>Figure 4.1</b> <i>In vitro</i> testing of lysine specificity of SpaB  | 100 |
| <b>Figure 4.2</b> Measurement of activity <sup>Cd</sup> SrtA of SpaB   | 101 |
| <b>Figure 4.3</b> Competition Reaction between <sup>N</sup> SpaA and SpaB with <sup>Cd</sup> SrtA <sup>Δ</sup> | 103 |
| <b>Figure 4.4</b> Structure of the SpaB Pilin  | 104 |
| <b>Figure 4.5</b> Biochemical study of the pilin motif in SpaB   | 106 |
| <b>Figure 4.6</b> SpaB does not appear to ligate to <sup>N</sup> SpaA through its CWSS                         | 108 |
| <b>Figure 4.7</b> Investigating how mutations to <sup>C</sup> SpaA's E-box affects SpaB incorporation          | 109 |

## Chapter 5: The Search for Sortase Inhibitors

### Section 5.2 NMR Structure Based Optimization of *Staphylococcus aureus*

#### Sortase A Pyridazinone Inhibitors

|  |     |
|--|-----|
| <b>Figure 5.2.1</b> NMR spectra of SrtA:2-salt complex   | 127 |
| <b>Figure 5.2.2.</b> NMR solution structure of the SrtA:2-salt complex                         | 129 |
| <b>Figure 5.2.3</b> LC-MS traces of the SrtA:2-salt complex                                    | 130 |
| <b>Figure 5.2.4</b> Overview of the molecular docking and experimental testing process         | 130 |
| <b>Figure 5.2.5</b> Inhibitors inactivation kinetics and effect on protein display             | 132 |
| <b>Figure 5.2.6</b> Comparison of the SrtA:2-salt structure with apo- and holo-SrtA Structures | 134 |
| <b>Figure 5.2.7</b> Docking poses of compounds 2-61 and 2-62                                   | 136 |

|  |     |
|--|-----|
| <b>Figure 5.2.S1</b> Overlay of the <i>in vitro</i> dose-response curves of 2-10 (red), 2-17 (orange), 2-54 (green) and 2-62 (blue). | 140 |
| <b>Figure 5.2.S2</b> General scheme  | 141 |
| <b>Section 5.3 A Cell-based Screen in <i>Actinomyces oris</i> to Identify Sortase Inhibitors</b>                                     |     |
| <b>Figure 5.3.1</b> Design and overall work-flow of cell-based inhibitor screen  | 151 |
| <b>Figure 5.3.2</b> Growth effects of the screened molecules   | 153 |
| <b>Figure 5.3.3</b> Chemical structures of preliminary hit molecules   | 154 |
| <b>Figure 5.3.4</b> Assessment of cell surface proteins  | 155 |
| <b>Figure 5.3.5</b> Detection of pili by electron microscopy   | 156 |
| <b>Figure 5.3.S1</b> Full length Gels for Figure 5.3.4   | 161 |

## List of Tables

### Chapter 2: Sortase-assembled pili in *Corynebacterium diphtheriae* are built using a latch mechanism

|   |    |
|---|----|
| <b>Table S1</b> Structural statistics of the solution structure of <sup>N</sup> SpaA-signal complex | 49 |
|---|----|

### Chapter 3: Kinetics and Optimization of the Lysine–Isopeptide Bond Forming Sortase Enzyme from *Corynebacterium diphtheriae*

|   |    |
|---|----|
| <b>Table 3.1</b> Kinetics of <sup>Cd</sup> SrtA Catalyzed Lysine-Isopeptide Formation | 72 |
|---|----|

### Chapter 4: Unique Structural and Kinetic Features of the Terminating Pilin Subunit from *Corynebacterium diphtheriae*

|   |     |
|---|-----|
| <b>Table 4.1</b> Kinetics of <sup>Cd</sup> SrtA <sup>Δ</sup> catalysis between pilin subunits | 110 |
|---|-----|

|  |     |
|--|-----|
| <b>Table 4.2</b> Thermostability of Sortase Proteins | 111 |
|--|-----|

|  |     |
|--|-----|
| <b>Table 4.3</b> Structural statistics of the solution structure of SpaB | 112 |
|--|-----|

|  |     |
|--|-----|
| <b>Table 4.4</b> DALI analysis of SpaB | 113 |
|--|-----|

### Chapter 5: The Search for Sortase Inhibitors

#### Section 5.2 NMR Structure Based Optimization of *Staphylococcus aureus*

#### Sortase A Pyridazinone Inhibitors

|  |     |
|--|-----|
| <b>Table 5.2.1</b> Statistics for the NMR modeled structure of SrtA bound to a pyridazinone inhibitor 2-salt | 128 |
|--|-----|

|   |     |
|---|-----|
| <b>Table 5.2.2</b> SrtA inhibition of the pyridazinone compound derivatives | 131 |
|---|-----|

|   |     |
|---|-----|
| <b>Table 5.2.3</b> Inactivation kinetics of pyridazinone compound derivatives | 132 |
|---|-----|

### **Section 5.3 A Cell-based Screen in *Actinomyces oris* to Identify Sortase Inhibitors**

**Table 5.3.1** Growth and inhibitory properties of the preliminary hit molecules 154



## Acknowledgements

I would like to thank my advisor, Professor Robert T. Clubb, for the opportunity to let me learn in his laboratory, for his innovative ideas, and for his motivation for when I needed encouragement. Additionally, I would like to thank my collaborators Professors Hung Ton-That, Joseph Loo, Michael Jung and the members of their lab for providing advice, resources, direction, and data for many of my projects. I thank the members of my dissertation committee, Professors Steven Clarke, Joseph Loo, and Jose Rodriguez, for providing me advice and feedback throughout my graduate career as well as reviewing this dissertation. I would like to thank the members of the Clubb lab for their support and friendship over the years, including Dr. Scott McConnell, Dr. Brendan Mahoney, Dr. Kat Ellis-Guardiola, Dr. Orlando Martinez, Dr. Brendan Amer, Dr. Grace Huang, Dr. Ramsay Macdonald, Jason Gosschalk, Dr. Albert Chan, Jess Soule, Andrew Goring, Allen Takayesu, Jordan Ford, Nikki Cheung, and Christine Minor. Thank you also to my wonderful undergrads Ana Alvarez, Justin Yu, and Jack Scully; I appreciate all your hard work and time you put into the lab. Most importantly, I would like to thank my family and friends for their unending kindness and giving me lots of food. Thank you for always making sure I was fed and checking in on me. Finally, I would like to thank the Cellular and Molecular Biology Training Grant for its financial support.

Chapter Two of this dissertation is a version of a published manuscript: Sortase-assembled pili in *Corynebacterium diphtheriae* are built using a latch mechanism. McConnell, S.A., McAllister, R.A., Amer, B.R., Mahoney, B.J., Sue, C.K., Chang, C., Ton-That, H., and Clubb, R. T. Proc. Nat. Acad. Sci. USA. 118,(12). Mar 2021. Reproduced with permission from the Proceedings of the National Academy of Sciences. We express appreciation to the Professor Ton-That and his laboratory for their valuable collaborative studies which contributed to this manuscript.

Chapter Three of this dissertation is a version of a published manuscript: Kinetics and Optimization of the Lysine-Isopeptide Bond Forming Activity of the Pilin Sortase from *Corynebacterium diphtheriae*. Sue, C. K., McConnell, S.A., Ellis-Guardiola, K., Muroski, J., McAllister, R. A., Yu, J., Alvarez, A., Ogorzalek Loo, R. R., Loo, J. A., Ton-That, H. and Clubb, R. T. *Bioconjugate Chemistry*. May 2020. Reproduced with permission from the *Journal of Bioconjugate Chemistry*. We express appreciation to Dr. Loo and members of his laboratory for their valuable collaborative studies which contributed to this manuscript.

Chapter Four of this dissertation is ongoing progress towards understanding the termination of the SpaA pili in *Corynebacterium diphtheriae* through the structural and kinetic understanding of the basal pilin SpaB. Work in this chapter will be used for a future publication. We express appreciation to Dr. Loo and members of his laboratory for their assistance.

Chapter Five of this dissertation is reformatted from two published manuscripts: NMR structure-based optimization of *Staphylococcus aureus* sortase A pyridazinone inhibitors. Albert H. Chan, Sung Wook Yi, Ethan M. Weiner, Brendan R. Amer, Christopher K. Sue, Jeff Wereszczynski, Carly A. Dillen, Silvia Senese, Jorge Z. Torres, J. Andrew McCammon, Lloyd S. Miller, Michael E. Jung, and Robert T. Clubb. *Chemical Biology and Drug Design*. February 2017, and A Cell-based Screen in *Actinomyces oris* to Identify Sortase Inhibitor. Jason E. Gosschalk, Chungyu Chang, Christopher K. Sue, Sara D. Siegel, Chenggang Wu, Michele D. Kattke, Sung Wook Yi, Robert Damoiseaux, Michael E. Jung, Hung Ton-That and Robert T. Clubb. *Scientific Reports*. May 2020. Reproduced with permission from *Journal of Chemical Biology and Drug Design* and *Scientific Reports*. We express appreciation to Professors Torres, Jung, and Ton-That and members of their laboratory for their assistance.

## Vita

Christopher Kenji Sue received primary and secondary education in Rancho Palos Verdes, California completing high school in 2012. Chris enrolled at Westmont College in Santa Barbara in the Fall of 2012. In 2016, he earned a Bachelor of Science degree in Chemistry with honors. At Westmont College he earned the summer research award in both 2014 and 2015. In 2016, Chris entered UCLA through the Chemistry Doctoral Program with a chemical biology specialty. While in graduate school, he received financial support from the Cellular and Molecular Biology Training Grant. Chris is an author on four papers and has presented on several posters.

### Publications Resulting From this Dissertation Research

Sortase-assembled pili in *Corynebacterium diphtheriae* are built via a latch-like mechanism. Scott A. McConnell, Rachel A. McAllister, Brendan Amer, Brendan Mahoney, **Christopher K. Sue**, Chungyu Chang, Hung Ton-That and Robert T. Clubb. Proceedings of the National Academy of Sciences. March 2021

Kinetics and Optimization of the Lysine-Isopeptide Bond Forming Activity of the Pilin Sortase from *Corynebacterium diphtheriae*. **Christopher K. Sue**, Scott A. McConnell, Ken Ellis-Guardiola, John M. Muroski, Rachel A. McAllister, Justin Yu, Ana I. Alvarez, Chungyu Chang, Rachel R. Ogorzalek Loo, Joseph A. Loo, Hung Ton-That, and Robert T. Clubb. Bioconjugate Chemistry. May 2020.

A Cell-based Screen in *Actinomyces oris* to Identify Sortase Inhibitor. Jason E. Gosschalk, Chungyu Chang, **Christopher K. Sue**, Sara D. Siegel, Chenggang Wu, Michele D. Kattke,

Sung Wook Yi, Robert Damoiseaux, Michael E. Jung, Hung Ton-That and Robert T. Clubb.  
Scientific Reports. May 2020.

NMR structure-based optimization of *Staphylococcus aureus* sortase A pyridazinone inhibitors.  
Albert H. Chan, Sung Wook Yi, Ethan M. Weiner, Brendan R. Amer, **Christopher K. Sue**, Jeff  
Wereszczynski, Carly A. Dillen, Silvia Senese, Jorge Z. Torres, J. Andrew McCammon, Lloyd S.  
Miller, Michael E. Jung, and Robert T. Clubb. Chemical Biology and Drug Design. February  
2017.

## **Chapter 1**

### **Introduction: The Roles of Sortase Enzymes in Bacterial Pilus Assembly**

## 1.1 Overview

The evolution of multidrug resistance among pathogenic bacteria threatens to destabilize our current arsenal of antibiotic treatments.<sup>1</sup> Due to overuse in agriculture and patient care, bacteria have gained resistance and become immune to almost all known antibiotic treatments.<sup>2-4</sup> By 2050, antibiotic-resistant bacterial infections are forecasted to cause more fatalities than cancer.<sup>5</sup> An example of this growing threat is the methicillin-resistant *Staphylococcus aureus* (MRSA), which in the United States causes more than 320,000 infections and 11,000 deaths annually.<sup>6</sup> In addition to the growing resistance to commonly used antibiotics, the number of new antibiotics developed and approved has slowed.<sup>4</sup> From 1980 to 1990, 30 new antibiotics were introduced for medical use, while only 13 were approved by the FDA between 2000 and 2014. Moreover, the majority of these “new” antibiotics are analogs of existing drug classes, so resistance mechanisms may still impart some immunity to these “novel” treatments.<sup>4</sup> Due to the lack of newly discovered antibiotic classes and growing resistance to currently used antibiotics, antibiotic resistant bacteria are a serious threat that needs to be addressed.<sup>4</sup> To combat this growing threat, there needs to be more study and development of antibiotics that target alternative pathways.

## 1.2 Introduction to Sortase Transpeptidase Enzymes

Many clinically important pathogens are gram-positive bacteria that use sortase enzymes to elaborate their surfaces with protein virulence factors.<sup>7</sup> Amongst the most alarming strains of antibiotic resistant bacteria according to the CDC are the ESKAPE pathogens.<sup>8</sup> Three of these: MRSA,  $\beta$ -lactam-resistant *Streptococcus pneumoniae*, and vancomycin-resistant *Enterococcus faecium* have been classified by the World Health Organization as high priority pathogens that require the development of new antibiotics.<sup>9</sup> These microbes and other gram-positive bacteria possess a single membrane that is surrounded by a cell wall containing a thick

peptidoglycan layer which protects the cell from extracellular stressors.<sup>10</sup> This organelle also protects these organisms from osmotic lysis while allowing for the passage of essential nutrients and serves as a scaffold for various proteins and glycopolymers.<sup>11</sup> In pathogens, these embedded proteins can function as virulence factors and have roles in nutrient acquisition, immune system modulation, and bacterial adhesion.<sup>12</sup>

Gram-positive bacteria in the Firmicutes and Actinobacteria phyla utilize sortase-dependent pathways to display surface proteins.<sup>13</sup> Sortases catalyze a transpeptidation reaction that either attaches surface proteins to the cell wall or assembles filamentous protein pili (**Figure 1.1**). Sortases that attach proteins to the cell wall link together a five residue C-terminal cell wall sorting signal (CWSS) and a reactive nucleophilic group found in lipid-II, a precursor used in cell envelope biosynthesis.<sup>14</sup> After sortase-mediated attachment to lipid-II, proteins are displayed on the surface when lipid-II is incorporated into the cell wall. Pilus assembly sortases operate through similar mechanisms by reacting with proteins containing a CWSS, but are distinct because they employ a nucleophilic lysine sidechain that is located within another protein as their second reactant. These unique sortases link together protein subunits to construct pili, long filamentous structures attached via covalent lysine isopeptide bonds, resulting in structures that have enormous tensile strength.<sup>15</sup>

Based on their primary sequences, there are at least eight types of sortase enzymes, designated as class A, B, C, D1, D2, E, F, and Marine sortases.<sup>16,17</sup> This introduction will primarily focus on class A and C sortases, which are the main focus of the research described herein. Class A enzymes attach proteins to the cell wall. Many of these sortase-attached surface proteins are classified as members of the microbial surface components recognizing adhesive matrix molecules (MSCRAMM) family and are virulence factors.<sup>18,19</sup> A diverse range of gram-positive bacteria within the Firmicutes phylum contain at least one class A sortase that acts as a house-keeping enzyme, attaching a large number of distinct proteins to the cell

surface.<sup>16,20–22</sup> In a number of pathogens, class A enzymes have been shown to be important for infectivity in mouse models, but eliminating sortase's activity does not affect bacterial growth in cell cultures.<sup>23–28</sup> Meanwhile, class C sortases polymerize pili on the surface of gram-positive bacteria by linking together protein subunits through lysine isopeptide bonds.<sup>29</sup> Removal of pili-catalyzing sortases results in diminished adherence to host cells. Furthermore, pili are potential vaccine targets due to their exposure on the surface.<sup>30,31</sup> Therefore, sortases are attractive targets for the development of anti-infectives that work by disrupting the ability for the microbe to establish an infection, but otherwise do not affect microbial growth outside the host. These anti-infective therapeutics could result in less selective pressure for evolutionary adaptations and create a new class of antibiotic therapies.<sup>32</sup>

There have been many excellent reviews on sortase enzymes.<sup>13,29,33–35</sup> In this introduction, I review what is known about two archetypal sortase enzymes, the class A sortase from *Staphylococcus aureus* (<sup>Sa</sup>SrtA) and the class C sortase from *Corynebacterium diphtheriae* (<sup>Cd</sup>SrtA), respectively. I also review recent research to discover sortase inhibitors.

### 1.3 Class A Sortase Enzymes Append Proteins to the Cell Wall

The first discovered and most well studied sortase is the class A enzyme from *S. aureus* (**Figure 1.1A**).<sup>36</sup> The catalytic mechanism of <sup>Sa</sup>SrtA currently forms the basis for our understanding for the sortase family (**Figure 1.2**). <sup>Sa</sup>SrtA substrates are transported through the inner membrane through the Sec translocon, which targets an N-terminal signal peptide consisting of 15 to 20 hydrophobic residues.<sup>37</sup> After translocation, signal peptidases cleave the signal peptide from the substrate, resulting in the substrate folding into its mature conformation.<sup>38</sup> Sortase substrates remain tethered to the membrane via a transmembrane helix within their CWSS, which also contains an exposed LPXTG peptide sequence that is recognized by sortase. Sortase catalysis occurs via a two-step process of thioesterification followed by peptide bond formation. During the first step, <sup>Sa</sup>SrtA recognizes the LPXTG



sequence and uses its catalytic cysteine residue (C184) to nucleophilically attack the carbonyl carbon between the threonine and glycine residues.<sup>39</sup> This results in the transient formation of an oxyanion tetrahedral intermediate that has been proposed to be stabilized by a nearby arginine residue (R197).<sup>40</sup> The catalytic histidine residue (H120) has then been postulated to donate a proton to the leaving group facilitating peptide bond cleavage and tetrahedral intermediate collapse into a longer-lived thio-acyl intermediate. In the semi-stable thio-acyl intermediate, the threonine at the C-terminal end of the protein is linked to the cysteine residue in sortase. Peptide bond formation then occurs when the primary amine group located on lipid-II cell wall intermediate enters the active site and nucleophilically attacks the thio-acyl linkage. In this reaction, H120 is thought to deprotonate the amine to facilitate its attack on the carbonyl carbon within the thio-acyl intermediate.<sup>41,42</sup> This creates a second tetrahedral intermediate that resolves into the final protein:lipid-II product of the reaction. If water nucleophilically attacks the containing the thio-acyl intermediate instead of attack by lipid-II, sortase acts as a protease cleaving the LPXTG sequence between the threonine and glycine residues via hydrolysis.

Insight into the molecular basis of <sup>Sa</sup>SrtA-mediated catalysis has been revealed by the determination of atomic structures of the enzyme in its free and peptide-bound states.<sup>43-45</sup> The enzymes in these structures are missing the N-terminal transmembrane helix which likely anchors the enzyme to the cell membrane and positions the catalytic domain towards the extracellular surface.<sup>43,46</sup> The catalytic domain of <sup>Sa</sup>SrtA and other sortases contains a closed eight-stranded  $\beta$ -barrel structure (**Figure 1.3A**). In the solution NMR structure of <sup>Sa</sup>SrtA the coordinates of the catalytic domain are precisely defined, except for 19 amino acids within a flexible loop that connects strands  $\beta_6$  to  $\beta_7$  ( $\beta_6/\beta_7$  loop). This unstructured loop transitions from an open to closed conformation in the presence of its CWSS peptide substrate, and has been shown to be important for substrate recognition.  $\text{Ca}^{2+}$  has also been shown to be important for <sup>Sa</sup>SrtA activity *in vitro* as its presence increases both hydrolysis and transpeptidation activity.<sup>47</sup>

NMR experiments show that  $\text{Ca}^{2+}$  binds to a pocket generated by the  $\beta 3/\beta 4$  and  $\beta 6/\beta 7$  loop regions. This binding quenches  $\beta 6/\beta 7$  loop motions and stabilizes it in the active configuration, explaining why the presence of  $\text{Ca}^{2+}$  increases activity. The  $\text{S}^{\text{a}}\text{SrtA}$  active site is composed of a triad of cysteine (C184), arginine (R197), and histidine (H120) residues positioned adjacent to one another within the active site.<sup>48,49</sup> The NMR structure of  $\text{S}^{\text{a}}\text{SrtA}$  bound to a sorting signal mimic revealed that the substrate binds to a groove proximal to the active site cysteine. Strands  $\beta 4$  and  $\beta 7$  compose the floor of the groove and the loops  $\beta 6/\beta 7$ ,  $\beta 7/\beta 8$ ,  $\beta 3/\beta 4$ , and  $\beta 2/\text{H}1$  form the walls.<sup>13,45</sup> A chimeric enzyme was constructed that swapped in the  $\beta 6/\beta 7$  loop from sortase B from *Staphylococcus aureus* for the loop region in  $\text{S}^{\text{a}}\text{SrtA}$  and changed the specificity  $\text{S}^{\text{a}}\text{SrtA}$  from LPXTG to sortase B's NPQTN CWSS.<sup>50</sup> This modification allowed cleavage of the NPQTN CWSS, but did not alter specificity for the nucleophile. Thus,  $\beta 6/\beta 7$  loop is an important determinant for dictating the enzyme's sorting signal specificity.

#### 1.4 Pilus Assembly by Class C Sortases

Pili (also called fimbriae) are long filamentous structures found on the surface of bacteria.<sup>34,51,52</sup> They mediate interactions with the extracellular environment and are frequently essential for attachment to host cells and intra- and inter-species bacterial interactions.<sup>53</sup> Bacteria in the Firmicutes and Actinobacteria phyla use class C sortase enzymes to build pili, which covalently crosslink the protein subunits of the pilus (called pilins) via lysine isopeptide bonds (**Figure 1.1B**).<sup>54</sup> These bonds confer high tensile strength allowing pili to endure significant forces.<sup>55-57</sup> Disruption to pilus formation in *C. diphtheriae* results in diminished adherence to host epithelial cells.<sup>30</sup> Pili have also been found to be essential for the formation of biofilms which are self-produced matrices of extracellular polymeric substrates.<sup>58</sup> These biofilms are important for host colonization by aiding in immune evasion and protection from antibiotics.<sup>59,60</sup> Studies in the gram-positive organism *Enterococcus faecalis* showed that mutations which prevented pilus formation result in reduced biofilm formation in cells and

attenuated virulence in a rat animal model of infection.<sup>58</sup> Pili components are also considered to be attractive antigens to produce vaccines to prevent infections as some pili are surface exposed and important for bacterial survival and host colonization.<sup>61</sup> For example, studies in which mice exposed to several hundred individual recombinant proteins from group B *Streptococcus* showed that immunization with pilin-derived antigens resulted in immunological protection against 12 GBS strains.<sup>31</sup>

The work described in this thesis focuses on the assembly of the SpaA pilus in the pathogen *C. diphtheriae*, which causes pharyngeal diphtheria. *C. diphtheriae* displays three distinct pili: SpaA-pili formed from SpaA, SpaB and SpaC pilins; SpaD-pili formed from SpaD, SpaE, and SpaF pilins; and SpaH-pili formed from SpaH, SpaG, and SpaI pilins.<sup>15,62,63</sup> Each type of pilus is assembled with one or more dedicated sortase enzymes that are co-expressed with the pilin components and are pilus-specific (e.g. sortases that build the SpaA pilus cannot assemble SpaH-pili and vice versa).<sup>63,64</sup> However, each type of pili is attached to the cell wall by the action of the same class A sortase, SrtF.

The archetypal SpaA pilus forms long (1-2  $\mu\text{m}$ ) and thin (1-2 nm in diameter) fibers, and functions to assist in binding to pharyngeal cells.<sup>15</sup> It is composed of three subunits: (i) a tip pilin, SpaC, (ii) a main shaft pilin, SpaA, and (iii) a basal anchor pilin, SpaB. The SpaA pilus is assembled by the class C sortase ( $\text{C}^{\text{d}}\text{SrtA}$ ) which catalyzes the formation of isopeptide bonds between pilins.<sup>30,65</sup> The assembly of the SpaA pilus begins with the Sec-dependent secretion of its pilins (SpaA, SpaB and SpaC) which remain bound to the extracellular surface via the membrane-spanning helix within their CWSS (**Figure 1.4**).<sup>12</sup>  $\text{C}^{\text{d}}\text{SrtA}$  begins polymerization when the sortase recognizes the LPLTG sorting signal found in the CWSS on the SpaC pilin.  $\text{C}^{\text{d}}\text{SrtA}$  cleaves the peptide bond between the threonine and glycine residues to form a thio-acyl  $\text{C}^{\text{d}}\text{SrtA}$ -SpaC intermediate. This intermediate then reacts with a second  $\text{C}^{\text{d}}\text{SrtA}$ -SpaA thio-acyl intermediate that is formed in a similar manner. In this reaction, the SpaA K190 lysine within the

$^{Cd}SrtA$ -SpaA thio-acyl intermediate nucleophilically attacks  $^{Cd}SrtA$ -SpaC, cleaving the thioester bond such that the K190 side chain in SpaA is joined to the C-terminal threonine residue in SpaC. The SpaC subunit is transferred to form a  $^{Cd}SrtA$ -SpaA-SpaC thio-acyl complex. After this step, SpaA proteins are progressively added to the base of the pilus in a similar manner in which the K190 nucleophile originating from a  $^{Cd}SrtA$ -SpaA thio-acyl intermediate attacks the thio-acyl linkage between another  $^{Cd}SrtA$  enzyme and the growing pilus. This process occurs 100-250 times in cells.<sup>62</sup> Polymerization is terminated when SpaB enters the reaction, presumably as a  $^{Cd}SrtF$ -SpaB acyl intermediate. In this reaction, SpaB's lysine K139 side chain is used as a nucleophile by the  $^{Cd}SrtA$  enzyme within the thioacyl intermediate that contains the nearly assembled pilus ( $^{Cd}SrtA$ -(SpaA)<sub>n</sub>-SpaC, where "n" represents the number of SpaA proteins that form the shaft of the pilus).<sup>66,67</sup> This final step transfers the pilus to the SrtF enzyme to create a SrtF-SpaB-(SpaA)<sub>n</sub>-SpaC intermediate. SrtF is capable of recognizing the diaminopimelic acid moiety within the cell wall precursor lipid II, using it as a nucleophile to create covalently linked Lipid II-SpaB-(SpaA)<sub>n</sub>-SpaC product.<sup>68</sup> Lipid II is a precursor for cell wall synthesis, such that transglycosylation and transpeptidation reactions that build the cell wall incorporate the pilus into the peptidoglycan mesh where it is covalently attached to the diaminopimelic group.<sup>69</sup> In cells lacking SrtF most of the pili are secreted into the milieu because they are not attached to the cell wall.<sup>66</sup> However, a small fraction does appear to be attached suggested that the  $^{Cd}SrtA$  can perform this function, albeit less efficiently than SrtF. Assembly of the SpaD- and SpaH-pili are believed to occur in a generally similar manner as the SpaA pilus. However, the shaft of these pili is constructed using two pilus sortases and attached to the cell wall by SrtF.<sup>70</sup>

The structure of SpaA pilin has been determined and is composed of three tandem Ig-type domains: the N-domain (residues 54-192), the M-domain (residues 193-351), and the C-domain (residues 352-484).<sup>71</sup> (PDB: 3HTL) The M- and C-domains contain stabilizing internal

isopeptide bonds between lysine  $\epsilon$ -amino groups and the carboxamide groups of asparagine (K199 and N321 in the M-domain and K363 and N482 in the C-domain), which have also been observed in other pilin structures.<sup>24</sup> The lysine and asparagine are located in a hydrophobic environment that favors a nonprotonated lysine and protonated asparagine and are positioned adjacent to a conserved glutamate residue that likely catalyzes the formation of the internal isopeptide bonds.<sup>71</sup> One of these catalytic glutamates (E446) in the C-domain of SpaA is contained in an E-box motif (YxLxETxAPxGY) which is conserved in many pilin proteins.<sup>65,71</sup> In addition, E446 has been shown to be essential for the attachment of SpaB and SpaC to the pilus, but the reason for this remains unknown.<sup>65</sup> SpaA's N-terminal domain does not contain an internal isopeptide bond but is present in other pilins, such as the N-terminal domain in the BcpA pilin from *Bacillus cereus*.<sup>72</sup> Based on studies of the *S. pyogenes* spy0128 pilin, it is generally believed that the function of these internal isopeptide bonds is to increase the tensile strength of pilins, as well as their thermostability and resistance to proteolysis. The reactive lysine (K190) in SpaA is part of a highly conserved pilin sequence motif (WxxxVxVYPK) that is found in many pilin proteins.<sup>54,65</sup> Chapter 2 discusses how residues in the pilin motif form a binding groove for the LPXTG sorting signal and how mutations in this sequence element affects rate of crosslinking *in vitro* and in cells.

In the SpaA pilus, incorporation of the SpaB pilin is thought to terminate pilus assembly and promote its attachment to the cell wall.<sup>73</sup> The SpaB primary sequence does not contain many of the key motifs found in other pilins, such as SpaA, that were thought to be essential for ligation to a specific lysine residue.<sup>67</sup> Yet despite lacking these motifs, SpaB has been found to make a specific lysine isopeptide bond with its K139 residue to a CWSS. It currently remains unclear how <sup>Cd</sup>SrtA recognizes SpaB and forms isopeptide bonds. Furthermore, EM studies from the Ton-That group (UCLA) have shown that SpaB is infrequently incorporated into the shaft of the pilus.<sup>62</sup> This is believed to be important in allowing host cell adhesion as SpaB and

SpaC have been shown to mediate adherence to pharyngeal epithelial cells.<sup>30,74</sup> If SpaB is present throughout the pilus, the SpaA pilus can potentially have multiple points that can adhere to host cells. The interspersed SpaB may facilitate a zippering mechanism to bring bacteria closer to the host cell surface. As such, further work on understanding how SpaB's K139 is recognized and how SpaB is incorporated into the pilus is described in Chapter 4.

The structure of the <sup>Cd</sup>SrtA shows a characteristic beta barrel fold similar to that found in the class A sortase from *S. aureus* (**Figure 1.3B**).<sup>43,67</sup> However, <sup>Cd</sup>SrtA and many other class C sortases are unique because they contain a N-terminal "lid" region that obscures the active site, as well as a C-terminal transmembrane helix. In many of the structures of class C enzymes the lid structure contains a DPW amino acid sequence that interacts with the active site (PDB: 5K9A).<sup>67</sup> In the crystal structure of <sup>Cd</sup>SrtA, the tryptophan residue (W83) participates in aromatic stacking with the active site cysteine (C222) and histidine (H160) residues, while the aspartic acid residue (D81) interacts with an active site arginine residue (R231). When both W83 and D81 are substituted with glycine, transpeptidation activity is observed *in vitro*. This is presumably because these mutations destabilize the structure of the lid, consistent with the absence of electron density for these atoms in the crystal structure of the mutant enzyme (PDB: 6BWE). Current models of its function suggest that the lid has an inhibitory function by masking the enzymes active site and binding surface for CWSS. The W83G and D81G substitutions destabilize the lid structure, thereby allowing substrate access and enabling <sup>Cd</sup>SrtA lysine isopeptide forming activity to be reconstituted *in vitro*. Work discussed in Chapter 3 reveals that the enzyme can be further activated by introducing additional destabilizing mutations into the lid. These mutants were also employed to gain insight into the structure of the crosslinked interface in the SpaA pilus (chapter 2).

## 1.5 Sortase Inhibitors

After the initial discovery of penicillin, a large arsenal of antibiotics have been developed to treat bacterial infections.<sup>75,76</sup> After decades of use, bacteria have evolved resistance to antibiotics, creating a pressing need for new therapeutics.<sup>76,77</sup> As previously mentioned, sortases are attractive drug targets because they are found in a wide range of gram-positive bacterial pathogens and have been shown to be important virulence factors. The *S. aureus* sortase A (<sup>Sa</sup>SrtA) has been the primary focus of inhibitor screening efforts because of the clinical importance of MRSA and other multi-drug resistant strains of *S. aureus*.<sup>8,78</sup> Therefore, many drug discovery studies have been conducted to identify a <sup>Sa</sup>SrtA inhibitor. Because sortase inhibitors have been screened for many years, this introduction will review the most common methods and approaches in discovering a <sup>Sa</sup>SrtA inhibitor over the past 5 years, and how they differ from the cell-based approach described in chapter 5.<sup>78,79</sup> Below I also provide a few examples in which these approaches were used. Many additional screens have also been performed and have been reviewed.<sup>78-80</sup>

The main strategies for discovering compounds that inhibit <sup>Sa</sup>SrtA have been high throughput screening using a Förster resonance energy transfer (FRET) assay, the creation of reactive peptide mimics, *in silico* computational screening, and cell-based screening using *Actinomyces oris*.<sup>78</sup> The majority of screens that have been published so far use a FRET assay to test sortase activity.<sup>81</sup> The FRET assay involves the use of a fluorophore and quencher pair that is separated by a peptide sequence corresponding to CWSS for <sup>Sa</sup>SrtA. When sortase is uninhibited, the linkage between the pair is hydrolyzed releasing the fluorophore and quencher resulting in an increase in fluorescence. The first published FRET-based screen was performed by the Kim group (Seoul National University) and identified an inhibitor containing a diacylacrylonitrile scaffold out of 1000 tested molecules.<sup>81</sup> Using structure-activity relationship (SAR) analyses they improved the potency of the original hit molecule from an IC<sub>50</sub> of 231 ±

6.530  $\mu\text{M}$  to  $9.244 \pm 0.474 \mu\text{M}$ . They then conducted computational docking to simulate the optimized molecule and found that it preferred to bind to a large hydrophobic binding pocket adjacent to the active site of the enzyme, and that it did not directly interact with the catalytic C184 residue in the enzyme. The Clubb group also utilized the FRET assay to screen ~30,000 molecules for *Sa*SrtA inhibitors.<sup>82</sup> Three molecules containing pyridazinone or pyrazolethione scaffolds were inhibitory, with  $K_i$  values in the micromolar range. Docking studies showed that pyridazinone molecules bound to a structured hydrophobic pocket that contains residues from the  $\beta 6/\beta 7$  loop and the  $\beta 8$  strand. The inhibitors were not bactericidal and also inhibited sortases from *Bacillus anthracis*, suggesting that they could be general sortase inhibitors. Work described in Chapter 5 of this dissertation describes the use of structural information to improve the potency of these pyridazinone-based inhibitors. These efforts led to molecules that were 70-fold more potent than the initial lead molecule.

Other groups have attempted to create reactive peptide mimics of the CWSS substrate in order to inhibit the enzyme.<sup>83</sup> The McCafferty group (Duke) developed a phosphinate-containing peptidomimetic  $\text{NH}_2\text{-YALPE-Ala}\Psi\{\text{PO}_2\text{H-CH}_2\}\text{Gly-EE-NH}_2$ . The peptidomimetic exhibited very poor inhibition of enzyme ( $\text{IC}_{50} = 10 \text{ mM}$ ), presumably because it did not accurately mimic the transition state. Peptide macrocycles have also been explored as *Sa*SrtA inhibitors and can bind the enzyme active sites with high affinity.<sup>84</sup> The Heinis group (Ecole Polytechnique Fédérale de Lausanne) developed a bicyclic peptide containing a Leu-Pro-Pro which attempted to imitate the LPXTG motif.<sup>85</sup> Their best bicyclic peptide (ACPLLPPCSLDCG) had an impressively low  $K_i$  value of  $1.5 \pm 0.4 \mu\text{M}$  based on FRET measurements, but when tested *in vivo* using a fluorescein-labeled LPETG substrate it was far less potent ( $\text{IC}_{50}$  of  $167 \mu\text{M}$ ). Its low potency was suggested to be caused by insufficient diffusion through the bacterial cell wall to the membrane where *Sa*SrtA is located.

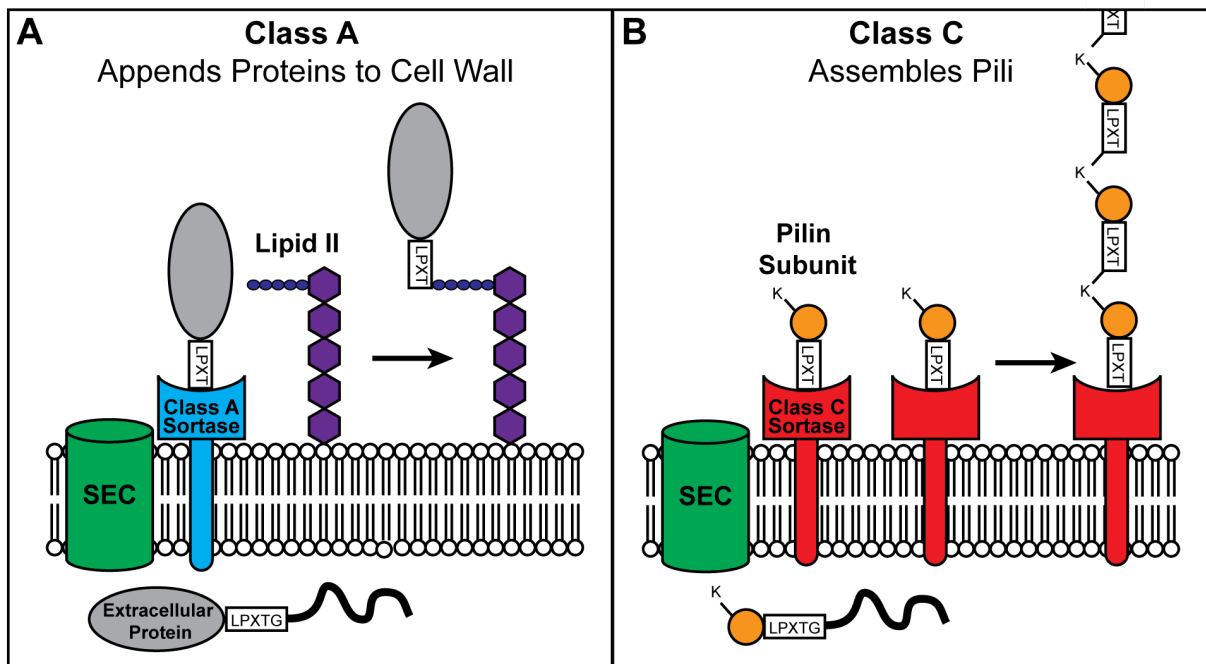


Computational screening methods which virtually dock potential inhibitors to a portion of <sup>Sa</sup>SrtA structure have also been used to identify and optimize sortase inhibitors. The first *in silico* screen for sortase was conducted by the Velu group (University of Alabama at Birmingham). They screened 150,000 compounds from Maybridge and Chembridge compound libraries against sortase's active site using the FlexX software package integrated into SYBYL.<sup>86</sup> This docking program identified 108 compounds as potential inhibitors, and using the FRET enzymatic assay the pool was narrowed to eight compounds with potentially strong inhibitor activity. Using SAR analysis, their best lead molecule was optimized and improved from an IC<sub>50</sub> of 75 ± 4.1 μM to 58 ± 4.9 μM. The Ng group (Singapore Institute of Food and Biotechnology) also screened a computational library that contained mostly natural products from the A\*STAR Natural Organism Library.<sup>87</sup> They identified two promising compounds, N1287 and N2576, which had fairly low IC<sub>50</sub> values of 24 ± 2.1 μM to 47 ± 4.2 μM, respectively. The molecules also prevent biofilm formation, but the minimum inhibitory concentrations of N1287 and N2576 are lower than their IC<sub>50</sub> values, which suggests that they may be inactivating processes other than <sup>Sa</sup>SrtA activity.

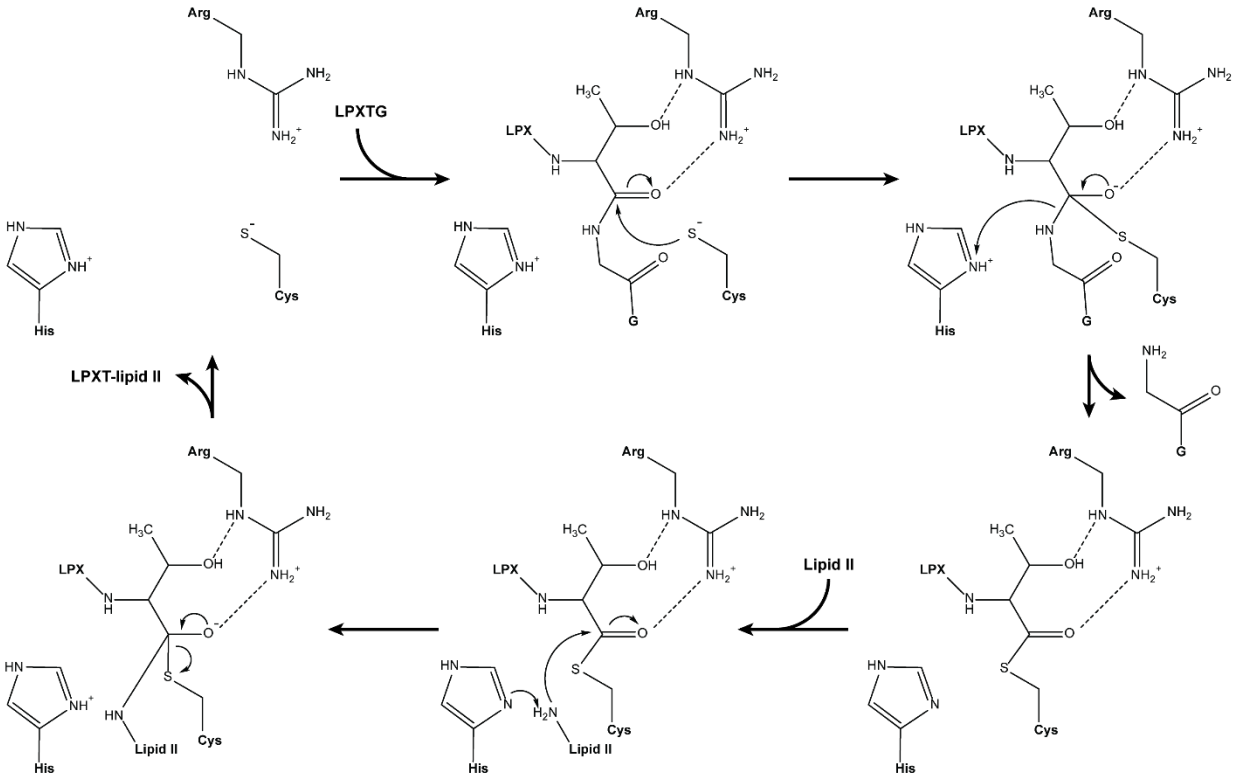
Due to the complexity of the bacterial cell wall, the FRET assay or *in silico* screening methods may be unable to accurately model the sortase system in its natural state. Thus, when these inhibitors are incubated with pathogenic organisms in a host animal, the molecule is unable to effectively inhibit the enzyme. Recently, *Actinomyces oris*, an oral bacterium, was found to require a functional sortase enzyme (<sup>Ao</sup>SrtA) in order to be viable in cell culture.<sup>88,89</sup> This is an unusual phenotype, and presumably occurs because the <sup>Ao</sup>SrtA specifically anchors a glycosylated surface protein to the cell wall. Elimination of <sup>Ao</sup>SrtA activity causes the accumulation of glycosylated proteins in the membrane leading to cell death via glycol-related stress. Using this unique growth dependence on <sup>Ao</sup>SrtA, we created an *in vivo* screen and counter-screen for inhibitors of <sup>Ao</sup>SrtA that worked by monitoring their effects on the growth

*A. oris*. Screening >200,000 molecules, we found three novel <sup>Sa</sup>SrtA inhibitors. Further discussion of this work is provided in Chapter 5.

## 1.6 Figures

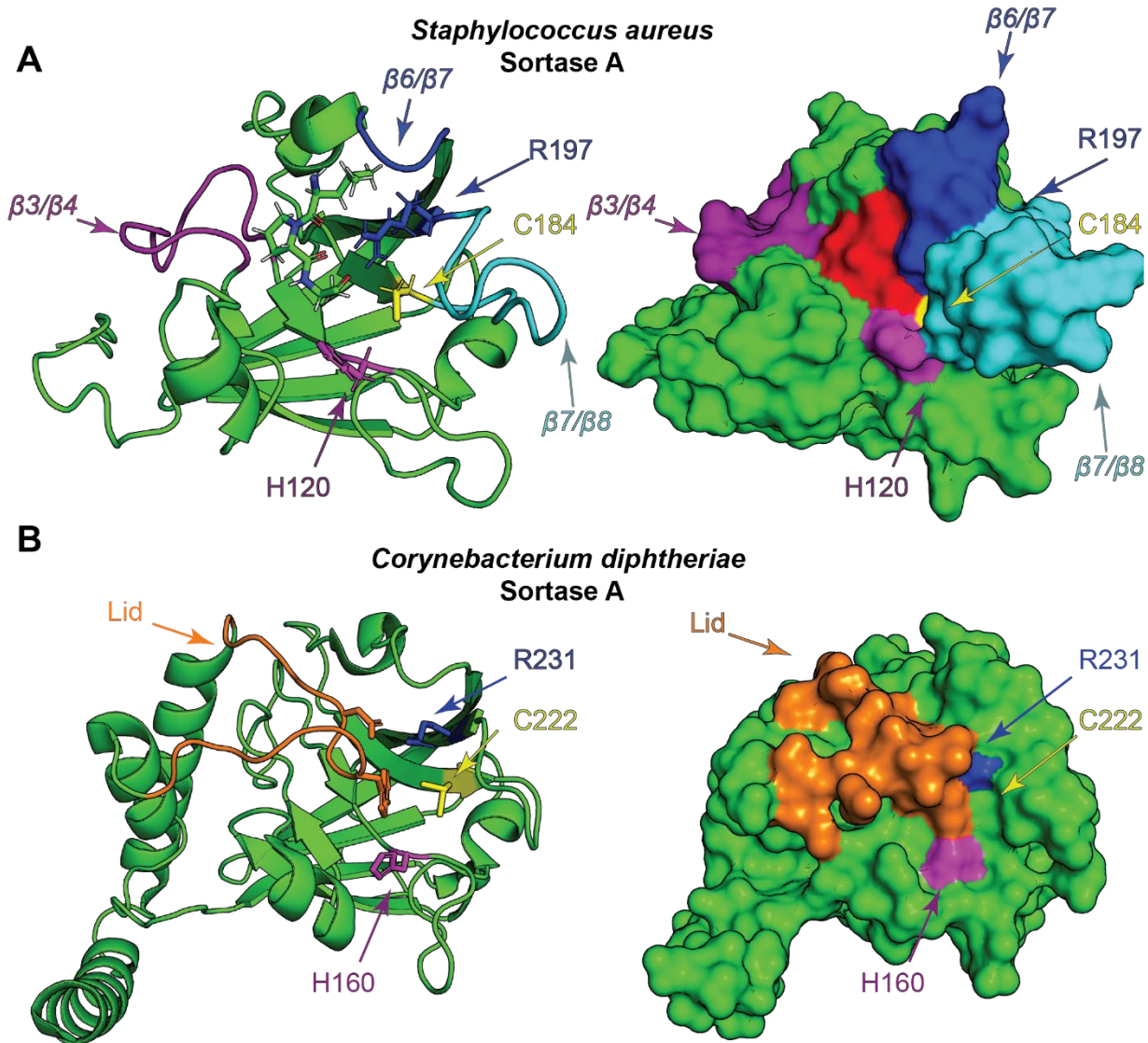


**Figure 1.1 Overview of class A and class C sortases.** (A) Substrates are first secreted across the cytoplasmic membrane through the SEC translocon. Class A sortases (blue) recognize proteins (pink) containing an LPXTG motif on their C-terminus and forms a thiol-acyl intermediate with the CWSS. Class A sortases then catalyze a reaction between the CWSS and a primary amine group found on a cell wall precursor, lipid II (purple). (B) Class C sortases (red) recognize pilin subunits (orange) containing an LPXTG CWSS, then a specific lysine residue found on the subunits, and link the CWSS to the lysine side chain's amine group. This process is repeated resulting in the polymerization of pili.



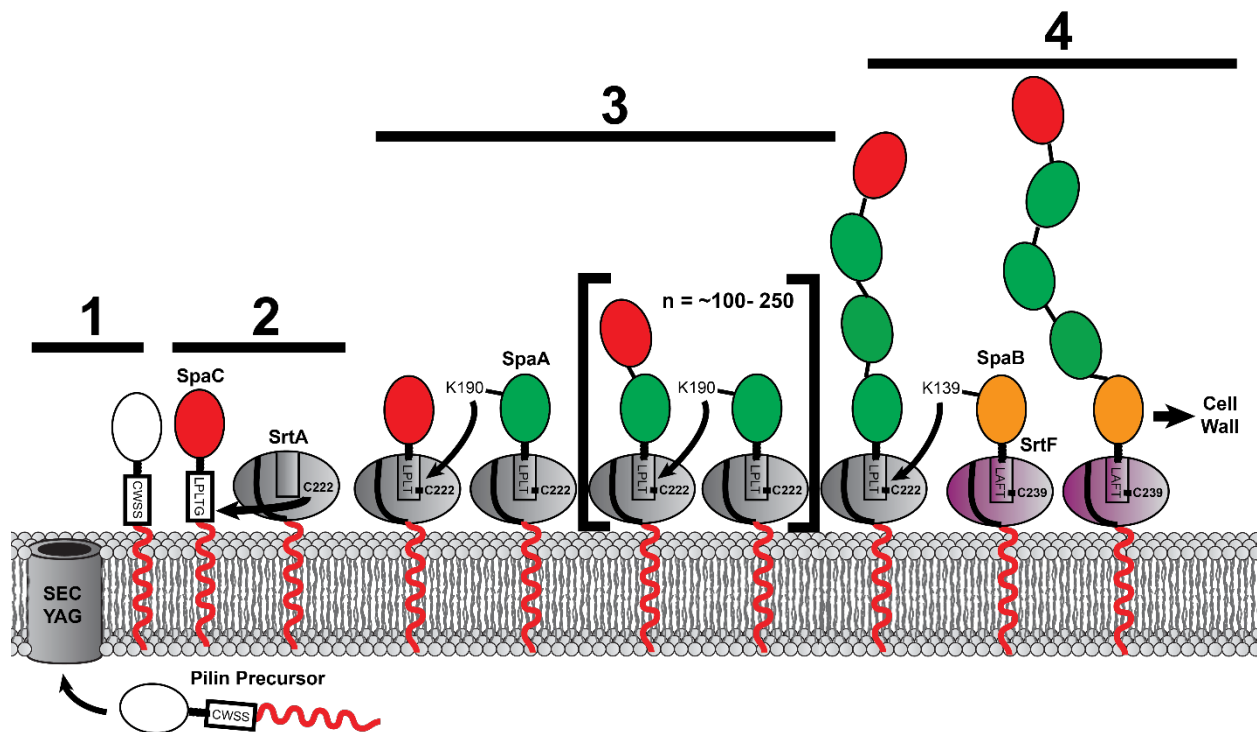
**Figure 1.2** The proposed molecular mechanism of class A Sortase enzymes from *S.*

*aureus*. Sortase transpeptidation follows a “ping-pong” mechanism where the catalytic cysteine residue recognizes a unique sorting signal motif and nucleophilically attacks the carboxyl group between the threonine and glycine of the CWSS forming a tetrahedral intermediate that is potentially stabilized by a positively-charged arginine residue via an oxyanion hole. The tetrahedral intermediate then collapses when the proton from histidine is transferred to the glycine forming a semi-stable thio-acyl intermediate between the cysteine and the threonine residue. Lipid II is the second substrate and contains an amine group that enters the active site. Histidine then deprotonates the amine group which helps to facilitate the substrate’s nucleophilic attack on the thio-acyl bond. This forms a second tetrahedral intermediate which collapses to create a new peptide bond. (Figured adapted from reference 13 with permission)



**Figure 1.3 Representative class A and C sortase structures.** All sortase structures contain an eight stranded beta barrel motif. (A) Structure of the class A *Staphylococcus aureus* Sortase A (PDB:2KID) showing three protein loops that control substrate recognition. (B) The structure of the class C *Corynebacterium diphtheriae* Sortase A (PDB: 5K9A). It has a similar fold as the class A enzyme, but contains a unique lid motif (colored in orange) that covers the active site and has previously hindered *in vitro* studies. Class C enzymes also contain a unique C-terminal

transmembrane helix (not shown). The active sites of both structures are composed of a histidine (colored in purple), an arginine (colored in blue) and a cysteine (colored in gold).



**Figure 1.4 Pilus polymerization is catalyzed by sortase enzymes in *Corynebacterium diphtheriae*.** (1) Pilin subunits containing an N-terminal signal peptide are directed and transported to the Sec translocon for secretion. The N-terminal signal peptide is cleaved by signal peptidases and pilin subunits are partially secreted but remain attached to the cell membrane via a transmembrane helix. (2) A pilin sortase ( $^{Cd}$ SrtA) recognizes SpaC through its extracellular 5-residue LPLTG motif. The cysteine residue on  $^{Cd}$ SrtA cleaves the LPLTG resulting in the formation of a thio-acyl intermediate. (3) A nearby SpaA molecule resolves the acyl-intermediate by nucleophilically attacking with its  $\epsilon$ -amine group on a reactive lysine (K190). This results in the formation of an isopeptide linkage between SpaA subunits and this process is repeated 100 to 250 times. (4) A SpaB subunit completes polymerization through the formation of a final isopeptide bond and is later ligated to the cell wall by SrtF, the housekeeping sortase.

## 1.7 References

- (1) Aslam, B.; Wang, W.; Arshad, M. I.; Khurshid, M.; Muzammil, S.; Rasool, M. H.; Nisar, M. A.; Alvi, R. F.; Aslam, M. A.; Qamar, M. U.; Salamat, M. K. F.; Baloch, Z. Antibiotic Resistance: A Rundown of a Global Crisis. *Infect. Drug Resist.* **2018**, *11*, 1645–1658. <https://doi.org/10.2147/IDR.S173867>.
- (2) Manyi-Loh, C.; Mamphweli, S.; Meyer, E.; Okoh, A. Antibiotic Use in Agriculture and Its Consequential Resistance in Environmental Sources: Potential Public Health Implications. *Molecules* **2018**, *23* (4). <https://doi.org/10.3390/molecules23040795>.
- (3) Klein, E. Y.; Van Boeckel, T. P.; Martinez, E. M.; Pant, S.; Gandra, S.; Levin, S. A.; Goossens, H.; Laxminarayan, R. Global Increase and Geographic Convergence in Antibiotic Consumption between 2000 and 2015. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (15), E3463–E3470. <https://doi.org/10.1073/pnas.1717295115>.
- (4) Abraham, E. P. The Antibiotics. *Compr. Biochem.* **1963**, *11* (4), 181–224. <https://doi.org/10.1016/B978-1-4831-9711-1.50022-3>.
- (5) Shankar, Pr. Review on Antimicrobial Resistance : Tackling Drug-Resistant Infections Globally. *Arch. Pharm. Pract.* **2016**, *7* (3), 110.
- (6) CDC. *Antibiotic Resistance Threats in the United States, 2019, Atlanta, GA: U.S. Department of Health and Human Services*; 2019.
- (7) Pishesha, N.; Ingram, J. R.; Ploegh, H. L. Sortase A: A Model for Transpeptidation and Its Biological Applications. *Annu. Rev. Cell Dev. Biol.* **2018**, *34*, 163–188. <https://doi.org/10.1146/annurev-cellbio-100617-062527>.
- (8) Mulani, M. S.; Kamble, E. E.; Kumkar, S. N.; Tawre, M. S.; Pardesi, K. R. Emerging Strategies to Combat ESKAPE Pathogens in the Era of Antimicrobial Resistance: A Review. *Front. Microbiol.* **2019**, *10* (APR), 539. <https://doi.org/10.3389/fmicb.2019.00539>.
- (9) WHO, W. H. O. WHO (2017) Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics. *Cad. Pesqui.* **2017**, *43* (148), 348–365.
- (10) Jubeh, B.; Breijyeh, Z.; Karaman, R. Resistance of Gram-Positive Bacteria to Current Antibacterial



- Agents and Overcoming Approaches. *Molecules* **2020**, *25* (12), 2888.  
<https://doi.org/10.3390/molecules25122888>.
- (11) Navarre, W. W.; Schneewind, O. Surface Proteins of Gram-Positive Bacteria and Mechanisms of Their Targeting to the Cell Wall Envelope. *Microbiol. Mol. Biol. Rev.* **1999**, *63* (1), 174–229.  
<https://doi.org/10.1128/membr.63.1.174-229.1999>.
- (12) Spirig, T.; Weiner, E. M.; Clubb, R. T. Sortase Enzymes in Gram-Positive Bacteria. *Mol. Microbiol.* **2011**, *82* (5), 1044–1059. <https://doi.org/10.1111/j.1365-2958.2011.07887.x>.
- (13) Jacobitz, A. W.; Kattke, M. D.; Wereszczynski, J.; Clubb, R. T. Sortase Transpeptidases: Structural Biology and Catalytic Mechanism. *Adv. Protein Chem. Struct. Biol.* **2017**, *109* (2), 223–264. <https://doi.org/10.1016/bs.apcsb.2017.04.008>.
- (14) Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O. Staphylococcus Aureus Sortase, an Enzyme That Anchors Surface Proteins to the Cell Wall. *Science (80-. )*. **1999**, *285* (5428), 760–763. <https://doi.org/10.1126/science.285.5428.760>.
- (15) Ton-That, H.; Schneewind, O. Assembly of Pili in Gram-Positive Bacteria. *Trends Microbiol.* **2004**, *12* (5), 228–234. <https://doi.org/10.1016/j.tim.2004.03.004>.
- (16) Comfort, D.; Clubb, R. T. A Comparative Genome Analysis Identifies Distinct Sorting Pathways in Gram-Positive Bacteria. *Infect. Immun.* **2004**, *72* (5), 2710–2722.  
<https://doi.org/10.1128/IAI.72.5.2710-2722.2004>.
- (17) Malik, A.; Kim, S. B. A Comprehensive in Silico Analysis of Sortase Superfamily. *J. Microbiol.* **2019**, *57* (6), 431–443. <https://doi.org/10.1007/s12275-019-8545-5>.
- (18) Foster, T. J.; Geoghegan, J. A.; Ganesh, V. K.; Höök, M. Adhesion, Invasion and Evasion: The Many Functions of the Surface Proteins of Staphylococcus Aureus. *Nature Reviews Microbiology*. Nat Rev Microbiol January 2014, pp 49–62. <https://doi.org/10.1038/nrmicro3161>.
- (19) Ann L Coker; Nalawansa, Dhanusha A. Pflum, M. K. 乳鼠心肌提取 HHS Public Access. *Physiol. Behav.* **2017**, *176* (5), 139–148. <https://doi.org/10.1002/bip.21472.Sortase>.
- (20) Pallen, M. J.; Lam, A. C.; Antonio, M.; Dunbar, K. An Embarrassment of Sortases- A Richness of Substrates? *Trends Microbiol.* **2001**, *9* (3), 97–101. [https://doi.org/10.1016/S0966-842X\(01\)01956-](https://doi.org/10.1016/S0966-842X(01)01956-)

- 4.
- (21) Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O. Staphylococcus Aureus Sortase, an Enzyme That Anchors Surface Proteins to the Cell Wall. *Science (80- )*. **1999**, *285* (5428), 760–763. <https://doi.org/10.1126/science.285.5428.760>.
- (22) Ton-That, H.; Liu, G.; Mazmanian, S. K.; Faull, K. F.; Schneewind, O. Purification and Characterization of Sortase, the Transpeptidase That Cleaves Surface Proteins of Staphylococcus Aureus at the LPXTG Motif. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (22), 12424–12429. <https://doi.org/10.1073/pnas.96.22.12424>.
- (23) Aminov, R. I. A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the Future. *Front. Microbiol.* **2010**, *1* (DEC), 1–7. <https://doi.org/10.3389/fmicb.2010.00134>.
- (24) Hae, J. K.; Coulibaly, F.; Clow, F.; Proft, T.; Baker, E. N. Stabilizing Isopeptide Bonds Revealed in Gram-Positive Bacterial Pilus Structure. *Science (80- )*. **2007**, *318* (5856), 1625–1628. <https://doi.org/10.1126/science.1145806>.
- (25) Bierne, H.; Mazmanian, S. K.; Trost, M.; Pucciarelli, M. G.; Liu, G.; Dehoux, P.; Jänsch, L.; Garcia-del Portillo, F.; Schneewind, O.; Cossart, P. Inactivation of the SrtA Gene in *Listeria Monocytogenes* Inhibits Anchoring of Surface Proteins and Affects Virulence. *Mol. Microbiol.* **2002**, *43* (4), 869–881. <https://doi.org/10.1046/j.1365-2958.2002.02798.x>.
- (26) Garandeau, C.; Réglie-Poupet, H.; Dubail, I.; Beretti, J. L.; Berche, P.; Charbit, A. The Sortase SrtA of *Listeria Monocytogenes* Is Involved in Processing of Internalin and in Virulence. *Infect. Immun.* **2002**, *70* (3), 1382–1390. <https://doi.org/10.1128/IAI.70.3.1382-1390.2002>.
- (27) Paterson, G. K.; Mitchell, T. J. The Role of *Streptococcus Pneumoniae* Sortase A in Colonisation and Pathogenesis. *Microbes Infect.* **2006**, *8* (1), 145–153. <https://doi.org/10.1016/j.micinf.2005.06.009>.
- (28) Kharat, A. S.; Tomasz, A. Inactivation of the SrtA Gene Affects Localization of Surface Proteins and Decreases Adhesion of *Streptococcus Pneumoniae* to Human Pharyngeal Cells in Vitro. *Infect. Immun.* **2003**, *71* (5), 2758–2765. <https://doi.org/10.1128/IAI.71.5.2758-2765.2003>.
- (29) Bhat, A. H.; Nguyen, M. T.; Das, A.; Ton-That, H. Anchoring Surface Proteins to the Bacterial Cell

- Wall by Sortase Enzymes: How It Started and What We Know Now. *Current Opinion in Microbiology*. Elsevier Ltd April 1, 2021, pp 73–79. <https://doi.org/10.1016/j.mib.2021.01.013>.
- (30) Mandlik, A.; Swierczynski, A.; Das, A.; Ton-That, H. Corynebacterium Diphtheriae Employs Specific Minor Pilins to Target Human Pharyngeal Epithelial Cells. *Mol. Microbiol.* **2007**, *64* (1), 111–124. <https://doi.org/10.1111/j.1365-2958.2007.05630.x>.
- (31) Maione, D.; Margarit, I.; Rinaudo, C. D.; Massignani, V.; Mora, M.; Scarselli, M.; Tettelin, H.; Brettoni, C.; Iacobini, E. T.; Rosini, R.; D'Agostino, N.; Miorin, L.; Buccato, S.; Mariani, M.; Galli, G.; Nogarotto, R.; Dei, V. N.; Vegni, F.; Fraser, C.; Mancuso, G.; Teti, G.; Madoff, L. C.; Paoletti, L. C.; Rappuoli, R.; Kasper, D. L.; Telford, J. L.; Grandi, G. Immunology: Identification of a Universal Group B Streptococcus Vaccine by Multiple Genome Screen. *Science (80-. )*. **2005**, *309* (5731), 148–150. <https://doi.org/10.1126/science.1109869>.
- (32) Maresso, A. W.; Schneewind, O. Sortase as a Target of Anti-Infective Therapy. *Pharmacol. Rev.* **2008**, *60* (1), 128–141. <https://doi.org/10.1124/pr.107.07110>.
- (33) Susmitha, A.; Bajaj, H.; Madhavan Nampoothiri, K. The Divergent Roles of Sortase in the Biology of Gram-Positive Bacteria. *Cell Surf.* **2021**, *7*. <https://doi.org/10.1016/j.tcs.2021.100055>.
- (34) Ramirez, N. A.; Das, A.; Ton-That, H. New Paradigms of Pilus Assembly Mechanisms in Gram-Positive Actinobacteria. *Trends Microbiol.* **2020**, *28* (12), 999–1009. <https://doi.org/10.1016/j.tim.2020.05.008>.
- (35) Spirig, T.; Weiner, E. M.; Clubb, R. T. Sortase Enzymes in Gram-Positive Bacteria. *Mol. Microbiol.* **2011**, *82* (5), 1044–1059. <https://doi.org/10.1111/j.1365-2958.2011.07887.x>.
- (36) Ton-That, H.; Faull, K. F.; Schneewind, O. Anchor Structure of Staphylococcal Surface Proteins: A Branched Peptide That Links the Carboxyl Terminus of Proteins to the Cell Wall. *J. Biol. Chem.* **1997**, *272* (35), 22285–22292. <https://doi.org/10.1074/jbc.272.35.22285>.
- (37) Schneewind, O.; Missiakas, D. Sec-Secretion and Sortase-Mediated Anchoring of Proteins in Gram-Positive Bacteria. *Biochim. Biophys. Acta - Mol. Cell Res.* **2014**, *1843* (8), 1687–1697. <https://doi.org/10.1016/j.bbamcr.2013.11.009>.
- (38) Sanchez, B. C.; Chang, C.; Wu, C.; Tran, B.; Ton-That, H. Electron Transport Chain Is

Biochemically Linked to Pilus Assembly Required for Polymicrobial Interactions and Biofilm Formation in the Gram-Positive Actinobacterium *Actinomyces Oris*. *MBio* **2017**, *8* (3).  
<https://doi.org/10.1128/mBio.00399-17>.

- (39) Kruger, R. G.; Otvos, B.; Frankel, B. A.; Bentley, M.; Dostal, P.; McCafferty, D. G. Analysis of the Substrate Specificity of the *Staphylococcus Aureus* Sortase Transpeptidase SrtA. *Biochemistry* **2004**, *43* (6), 1541–1551. <https://doi.org/10.1021/bi035920j>.
- (40) Jacobitz, A. W.; Wereszczynski, J.; Yi, S. W.; Amer, B. R.; Huang, G. L.; Nguyen, A. V.; Sawaya, M. R.; Jung, M. E.; McCammon, J. A.; Clubb, R. T. Structural and Computational Studies of the *Staphylococcus Aureus* Sortase B-Substrate Complex Reveal a Substrate-Stabilized Oxyanion Hole. *J. Biol. Chem.* **2014**, *289* (13), 8891–8902. <https://doi.org/10.1074/jbc.M113.509273>.
- (41) Mazmanian, S. K.; Ton-That, H.; Schneewind, O. Sortase-Catalysed Anchoring of Surface Proteins to the Cell Wall of *Staphylococcus Aureus*. *Mol. Microbiol.* **2001**, *40* (5), 1049–1057. <https://doi.org/10.1046/j.1365-2958.2001.02411.x>.
- (42) Perry, A. M.; Ton-That, H.; Mazmanian, S. K.; Schneewind, O. Anchoring of Surface Proteins to the Cell Wall of *Staphylococcus Aureus*. III. Lipid II Is an in Vivo Peptidoglycan Substrate for Sortase-Catalyzed Surface Protein Anchoring. *J. Biol. Chem.* **2002**, *277* (18), 16241–16248. <https://doi.org/10.1074/jbc.M109194200>.
- (43) Ilangovan, U.; Ton-That, H.; Iwahara, J.; Schneewind, O.; Clubb, R. T. Structure of Sortase, the Transpeptidase That Anchors Proteins to the Cell Wall of *Staphylococcus Aureus*. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (11), 6056–6061. <https://doi.org/10.1073/pnas.101064198>.
- (44) Zong, Y.; Bice, T. W.; Ton-That, H.; Schneewind, O.; Narayana, S. V. L. Crystal Structures of *Staphylococcus Aureus* Sortase A and Its Substrate Complex. *J. Biol. Chem.* **2004**, *279* (30), 31383–31389. <https://doi.org/10.1074/jbc.M401374200>.
- (45) Suree, N.; Liew, C. K.; Villareal, V. A.; Thieu, W.; Fadeev, E. A.; Clemens, J. J.; Jung, M. E.; Clubb, R. T. The Structure of the *Staphylococcus Aureus* Sortase-Substrate Complex Reveals How the Universally Conserved LPXTG Sorting Signal Is Recognized. *J. Biol. Chem.* **2009**, *284* (36), 24465–24477. <https://doi.org/10.1074/jbc.M109.022624>.

- (46) Zong, Y.; Bice, T. W.; Ton-That, H.; Schneewind, O.; Narayana, S. V. L. Crystal Structures of Staphylococcus Aureus Sortase A and Its Substrate Complex. *J. Biol. Chem.* **2004**, *279* (30), 31383–31389. <https://doi.org/10.1074/jbc.M401374200>.
- (47) Naik, M. T.; Suree, N.; Ilangovan, U.; Liew, C. K.; Thieu, W.; Campbell, D. O.; Clemens, J. J.; Jung, M. E.; Clubb, R. T. Staphylococcus Aureus Sortase a Transpeptidase: Calcium Promotes Sorting Signal Binding by Altering the Mobility and Structure of an Active Site Loop. *J. Biol. Chem.* **2006**, *281* (3), 1817–1826. <https://doi.org/10.1074/jbc.M506123200>.
- (48) Frankel, B. A.; Tong, Y.; Bentley, M. L.; Fitzgerald, M. C.; McCafferty, D. G. Mutational Analysis of Active Site Residues in the Staphylococcus Aureus Transpeptidase SrtA. *Biochemistry* **2007**, *46* (24), 7269–7278. <https://doi.org/10.1021/bi700448e>.
- (49) Ton-That, H.; Mazmanian, S. K.; Alksne, L.; Schneewind, O. Anchoring of Surface Proteins to the Cell Wall of Staphylococcus Aureus. Cysteine 184 and Histidine 120 of Sortase Form a Thiolate-Imidazolium Ion Pair for Catalysis. *J. Biol. Chem.* **2002**, *277* (9), 7447–7452. <https://doi.org/10.1074/jbc.M109945200>.
- (50) Bentley, M. L.; Gaweska, H.; Kielec, J. M.; McCafferty, D. G. Engineering the Substrate Specificity of Staphylococcus Aureus Sortase A: The B6/B7 Loop from SrtB Confers Npqtn Recognition to SrtA. *J. Biol. Chem.* **2007**, *282* (9), 6571–6581. <https://doi.org/10.1074/jbc.M610519200>.
- (51) Kline, K. A.; Dodson, K. W.; Caparon, M. G.; Hultgren, S. J. A Tale of Two Pili: Assembly and Function of Pili in Bacteria. *Trends Microbiol.* **2010**, *18* (5), 224–232. <https://doi.org/10.1016/j.tim.2010.03.002>.
- (52) Thanassi, D. G.; Bliska, J. B.; Christie, P. J. Surface Organelles Assembled by Secretion Systems of Gram-Negative Bacteria: Diversity in Structure and Function. *FEMS Microbiol. Rev.* **2012**, *36* (6), 1046–1082. <https://doi.org/10.1111/j.1574-6976.2012.00342.x>.
- (53) Mandlik, A.; Swierczynski, A.; Das, A.; Ton-That, H. Pili in Gram-Positive Bacteria: Assembly, Involvement in Colonization and Biofilm Development. *Trends Microbiol.* **2008**, *16* (1), 33–40. <https://doi.org/10.1016/j.tim.2007.10.010>.
- (54) Bradshaw, W. J.; Davies, A. H.; Chambers, C. J.; Roberts, A. K.; Shone, C. C.; Acharya, K. R.

- Molecular Features of the Sortase Enzyme Family. *FEBS J.* **2015**, *282* (11), 2097–2114.  
<https://doi.org/10.1111/febs.13288>.
- (55) Dufrière, Y. F.; Viljoen, A. Binding Strength of Gram-Positive Bacterial Adhesins. *Front. Microbiol.* **2020**, *11*, 1457. <https://doi.org/10.3389/fmicb.2020.01457>.
- (56) Miller, E.; Garcia, T.; Hultgren, S.; Oberhauser, A. F. The Mechanical Properties of E. Coli Type 1 Pili Measured by Atomic Force Microscopy Techniques. *Biophys. J.* **2006**, *91* (10), 3848–3856.  
<https://doi.org/10.1529/biophysj.106.088989>.
- (57) Manetti, A. G. O.; Zingaretti, C.; Falugi, F.; Capo, S.; Bombaci, M.; Bagnoli, F.; Gambellini, G.; Bensi, G.; Mora, M.; Edwards, A. M.; Musser, J. M.; Graviss, E. A.; Telford, J. L.; Grandi, G.; Margarit, I. Streptococcus Pyogenes Pili Promote Pharyngeal Cell Adhesion and Biofilm Formation. *Mol. Microbiol.* **2007**, *64* (4), 968–983. <https://doi.org/10.1111/j.1365-2958.2007.05704.x>.
- (58) Nallapareddy, S. R.; Singh, K. V.; Sillanpää, J.; Garsin, D. A.; Höök, M.; Erlandsen, S. L.; Murray, B. E. Endocarditis and Biofilm-Associated Pili of Enterococcus Faecalis. *J. Clin. Invest.* **2006**, *116* (10), 2799–2807. <https://doi.org/10.1172/JCI29021>.
- (59) Parsek, M. R.; Singh, P. K. Bacterial Biofilms: An Emerging Link to Disease Pathogenesis. *Annu. Rev. Microbiol.* **2003**, *57*, 677–701. <https://doi.org/10.1146/annurev.micro.57.030502.090720>.
- (60) Bowler, P.; Murphy, C.; Wolcott, R. Biofilm Exacerbates Antibiotic Resistance: Is This a Current Oversight in Antimicrobial Stewardship? *Antimicrob. Resist. Infect. Control* **2020**, *9* (1), 1–5.  
<https://doi.org/10.1186/s13756-020-00830-6>.
- (61) Nuccitelli, A.; Cozzi, R.; Gourlay, L. J.; Donnarumma, D.; Necchi, F.; Norais, N.; Telford, J. L.; Rappuoli, R.; Bolognesi, M.; Maione, D.; Grandi, G.; Rinaudo, C. D. Structure-Based Approach to Rationally Design a Chimeric Protein for an Effective Vaccine against Group B Streptococcus Infections. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (25), 10278–10283.  
<https://doi.org/10.1073/pnas.1106590108>.
- (62) Ton-That, H.; Schneewind, O. Assembly of Pili on the Surface of Corynebacterium Diphtheriae. *Mol. Microbiol.* **2003**, *50* (4), 1429–1438. <https://doi.org/10.1046/j.1365-2958.2003.03782.x>.

- (63) Gaspar, A. H.; Ton-That, H. Assembly of Distinct Pilus Structures on the Surface of *Corynebacterium Diphtheriae*. *J. Bacteriol.* **2006**, *188* (4), 1526–1533.  
<https://doi.org/10.1128/JB.188.4.1526-1533.2006>.
- (64) Swierczynski, A.; Ton-That, H. Type III Pilus of *Corynebacteria*: Pilus Length Is Determined by the Level of Its Major Pilin Subunit. *J. Bacteriol.* **2006**, *188* (17), 6318–6325.  
<https://doi.org/10.1128/JB.00606-06>.
- (65) Ton-That, H.; Marraffini, L. A.; Schneewind, O. Sortases and Pilin Elements Involved in Pilus Assembly of *Corynebacterium Diphtheriae*. *Mol. Microbiol.* **2004**, *53* (1), 251–261.  
<https://doi.org/10.1111/j.1365-2958.2004.04117.x>.
- (66) Mandlik, A.; Das, A.; Ton-That, H. The Molecular Switch That Activates the Cell Wall Anchoring Step of Pilus Assembly in Gram-Positive Bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (37), 14147–14152. <https://doi.org/10.1073/pnas.0806350105>.
- (67) Chang, C.; Amer, B. R.; Osipiuk, J.; McConnell, S. A.; Huang, I. H.; Hsieh, V.; Fu, J.; Nguyen, H. H.; Muroski, J.; Flores, E.; Loo, R. R. O.; Loo, J. A.; Putkey, J. A.; Joachimiak, A.; Das, A.; Clubb, R. T.; Ton-That, H. In Vitro Reconstitution of Sortase-Catalyzed Pilus Polymerization Reveals Structural Elements Involved in Pilin Cross-Linking. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (24), E5477–E5486. <https://doi.org/10.1073/pnas.1800954115>.
- (68) Swaminathan, A.; Mandlik, A.; Swierczynski, A.; Gaspar, A.; Das, A.; Ton-That, H. Housekeeping Sortase Facilitates the Cell Wall Anchoring of Pilus Polymers in *Corynebacterium Diphtheriae*. *Mol. Microbiol.* **2007**, *66* (4), 961–974. <https://doi.org/10.1111/j.1365-2958.2007.05968.x>.
- (69) Kumar, S.; Mollo, A.; Kahne, D.; Ruiz, N. The Bacterial Cell Wall: From Lipid II Flipping to Polymerization. *Chem. Rev.* **2021**. <https://doi.org/10.1021/acs.chemrev.1c00773>.
- (70) Gaspar, A. H.; Ton-That, H. Assembly of Distinct Pilus Structures on the Surface of *Corynebacterium Diphtheriae*. *J. Bacteriol.* **2006**, *188* (4), 1526–1533.  
<https://doi.org/10.1128/JB.188.4.1526-1533.2006>.
- (71) Hae, J. K.; Paterson, N. G.; Gaspar, A. H.; Ton-That, H.; Baker, E. N. The *Corynebacterium Diphtheriae* Shaft Pilin SpaA Is Built of Tandem Ig-like Modules with Stabilizing Isopeptide and

- Disulfide Bonds (Proceedings of the National Academy of Sciences of the United States of America (2009) 106, 40, (16967-16971) 10.1073/Pnas. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106 (43), 18427. <https://doi.org/10.1073/pnas.0911293106>.
- (72) Budzik, J. M.; Marraffini, L. A.; Souda, P.; Whitelegge, J. P.; Faull, K. F.; Schneewind, O. Amide Bonds Assemble Pili on the Surface of Bacilli. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105 (29), 10215–10220. <https://doi.org/10.1073/pnas.0803565105>.
- (73) Mandlik, A.; Das, A.; Ton-That, H. The Molecular Switch That Activates the Cell Wall Anchoring Step of Pilus Assembly in Gram-Positive Bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105 (37), 14147–14152. <https://doi.org/10.1073/pnas.0806350105>.
- (74) Chang, C.; Mandlik, A.; Das, A.; Ton-That, H. Cell Surface Display of Minor Pilin Adhesins in the Form of a Simple Heterodimeric Assembly in *Corynebacterium Diphtheriae*. *Mol. Microbiol.* **2011**, 79 (5), 1236–1247. <https://doi.org/10.1111/j.1365-2958.2010.07515.x>.
- (75) Fleming, A. On the Antibacterial Action of Cultures of a *Penicillium*, with Special Reference to Their Use in the Isolation of *B. Influenzae*. 1929. *Bull. World Health Organ.* **2001**, 79 (8), 780–790. <https://doi.org/10.1093/clinids/2.1.129>.
- (76) Brown, E. D.; Wright, G. D. Antibacterial Drug Discovery in the Resistance Era. *Nature* **2016**, 529 (7586), 336–343. <https://doi.org/10.1038/nature17042>.
- (77) Davies, J. Origins and Evolution of Antibiotic Resistance. *Microbiologia* **1996**, 12 (1), 9–16. <https://doi.org/10.1128/mnbr.00016-10>.
- (78) Zrelavs, N.; Kurbatska, V.; Rudevica, Z.; Leonchiks, A.; Fridmanis, D. Sorting out the Superbugs: Potential of Sortase a Inhibitors among Other Antimicrobial Strategies to Tackle the Problem of Antibiotic Resistance. *Antibiotics* **2021**, 10 (2), 1–14. <https://doi.org/10.3390/antibiotics10020164>.
- (79) Suree, N.; Jung, M.; Clubb, R. Recent Advances Towards New Anti-Infective Agents That Inhibit Cell Surface Protein Anchoring in *Staphylococcus Aureus* and Other Gram-Positive Pathogens. *Mini-Reviews Med. Chem.* **2007**, 7 (10), 991–1000. <https://doi.org/10.2174/138955707782110097>.
- (80) Zhang, J.; Liu, H.; Zhu, K.; Gong, S.; Dramsi, S.; Wang, Y. T.; Li, J.; Chen, F.; Zhang, R.; Zhou, L.; Lan, L.; Jiang, H.; Schneewind, O.; Luo, C.; Yang, C. G. Antiinfective Therapy with a Small



- Molecule Inhibitor of Staphylococcus Aureus Sortase. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (37), 13517–13522. <https://doi.org/10.1073/pnas.1408601111>.
- (81) Oh, K. B.; Kim, S. H.; Lee, J.; Cho, W. J.; Lee, T.; Kim, S. Discovery of Diarylacrylonitriles as a Novel Series of Small Molecule Sortase A Inhibitors. *J. Med. Chem.* **2004**, *47* (10), 2418–2421. <https://doi.org/10.1021/jm0498708>.
- (82) Suree, N.; Yi, S. W.; Thieu, W.; Marohn, M.; Damoiseaux, R.; Chan, A.; Jung, M. E.; Clubb, R. T. Discovery and Structure-Activity Relationship Analysis of Staphylococcus Aureus Sortase A Inhibitors. *Bioorganic Med. Chem.* **2009**, *17* (20), 7174–7185. <https://doi.org/10.1016/j.bmc.2009.08.067>.
- (83) Kruger, R. G.; Barkallah, S.; Frankel, B. A.; McCafferty, D. G. Inhibition of the Staphylococcus Aureus Sortase Transpeptidase SrtA by Phosphinic Peptidomimetics. *Bioorganic Med. Chem.* **2004**, *12* (13), 3723–3729. <https://doi.org/10.1016/j.bmc.2004.03.066>.
- (84) Angelini, A.; Cendron, L.; Chen, S.; Touati, J.; Winter, G.; Zanotti, G.; Heinis, C. Bicyclic Peptide Inhibitor Reveals Large Contact Interface with a Protease Target. *ACS Chem. Biol.* **2012**, *7* (5), 817–821. <https://doi.org/10.1021/cb200478t>.
- (85) Rentero Rebollo, I.; McCallin, S.; Bertoldo, D.; Entenza, J. M.; Moreillon, P.; Heinis, C. Development of Potent and Selective S. Aureus Sortase A Inhibitors Based on Peptide Macrocycles. *ACS Med. Chem. Lett.* **2016**, *7* (6), 606–611. <https://doi.org/10.1021/acsmchemlett.6b00045>.
- (86) Chenna, B. C.; Shinkre, B. A.; King, J. R.; Lucius, A. L.; Narayana, S. V. L.; Velu, S. E. Identification of Novel Inhibitors of Bacterial Surface Enzyme Staphylococcus Aureus Sortase A. *Bioorganic Med. Chem. Lett.* **2008**, *18* (1), 380–385. <https://doi.org/10.1016/j.bmcl.2007.10.051>.
- (87) Thappeta, K. R. V.; Zhao, L. N.; Nge, C. E.; Crasta, S.; Leong, C. Y.; Ng, V.; Kanagasundaram, Y.; Fan, H.; Ng, S. B. In-Silico Identified New Natural Sortase a Inhibitors Disrupt s. Aureus Biofilm Formation. *Int. J. Mol. Sci.* **2020**, *21* (22), 1–18. <https://doi.org/10.3390/ijms21228601>.
- (88) Siegel, S. D.; Amer, B. R.; Wu, C.; Sawaya, M. R.; Gosschalk, J. E.; Clubb, R. T.; Ton-That, H. Structure and Mechanism of LcpA, a Phosphotransferase That Mediates Glycosylation of a Gram-

Positive Bacterial Cell Wall-Anchored Protein. *MBio* **2019**, *10* (1).

<https://doi.org/10.1128/mBio.01580-18>.

- (89) Wu, C.; Huang, I. H.; Chang, C.; Reardon-Robinson, M. E.; Das, A.; Ton-That, H. Lethality of Sortase Depletion in *Actinomyces Oris* Caused by Excessive Membrane Accumulation of a Surface Glycoprotein. *Mol. Microbiol.* **2014**, *94* (6), 1227–1241. <https://doi.org/10.1111/mmi.12780>.

## Chapter 2

### **Sortase-assembled Pili in *Corynebacterium diphtheriae* are Built Using a Latch Mechanism**

## 2.1 Overview

Many gram-positive bacteria have long, filamentous pili on their surface that assist in host colonization by mediating microbe-microbe interactions and host tissue adhesion. These pili are constructed by linking individual subunits and covalently crosslinking them together via lysine-isopeptide bonds. Sortase enzymes catalyze these bonds by recognizing and joining a cell wall sorting signal with a specific lysine residue found on pilin subunits. Although many individual structures of isolated pilin subunits have been discovered, there have been very few structural studies of cross-linked stabilized pilin structures. Using a reconstituted pilus assembly system from *Corynebacterium diphtheria*, we studied the N-terminal domain of the SpaA pilus subunit SpaA (<sup>N</sup>SpaA) when it was crosslinked to a peptide containing the LPLTG sorting signal. Our work uncovered the solution structure and dynamics of the crosslinked interface and showed that when SpaA pilins are linked together that they form a ridged interface. This sheds light on the way sortase recognizes the reactive lysine found on <sup>N</sup>SpaA. We propose that ligation between SpaA pilins requires the use of a latch mechanism involving the highly mobile AB loop. We show that the AB loop becomes rigid upon binding the sorting signal. Based on our structural and biochemical studies, we postulate that the AB loop mediates interactions between *C. diphtheria*'s sortase and the SpaA pilus subunit and functions as a key structural motif for the identification of the reactive lysine residue in SpaA. My contributions to this paper were assisting the HPLC kinetic measurements of <sup>N</sup>SpaA mutants and measuring the resistance to proteolysis from trypsin of <sup>N</sup>SpaA mutants conjugated to C-terminal domain of the SpaA pilin.

This chapter is reformatted with permission from a peer-reviewed journal article:  
Sortase-assembled pili in *Corynebacterium diphtheriae* are built using a latch mechanism.  
McConnell, S.A., McAllister, R.A., Amer, B.R., Mahoney, B.J., Sue, C.K., Chang, C., Ton-That, H., and Clubb, R. T. Proc. Nat. Acad. Sci. USA. **118**,(12). Mar 2021.

## **2.2 Sortase-assembled pili in *Corynebacterium diphtheriae* are Built Using a Latch Mechanism**



# Sortase-assembled pili in *Corynebacterium diphtheriae* are built using a latch mechanism

Scott A. McConnell<sup>a</sup>, Rachel A. McAllister<sup>a</sup>, Brendan R. Amer<sup>a</sup>, Brendan J. Mahoney<sup>a</sup>, Christopher K. Sue<sup>a</sup>, Chungyu Chang<sup>b,c</sup>, Hung Ton-That<sup>b,c</sup>, and Robert T. Clubb<sup>a,c,1</sup>

<sup>a</sup>Department of Chemistry and Biochemistry and the University of California, Los Angeles-US Department of Energy Institute of Genomics and Proteomics, University of California, Los Angeles, CA 90095; <sup>b</sup>Division of Oral Biology and Medicine, University of California, Los Angeles, CA 90095; and <sup>c</sup>Molecular Biology Institute, University of California, Los Angeles, CA 90095

Edited by Scott J. Hultgren, Washington University School of Medicine, St. Louis, MO, and approved February 4, 2021 (received for review September 23, 2020)

**Gram-positive bacteria assemble pili (fimbriae) on their surfaces to adhere to host tissues and to promote polymicrobial interactions. These hair-like structures, although very thin (1 to 5 nm), exhibit impressive tensile strengths because their protein components (pilins) are covalently crosslinked together via lysine-isopeptide bonds by pilus-specific sortase enzymes. While atomic structures of isolated pilins have been determined, how they are joined together by sortases and how these interpilin crosslinks stabilize pilus structure are poorly understood. Using a reconstituted pilus assembly system and hybrid structural biology methods, we elucidated the solution structure and dynamics of the crosslinked interface that is repeated to build the prototypical SpaA pilus from *Corynebacterium diphtheriae*. We show that sortase-catalyzed introduction of a K190-T494 isopeptide bond between adjacent SpaA pilins causes them to form a rigid interface in which the LPLTG sorting signal is inserted into a large binding groove. Cellular and quantitative kinetic measurements of the crosslinking reaction shed light onto the mechanism of pilus biogenesis. We propose that the pilus-specific sortase in *C. diphtheriae* uses a latch mechanism to select K190 on SpaA for crosslinking in which the sorting signal is partially transferred from the enzyme to a binding groove in SpaA in order to facilitate catalysis. This process is facilitated by a conserved loop in SpaA, which after crosslinking forms a stabilizing latch that covers the K190-T494 isopeptide bond. General features of the structure and sortase-catalyzed assembly mechanism of the SpaA pilus are likely conserved in Gram-positive bacteria.**

pili | sortase | Gram positive | lysine isopeptide bond | integrative structural biology

The cellular surface of many bacteria is elaborated with thin appendages called pili (also called fimbriae) which have a range of roles including twitching motility, conjugation, immunomodulation, biofilm formation, and adherence (1, 2). These long proteinaceous fibers are key virulence factors that mediate initial host-pathogen interactions, which are subsequently strengthened by more intimate contacts from shorter pili and cell wall-attached adhesins (1–13). As the infection progresses, pili also facilitate biofilm formation, protecting invading microbes from host immune clearance and exogenous antibiotics (1, 5, 11, 14). Gram-positive bacteria display very thin (1 to 5 nm) (15) hair-like pili that nevertheless possess enormous tensile strength because their protein components are crosslinked together by lysine-isopeptide bonds. These crosslinked fibers are displayed by a wide range of pathogenic and commensal Gram-positive bacteria, but their structures and mechanism of assembly remain poorly understood (2, 5, 7–10).

Pili in Gram-positive bacteria are assembled by pilus-specific sortase enzymes that crosslink the pilus subunits (called pilins) together via lysine-isopeptide bonds. Our current understanding of this process has been significantly advanced by studies of the SpaA pilus in *Corynebacterium diphtheriae*, a pathogen that causes pharyngeal diphtheria (5, 8, 10, 16). The SpaA pilus mediates

adherence to the pharyngeal epithelium and consists of three types of pilins: the pilus shaft is formed by SpaA, and the tip and base are formed by SpaC and SpaB, respectively (17). The *C. diphtheriae* pilus-specific sortase (<sup>Cd</sup>SrtA) assembles the pilus by catalyzing a repetitive, irreversible transpeptidation reaction that covalently links pilin subunits together via an isopeptide bond. The shaft of the pilus is formed by ~100 to 250 crosslinked SpaA pilins (16). <sup>Cd</sup>SrtA-catalyzed SpaA polymerization begins when SpaA prepilin proteins containing an N-terminal signal peptide sequence are exported via the Sec pathway and retained on the extracellular surface via a C-terminal cell wall sorting signal (CWSS) (18). <sup>Cd</sup>SrtA then crosslinks SpaA proteins together via a two-step process. First, an LPLTG sorting signal sequence within the CWSS is cleaved between Thr and Gly residues by the sortase, generating a thioacyl-linked <sup>Cd</sup>SrtA-SpaA intermediate in which the enzyme's active site cysteine residue is covalently linked to the carbonyl atom of the sorting signal threonine. In the second step, a lysine ε-amine group originating from another SpaA pilin attacks the thioacyl linkage in the <sup>Cd</sup>SrtA-SpaA intermediate, thereby joining distinct SpaA proteins together via a K190-T494 isopeptide bond (Fig. 1A). The reactive lysine in SpaA is housed within the N-terminal domain and is part of a highly conserved WxxxVxVYPK sequence motif that is found in many pilin proteins (17). The shaft of the pilus is constructed by repeating this two-step process, and a similar <sup>Cd</sup>SrtA-catalyzed reaction is used to add the SpaC tip pilin to SpaA. Pilus assembly is completed by incorporating the SpaB basal pilin, which promotes pilus attachment to the cell wall using a distinct housekeeping sortase <sup>Cd</sup>SrtF

BIOCHEMISTRY

## Significance

**Bacteria display hair-like structures called pili that enable them to adhere to host tissues, surfaces, and other bacteria. Understanding how microbes assemble and display pili is important, as these structures play a key role in host-pathogen interactions that cause disease. In this study, we have gained insight into how pathogenic Gram-positive bacteria assemble pili by visualizing the intersubunit covalent crosslink that stabilizes these fibrous structures and by investigating how pili are assembled by specialized pilus-specific sortase enzymes. We propose that pili are assembled via a conserved "latch" mechanism that is critical for driving the crosslinking of specific sites on the pilus.**

Author contributions: S.A.M., C.C., H.T.-T., and R.T.C. designed research; S.A.M., R.A.M., B.R.A., B.J.M., C.K.S., and C.C. performed research; S.A.M., R.A.M., B.R.A., B.J.M., C.K.S., C.C., H.T.-T., and R.T.C. analyzed data; and S.A.M. and R.T.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

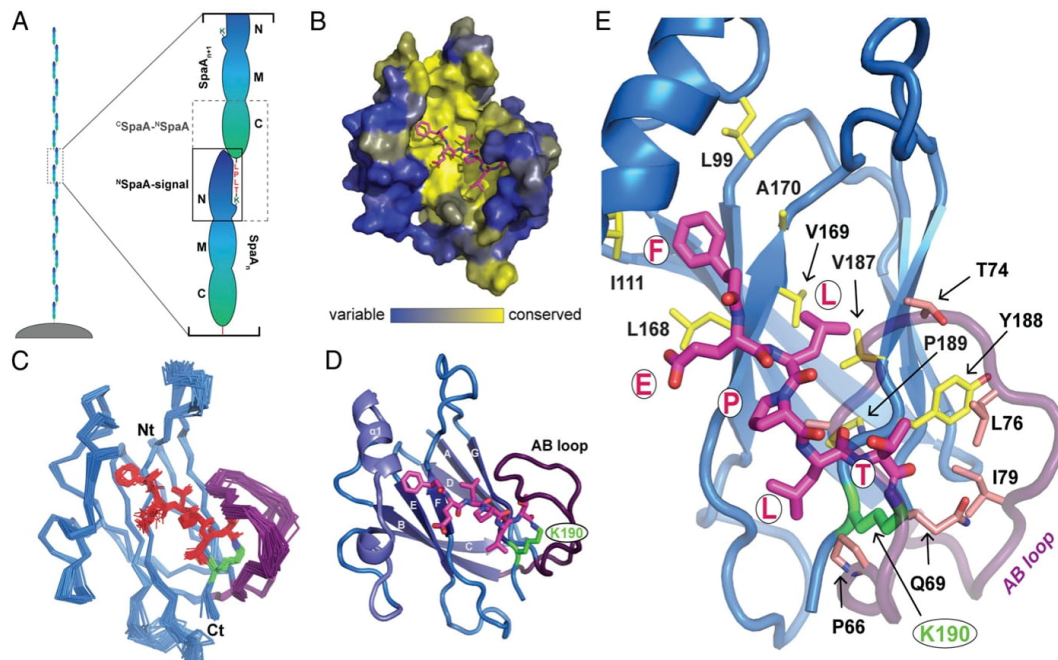
Published under the PNAS license.

<sup>1</sup>To whom correspondence may be addressed. Email: rclubb@mbi.ucla.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2019649118/-DCSupplemental>.

Published March 15, 2021.

Downloaded from <https://www.pnas.org> by UCLA on March 11, 2022 from IP address 128.97.39.237.



**Fig. 1.** Structure of the  $^{15}\text{SpaA}$ -signal peptide complex. (A) Schematic of pilus polymerization with full-length SpaA molecules. An expanded view of the two portions of the crosslinked SpaA polymer investigated in this study,  $^{15}\text{SpaA}$ - $^{15}\text{SpaA}$  complex and  $^{15}\text{SpaA}$ -signal, are boxed in gray dashed lines and solid black lines, respectively. (B) The  $^{15}\text{SpaA}$ -signal peptide complex is represented in surface representation with relative conservation of each residue indicated by a color gradient ranging from highly variable positions (blue) to highly conserved residues (yellow). The peptide (magenta sticks) is docked into a highly conserved, nonpolar binding groove on SpaA. (C) A bundle of the 40 lowest energy structures of the SpaA-signal complex is displayed. The backbone of the  $^{15}\text{SpaA}$  domain is represented by blue ribbons. The last five residues of the sorting signal peptide are depicted as red sticks, and Lys190 is shown as green sticks. (D) Secondary structural elements of the NMR structure are highlighted. (E) An expanded view of the peptide binding interface shows how the peptide is bound in the cleft of  $^{15}\text{SpaA}$ . Residues on SpaA exhibiting intermolecular NOEs to the peptide are shown as sticks. Interacting residues in the core of the domain and within the AB loop are colored yellow and pink, respectively.

(3). Pilus biogenesis is thought to occur within “pilusosomes” on the cell surface, at which pilin substrates and pilus-specific sortases colocalize to facilitate rapid polymerization (18).

Despite their importance in bacterial physiology and pathogenesis, only structures of isolated, noncrosslinked pilins have been determined at atomic-level resolution (19). This is because it has been challenging to obtain homogenous crosslinked pili that are suitable for biophysical analyses and because Gram-positive pili are thin and flexible, making them difficult to study using cryogenic electron microscopy and X-ray crystallography. Crystal structures of isolated pilins have revealed that they contain IgG-like Cna-type domains and frequently one or more spontaneously forming intradomain isopeptide bonds that impart significant resistance to mechanical forces (19–21). Internal isopeptide bond linkages exist as either D- or E-type and are extremely stabilizing, allowing pilin domains to withstand the highest unfolding forces yet reported for a globular protein (20). Atomic-level structures of sortase crosslinked pilins have yet to be visualized, but a transmission electron microscopy study of the *Streptococcus pneumoniae* RrgB pilus enabled the periodicity and polarity of individual subunits within the pilus fiber to be determined (22). This work revealed that the subunits in the pilus are arranged in a head-to-tail manner, enabling sortase-catalyzed isopeptide crosslinking between the lysine and LPxTG motifs located at the N- and C-terminal ends of the pilin, respectively. In crystals, similar head-to-tail packing

arrangements are observed, but whether these lattice interactions are also present in the intact pilus is not known.

In this study, we used a recently developed *in vitro* pilus assembly system and hybrid structural biology methods to gain insight into the structure and biogenesis mechanism of the SpaA pilus from *C. diphtheriae*. We first determined the NMR structure of the N-terminal domain of SpaA crosslinked to the sorting signal peptide and then used small angle X-ray scattering (SAXS), NMR, and crystallographic data to model the structure of the isopeptide-linked SpaA-SpaA building block that is repeated to construct the pilus shaft. We show that crosslinking is accompanied by a large disordered-to-ordered structural change in the SpaA pilin, which forms an interpilin interface that differs markedly from packing interactions observed in crystals of the isolated SpaA. Quantitative measurements of kinetics of the sortase-catalyzed transpeptidation reaction suggest that the enzyme uses a latch mechanism to select the appropriate lysine residue on SpaA for interpilin crosslinking.

## Results

**NMR Structure of the Crosslinked  $^{15}\text{SpaA}$ -Signal Complex.** To learn how Gram-positive pili are stabilized by interpilin lysine-isopeptide bonds, we examined how these crosslinks ligate SpaA pilins together to construct the shaft of the *C. diphtheriae* SpaA pilus (7, 16). The SpaA shaft pilin contains three autonomously folded domains: N-terminal ( $^{15}\text{SpaA}$ , residues 53 to 195), middle

(<sup>M</sup>SpaA, residues 196 to 349), and C-terminal (<sup>C</sup>SpaA, residues 350 to 500) domains (23). SpaA pilins are joined together via interpilin crosslinks that connect the <sup>N</sup>SpaA and <sup>C</sup>SpaA domains; a lysine–isopeptide bond links the side chain ε-amine group of K190 within <sup>N</sup>SpaA to the carbonyl group of the T494 residue present in a LPLTG sorting signal sequence that immediately follows <sup>C</sup>SpaA in the primary sequence (Fig. 1A). Previously, we demonstrated that a mutationally activated <sup>Cd</sup>SrtA enzyme covalently crosslinks peptides containing the LPLTG sorting signal sequence to <sup>N</sup>SpaA, a process that mimics the reaction that is repeated to build the shaft of the pilus (24–26). We first employed this enzyme to produce the <sup>N</sup>SpaA–signal complex, in which the K190 side chain in <sup>N</sup>SpaA is joined via an isopeptide bond to the threonine residue in a sorting signal peptide (KNAGFELPLT peptide that corresponds to residues K485 to T494 in <sup>C</sup>SpaA) (SI Appendix, Fig. S1). Heteronuclear multidimensional NMR spectroscopy was then used to determine the atomic structure of the complex using a total 2,070 experimental restraints, including 66 intermolecular nuclear Overhauser effect (NOE) distance restraints (SI Appendix, Table S1 and Fig. S3). The structure of the complex is well defined by the NMR data, as the backbone and heavy atom coordinates of residues T3 to Q192 and G488 to T494 in the ensemble can be superimposed to the average structure with a RMSD of 0.48 ± 0.05 and 0.81 ± 0.04 Å, respectively (Fig. 1C).

<sup>N</sup>SpaA adopts a CnaB-type fold that binds the crosslinked sorting signal via a large groove formed by residues within strands βF and βG (Fig. 1D). The bound signal contains a characteristic kink at its single proline residue, causing it to form an L-shaped structure that spans from the K190 attachment site to a wedge-shaped opening between the FG loop and helix α1 (27, 28). The conserved LPLT residues in the sorting signal form nonpolar interactions with a conserved surface on <sup>N</sup>SpaA and bury ~630 Å<sup>2</sup> of solvent-exposed surface area (Fig. 1B). A detailed summary of these interactions is provided in SI Appendix, Fig. S5. The C-terminal T494 residue in the signal is joined via an isopeptide linkage to the side chain of K190 located at the end of strand G and partially masked from the solvent by residues that connect strands A and B (called the AB loop) (Fig. 1D). The AB loop extends over the P492–L293–T494 portion of the sorting signal, contacting the T494 methyl group via interactions with the side chains of Q69, L76, and I79 (Fig. 1E). Interestingly, G73 in the AB loop is highly conserved among proteins that contain the pilin motif (SI Appendix, Fig. S4) and is located at the tip of the loop where the chain reverses direction and is in close contact with the bound sorting signal. The positioning of the N-terminal end of the structured portion of the signal is defined by intermolecular NOEs to the aromatic side chain of F489, which is nestled into a hydrophobic region positioned directly underneath the first turn of α1. Signal residues N-terminal to F489 are disordered and exit the binding groove via an opening between the FG loop and helix α1, which must therefore form the interpilin interface in the shaft of the SpaA pilus.

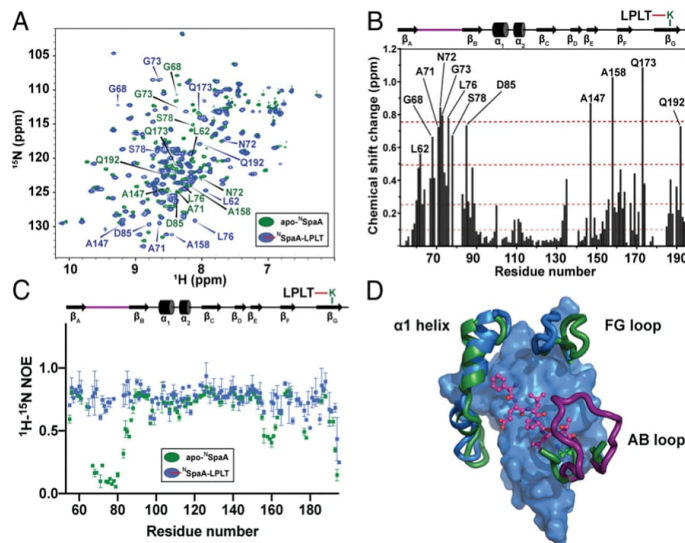
**Crosslinking Triggers the Closure of a Stabilizing Latch over the Interpilin Linkage.** A comparison of the structure of the <sup>N</sup>SpaA–signal complex with a previously determined 1.6-Å crystal structure of unmodified SpaA protein reveals striking conformational differences (Fig. 2D). While the apo- and complexed-forms of the <sup>N</sup>SpaA domain adopt generally similar tertiary structures (their backbone coordinates can be superimposed with a RMSD of 2.3 Å), crosslinking causes a significant rearrangement in the AB loop, as well as more subtle changes in the positioning of the FG loop and α1 helix. In the complex, the AB loop rests against the body of the protein, encapsulating the K190–T494 isopeptide linkage, while in the structure of the unmodified SpaA protein, coordinates for residues Q69 to I79 in the AB loop are missing because they exhibit scant electron density (23). To determine if

the AB loop undergoes a disordered-to-ordered transition upon signal attachment by sortase, we assigned the backbone chemical shifts of apo-<sup>N</sup>SpaA and acquired <sup>1</sup>H–<sup>15</sup>N steady-state NOE relaxation data for apo-<sup>N</sup>SpaA and the <sup>N</sup>SpaA–signal complex. Consistent with the NMR structure of the complex, the largest differences in the backbone chemical shifts occur for residues that form the signal binding groove and the AB loop (Fig. 2A and B). Interestingly, the <sup>1</sup>H–<sup>15</sup>N steady-state NOE data reveal that signal attachment significantly retards motions in the AB loop, as residues M63 to G86 in apo-<sup>N</sup>SpaA exhibit small-magnitude steady-state NOEs indicative of high mobility, whereas in the <sup>N</sup>SpaA–signal complex, they are rigid with values around 0.8 (Fig. 2C and SI Appendix, Fig. S2). Covalent signal attachment also quenches motions on the opposite side of the binding pocket, as similar, albeit smaller trends are observed for residues in the FG loop that contact the N-terminal end of the sorting signal near the interpilin interface. Limited proteolysis experiments of apo-<sup>N</sup>SpaA and the <sup>N</sup>SpaA–signal complex indicate that the protein in the complex is ~42% more resistant to proteolytic degradation after 24 h (SI Appendix, Fig. S6A). The <sup>N</sup>SpaA–signal complex is also slightly more thermostable based on differential scanning calorimetry experiments (its Δ<sub>u</sub>G° increases by ~0.9 kJ) (SI Appendix, Fig. S6B and C) (29, 30). Taken together, these data indicate that the AB loop becomes ordered upon pilin crosslinking, forming a latch structure that shields the isopeptide linkage and stabilizes <sup>N</sup>SpaA.

**Solution Structure of the Interpilin Interface.** To gain insight into the structure and dynamics of the interpilin interface that is repeated to build the SpaA pilus, we used the activated <sup>Cd</sup>SrtA enzyme to generate a crosslinked <sup>C</sup>SpaA–<sup>N</sup>SpaA dimer (M.W. 31 kDa); in the dimer, <sup>N</sup>SpaA is crosslinked via its K190 residue to the sorting signal that resides in a 11–amino acid C-terminal tail that immediately follows the <sup>C</sup>SpaA domain (Fig. 1A). NMR spectra were acquired using samples of the dimer in which either the <sup>C</sup>SpaA or <sup>N</sup>SpaA domains were selectively labeled with nitrogen-15. A comparison with the spectra of the corresponding isolated domains reveals that crosslinking causes substantial chemical shift changes, suggesting that the domains pack against one another in the dimer (SI Appendix, Fig. S7A and B). This is substantiated by molecular correlation time (τ<sub>c</sub>) measurements using NMR <sup>15</sup>N relaxation data, as the τ<sub>c</sub> of the <sup>C</sup>SpaA–<sup>N</sup>SpaA dimer is 18.0 ns, much longer than expected if the domains were simply connected by a flexible linker that enabled them to freely reorient (the τ<sub>c</sub> values of the isolated <sup>N</sup>SpaA and <sup>C</sup>SpaA domains are 8.7 and 9.5 ns, respectively) (SI Appendix, Fig. S7C). SAXS data of the crosslinked dimer also indicate that it is generally inflexible as evidenced by the distance distribution (SI Appendix, Fig. S8A) and normalized Kratky plots of the data (SI Appendix, Fig. S8B). The calculated radius of gyration is smaller than expected for a freely reorienting dimer, and the Kratky plot is characterized by features which correspond to a nonflexible structure. Thus, both NMR and SAXS analyses are in agreement and indicate that the domains within the dimer are immobilized with respect to one another.

The solution structure of the crosslinked <sup>C</sup>SpaA–<sup>N</sup>SpaA dimer was modeled using an integrated approach that employed SAXS, NMR, and crystallographic data. An initial model of the <sup>C</sup>SpaA–<sup>N</sup>SpaA complex was constructed using the crystal and NMR structures of <sup>C</sup>SpaA (Protein Data Bank [PDB]: 3HR6) and <sup>N</sup>SpaA–signal (this work), respectively. SAXS data were then employed to drive multistate rigid body modeling of the complex using the MultiFOXS approach (31). To account for potential interdomain flexibility, 10,000 models were generated, and residues connecting the <sup>C</sup>SpaA and <sup>N</sup>SpaA domains (K483–A488) were allowed to move freely during the calculations (Methods). Models of the <sup>C</sup>SpaA–<sup>N</sup>SpaA dimer in which the domains adopt related orientations with respect to one another best fit the





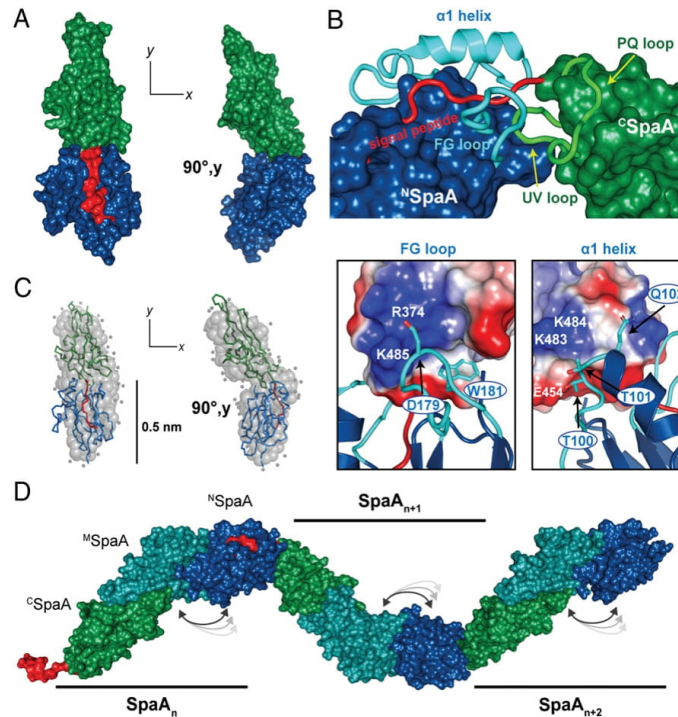
**Fig. 2.** The AB loop undergoes a disordered-to-ordered transition upon crosslinking. (A)  $^1\text{H}$ - $^{15}\text{N}$  HSQC correlation spectra of apo- $^{15}\text{N}$ SpaA (green) and  $^{15}\text{N}$ SpaA-signal complex (blue) are overlaid. The positions of residues with differences larger than 0.5 ppm in composite chemical shift are indicated on the plot. (B) Chemical shift perturbations (CSPs) for each residue are plotted with respect to primary sequence. CSPs are binned into >0.75 ppm, 0.75 to 0.5 ppm, 0.5 to 0.25 ppm, and 0.1 to 0.25 ppm (indicated by red dotted lines of increasing transparency). (C) Heteronuclear NOE data are graphed as a function of primary sequence for apo-SpaA (green) and SpaA-signal (blue). (D) A surface rendering of the complex, with the AB loop (purple), FG loop, and  $\alpha 1$  helix (blue) highlighted with cartoon representations and the signal peptide shown as magenta sticks. The apo- $^{15}\text{N}$ SpaA structure is aligned to the NMR complex, and the corresponding secondary elements are shown in green to highlight conformational shifts. The backbone coordinates can be superimposed with a RMSD of 2.3 Å.

SAXS data (Fig. 3A and *SI Appendix*, Fig. S8 C and D). This domain orientation is further supported by lower resolution ab initio modeling of the scattering data using the program GASBOR, which yields an elongated, kinked dimeric structure that is similar to those obtained by MultiFOXS (Fig. 3C). In the MultiFOXS model of the crosslinked  $^{\text{C}}\text{SpaA}$ - $^{\text{N}}\text{SpaA}$  dimer that best fits the SAXS data, the FG loop in  $^{\text{N}}\text{SpaA}$  is inserted between the UV and PQ loops in  $^{\text{C}}\text{SpaA}$ , thereby positioning the sorting signal following  $^{\text{C}}\text{SpaA}$  within  $^{\text{N}}\text{SpaA}$ 's peptide binding groove (Fig. 3 B, Top). The protein-protein interface buries 1,270 Å<sup>2</sup> of solvent-exposed surface area (32) and is further stabilized by interactions between the N-terminal portion of helix  $\alpha 1$  in  $^{\text{N}}\text{SpaA}$  and UV loop in  $^{\text{C}}\text{SpaA}$ . This packing arrangement explains why the W181 residue within the WxxxVxVYPK pilin motif is conserved, as its indole side chain presumably plays a key structural role in stabilizing the SpaA-SpaA interface by packing into a hydrophobic surface formed by residues A450 and Y453 in  $^{\text{C}}\text{SpaA}$ 's UV loop (Fig. 3 B, Bottom Left). Stabilizing electrostatic interactions are predicted to surround this nonpolar interface. On one side, the negatively charged D179 side chain in  $^{\text{N}}\text{SpaA}$ 's FG loop is packed against a cationic surface on  $^{\text{C}}\text{SpaA}$  formed by residues K485 (signal peptide) and R374 (PQ loop) (Fig. 3 B, Bottom Left), while on the other side, hydrophilic interactions occur between residues at the N-terminal end of  $\alpha 1$  helix in  $^{\text{N}}\text{SpaA}$  (T99, T100, and Q101) and residues located in  $^{\text{C}}\text{SpaA}$ 's UV loop (K483, K484, and E454) (Fig. 3 B, Bottom Right). Notably, the SAXS-derived solution structure of the  $^{\text{C}}\text{SpaA}$ - $^{\text{N}}\text{SpaA}$  dimer presented here differs markedly from the head-to-tail packing arrangement observed in crystals of the isolated SpaA protein (23), and only the SAXS model is compatible with NMR chemical shift mapping data (described in *SI Appendix*, Fig. S9).

The biological relevance of the SAXS-derived model of the interpilin interface is substantiated by in vitro and in vivo data.

Bacteria overexpressing SpaA proteins containing L99R, P172R, D179R, or L182R single-amino acid substitutions at the predicted interface are impaired in pili display when assessed by immunoelectron microscopy or immunoblotting of fractionated cells with a specific antibody against SpaA (*SI Appendix*, Fig. S10). All interfacial substitutions reduce the abundance of surface-displayed pili when assessed by microscopy, with pili harboring P172R and D179R substitutions showing largest defects. The SpaA variants altering the presumed interface were further assessed for their ability to serve as a substrate in the sortase-catalyzed cross-linking reaction using an in vitro gel-based assay that monitors the production of the  $^{\text{C}}\text{SpaA}$ - $^{\text{N}}\text{SpaA}$  dimer (*SI Appendix*, Fig. S11) (24). Consistent with formation of the interpilin interface being required for efficient transpeptidation, a densitometry analysis of the domain coupling reactions reveals that SpaA proteins harboring interfacial L99R, P172R, and L182R point substitutions produce less crosslinked product (*SI Appendix*, Fig. S11B). Interestingly, the integrity of the interpilin interface in the SAXS-derived model appears to be an important determinant for overall stability. This is because trypsin digestion studies reveal that the  $^{\text{C}}\text{SpaA}$ - $^{\text{N}}\text{SpaA}$  dimer is significantly more resistant to proteolytic degradation than the  $^{\text{N}}\text{SpaA}$ -signal complex, and introducing a single L99D substitution at the interface is destabilizing as it increases  $^{\text{C}}\text{SpaA}$ - $^{\text{N}}\text{SpaA}$  dimer susceptibility to proteolysis (*SI Appendix*, Fig. S11C). In all, the results of both biochemical and cellular experiments support the SAXS-derived model of the interpilin interface and suggest that its formation is required for efficient transpeptidation and resistance to proteolytic degradation.

**The Sorting Signal Must Be Partially Transferred to SpaA to Initiate Crosslinking.** Guided by the solution structures of the  $^{\text{N}}\text{SpaA}$ -signal complex and the crosslinked  $^{\text{C}}\text{SpaA}$ - $^{\text{N}}\text{SpaA}$  dimer, we employed in vitro crosslinking and cellular assays to probe the mechanism of pilus biogenesis. Initially,  $^{\text{N}}\text{SpaA}$  proteins containing



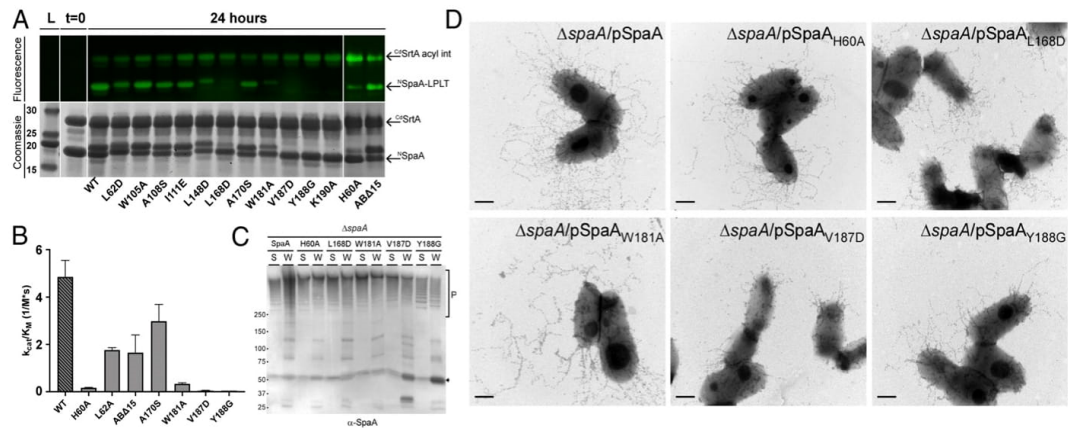
**Fig. 3.** SAXS model of the crosslinked interpilin interface. (A) The best fit single-conformation rigid body model from the MultiFOXS calculation ( $^c$ SpaA, green;  $^N$ SpaA, blue; signal peptide, red). (B) Details of the interface formed by the solution structure of the  $^c$ SpaA- $^N$ SpaA complex. (Top) The cores of the two domains and the interacting loops are shown as surface and cartoon representations, respectively. (Bottom) Expanded views of two interactions to  $^c$ SpaA mediated by the FG loop and  $\alpha 1$  helix of  $^N$ SpaA. (C) GASBOR ab initio model from solution scattering, shown as gray spheres. The rigid body model calculated by MultiFOXS is aligned with the GASBOR model and depicted in ribbon representation. Left and right shapes are the same GASBOR models rotated counterclockwise by  $90^\circ$  around y-axis. (D) Model of a dimer of full-length SpaA molecules comprising the pilus shaft. Coordinates from the crystal structure of full-length SpaA molecules (PDB: 3HR6) were arranged in head-to-tail arrangements according to the interface determined by the MultiFOXS model of the  $^c$ SpaA- $^N$ SpaA structure (color gradient blue to green from N term to C term).

amino acid substitutions at conserved sites within the SpaA-SpaA interface were tested for their ability to serve as substrates for  $^c$ SrtA using an in vitro gel-based assay that monitors the covalent attachment of a sorting signal peptide fluorophore to  $^N$ SpaA (Fig. 4A) (24). Some of the largest defects in transpeptidation activity occur when  $^N$ SpaA residues that contact the LPLT sorting signal in the  $^N$ SpaA-signal complex are altered (L168D and V187D) or when the K190 nucleophile and residues immediately adjacent to it are changed (H60A, Y188G, and K190A).

Variants exhibiting significant defects in reactivity were further evaluated using a newly developed high performance liquid chromatography (HPLC)-based assay that quantitatively measures the steady-state kinetics of crosslinking (26) (Fig. 4B and *SI Appendix*, Fig. S12). In the HPLC-based assay, the sorting signal peptide is held in excess, such that effects of  $^N$ SpaA substitutions on the second step of transpeptidation are revealed (i.e., the rate at which the K190 lysine  $\epsilon$ -amine group in  $^N$ SpaA attacks the thioacyl-linked  $^c$ SrtA-SpaA intermediate). The V187D and Y188G variants exhibit the largest defects in transpeptidation and alter side chains that are positioned immediately proximal to the K190 nucleophile ( $k_{cat}/K_M$  values are less than 1% of wild-type protein) (Fig. 4B). Both substitutions reduce catalytic turnover, suggesting that they are needed to properly form the active site used to catalyze the isopeptide bond. The H60A variant also exhibits reduced catalytic turnover, presumably because its imidazole ring stabilizes the

positioning of the Y188 sidechain through pi-stacking interactions. Interestingly, disrupting contacts between  $^N$ SpaA and residues in the sorting signal that are positioned distal to the site of isopeptide formation also reduces the rate of transpeptidation by increasing the  $K_M$  (A170S and W181A). This suggests that when the thioacyl  $^c$ SrtA-SpaA intermediate formed in the first step of catalysis encounters  $^N$ SpaA, the sorting signal bound to the  $^c$ SrtA's active site must move from the enzyme into the binding groove on  $^N$ SpaA to form a catalytically active complex that performs the final step of transpeptidation.

Bacteria-expressing SpaA proteins containing single-amino acid substitutions in the sorting signal binding groove also show defects in pilus display. As shown in Fig. 4C, immunoblotting analysis of protein samples collected from the culture medium (S) and cell wall (W) fractions of corynebacterial cells expressing wild-type SpaA produced abundant SpaA polymers detected in the cell wall fractions, with some polymers secreted into the extracellular milieu. These protein polymers (P) have high molecular weights as they were not well separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) electrophoresis. Strains expressing SpaA with H60A, L168D, and W181A still produced SpaA polymers, albeit less abundantly as compared to wild-type SpaA. Consistent with the in vitro analysis, mutants V187D and Y188G exhibited a significant pilus assembly defect, with reduced pilus polymerization and accumulation of SpaA



**Fig. 4.** In vitro and in vivo validation of key residues on the SpaA acceptor domain. (A) Gel fluorescence assay to rapidly screen a library of  $^{15}N$ SpaA variants. The Top shows an SDS-PAGE gel visualized by FITC fluorescence, indicating the presence of  $^{15}N$ SrtA (top band) or  $^{15}N$ SpaA (bottom band) conjugated to fluorescent peptide. The Bottom shows the same SDS-PAGE gel visualized by Coomassie staining in order to visualize the total protein composition of each lane. Peptide-labeled  $^{15}N$ SpaA variants typically have slightly lower electrophoretic mobility than the corresponding apo- $^{15}N$ SpaA variant. (B) Bar graph comparing the catalytic efficiency ( $k_{cat}/K_M$ ) of each mutant to the corresponding kinetic parameters of wild-type  $^{15}N$ SpaA. (C) Cells of the *C. diphtheriae*  $\Delta spaA$  mutant expressing wild-type SpaA or individual SpaA mutants from a plasmid were grown to midlog phase and subjected to cell fractionation. Protein samples collected from the culture medium (S) and cell wall (W) fractions were analyzed by immunoblotting with specific antibodies against SpaA ( $\alpha$ -SpaA). Molecular mass markers in kDa and SpaA polymers (P) and monomer (arrow) are indicated. (D) Cells of strains used in (C) were immobilized on carbon-coated nickel grids and stained with  $\alpha$ -SpaA followed by IgG-conjugated 12 nm gold particles and 1% uranyl acetate prior to electron microscopy. (Scale bars of 0.5  $\mu$ m.)

precursors. To corroborate the fractionation results, the same set of strains was analyzed by immunoelectron microscopy, whereby cells were stained by SpaA antibodies, followed by staining with 12 nm gold particles conjugated to IgG. Consistent with the Western blotting analysis and in vitro kinetic measurements, severe effects are observed when V187D and Y188G substitutions are introduced near the site of K190 crosslinking. These variants are unable to assemble long and abundant pili as compared to the wild-type SpaA (Fig. 4D, V187D and Y188G panels) while less severe effects are observed for the H60A mutant strain that exhibited higher activity in vitro (Fig. 4D, compare SpaA and H60A panels). While V187D and Y188G  $^{15}N$ SpaA variants are folded similar to the wild-type protein, these nonconservative alterations can be expected to significantly perturb the positioning of nearby side chains. Finally, A170S and W181A mutants designed to alter residues that contact the sorting signal but are positioned distal to the site of crosslinking also produced fewer pili as compared to the wild-type SpaA. The L168D mutant assembled short pili, and the W181A mutant appeared to be fragile, with broken pili surrounding the cells (Fig. 4D, L168D and W181A panels). Thus, both the in vitro and cellular data suggest that signal transfer from the enzyme to the sorting signal pocket on SpaA is required for efficient catalysis, as mutation of this surface impairs crosslinking. This explains previously reported findings that  $C^{15}SrtA$  can only crosslink sorting signals to K190 when it is housed in a structurally intact  $^{15}N$ SpaA domain (24).

### Discussion

Using a recently developed in vitro assembly reaction (24, 25) and integrative structural biology methods, we determined the structure and dynamics of the lysine-isopeptide bond crosslinked  $C^{15}SpaA$ - $^{15}N$ SpaA interface that is repeated to build the shaft of the *C. diphtheriae* SpaA pilus (Fig. 1A). The structure of  $^{15}N$ SpaA covalently attached to the  $C^{15}SpaA$  sorting signal was first determined by NMR, and then SAXS data of the SpaA-SpaA complex was used in conjunction with the crystal structure of the  $C^{15}SpaA$  domain to model the solution structure of the interpilin interface

that is formed by sortase crosslinking of the  $C^{15}SpaA$  and  $^{15}N$ SpaA domains. This work reveals that sortase crosslinking immobilizes the pilin subunits, triggering the formation of an extensive interpilin interface in which the sorting signal following  $C^{15}SpaA$  is inserted into a nonpolar groove on  $^{15}N$ SpaA (Fig. 3A). Residues within the conserved WxxxVxVYYPK pilin motif line the binding groove and when altered slow transpeptidation in vitro and in cells (Fig. 4). Our results shed light onto the structure and dynamics of the shaft of the SpaA pilus, which can be modeled using our structure of the  $C^{15}SpaA$ - $^{15}N$ SpaA complex and a previously reported structure of the intact SpaA protein (Fig. 3D) (23). The shaft is formed by SpaA pilins that are arrayed in a head-to-tail manner with successive sortase-installed interpilin isopeptide bonds positioned on opposite faces of the polymer and each crosslinked SpaA-SpaA unit forming a “S” shape because of a  $\sim 140^\circ$  kink at the  $C^{15}SpaA$ - $^{15}N$ SpaA junction. The SpaA pilus and other Gram-positive pili are presumably flexible, as they appear as nonlinear, hair-like structures in transmission electron micrographs (Fig. 4D) (16, 22, 33). This flexibility likely originates from intrapilin motions that occur between the N-terminal and middle domains, as crystal structures of isolated pilins have revealed a small interface between the domains that allow them to adopt different positions with respect to one another (33–39). The strongest evidence comes from studies of the GG-SpaD shaft pilin from *Lactobacillus rhamnosus*, as its N-terminal domain adopts a range of bent conformations relative to the body of the protein (40). Some flexibility in the SpaA pilus may also originate from motions at the interpilin  $C^{15}SpaA$ - $^{15}N$ SpaA interface, as it is primarily formed by hydrophilic interactions between the proteins (Fig. 3B). However, these motions are presumably modest, since the normalized Kratky plot of the SAXS data for the  $C^{15}SpaA$ - $^{15}N$ SpaA complex reveals a clear bell-shaped curve at low  $q$  values with a maximum peak height of 1.32 at a peak position of 2.29, which are only slightly larger than expected for a rigid, compact structure (SI Appendix, Fig. S8) (41, 42). When adorned with its SpaC tip pilin, the SpaA pilus adheres *C. diphtheriae* to human pharyngeal cells, preventing disengagement of the microbe by withstanding

significant pulling forces caused by coughing, sneezing, mucociliary flow, etc. (43, 44) The SpaA pilus can withstand these forces due to the optimal positioning of intra- and intermolecular lysine–isopeptide bond crosslinks (45). The model of the pilus fiber presented here (Figs. 1D and 3B) is in agreement with previous single-molecule pulling experiments, which concluded that the load bearing spine of the pilus only goes through the middle (<sup>M</sup>SpaA) and C-terminal (<sup>C</sup>SpaA) domains that contain spontaneously forming intrapilin isopeptide linkages that are capable of withstanding a pulling force of ~525 pN (21, 23, 45).

To assemble pili, <sup>Cd</sup>SrtA and other pilus-specific sortases select for crosslinking a single lysine on the surface of their pilin substrates. As sortases are relatively small enzymes, how this specificity is achieved has remained unclear. Our results suggest that selectivity is achieved using a latching mechanism that relies upon tertiary structural features present in the SpaA protein. Fig. 5 shows a working model of lysine–isopeptide bond forming step catalyzed by <sup>Cd</sup>SrtA that adds a single SpaA protein to the shaft of the pilus. This reaction forms a K190-T494 lysine–isopeptide bond between SpaA proteins, connecting the sorting signal (red) following the <sup>C</sup>SpaA domain (green) to the K190 amine group in <sup>N</sup>SpaA (blue) (Fig. 5A). Presumably, two thioacyl-linked enzyme–substrate intermediates mediate this interaction and are tethered to the membrane via their respective <sup>Cd</sup>SrtA enzymes (7, 18). The growing (SpaA)<sub>n</sub> polymer is housed in a <sup>Cd</sup>SrtA-(SpaA)<sub>n</sub> intermediate in which the carbonyl group in residue T494 of the C-terminal sorting signal on the polymer is attached via a thioacyl bond to <sup>Cd</sup>SrtA's active site cysteine. New SpaA proteins enter the reaction as similarly bonded thioacyl enzyme–substrate intermediates (<sup>Cd</sup>SrtA-SpaA) after their sorting signals are nucleophilically attacked by the enzyme's active site cysteine residue. A single protein is then added to the shaft when the reaction intermediates form a ternary complex that enables the K190 nucleophile on <sup>Cd</sup>SrtA-SpaA to resolve the thioacyl bond in the <sup>Cd</sup>SrtA-(SpaA)<sub>n</sub>. During this process, the growing pilus is transferred from one enzyme to another and is then poised to react with a new <sup>Cd</sup>SrtA-SpaA intermediate to continue the polymerization reaction.

Our results suggest that the isopeptide bond-forming reaction occurs through a latch mechanism in which selectivity for K190 is achieved by first requiring that the sorting signal be transferred from the enzyme to <sup>N</sup>SpaA in order to initiate catalysis (Fig. 5B). The most parsimonious orientation of the enzyme–substrate reactants in the ternary complex aligns the sorting signal binding pockets on the enzyme and <sup>N</sup>SpaA, enabling a simple translation movement to transfer the signal between the proteins. This orientation explains why the AB loop is flexible, as it can readily be displaced outwards to allow K190 access to the enzyme's active site. Moreover, this arrangement positions <sup>Cd</sup>SrtA's β7/β8 loop near K190, providing a rationale for why many of its residues are highly conserved and important for catalysis (25). We surmise that within the ternary complex, the sorting signal must be partially transferred to the binding groove on <sup>N</sup>SpaA in order to activate K190 for catalysis. This is because several <sup>N</sup>SpaA variants that alter contacts to sorting signal residues that are positioned distal to the site of crosslinking slow transpeptidation by increasing the enzyme's K<sub>M</sub> for <sup>N</sup>SpaA (Fig. 4B, W181A/A170S). Partial signal transfer would act as a zipper, juxtaposing K190 and the <sup>Cd</sup>SrtA-(SpaA)<sub>n</sub> thioacyl bond, potentially creating a microenvironment that deprotonates the ε-amine for nucleophilic attack on the thioacyl linkage. This would seem essential, since in the isolated <sup>N</sup>SpaA protein the AB loop is dynamic (Fig. 2C) and thus transiently exposes the side chain of K190 to solvent such that it presumably adopts a protonated, nonnucleophilic state at physiological pHs (its calculated pK<sub>a</sub> is ~10.1) (46–48). As the isopeptide bond forms, our NMR data reveal that the AB latch closes, undergoing a disordered-to-ordered transition that shields the bond and stabilizes the protein. Latch closure may also help

drive the dissolution of the ternary complex, freeing the transferred polymer for another round of catalysis. It seems likely that other sortases will build pili through a similar mechanism in which the sorting signal is first partially transferred to a binding groove on the pilin substrate to facilitate crosslinking to a specific lysine residue. While not universal, the latch may also be a conserved feature in many sortase-catalyzed pilus biogenesis reactions, as structures of several shaft pilins solved in their apo-states also contain disordered AB loops (20, 23, 36, 40, 49–53). As pili in Gram-positive bacteria are important virulence factors, the results reported here could be useful in guiding the development of antibiotics that work by inhibiting pilus assembly.

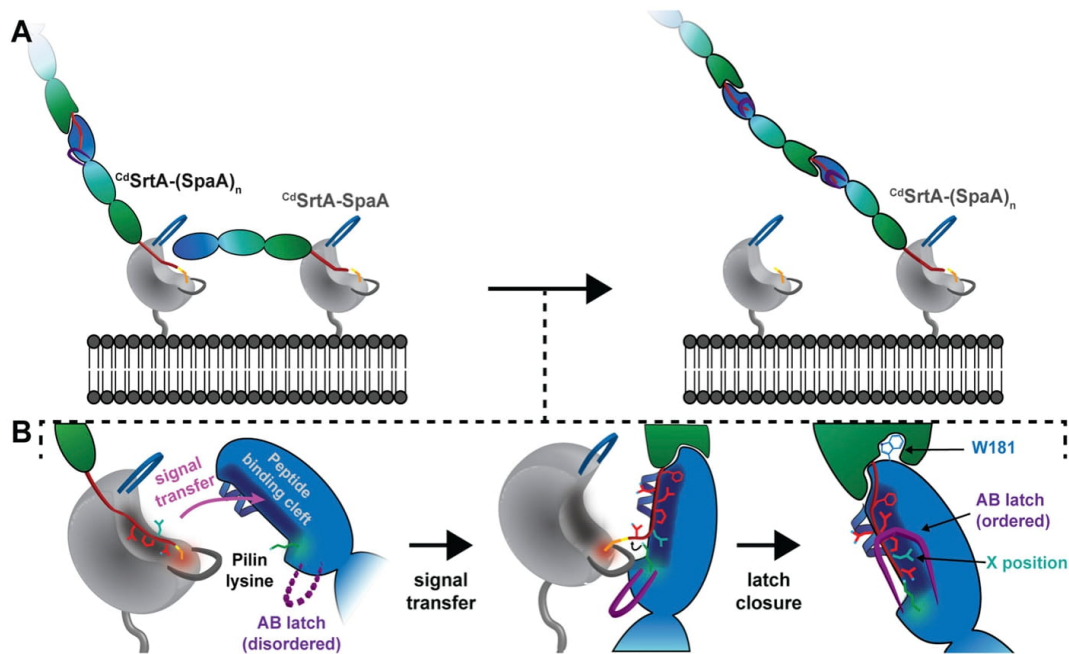
## Methods

### Production of the Crosslinked <sup>N</sup>SpaA-Signal Complex and <sup>C</sup>SpaA-<sup>N</sup>SpaA Dimer.

The amino-terminal domain of the SpaA protein from *C. diphtheriae* (<sup>N</sup>SpaA, residues E53 to S195) was produced and purified as described previously (24, 25). The <sup>N</sup>SpaA-signal complex was generated by enzymatic covalent ligation of synthetic peptide to [<sup>13</sup>C,<sup>15</sup>N] <sup>N</sup>SpaA using a previously described activated variant of <sup>Cd</sup>SrtA (24, 26) (<sup>Cd</sup>SrtA<sup>3M</sup>, residues N37 to Q257, containing D81G/W83G/N85A mutations). Complete modification of <sup>N</sup>SpaA to its cognate sorting signal peptide occurred after incubation of 100 μM <sup>Cd</sup>SrtA<sup>3M</sup>-His<sub>6</sub>, 100 μM <sup>N</sup>SpaA, and 1 mM synthetic peptide derived from the SpaA sorting signal motif (KNAGFELPLTGGSGRI) (Peptide 2.0) in modification buffer (50 mM Tris pH 8.0, 300 mM NaCl, and 5 mM TCEP) for 24 h at room temperature. Sortase and unreacted peptide were removed from the reaction by HisPur Co<sup>2+</sup> purification and subsequent concentration by Amicon spin filters with a 10 kDa molecular weight cut-off (MWCO). Complex formation was confirmed by matrix assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry as well as SDS-PAGE analysis. The sample was exchanged into NMR buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 6.0, 100 mM NaCl, 8% D<sub>2</sub>O, and 0.01% Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>) and diluted to a concentration of 1.2 mM for NMR studies. Subsequently, the sample was lyophilized and redissolved into 100% deuterated NMR buffer for additional NMR studies. The cross-linked <sup>C</sup>SpaA-<sup>N</sup>SpaA dimer was prepared as described for the <sup>N</sup>SpaA-signal complex but employed 300 μM <sup>C</sup>SpaA (SpaA, residues R350 to I500) instead of peptide. To assess the relative reactivity of <sup>N</sup>SpaA variants with interfacial point mutations, the reactions were stopped at indicated timepoints and product formation was assessed by densitometry using ImageJ (54). For SAXS analysis, the wild-type dimer was purified from the reaction components using HisPur Co<sup>2+</sup> immobilized metal affinity chromatography (IMAC) (Thermo Fisher Scientific) and size exclusion chromatography (Superdex Increase 10/300 GL, GE Healthcare). For NMR samples of the dimer, either <sup>N</sup>SpaA or <sup>C</sup>SpaA was uniformly labeled with nitrogen-15, and the other component was expressed in natural abundance nitrogen-14 media.

**Immunoelectron Microscopy and Cell Fractionation Studies.** Cells of the *C. diphtheriae* ΔspaA mutant expressing wild-type SpaA or individual SpaA mutants from a plasmid were grown on Heart Infusion Broth (HIB) agar plates. A full loop of each overnight colony was collected, suspended in, and gently washed once with phosphate-buffered saline (PBS) buffer. A drop (7 μL) of bacterial suspension in PBS was immobilized on the carbon-coated nickel grids for immunogold labeling as previously reported (25, 55). Cells were stained with antibodies against SpaA (α-SpaA; 1:100 dilution), followed by IgG antibodies conjugated to 12 nm colloid gold particles. Cells were then stained with 1% uranyl acetate prior to analysis using a JEOL JEM1200 electron microscope. For fractionation studies, cells were cultured in HIB media supplemented with 30 μg/mL kanamycin and grown to mid log phase (OD<sub>600</sub> = 0.5 to 0.6). Cells were normalized to OD<sub>600</sub> at 1 before harvest. The harvested cells were fractionated into medium (S) and cell wall-associated (W) fractions as described previously (4). The samples were separated by 3 to 14% gradient SDS-PAGE and analyzed by immunoblotting with α-SpaA antibody.

**NMR Structure Determination and Relaxation Measurements.** NMR spectra were collected at 298 K on Bruker Avance III HD 600 MHz and Avance NEO 800 MHz spectrometers equipped with triple resonance probes. NMR data were processed with NMRPipe (56), and analyzed using CARRA (57) (version 1.8.4), XIPP (58) (version 1.19.6 p0), and NMRFAM-Sparky (59). <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N protein chemical shifts were assigned using the following experiments: <sup>15</sup>N-HSQC, <sup>13</sup>C-HSQC, HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HBHA(CO)NH, HNHA, HNHB, CC(CO)NH, H(CCCO)NH, HCCH-COSY, HCCH-TOCSY, and <sup>15</sup>N-TOCSY. Chemical shifts of the unlabeled (natural abundance) sorting signal peptide



**Fig. 5.** Latch mechanism of sortase-catalyzed pilus biogenesis. (A) The characteristic N-terminal “lid” appendage (blue) of  $^{cd}SrtA$  occludes the catalytic cysteine residue. During the acylation step, the lid must be opened to allow substrate entry to the active site, which allows for the formation of an enzyme-SpaA acyl intermediate. The sorting signal peptide (red) is bound in the  $^{cd}SrtA$  binding pocket (dark gray). (B) The acyl reaction intermediate encounters another SpaA molecule on the cell surface and the molecules are arranged such that the peptide docking sites and reactive cysteine and lysine of  $^{cd}SrtA$  and SpaA, respectively, are juxtaposed. This ternary complex in which both SpaA substrates are bound to the pilin polymerase is called the attack complex. Prior to transpeptidation, the signal peptide is partially transferred to the binding groove of the SpaA acceptor substrate. After the signal peptide binds efficiently to SpaA, the pilin lysine (K190, green) nucleophilically attacks the acyl linkage, resolving the intermediate and resulting in a SpaA-SpaA isopeptide linkage. The orientation of the molecules suggests that the AB loop of the SpaA acceptor (purple) may engage in loop-loop interactions with the  $\beta 7/\beta 8$  loop of  $^{cd}SrtA$  (gray). Following transpeptidation, the previously disordered AB loop collapses into a rigid latch-like conformation over the isopeptide linkage, which dissolves the ternary complex and provides additional stability to the linkage site.

were assigned using: two-dimensional (2D) (F1,F2)  $^{13}C$ -filtered NOESY and 2D (F1)  $^{13}C$ ,  $^{15}N$ -filtered TOCSY experiments (60). Protein NOE distance restraints were acquired from  $^{15}N$ - and  $^{13}C$ -edited NOESY spectra (120 ms mixing time), and intermolecular restraints were obtained from three-dimensional (3D) (F1)  $^{13}C$ ,  $^{15}N$ -filtered (F2)  $^{13}C$ -edited NOESY-HSQC, (F1)  $^{13}C$ ,  $^{15}N$ -filtered (F2)  $^{15}N$ -edited NOESY-HSQC, and 2D (F1)  $^{13}C$ -filtered NOESY spectra.  $\psi$  and  $\phi$  dihedral restraints were obtained from secondary  $^{13}C$  chemical shifts as calculated by TALOS-N (61) and  $^3J_{HNH\alpha}$  measurements from the HNHA spectrum. Additional  $\psi$  angle restraints were obtained from analysis of  $^{15}N$ -edited NOESY spectrum. Rotamer assignments and  $\gamma_1$  angle restraints for  $\beta$ -methylene protons were obtained through analysis of  $^{15}N$ -TOCSY, HNHB, HN(CO)HB, and  $^{15}N$ -ROESY spectra.

Structures were determined using the program XPLOR-NIH (62, 63). Initially, NOE cross peaks in the 3D  $^{15}N$ -edited NOESY-HSQC and  $^{13}C$ -edited NOESY-HSQC spectra were assigned automatically using the program UNIO10 (64, 65). The NOESY data were then manually inspected using the program Xipp (58) to verify all cross peak assignments and to identify additional distance restraints. An iterative procedure was used to refine the structure of the protein-peptide complex. In the final round of calculations, 200 structures were generated, which yielded a total of 50 with no NOE, dihedral angle, or scalar coupling violations greater than 0.5 Å, 5°, or 2 Hz, respectively. The structures were sorted based on lowest overall energy, and the top 40 were selected as the ensemble to represent the structure of  $^{15}N$ -SpaA-signal and have been deposited in the PDB (code 7K7F). The programs MOLMOL (66) and PyMOL (67) were used to generate figures.

The  $^{15}N$  relaxation data were collected using 1 mM  $^{15}N$ ,  $^{13}C$ -labeled samples of the apo- and  $^{15}N$ -SpaA-signal complex dissolved in  $H_2O$  on a Bruker Avance 600 MHz NMR spectrometer equipped with a triple resonance cryogenic probe. Data were analyzed using SPARKY (59) and included:  $^{15}N$  longitudinal

relaxation rates ( $R_1$ ), transverse relaxation rates ( $R_2$ ), and  $\{^1H\}$ - $^{15}N$  heteronuclear NOEs. Quantifiable relaxation data could be measured for all parameters for 84 and 97 of 143 backbone amides for the apo- and  $^{15}N$ -SpaA-signal complexes, respectively. For inclusion in the calculations, data from each residue must meet the following criteria: isolated  $^1H$ - $^{15}N$  cross peaks and  $\{^1H\}$ - $^{15}N$  NOE values of  $>0.6$ . For backbone  $R_1$  ( $^{15}N$ ) and  $R_2$  ( $^{15}N$ ) measurements at 600 MHz, the same relaxation delays were used for both apo- and  $^{15}N$ -SpaA-signal complex samples.  $R_1$  ( $^{15}N$ ) measurements used delays of  $T = 1500, 1000, 500, 300$  (x2), 100, and 50 ms.  $R_2$  ( $^{15}N$ ) measurements used delays of 17 (x2), 34, 51, 68, 85, 119 (x2), 153, and 170 ms.

To calculate expected rotational correlation times based on molecular weight, the following relationship between hydrodynamic radius and protein molecular weight was employed:

$$r \approx \sqrt[3]{\frac{3M}{4\pi\rho N_a}} + r_w$$

in which  $M$  = molecular weight of the protein,  $\rho$  = the average density for proteins ( $1.37 \text{ g/cm}^3$ ),  $N_a$  = Avogadro's number, and  $r_w$  = hydration radius ( $1.6$  to  $3.2 \text{ \AA}$ ) (68). After calculating the hydrodynamic radius of the protein of interest, and assuming a spherical approximation, the Stokes' law equation was used to calculate an expected rotational correlation time:

$$\tau_c = \frac{4\pi\eta r^3}{3kT}$$

in which  $\eta$  = the viscosity of the solvent,  $r$  = hydrodynamic radius (calculated above),  $k$  = Boltzmann constant, and  $T$  = acquisition temperature.

Experimental values of  $\tau_c$  for the complex and isolated domains were estimated using the ratio of  $T_1$  and  $T_2$  NMR relaxation rates (69).

**Differential Scanning Fluorimetry and Protease Sensitivity Measurements.** The melting temperature and thermodynamic parameters were extracted from Differential Scanning Fluorimetry data by a method described previously (30). Briefly,  $^{15}\text{N}$ SpaA proteins were diluted to 50  $\mu\text{M}$  in assay buffer (50 mM Tris pH 8.0, 300 mM NaCl), supplemented with 15x SYPRO Orange (Sigma-Aldrich) at a total volume of 20  $\mu\text{L}$ . Thermal denaturation reactions were run on a CFX Connect qPCR system (Bio-Rad). A heating rate of 0.2  $^\circ\text{C}/\text{min}$  was employed from 4 to 95  $^\circ\text{C}$ , and fluorescence measurements (excitation at 525  $\pm$  10 nm, detection at 570  $\pm$  10 nm) were acquired after each 0.5 $^\circ$  step increase. The melting temperature for each protein was determined by the first Derivative method after averaging the three replicate measurements. The  $T_m$  is defined as the midpoint of the transition from folded to unfolded and is identified spectroscopically as the temperature where the rate of fluorescence increases with respect to temperature is greatest. The  $T_m$  was then used to calculate the equilibrium constant of unfolding, as previously described (30). For the limited proteolysis experiments, either  $^{15}\text{N}$ SpaA, the  $^{15}\text{N}$ SpaA-signal complex, or the  $^{15}\text{N}$ SpaA- $^{15}\text{N}$ SpaA dimer was dissolved in assay buffer (50 mM Tris pH 8.0, 300 mM NaCl) at a concentration of 20  $\mu\text{M}$  in a volume of 100  $\mu\text{L}$ . Trypsin protease stock solutions were created as described by the manufacturer (Sigma-Aldrich). A total of 0 to 200 ng of trypsin protease (or 400 ng for  $^{15}\text{N}$ SpaA- $^{15}\text{N}$ SpaA dimer reactions) was added to the reactions and incubated at 37  $^\circ\text{C}$ . Samples from each reaction were taken after 6 and 24 h, separated by SDS-PAGE, and analyzed by densitometry.

**Quantitative Transpeptidation Measurements.** The gel-based fluorescence assays were performed at room temperature as previously described (24, 26). All reactions included 25  $\mu\text{M}$   $^{64}\text{SrT}A^+$ , 250  $\mu\text{M}$  fluorescent sorting signal peptide (fluorescein isothiocyanate [FITC]-KNAGFELPLTGGSGRI) and 25  $\mu\text{M}$  wild-type or variant  $^{15}\text{N}$ SpaA. Time points were taken after 24 h, and the resultant reaction mixture was separated by SDS-PAGE. The protein gels then washed in ddH<sub>2</sub>O and fluorescence data were acquired with a Pharos FX gel imager (Bio-Rad). FITC was detected by excitation with a 488 nm laser line and detection with a 515 to 545 nm emission filter. The same gel was then stained with Coomassie to visualize the total protein content of each lane. Quantification of the kinetic parameters of transpeptidation was carried out by separation of the reaction components by reverse-phase HPLC at various time points and substrate concentrations. Each reaction was incubated at room temperature, and proteins were dissolved in assay buffer (50 mM Tris pH 8.0, 300 mM NaCl), containing 50  $\mu\text{M}$   $^{64}\text{SrT}A$   $\Delta 79$ -87, 1 mM sorting signal peptide (FELPLTGGSG), 5 mM DTT, and 50 to 300  $\mu\text{M}$  wild-type or mutant  $^{15}\text{N}$ SpaA. Different variants were incubated either 3.5 h (wild-type, L62D, AB $\Delta$ 15, and A1705) or 16 h (H60A, W181A, V187D, and Y188G) depending on reactivity. The reactions were stopped by flash freezing in liquid nitrogen after incubation. Each reaction condition was run in duplicate. A total of 25  $\mu\text{L}$  of each reaction was injected onto a Water Symmetry 300 C4 HPLC column (4.6  $\times$  150 mm, 5  $\mu\text{m}$  particle size). Proteins were eluted by applying a gradient from 35 to 46% acetonitrile (with 0.1% trifluoroacetic acid) over 12 min at a flow rate of 1 mL/min. Elution of proteins was monitored by absorbance at 215 nm. Peak height of each elution in the HPLC trace was measured by integration of peak areas using Graphical Analysis (Vernier). Data were plotted as Lineweaver-Burk in order to calculate kinetic parameters for each  $^{15}\text{N}$ SpaA variant.

**SAXS Analysis.** Scattering data were generated at the SIBYLS beamline (Advanced Light Source [ALS], Lawrence Berkeley National Laboratory) (70).

1. C. Danne, S. Dramsi, Pili of gram-positive bacteria: Roles in host colonization. *Res. Microbiol.* **163**, 645–658 (2012).
2. N. A. Ramirez, A. Das, H. Ton-That, New paradigms of pilus assembly mechanisms in gram-positive actinobacteria. *Trends Microbiol.* **28**, 999–1009 (2020).
3. A. Mandlik, A. Das, H. Ton-That, The molecular switch that activates the cell wall anchoring step of pilus assembly in gram-positive bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 14147–14152 (2008).
4. C. Chang, A. Mandlik, A. Das, H. Ton-That, Cell surface display of minor pilin adhesins in the form of a simple heterodimeric assembly in *Corynebacterium diphtheriae*. *Mol. Microbiol.* **79**, 1236–1247 (2011).
5. A. Mandlik, A. Swierczynski, A. Das, H. Ton-That, Pili in gram-positive bacteria: Assembly, involvement in colonization and biofilm development. *Trends Microbiol.* **16**, 33–40 (2008).
6. M. K. Hospenthal, T. R. D. Costa, G. Waksman, A comprehensive guide to pilus biogenesis in Gram-negative bacteria. *Nat. Rev. Microbiol.* **15**, 365–379 (2017).
7. H. Ton-That, O. Schneewind, Assembly of pili in Gram-positive bacteria. *Trends Microbiol.* **12**, 228–234 (2004).

Purified  $^{15}\text{N}$ SpaA- $^{15}\text{N}$ SpaA complex (10 mg/mL) was dissolved in size-exclusion chromatography (SEC) buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 6.5, 100 mM NaCl, and 0.01% NaN<sub>3</sub>) and applied to a Shodex KW-802.5 SEC column for SEC-SAXS. Scattering of the buffer without protein was obtained using SAXS data from the SEC run where no protein was eluted and was subtracted from the merged data from the frames corresponding to the elution of the complex. Radius of gyration ( $R_g$ ) and maximal particle dimension ( $D_{\text{max}}$ ) were calculated by Guinier analysis [BioXTAS RAW (71)] and GNOM (ATSAS software package), respectively. Calculated from the Guinier approximation, the radius of gyration ( $R_g$ ) and forward scattering intensity ( $I(0)$ ) were determined to be 25.1  $\text{\AA}$  and 51.0  $\text{\AA}$ , respectively. From the distance distribution function, the  $D_{\text{max}}$  and Porod volume were calculated to be 8.7 nm and  $3.72 \times 10^4 \text{\AA}^3$ , respectively. The rod-like conformation of the dimer can be inferred from these low-resolution structural parameters, as a spherical with 287 residues would be expected to have a  $R_g$  and  $D_{\text{max}}$  of 19.8  $\text{\AA}$  and 5.1 nm, respectively ( $R_g \approx 3\sqrt{n}$  and  $D_{\text{max}} \approx 2.6R_g$ ) (41). Inspection of the normalized Kratky plot of the SAXS data reveals a clear bell-shaped curve at low  $q$  values with a maximum peak height at 1.32 at a peak position of 2.29, which indicates a small contribution from disordered regions of the complex (idealized peak height and position are  $3/e \approx 1.1$  and  $qR_g = \sqrt{3} \approx 1.73$ , respectively) (41, 42). The program GASBOR (72) was used to calculate low-resolution ab initio models in which each residue of the protein is represented as dummy residues (DRs), starting from a random distribution inside a search box with long axis of diameter  $D_{\text{max}}$  followed by a simulated annealing protocol to condense the DR distribution to fit the experimental scattering data.

For rigid body modeling, we began by generating the  $^{15}\text{N}$ SpaA- $^{15}\text{N}$ SpaA isopeptide dimer starting structure by merging the crystal and NMR coordinates of  $^{15}\text{N}$ SpaA (PDB:3HR6) and  $^{15}\text{N}$ SpaA-signal (this work), respectively, into a single coordinate file using PyMOL (67). The coordinates were energy minimized in GROMACS (73) to remove steric clashes or inappropriate geometries. Based on the lack of electron density in the crystal structure in positions C-terminal to K484 and lack of defined peptide orientation in our NMR complex N-terminal to F489, the positions in the sorting signal from K483-A488 were defined as flexible residues and generated 10,000 conformations which explored the conformational space available through rotation of those backbone dihedrals using a Rapidly Exploring Random Tree search algorithm in the Integrative Modeling Platform software package (74). A SAXS profile was then generated for each model using the FOXS method and the best scoring multistate models were enumerated. The MultiFOX algorithm predicted two single-state models, one with a significantly better fit to the experimental SAXS data ( $\chi^2 = 0.89$ , 1.24). For subsequent analysis, we chose the model with the superior  $\chi^2$  value. No multistate models were predicted, ruling out the possibility that the complex consists of an ensemble of conformations in solution.

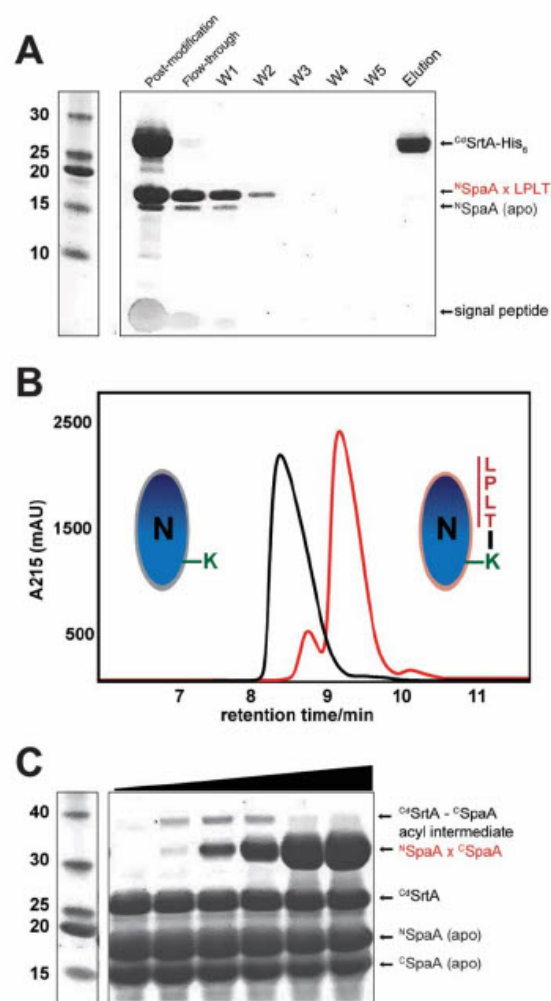
**Data Availability.** Protein structure data have been deposited in PDB (7K7F).

**ACKNOWLEDGMENTS.** We thank members of the R.T.C. and H.T.-T. laboratories for discussions and reviewing this manuscript. This research was supported by funding from the NIH under the Award Numbers AI52217 (to R.T.C.), DE025015 and DE017382 (to H.T.-T.), and T32 GM007185 and the UCLA Graduate Division Dissertation Year Fellowship (to S.A.M.). We also acknowledge NIH shared instrumentation Grants S10OD025073 and S10OD016336 and support from the US Department of Energy (DE-FC03-02ER63421). SAXS data were collected at SIBYLS, which is supported by the DOE-BER IDAT, DE-AC02-05CH11231, and National Institute of General Medical Sciences ALS-ENABLE (P30 GM124169 and S10OD018483).

8. J. L. Telford, M. A. Barocchi, I. Margarit, R. Rappuoli, G. Grandi, Pili in gram-positive pathogens. *Nat. Rev. Microbiol.* **4**, 509–519 (2006).
9. J. R. Scott, D. Zähler, Pili with strong attachments: Gram-positive bacteria do it differently. *Mol. Microbiol.* **62**, 320–330 (2006).
10. H. J. Kang, E. N. Baker, Structure and assembly of gram-positive bacterial pili: Unique covalent polymers. *Curr. Opin. Struct. Biol.* **22**, 200–207 (2012).
11. K. A. Kline, K. W. Dodson, M. G. Caparon, S. J. Hultgren, A tale of two pili: Assembly and function of pili in bacteria. *Trends Microbiol.* **18**, 224–232 (2010).
12. T. Proft, E. N. Baker, Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. *Cell. Mol. Life Sci.* **66**, 613–635 (2009).
13. R. Fronzes, H. Remaut, G. Waksman, Architectures and biogenesis of non-flagellar protein appendages in Gram-negative bacteria. *EMBO J.* **27**, 2271–2280 (2008).
14. T. Abee, A. T. Kovács, O. P. Kuipers, S. van der Veen, Biofilm formation and dispersal in Gram-positive bacteria. *Curr. Opin. Biotechnol.* **22**, 172–179 (2011).
15. N. Van Gerven, G. Waksman, H. Remaut, "Pili and flagella: Biology, structure, and biotechnological applications" in *Progress in Molecular Biology and Translational Science*, S. Howorka, Ed. (Elsevier B.V., 2011), 103, pp. 21–72.

16. H. Ton-That, O. Schneewind, Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Mol. Microbiol.* **50**, 1429–1438 (2003).
17. H. Ton-That, L. A. Marraffini, O. Schneewind, Sortases and pilin elements involved in pilus assembly of *Corynebacterium diphtheriae*. *Mol. Microbiol.* **53**, 251–261 (2004).
18. I. K. Guttilla *et al.*, Acyl enzyme intermediates in sortase-catalyzed pilus morphogenesis in gram-positive bacteria. *J. Bacteriol.* **191**, 5603–5612 (2009).
19. B. Khare, S. V. L. Narayana, Pilus biogenesis of Gram-positive bacteria: Roles of sortases and implications for assembly. *Protein Sci.* **26**, 1458–1473 (2017).
20. H. J. Kang, F. Coulibaly, F. Clow, T. Proft, E. N. Baker, Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. *Science* **318**, 1625–1628 (2007).
21. H. J. Kang, E. N. Baker, Intramolecular isopeptide bonds: Protein crosslinks built for stress? *Trends Biochem. Sci.* **36**, 229–237 (2011).
22. M. Hilleringmann *et al.*, Molecular architecture of *Streptococcus pneumoniae* TIGR4 pili. *EMBO J.* **28**, 3921–3930 (2009).
23. H. J. Kang, N. G. Paterson, A. H. Gaspar, H. Ton-That, E. N. Baker, The *Corynebacterium diphtheriae* shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 16967–16971 (2009).
24. S. A. S. A. McConnell *et al.*, Protein labeling via a specific lysine-isopeptide bond using the pilin polymerizing sortase from *Corynebacterium diphtheriae*. *J. Am. Chem. Soc.* **140**, 8420–8423 (2018).
25. C. Chang *et al.*, In vitro reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5477–E5486 (2018).
26. C. K. Sue *et al.*, Kinetics and optimization of the lysine-isopeptide bond forming sortase enzyme from *Corynebacterium diphtheriae*. *Bioconjug. Chem.* **31**, 1624–1634 (2020).
27. A. H. Chan *et al.*, Structure of the bacillus anthracis sortase A enzyme bound to its sorting signal: A flexible amino-terminal appendage modulates substrate access. *J. Biol. Chem.* **290**, 25461–25474 (2015).
28. N. Suree *et al.*, The structure of the *Staphylococcus aureus* sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. *J. Biol. Chem.* **284**, 24465–24477 (2009).
29. B. Baljinyam, M. Ronzetti, A. Yasgar, A. Simeonov, Applications of differential scanning fluorimetry and related technologies in characterization of protein-ligand interactions. *Methods Mol. Biol.* **2089**, 47–68 (2020).
30. T. A. Wright, J. M. Stewart, R. C. Page, D. Konkolewicz, Extraction of thermodynamic parameters of protein unfolding using parallelized differential scanning fluorimetry. *J. Phys. Chem. Lett.* **8**, 553–558 (2017).
31. D. Schneidman-Duhovny, M. Hammel, J. A. Tainer, A. Sali, FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W424–W429 (2016).
32. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 714–797 (2007).
33. G. Spraggon *et al.*, Supramolecular organization of the repetitive backbone unit of the *Streptococcus pneumoniae* pilus. *PLoS One* **5**, e10919 (2010).
34. J. M. Budzik *et al.*, Intramolecular amide bonds stabilize pili on the surface of bacilli. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19992–19997 (2009).
35. R. Cozzi *et al.*, Structure and assembly of group B streptococcus pilus 2b backbone protein. *PLoS One* **10**, e0125875 (2015).
36. H. J. Kang *et al.*, A slow-forming isopeptide bond in the structure of the major pilin SpaD from *Corynebacterium diphtheriae* has implications for pilus assembly. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 1190–1201 (2014).
37. A. Mishra *et al.*, Two autonomous structural modules in the fimbrial shaft adhesin FimA mediate Actinomyces interactions with streptococci and host cells during oral biofilm development. *Mol. Microbiol.* **81**, 1205–1220 (2011).
38. K. Vengadesan, X. Ma, P. Dwivedi, H. Ton-That, S. V. L. Narayana, A model for group B *Streptococcus* pilus type 1: The structure of a 35-kDa C-terminal fragment of the major pilin GB580. *J. Mol. Biol.* **407**, 731–743 (2011).
39. A. Nuccitelli *et al.*, Structure-based approach to rationally design a chimeric protein for an effective vaccine against Group B *Streptococcus* infections. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10278–10283 (2011).
40. P. Chaurasia, S. Pratap, A. Palva, I. von Ossowski, V. Krishnan, Bent conformation of a backbone pilin N-terminal domain supports a three-stage pilus assembly mechanism. *Commun. Biol.* **1**, 94 (2018).
41. D. Durand *et al.*, NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **169**, 45–53 (2010).
42. V. Receveur-Brechot, D. Durand, How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr. Protein Pept. Sci.* **13**, 55–75 (2012).
43. E. A. Rogers, A. Das, H. Ton-That, Adhesion by pathogenic *Corynebacterium*. *Adv. Exp. Med. Biol.* **715**, 91–103 (2011).
44. A. Mandlik, A. Swierczynski, A. Das, H. Ton-That, *Corynebacterium diphtheriae* employs specific minor pilins to target human pharyngeal epithelial cells. *Mol. Microbiol.* **64**, 111–124 (2007).
45. D. J. Echelman *et al.*, CnaA domains in bacterial pili are efficient dissipaters of large mechanical shocks. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2490–2495 (2016).
46. K. P. Kilambi, J. J. Gray, Rapid calculation of protein pKa values using Rosetta. *Biophys. J.* **103**, 587–595 (2012).
47. S. Lyskov *et al.*, Serverification of molecular modeling applications: The Rosetta online server that includes everyone (ROSIE). *PLoS One* **8**, e63906 (2013).
48. R. Nagarajan *et al.*, PDBparam: Online resource for computing structural parameters of proteins. *Bioinform. Biol. Insights* **10**, 73–80 (2016).
49. K. Persson, A. Esberg, R. Claesson, N. Strömberg, The pilin protein FimP from actinomyces oris: Crystal structure and sequence analyses. *PLoS One* **7**, e48364 (2012).
50. P. G. Young *et al.*, Structural conservation, variability, and immunogenicity of the T6 backbone pilin of serotype M6 *Streptococcus pyogenes*. *Infect. Immun.* **82**, 2949–2957 (2014).
51. L. El Mortaji *et al.*, The full-length *Streptococcus pneumoniae* major pilin RrgB crystallizes in a fiber-like structure, which presents the D1 isopeptide bond and provides details on the mechanism of pilus polymerization. *Biochem. J.* **441**, 833–841 (2011).
52. P. Chaurasia, S. Pratap, I. von Ossowski, A. Palva, V. Krishnan, New insights about pilus formation in gut-adapted *Lactobacillus rhamnosus* GG from the crystal structure of the SpaA backbone-pilin subunit. *Sci. Rep.* **6**, 28664 (2016).
53. M. M. Shaik *et al.*, A structural snapshot of type II pilus formation in *Streptococcus pneumoniae*. *J. Biol. Chem.* **290**, 22581–22592 (2015).
54. C. T. Rueden *et al.*, ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529 (2017).
55. C. Chang, I. H. Huang, A. P. A. Hendrickx, H. Ton-That, Visualization of Gram-positive bacterial pili. *Methods Mol. Biol.* **966**, 77–95 (2013).
56. F. Delaglio *et al.*, NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
57. R. Keller, *The Computer Aided Resonance Assignment Tutorial* (Cantina Verlag, Goldau, Switzerland, 2004).
58. D. S. Garrett, M. Cai, G. M. Clore, XIPP: Multi-dimensional NMR analysis software. *J. Biomol. NMR* **74**, 9–25 (2020).
59. W. Lee, M. Tonelli, J. L. Markley, NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).
60. G. M. Clore, A. Bax, A. M. Gronenborn, Stereospecific assignment of  $\beta$ -methylene protons in larger proteins using 3D 15N-separated Hartmann-Hahn and 13C-separated rotating frame Overhauser spectroscopy. *J. Biomol. NMR* **1**, 13–22 (1991).
61. Y. Shen, A. Bax, Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013).
62. C. D. Schwieters, J. J. Kuszewski, G. Marius Clore, Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **48**, 47–62 (2006).
63. C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. M. Clore, The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73 (2003).
64. T. Herrmann, P. Güntert, K. Wüthrich, Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209–227 (2002).
65. T. Herrmann, P. Güntert, K. Wüthrich, Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR* **24**, 171–189 (2002).
66. R. Koradi, M. Billeter, K. Wüthrich, MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55, 29–32 (1996).
67. W. L. DeLano, The PyMOL Molecular Graphics System, Version 2.3. Schrödinger LLC www.pymol.org (2020). 10.1038/hr.2014.17.
68. J. Cavanagh, *Protein NMR Spectroscopy: Principles and Practice* (Academic Press, 2007).
69. L. E. Kay, D. A. Torchia, A. Bax, Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry* **28**, 8972–8979 (1989).
70. S. Classen *et al.*, Implementation and performance of SIBYLS: A dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the advanced light source. *J. Appl. Cryst.* **46**, 1–13 (2013).
71. J. B. Hopkins, R. E. Gillilan, S. Skou, BioXTAS RAW: Improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J. Appl. Cryst.* **50**, 1545–1553 (2017).
72. D. I. Svergun, M. V. Petoukhov, M. H. J. Koch, Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001).
73. S. M. Hacker *et al.*, Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **9**, 1181–1190 (2017).
74. D. Russel *et al.*, Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* **10**, e1001244 (2012).

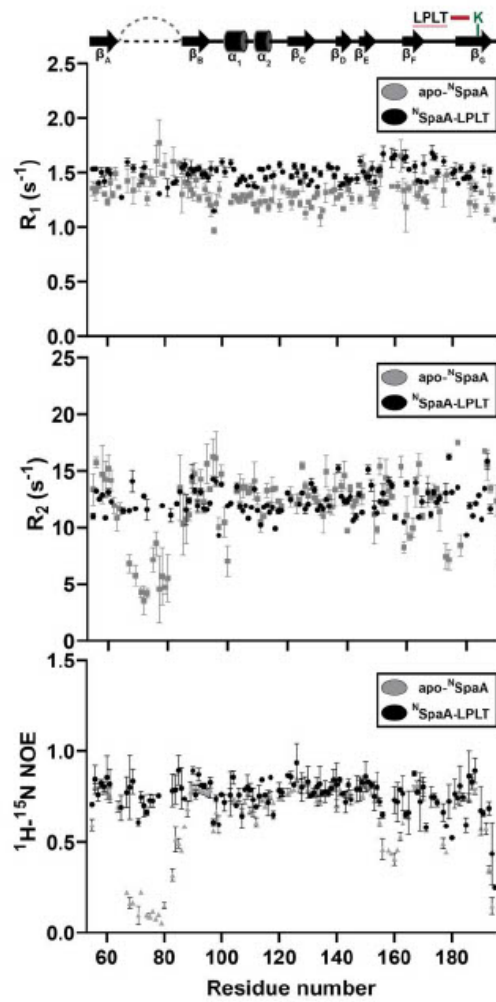
## 2.3 Supplemental Information



**Figure S1: Enzymatic production of the  $^{15}\text{N}$ SpaA-signal complexes.** A) Signal peptide modification and purification was monitored by SDS-PAGE.  $^{13}\text{C}$ SrtA<sup>A</sup>-His<sub>6</sub> was incubated with  $^{15}\text{N}$ SpaA and the KNAGFELPLTGGSGRI signal peptide for 24 h resulting in nearly complete modification of the  $^{15}\text{N}$ SpaA domain.  $^{13}\text{C}$ SrtA<sup>A</sup>-His<sub>6</sub> and free peptide were



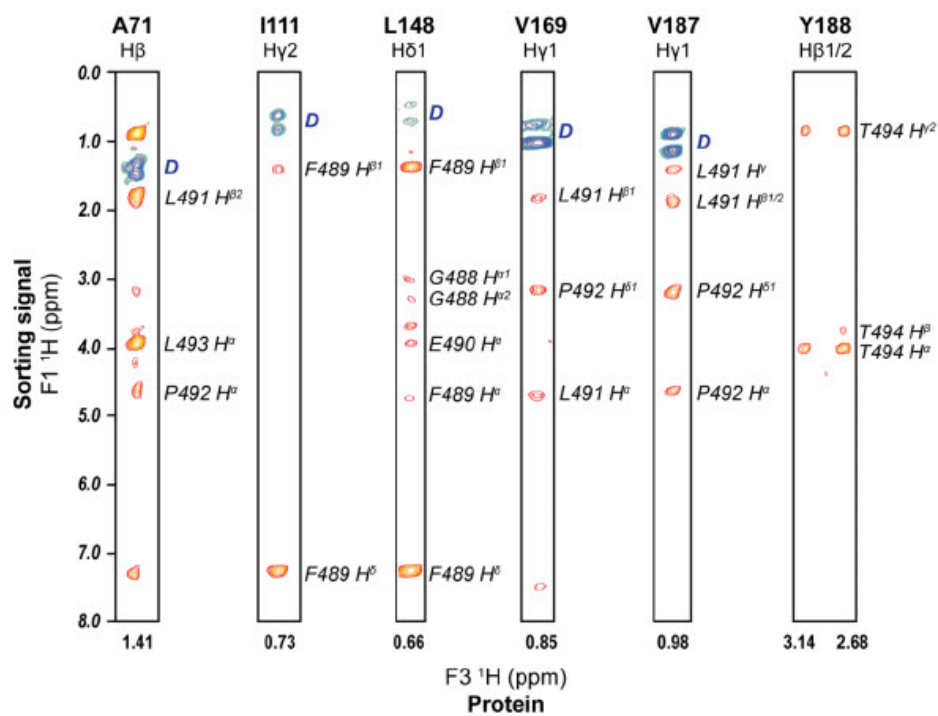
subsequently purified from the sample by application of the reaction to a  $\text{Co}^{2+}$  IMAC resin and spin filtration. The  $^{15}\text{N}$ SpaA domain eluted during the wash steps, while  $^{13}\text{C}$ SrtA<sup>A</sup>-His<sub>6</sub> was retained on the column until elution with 200 mM imidazole. Lane 1, full reaction after incubation; Lane 2, flowthrough from IMAC column containing  $^{15}\text{N}$ SpaA and signal peptide; Lane 3-7, 10 mL washes; Lane 8, elution with 50 mL buffer supplemented with 200 mM imidazole containing His<sub>6</sub>-tagged sortase. B) Reversed-phase HPLC analysis of the apo- $^{15}\text{N}$ SpaA and  $^{15}\text{N}$ SpaA-signal complex. The reaction mixture was separated on a Waters Symmetry 300 C4 HPLC column after a 24 h incubation with sortase, resulting in >95% modification. C) SDS-PAGE analysis of the  $^{15}\text{N}$ SpaA-signal complex.  $^{13}\text{C}$ SrtA<sup>A</sup> was incubated with  $^{15}\text{N}$ SpaA and  $^{13}\text{C}$ SpaA domains and the reactions were monitored over time (t=0, 5 min, 30 min, 2 h, 24 h, and 48 h). Appearance of a high molecular weight band on the reducing SDS-PAGE gel indicates the formation of the covalent  $^{13}\text{C}$ SpaA- $^{15}\text{N}$ SpaA species.



**Figure S2: NMR relaxation data for apo-<sup>15</sup>N SpaA and <sup>15</sup>N SpaA-signal peptide complex.**

Panels A-C show scatter plots of the relaxation data for apo-<sup>15</sup>N SpaA and the <sup>15</sup>N SpaA-signal peptide complex. <sup>15</sup>N longitudinal ( $R_1$ ; A), and transverse ( $R_2$ ; B) relaxation rates, and  $\{^1\text{H}\}$ -<sup>15</sup>N

heteronuclear NOEs (hetNOE; C) are plotted as a function of residue number. The secondary structures present in the protein are displayed on top of the first plot.



**Figure S3: Representative intermolecular NOE data.** Strips from a three-dimensional (F1)  $^{13}\text{C}$ ,  $^{15}\text{N}$ -filtered, (F2)  $^{13}\text{C}$  -edited NOESY-HSQC spectrum of the  $^{\text{N}}\text{SpaA}$ -signal complex. The residue (top) and proton chemical shift (bottom) from isotopically-labeled  $^{\text{N}}\text{SpaA}$  protein that gives rise to the set of intermolecular NOEs are indicated for each data strip. The cross peak from the unlabeled signal peptide is labeled to the right of each peak. The diagonal peak is indicated by a D.

**Table S1.** Structural statistics of the solution structure of <sup>N</sup>SpaA-signal complex.

|  | $\langle SA \rangle^a$ | $\langle \overline{SA} \rangle^a$ |
|--|------------------------|-----------------------------------|
| Root mean square deviations                                |                        |                                   |
| NOE interproton distance restraints (Å) <sup>b</sup>       |                        |                                   |
| All (2070)   | 0.036 ± 0.002          | 0.034                             |
| Intermolecular (66)  | 0.033 ± 0.006          | 0.030                             |
| Dihedral angles restraints (°) <sup>c</sup> (305)          | 0.515 ± 0.074          | 0.639                             |
| <sup>3</sup> J <sub>HNA</sub> coupling constants (Hz) (92) | 0.774 ± 0.030          | 0.882                             |
| Secondary <sup>13</sup> C shifts (p.p.m.)                  |                        |                                   |
| <sup>13</sup> C <sub>α</sub> (125)                         | 1.072 ± 0.174          | 1.054                             |
| <sup>13</sup> C <sub>β</sub> (125)                         | 1.410 ± 0.175          | 1.340                             |
| Deviation from idealized covalent geometry                 |                        |                                   |
| bonds (Å)  | 0.011 ± 0.000          | 0.005                             |
| angles (°)   | 0.724 ± 0.010          | 0.633                             |
| impropers (°)  | 0.391 ± 0.015          | 0.528                             |
| PROCHECK results (%) <sup>d</sup>                          |                        |                                   |
| most favorable region                                      | 87.8 ± 1.3             | 84.3                              |
| additionally allowed region                                | 11.0 ± 1.4             | 13.2                              |
| generously allowed region                                  | 1.6 ± 0.6              | 2.5                               |
| disallowed region  | 0.0 ± 0.0              | 0.0                               |
| Coordinate Precision (Å) <sup>e</sup>                      |                        |                                   |
| Protein backbone   | 0.48 ± 0.05            |                                   |
| Protein heavy atoms  | 0.81 ± 0.04            |                                   |

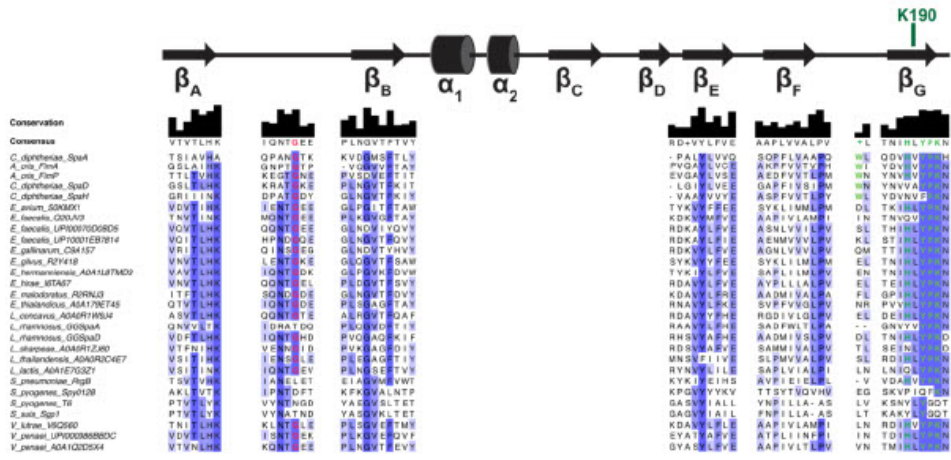
<sup>a</sup>  $\langle SA \rangle^a$  represents an ensemble of the 20 best structures calculated by simulated annealing.  $\langle \overline{SA} \rangle^a$  represents the average energy-minimized structure. The number of terms for each restraint is given in parentheses. None of the structures exhibited distance violations greater than 0.5 Å, dihedral angle violations greater than 5°, coupling constant violations greater than 2 Hz

<sup>b</sup> Distance restraints: 570 sequential, 233 medium (2 ≤ residue separation ≤ 4) and 795 long range (>4 residues apart)

<sup>c</sup> The experimental dihedral angle restraints were as follows: 130 φ, 131 ψ, and 44 χ<sub>1</sub> angular restraints

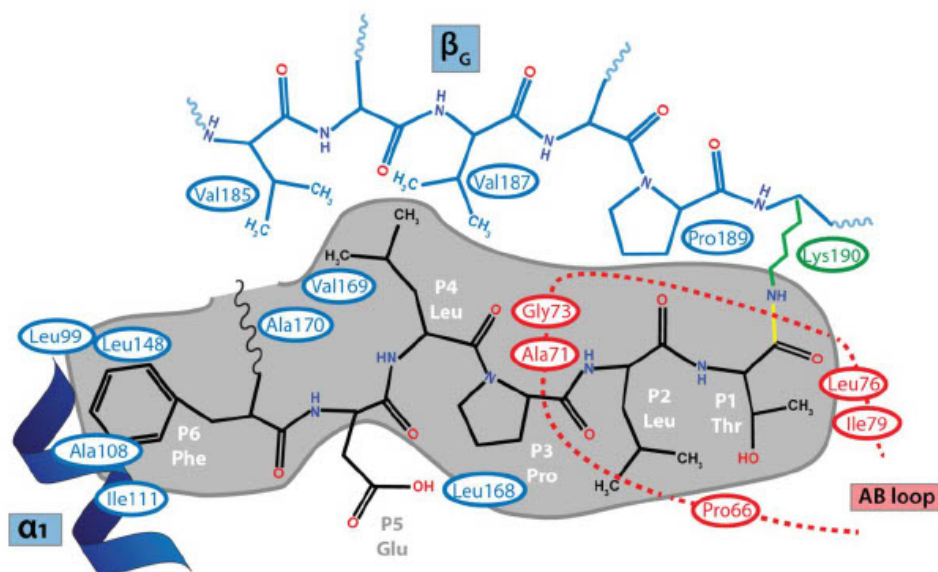
<sup>d</sup> PROCHECK-NMR<sup>1</sup> data includes residues 52-195 of <sup>N</sup>SpaA and residues 488-494 of the signal peptide. For the structured regions of the protein and peptide, 98 ± 0.1 % of the residues were in the favored or allowed regions of the Ramachandran plot

<sup>e</sup> The coordinate precision is defined as the average atomic root mean square deviation (rmsd) of the 20 individual SA structures and their mean coordinates. These values are for residues 54-192 of <sup>N</sup>SpaA and residues 488-494 in the signal peptide. Backbone atoms refers to the N, C<sup>α</sup>, and C' atoms



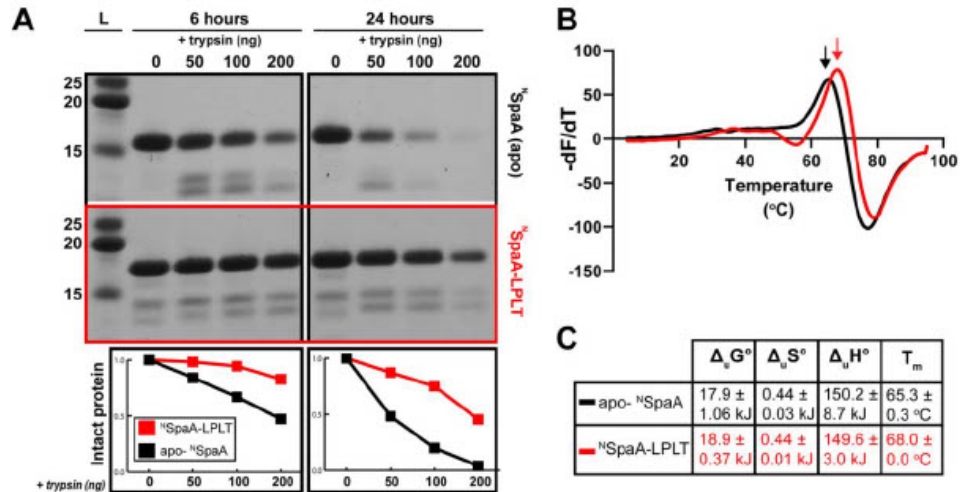
**Figure S4: Multiple sequence alignment of pilin proteins.** The figure shows the aligned primary sequences of N-terminal domains in pilin proteins that are known to be covalently crosslinked to sorting signals via lysine isopeptide bonds. Residues important for crosslinking are color coded. Green, pilin motif residues; and red, conserved residues in the AB loop that contact the signal peptide. The alignment was generated using the T-Coffee alignment algorithm as implemented in the program JalView. The location of secondary structural elements of *N*SpaA and the reactive K190 residue are shown above the alignment. The following sequences are shown: *Cdiphtheriae\_SpaA* [CAB0816384], *Aoris\_FimA* [PKY85323], *Aoris\_FimP* [AAD28827], *Cdiphtheriae\_SpaD* [OWO35423], *Cdiphtheriae\_SpaH* [CAB0527416], *Eavium\_S0KMX1* [OJG21096], *Efaecalis\_Q20JV3* [WP\_010815870], *Efaecalis\_UP100070D0BD5* [WP\_057090439], *Efaecalis\_UP10001EB7814* [WP\_077836692], *Egallinarum\_C9A157* [WP\_123866913], *Egilvus\_R2Y418* [OJG42114], *Ehermanniensis\_A0A1L8TMD3* [OJG45383], *Ehirae\_I6TA67* [OJG52668], *Emalodoratus\_R2RNJ3* [OJG57920],

Ethialandicus\_A0A179ET45 [OJG95917], Lconcavus\_A0A0R1W9J4 [WP\_057823694],  
Lrhamnosus\_GGSpaA [PTS45943], Lrhamnosus\_GGSpaD [OAU92240],  
Lsharpae\_A0A0R1ZJ60 [KRM54984], Lthailandensis\_A0A0R2C4E7 [WP\_054750861],  
Llactis\_A0A1E7G3Z1 [PCS20467], Spneumoniae\_RrgB [SND77541],  
Spyogenes\_Spy0128 [BAQ50580], Spyogenes\_T6 [ESA45498], Ssuis\_Sgp1  
[ALA29851], Vlutrae\_V6Q560 [WP\_023606321], Vpenaei\_UPI000986BBDC  
[WP\_077275831], Vpenaei\_A0A1Q2D5X4 [WP\_077275817].



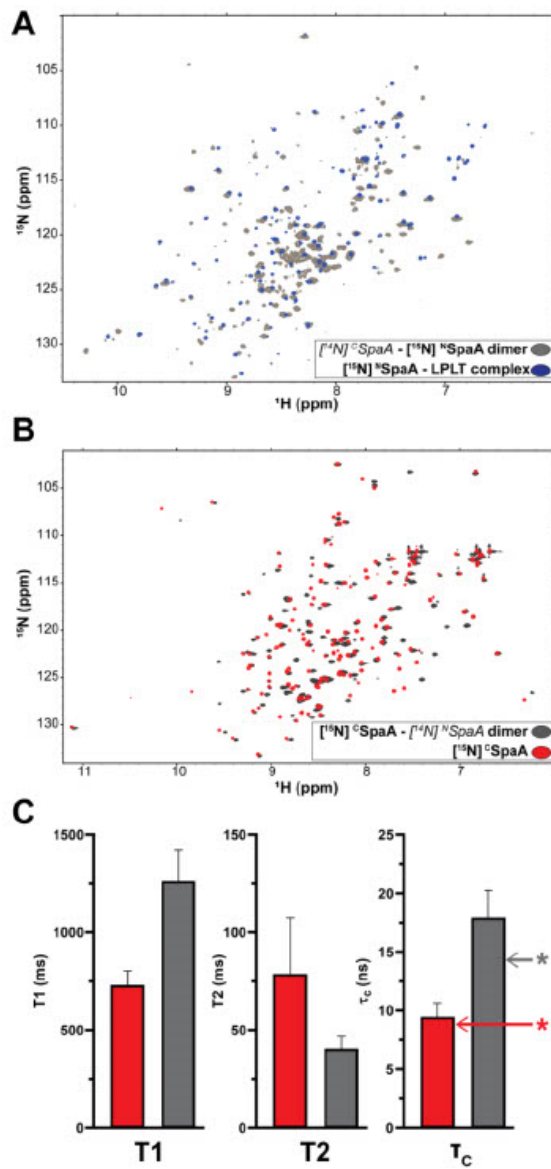
**Figure S5: Contact map summarizing protein-peptide interactions in the NMR structure of the <sup>15</sup>SpaA-signal complex.** Schematic representation of the hydrophobic contacts between the signal peptide and <sup>15</sup>SpaA. The crosslinked peptide is shown as black lines. Contacting residues in <sup>15</sup>SpaA are indicated in blue, while AB loop contacts are highlighted in red. Residues within  $\beta$ -strand G are shown explicitly, while other contacts in the binding groove are indicated by ovals at their positions relative to the peptide, labeled with their residue name. The program MOLMOL was used to analyze the NMR structure of the average energy minimized structure of the <sup>15</sup>SpaA-signal complex.



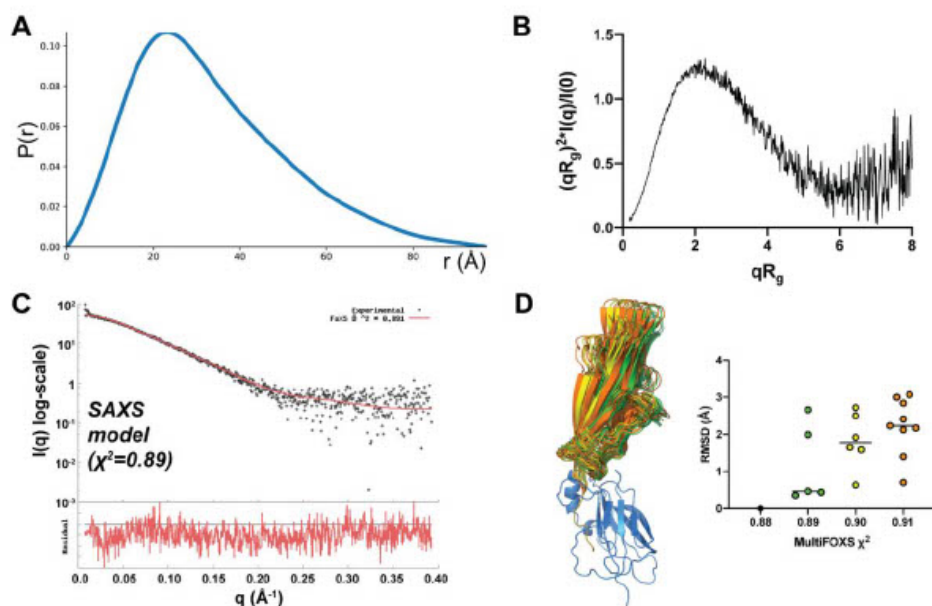


**Figure S6: <sup>N</sup>SpaA thermostability and proteolytic stability.** The <sup>N</sup>SpaA-signal complex exhibits enhanced thermo- and proteolytic stability. A) SDS-PAGE data showing the results of limited proteolysis experiments comparing the proteolytic stability of the apo- and peptide-conjugated forms of <sup>N</sup>SpaA. Each lane shows the amount of protein degradation that occurs when increasing amounts of trypsin are added (0, 50, 100, 200 ng). Left and right panels correspond to data collected after 6- and 24-hour exposure to trypsin, respectively. The bottom panels show the band intensities corresponding to intact protein for each experiment, normalized to the no trypsin control for each experiment. B) Protein unfolding is measured by differential scanning fluorimetry which tracks increased fluorescence of the SYPRO fluorophore upon protein unfolding. Melting curves were obtained as the average of four replicates. The first derivative curves of the DSF experiments are plotted in black for apo-<sup>N</sup>SpaA and red for <sup>N</sup>SpaA-signal complex. The

maxima of the derivative plot correspond to the melting temperature, indicated by red or black arrows. C) Fitting of thermal unfolding data to the relationship  $\Delta_u G = -RT \ln K_u$  yields thermodynamic parameters for each protein. Calculated  $\Delta_u G$  values were calculated based on unfolding measured at temperatures representing 10-50% unfolding<sup>2,3</sup>.

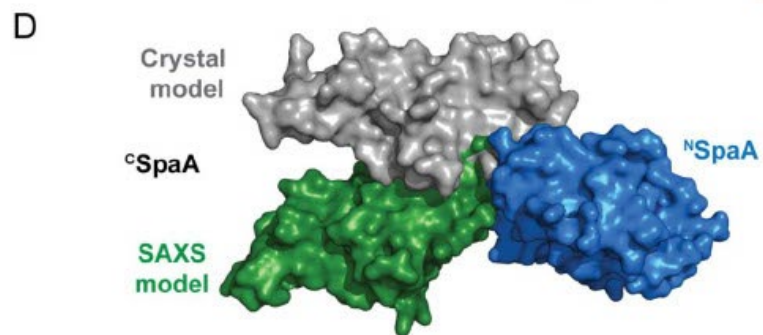
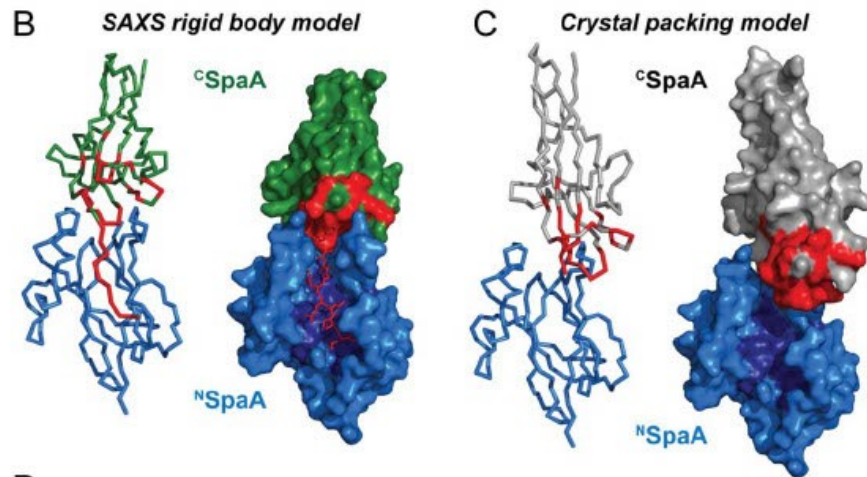
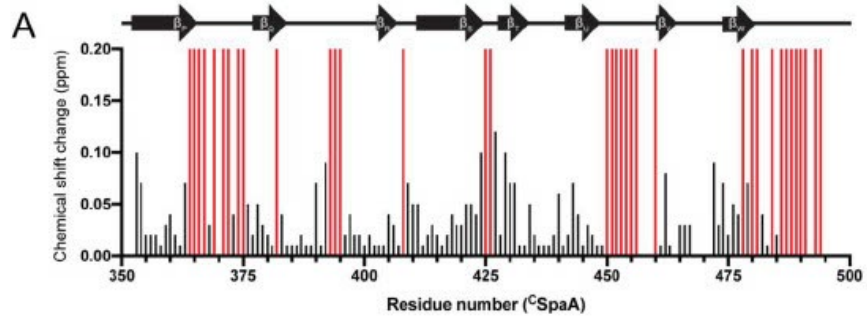


**Figure S7: NMR relaxation data indicates that the  $^{13}\text{C}\text{SpaA} - ^{15}\text{N}\text{SpaA}$  dimer is a rigid unit.** A)  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra for the  $[^{15}\text{N}]\text{SpaA}$ -LPLT complex (blue) overlaid with the spectrum of the  $[^{14}\text{N}]^{13}\text{C}\text{SpaA}$ -  $[^{15}\text{N}]^{15}\text{N}\text{SpaA}$  dimer (gray). B)  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra for monomeric  $[^{15}\text{N}]^{13}\text{C}\text{SpaA}$  (red) and  $[^{15}\text{N}]^{13}\text{C}\text{SpaA}$ -  $[^{14}\text{N}]^{15}\text{N}\text{SpaA}$  dimer (gray) are overlaid. C) Average longitudinal ( $T_1$ ) and transverse ( $T_2$ ) relaxation times for well resolved residues of  $[^{15}\text{N}]^{13}\text{C}\text{SpaA}$  (red) and  $[^{15}\text{N}]^{13}\text{C}\text{SpaA}$ -  $[^{14}\text{N}]^{15}\text{N}\text{SpaA}$  dimer (gray) are presented. Global rotational correlation times ( $\tau_c$ ) for each molecule (calculated from  $T_1/T_2$  values) are plotted. Asterisks indicate the expected correlation times for globular molecules of the molecular weights corresponding to the isolated  $^{13}\text{C}\text{SpaA}$  protein (red) and the  $^{13}\text{C}\text{SpaA}$ - $^{15}\text{N}\text{SpaA}$  dimer (gray).



**Figure S8: Small angle X-ray scattering data of the <sup>C</sup>SpaA-<sup>N</sup>SpaA complex.** SAXS data for the crosslinked <sup>C</sup>SpaA-<sup>N</sup>SpaA complex. A) Distance distribution,  $p(r)$  versus distance ( $r$ ). B) A dimensionless Kratky plot of the SAXS data. C) Experimental raw x-ray scattering data (black) shows agreement to the theoretical profile calculated from the MultiFOXS rigid body model of the complex (red)<sup>4</sup>. Below the figure are shown the residuals as a function of  $q$  value ( $(\text{experimental} - \text{theoretical}) / \text{error}$ ). The model fits the data with a  $\chi^2$  of 0.89. D) MultiFOXS models consist of several calculated conformations which are clustered based on structural similarity. To assess the structural heterogeneity of the best fit bundle of structures in our model, the clustering threshold was reduced ( $t=0.001$ ) to reveal additional clusters with subtle differences in conformation. The top 9 model clusters (of 1550 total calculated clusters) with  $\chi^2$  fits from 0.88-0.91 were analyzed

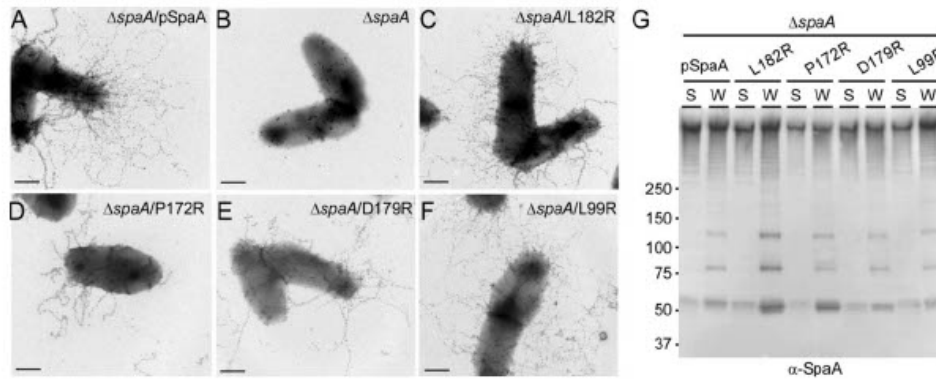
by comparing the all-atom RMSD of each of the structures to the best fit structure ( $\chi^2 = 0.88$ ). The RMSD ranges for clusters in the first four  $\chi^2$  bins are displayed to the right of the structural bundle. For the first 9 clusters, the RMSD ranges from 0.36 to 3.07 Å.



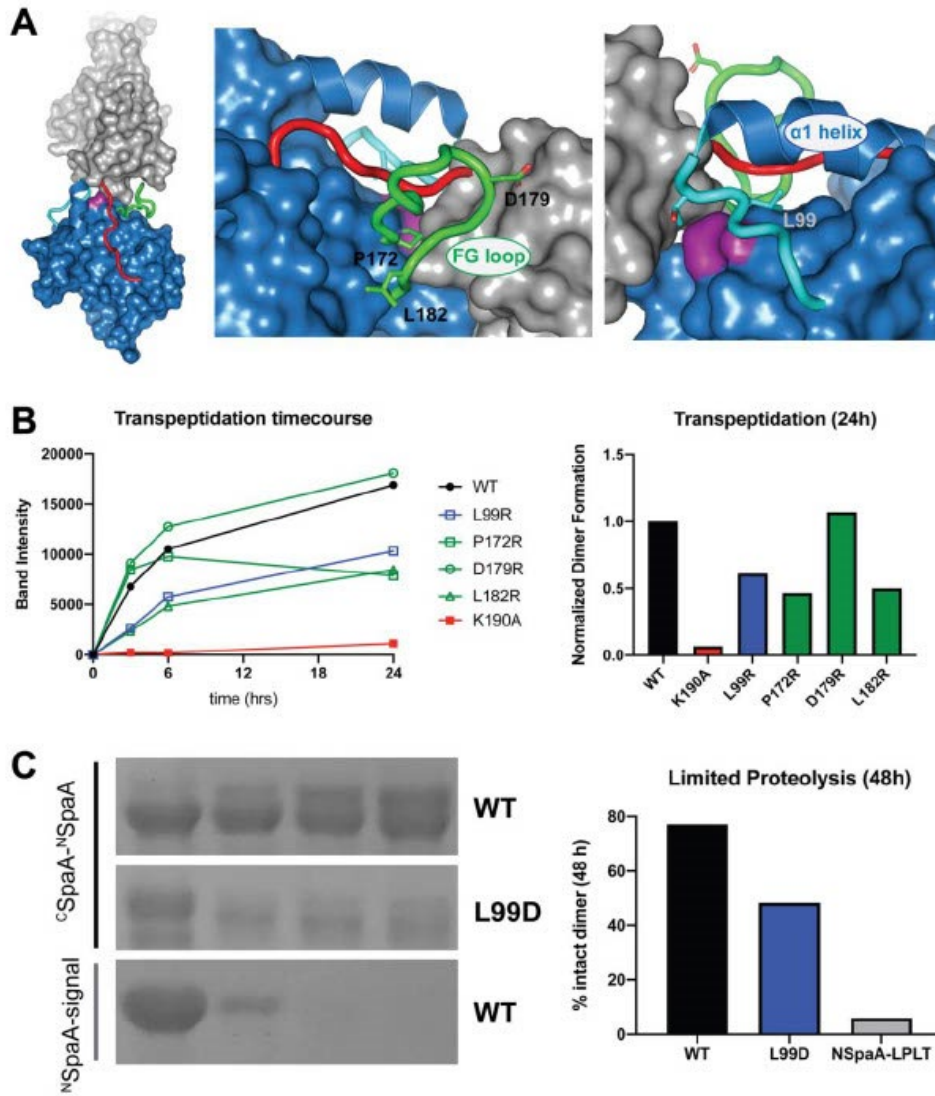
**Figure S9: The structure of the  $^{13}\text{C}$ SpaA- $^{15}\text{N}$ SpaA dimer is compatible with NMR data and distinct from packing interactions observed in crystals.** A) The protein-protein contact interface in the  $^{13}\text{C}$ SpaA- $^{15}\text{N}$ SpaA dimer was mapped by determining NMR chemical shift perturbations (CSPs) between  $^{15}\text{N}$   $^{13}\text{C}$ SpaA and  $^{15}\text{N}$   $^{13}\text{C}$ SpaA -  $^{14}\text{N}$   $^{15}\text{N}$ SpaA dimer in which the  $^{13}\text{C}$ SpaA is selectively labeled with nitrogen-15. The backbone chemical shifts assignments ( $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$ ) of apo- $^{13}\text{C}$ SpaA were assigned through analysis of the following triple resonance experiments:  $^{15}\text{N}$ -HSQC, HNCACB, CBCA(CO)NH, HNCO and HN(CA)CO. The composite CSPs (calculated as  $\sqrt{[\delta_H^2 + (\alpha \cdot \delta_N)^2]}$ , where the scaling factor,  $\alpha$ , is set to 0.154) are plotted with respect to residue number of  $^{13}\text{C}$ SpaA<sup>5</sup>. Chemical shifts that could not be mapped to the complex spectrum due to large deviations were arbitrarily set to 0.2 ppm and highlighted in red. B) The largest CSPs on  $^{13}\text{C}$ SpaA are mapped onto the structure of the  $^{13}\text{C}$ SpaA- $^{15}\text{N}$ SpaA SAXS model dimer and colored red ( $^{15}\text{N}$ SpaA (blue) and  $^{13}\text{C}$ SpaA (green)). In the panel, the structure is shown in ribbon (left) and surface (right) representations. C) As in panel (B), but the largest CSPs are mapped on to the packing interface observed in the crystal of the isolated protein and the  $^{13}\text{C}$ SpaA domain is colored gray. A comparison of panels (B) and (C) clearly shows that the binding surface identified by NMR CSPs is more compatible with the SAXS-derived structure of the  $^{13}\text{C}$ SpaA- $^{15}\text{N}$ SpaA dimer. D) Comparison of the MultiFOXS rigid body model of the  $^{13}\text{C}$ SpaA- $^{15}\text{N}$ SpaA dimer with packing interactions observed in crystals of the isolated SpaA protein (PDB 3HR6). The  $^{15}\text{N}$ SpaA domains (blue) are aligned to emphasize differences in the orientation of the  $^{13}\text{C}$ SpaA domain in the  $^{13}\text{C}$ SpaA- $^{15}\text{N}$ SpaA dimer in the SAXS-derived solution conformation (green) and crystal packing orientation (gray). The non-crosslinked  $^{15}\text{N}$ SpaA



and <sup>c</sup>SpaA domains in the crystal are rotated by ~25° and separated by an additional ~9 Å as compared to the solution structure.

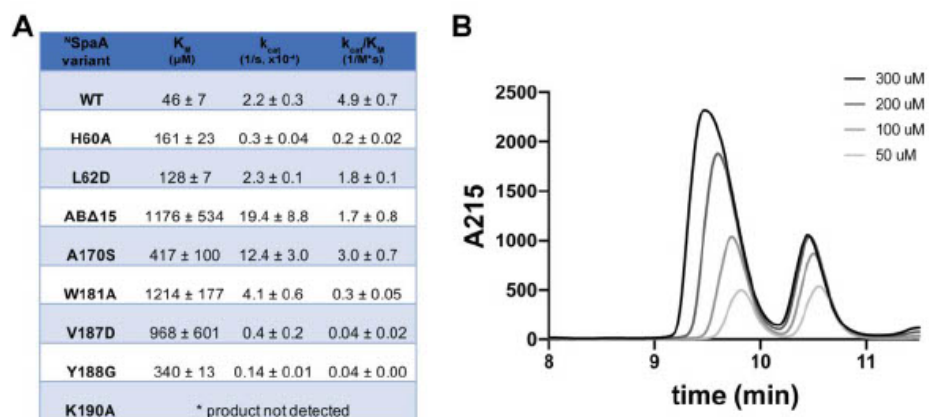


**Figure S10: *In vivo* biochemical and electron microscopic analysis of <sup>C</sup>SpaA-<sup>N</sup>SpaA interfacial mutants.** (A-F) Electron microscopy images of *C. diphtheriae* cells expressing wild-type and SpaA proteins containing single amino acid substitutions. Cells of the *C. diphtheriae*  $\Delta spaA$  mutant or those expressing wild-type SpaA or individual SpaA mutants from a plasmid were immobilized on carbon-coated nickel grids. They were then stained with  $\alpha$ -SpaA, followed by IgG-conjugated 12-nm gold particles and 1% uranyl acetate prior to electron microscopy; scale bars of 0.5  $\mu m$ . (G) Cells of strains used in A to F, except B were grown to mid-log phase and subjected to cell fractionation. Protein samples collected from the culture medium (S) and cell wall (W) fractions were analyzed by immunoblotting with specific antibodies against SpaA ( $\alpha$ -SpaA).



**Figure S11: *In vitro* analysis of transpeptidation rates and proteolytic stability of  $^c\text{SpaA}$ - $^N\text{SpaA}$  interfacial mutants.** A) The  $^c\text{SpaA}$ - $^N\text{SpaA}$  dimer model is shown with the location of interfacial residues indicated. The FG loop and  $\alpha 1$  helix of  $^N\text{SpaA}$  are

colored green and cyan, respectively. B) Transpeptidation kinetics of <sup>N</sup>SpaA variants containing single amino acid mutations at the <sup>C</sup>SpaA-<sup>N</sup>SpaA interface were analyzed (L99R, P172R, D179R and L182R), as well as a K190A variant that removes the nucleophile. Product formation after 0, 3, 6 and 24 hours was quantified by SDS-PAGE and densitometry measurements. Reaction time-course and endpoint product formation (24 h) are shown in the left and right panels, respectively. C) Proteolytic degradation of the <sup>C</sup>SpaA-<sup>N</sup>SpaA and <sup>N</sup>SpaA-signal complexes. Two variants of the <sup>C</sup>SpaA-<sup>N</sup>SpaA complex are shown that contain either the wild-type (WT) <sup>N</sup>SpaA or <sup>N</sup>SpaA containing a L99D substitution. The <sup>C</sup>SpaA-<sup>N</sup>SpaA and <sup>N</sup>SpaA-signal complexes were subjected to limited proteolysis with trypsin. SDS-PAGE time points of the proteolysis reaction over 48h and comparison of endpoint percentages of intact complexes are shown in the left and right panels, respectively.



**Figure S12: Expanded HPLC kinetics data.** A) Steady state kinetics parameters are presented. All kinetics data are approximations because saturating conditions could not be achieved during to practical limitations<sup>6</sup>. Transpeptidation steady-state kinetic parameters were determined by monitoring the rate at which the enzyme ligated an FELPLTGGSG peptide to the <sup>15</sup>SpaA domain via a lysine–isopeptide bond. B) Representative HPLC chromatogram demonstrates that the production of <sup>15</sup>SpaA-LPLT complex can be measured by its increased retention time on a reversed phase HPLC gradient using a C4 column.

## SI References

1. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
2. Baljinnyam, B., Ronzetti, M., Yasgar, A. & Simeonov, A. Applications of Differential Scanning Fluorometry and Related Technologies in Characterization of Protein-Ligand Interactions. *Methods Mol. Biol.* **2089**, 47–68 (2020).
3. Wright, T. A., Stewart, J. M., Page, R. C. & Konkolewicz, D. Extraction of Thermodynamic Parameters of Protein Unfolding Using Parallelized Differential Scanning Fluorimetry. *J. Phys. Chem. Lett.* **8**, 553–558 (2017).
4. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, (2016).
5. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).
6. Sue, C. K. *et al.* Kinetics and Optimization of the Lysine-Isopeptide Bond Forming Sortase Enzyme from *Corynebacterium diphtheriae*. *Bioconjug. Chem.* **31**, 1624–1634 (2020).

## **Chapter 3**

### **Kinetics and Optimization of the Lysine-Isopeptide Bond-Forming Activity of the Pilin Sortase from *Corynebacterium diphtheriae***

### 3.1 Overview

The cell wall of many gram-positive pathogens such as *Corynebacterium diphtheriae*, *Streptococcus pyogenes*, *Actinomyces oris*, and *Streptococcus pneumoniae* has proteinaceous pili structures on the surface that mediate bacterial adhesion to host tissue. These gram-positive bacteria use a sortase transpeptidase enzyme to assemble these pili by linking protein subunits (pilins) together via lysine isopeptide bonds. Previous studies of the pilus assembly enzyme from *Corynebacterium diphtheriae* ( $C^d$ SrtA) found a lid motif that obscures the active site and is the reason for its inability to reconstitute the enzyme's *in vitro* activity. Two residues found in this lid motif, D81 and W83 were found to be interacting with the active site residues through electrostatic interactions and aromatic stacking. (PDB: 5K9A) Mutations of these residues to glycines resulted in the reconstitution of activity and the ability to study the enzyme *in vitro*. Previous *in vitro* studies of  $C^d$ SrtA involve the use of sodium dodecyl sulfate-polyacrylamide gel electrophoresis which monitored the reactions between pilin subunits but prevented accurate quantification of  $C^d$ SrtA's kinetics. To overcome this challenge, we developed an HPLC assay that separated and quantified the reaction catalyzed by  $C^d$ SrtA. This allowed the monitoring of the ligation reaction between a peptide that contained an LPLTG cell wall sorting sequence motif and the N-terminal domain of SpaA ( $^N$ SpaA) from *C. diphtheriae*. To gain a greater understanding of the purpose of  $C^d$ SrtA's lid motif, we deleted the lid motif from  $C^d$ SrtA to see if it would have any effect on its activity. We surprisingly found that this modified version of the  $C^d$ SrtA's lid, which we have named  $C^d$ SrtA $^{\Delta}$  is 7-fold more active than previously  $C^d$ SrtA enzymes. To understand why  $C^d$ SrtA $^{\Delta}$  is more active, we used liquid-chromatography mass spectrometry and found that the formation of the  $C^d$ SrtA-peptide acyl-intermediate was the slow step in the reaction. A comparison between  $C^d$ SrtA $^{\Delta}$  and other  $C^d$ SrtA enzymes found that  $C^d$ SrtA $^{\Delta}$  is faster at forming this acyl-intermediate and we postulate that this is the reason for its improved activity. My contributions to this paper were measuring the Michaelis-Menten Kinetics, testing the



hydrolysis activity of the enzyme, and monitoring acyl-intermediate formation through liquid chromatography-mass spectrometry.

This chapter is reformatted with permission from a peer-reviewed journal article: Kinetics and Optimization of the Lysine-Isopeptide Bond Forming Activity of the Pilin Sortase from *Corynebacterium diphtheriae*. Sue, C. K., McConnell, S.A., Ellis-Guardiola, K., Muroski, J., McAllister, R. A., Yu, J., Alvarez, A., Ogorzalek Loo, R. R., Loo, J. A., Ton-That, H. and Clubb, R. T. Bioconjugate Chemistry. May 2020.

### **3.2 Kinetics and Optimization of the Lysine-Isopeptide Bond Forming Activity of the Pilin Sortase from *Corynebacterium diphtheria***

## Kinetics and Optimization of the Lysine–Isopeptide Bond Forming Sortase Enzyme from *Corynebacterium diphtheriae*

Christopher K. Sue, Scott A. McConnell, Ken Ellis-Guardiola, John M. Muroski, Rachel A. McAllister, Justin Yu, Ana I. Alvarez, Chungyu Chang, Rachel R. Ogorzalek Loo, Joseph A. Loo, Hung Ton-That, and Robert T. Clubb\*



Cite This: *Bioconjugate Chem.* 2020, 31, 1624–1634



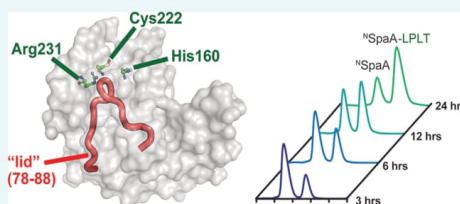
Read Online

ACCESS |

Metrics & More

Article Recommendations

**ABSTRACT:** Site-specifically modified protein bioconjugates have important applications in biology, chemistry, and medicine. Functionalizing specific protein side chains with enzymes using mild reaction conditions is of significant interest, but remains challenging. Recently, the lysine–isopeptide bond forming activity of the sortase enzyme that builds surface pili in *Corynebacterium diphtheriae* ( $C^d$ SrtA) has been reconstituted *in vitro*. A mutationally activated form of  $C^d$ SrtA was shown to be a promising bioconjugating enzyme that can attach Leu-Pro-Leu-Thr-Gly peptide fluorophores to a specific lysine residue within the N-terminal domain of the SpaA protein ( $^N$ SpaA), enabling the labeling of target proteins that are fused to  $^N$ SpaA. Here we present a detailed analysis of the  $C^d$ SrtA catalyzed protein labeling reaction. We show that the first step in catalysis is rate limiting, which is the formation of the  $C^d$ SrtA-peptide thioacyl intermediate that subsequently reacts with a lysine  $\epsilon$ -amine in  $^N$ SpaA. This intermediate is surprisingly stable, limiting spurious proteolysis of the peptide substrate. We report the discovery of a new enzyme variant ( $C^d$ SrtA $^{\Delta}$ ) that has significantly improved transpeptidation activity, because it completely lacks an inhibitory polypeptide appendage (“lid”) that normally masks the active site. We show that the presence of the lid primarily impairs formation of the thioacyl intermediate and not the recognition of the  $^N$ SpaA substrate. Quantitative measurements reveal that  $C^d$ SrtA $^{\Delta}$  generates its cross-linked product with a catalytic turnover number of  $1.4 \pm 0.004 \text{ h}^{-1}$  and that it has apparent  $K_M$  values of  $0.16 \pm 0.04$  and  $1.6 \pm 0.3 \text{ mM}$  for its  $^N$ SpaA and peptide substrates, respectively.  $C^d$ SrtA $^{\Delta}$  is 7-fold more active than previously studied variants, labeling >90% of  $^N$ SpaA with peptide within 6 h. The results of this study further improve the utility of  $C^d$ SrtA as a protein labeling tool and provide insight into the enzyme catalyzed reaction that underpins protein labeling and pilus biogenesis.



### INTRODUCTION

New methods are needed to create protein bioconjugates that can be used as therapeutics, imaging tools, diagnostic reagents, and materials.<sup>1–5</sup> Labeling specific sites on the protein is often preferred as it enables the construction of well-defined antibody–drug conjugates, small molecule- and fluorophore-labeled proteins for biophysical experiments, orientation-specific protein immobilization and the preparation of ordered, multifunctional protein complexes.<sup>6–8</sup> A variety of protein modification strategies have been developed that exhibit varying degrees of site-selectivity, efficiency, and ease of use. They range from chemical approaches that leverage the reactivity of amino acid specific functional groups (e.g., cysteine and lysine modifications) to highly selective, but less facile methods that require the incorporation of non-natural amino acids to facilitate bio-orthogonal conjugation chemistries (e.g., azide or alkyne-containing residues for click chemistry).<sup>9,10</sup> Bioconjugating enzymes (e.g., ligase, transferases, etc.) are particularly attractive for site-specific protein

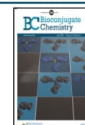
labeling, because they can be employed using mild reaction conditions, and in principle can be highly selective.<sup>11,12</sup> The sortase A enzyme from *Staphylococcus aureus* ( $S^a$ SrtA) is one of the most widely used bioconjugating enzymes.<sup>12–14</sup> It has been successfully deployed to catalyze protein–protein ligations and backbone cyclization, and to modify proteins with peptides, lipids, sugars, and small molecules.<sup>15–21</sup> However,  $S^a$ SrtA bioconjugations do not readily modify protein side chains and are almost exclusively restricted to altering only the N- or C-terminus of a protein.

Recently, we demonstrated that the pilus-specific sortase from *Corynebacterium diphtheriae* ( $C^d$ SrtA) can be used to

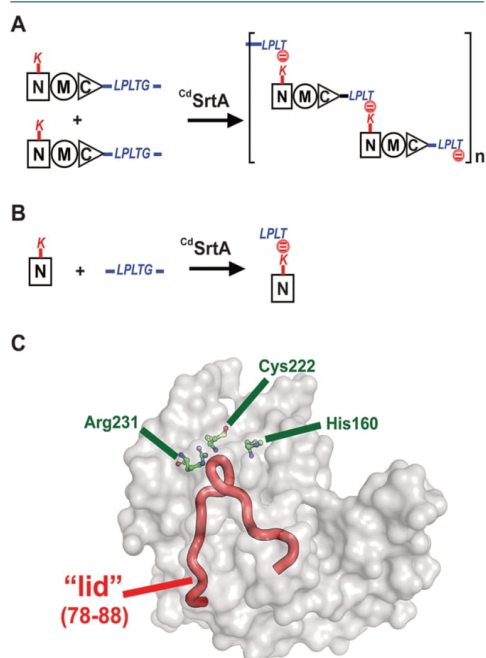
Received: March 25, 2020

Revised: May 7, 2020

Published: May 12, 2020



attach a peptide fluorophore via an isopeptide bond to a specific lysine residue within a protein.<sup>22</sup> Although  $C^d$ SrtA and  $S^d$ SrtA are both members of the sortase-superfamily of cysteine transpeptidases, they have distinct substrate specificities.<sup>23</sup>  $S^d$ SrtA is a class A sortase that catalyzes formation of backbone–backbone peptide bonds, whereas  $C^d$ SrtA is a class C sortase that joins molecules together via lysine–isopeptide bonds.<sup>24,25</sup>  $C^d$ SrtA assembles the SpaA pilus in *C. diphtheriae* by cross-linking SpaA “pilin” subunits via a lysine–isopeptide bond.<sup>26,27</sup> In this process, a lysine residue (K190) on one SpaA pilin is joined to the C-terminal LPLTG sorting signal located in a second SpaA pilin.<sup>26</sup> Repetition of this reaction forms the SpaA pilus, which is approximately 1–2  $\mu$ m in length and further elaborated with unique tip and basal pilin proteins (Figure 1A).<sup>27,28</sup> *In vitro*, the native  $C^d$ SrtA enzyme is enzymatically inactive because it contains a polypeptide appendage that occludes its active site, called a “lid” (Figure 1C). However,  $C^d$ SrtA variants containing destabilizing amino acid substitutions in the lid exhibit *in vitro* activity.<sup>28</sup> The most



**Figure 1.** *C. diphtheriae*  $C^d$ SrtA pilin sortase catalyzes lysine isopeptide bond formation. (A) Schematic showing the pilin polymerization reaction catalyzed by  $C^d$ SrtA. The enzyme creates the SpaA pilus by polymerizing SpaA pilin proteins. In the reaction, it recognizes lysine (K190) side chain nucleophile within the N-terminal domain of SpaA ( $N^d$ SpaA), and it joins the backbone threonine carbonyl carbon atom located in the C-terminal LPLTG sorting signal located within another SpaA protein. This reaction is repeated to construct the SpaA pilus that mediates bacterial adhesion. (B) Schematic of the reaction used to monitor lysine isopeptide bond formation. In this assay, the  $C^d$ SrtA enzyme ligates the isolated  $N^d$ SpaA domain to the peptide containing the LPLTG sorting signal (FELPLTGGSG). (C) Structure of  $C^d$ SrtA showing H160, C222, and R231 active site residues. The “Lid” is highlighted in red (residues P77 to S89).

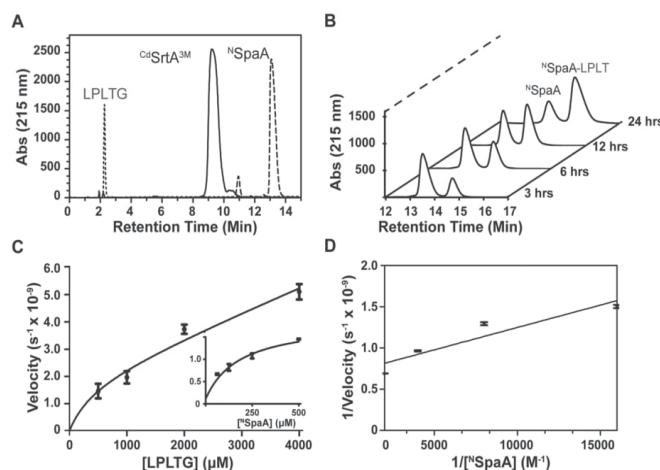
active form of the enzyme thus far discovered is  $C^d$ SrtA<sup>3M</sup>, which contains residues N37–Q257 in  $C^d$ SrtA and D81G/W83G/N85A substitutions in the lid.<sup>22</sup>  $C^d$ SrtA<sup>3M</sup> is a promising bioconjugation tool, as it can be used to selectively modify proteins harboring the N-terminal domain of SpaA ( $N^d$ SpaA) with peptide fluorophores. Modification via lysine–isopeptide bonds is attractive, as these linkages may be less susceptible to proteolysis and enzymatic reversibility.

In this study, we developed an HPLC-based assay to measure for the first time the kinetics of catalysis, and we have used the assay to identify a new  $C^d$ SrtA variant that has improved bioconjugation activity. In particular, we show that (i) the bioconjugation reaction rate is limited by the formation of an enzyme–acyl intermediate with the LPLTG sorting signal, (ii) the enzyme preferentially recognizes nonpolar amino acids at the X position within the sorting signal, (iii) unlike  $S^d$ SrtA,  $C^d$ SrtA exhibits minimal proteolytic activity, (iv) amino acid substitutions introduced into the lid accelerate catalysis by facilitating enzyme–acyl intermediate formation, and (v) complete removal of the lid further activates the enzyme. These results increase the *in vitro* utility of  $C^d$ SrtA as a bioconjugation tool to modify proteins and provides new insight into the enzymatic reaction that underpins the construction of surface pili in Gram-positive bacteria.

## RESULTS AND DISCUSSION

### Kinetics of Lysine–Isopeptide Bond Formation.

Previously, we monitored the lysine–isopeptide bond forming activity of  $C^d$ SrtA<sup>3M</sup> using SDS-PAGE.<sup>22</sup> However, the kinetics of this process could not be accurately determined because the reactants and products were difficult to separate and quantify. To overcome this problem, we developed an HPLC-based transpeptidation assay that monitors the ability of  $C^d$ SrtA<sup>3M</sup> to join the N-terminal domain from SpaA ( $N^d$ SpaA, residues E30–S195 of SpaA) to a peptide containing its C-terminal sorting signal (FELPLTGGSG, hereafter called LPLTG peptide). This reaction represents an isolated ligation event in the polymerization reaction by producing a  $N^d$ SpaA–LPLT product in which the K190 side chain in  $N^d$ SpaA is joined via an isopeptide bond to the threonine carbonyl group in the FELPLT peptide (Figure 1B). The reactants and products are readily separated by reverse-phase HPLC (Figure 2A). Moreover, this procedure enables facile monitoring of the time-dependent conversion of the protein substrate ( $N^d$ SpaA) into the cross-linked protein–peptide product ( $N^d$ SpaA–LPLT) (Figure 2B). The identity of the product and the location of its isopeptide linkage were previously confirmed by LC-MS/MS.<sup>22,28</sup> Initially, for each substrate ( $N^d$ SpaA and the LPLTG peptide) the dependence of the reaction velocity on substrate concentration was determined (Figure 2C). This analysis reveals that  $N^d$ SpaA and peptide substrates do not saturate the enzyme even when they are present at concentrations of 500  $\mu$ M (Figure 2C, insert) and 4 mM (Figure 2C, main), respectively. Because using higher concentrations of each substrate is not practical, we determined apparent steady-state parameters using subsaturating amounts of each substrate (representative data is shown in Figure 2D). Two sets of Michaelis–Menten parameters were obtained. Either the concentration of  $N^d$ SpaA was varied from 62.5 to 500  $\mu$ M with the amount of LPLTG peptide held constant at 1 mM, or the concentration of the LPLTG peptide was varied from 0.5 to 4 mM, while the concentration of  $N^d$ SpaA was held fixed at 500  $\mu$ M. These measurements revealed that  $C^d$ SrtA<sup>3M</sup> catalyzes isopeptide



**Figure 2.**  $C^d$ SrtA transpeptidation assay: (A) Superimposed reversed-phase HPLC traces showing the separation of the sorting signal peptide (LPLTG, short dashes),  $C^d$ SrtA<sup>3M</sup> (solid line), and  $N^S$ paA (long dashes). (B) Representative HPLC traces that track the progress of the reaction. Peaks corresponding to  $N^S$ paA and  $N^S$ paA-LPLT are shown. The reactions were sampled at 3, 6, 12, and 24 h, and contained 100  $\mu$ M of enzyme, 200  $\mu$ M of  $N^S$ paA, and 1 mM of LPLTG-peptide. (C) Plots showing the measured velocity versus substrate concentration for the LPLTG peptide (the sorting signal) and  $N^S$ paA (inset). A concentration range of 500  $\mu$ M to 4 mM, and 62.5  $\mu$ M to 500  $\mu$ M were used for the LPLTG peptide and  $N^S$ paA, respectively. Initial velocities were determined after 3 h, as described in the Methods section. (D) Lineweaver–Burk graph graphing showing kinetics data for  $C^d$ SrtA<sup>3M</sup>. The  $k_{cat}$  and  $K_M$  values were determined from a linear fit of this data.

**Table 1.** Kinetics of  $C^d$ SrtA Catalyzed Lysine–Isopeptide Formation<sup>a</sup>

|                          | $k_{cat} \times 10^{-5}$ (s <sup>-1</sup> ) <sup>b</sup> | $^N K_M \times 10^{-4}$ (M) | $^S K_M \times 10^{-4}$ (M) | $k_{cat}/^N K_M$ (s <sup>-1</sup> M <sup>-1</sup> ) |
|--------------------------|--|-----------------------------|-----------------------------|---|
| $C^d$ SrtA               | n.d. <sup>c</sup>  | n.d.                        | n.d.                        | n.d.  |
| $C^d$ SrtA <sup>3M</sup> | 5.6 ± 0.8  | 0.7 ± 0.1                   | 20 ± 10                     | 0.7 ± 0.1   |
| H160A                    | 3.1 ± 0.4  | 0.43 ± 0.05                 | —                           | 0.72 ± 0.08   |
| C222A                    | n.d.   | n.d.                        | —                           | n.d.  |
| R231A                    | n.d.   | n.d.                        | —                           | n.d.  |
| $C^d$ SrtA <sup>Δ</sup>  | 40 ± 0.1   | 1.6 ± 0.4                   | 16 ± 3                      | 2.5 ± 0.6   |
| H160A                    | 2.5 ± 0.6  | 0.70 ± 0.02                 | —                           | 0.36 ± 0.09   |
| C222A                    | n.d.   | n.d.                        | —                           | n.d.  |
| R231A                    | n.d.   | n.d.                        | —                           | n.d.  |
| $S^d$ SrtA <sup>d</sup>  | 1600 ± 100   | 1.8 ± 0.1                   | 73.3 ± 10.1 <sup>e</sup>    | 86 ± 5  |

<sup>a</sup>All kinetics are approximations as saturating concentrations were not able to be measured. <sup>b</sup>Transpeptidation steady-state kinetic parameters for  $C^d$ SrtA were determined by the monitoring rate at which the enzyme ligated an FELPLTGGSG peptide to the  $N^S$ paA domain via a lysine–isopeptide bond. <sup>c</sup>n.d., not determined because an insufficient amount of product was detectable. <sup>d</sup>Transpeptidation steady-state kinetic parameters for  $S^d$ SrtA were determined by monitoring the rate at which the enzyme ligated GGG and FELPLTGGSG peptides via a backbone peptide bond. Reported values are the average from three measurements, and the error is the standard deviation. <sup>e</sup>Values are reported from Frankel et al. (2005) and measures reactions between an Abz-LPETG-Dap(Dnp)-NH<sub>2</sub> and pentaglycine.<sup>29</sup> <sup>N</sup>Refers to transpeptidation kinetics measure when  $N^S$ paA is varied and FELPLTGGSG concentration is held constant. <sup>S</sup>Refers to when FELPLTGGSG peptide is varied and  $N^S$ paA is held constant.

bond formation with apparent  $K_M$  values for  $N^S$ paA ( $^N K_M$ ) and the LPLTG peptide ( $^S K_M$ ) of 70 ± 10  $\mu$ M and 2 ± 1 mM, respectively (Table 1). Similar turnover numbers were measured in each experiment, with a maximal value of 26 ± 10 ( $\times 10^{-5}$  s) (obtained when  $N^S$ paA is held constant at 500  $\mu$ M and the LPLTG peptide is varied).

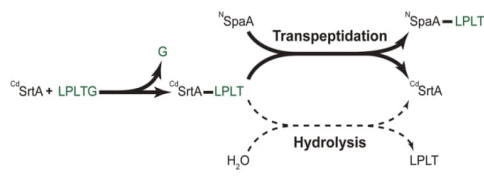
**Formation of the  $C^d$ SrtA-LPLT Thioacyl Intermediate Is a Rate Limiting Step in Catalysis and Can Be Accelerated by Completely Removing the Lid.**  $C^d$ SrtA<sup>3M</sup> is the most active form of  $C^d$ SrtA thus far discovered and contains three substitutions in the inhibitory lid structure (D81G, W83G, N85A).<sup>22</sup> We reasoned that an enzyme variant with the entire lid deleted might exhibit even higher transpeptidation activity by completely unmasking the active

site. Inspection of the crystal structure  $C^d$ SrtA suggests that lid residues I78 to A88 can be deleted without disrupting its tertiary structure, as the remaining P77 and S89 amino acids are positioned adjacent to one another in three-dimensional space.<sup>28</sup> Indeed, the measured steady-state kinetic parameters for a lid-deletion  $C^d$ SrtA<sup>Δ</sup> (residues N37–Q257 of  $C^d$ SrtA in which the amino acids I78 to A88 are deleted) reveal that it is more active than  $C^d$ SrtA<sup>3M</sup>; there is a 7-fold improvement in the apparent  $k_{cat}$  while the  $^N K_M$  and  $^S K_M$  values are only modestly affected (Table 1).

To address why  $C^d$ SrtA<sup>Δ</sup> is catalytically more active than  $C^d$ SrtA<sup>3M</sup>, we investigated how lid removal affected catalysis. By analogy to the prototypical  $S^d$ SrtA enzyme,  $C^d$ SrtA presumably catalyzes lysine–isopeptide bond formation via a two-step

process (Scheme 1).<sup>23,29–31</sup> In the transpeptidation mechanism,  $\text{C}^{\text{d}}\text{SrtA}$ 's C222 thiol presumably functions as a

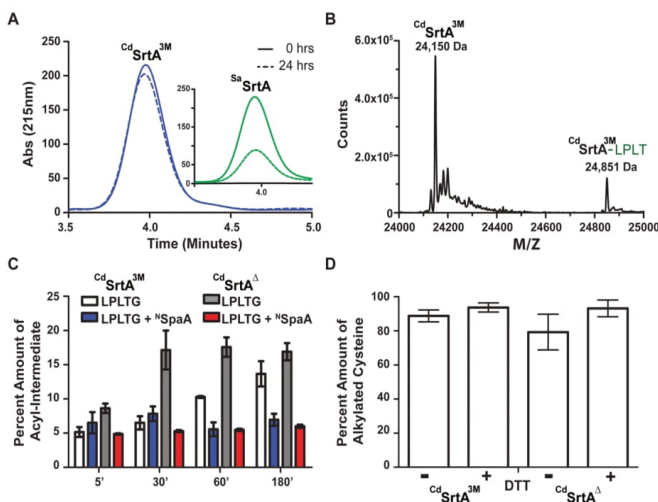
**Scheme 1. Schematic Showing the Overall Mechanism of  $\text{C}^{\text{d}}\text{SrtA}$ -Catalyzed Isopeptide Formation (Top) and a Potential Hydrolytic Side Reaction (Bottom)**



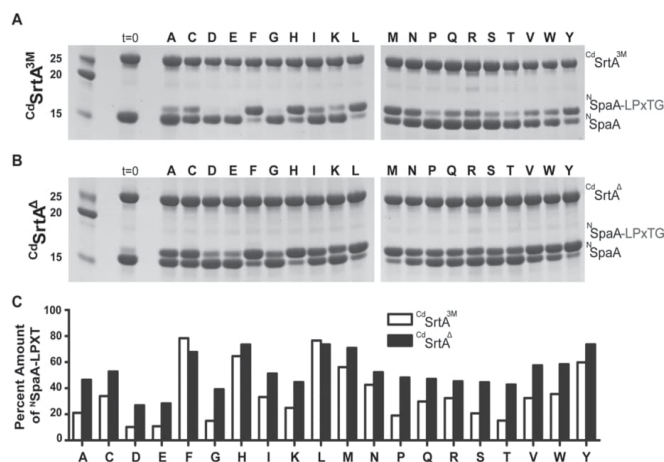
nucleophile to attack the threonine carbonyl carbon in the LPLTG sorting signal to form a  $\text{C}^{\text{d}}\text{SrtA}$ -LPLT thioacyl intermediate. Then, the enzyme recognizes the K190 side chain amine group located within  $\text{N}^{\text{SpaA}}$ , which resolves the thioacyl intermediate to form the lysine–isopeptide linked  $\text{N}^{\text{SpaA}}$ -LPLT product. In  $\text{S}^{\text{d}}\text{SrtA}$ , a side reaction also occurs in which a water molecule attacks the thioacyl intermediate instead of a primary amine, resulting in the hydrolysis of the intermediate to release the LPLT peptide.<sup>29</sup> In this side reaction, the enzyme effectively functions as a protease, cleaving the LPLTG peptide substrate at the peptide bond that joins the threonine and glycine residues.  $\text{C}^{\text{d}}\text{SrtA}$  has also been shown to proteolyze its LPLTG peptide substrate, but the kinetics and extent of proteolysis have not been rigorously measured.<sup>1,22,28</sup>

To determine if differences in the rate of the hydrolytic side reaction cause the  $\text{C}^{\text{d}}\text{SrtA}^{\Delta}$  and  $\text{C}^{\text{d}}\text{SrtA}^{3\text{M}}$  enzymes to produce differing amounts of transpeptidation product, we measured the ability of each enzyme to proteolyze the LPLTG peptide substrate using reversed-phase HPLC. In these reactions, only the enzyme and LPLTG peptide are present. Interestingly, even though  $\text{C}^{\text{d}}\text{SrtA}^{\Delta}$  and  $\text{C}^{\text{d}}\text{SrtA}^{3\text{M}}$  catalyze transpeptidation at 25 °C, the rate at which the hydrolytic side reaction occurs at this temperature is very slow for both enzymes, with less than 5% of the LPLTG peptide substrate consumed by the enzyme (Figure 3A, left). In fact, because this assay employed excess amounts of peptide relative to enzyme (50  $\mu\text{M}$  enzyme and 500  $\mu\text{M}$  peptide), much of the observed peptide consumption can be attributed to formation of the thioacyl intermediate and not to repeated rounds of proteolysis. This finding is in marked contrast to the archetypal  $\text{S}^{\text{d}}\text{SrtA}$  enzyme, which proteolyzes more than 60% of its sorting signal peptide substrate when identical reaction conditions are used (Figure 3A, right). Thus, we conclude that for both the  $\text{C}^{\text{d}}\text{SrtA}^{\Delta}$  and  $\text{C}^{\text{d}}\text{SrtA}^{3\text{M}}$  enzymes the transpeptidation pathway is dominant and the hydrolytic side reaction occurs only to a minor extent.

We wondered whether the superior transpeptidation activity of  $\text{C}^{\text{d}}\text{SrtA}^{\Delta}$  originated from its ability to form the enzyme–LPLT thioacyl intermediate more rapidly than  $\text{C}^{\text{d}}\text{SrtA}^{3\text{M}}$  (Scheme 1). For both the  $\text{C}^{\text{d}}\text{SrtA}^{\Delta}$  and  $\text{C}^{\text{d}}\text{SrtA}^{3\text{M}}$  enzymes, this intermediate is readily observable in LC-MS mass spectra when they are incubated with the LPLTG peptide (Figure 3B). This finding is consistent with the low proteolytic activity of each enzyme and suggests that the two steps of catalysis are independent of one another—each enzyme can form and



**Figure 3.** Characterization of  $\text{C}^{\text{d}}\text{SrtA}^{3\text{M}}$  and  $\text{C}^{\text{d}}\text{SrtA}^{\Delta}$ . (A) Representative reverse phase HPLC trace showing the change in the concentration of LPLTG peptide due to proteolysis by  $\text{C}^{\text{d}}\text{SrtA}^{3\text{M}}$  (blue) and  $\text{S}^{\text{d}}\text{SrtA}^{\text{WT}}$  (green) at 25 °C after 0 (solid lines) and 24 h (dashed lines). (B) Mass deconvolution of LC-MS data of  $\text{C}^{\text{d}}\text{SrtA}^{3\text{M}}$ . The acyl-intermediate at 24,851 Da is approximately 700 Da higher than where the enzyme is at 24,150 Da. (C) Comparison of the amounts of acyl-intermediate in the presence and absence of  $\text{N}^{\text{SpaA}}$  measured over a period of three hours.  $\text{C}^{\text{d}}\text{SrtA}^{3\text{M}}$  without  $\text{N}^{\text{SpaA}}$  (white) in comparison to  $\text{C}^{\text{d}}\text{SrtA}^{\Delta}$  without  $\text{N}^{\text{SpaA}}$  (gray) shows the faster formation of the  $\text{C}^{\text{d}}\text{SrtA}^{\Delta}$  acyl-intermediate. In contrast, both enzymes form these intermediates at similar rates when  $\text{N}^{\text{SpaA}}$  is present ( $\text{C}^{\text{d}}\text{SrtA}^{3\text{M}}$  (blue) and  $\text{C}^{\text{d}}\text{SrtA}^{\Delta}$  (red)). (D) Data showing the amount of reduced active site C222 thiol in the  $\text{C}^{\text{d}}\text{SrtA}^{3\text{M}}$  and  $\text{C}^{\text{d}}\text{SrtA}^{\Delta}$  enzymes. Freshly produced samples of each enzyme were either treated with an excess amount of DTT or a buffer control for one hour. The proteins were then digested with trypsin, reacted with iodoacetamide, and the extent of cysteine alkylation determined by mass spectrometry. The experiments were performed in triplicate.



**Figure 4.** Role of the X residue in transpeptidation. (A) SDS-PAGE analysis of product formation when  $CdSrtA^{3M}$  was reacted with  $NSpaA$ , and a series of LPXTG peptides in which the identity of the amino acid at the X position was systematically varied. Reactions (200  $\mu$ M enzyme, 200  $\mu$ M  $NSpaA$ , 5 mM DTT, and 1 mM Peptide) were measured after 24 h. The most reactive sorting signal peptides contained nonpolar X residues, while those containing polar residues were less reactive. (B) As in panel (A), but  $CdSrtA^{\Delta}$  was used instead of  $CdSrtA^{3M}$ . Similar trends in activity are observed. (C) Histogram plot showing the amount of product created for each peptide in the library. The fraction of  $NSpaA$  converted to  $NSpaA-LPxTG$  is shown. The data was obtained by analyzing the SDS-PAGE data (panels A and B) using the program ImageJ.

maintain the enzyme–acyl intermediate in the absence of the  $NSpaA$  nucleophile. Moreover, it is compatible with our previously published cellular studies of class C sortases in which long-lived and stable acyl–enzyme intermediates were established to be important for catalysis.<sup>32</sup> To determine if  $CdSrtA^{\Delta}$  and  $CdSrtA^{3M}$  differed in their ability to perform the first step of catalysis, we tracked thioacyl intermediate formation at various times after mixing the enzymes with the LPLTG peptide (25  $\mu$ M and 1 mM of enzyme and LPLTG peptide, respectively). The intermediate formation was followed over a 3 h period; the same duration was used to measure transpeptidation activity. This analysis revealed that  $CdSrtA^{\Delta}$  forms the thioacyl intermediate more rapidly than  $CdSrtA^{3M}$ , with  $\sim 17 \pm 2\%$  of  $CdSrtA^{\Delta}$  joined to the peptide after 30 min, while it takes up to 3 h for  $CdSrtA^{3M}$  to form similar levels of this reaction intermediate (Figure 3C). Interestingly, when similar time course experiments are performed in the presence of both  $NSpaA$  and LPLTG substrates, significantly lower amounts of thioacyl intermediate are observed for each type of enzyme ( $\sim 5\%$  for both  $CdSrtA^{\Delta}$  and  $CdSrtA^{3M}$ ). As the reaction conditions are identical to those used to measure transpeptidation activity, this data suggests that the formation of the acyl-intermediate is rate-limiting. Thus, we conclude that improved transpeptidation activity of  $CdSrtA^{\Delta}$  results from its ability to form the acyl-intermediate at a faster rate than  $CdSrtA^{3M}$ .

**The Identity of the “X” Residue within the LPXTG Sorting Signal Affects the Rate of Transpeptidation.** Structural and computational studies of class A and B sortases bound to their respective sorting signals have revealed that they do not recognize the side chain of the X residue within their respective LPXTG-type sorting signal substrates because it projects away from the enzyme into the solvent.<sup>23,33,34</sup> The structures are consistent with detailed substrate specificity analyses of the class A  $SaSrtA$  enzyme, which revealed that

LPXTG sorting signals containing any amino acid at the X position can be used as substrates.<sup>35</sup> However,  $CdSrtA$  and other class C sortases are unique, because they contain a lid structure whose proximity to the active site could affect recognition of the LPXTG sorting signal (Figure 1C).<sup>33</sup> Indeed, our prior studies of  $CdSrtA^{3M}$  revealed that a leucine to alanine substitution at the X position of the LPLTG sorting signal slowed transpeptidation.<sup>22</sup> To investigate this issue in greater detail, the  $CdSrtA^{\Delta}$  and  $CdSrtA^{3M}$  enzymes were tested for their ability to utilize as transpeptidation substrates 20 distinct LPXTG peptides in which the X position was varied. For these studies, the  $CdSrtA^{\Delta}$  and  $CdSrtA^{3M}$  enzymes were incubated with  $NSpaA$  and each member of the peptide library, and the amount of cross-linked product was then determined by SDS-PAGE (Figure 4A,B). Significant variation in reactivity is observed for the different library members. However, in nearly all cases,  $CdSrtA^{\Delta}$  is more active than  $CdSrtA^{3M}$ , consistent with the steady-state kinetic measurements that employed the LPLTG peptide (Table 1). Interestingly, both enzymes exhibit similar sequence preferences. In particular, they preferentially use sorting signals containing phenylalanine, histidine, methionine, tyrosine, and leucine at the X position, while their least reactive substrates contain negatively charged side chains at this site. In all cases, peptides containing leucine at the X position are very reactive, explaining why this amino acid is present in the native LPLTG substrate present in  $SpaA$ . The X position-dependent activity of  $CdSrtA$  is distinct from what has been observed for  $SaSrtA$ , as similar peptide library studies have shown that after 24 h exposure all peptides in the library are processed to a similar extent by  $SaSrtA$ .<sup>35</sup> The molecular basis underlying the observed variation in peptide reactivity is not known, but it is not caused by the presence of the lid as similar trends in activity are observed for  $CdSrtA^{3M}$  and  $CdSrtA^{\Delta}$ .

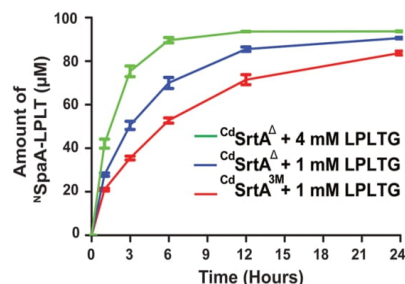
**Cysteine and Arginine Active Site Residues Are Required for Catalysis.** Based on sequence homology with

the well-studied  $^{54}\text{SrtA}$  sortase, three conserved residues in  $^{\text{Cd}}\text{SrtA}$  are presumably required for catalysis: H160, C222, and R231 (Figure 1B).<sup>36,37</sup> However, their role in catalyzing lysine–isopeptide bond transpeptidation *in vitro* has not been determined. We therefore used the HPLC-assay to measure the transpeptidation activities of  $^{\text{Cd}}\text{SrtA}^{3\text{M}}$  and  $^{\text{Cd}}\text{SrtA}^{\Delta}$  enzymes containing alanine substitutions at these sites. For both enzyme variants, C222A and R231A substitutions completely abrogate transpeptidation activity, demonstrating that they have critical functions in catalysis (Table 1). However,  $^{\text{Cd}}\text{SrtA}^{3\text{M}}$  and  $^{\text{Cd}}\text{SrtA}^{\Delta}$  enzymes harboring H160A substitutions retain some enzyme activity, exhibiting similar turnover numbers that are reduced by 44% and 94% as compared to their native forms, respectively. In the well-studied  $^{54}\text{SrtA}$  transpeptidation reaction the analogous histidine, cysteine, and arginine residues are essential for catalysis *in vitro* (R197, C184, and H120 in the  $^{54}\text{SrtA}$  sequence).<sup>36</sup> The cysteine thiol functions as a nucleophile, while the arginine guanidinium group (R197 in  $^{54}\text{SrtA}$ ) has been proposed to facilitate catalysis by stabilizing oxyanion tetrahedral intermediates.<sup>23</sup> Our finding that the C222A and R231A sortase variants are completely inactive is consistent with these residues having similar functions.<sup>28</sup> It is also compatible with pH dependence of the transpeptidation reaction, which occurs most rapidly at pH values between 7.5 and 8. We were surprised that  $^{\text{Cd}}\text{SrtA}$  containing a H160A substitution retained some activity, since the analogous alteration in  $^{54}\text{SrtA}$  disrupts transpeptidation *in vitro*.<sup>36</sup> In  $^{54}\text{SrtA}$ , the histidine side chain has been proposed to function as a general acid and base, protonating the amine of the leaving group glycine residue as the threonine–glycine peptide bond in the LPLTG sorting signal is broken, and facilitating the last step of catalysis by deprotonating the incoming amine nucleophile that resolves the thioacyl intermediate to produce the final transpeptidation product (Scheme 1).<sup>29,38</sup> Whole cell studies have shown that  $^{\text{Cd}}\text{SrtA}$  containing H160A are incapable of assembling surface pili.<sup>28</sup> The residual activity observed in the H160A enzyme suggests that residues in addition to H160 in  $^{\text{Cd}}\text{SrtA}$  may facilitate these steps *in vitro*, albeit less efficiently.

We wondered whether the superior activity of  $^{\text{Cd}}\text{SrtA}^{\Delta}$  relative to  $^{\text{Cd}}\text{SrtA}^{3\text{M}}$  could be attributed to differences in the oxidation state of the C222 sulfhydryl group that is influenced by the presence of the lid. This is because the sulfhydryl group can in principle become oxidized to unreactive disulfide, sulfenic, sulfinic, and sulfonic forms that are non-nucleophilic.<sup>39</sup> No significant differences in disulfide formation were observed for the two proteins in nonreducing SDS-PAGE and MALDI experiments. To investigate whether they formed more oxidized states, freshly purified  $^{\text{Cd}}\text{SrtA}^{\Delta}$  and  $^{\text{Cd}}\text{SrtA}^{3\text{M}}$  enzymes were incubated for one hour in buffer A (50 mM Tris-HCl, 300 mM NaCl, at a pH of 8.0) with, and without, a reducing agent (5 mM DTT). The amount of C222 (the protein's only cysteine) present in the active thiol form was then determined by adding iodoacetamide, digesting with trypsin, and analyzing by mass spectrometry. The percent of reduced thiol was calculated by measuring the amount of C222 in each enzyme that was derivatized with carboxyamidomethyl as compared to the total amount of C222 in its various oxidation states. In the absence of reducing agent,  $89 \pm 4$  and  $79 \pm 10\%$  of  $^{\text{Cd}}\text{SrtA}^{3\text{M}}$  and  $^{\text{Cd}}\text{SrtA}^{\Delta}$  contain a reactive thiol, respectively. Moreover, only small increases in the amount of reactive thiol are observed when the DTT reducing agent is

present; under these conditions,  $94 \pm 3\%$  and  $93 \pm 5\%$   $^{\text{Cd}}\text{SrtA}^{3\text{M}}$  and  $^{\text{Cd}}\text{SrtA}^{\Delta}$  are reactive, respectively (Figure 3D). Thus, the C222 sulfhydryl in freshly purified  $^{\text{Cd}}\text{SrtA}^{3\text{M}}$  and  $^{\text{Cd}}\text{SrtA}^{\Delta}$  primarily exists in a reduced, transpeptidation competent state. A caveat is that ionization efficiencies may differ for alkylated and nonalkylated forms of a cysteine-containing peptide. Nevertheless, the similarity between the  $+/-\text{DTT}$  values implies that C222 is primarily reduced. It should be noted that we have found that C222 can become oxidized and unreactive if the enzymes are stored for more than several weeks, and their activity can be restored by incubating them with DTT. Therefore, including DTT in the labeling reactions is recommended as a precaution.

**Improved Protein Lysine Labeling Using  $^{\text{Cd}}\text{SrtA}^{\Delta}$ .** Having defined substrate and reaction conditions that are optimal for activity, we directly compared the peptide labeling efficiencies of  $^{\text{Cd}}\text{SrtA}^{3\text{M}}$  with the newer  $^{\text{Cd}}\text{SrtA}^{\Delta}$  enzyme. Consistent with our steady-state kinetic analyses, a temporal analysis using identical conditions clearly shows that  $^{\text{Cd}}\text{SrtA}^{\Delta}$  (blue trace) produces more cross-linked product than



**Figure 5.** Labeling activity of  $^{\text{Cd}}\text{SrtA}^{3\text{M}}$  and  $^{\text{Cd}}\text{SrtA}^{\Delta}$ . Reactions containing each enzyme were performed using identical conditions: 100  $\mu\text{M}$  enzyme, 100  $\mu\text{M}$   $^{\text{N}}\text{SpaA}$ , and either 1 or 4 mM LPLTG peptide.  $^{\text{Cd}}\text{SrtA}^{\Delta}$  exhibits superior labeling activity at all measured time points. The experiments were performed in triplicate.

$^{\text{Cd}}\text{SrtA}^{3\text{M}}$  (red trace) (Figure 5) (100  $\mu\text{M}$  enzyme, 100  $\mu\text{M}$   $^{\text{N}}\text{SpaA}$  and 1 mM LPLTG peptide at 25 °C). In particular, when  $^{\text{Cd}}\text{SrtA}^{\Delta}$  is used,  $\sim 90\%$  of  $^{\text{N}}\text{SpaA}$  is modified with peptide within 12 h, a higher amount than is achieved with  $^{\text{Cd}}\text{SrtA}^{3\text{M}}$  even after 24 h. Notably, even faster labeling can be achieved using  $^{\text{Cd}}\text{SrtA}^{\Delta}$  by increasing the concentration of LPLTG peptide in the reaction to 4 mM, which enables  $90 \pm 2\%$  of  $^{\text{N}}\text{SpaA}$  to be modified within 6 h (Figure 5, green trace). This finding is compatible with the relatively high  $^{\text{S}}K_{\text{M}}$  ( $2 \pm 1$  mM), which makes it challenging to saturate the enzyme with sorting signal substrate (Table 1). As  $^{\text{Cd}}\text{SrtA}^{3\text{M}}$  has previously been shown to be capable of labeling  $^{\text{N}}\text{SpaA}$  with a peptide fluorophore,  $^{\text{Cd}}\text{SrtA}^{\Delta}$  can be expected to more active in these labeling reactions as well.<sup>22</sup> Similar comparisons were performed at 37 °C instead of 25 °C.  $^{\text{Cd}}\text{SrtA}^{\Delta}$  becomes less active at elevated temperatures because it is less thermostable than  $^{\text{Cd}}\text{SrtA}^{3\text{M}}$ . However, both the rate and yield of product formed by  $^{\text{Cd}}\text{SrtA}^{\Delta}$  at 25 °C are superior to what  $^{\text{Cd}}\text{SrtA}^{3\text{M}}$  produces at 37 °C. Thus,  $^{\text{Cd}}\text{SrtA}^{\Delta}$  should be used for labeling reactions that are performed at room temperature.

**Comparison with the Prototypical  $^{54}\text{SrtA}$  Sortase.**  $^{54}\text{SrtA}$  is the best-studied member of the sortase superfamily

and is routinely used to modify the N- and C-termini of proteins.<sup>13,23</sup>  $^{135}\text{Cd}$ SrtA and  $^{68}\text{Zn}$ SrtA catalyze mechanistically related transpeptidation reactions, but differ in the type of nucleophile that they use, i.e., a lysine  $\epsilon$ -amine within  $^{\text{N}}\text{SpaA}$  and the N-terminal amine group within an oligoglycine peptide, respectively. In order to directly compare their activities, an established HPLC assay was used to measure the rate at which  $^{68}\text{Zn}$ SrtA joins the LPLTG peptide previously used in studies of  $^{135}\text{Cd}$ SrtA to its triglycine substrate ( $\text{Gly}_3$ ), a reaction that creates a FELPTGGGG peptide product (Table 1).<sup>29,40</sup> The steady-state kinetic values measured from this analysis are generally consistent with published values that used a much shorter fluorogenic labeled peptide.<sup>29</sup> Our direct comparison reveals that  $^{68}\text{Zn}$ SrtA forms backbone peptide bonds  $\sim 40$ -times faster than  $^{135}\text{Cd}$ SrtA $^{\Delta}$  creates lysine–isopeptide bonds (Table 1). Interestingly, both enzymes exhibit generally similar  $K_M$  values for their substrates, with each exhibiting millimolar  $K_M$  values for the sorting signal substrate ( $^{135}\text{Cd}$ SrtA $^{\Delta}$ ,  $1.6 \pm 0.3$  mM;  $^{68}\text{Zn}$ SrtA,  $7.33 \pm 1.01$  mM), and approximately  $\sim 10$ -fold lower  $K_M$  values for their respective amine nucleophiles ( $^{135}\text{Cd}$ SrtA $^{\Delta}$ ,  $160 \pm 30$   $\mu\text{M}$ ;  $^{68}\text{Zn}$ SrtA,  $180 \pm 10$   $\mu\text{M}$ ).<sup>29</sup> In addition, based on prior studies by the McCafferty group and the studies presented here (Figure 3), for both  $^{68}\text{Zn}$ SrtA and  $^{135}\text{Cd}$ SrtA the first step of catalysis is rate limiting—the formation of the thioacyl intermediate (Scheme 1).<sup>29,40</sup> It is unclear why this step is so slow in both enzymes. However, it has been shown that  $^{68}\text{Zn}$ SrtA primarily exists in an inactive form in which only a small fraction (ca. 0.06%) of enzyme contains catalytically capable Cys184 thiolate and His120 imidazolium forms that are capable of reacting with the sorting signal to form the thioacyl intermediate.<sup>30,41,42</sup> Whether the active site of  $^{135}\text{Cd}$ SrtA also primarily exists in a catalytically dormant state remains to be determined. Startlingly, we found that  $^{135}\text{Cd}$ SrtA is quite inefficient at proteolyzing the LPLTG substrate, which is in stark contrast to previously reported studies of  $^{68}\text{Zn}$ SrtA that have shown that it catalyzes this reaction with a  $k_{\text{cat}}$  of  $0.28 \pm 0.02$   $\text{s}^{-1}$  (Figure 3A).<sup>29</sup> Presumably this difference originates from distinct active site features that enable  $^{68}\text{Zn}$ SrtA to use water as a nucleophile much more efficiently than  $^{135}\text{Cd}$ SrtA (Scheme 1). This idea is substantiated by our observation that the thioacyl intermediate in the  $^{135}\text{Cd}$ SrtA reaction forms to an appreciable extent in the absence of the nucleophile (Figure 3C). The reduced proteolytic activity of  $^{135}\text{Cd}$ SrtA is presumably advantageous, limiting the release of partially assembled pili from the bacterial cell surface. Similar to  $^{135}\text{Cd}$ SrtA, under certain conditions  $^{68}\text{Zn}$ SrtA can catalyze formation of a lysine isopeptide bond.<sup>43–45</sup> A thorough study by Dasgupta and colleagues demonstrated that  $^{68}\text{Zn}$ SrtA can join the sorting signal and lysine-containing peptides together via an isopeptide linkage.<sup>44</sup> However, the reaction was inefficient and exhibited poor substrate specificity; even after 12 h, the ligation reaction was incomplete and a range of distinct isopeptide-linked products were generated. Moreover, during these reactions a significant amount of the sorting signal was proteolyzed, suggesting that  $^{68}\text{Zn}$ SrtA does not discriminate between lysine and water nucleophiles. This is in contrast to  $^{135}\text{Cd}$ SrtA $^{\Delta}$  and  $^{135}\text{Cd}$ SrtA $^{3M}$ , which produce lysine isopeptide-linked products at high yields with only limited proteolysis of the sorting signal substrate. A current limitation of the  $^{135}\text{Cd}$ SrtA-labeling system compared to  $^{68}\text{Zn}$ SrtA is that it requires the target protein to be expressed as a fusion to  $^{\text{N}}\text{SpaA}$  to achieve maximum labeling. However,

because  $^{\text{N}}\text{SpaA}$  is a small protein its presence is unlikely to affect the function of the target protein.

In conclusion, we have characterized the *in vitro* kinetics and mechanism of lysine–isopeptide bond forming activity of  $^{135}\text{Cd}$ SrtA, and discovered  $^{135}\text{Cd}$ SrtA $^{\Delta}$  which is  $\sim 7$ -fold more active than previously reported enzyme variants. This bioconjugation activity is beginning to rival that of microbial transglutaminases from *Streptomyces mobaraensis*, which can be used to join biomolecules isopeptide linkages between glutamine and lysine side chains.<sup>46</sup> Although promising, these enzymes have not gained wide usage in site-specific protein labeling, presumably because of their penchant to catalyze spurious ligations.<sup>46–49</sup> In contrast,  $^{135}\text{Cd}$ SrtA exhibits a high level of specificity for its substrates, ligating peptides containing a LPXTG sequence to a specific lysine residue within the  $^{\text{N}}\text{SpaA}$  domain. The molecular basis of specificity for the K190 side chain remains unknown, but presumably originates from protein–protein interactions between the  $^{\text{N}}\text{SpaA}$  and  $^{135}\text{Cd}$ SrtA-LPLT thioacyl intermediate that function to properly position the nucleophile for catalysis. It may also arise from  $^{\text{N}}\text{SpaA}$  structural features that provide an environment for K190 that lower its  $pK_a$ . Our mechanistic analysis also provides insight into the function of the lid, which is widely conserved in sortase enzymes that assemble pili. We show that *in vitro* its presence primarily affects the rate of thioacyl intermediate formation, and that it does not have a significant role in recognizing  $^{\text{N}}\text{SpaA}$  or the X residue within the LP(X)TG sorting signal. Further improvements in  $^{135}\text{Cd}$ SrtA-mediated labeling activity may also be possible, as we estimate that current versions of the enzyme catalyze *in vitro* transpeptidations  $\sim 10^2$ – $10^3$  times more slowly than the native enzyme when it is located on the cell surface. A number of approaches could be used to improve the kinetics of protein labeling, including discovering variants of  $^{135}\text{Cd}$ SrtA $^{\Delta}$  that have superior thermostability and employing strategies that increase the effective substrate concentration by either immobilizing the reactants or fusing the enzyme to its substrates.<sup>50–52</sup> Finally, further improvements may also be achieved by obtaining a greater understanding of the process of substrate recognition and catalysis.

## METHODS

**Protein Reagents.** Purified  $^{135}\text{Cd}$ SrtA $^{3M}$  pilin sortase (residues N37–Q257 of SrtA from *C. diphtheriae*) and enzyme variants were expressed and purified as described previously.<sup>22</sup> Briefly, proteins were expressed from a pE-SUMO (Life-sensors) plasmid in *E. coli* BL21 (DE3) cells. The cells were grown up in LB supplemented with 500  $\mu\text{g}/\text{mL}$  of Kanamycin at 37 °C until they reached an  $\text{OD}_{600}$  of  $\sim 0.6$ . The cells were induced with 1 M IPTG and then left to express at 17 °C for 8 to 12 h. After, the cells were removed from the incubator and pelleted at 8670g for 15 min. The pellets were then dissolved in a buffer of 50 mM Tris-HCl, 300 mM NaCl, at a pH of 8.0 (lysis buffer). Subsequently, the cells were lysed using high-pressure emulsification and then fractionated via centrifugation at 22,720g for 50 min. Afterward, the cell lysate was purified via IMAC- $\text{Co}^{2+}$  purification. Proteins were eluted from the resin using a lysis buffer supplemented with 200 mM Imidazole. The His $_{6x}$ -SUMO tags were removed via treatment by His $_{6x}$ -Ulp1 protease at 1 mg/mL and subsequent purification by IMAC- $\text{Co}^{2+}$ . Afterward, the protein was purified by size exclusion chromatography via the AKTA Pure (GE) and with Superdex 75pg resin. Protein purity was confirmed by SDS-PAGE. pE-SUMO expression plasmids encoding  $^{135}\text{Cd}$ SrtA $^{3M}$  variants were



created using standard molecular biology methods and confirmed by nucleotide sequencing.  $^{15}\text{N}$ SpaA (residues E30 to S195) and  $^{35}\text{S}$ SrtA (*S. aureus* Sortase A, residues Q60–K206) were purified as described previously.<sup>28,53</sup> All purified enzymes were stored at  $-20\text{ }^{\circ}\text{C}$  in buffer A (50 mM Tris-HCl, 300 mM NaCl, at a pH of 8.0) supplemented with 40% glycerol. The FELPLTGGSG peptide (LPLTG peptide) used in the transpeptidation and hydrolysis assays was synthesized by Peptide 2.0.

**Transpeptidation Assays.** An HPLC-based assay was developed to quantify the kinetics of  $^{15}\text{N}$ SpaA catalyzed lysine–isopeptide bond formation. In this assay, the  $^{15}\text{N}$ SpaA protein containing the reactive lysine is ligated to a FELPLTGGSG peptide (LPLTG, where the underlined residues correspond to the sorting signal) by the pilin sortase, followed by quantification using a HPLC C4 column. Reactions were performed in 100  $\mu\text{L}$  volumes and contained 25  $\mu\text{M}$  of pilin sortase (either  $^{15}\text{N}$ SpaA<sup>3M</sup> or  $^{15}\text{N}$ SpaA<sup>A</sup>), DTT (5 mM), either constant or variable amounts of LPLTG peptide (1 mM or 0.5 to 4 mM), and either constant or variable amounts of  $^{15}\text{N}$ SpaA (500  $\mu\text{M}$ , 62.5 to 500  $\mu\text{M}$ ). All components were dissolved in buffer A. At these substrate concentrations, an estimated  $K_M$  for  $^{15}\text{N}$ SpaA and LPLTG was determined. Reactions were initiated by adding the pilin sortase from a 2 mM stock solution, incubated for 3 h at  $25\text{ }^{\circ}\text{C}$  and then flash-frozen with liquid  $\text{N}_2$  and stored at  $-20\text{ }^{\circ}\text{C}$ . The reactions were analyzed using a Phenomenex C4 column (5  $\mu\text{m}$ ,  $4.6 \times 150\text{ mm}$ ) and with an initial dwell time of 3 min at 36%  $\text{CH}_3\text{CN}/0.1\%$  TFA followed by a linear gradient from 36% to 46%  $\text{CH}_3\text{CN}/0.1\%$  TFA for 10 min at 1 mL/min was applied. The column was subsequently flushed with high concentrations of  $\text{CH}_3\text{CN}/0.1\%$  at 1 mL/min.  $^{15}\text{N}$ SpaA containing peaks were detected at 215 nm and the amount of substrate converted to product was calculated by integrating the area under the HPLC traces. The identity of each peak in the HPLC chromatogram was confirmed via MALDI-TOF MS. The activity of  $^{15}\text{N}$ SpaA was compared to  $^{35}\text{S}$ SrtA, which catalyzes a transpeptidation reaction that forms a backbone–backbone peptide bond between LPXTG and oligoglycine peptides.  $^{35}\text{S}$ SrtA transpeptidation activity was measured as described previously.<sup>40</sup> These reactions were performed in an identical manner to the  $^{15}\text{N}$ SpaA reaction described above, except that they contained 25  $\mu\text{M}$   $^{35}\text{S}$ SrtA instead of  $^{15}\text{N}$ SpaA and were supplemented with 10 mM calcium in Buffer A, and triglycine (Gly<sub>3</sub> peptide) (62.5  $\mu\text{M}$  to 1 mM) instead of  $^{15}\text{N}$ SpaA. Reactions were quenched 15 min after mixing by adding an equal amount of 1 N hydrochloric acid (HCl). The reaction products were separated by HPLC using a Phenomenex C18 column (10  $\mu\text{m}$ ,  $4.6 \times 150\text{ mm}$ ) and a linear gradient from 26% to 30%  $\text{CH}_3\text{CN}/0.1\%$  TFA over 8 min (1 mL/min). All HPLC experiments were performed on an Agilent 1100 HPLC. For both the  $^{35}\text{S}$ SrtA and  $^{15}\text{N}$ SpaA reactions kinetic parameters were obtained by fitting the data with SigmaPlot 12.0.

Two types of transpeptidation assays were used to investigate sorting signal specificity of the  $^{15}\text{N}$ SpaA<sup>3M</sup> and  $^{15}\text{N}$ SpaA<sup>A</sup> enzymes for the X residue within the LPXTG sorting signal. A total of 20 peptides were tested in which the X residues in the FELPXTGGSG was varied (Peptide 2.0). Reactions were performed in buffer A with a total volume of 35  $\mu\text{L}$ . The reactions contained: either  $^{15}\text{N}$ SpaA<sup>3M</sup> or  $^{15}\text{N}$ SpaA<sup>A</sup> (200  $\mu\text{M}$ ), DTT (5 mM),  $^{15}\text{N}$ SpaA (200  $\mu\text{M}$ ), and one of the FELPXTGGSG peptides (1 mM, Peptide 2.0). Transpeptidation reactions performed at  $25\text{ }^{\circ}\text{C}$  for 24 h, and then

quenched by flash freezing in liquid  $\text{N}_2$  and stored at  $-20\text{ }^{\circ}\text{C}$ . Five microliters of each reaction was diluted 4 times in SDS loading buffer and separated using a 12% SDS-PAGE gel, then visualized by Coomassie staining. The resulting bands were analyzed with ImageJ with the zero hour time point used as a control for activity.

#### Hydrolysis and Cysteine Oxidation Measurements

The hydrolytic activity of  $^{35}\text{S}$ SrtA and  $^{15}\text{N}$ SpaA variants was determined using an HPLC-based assay that monitors the ability of each enzyme to cleave the LPLTG peptide between the threonine and glycine residues. Reactions were performed in buffer A supplemented with 10 mM calcium and contained a total volume of 100  $\mu\text{L}$ : sortase (50  $\mu\text{M}$ ), LPLTG (500  $\mu\text{M}$ ), and DTT (5 mM). Reactions were incubated at  $25\text{ }^{\circ}\text{C}$  for 24 h, and then quenched by adding an equal volume of 1 N HCl. Reaction products were separated on a Waters C18 Column (10  $\mu\text{m}$ ,  $4.6 \times 150\text{ mm}$ ) using a linear gradient from 26% to 30%  $\text{CH}_3\text{CN}/0.1\%$  TFA over 8 min at 1 mL/min. The reaction was monitored at 215 nm and the identity of each peak in the HPLC chromatogram was confirmed via LC-MS.

The oxidation status of the active site cysteine by monitoring susceptibility to iodoacetamide alkylation.  $^{15}\text{N}$ SpaA<sup>3M</sup> and  $^{15}\text{N}$ SpaA<sup>A</sup> were purified and stored in buffer A. Enzymes were then either treated to a final concentration of 5 mM of DTT or a buffer control for 1 h before being frozen at  $-20\text{ }^{\circ}\text{C}$ . Subsequently, the proteins were defrosted and alkylated in 25 mM iodoacetamide and exchanged four times into 100 mM ammonium bicarbonate buffer using Amicon 3 kDa centrifugal filters.<sup>54</sup> Trypsin digestion was performed overnight at  $37\text{ }^{\circ}\text{C}$ . C18 Stage Tips were used for desalting prior to tandem mass spectrometry. Peptides were separated and measured on an EASY-Spray HPLC column (25 cm  $\times$  75  $\mu\text{m}$  ID packed with PepMap RSLC C18, 2  $\mu\text{m}$  particles, Thermo Scientific) with an online Easy-nLC 100 chromatography system to a Orbitrap mass spectrometer (Q-Exactive Orbitrap, Thermo Scientific). Precursor ions were selected using data-dependent acquisition (top 10) and fragmented using collision induced dissociation (CID) at a normalized collision energy of 27. Raw MS/MS files were converted to mgf format (Thermo Proteome Discoverer, Thermo Scientific ver. 1.4) and were searched against a sequence database using MASCOT (Matrix Science). Searches employed variable cysteine carbamidomethylation and methionine oxidation. The precursor mass accuracy was set to 10 ppm, while that for product ions was set to 0.02 da. Once identified, the fully digested peptides' intensities were quantified from the area under the curves. Intensities of carbamidomethylated peptides were normalized by dividing against the summed intensity of the three most abundant peptides identified.

**Acyl-Intermediate Detection with LC-MS.** LC-MS reactions to compare acyl formation with and without the presence of  $^{15}\text{N}$ SpaA were performed in buffer A and contained a total volume of 100  $\mu\text{L}$ : sortase (25  $\mu\text{M}$ ), DTT (5 mM), LPLTG (1 mM), and  $^{15}\text{N}$ SpaA (250  $\mu\text{M}$ ) or buffer. All reactions were incubated at  $25\text{ }^{\circ}\text{C}$  and 10  $\mu\text{L}$  time points were removed and frozen with liquid  $\text{N}_2$  before being stored at  $-20\text{ }^{\circ}\text{C}$ . An experiment containing only the enzyme (no peptide added) was performed to provide an external standard for the amount of unmodified enzyme in the mass spectrum. Because the  $^{15}\text{N}$ SpaA-LPLT acyl intermediate cannot readily be separated from unmodified  $^{15}\text{N}$ SpaA by chromatography, its amount in the assay is estimated from the mass spectrometry data by assuming that the acyl-intermediate and unmodified enzyme

ionize to similar extents. Before being run on the LC-MS system, samples were then diluted with 90  $\mu$ L of 200 mM L-tryptophan (internal standard) and then measured on a Zorbax 300SB-C3 (3.5  $\mu$ m, 3.0  $\times$  150 mm) with an Agilent 6530 Q-TOF and Agilent 1260 Infinity HPLC with a gradient of 30–99% over 6 min at 0.8 mL/min. The data was analyzed with Agilent MassHunter Qualitative Analysis. Note that, although the sequence of <sup>N</sup>SpaA has been verified, the mass of <sup>N</sup>SpaA and the acyl intermediate in the spectrum is 88 Da larger than the mass predicted based on the primary sequence. The origin of this difference is not known. The amount of enzyme and acyl-intermediate was calculated by integrating the area under the curve for each peak.

## AUTHOR INFORMATION

### Corresponding Author

**Robert T. Clubb** – Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, and Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California 90095, United States; [orcid.org/0000-0001-5718-3985](https://orcid.org/0000-0001-5718-3985); Phone: (+1) 310 206 2334; Email: [rclubb@mbi.ucla.edu](mailto:rclubb@mbi.ucla.edu); Fax: (+1) 310 206 4779

### Authors

**Christopher K. Sue** – Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States

**Scott A. McConnell** – Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States

**Ken Ellis-Guardiola** – Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States

**John M. Muroski** – Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States

**Rachel A. McAllister** – Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States

**Justin Yu** – Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States

**Ana I. Alvarez** – Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States

**Chungyu Chang** – Molecular Biology Institute and the Division of Oral Biology and Medicine, School of Dentistry, University of California, Los Angeles, Los Angeles, California 90095, United States

**Rachel R. Ogorzalek Loo** – Department of Chemistry and Biochemistry and UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States

**Joseph A. Loo** – Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California, Los

Angeles, Los Angeles, California 90095, United States;

[orcid.org/0000-0001-9989-1437](https://orcid.org/0000-0001-9989-1437)

**Hung Ton-That** – Molecular Biology Institute and the Division of Oral Biology and Medicine, School of Dentistry, University of California, Los Angeles, Los Angeles, California 90095, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.bioconjchem.0c00163>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research program under Award Number DE-FC02-02ER63421 and National Institutes of Health Grants AIS2217 (R.T.C. and H.T.-T.), DE025015 (H.T.-T.) and GM103479 (J.A.L.). C.K.S. and S.A.M. were supported by a Cellular and Molecular Biology Training Grant (Ruth L. Kirschstein National Research Service Award GM007185). J.M. was supported by a UCLA Molecular Biology Institute Whitcome Fellowship (S10OD016336) NMR equipment used in this research was purchased using funds from shared equipment grant NIH S10OD016336.

## REFERENCES

- (1) Agarwal, P., and Bertozzi, C. R. (2015) Site-Specific Antibody “Drug Conjugates: The Nexus of Bioorthogonal Chemistry, Protein Engineering, and Drug Development. *Bioconjugate Chem.* 26 (2), 176–192.
- (2) Chudasama, V., Maruani, A., and Caddick, S. (2016) Recent advances in the construction of antibody drug conjugates. *Nat. Chem.* 8 (2), 114–119.
- (3) Hoyt, E. A., Cal, P. M. S. D., Oliveira, B. L., and Bernardes, G. A. J. L. (2019) Contemporary approaches to site-selective protein modification. *Nature Reviews Chemistry* 3 (3), 147–171.
- (4) Lagasse, H. A. D., Alexaki, A., Simhadri, V. L., Katagiri, N. H., Jankowski, W., Sauna, Z. E., and Kimchi-Sarfaty, C. (2017) Recent advances in (therapeutic protein) drug development. *F1000Research* 6, 113–113.
- (5) Specht, E. A., Braselmann, E., and Palmer, A. E. (2017) A Critical and Comparative Review of Fluorescent Tools for Live-Cell Imaging. *Annu. Rev. Physiol.* 79 (1), 93–117.
- (6) Sochaj, A. M., Swiderska, K. W., and Otlewski, J. (2015) Current methods for the synthesis of homogeneous antibody drug conjugates. *Biotechnol. Adv.* 33 (6), 775–784.
- (7) Matsumoto, T., Tanaka, T., and Kondo, A. (2012) Enzyme-mediated methodologies for protein modification and bioconjugate synthesis. *Biotechnol. J.* 7 (9), 1137–1146.
- (8) Mohamad, N. R., Marzuki, N. H. C., Buang, N. A., Huyop, F., and Wahab, R. A. (2015) An overview of technologies for immobilization of enzymes and surface analysis techniques for immobilized enzymes. *Biotechnol. Biotechnol. Equip.* 29 (2), 205–220.
- (9) Krall, N., da Cruz, F. P., Boutourel, O., and Bernardes, G. A. J. L. (2016) Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* 8 (2), 103–113.
- (10) Spicer, C. D., and Davis, B. G. (2014) Selective chemical protein modification. *Nat. Commun.* 5 (1), 4740.
- (11) Li, X., Fang, T., and Boons, G.-J. (2014) Preparation of Well-Defined Antibody Drug Conjugates through Glycan Remodeling and Strain-Promoted Azide-Alkyne Cycloadditions. *Angew. Chem., Int. Ed.* 53 (28), 7179–7182.
- (12) Zhang, Y., Park, K.-Y., Suazo, K. F., and Distefano, M. D. (2018) Recent progress in enzymatic protein labelling techniques and their applications. *Chem. Soc. Rev.* 47 (24), 9106–9136.

- (13) Antos, J. M., Truttmann, M. C., and Ploegh, H. L. (2016) Recent advances in sortase-catalyzed ligation methodology. *Curr. Opin. Struct. Biol.* 38, 111–118.
- (14) Schmohl, L., and Schwarzer, D. (2014) Sortase-mediated ligations for the site-specific modification of proteins. *Curr. Opin. Chem. Biol.* 22, 122–128.
- (15) Popp, M. W.-L., and Ploegh, H. L. (2011) Making and Breaking Peptide Bonds: Protein Engineering Using Sortase. *Angew. Chem., Int. Ed.* 50 (22), 5024–5032.
- (16) Samantary, S., Marathe, U., Dasgupta, S., Nandicoori, V. K., and Roy, R. P. (2008) Peptide $\alpha$ Sugar Ligation Catalyzed by Transpeptidase Sortase: A Facile Approach to Neoglycoconjugate Synthesis. *J. Am. Chem. Soc.* 130 (7), 2132–2133.
- (17) Mazmanian, S. K., Liu, G., Ton-That, H., and Schneewind, O. (1999) *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science* 285 (5428), 760–3.
- (18) Antos, J. M., Chew, G.-L., Guimaraes, C. P., Yoder, N. C., Grotenbreg, G. M., Popp, M. W.-L., and Ploegh, H. L. (2009) Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J. Am. Chem. Soc.* 131 (31), 10800–10801.
- (19) Popp, M. W., Dougan, S. K., Chuang, T. Y., Spooner, E., and Ploegh, H. L. (2011) Sortase-catalyzed transformations that improve the properties of cytokines. *Proc. Natl. Acad. Sci. U. S. A.* 108 (8), 3169–74.
- (20) Tsukiji, S., and Nagamune, T. (2009) Sortase-mediated ligation: a gift from Gram-positive bacteria to protein engineering. *ChemBioChem* 10 (5), 787–98.
- (21) Fottner, M., Brunner, A.-D., Bittl, V., Horn-Ghetko, D., Jussupow, A., Kaila, V. R. I., Bremm, A., and Lang, K. (2019) Site-specific ubiquitylation and SUMOylation using genetic-code expansion and sortase. *Nat. Chem. Biol.* 15 (3), 276–284.
- (22) McConnell, S. A., Amer, B. R., Muroski, J., Fu, J., Chang, C., Ogorzalek Loo, R. R., Loo, J. A., Osipiuk, J., Ton-That, H., and Clubb, R. T. (2018) Protein Labeling via a Specific Lysine-Isopeptide Bond Using the Pilin Polymerizing Sortase from *Corynebacterium diphtheriae*. *J. Am. Chem. Soc.* 140 (27), 8420–8423.
- (23) Jacobitz, A. W., Katke, M. D., Wereszczynski, J., and Clubb, R. T. (2017) Sortase Transpeptidases: Structural Biology and Catalytic Mechanism. *Adv. Protein Chem. Struct. Biol.* 109, 223–264.
- (24) Kang, H. J., and Baker, E. N. (2009) Intramolecular isopeptide bonds give thermodynamic and proteolytic stability to the major pilin protein of *Streptococcus pyogenes*. *J. Biol. Chem.* 284 (31), 20729–37.
- (25) Comfort, D., and Clubb, R. T. (2004) A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria. *Infect. Immun.* 72 (5), 2710–22.
- (26) Echelman, D. J., Alegre-Cebollada, J., Badilla, C. L., Chang, C., Ton-That, H., and Fernandez, J. M. (2016) CnaA domains in bacterial pili are efficient dissipaters of large mechanical shocks. *Proc. Natl. Acad. Sci. U. S. A.* 113 (9), 2490–5.
- (27) Ton-That, H., and Schneewind, O. (2003) Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Mol. Microbiol.* 50 (4), 1429–38.
- (28) Chang, C., Amer, B. R., Osipiuk, J., McConnell, S. A., Huang, I. H., Hsieh, V., Fu, J., Nguyen, H. H., Muroski, J., Flores, E., Ogorzalek Loo, R. R., Loo, J. A., Putkey, J. A., Joachimiak, A., Das, A., Clubb, R. T., and Ton-That, H. (2018) In vitro reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking. *Proc. Natl. Acad. Sci. U. S. A.* 115 (24), E5477–E5486.
- (29) Frankel, B. A., Kruger, R. G., Robinson, D. E., Kelleher, N. L., and McCafferty, D. G. (2005) *Staphylococcus aureus* sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry* 44 (33), 11188–200.
- (30) Clancy, K. W., Melvin, J. A., and McCafferty, D. G. (2010) Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. *Biopolymers* 94 (4), 385–96.
- (31) Bradshaw, W. J., Davies, A. H., Chambers, C. J., Roberts, A. K., Shone, C. C., and Acharya, K. R. (2015) Molecular features of the sortase enzyme family. *FEBS J.* 282 (11), 2097–114.
- (32) Guttilla, I. K., Gaspar, A. H., Swierczynski, A., Swaminathan, A., Dwivedi, P., Das, A., and Ton-That, H. (2009) Acyl enzyme intermediates in sortase-catalyzed pilus morphogenesis in gram-positive bacteria. *J. Bacteriol.* 191 (18), 5603–12.
- (33) Jacobitz, A. W., Naziga, E. B., Yi, S. W., McConnell, S. A., Peterson, R., Jung, M. E., Clubb, R. T., and Wereszczynski, J. (2016) The “Lid” in the *Streptococcus pneumoniae* SrtC1 Sortase Adopts a Rigid Structure that Regulates Substrate Access to the Active Site. *J. Phys. Chem. B* 120, 8302.
- (34) Chan, A. H., Yi, S. W., Terwilliger, A. L., Maresso, A. W., Jung, M. E., and Clubb, R. T. (2015) Structure of the *Bacillus anthracis* Sortase A Enzyme Bound to Its Sorting Signal: A Flexible Amino-Terminal Appenage Modulates Substrate Access. *J. Biol. Chem.* 290 (42), 25461–74.
- (35) Kruger, R. G., Otvos, B., Frankel, B. A., Bentley, M., Dostal, P., and McCafferty, D. G. (2004) Analysis of the Substrate Specificity of the *Staphylococcus aureus* Sortase Transpeptidase SrtA. *Biochemistry* 43 (6), 1541–1551.
- (36) Frankel, B. A., Tong, Y., Bentley, M. L., Fitzgerald, M. C., and McCafferty, D. G. (2007) Mutational analysis of active site residues in the *Staphylococcus aureus* transpeptidase SrtA. *Biochemistry* 46 (24), 7269–78.
- (37) Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O., and Narayana, S. V. (2004) Crystal structures of *Staphylococcus aureus* sortase A and its substrate complex. *J. Biol. Chem.* 279 (30), 31383–9.
- (38) Huang, X., Aulabaugh, A., Ding, W., Kapoor, B., Alksne, L., Tabei, K., and Ellestad, G. (2003) Kinetic mechanism of *Staphylococcus aureus* sortase SrtA. *Biochemistry* 42 (38), 11307–15.
- (39) Alcock, L. J., Perkins, M. V., and Chalker, J. M. (2018) Chemical methods for mapping cysteine oxidation. *Chem. Soc. Rev.* 47 (1), 231–268.
- (40) Kruger, R. G., Dostal, P., and McCafferty, D. G. (2004) Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the *Staphylococcus aureus* sortase transpeptidase SrtA. *Anal. Biochem.* 326 (1), 42–8.
- (41) Weiner, E. M., Robson, S., Marohn, M., and Clubb, R. T. (2010) The Sortase A enzyme that attaches proteins to the cell wall of *Bacillus anthracis* contains an unusual active site architecture. *J. Biol. Chem.* 285 (30), 23433–43.
- (42) Connolly, K. M., Smith, B. T., Pilpa, R., Ilangovan, U., Jung, M. E., and Clubb, R. T. (2003) Sortase from *Staphylococcus aureus* does not contain a thiolate-imidazolium ion pair in its active site. *J. Biol. Chem.* 278 (36), 34061–5.
- (43) Bellucci, J. J., Bhattacharyya, J., and Chilkoti, A. (2014) A Noncanonical Function of Sortase Enables Site-Specific Conjugation of Small Molecules to Lysine Residues in Proteins. *Angew. Chem., Int. Ed.* 54 (2), 441–445.
- (44) Dasgupta, S., Samantary, S., Sahal, D., and Roy, R. P. (2011) Isopeptide ligation catalyzed by quintessential sortase A: mechanistic cues from cyclic and branched oligomers of indolicidin. *J. Biol. Chem.* 286 (27), 23996–24006.
- (45) Mohlmann, S., Mahlert, C., Greven, S., Scholz, P., and Harrenga, A. (2011) In vitro Sortagging of an Antibody Fab Fragment: Overcoming Unproductive Reactions of Sortase with Water and Lysine Side Chains. *ChemBioChem* 12 (11), 1774–1780.
- (46) Deweid, L., Avrutina, O., and Kolmar, H. (2019) Microbial transglutaminase for biotechnological and biomedical engineering. *Biol. Chem.* 400, 257.
- (47) Malešević, M., Migge, A., Hertel, T. C., and Pietzsch, M. (2015) A Fluorescence-Based Array Screen for Transglutaminase Substrates. *ChemBioChem* 16 (8), 1169–1174.
- (48) Jeger, S., Zimmermann, K., Blanc, A., Grünberg, J., Honer, M., Hunziker, P., Struthers, H., and Schibli, R. (2010) Site-Specific and Stoichiometric Modification of Antibodies by Bacterial Transglutaminase. *Angew. Chem., Int. Ed.* 49 (51), 9995–9997.

(49) Maullu, C., Raimondo, D., Caboi, F., Giorgetti, A., Sergi, M., Valentini, M., Tonon, G., and Tramontano, A. (2009) Site-directed enzymatic PEGylation of the human granulocyte colony-stimulating factor. *FEBS J.* 276 (22), 6741–6750.

(50) Amer, B. R., Macdonald, R., Jacobitz, A. W., Liauw, B., and Clubb, R. T. (2016) Rapid addition of unlabeled silent solubility tags to proteins using a new substrate-fused sortase reagent. *J. Biomol. NMR* 64 (3), 197–205.

(51) Policarpo, R. L., Kang, H., Liao, X., Rabideau, A. E., Simon, M. D., and Pentelute, B. L. (2014) Flow-Based Enzymatic Ligation by Sortase A. *Angew. Chem., Int. Ed.* 53 (35), 9203–9208.

(52) Witte, M. D., Wu, T., Guimaraes, C. P., Theile, C. S., Blom, A. E. M., Ingram, J. R., Li, Z., Kundrat, L., Goldberg, S. D., and Ploegh, H. L. (2015) Site-specific protein modification using immobilized sortase in batch and continuous-flow systems. *Nat. Protoc.* 10 (3), 508–516.

(53) Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O., and Clubb, R. T. (2001) Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U. S. A.* 98 (11), 6056–61.

(54) Melvin, J. A., Murphy, C. F., Dubois, L. G., Thompson, J. W., Moseley, M. A., and McCafferty, D. G. (2011) *Staphylococcus aureus* Sortase A Contributes to the Trojan Horse Mechanism of Immune Defense Evasion with Its Intrinsic Resistance to Cys184 Oxidation. *Biochemistry* 50 (35), 7591–7599.

## Chapter 4

### Unique Structural and Kinetic Features of the Terminating Pilin

Subunit from *Corynebacterium diphtheriae*

## 4.1 Overview

Gram-positive bacteria utilize long filamentous pili on their surfaces to form biofilms and to adhere to host tissues, abiotic surfaces, and other bacteria. These pili are made of individual subunits called pilins and are linked together through a sortase-dependent mechanism that connects the pilins via lysine isopeptide bonds. Previous studies have explored how sortases crosslink SpaA pilins, which compose the majority of the pilus shaft. Here, using a reconstituted pilus assembly system from *Corynebacterium diphtheriae* and NMR, we have investigated how SrtA adds SpaB pilins to the pilus, which terminates pilus assembly resulting in the transfer of the pilus to the cell wall.

We have discovered that: (i)  $^{Cd}$ SrtA terminates pilus assembly by catalyzing the formation of an isopeptide bond between K139 on SpaB and T494 on SpaA, (ii)  $^{Cd}$ SrtA catalyzes the formation of the bond between SpaB and SpaA more slowly than it creates isopeptide bonds that form the shaft of the pilus ( $k_{cat}$  values of  $7.0 \pm 0.4$  and  $40.0 \pm 0.1$  ( $10^{-5}$  ( $s^{-1}$ )) for SpaB and SpaA, respectively, (iii) SpaA and SpaB sortase substrates adopt similar atomic structures despite sharing 17% of sequence homology (iv) both SpaA and SpaB contain a disordered AB loop that is expected to reside at the enzyme-substrate interface. However, the loops are distinctly positioned on the surfaces of the pilin suggesting that the AB loop may be a key determinant for controlling the rate of reactivity with  $^{Cd}$ SrtA and (v) we present evidence that SpaB has a high affinity for  $^{Cd}$ SrtA, suggesting that even though it is less reactive, on the cell surface it may act to inhibit pilus polymerization to promote pilus assembly termination. Combined, this work provides mechanistic insight into pilus assembly by shedding light on how the reaction is terminated. Research in this chapter will be published in the future.

## 4.2 Introduction

Adhesive protein polymers called “pili” or “fimbriae” project from the surface of bacteria and are frequently used as virulence factors that have a wide range of functions, including among others mediating initial host-pathogen interaction, twitching motility, conjugation, and immunomodulation.<sup>1-3</sup> Gram-positive bacteria such as *Actinomyces oris*, *Enterococcus faecalis*, and *Bacillus cereus*, use sortase enzymes to build long (1-5 micron) and thin (1-5 nm) pili.<sup>4-6</sup> Sortases catalyze the formation of unique lysine-isopeptide bonds between pilin subunits and these bonds confer extensive tensile strength and resistance to proteolysis.<sup>7,8</sup> Microbes with impaired pilus assembly exhibit diminished adherence to host tissues and attenuated bacterial virulence.<sup>9,10</sup> Due to this key functionality and surface accessibility, the pilus assembly reaction is a potential drug target for novel antibiotic therapeutics. Despite the presence of pili on a wide variety of gram-positive commensal and pathogenic bacterial species, the molecular basis through which pili are assembled remains poorly understood.<sup>11</sup>

The most well understood pilus assembly system is from *Corynebacterium diphtheriae*, the main cause of pharyngeal diphtheria, and displays three types of pili: SpaA, SpaD and SpaH.<sup>12-14</sup> Of these three pili, assembly of the SpaA pilus that mediates adherence to the pharyngeal epithelium is best understood.<sup>1,15</sup> The SpaA pilus is constructed from three types of pilin subunits: the main shaft pilin SpaA, the tip pilin SpaC, and the basal pilin SpaB, whose incorporation terminates the polymerization reaction to promote pilus attachment to the cell wall.<sup>12,16</sup> The sortase A (<sup>Cd</sup>SrtA) enzyme assembles the SpaA pilus which builds the shaft of the pilus by catalyzing a repetitive, irreversible transpeptidation reaction that joins a C-terminal cell wall sorting signal (CWSS) located on the C-domain of a SpaA pilin to a conserved lysine side chain (K190) located on the N-domain of another SpaA protein. The enzyme uses similar

chemistry to attach the SpaC and SpaB proteins to the tip and base of the pilus. The details of this process are described in Chapter 1.

Work described in this chapter focuses on the SpaB pilin, whose incorporation into the pilus terminates the assembly reaction and controls the length of the pilus.<sup>15</sup> In the termination reaction, the <sup>Cd</sup>SrtA enzyme recognizes SpaB's K139 and adds it to the base of the growing pilus shaft.<sup>15</sup> This process is stoichiometrically controlled, as a SpaB knockout strain of *C. diphtheriae* exhibits extra-long pili and excess amounts of SpaB result in truncated pili.<sup>15</sup> SpaB has also been found to be interspersed within the shaft of the pilus through an unknown mechanism.<sup>17</sup> SpaB and SpaC are required for adherence to pharyngeal cells and a previous study by the Ton-That lab found monomeric pilins of SpaB and SpaC and heterodimer complexes composed of SpaB-SpaC on the cell surface.<sup>18</sup> In their study, they hypothesize that these monomeric and dimeric pilins mediate close contact to *C. diphtheriae* after longer pili initiate contact. SpaB spaced throughout the pilus may serve a similar purpose by providing multiple points of contact for *C. diphtheriae* to bind to its host surface.<sup>16</sup> In order to further understand the termination reaction, we probed the kinetics of <sup>Cd</sup>SrtA catalyzing a reaction of SpaB to the CWSS of SpaA, and determined the solution structure of SpaB. These results expand our understanding of the role SpaB in pilus biogenesis and suggest that the pilus may be flexibly linked to the cell wall.

### **4.3 Results and Discussion**

#### Reactive lysine residue and its recognition by sortase

Sortase catalysis is composed of a two-step process with an acyl-intermediate formation step between <sup>Cd</sup>SrtA and the LPLTG CWSS, and a transpeptidation step linking the threonine on the CWSS to the reactive lysine residue found on SpaA or SpaB.<sup>19</sup> It is currently unclear how sortase crosslinks the SpaB pilin to base of the pilus. Previously reported *in vivo* and *in vitro* studies suggested that SpaB is crosslinked to SpaA via an isopeptide bond between the side



chain of K139 within SpaB and the C-terminal T494 residue within SpaA.<sup>20,21</sup> However, the kinetics of this process or the enzyme's selectivity for K139 are poorly understood. To investigate this, we purified SpaB (residues Q25-V150) which contains the reactive lysine (K139) residue and a construct that contains the knockout of the K139, SpaB-K139A. The SpaB constructs lack the N-terminal residues that contain the signal peptide and the C-terminal sorting signal of LAFTG. They were expressed in *Escherichia coli* and purified by metal affinity chromatography. Reactions were set up by mixing together SpaB (300  $\mu$ M) or SpaB-K139A (300  $\mu$ M) with  $^{14}$ C-SrtA $^{\Delta}$  (100  $\mu$ M), a sortase from *C. diphtheriae* where the residues N37-Q257 are expressed and the amino acids I78 to A88 are deleted, and  $^{14}$ CSpaA (300  $\mu$ M), the C-terminal domain of SpaA from residues 350 to 500 which contains the CWSS LPLTG motif.<sup>22</sup> A control using  $^{14}$ CSpaA (300  $\mu$ M), the N-terminal domain of SpaA from residues 30 to 194, which contains the K190 nucleophile, was also setup with  $^{14}$ C-SrtA $^{\Delta}$  (100  $\mu$ M), and  $^{14}$ CSpaA (300  $\mu$ M). The reactions were incubated at 25°C and timepoints were taken after 0, 24, and 48 hours of incubation and stopped by the addition of SDS-containing sample buffer. Timepoints were then separated and analyzed by SDS/PAGE and Coomassie staining. In the reaction containing SpaB,  $^{14}$ CSpaA, and  $^{14}$ C-SrtA $^{\Delta}$ , a strong band formed around the 30 kDa weight marker after 24 hours. This mass approximately corresponds to  $^{14}$ CSpaA-SpaB ligated together, while the reaction containing SpaB-K139A did not contain a band at a similar weight for the  $^{14}$ CSpaA-SpaB-K139A complex. **(Figure 4.1A)** To ensure that no other lysines were potentially reacting with  $^{14}$ CSpaA, we mutated the other lysine residues found in SpaB (SpaB-K53A and SpaB-K115A) and incubated them with  $^{14}$ CSpaA and  $^{14}$ C-SrtA $^{\Delta}$  in an identical manner as before. **(Figure 4.1B)** Similar to the SpaB containing no mutations, we saw a band that corresponded to a SpaB- $^{14}$ CSpaA product. Working with Janine Fu from the Loo Lab at UCLA, the linkage was further verified through the use of LC-MS/MS which was able to detect the specific linkage of SpaB's K139 to the LPLTG.

(**Figure 4.1C**) Thus, we have demonstrated that K139 is the nucleophilic lysine residue responsible for ligation between SpaB and SpaA.

#### Kinetics of SpaB's lysine-isopeptide bond formation compared to <sup>N</sup>SpaA

We have previously characterized <sup>Cd</sup>SrtA's kinetics with <sup>N</sup>SpaA and a peptide that contained SpaA's CWSS (FELPLTGGSG), hereafter called the LPLTG peptide.<sup>22</sup> To investigate if there were any differences in kinetics between <sup>N</sup>SpaA and SpaB, identical transpeptidation reactions were created with SpaB. We used identical conditions to our previously reported kinetic data in Chapter 3 and incubated SpaB at varying concentrations with 1 mM of LPLTG peptide and 25  $\mu\text{M}$  <sup>Cd</sup>SrtA <sup>$\Delta$</sup> .<sup>22</sup> 10  $\mu\text{L}$  aliquots of each reaction were taken after 3 hours of incubating at 25°C and frozen with liquid N<sub>2</sub> to stop the reaction. The reactants and products were readily separated by reverse phase HPLC where we monitored the conversion of reactant (SpaB) into the product (SpaB-LPLT). The identity of the reactants and products was confirmed by LC-MS/MS. Michaelis-Menten kinetics were obtained with SpaB concentration varied from 62.5 to 500  $\mu\text{M}$  (**Figure 4.2A**). The measurements revealed that <sup>Cd</sup>SrtA <sup>$\Delta$</sup>  has a  $K_M$  of  $32 \pm 2 \mu\text{M}$  and a  $k_{\text{cat}}$  of  $7.0 \pm 0.4 \times 10^{-5} \text{ s}^{-1}$  (**Table 4.1**) when SpaB is used as a substrate. Comparing the  $k_{\text{cat}}$  between <sup>N</sup>SpaA and SpaB, SpaB is around 5.7 times less reactive than <sup>N</sup>SpaA.<sup>22</sup> This slower reactivity makes sense in light of SpaB's purpose as the terminating pilus. If SpaB had a faster or equivalent  $k_{\text{cat}}$  to SpaA, it would most likely result in extremely short pili that may be unable to mediate adhesion to host tissues. Therefore, the slower kinetics of SpaB must be essential in modulating the length of the pilus and preventing the accidental early termination of pilus assembly.

To investigate the SpaB-LPLT complex for structural studies, we wanted to see if SpaB incubated over a longer period with LPLTG peptide and SrtA would generate >90% product similar to <sup>N</sup>SpaA.<sup>22,23</sup> We mixed SpaB with <sup>Cd</sup>SrtA <sup>$\Delta$</sup>  and 4 mM of the LPLTG peptide and incubated it at 25°C for more than 24 hours. Taking samples from these reactions and injecting

them onto the HPLC, we found that only 50-70% of SpaB is ligated to the LPLTG peptide compared to nearly 90-95% of <sup>N</sup>SpaA (**Figure 4.2B**). Since we initially expected it to modify to similar levels as <sup>N</sup>SpaA, we investigated to see if the lack of modification was due to the removal of the inhibitory lid region (residues I78 to A88) on <sup>Cd</sup>SrtA. The inhibitory lid on <sup>Cd</sup>SrtA has been previously shown to inhibit activity *in vitro* and its true purpose has yet to be elucidated.<sup>20</sup> We therefore used another construct <sup>Cd</sup>SrtA<sup>3M</sup> (residues N37-Q257 of <sup>Cd</sup>SrtA with three mutations D81G/W83G/N85A) which contains the lid region but has three mutations in the lid region that reconstitute *in vitro* activity.<sup>23</sup> Testing with <sup>Cd</sup>SrtA<sup>3M</sup> under identical conditions as before, we attempted to react SpaB with LPLTG peptide and found no major differences between when we used <sup>Cd</sup>SrtA<sup>3M</sup> and <sup>Cd</sup>SrtA<sup>Δ</sup>. Also, to ensure that the enzymes were not unfolded, we conducted SYPRO Orange thermostability assays similar to what was done in Chapter 2 (**Table 4.2**). At present, we are unsure what is preventing full peptide modification to <sup>N</sup>SpaA by <sup>Cd</sup>SrtA<sup>3M</sup>.

Previously in Chapter 3, I was able to show that the first step in sortase catalysis was the slow step when <sup>N</sup>SpaA and the LPLTG were used as substrates. This was done by using liquid chromatography-mass spectrometry (LC-MS) which tracked the formation of the acyl-intermediate complex (<sup>Cd</sup>SrtA-LPLT) between <sup>Cd</sup>SrtA and the LPLTG peptide. Our results show that when <sup>N</sup>SpaA was introduced, <sup>Cd</sup>SrtA-LPLT concentration remained very low and we believe that <sup>N</sup>SpaA's lysine residue immediately reacts with <sup>Cd</sup>SrtA-LPLT when it was present. This provided evidence that the formation of the acyl-intermediate was the slow step in the reaction. Due to SpaB having lower  $k_{cat}$  compared to <sup>N</sup>SpaA, we wanted to test and see if the first step continued to be the slow step in the reaction. The assay was replicated as before, 25  $\mu$ M of <sup>Cd</sup>SrtA<sup>Δ</sup> with 1mM of the LPLTG peptide and either SpaB (250  $\mu$ M), <sup>N</sup>SpaA (250  $\mu$ M), or buffer was added (**Figure 4.2C**). The reactions were then incubated at 25°C and 10  $\mu$ L time points were taken out at 5, 30, 60, and 180 minutes. Similar to our findings before, we found that when SpaB or <sup>N</sup>SpaA is present, acyl-intermediate formation remains around 4-5% for three hours,

but when SpaB is absent, acyl-intermediate concentration increases over time. This agrees well with our previous finding that acyl-intermediate formation is the rate limiting step in the reaction.

#### Comparing $^{Cd}$ SrtA reaction preference of between $^{N}$ SpaA and SpaB

Since SpaB and  $^{N}$ SpaA both contain a highly specific lysine that reacts with the same CWSS, we wanted to test if  $^{Cd}$ SrtA $^{\Delta}$  had a particular preference for SpaA or SpaB's lysine nucleophile.<sup>20,21</sup> We therefore constructed a reaction that contained  $^{N}$ SpaA and SpaB at equal molar concentrations of 200  $\mu$ M, and 1 mM of the LPLTG peptide (**Figure 4.3A**). Reactions were initiated when  $^{Cd}$ SrtA $^{\Delta}$  (25  $\mu$ M) was added and were separated using reverse phase HPLC on a C4 column. Controls were run containing either  $^{N}$ SpaA or SpaB and LPLTG peptide and  $^{Cd}$ SrtA $^{\Delta}$  at identical concentrations as above and additional buffer. The amount of SpaB-LPLT product generated in either the presence or absence of  $^{N}$ SpaA showed only small differences within the first three hours (**Figure 4.3B**). This is in stark contrast to the amount of  $^{N}$ SpaA-LPLT produced when SpaB is also present in the reaction. When SpaB is present, the formation of  $^{N}$ SpaA-LPLT initially slows down (**Figure 4.3C**). This suggests that SpaB preferentially reacts with  $^{Cd}$ SrtA-LPLT as compared to  $^{N}$ SpaA. We then wanted to see if SpaB containing a lysine knockout mutation (K139A) also slowed  $^{N}$ SpaA-LPLT product formation. After setting up identical reactions as above, but with SpaB-K139A, we found that  $^{N}$ SpaA-LPLT formation was also slowed as compared to reactions that only contained  $^{N}$ SpaA,  $^{Cd}$ SrtA $^{\Delta}$ , and the LPLTG peptide. We conclude that when both SpaA and SpaB are present that SpaB is more reactive than SpaA. This is consistent with enzyme's lower  $K_M$  value for SpaB as compared to  $^{N}$ SpaA. This suggests that SpaB binds to the sortase enzyme in a way that limits the activity of sortase toward  $^{N}$ SpaA, such that SpaB acts as a competitive inhibitor.

## NMR structure of SpaB

To gain a greater understanding of the differences between SpaB and <sup>N</sup>SpaA, heteronuclear multidimensional NMR spectroscopy was used to determine the atomic structure of SpaB. The structure was determined using 1,273 experimental restraints (**Table 4.3**). The structure's backbone was defined between N28-D142 to an average structure with an RMSD of  $0.5 \pm 0.05$  Å. The structure of SpaB consists of a single domain and adopts a CnaB-type fold. (**Figure 4.4A**) The CnaB-type fold contains a core  $\beta$ -sandwich which is composed of a set of three beta sheet strands (GAD), a set of four beta sheet strands (CBEF), and a single  $\alpha$ -helix (**Figure 4.4B, C**). It has many similar features found on CnaB-type folds like <sup>N</sup>SpaA, such as an AB loop and the location of the reactive lysine residue found on the last  $\beta$ G loop.<sup>24</sup> As expected, K139 is found on the surface as a solvent-exposed lysine. Heteronuclear NOE data suggest that the structure is well-ordered with the exception of the N and C-terminal tails and the AB Loop spanning residues 38 to 50 (**Figure 4.4D**). This loop is similar to the one found on <sup>N</sup>SpaA and other CnaB folded pilins subunits, but it is notably positioned further away from the SpaB's K139. We have previously shown in Chapter 2 that the AB loop on <sup>N</sup>SpaA is mobile and flexible when not bound to a CWSS, but becomes rigid and ordered in the peptide-bound complex.

A DALI analysis was performed to identify structural homologues to SpaB.<sup>25</sup> The results are shown in **Table 4.4**. Generally, a DALI Z-score above 2 indicates a similar fold, whereas scores  $> \sim 5$  indicate a much higher-level structural homology. Our analysis reveals that SpaB shares only limited structural homology to proteins in the data base with the most closely related structural homologs having Z-scores ranging from 4.5-6.5. Interesting, all but one of the closet structural homologs correspond to the N-terminal domain of a pilin proteins that serve as substrates for a sortase enzyme. Moreover, all of the homologous structures adopt CnaB type fold, but differ radically in loop positioning causing lower structural homology scores. There are five previously solved basal pilin structures: RrgC (PDB: 4QQ1) from *Streptococcus pneumonia*

that contains three-domains FctB (PDB: 3KLQ) from *Streptococcus pyogenes*, GBS52 (PDB:3PHS) from *Streptococcus agalactiae*, *Lactobacillus rhammosus* GG's SpaB (PDB:7CBS) and SpaE (6JCH).<sup>25-29</sup> Of these only two (PDB:7CBS and 6JCH) are most structurally homologous to SpaB. Collectively, our results indicate that there is great variability the in the structures basal subunits within Gram-positive pili.

Current understanding of lysine recognition in pilin subunits is thought to be dependent on two key factors found in SpaA: (1) the presence of the pilin motif, WxxxVxVYPK, a highly conserved sequence of amino acids, and (2) the AB loop which mediates the ternary complex. SpaB only contains the conserved proline and lysine residue found in the pilin motif.<sup>14,30</sup> Based on our solved structure, the residues that would typically correspond to the pilin motif in SpaB do not appear to line any significant groove region found in <sup>N</sup>SpaA. <sup>N</sup>SpaA's AB loop is believed to help mediate the linkage between pilin subunits by aiding in the identification of <sup>N</sup>SpaA's K190 residue and stabilizing the structure. At this time, it is currently unclear if SpaB's AB loop is positioned in a way that allows for a similar complex to occur. The loop is just as flexible, but does not appear to be able to position itself over K139 in a similar fashion to <sup>N</sup>SpaA's AB Loop. This difference may be part of the reason for why SpaB's lysine is less reactive than <sup>N</sup>SpaA and could be part of how SpaB unproductively binds to <sup>Cd</sup>SrtA inhibiting the enzyme. Future work will be required into investigating and seeing if mutations to the AB loop result in dramatically alter activity.

#### SpaB does not need the pilin motif for lysine recognition

Previous studies have found that many pilin subunits that form lysine isopeptide bonds contain a highly conserved pilin motif.<sup>14,31</sup> This pilin motif contains the lysine residue which acts as a nucleophile for the transpeptidation reaction and has been previously thought to be important for recognition.<sup>20,32</sup> Structural data of <sup>N</sup>SpaA found that the pilin motif lines a nonpolar

groove region in <sup>N</sup>SpaA which the LPLTG peptide sits above (**Figure 4.5A**). Residues within the pilin motif have been shown in Chapter 2 to interact with the CWSS and mediate its binding between this groove region. Mutations that target the pilin motif and residues within the groove in <sup>N</sup>SpaA were previously found to disrupt catalytic activity both *in vitro* and *in vivo*.<sup>30</sup> SpaB notably lacks the conserved residues found in other *C. diphtheriae* pilins such as SpaA, SpaH, and SpaD, and the structure does not appear to contain a groove region (**Figure 4.5B** and **Figure 4.4C**). The lack of groove appears to be due to a different positioning of  $\beta$ F and  $\beta$ G which are spaced further apart from one another compared to <sup>N</sup>SpaA. Further analysis of the electrostatics of SpaB and <sup>N</sup>SpaA show that <sup>N</sup>SpaA contains a positive charge that lines the groove all the way to K190 (**Figure 4.5C**).<sup>33</sup> Areas proximal to K139 appear to also contain this positively charged surface, but with negatively charged regions closer than on <sup>N</sup>SpaA. Similar analysis of <sup>Cd</sup>SrtA reveal those surface areas close to C222 contain a highly negative charge (PDB: 5K9A).<sup>20</sup> These opposing electrostatic charges between pilin subunits and <sup>Cd</sup>SrtA may aid in the recognition the lysine residue.

Despite SpaB's differences to <sup>N</sup>SpaA, SpaB is able to form the lysine isopeptide bond with K139. Therefore, we attempted to see if the insertion of the pilin motif into SpaB would cause any changes in activity with <sup>Cd</sup>SrtA. To test this hypothesis, we inserted the missing conserved residues into SpaB's pilin motif region and created a SpaB tetramutant, SpaB<sup>4M</sup> (T130W, M134V, L136V, and R137Y). We then tested the reactivity of SpaB and SpaB<sup>4M</sup> (300 $\mu$ M) with <sup>Cd</sup>SrtA $\Delta$  (100  $\mu$ M) and <sup>C</sup>SpaA (300 $\mu$ M) and took timepoints at 0, 24, 48, and 72 hours. The reactions were stopped by the addition of an SDS-containing buffer. Comparing SpaB and SpaB<sup>4M</sup>, there does not appear to be any difference between the two SpaB variants. (**Figure 4.5D**) We then wanted to see if there would be a noticeable difference when the LPLTG peptide was used instead of <sup>C</sup>SpaA, so we incubated SpaB and SpaB<sup>4M</sup> (300  $\mu$ M) with <sup>Cd</sup>SrtA $\Delta$  (100  $\mu$ M) and LPLTG peptide (300  $\mu$ M). (**Figure 4.5E**). These reactions were incubated for 72

hours, after which quantification by HPLC appeared to show that only 40% of SpaB or SpaB<sup>4M</sup> was catalyzed into SpaB-LPLT product. These results show that the pilin motif which was found to be important in SpaA, does not affect the formation of the isopeptide bond between SpaB and the CWSS when conserved residues are inserted into SpaB.

#### SpaB incorporation throughout the pilus

Previous data from the Ton-That lab showed that in addition to SpaB's traditional role as the terminating pilin, SpaB subunits can be found throughout the length of the pilus.<sup>14,34</sup> We have hypothesized two potential methods of SpaB incorporation throughout the pilus. The first method is the house-keeping sortase enzyme <sup>Cd</sup>SrtF in *C. diphtheriae*, which recognizes SpaB's CWSS (LAFTG). <sup>Cd</sup>SrtF typically links SpaB and the completed pili to cell wall precursors. Because the reaction between SpaB and the cell wall precursors are similar, <sup>Cd</sup>SrtF could potentially ligate SpaB to SpaA's lysine residue and would explain why SpaB is found throughout the pilus. In order to test this, we created a SpaB<sup>FL</sup>-K139A (Q25-L182) construct containing both its CWSS (LAFTG) and lysine knockout (K139A) which would prevent any homotypic reactions with SpaB's lysine. A reaction containing SpaB<sup>FL</sup>-K139A (300 μM), <sup>N</sup>SpaA (300 μM) and either <sup>Cd</sup>SrtA<sup>A</sup>, <sup>Cd</sup>SrtA<sup>3M</sup>, or <sup>Cd</sup>SrtF, (a sortase where residues A57-G273 were expressed and contained an N-terminal His6x tag) (100 μM) were mixed together and incubated at 25°C. **(Figure 4.6A)** Aliquots of the reaction were taken at 0, 24, 48, and 72 hours were taken and quenched by the addition of SDS-containing sample buffer. Samples were separated out on an SDS-PAGE gel. Analysis of the gel did not show any formation of <sup>N</sup>SpaA- SpaB<sup>FL</sup>-K139A complex. **(Figure 4.6B)** As of the writing of this thesis, it remains unclear as to why we do not see any activity and we propose that future experiments in trying to uncover this mechanism test the activity of <sup>Cd</sup>SrtF with peptides instead of the full SpaB pilin subunit.



The second potential method for how SpaB is incorporated throughout the pilus is through the ligation of SpaB-K139 to SpaA's E446 residue or nearby residues. Previous data from the Ton-That lab show *in vivo* that SpaA residue E446 is essential for the insertion of SpaC and SpaB into the pilus.<sup>14</sup> E446 is part of a conserved sequence motif known as the E-box motif (YxLxETxAPxGY) which contains a highly conserved glutamate. Mutations of E446 to an alanine or arginine prevent SpaB insertion into the pilus *in vivo*.<sup>14</sup> Structural studies of SpaA suggest that E446 helps to catalyze an internal isopeptide bond between K363 and N482 (**Figure 4.7A**). It is inferred that the structural stability created by this internal isopeptide bond is key to the insertion of SpaB and SpaC into the pilus.<sup>14</sup> To further study the E-box motif, we created and expressed a <sup>C</sup>SpaA-E446A mutant and tested it *in vitro* by reacting it with <sup>N</sup>SpaA and SpaB and SpaB<sup>FL</sup> (**Figure 4.7B**). Interestingly, SpaB and SpaB<sup>FL</sup> were able to react with <sup>C</sup>SpaA-E446A according to our SDS-PAGE results. This data seems to contradict the previous *in vivo* data, so we wanted to see if the inability for ligation to SpaB is related to <sup>C</sup>SpaA stability. To do so, we used an SYPRO Orange thermostability assay to test the overall stability of <sup>C</sup>SpaA with and without the E446A mutation (**Figure 4.7C, D**). <sup>C</sup>SpaA-E446A had two separate T<sub>m</sub> values at 56.2 ± 0.4 °C and 74.6 ± 0.2 °C while <sup>C</sup>SpaA only had one at 74.4 ± 0.2 °C. There are two potential possibilities for why <sup>C</sup>SpaA-E446A has two separate unfolding events. The first is that a portion of <sup>C</sup>SpaA-E446A first unfolds before the entire domain unfolds. The second is that there are two populations of <sup>C</sup>SpaA-E446A where some form an internal isopeptide bond between K363 and N482 and some do not. Whatever the reason, the instability or lack of internal isopeptide bond, may prevent SpaA's CWSS from being ligated to SpaB as it disrupts <sup>Cd</sup>SrtA's ability to recognize SpaB's lysine residue. Our experiments unfortunately have not solved the mystery of how SpaB is inserted throughout the pilus. Future experiments that focus on <sup>Cd</sup>SrtF may be the key to understanding how SpaB is incorporated.

## Conclusions

This work lays the foundation for further understanding the termination of pilin polymerization. Using our recently developed HPLC method and structural biology methods, we have further elucidated the termination mechanisms behind the formation of the SpaA pilus. This work reveals that the <sup>Cd</sup>SrtA sortase has a lower  $k_{cat}$  when SpaB is ligated to a LPLTG peptide compared to <sup>N</sup>SpaA. Through the use of LC-MS, we determined that sortase's slowest step remains the formation of the <sup>Cd</sup>SrtA-LPLT acyl-intermediate. Our work also showed us that <sup>Cd</sup>SrtA preferentially links SpaB's lysine residue over <sup>N</sup>SpaA's. SpaB can also bind to <sup>Cd</sup>SrtA in a non-productive manner that prevents <sup>N</sup>SpaA from binding. This posits a potentially new method of pilus polymerization length control where high concentrations of SpaB modulate the activity of <sup>Cd</sup>SrtA, resulting in pili of a specific length.

By solving the structure of SpaB, we have found that despite a large difference in primary sequence, SpaB has a similar topology to <sup>N</sup>SpaA. This topological similarity may be the primary reason why <sup>Cd</sup>SrtA recognizes both SpaA and SpaB's lysine residues, as SpaB does not contain nor require the pilin motif which is conserved in many other pilin subunits. Furthermore, our experiments show that insertion of the pilin motif into SpaB showed no changes in the reactivity of SpaB with <sup>C</sup>SpaA or LPLTG peptide. This warrants further work to uncover how <sup>Cd</sup>SrtA recognizes K139 and if the AB loop, which differs in SpaB, is involved in this recognition mechanism.

Despite our efforts to understand how SpaB is inserted throughout the pilus instead of solely at the base, it remains unclear how this occurs. Our initial attempts to use <sup>Cd</sup>SrtF as the catalyst in ligating SpaB's LAFTG CWSS to <sup>N</sup>SpaA's lysine have found no evidence of the formation of a <sup>N</sup>SpaA-SpaB dimer. We are unsure if this lack of dimer formation is due to <sup>Cd</sup>SrtF being inactive *in vitro*, or if <sup>Cd</sup>SrtF is unable recognize <sup>N</sup>SpaA's lysine residue. We then investigated to see if SpaA's conserved E box residue E446 is involved in SpaB's insertion into

the pilus by mutating E447 in a  $^{13}\text{C}$ SpaA construct. Our results show that mutation of SpaA's E446 does not prevent ligation to SpaB *in vitro* which differs from previously reported *in vivo* data.<sup>14</sup> We reason that this discrepancy from the *in vivo* results is most likely due to structural disruptions from either the lack one of the internal isopeptide bonds. Our work sets the foundation for understanding pilin polymerization termination and may aid in the fight against gram-positive bacteria.

#### **4.4 Materials and Methods**

##### Expression of protein reagents

All protein constructs were cloned and expressed using standard methods as described in (20).  $^{13}\text{C}$ SpaA constructs contain an N-terminal 6xHis tag in a pMAPLE4 expression vector. All other protein constructs contained an N-terminal small ubiquitin-like modifier (SUMO) solubility domain with a 6xHis tag in the pSUMO expression vector. Proteins were expressed in *Escherichia coli* BL21 (DE3) cells. Cultures were grown in miller LB broth (Thermo Fisher) and were supplemented with 50  $\mu\text{g}/\text{mL}$  of kanamycin sulfate (GoldBio), and cultures were grown at 37°C in a shaking incubator to the log growth phase ( $\text{OD}_{600}$  of 0.6 to 0.8) before being induced with 1 mM isopropyl  $\beta$ -d-1-thiogalactopyranoside (IPTG) (GoldBio). Cultures were then moved to low temperature orbital shakers and incubated at 17°C for 16 hours. For isotopically labeled samples, cell pellets were exchanged into M9 media supplemented with  $^{13}\text{C}$ -labeled glucose and  $^{15}\text{N}$ -labeled ammonium chloride before being induced with IPTG. Cells were then pelleted by spinning in a centrifuge at  $\sim 8,500 \times g$  for 15 min. Pellets were then redissolved in 50 mM Tris-HCl and 300 mM NaCl, at pH 8.0 (Buffer A). Subsequently, the cells were lysed using high-pressure emulsification and then fractionated via centrifugation at  $\sim 23,000 \times g$  for 50 min. Afterward, the cell lysate was purified via IMAC- $\text{Co}^{2+}$  purification. Proteins were eluted from the resin using a lysis buffer supplemented with 200 mM imidazole. The 6xHis-SUMO tags were removed via treatment by 6xHis-Ulp1 protease at 1 mg/mL and subsequent purification by

IMAC-Co<sup>2+</sup>. Afterward, the protein was purified by size exclusion chromatography using an AKTA Pure FPLC system (GE) and a column packed with Superdex 75 PG resin. Protein purity was confirmed by SDS-PAGE. pE-SUMO expression plasmids were created using standard molecular biology methods and confirmed by nucleotide sequencing. All purified enzymes were stored at -20 °C in buffer A supplemented with 40% glycerol. The FELPLTGGSG peptide (LPLTG peptide) used in the transpeptidation and hydrolysis assays was synthesized by Peptide 2.0.

### SDS-PAGE Transpeptidation Assay

Gel-based assays were conducted in a similar manner to references 20 and 23. In this assay, <sup>C</sup>SpaA containing a CWSS was ligated to either SpaB or <sup>N</sup>SpaA by <sup>Cd</sup>SrtA<sup>Δ</sup> and reactions were separated by SDS-PAGE. Reactions were performed in 100 μL volumes and contained 100 μM of <sup>Cd</sup>SrtA<sup>Δ</sup>, DTT (5 mM), 200 μM of <sup>C</sup>SpaA, and 200 μM of SpaB or <sup>N</sup>SpaA. All components were dissolved in buffer A, and reactions were initiated by adding the <sup>Cd</sup>SrtA<sup>Δ</sup>. Reactions were conducted at 25 °C and timepoints were taken at 0, 24, 48, and 72 hours. Samples were then separated by SDS-PAGE.

Samples containing <sup>Cd</sup>SrtF were conducted in a similar manner as above except that <sup>Cd</sup>SrtF was used instead of <sup>Cd</sup>SrtA<sup>Δ</sup> and <sup>C</sup>SpaA was replaced with a SpaB construct containing a CWSS (LAFTG) and a K139A mutation.

### HPLC transpeptidation assay

An HPLC-based assay was developed to quantify the kinetics of <sup>Cd</sup>SrtA when SpaB is used as a substrate. In this assay, SpaB containing the reactive lysine is ligated to a FELPLTGGSG peptide (LPLTG peptide) by <sup>Cd</sup>SrtA<sup>Δ</sup> and reactions were quantified using a Phenomenex C4 column (5 μm, 4.6 x 150 mm). Reactions were performed in 100 μL volumes

and contained 25  $\mu\text{M}$  of  $^{125}\text{I}$ -SrtA<sup>A</sup>, DTT (5 mM), 1 mM of LPLTG peptide, and variable amounts of SpaB (62.5 to 500  $\mu\text{M}$ ). All components were dissolved in buffer A, and reactions were initiated by adding the  $^{125}\text{I}$ -SrtA<sup>A</sup>, incubated for 3 h at 25 °C and then flash-frozen with liquid N<sub>2</sub> and stored at -20 °C. The reactions were analyzed using a Phenomenex C4 column (5  $\mu\text{m}$ , 4.6  $\times$  150 mm) and with an initial dwell time of 3 min at 30% acetonitrile/0.1% TFA followed by a linear gradient from 30% to 40% CH<sub>3</sub>CN/0.1% TFA for 23 min at 1 mL/min. The column was subsequently flushed with high concentrations of CH<sub>3</sub>CN/0.1% TFA at 1 mL/min. SpaB containing peaks were detected at 215 nm and the amount of substrate converted to product was calculated by integrating the area under the HPLC traces. The identity of each peak in the HPLC chromatogram was confirmed via LC-MS. All HPLC experiments were performed on an Agilent 1100 HPLC. Reactions' kinetic parameters were obtained by fitting the data with SigmaPlot 12.0.

#### Acyl-Intermediate Detection using LC-MS

LC-MS reactions to compare acyl formation with and without the presence of SpaB and <sup>15</sup>N-SpaA were performed in buffer A and contained a total volume of 100  $\mu\text{L}$ : sortase (25  $\mu\text{M}$ ), DTT (5 mM), LPLTG (1 mM), and <sup>15</sup>N-SpaA (250  $\mu\text{M}$ ) or buffer. All reactions were incubated at 25 °C and 10  $\mu\text{L}$  time points were removed and frozen with liquid N<sub>2</sub> before being stored at -20 °C. An experiment containing only the enzyme (no peptide added) was performed to provide an external standard for the amount of unmodified enzyme in the mass spectrum. Because the  $^{125}\text{I}$ -SrtA-LPLT acyl intermediate cannot readily be separated from unmodified  $^{125}\text{I}$ -SrtA by chromatography, its amount in the assay is estimated from the mass spectrometry data by assuming that the acyl-intermediate and unmodified enzyme ionize to similar extents. Before being run on the LC-MS system, samples were diluted with 90  $\mu\text{L}$  of 200 mM L-tryptophan (internal standard) and then measured on a Zorbax 300SB-C3 (3.5  $\mu\text{m}$ , 3.0  $\times$  150 mm) with an

Agilent 6530 QTOF and Agilent 1260 Infinity HPLC with a gradient of 30– 99% of water to acetonitrile over 6 min at 0.8 mL/min. The data was analyzed with Agilent MassHunter Qualitative Analysis. The amount of enzyme and acyl-intermediate was calculated by integrating each peak.

### Competition experiment

Reactions were conducted in 100  $\mu$ L volumes with 1 mM FELPLTGGSG, 25  $\mu$ M  $^{Cd}SrTA^{\Delta}$ , 200  $\mu$ M either  $^{N}SpaA$ , SpaB-APNV, and/or SpaB-K139A, 5 mM DTT in buffer A. Reaction samples were taken after 1, 3, 6, and 24 hours and quenched by freezing at -20  $^{\circ}$ C to be analyzed by HPLC. To analyze the samples, a Phenomenex C4 column (5  $\mu$ m, 4.6  $\times$  150 mm) was used with a 3-minute dwell time of 30% CH<sub>3</sub>CN/0.1% TFA and subsequent linear gradient from 30% CH<sub>3</sub>CN/0.1% TFA to 44% CH<sub>3</sub>CN/0.1% TFA applied over 28 minutes with a flow rate of 1 mL/min. The column was then flushed with high concentrations of CH<sub>3</sub>CN/0.1% TFA to remove any remaining protein. Peaks corresponding to  $^{N}SpaA$  and SpaB were detected at 215 nm and were quantified by integration.

### NMR structural determination

Structure determination followed a similar protocol to reference 30. NMR spectra were collected at 298 K on Bruker Avance III HD 600 MHz and Avance NEO 800MHz spectrometers equipped with triple resonance probes. NMR data were processed with NMRPipe (35), and analyzed using CARA (36) (version 1.8.4), XIPP (37) (version 1.19.6 p0), and NMRFAM-Sparky (38).  $^1H$ ,  $^{13}C$ , and  $^{15}N$  protein chemical shifts were assigned using the following experiments:  $^{15}N$ -HSQC,  $^{13}C$ -HSQC, HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HBHA(CO)NH, HNHA, HNHB, CC(CO)NH, H(CCCO)NH, HCCH-COSY, HCCH-TOCSY, and  $^{15}N$ -TOCSY.

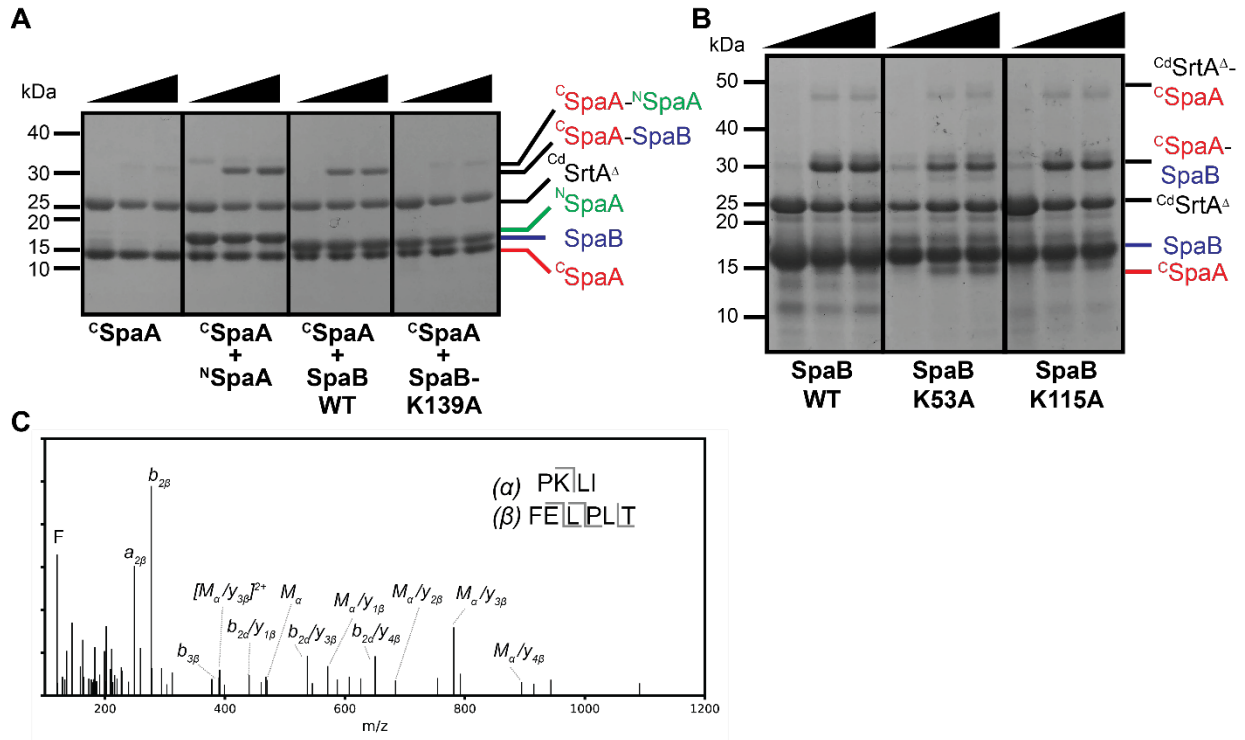
### Identification of the isopeptide bond in SpaB by mass spectrometry

Protein digestion and isopeptide bond identification were performed according to previous protocols (39). SpaB reactions were prepared in volumes of 100  $\mu\text{L}$  by incubating 200  $\mu\text{M}$  of SpaB, 100  $\mu\text{M}$  of  $^{13}\text{C}_6\text{SrtA}^{\text{A}}$ , 4 mM of the LPLTG peptide and 5 mM DTT in buffer A. Samples were allowed to incubate at 25°C for 48 hours before a 10  $\mu\text{L}$  sample was taken. Reactions were then evaluated as described in the HPLC transpeptidation reaction to ensure that ligation had taken place. Samples were then refrozen in -20°C. Mass spectrometry samples were analyzed similar to reference 20.

### Differential Scanning Fluorimetry

The melting temperatures were extracted from Differential Scanning Fluorimetry data by a method described previously (30). Briefly, proteins were diluted to 100  $\mu\text{M}$  in buffer A, supplemented with 15x SYPRO Orange (Sigma-Aldrich) at a total volume of 20  $\mu\text{L}$ . Thermal denaturation reactions were run on a CFX Connect qPCR system (Bio-Rad). A heating rate of 0.2 °C/min was employed from 4 to 95 °C, and fluorescence measurements (excitation at 525  $\pm$  10 nm, detection at 570  $\pm$  10 nm) were acquired after each 0.5 °C step increase. The melting temperature for each protein was determined by the first derivative method after averaging the three replicate measurements. The  $T_m$  is defined as the midpoint of the transition from folded to unfolded state and is identified spectroscopically as the temperature where the rate of fluorescence increases with respect to temperature is greatest. The  $T_m$  was then used to calculate the equilibrium constant of unfolding, as previously described (40).

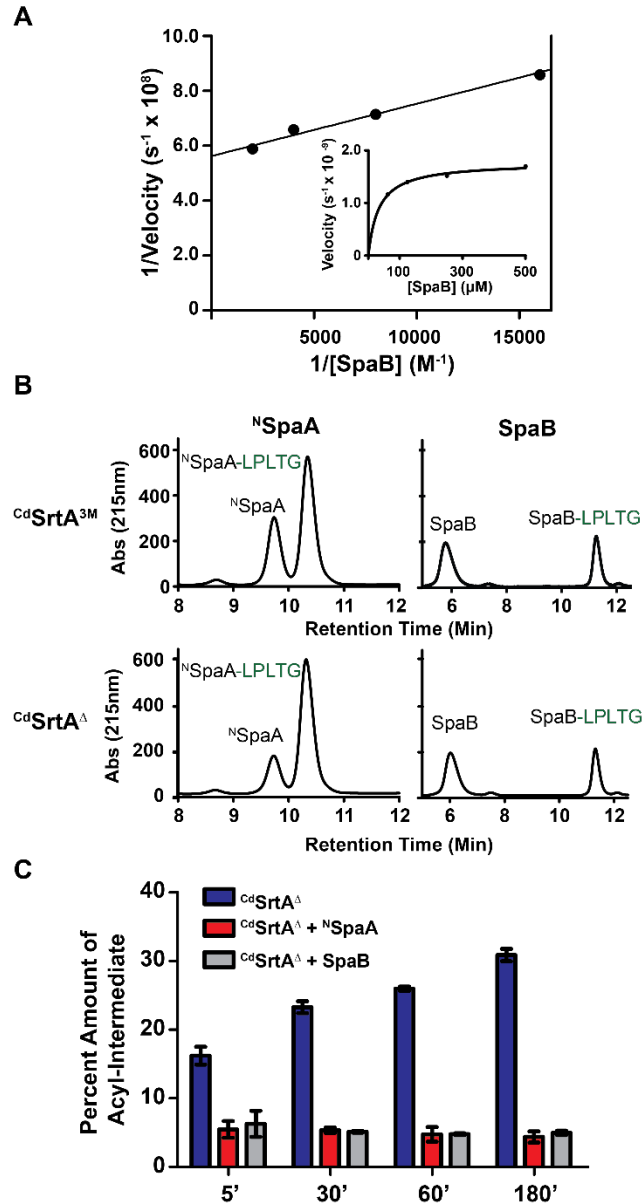
## 4.5 Figures



Note that hours are 0, 24, and 48 for all gels

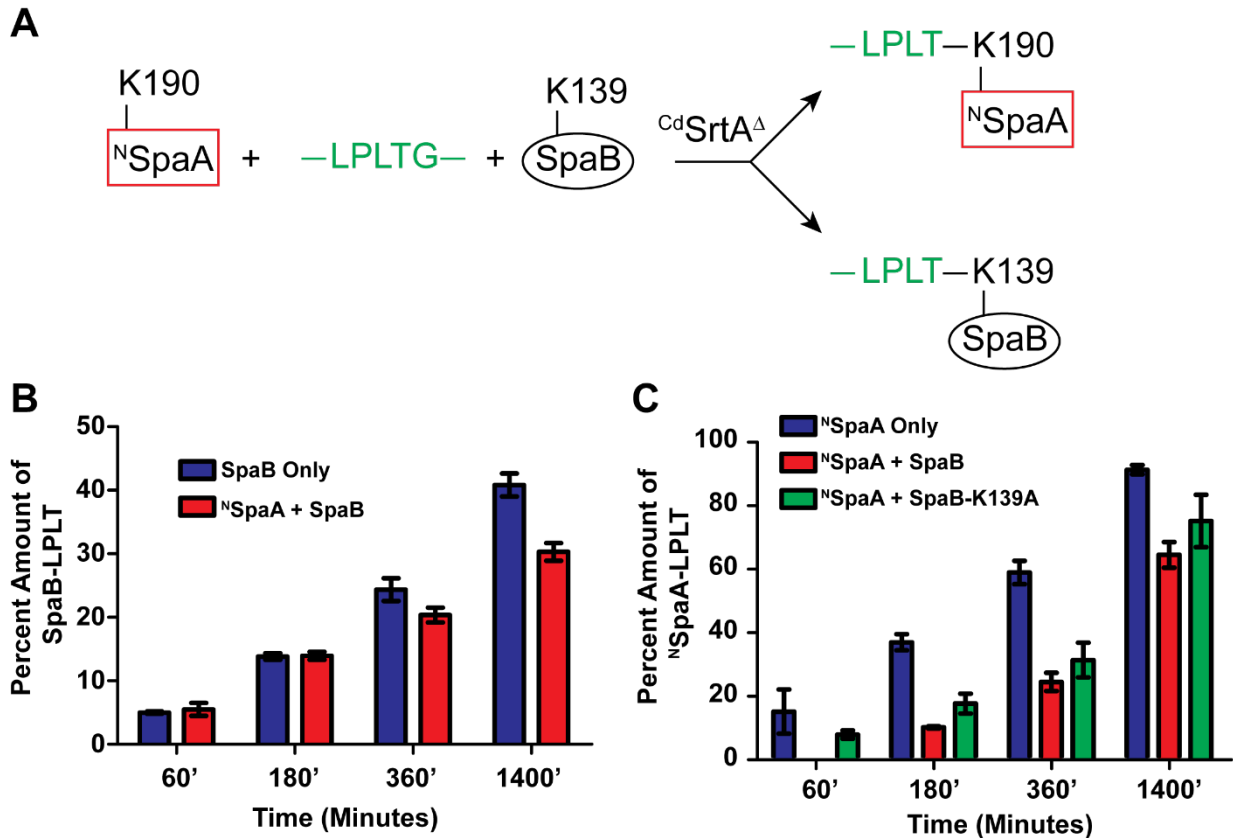
**Figure 4.1** *In vitro* testing of lysine specificity of SpaB (A) *In vitro* reconstitution of SpaB and <sup>N</sup>SpaA reaction to <sup>C</sup>SpaA. When K139 of SpaB was mutated to alanine (K139A), formation of <sup>C</sup>SpaA-SpaB was abrogated. The reaction samples were carried out at 25°C and analyzed by SDS/PAGE and blue silver staining after 0, 24, and 48 hours of incubation. (B) Reactions of additional SpaB lysine mutants K53A and K115A with <sup>C</sup>SpaA show complex formation similarly to wild-type SpaB. (C) The isopeptide bond between the SpaB K139 and the LPLTG peptide was identified by LC-MS/MS. Data shown is the m/z tandem mass spectrum of the linked peptide to SpaB K139 (sequence shown in insert).



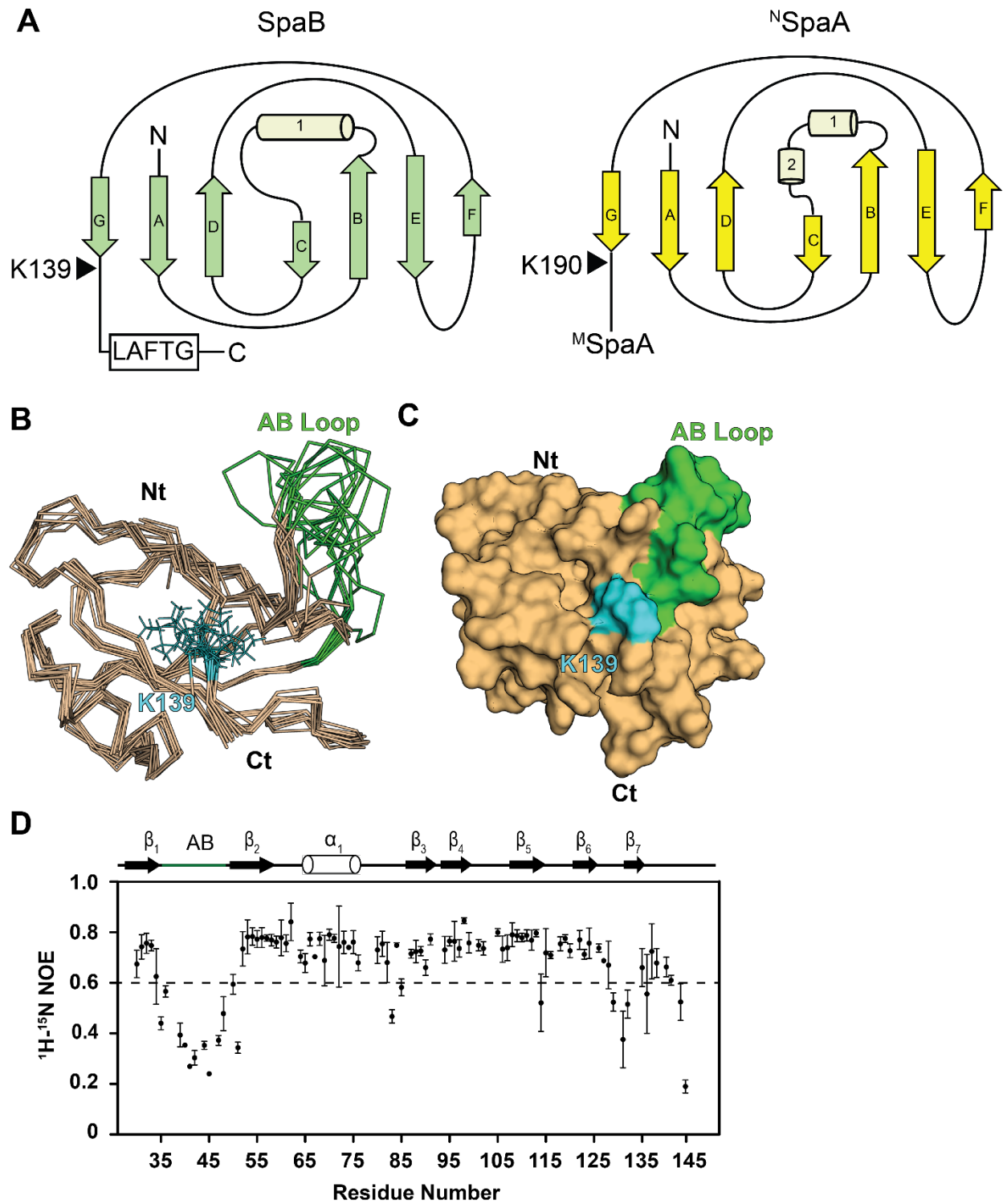


**Figure 4.2 Measurement of activity  $\text{CdSrtA}$  of SpaB.** (A) A Lineweaver-Burk plot graphing the initial velocities versus the increasing concentrations of SpaB from 62.5  $\mu\text{M}$  to 500  $\mu\text{M}$ .  $k_{\text{cat}}$  and  $K_{\text{M}}$  were estimated from a linear trend line fit. ( $n=3$ ) Insert depicts a velocity vs substrate graph. (B) A representative trace showing the overall product formation of  $^{\text{N}}\text{SpaA}$  versus SpaB when  $\text{CdSrtA}^{3\text{M}}$  or  $\text{CdSrtA}^{\Delta}$  are used after 24 hours. The percent amount of SpaB modified after 24 hours is approximately 60-70%. (C) A comparison of the amount of acyl-intermediate formed with  $\text{CdSrtA}^{\Delta}$  in the presence of the LPLTG peptide or in the presence of the LPLTG peptide and

<sup>N</sup>SpaA or SpaB. (n=3) When <sup>Cd</sup>SrtA<sup>Δ</sup> is incubated with LPLTG peptide, a complex of <sup>Cd</sup>SrtA<sup>Δ</sup>-LPLT peptide complex forms and is detectable with LC-MS. In the absence of <sup>N</sup>SpaA or SpaB, <sup>Cd</sup>SrtA<sup>Δ</sup>-LPLT concentration increases over time to more than 30%. When <sup>N</sup>SpaA or SpaB are present <sup>Cd</sup>SrtA<sup>Δ</sup>-LPLT concentration remains below 5%.

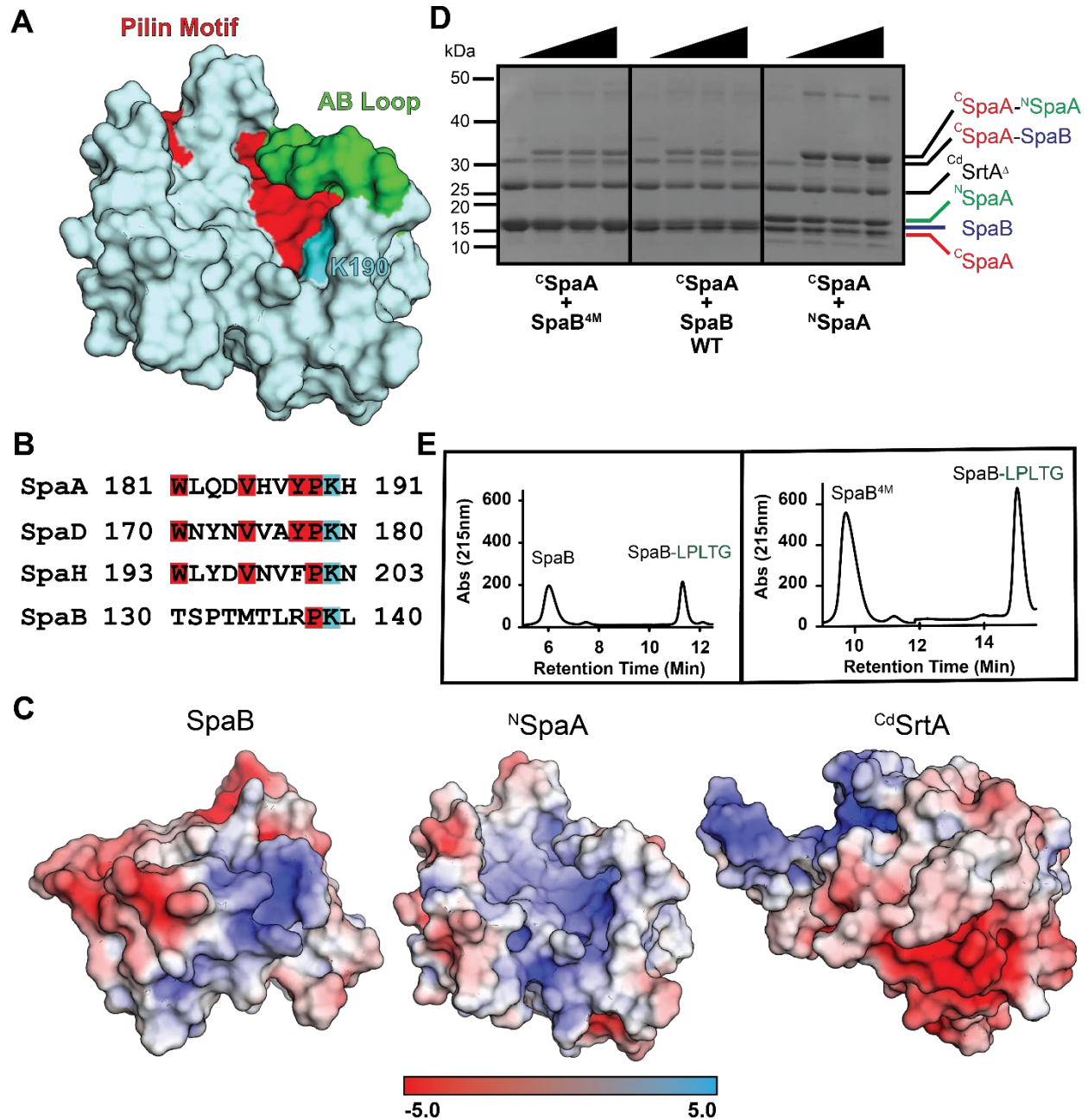


**Figure 4.3 Competition Reaction between <sup>N</sup>SpaA and SpaB with <sup>Cd</sup>SrtA<sup>Δ</sup>.** (A) A diagram depicting the competition reaction between <sup>N</sup>SpaA (200 μM) and SpaB (200 μM) with LPLTG peptide (1 mM) and <sup>Cd</sup>SrtA<sup>Δ</sup> (25 μM). <sup>Cd</sup>SrtA<sup>Δ</sup> catalysis results in either the formation of <sup>N</sup>SpaA-LPLT linked at its lysine residue (K190) or SpaB-LPLT linked at its lysine (K139). (B) The amount of SpaB-LPLT in the absence (blue) or presence (red) of <sup>N</sup>SpaA over a period of 24 hours. (n = 3) The amount of SpaB-LPLT when <sup>N</sup>SpaA is present does not appear to differ until after 3 hours. (C) The amount of <sup>N</sup>SpaA-LPLT in the absence (blue), presence of SpaB (red) or SpaB-K139A (green) over a period of 24 hours. <sup>N</sup>SpaA-LPLT is not detected during the first hour of reaction when SpaB was present. After 3 hours, <sup>N</sup>SpaA-LPLT product formation is 3.8 times lower when SpaB is present, and 2.1 times lower when SpaB-K139A is present.



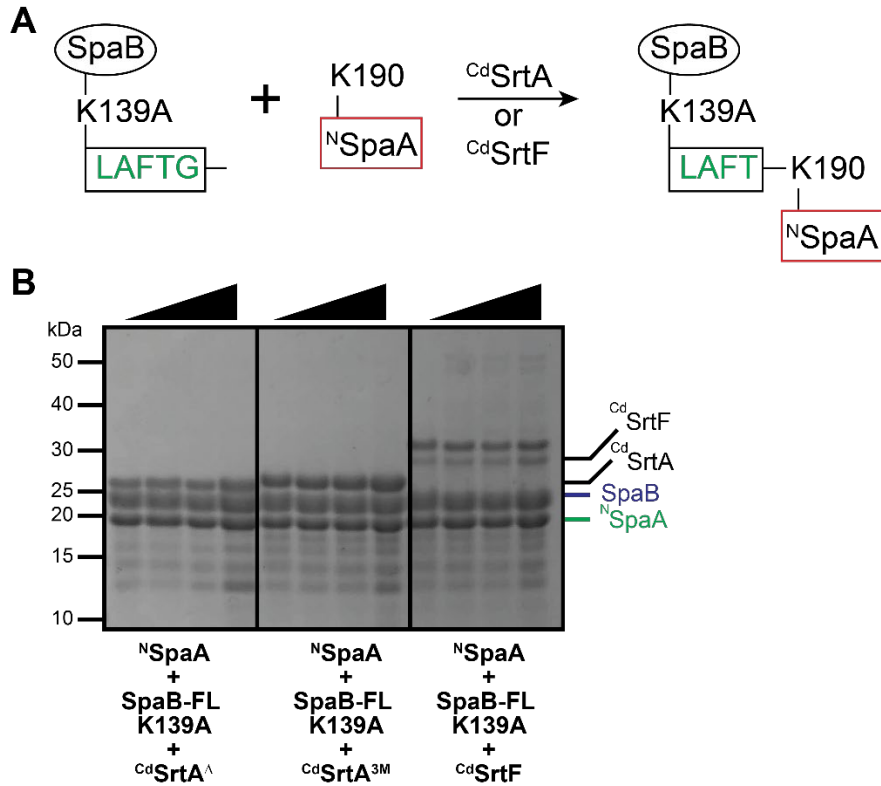
**Figure 4.4 Structure of the SpaB pilin.** (A) Schematic representation of the fold topology of SpaB compared to the fold topology of  $^{\text{N}}\text{SpaA}$ .  $\beta$  strands are designated A through G and the lone  $\alpha$ -helix is labeled as 1. (B) A bundle of the 20 lowest energy structures of SpaB. Tan color

represents the backbone of the SpaB, and the reactive lysine residue K139 is colored cyan. The green color highlights the AB loop located between  $\beta$ -sheet stands A and B. The SpaB's AB loop appears perpendicular to the LPLT peptide. (C) One of the SpaB structures represented in a surface view. (D) Heteronuclear NOE data is graphed as a function of primary sequence for SpaB. Flexible residues are located below the dashed lined at 0.6. The AB loop residues (38 to 50) are found beneath the 0.6 value and are therefore thought to be highly flexible.



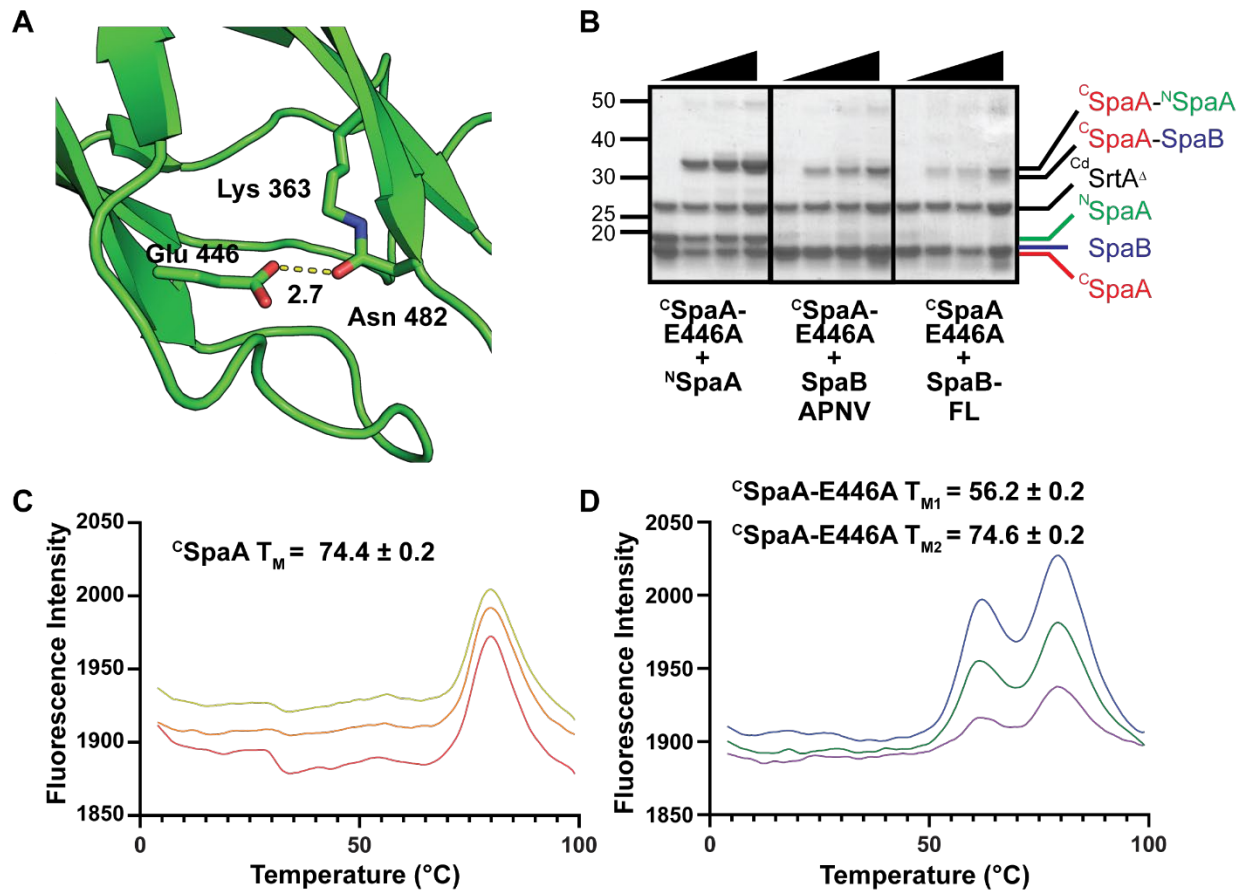
**Figure 4.5 Biochemical study of the pilin motif in SpaB.** (A) The NMR structure of <sup>N</sup>SpaA (PDB: 7K7F). The pilin motif is colored in red and lines the groove region that is formed by  $\beta$ F and  $\beta$ G. The AB loop is colored green and covers the reactive lysine. (B) Multiple sequence alignment of the pilin motif of SpaA, SpaD, SpaH and SpaB pilin proteins from *C. diphtheriae*. Residues that have been found to be conserved are highlighted in red and the reactive lysine

residue is highlighted in blue. (C) The electrostatic surface potentials were calculated with APBS Electrostatics from one of the SpaB structures, <sup>N</sup>SpaA (PDB:7K7F), and <sup>Cd</sup>SrtA (PDB:5K9A). The groove region found in <sup>N</sup>SpaA is lined by a highly positive charge. SpaB does contain a positively charged region proximal to K139, but it is notably missing any groove region. Partial negative charge regions are also found where a groove would be located in SpaB. <sup>Cd</sup>SrtA contains a negative charge region proximal to the C222. (D) Reactions of <sup>C</sup>SpaA with SpaB, SpaB<sup>4M</sup>, and <sup>N</sup>SpaA were run on an SDS-PAGE gel with timepoints taken at 0, 24, 48, and 72 hours. There does not appear to be any major difference between SpaB with and without the pilin motif. (E) Reactions containing SpaB and SpaB<sup>4M</sup> were incubated at 25°C for 72 hours. Samples were then taken and injected onto the HPLC. SpaB<sup>4M</sup> showed no major difference in the amount of SpaB-LPLT product formed compared to samples without the pilin motif.



**Figure 4.6 SpaB does not appear to ligate to <sup>N</sup>SpaA through its CWSS.** (A) A schematic depicting how SpaB may get ligated to <sup>N</sup>SpaA through SpaB's CWSS. (B) Samples attempting to ligate SpaB's CWSS with <sup>N</sup>SpaA with three different sortases. Timepoints were taken at 0, 24, 48, and 72 hours and run on an SDS-PAGE gel. There did not appear to any peaks that corresponded to a ligation between <sup>N</sup>SpaA and SpaB<sup>FL</sup>.





**Figure 4.7 Investigating how mutation to the <sup>c</sup>SpaA E-box affects SpaB incorporation.** (A) The internal isopeptide bond of SpaA between K363 and N482 and the E446 residue (PDB: 3HR6). (B) Reactions of <sup>c</sup>SpaA-E446A with <sup>N</sup>SpaA, SpaB, and SpaB<sup>FL</sup> were run on an SDS-PAGE Gel with timepoints taken at 0, 24, 48, and 72 hours. <sup>c</sup>SpaA-E446A was able to react with all three substrates. (C) Three traces of fluorescence intensity versus temperature of <sup>c</sup>SpaA. It has one main inflection at  $74.4 \pm 0.2$  °C (D) Three traces of fluorescence intensity versus temperature of <sup>c</sup>SpaA-E446A. There appears to have two separate inflections with two separate thermal shifts at  $56.2 \pm 0.4$  °C and  $74.6 \pm 0.2$  °C.

## 4.6 Tables

**Table 4.1. Kinetics of <sup>Cd</sup>SrtA<sup>Δ</sup> Catalysis between Pilin Subunits**

|                                | $k_{cat} \times 10^{-5} \text{ (s}^{-1}\text{)}$ | $K_M \times 10^{-5} \text{ (s}^{-1}\text{)}^a$ | $k_{cat} \times 10^{-5} \text{ (s}^{-1}\text{)}^b$ | $k_{cat}/K_M \text{ (s}^{-1} \text{ M}^{-1}\text{)}$ |
|--------------------------------|--|--|--|--|
| <sup>N</sup> SpaA <sup>c</sup> | 40 ± 0.1   | 1.6 ± 0.4                                      | 20 ± 0.4   | 2.5 ± 0.6  |
| SpaB <sup>e</sup>              | 7.0 ± 0.4  | 0.32 ± 0.02                                    | n/a <sup>d</sup>                                   | 2.1 ± 0.1  |

All kinetics are approximations.

<sup>a</sup> Refers to transpeptidation kinetics when either <sup>N</sup>SpaA or SpaB concentrations are varied and the FELPLTGGSG concentration was held constant at 1 mM.

<sup>b</sup> Refers to transpeptidation kinetics when the FELPLTGGSG concentration was varied and <sup>N</sup>SpaA was held constant at 500 μM.

<sup>c</sup>Values are reported from Sue *et al* (2020)

<sup>d</sup>n.d., not determined because SpaB concentration was held constant when FELPLTGGSG concentration was varied.

<sup>e</sup>Transpeptidation steady-state kinetics parameters for <sup>Cd</sup>SrtA<sup>Δ</sup> were determined by the monitoring rate at which the enzyme ligated an FELPLTGGSG peptide to the SpaB via a lysine-iso peptide bond.

**Table 4.2. SYPRO Orange Thermostability of Sortase Proteins**

|                          | CdSrtA <sup>WT</sup> | CdSrtA <sup>2M</sup> | CdSrtA <sup>3M</sup> | CdSrtA <sup>Δ</sup> |
|--------------------------|----------------------|----------------------|----------------------|---------------------|
| Melting Temperature (°C) | 44.0 ± 0.5           | 37.6 ± 0.2           | 36.7 ± 0.3           | 26.8 ± 0.3          |

**Table 4.3. Structural statistics of the solution structure of SpaB**

|  | $\langle SA \rangle^a$ |
|--|------------------------|
| Root mean square deviation                           |                        |
| NOE Interproton distance restraints (Å) <sup>b</sup> |                        |
| All (1277)   | 0.036 ± 0.003          |
| Dihedral angles restraints (°) <sup>c</sup> (192)    | 0.73 ± 0.08            |
| Deviation from idealized covalent geometry           |                        |
| bonds (Å)  | 0.0110 ± 0.00004       |
| angles (°)   | 0.73 ± 0.01            |
| impropers (°)  | 0.036 ± 0.02           |
| PROCHECK results (%) <sup>d</sup>                    |                        |
| most favorable region                                | 71.5 ± 2.8             |
| additionally allowed region                          | 24.4 ± 3.5             |
| generously allowed region                            | 4.01 ± 0.7             |
| disallowed region                                    | 0.0 ± 0.0              |
| Coordinate Precision (Å) <sup>e</sup>                |                        |
| Protein backbone                                     | 0.56 ± 0.06            |
| Protein heavy atoms                                  | 1.51 ± 0.35            |

<sup>a</sup>  $\langle SA \rangle^a$  represents an ensemble of the 20 best structures calculated by simulated annealing. The number of terms for each restraint is given in parentheses. None of the structures exhibited distance violations greater than 0.5 Å, dihedral angle violations greater than 5°, coupling constant violations greater than 2 Hz

<sup>b</sup> Distance restraints: 410 sequential, 131 medium ( $2 \leq$  residue separation  $\leq 4$ ) and 412 long range ( $>4$  residues apart)

<sup>c</sup> The experimental dihedral angle restraints were as follows: 95  $\phi$ , and 97  $\psi$ , angular restraints

<sup>d</sup> PROCHECK-NMR1 data includes residues 28-142 of SpaB. For the structured regions of the protein and peptide,  $98 \pm 0.1$  % of the residues were in the favored or allowed regions of the Ramachandran plot

<sup>e</sup> The coordinate precision is defined as the average atomic root mean square deviation (rmsd) of the 20 individual SA structures and their mean coordinates. These values are for residues 28-142 of SpaB. Backbone atoms refers to the N, C $^\alpha$ , and C' atoms

**Table 4.4 DALI Analysis of SpaB**

|    | Chain | Z-score | RMSD | Alignment Length (aa) | No. of Residues (aa) | Sequence Identity (%) | PDB  | Description  |
|----|-------|---------|------|-----------------------|----------------------|-----------------------|------|--|
| 1  | A     | 6.4     | 4.1  | 101                   | 449                  | 17                    | 3uxf | Structure of the fimbrial protein FimP from <i>Actinomyces oris</i>  |
| 2  | A     | 6.3     | 3.2  | 99                    | 470                  | 17                    | 4p0d | Structure of the T6 backbone pilin of serotype M6 <i>Streptococcus pyogenes</i>                              |
| 3  | A     | 5.8     | 3.8  | 92                    | 143                  | 14                    | 7k7f | Structure of the <i>Corynebacterium diphtheriae</i> SpaA Pilin-Signal Peptide Complex                        |
| 4  | A     | 5.8     | 3.6  | 93                    | 132                  | 17                    | 7cbs | Structure of SpaB basal pilin from <i>Lactobacillus rhamnosus</i> GG   |
| 5  | A     | 5.1     | 3.3  | 70                    | 522                  | 21                    | 6fwv | Structure of the <i>Bacillus anthracis</i> TIE protein   |
| 6  | A     | 5       | 3.2  | 92                    | 331                  | 11                    | 4oq1 | Structure of the <i>Ancillary Pilin RrgC</i> of <i>Streptococcus pneumoniae</i>                              |
| 7  | E     | 5       | 3.4  | 73                    | 202                  | 11                    | 4fxt | Structure of a DUF3823 family protein from <i>Bacteroides ovatus</i> ATCC                                    |
| 8  | A     | 4.8     | 3.6  | 75                    | 242                  | 16                    | 4eiu | Structure of a DUF3823 family protein (BACUNI_03093) from <i>Bacteroides uniformis</i> ATCC 8492             |
| 9  | B     | 4.5     | 3.4  | 80                    | 161                  | 13                    | 4gqz | Structure of the periplasmic copper-binding protein CueP from <i>Salmonella enterica</i> serovar Typhimurium |
| 10 | A     | 4.5     | 4.4  | 96                    | 327                  | 14                    | 6jch | Structure of SpaE basal pilin from <i>Lactobacillus rhamnosus</i> GG   |

## 4.7 References

- (1) Ramirez, N. A.; Das, A.; Ton-That, H. *Trends Microbiol.* **2020**, *28*, 999–1009.
- (2) Kline, K. A.; Dodson, K. W.; Caparon, M. G.; Hultgren, S. J. *Trends Microbiol.* **2010**, *18*, 224–232.
- (3) Danne, C.; Dramsi, S. *Res. Microbiol.* **2012**, *163*, 645–658.
- (4) Telford, J. L.; Barocchi, M. A.; Margarit, I.; Rappuoli, R.; Grandi, G. *Nat. Rev. Microbiol.* **2006**, *4*, 509–519.
- (5) Van Gerven, N.; Waksman, G.; Remaut, H. *Prog. Mol. Biol. Transl. Sci.* **2011**, *103*, 21–72.
- (6) Bradshaw, W. J.; Davies, A. H.; Chambers, C. J.; Roberts, A. K.; Shone, C. C.; Acharya, K. R. *FEBS J.* **2015**, *282*, 2097–2114.
- (7) Dufrière, Y. F.; Viljoen, A. *Front. Microbiol.* **2020**, *11*, 1457.
- (8) Manetti, A. G. O.; Zingaretti, C.; Falugi, F.; Capo, S.; Bombaci, M.; Bagnoli, F.; Gambellini, G.; Bensi, G.; Mora, M.; Edwards, A. M.; Musser, J. M.; Graviss, E. A.; Telford, J. L.; Grandi, G.; Margarit, I. *Mol. Microbiol.* **2007**, *64*, 968–983.
- (9) Becherelli, M.; Manetti, A. G. O.; Buccato, S.; Viciani, E.; Ciucchi, L.; Mollica, G.; Grandi, G.; Margarit, I. *Mol. Microbiol.* **2012**, *83*, 1035–1047.
- (10) Nallapareddy, S. R.; Singh, K. V.; Sillanpää, J.; Garsin, D. A.; Höök, M.; Erlandsen, S. L.; Murray, B. E. *J. Clin. Invest.* **2006**, *116*, 2799–2807.
- (11) Mandlik, A.; Swierczynski, A.; Das, A.; Ton-That, H. *Trends Microbiol.* **2008**, *16*, 33–40.
- (12) Ton-That, H.; Schneewind, O. *Mol. Microbiol.* **2003**, *50*, 1429–1438.

- (13) Gaspar, A. H.; Ton-That, H. *J. Bacteriol.* **2006**, *188*, 1526–1533.
- (14) Ton-That, H.; Marraffini, L. A.; Schneewind, O. *Mol. Microbiol.* **2004**, *53*, 251–261.
- (15) Mandlik, A.; Das, A.; Ton-That, H. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 14147–14152.
- (16) Mandlik, A.; Swierczynski, A.; Das, A.; Ton-That, H. *Mol. Microbiol.* **2007**, *64*, 111–124.
- (17) Ton-That, H.; Schneewind, O. *Trends Microbiol.* **2004**, *12*, 228–234.
- (18) Chang, C.; Mandlik, A.; Das, A.; Ton-That, H. *Mol. Microbiol.* **2011**, *79*, 1236–1247.
- (19) Jacobitz, A. W.; Kattke, M. D.; Wereszczynski, J.; Clubb, R. T. *Adv. Protein Chem. Struct. Biol.* **2017**, *109*, 223–264.
- (20) Chang, C.; Amer, B. R.; Osipiuk, J.; McConnell, S. A.; Huang, I. H.; Hsieh, V.; Fu, J.; Nguyen, H. H.; Muroski, J.; Flores, E.; Loo, R. R. O.; Loo, J. A.; Putkey, J. A.; Joachimiak, A.; Das, A.; Clubb, R. T.; Ton-That, H. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E5477–E5486.
- (21) Mandlik, A.; Das, A.; Ton-That, H. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 14147–14152.
- (22) Sue, C. K.; McConnell, S. A.; Ellis-Guardiola, K.; Muroski, J. M.; McAllister, R. A.; Yu, J.; Alvarez, A. I.; Chang, C.; Ogorzalek Loo, R. R.; Loo, J. A.; Ton-That, H.; Clubb, R. T. *Bioconjug. Chem.* **2020**, *31*, 1624–1634.
- (23) McConnell, S. A.; Amer, B. R.; Muroski, J.; Fu, J.; Chang, C.; Ogorzalek Loo, R. R.; Loo, J. A.; Osipiuk, J.; Ton-That, H.; Clubb, R. T. *J. Am. Chem. Soc.* **2018**, *140*, 8420–8423.
- (24) Hae, J. K.; Paterson, N. G.; Gaspar, A. H.; Ton-That, H.; Baker, E. N. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 18427.

- (25) Megta, A. K.; Pratap, S.; Kant, A.; Palva, A.; von Ossowski, I.; Krishnan, V. *Curr. Res. Struct. Biol.* **2020**, *2*, 229–238.
- (26) Linke, C.; Young, P. G.; Kang, H. J.; Bunker, R. D.; Middleditch, M. J.; Caradoc-Davies, T. T.; Proft, T.; Baker, E. N. *J. Biol. Chem.* **2010**, *285*, 20381–20389.
- (27) Krishnan, V.; Gaspar, A. H.; Ye, N.; Mandlik, A.; Ton-That, H.; Narayana, S. V. L. *Structure* **2007**, *15*, 893–903.
- (28) Megta, A. K.; Mishra, A. K.; Palva, A.; von Ossowski, I.; Krishnan, V. *J. Struct. Biol.* **2019**, *207*, 74–84.
- (29) Shaik, M. M.; Maccagni, A.; Tourcier, G.; Di Guilmi, A. M.; Dessen, A. *J. Biol. Chem.* **2014**, *289*, 16988–16997.
- (30) McConnell, S. A.; McAllister, R. A.; Amer, B. R.; Mahoney, B. J.; Sue, C. K.; Chang, C.; Ton-That, H.; Clubb, R. T. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*.
- (31) Ton-That, H.; Faull, K. F.; Schneewind, O. *J. Biol. Chem.* **1997**, *272*, 22285–22292.
- (32) Khare, B.; V. L. Narayana, S. *Protein Sci.* **2017**, *26*, 1458–1473.
- (33) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; et al. *Protein Sci.* **2018**, *27*, 112–128.
- (34) Ton-That, H.; Schneewind, O. *Mol. Microbiol.* **2003**, *50*, 1429–1438.
- (35) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293.
- (36) Keller, R. The Computer Aided Resonance Assignment Tutorial. Cantina Verlag: Goldau



Switzerland 2004.

- (37) Garrett, D. S.; Cai, M.; Clore, G. M. *J. Biomol. NMR* **2020**, *74*, 9–25.
- (38) Lee, W.; Tonelli, M.; Markley, J. L. *Bioinformatics* **2015**, *31*, 1325–1327.
- (39) Chang, C.; Wu, C.; Osipiuk, J.; Siegel, S. D.; Zhu, S.; Liu, X.; Joachimiak, A.; Clubb, R. T.; Das, A.; Ton-That, H. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 18041–18049.
- (40) Nettleship, J. E.; Brown, J.; Groves, M. R.; Geerlof, A. *Methods Mol. Biol.* **2008**, *426*, 299–318.

## **Chapter 5**

### **The Search for Sortase Inhibitors**

## 5.1 Overview

Antibiotic resistant gram-positive bacteria are one of the many growing threats to modern medicine. One of the most dangerous of these bacteria is the methicillin resistant strain of *Staphylococcus aureus* which is one of the main hospital-acquired infections and causes around 11,000 deaths a year. Therefore, there is a vital need to develop new antimicrobial agents that can combat these infections. A promising target for therapeutic development is the Sortase A (<sup>Sa</sup>SrtA) enzyme in *Staphylococcus aureus*, which appends protein virulence factors to its cell wall. Knockout of <sup>Sa</sup>SrtA has resulted in the inability for the bacteria to establish animal infections. In this chapter, I describe my contribution to a collaborative effort to discover sortase inhibitors. First, I will talk about my contribution to the optimization of pyridazinone-based sortase inhibitors led by Dr. Albert Chan. Starting from a previously discovered pyridazinone scaffold, we developed a soluble analog inhibitor. This analog inhibitor was incubated with the <sup>Sa</sup>SrtA enzyme and we determined the solution structure of <sup>Sa</sup>SrtA bound to the pyridazinone inhibitor. Using this structural data and *in silico* simulations, we created several analogs and tested their ability to inhibit sortase with a Förster resonance energy transfer (FRET) assay. One of these analogs showed a ~70-fold increase in activity compared to the original lead molecule. My contribution to these efforts was to measure the half maximal rate of inhibition of these compounds by a FRET assay.


In the second half of Chapter Five, I discuss my contribution to a novel high throughput screen that utilized a cell-based screening approach led by Jason Gossachalk. This method involved the use of *Actinomyces oris*, a Gram-positive bacterium that colonizes the human oral cavity and generates biofilm layers. *A. oris* utilizes a SrtA ( <sup>Ao</sup> SrtA) which anchors glycosylated surface protein A (GspA) to the cell wall. Recently, the Ton-That lab discovered that the viability of *A. oris* was dependent on the activity of  <sup>Ao</sup> SrtA and it is believed that cell death was caused due to the accumulation of glycosylated GspA in the membrane. As of this writing, *A. oris* is

believed to be the only known organisms that exhibits any sortase-dependent growth phenotype. Utilizing this phenotype, we developed a cell-based assay that screened for sortase inhibitors with wild type *A. oris* and a  $\Delta srtA/\Delta gspA$  strain. Small molecules were screened against the WT strain and inhibition growth defects were then counter screened against the  $\Delta srtA/\Delta gspA$  strain. Subsequent hits were then screened against an *in vitro* half maximal inhibitor compound against <sup>Ao</sup>SrtA and <sup>Sa</sup>SrtA enzymes and protein display in *A. oris* cells. My contributions involved running the preliminary screen and the counter screen and testing the hit validation method for both the IC<sub>50</sub> against the *Staphylococcus aureus* <sup>Sa</sup>SrtA FRET assay.

This chapter is reformatted with permission from a peer-reviewed journal articles: NMR structure-based optimization of *Staphylococcus aureus* sortase A pyridazinone inhibitors. Albert H. Chan, Sung Wook Yi, Ethan M. Weiner, Brendan R. Amer, Christopher K. Sue, Jeff Wereszczynski, Carly A. Dillen, Silvia Senese, Jorge Z. Torres, J. Andrew McCammon, Lloyd S. Miller, Michael E. Jung, and Robert T. Clubb. Chemical Biology and Drug Design. February 2017. and A Cell-based Screen in *Actinomyces oris* to Identify Sortase Inhibitor. Jason E. Gosschalk, Chungyu Chang, Christopher K. Sue, Sara D. Siegel, Chenggang Wu, Michele D. Kattke, Sung Wook Yi, Robert Damoiseaux, Michael E. Jung, Hung Ton-That and Robert T. Clubb. Scientific Reports. May 2020.

## **5.2.1 NMR structure-based optimization of *Staphylococcus aureus* sortase A pyridazinone inhibitors**

## NMR structure-based optimization of *Staphylococcus aureus* sortase A pyridazinone inhibitors

Albert H. Chan<sup>1,2,3</sup>  | Sung Wook Yi<sup>1</sup> | Ethan M. Weiner<sup>1,2,3</sup> | Brendan R. Amer<sup>1,2,3</sup> | Christopher K. Sue<sup>1</sup> | Jeff Wereszczynski<sup>4</sup> | Carly A. Dillen<sup>5</sup> | Silvia Senese<sup>1</sup> | Jorge Z. Torres<sup>1</sup> | J. Andrew McCammon<sup>6,7,8</sup> | Lloyd S. Miller<sup>5</sup> | Michael E. Jung<sup>1</sup> | Robert T. Clubb<sup>1,2,3</sup>

<sup>1</sup>Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, USA

<sup>2</sup>UCLA-DOE Institute of Genomics and Proteomics, University of California, Los Angeles, Los Angeles, CA, USA

<sup>3</sup>Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA, USA

<sup>4</sup>Department of Physics and Center for Molecular Study of Condensed Soft Matter, Illinois Institute of Technology, Chicago, IL, USA

<sup>5</sup>Department of Dermatology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

<sup>6</sup>Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, USA

<sup>7</sup>Howard Hughes Medical Institute, University of California, San Diego, La Jolla, CA, USA

<sup>8</sup>Department of Pharmacology, University of California, San Diego, La Jolla, CA, USA

### Correspondence

Robert T. Clubb and Michael E. Jung, Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, USA.  
Emails: rclubb@mbi.ucla.edu; jung@chem.ucla.edu

### Funding information

National Institute of Allergy and Infectious Diseases, Grant/Award Number: AI52217; National Institute of General Medical Sciences, Grant/Award Number: GM008496 and GM31749

*Staphylococcus aureus* is a leading cause of hospital-acquired infections in the USA and is a major health concern as methicillin-resistant *S. aureus* and other antibiotic-resistant strains are common. Compounds that inhibit the *S. aureus* sortase (SrtA) cysteine transpeptidase may function as potent anti-infective agents as this enzyme attaches virulence factors to the bacterial cell wall. While a variety of SrtA inhibitors have been discovered, the vast majority of these small molecules have not been optimized using structure-based approaches. Here we have used NMR spectroscopy to determine the molecular basis through which pyridazinone-based small molecules inhibit SrtA. These inhibitors covalently modify the active cysteine thiol and partially mimic the natural substrate of SrtA by inducing the closure of an active site loop. Computational and synthetic chemistry methods led to second-generation analogues that are ~70-fold more potent than the lead molecule. These optimized molecules exhibit broad-spectrum activity against other types of class A sortases, have reduced cytotoxicity, and impair SrtA-mediated protein display on *S. aureus* cell surface. Our work shows that pyridazinone analogues are attractive candidates for further development into anti-infective agents, and highlights the utility of employing NMR spectroscopy and solubility-optimized small molecules in structure-based drug discovery.

### KEYWORDS

molecular docking, molecular dynamics, NMR, protein structure, protein-inhibitor complex, Sortase, SrtA, *Staphylococcus aureus*, transpeptidase

**Abbreviations:** AA EK, aryl ( $\beta$ -amino)ethyl ketone; Ba-SrtA, *Bacillus anthracis* sortase A; Ba-SrtB, *Bacillus anthracis* sortase B; MRSA, methicillin-resistant *Staphylococcus aureus*; SrtA, *Staphylococcus aureus* sortase A.

## 1 | INTRODUCTION

*Staphylococcus aureus* is a leading cause of hospital- and community-acquired infections in the USA. This bacterial pathogen is estimated to cause 11,000 deaths annually in the USA. It inflicts a wide range of life-threatening diseases such as pneumonia, meningitis, osteomyelitis, endocarditis, toxic shock syndrome, bacteremia, and sepsis.<sup>[1]</sup> *S. aureus* skin and soft tissue infections also represent a major clinical problem as they result in over 11 million outpatient and emergency room visits, and close to 500,000 hospital admissions per year in the USA.<sup>[2,3]</sup> The rise of methicillin-resistant *S. aureus* (MRSA) is a major health problem that has created a pressing need for new antibiotics. MRSA harbors genes that confer resistance to  $\beta$ -lactam antibiotics and is now endemic in hospitals. *S. aureus* has also developed resistance to other antibiotics, including last resort and newer generation drugs, such as vancomycin, daptomycin, and linezolid.<sup>[4–6]</sup> The *S. aureus* sortase A (SrtA) enzyme is an attractive molecular target for the development of novel antibiotics. This is because this enzyme covalently attaches a range of protein virulence factors to the surface of *S. aureus* that play critical roles in the infection process, including promoting bacterial adhesion to host tissues, acquisition of essential nutrients, and the evasion and suppression of the immune response.<sup>[7,8]</sup> Several studies have shown that *srtA*<sup>−</sup> *S. aureus* mutants have reduced virulence in animal models of infection, presumably because their surfaces are devoid of key protein factors required to colonize host tissue and evade the immune response.<sup>[9,10]</sup> SrtA-related enzymes are also used by other clinically important pathogens to display factors that are required for their virulence (e.g., *Enterococcus faecalis*, *Listeria monocytogenes*, *Bacillus anthracis*, *Streptococcus pyogenes*, and *Streptococcus pneumoniae*).<sup>[10]</sup> Thus, small-molecule SrtA enzyme inhibitors could function as potent anti-infective agents to treat MRSA and infections caused by other types of Gram-positive bacteria.

SrtA resides on the extracellular membrane where it covalently attaches proteins to peptidoglycan by catalyzing a transpeptidation that joins proteins containing a C-terminal cell wall sorting signal (CWSS) to the cross-bridge peptide.<sup>[11]</sup> Besides its important role in pathogenesis, SrtA has several other properties that make it an attractive drug target: (i) It has no human homolog, reducing the likelihood of off-target effects, (ii) it is located on the bacterial surface such that inhibitors do not need to cross the cell membrane, thereby reducing potential cytotoxicity, and (iii) it is not required for the growth of *S. aureus* and other clinically important microbes when they are outside their human host.<sup>[12]</sup> Therefore, SrtA inhibitors that selectively target virulence mechanisms could have a distinct advantage over conventional antibiotics, as they may not induce the same selective pressures that lead to drug resistance.<sup>[13,14]</sup>

Several research groups have sought to identify small-molecule SrtA inhibitors that could be developed into therapeutics.<sup>[10,15]</sup> The small molecules that have thus far been identified include natural products and synthetic molecules identified by screening compound libraries, and rationally designed molecules that mimic the substrate or transition state intermediates.<sup>[10]</sup> In addition, virtual screening methods have been employed to identify inhibitors using the structure of the apo- or substrate-bound forms of the enzyme.<sup>[10]</sup> However, a clinically useful SrtA inhibitor has yet to be developed. A major obstacle hindering drug development has been the difficulty in applying structure-based methods to optimize SrtA inhibitors. This is because the active site of SrtA is structurally disordered in its apo-state making it difficult to model drug–enzyme interactions computationally and potentially hindering the application of X-ray crystallography that require crystallization of the SrtA–inhibitor complex. At present, only Zhulenkov and colleagues have used NMR spectroscopy to visualize how SrtA binds to an inhibitor, a benzisothiazolinone-based small molecule that irreversibly modifies the enzyme.<sup>[16]</sup> However, the structure of the inhibitor–SrtA complex was determined at low resolution. Maresso, et al. have also structurally characterized covalent inhibitors of the aryl ( $\beta$ -amino)ethyl ketone class by co-crystallizing the inhibitors with the sortase B from *Bacillus anthracis* (Ba-SrtB).<sup>[17]</sup> However, Ba-SrtB may not be a good platform from which to optimize *S. aureus* SrtA inhibitors, as SrtA and Ba-SrtB have distinct active site structures and recognize different sorting signal substrates; SrtA and Ba-SrtA recognize LPXTG and NPQTN sorting signals, respectively.<sup>[18]</sup> Thus, while many compounds inhibit the activity of SrtA in vitro, a lack of structural data to rationally optimize these molecules for clinical applications has been problematic.

Using high-throughput screening (HTS) methods, we previously identified pyridazinone-based molecules that are potent inhibitors of SrtA.<sup>[19]</sup> However, these molecules were not optimized using structure-based approaches because they were poorly soluble in water, which made it difficult to experimentally determine the three-dimensional structure of the enzyme–inhibitor complex, and because conformational disorder in the SrtA's active site made it difficult to computationally model enzyme–inhibitor interactions. We have now overcome this problem using a highly soluble pyridazinone analogue whose complex with SrtA could be determined using NMR spectroscopy. We show that pyridazinone inhibitors partially mimic the natural substrate, by inducing a conformational change in the active site that incompletely closes a key active site loop, resulting in a final structure that is reminiscent of both the apo- and sorting signal-bound forms of the enzyme. Guided by computational studies of the complex, a series of pyridazinone analogues were synthesized and their efficacy evaluated in vitro and in cell culture. This

work led to the discovery of 2-(3-fluorophenyl)-4-(3-hydroxypropoxy)-5-mercaptopyridazin-3(2H)-one, which inhibits SrtA-mediated protein display in *S. aureus* and improves inhibitory activity over the lead molecule 70-fold ( $IC_{50}$  value of  $21 \pm 14$  nM). The use of a solubility-optimized inhibitor analogue for NMR studies of the complex was critical and may be a generally useful strategy to study other enzyme-inhibitor interactions.

## 2 | METHODS AND MATERIALS

### 2.1 | Preparation of the SrtA-inhibitor complex for NMR and mass spectrometry studies

The catalytic domain of SrtA (residues Gln<sup>60</sup>-Lys<sup>206</sup>) was prepared as described previously.<sup>[20]</sup> To form a SrtA:inhibitor complex, 0.5 mM uniformly <sup>15</sup>N- and <sup>13</sup>C-labeled SrtA was incubated with 5 mM of sodium thiolate 4-ethoxy-5-mercapto-2-phenylpyridazin-3(2H)-one (hereafter referred as **2-salt**) in 50 mM Tris, pH 6.4, 150 mM NaCl, 20 mM CaCl<sub>2</sub> for 72 hr at room temperature. The complex was then dialyzed into NMR buffer that consisted of 50 mM Tris, pH 6.4, 150 mM NaCl, and 20 mM CaCl<sub>2</sub>. Two NMR samples were studied that each contained 1.5 mM SrtA-inhibitor complex dissolved in NMR buffer and either 7% or 99.999% D<sub>2</sub>O. Mass spectrometry was used to verify that the inhibitor forms a disulfide bond with Cys<sup>184</sup> in SrtA. The SrtA:**2-salt** complex was first digested with trypsin by incubating 20 μM <sup>15</sup>N/<sup>13</sup>C-SrtA:**2-salt** complex with 5 μg/ml trypsin for 24 hr at 37°C. The digestion reaction mixture was then split into two aliquots. To one of the aliquots, DTT was added to a final concentration of 5 mM. Both aliquots were then subjected to LC-MS analysis to monitor the masses of the cleaved peptides. LC-MS experiments were carried out on a Waters Acquity UPLC connected to a Waters LCT-Premier XE Time of Flight Instrument controlled by MassLynx 4.1 software (Waters Corp., Milford, MA, USA). Digested protein samples in 20 mM HEPES and 5 mM CaCl<sub>2</sub> were separated using an Phenomenex Hydro-RP column (3.0 × 50 mm, 4 μm packing) and were eluted with a gradient of 2%–50% solvent B over 10 min using a flow rate of 0.425 ml/min (solvent A: water, solvent B: acetonitrile, both with 0.3% formic acid). The mass spectrometer was equipped with a multimode source operated in the electrospray mode. Mass spectra were recorded from a mass of 70–2,000 Da. Capillary voltage was set to 1700 V and the source/desolvation gas temperatures were 120°C/350°C, respectively. The ion abundance values for product ions at *m/z* 819.5, 942.5, 1638, and 1886 were monitored by generating extracted ion chromatograms using a 0.5-Da mass window and integrating the peaks of interest. These ions had retention time values of 5.09, 6.86, 5.09, and 6.86 min, respectively.

### 2.2 | NMR spectroscopy and structure determination

NMR spectra of the SrtA:inhibitor complex were acquired at 298 K on Bruker Avance 500-, 600-, and 800-MHz spectrometers equipped with triple resonance cryogenic probes. NMR spectra were processed using NMRPipe<sup>[21]</sup> and analyzed using the PIPP<sup>[22]</sup> and CARRA (version 1.8.4)<sup>[23]</sup> software packages. Chemical shift assignments (<sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N) of SrtA were obtained by analyzing the following experiments: HNCA, HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, <sup>15</sup>N-edited TOCSY, HNHA, HNHB, HBHA(CO)NH, HCCH-TOCSY, HCCH-COSY, (HB)CB(CGCDCE)HE, and (HB)CB(CGCD)HD (reviewed in<sup>[24,25]</sup>). Chemical shift assignments for the inhibitor were obtained by analyzing a two-dimensional (F2) <sup>13</sup>C-filtered NOESY spectrum. The majority of φ and ψ dihedral angle restraints were obtained using the program TALOS+.<sup>[26]</sup> Additional backbone φ angle restraints were obtained by analyzing HNHA spectra.<sup>[27]</sup> Intramolecular protein distance restraints were obtained from three-dimensional <sup>15</sup>N- and <sup>13</sup>C-edited NOESY spectra. Intermolecular distance restraints were obtained by analyzing two-dimensional (F2) <sup>13</sup>C-filtered NOESY and <sup>13</sup>C-edited NOESY-HSQC spectra of the complex.

NOE assignments were initially obtained automatically using the programs ATNOS/CANDID controlled by UNIO<sup>[28,29]</sup> utilizing XPLOR-NIH (version 2.23).<sup>[30]</sup> NOE assignments were then verified manually for the active site residues (Val<sup>161</sup>-Asp<sup>175</sup> [β6/β7 loop], Thr<sup>183</sup>-Lys<sup>196</sup> [β7/β8 loop], and side chains of Leu<sup>97</sup>, Ser<sup>116</sup>, His<sup>120</sup>, Thr<sup>180</sup>, Ile<sup>182</sup>, and Val<sup>201</sup>) by inspecting the NOESY data. During inspection of the NOESY data, additional NOE restraints involving active site residues were identified and included in subsequent structure calculations. Final structures were calculated using XPLOR-NIH based on the lowest energy conformer of the previously solved apo-SrtA NMR structure (PDB 1IJA).<sup>[20]</sup> During the course of the structure calculations, backbone and side chain atoms of active site residues Val<sup>161</sup>-Asp<sup>175</sup> (β6/β7 loop) and Thr<sup>183</sup>-Lys<sup>196</sup> (β7/β8 loop), as well as side chain atoms of Leu<sup>97</sup>, Ser<sup>116</sup>, His<sup>120</sup>, Thr<sup>180</sup>, Ile<sup>182</sup>, Arg<sup>197</sup>, Ile<sup>199</sup>, and Val<sup>201</sup>, were allowed to move while all other SrtA atoms were held fixed in space. Residues within the β6/β7 and β7/β8 loops were allowed to move because they form the surface surrounding the presumed inhibitor binding site, and previous crystal and NMR structural studies of apo- and substrate (LPAT\*)-bound SrtA reveal these loops adopt distinct conformations as a result of substrate binding.<sup>[20,31,32]</sup> Mobility of other active site residues was restricted to their side chains because while they make contacts with the sorting signal substrate in the SrtA:LPAT\* structure, they do not demonstrate significant backbone rearrangement. Non-active site residues were kept rigid because they show little difference (RMSD = 0.8 Å) between the apo- and substrate-bound SrtA



structures, and such structural information is of little value for structure-based drug design of a competitive inhibitor. In the final set of calculations, a thiol group within the inhibitor was attached to Cys<sup>184</sup> through a disulfide bond. A single hydrogen bond between the backbone carbonyl of Gly<sup>167</sup> and the backbone amide of Asp<sup>170</sup> was also used to stabilize the 3<sub>10</sub>-helix within the  $\beta$ 6/ $\beta$ 7 loop, and was substantiated by characteristic NOE patterns from Gly<sup>167</sup>-Asp<sup>170</sup>. A total of 50 structures were calculated, of which 46 had no NOE, dihedral angle, or scalar coupling violations greater than 0.5 Å, 5°, or 2 Hz, respectively. Of these, 20 structures with the lowest overall energy were chosen to represent the structure of the SrtA-inhibitor complex. The programs MOLMOL<sup>[33]</sup> and PYMOL<sup>[34]</sup> were used to generate figures.

### 2.3 | In silico screening of pyridazinone analogues

Ligand preparation, receptor preparation, grid generation, and docking were all conducted with Schrödinger Suite 2011 (Schrödinger LLC, New York, NY, USA). Derivatives of 4-ethoxy-5-mercapto-2-phenylpyridazin-3(2*H*)-one for docking experiments were generated with the CombiGlide application. In CombiGlide, 5-mercapto-2-phenylpyridazin-3(2*H*)-one was defined as the core, and 225, 104, 104, 104, 32, 32, and 22 substituents were created at R2, R3, R4, R5, R6, R7, and R8 sites, respectively (see Figure 4 for definition of the R sites). Substituents were chosen based on the NMR structure of the complex (e.g., polar substituents were added to sites on the inhibitor scaffold that projected into the solvent). The substituents at the R2 site included various polar and non-polar groups of different sizes, as well as a series of substituents with the general formula  $-\text{O}(\text{CH}_2)_n\text{X}$ , where  $n = 1-4$  and X is a polar group. The substituents at R3, R4, and R5 are predominantly polar or positively charged. R6, R7, and R8 contained mostly small hydrophobic substituents. The receptor was processed using the Protein Preparation Wizard, which employs a restrained, partial energy minimization.<sup>[35]</sup> The disulfide bond between SrtA and the inhibitor, and the side chain atoms of Cys<sup>187</sup> were removed to prevent steric clashes during docking. Grids were generated by Glide with the grid box set around the inhibitor using default settings. The inhibitor was excluded in the grid calculations. A docking restraint was set up such that the position of the 5-mercapto sulfur atom of the inhibitor derivatives was restricted within 1 Å of the inhibitor sulfur atom in the NMR structure. Docking was carried out with Glide using XP settings.<sup>[36-38]</sup> Initially, a total of 623 ligands were docked onto the NMR structure of the protein in the SrtA:inhibitor complex (the co-ordinates of **2-salt** were removed prior to docking). After the first round of docking, the best substituents at each site were selected based on two criteria: 1) their docking score was at least one standard

deviation above the mean docking score of the control compound 4-ethoxy-5-mercapto-2-phenylpyridazin-3(2*H*)-one, and 2) the substituent made a specific interaction with the protein (e.g., a hydrogen bond). This reduced the number of substituents at R2, R3, R4, R5, R6, R7, and R8 to 19, 8, 12, 7, 4, 1 (hydrogen), and 4, respectively. CombiGlide was then used to generate 2,688 compounds with all possible combinations of substituents at R3, R4, R5, R6, and R7, with R2 restricted to ethoxy and R8 restricted to hydrogen. These 2,688 compounds were docked to the NMR structure, and the best substituents were selected as described above. This reduced the number of substituents at R2, R3, R4, R5, R6, R7, and R8 to 19, 3, 3, 1 (hydrogen), 4, 1 (hydrogen), and 4, respectively. CombiGlide was then used again to generate 2,736 compounds with all possible combinations of substituents, and the compounds were again docked to the NMR structure of the protein. Using the same selection criteria, 12, 3, 3, 1 (hydrogen), 4, 1 (hydrogen), and 1 (hydrogen) substituents were chosen at R2, R3, R4, R5, R6, R7, and R8, respectively. Finally, CombiGlide was then used to generate 432 compounds that contained all possible combinations of the substituents.

The top 432 molecules were docked to the enzyme using a procedure that accounts for SrtA mobility. In this procedure, protein motion was first simulated using molecular dynamics (MD) calculations. Bond, angle, and torsion parameters for the inhibitor were derived from the generalized Amber force field (GAFF), using the Antechamber program in Amber.<sup>[39,40]</sup> Atomic partial charges were derived from RESP<sup>[41]</sup> fitting of Gaussian 09<sup>[42]</sup> calculated electrostatic potentials at the Hartree-Fock/6-31G\* level. The procedures of MD simulations and clustering were the same as those described in.<sup>[43]</sup> Briefly, a 100-ns conventional MD simulation was performed on the SrtA:inhibitor complex using the AMBER99SB-ILDN force field with the simulation package NAMD.<sup>[44,45]</sup> Eight hundred frames at regularly spaced intervals were extracted from the last 80 ns of the MD simulation. These frames were aligned by the protein C $\alpha$  atoms in the active site and clustered by root mean square deviation (RMSD) conformational clustering using the GROMOS algorithm as implemented in GROMACS 4.5.<sup>[46]</sup> With an RMSD cutoff of 1.40 Å, 25 clusters were obtained, and the centroid member of each cluster was selected to represent each cluster. Subsequently, the top 432 compounds were docked to each of the 25 representative centroid structures, as well as the NMR structure. Procedures used for receptor preparation, grid generation, and docking are the same as those described above. To evaluate the docking results, compounds were ranked by the best docking score they obtained from any of the docking calculations to the NMR or 25 centroid conformers. The top 43 compounds were selected as candidates for synthesis and experimental testing.

## 2.4 | FRET-based enzymatic assay

The ability of the compounds to inhibit the activity of SrtA was determined using an established Förster resonance energy transfer (FRET) assay.<sup>[19,20]</sup> The catalytic domains of SrtA (residues Gln<sup>60</sup>-Lys<sup>206</sup>) and Ba-SrtA (residues Asp<sup>57</sup>-Lys<sup>210</sup>) were prepared as described previously.<sup>[20,47]</sup> Briefly, in the IC<sub>50</sub> assay, 20  $\mu$ l of Sa-SrtA (final assay concentration of 1  $\mu$ M in FRET buffer: 20 mM HEPES, 5 mM CaCl<sub>2</sub>, 0.05% v/v Tween-20, pH 7.5) was incubated with 1  $\mu$ l of test compound solution (dissolved in 100% DMSO, final assay concentration of 0.08–400  $\mu$ M) for 1 hr at room temperature. Subsequently, 30  $\mu$ l of substrate solution, which consists of the self-quenched fluorogenic peptide Abz-LPETG-Dap(Dnp)-NH<sub>2</sub> (32  $\mu$ M final assay concentration) (Peptide 2.0 Inc., Chantilly, VA, USA) dissolved in FRET buffer, was added to the mixture. Fluorescence was read immediately using an Infinite<sup>®</sup> M1000 PRO (Tecan US Inc., Morrisville, NC, USA) plate reader with the excitation and emission wavelengths set at 335 and 420 nm, respectively. The IC<sub>50</sub> assay used to test Ba-SrtA inhibition was similar with the following adjustments: Final assay concentration of Ba-SrtA was 10  $\mu$ M, final assay concentration of substrate was 100  $\mu$ M, and the FRET buffer was composed of 20 mM HEPES, 0.05% v/v Tween-20, pH 7.5. IC<sub>50</sub> values were calculated by fitting three independent sets of data to Equation 1 using SIGMAPLOT 6.0 (SPSS, Inc., Chicago, IL, USA):

$$\frac{v_i}{v_0} = \frac{1}{1 + \left(\frac{[I]}{IC_{50}}\right)^h} \quad (1)$$

where  $v_i$  and  $v_0$  are initial velocity of the reaction in the presence and absence of inhibitor at concentration  $[I]$ , respectively. The term  $h$  is Hill's coefficient.<sup>[48]</sup>

For inhibitors that displayed IC<sub>50</sub> values lower than half of the enzyme concentration used, data were fit to Morrison's quadratic equation (Equation 2) to calculate the apparent dissociation constant ( $K_i^{app}$ ):

$$\frac{v_i}{v_0} = 1 - \frac{([E]_T + [I] + K_i^{app}) - \sqrt{([E]_T + [I] + K_i^{app})^2 - 4[E]_T[I]}}{2[E]_T} \quad (2)$$

where  $v_i$  and  $v_0$  are initial velocity of the reaction in the presence and absence of inhibitor at concentration  $[I]$ , respectively.  $[E]_T$  is the total active enzyme concentration, and  $K_i^{app}$  is the apparent dissociation constant for the enzyme–inhibitor complex.<sup>[48]</sup>

To determine the rate of inhibition of SrtA by compounds 2–17, 2–54, and 2–62, the FRET assay we used for IC<sub>50</sub> or  $K_i^{app}$  determination was modified such that the Abz-LPETG-Dap(Dnp)-NH<sub>2</sub> substrate and inhibitor of various concentrations (final concentration 2.5–14  $\mu$ M) were added to the

enzyme at the same time. Fluorescence was read every 8 s for a total of 30 min. The reaction progress curve was fit to Equation 3 to determine  $k_{obs}$ , the rate constant for conversion from the initial velocity phase to full inhibition:

$$P = \frac{v_i}{k_{obs}} [1 - \exp(-k_{obs}t)] \quad (3)$$

where  $P$  is the baseline-corrected fluorescence value,  $v_i$  is the initial velocity, and  $t$  is time.<sup>[48]</sup> These  $k_{obs}$  measurements were then plotted against inhibitor concentration and fit to Equation 4 to obtain  $k_{inact}$  and  $K_I$  values.

$$k_{obs} = \frac{k_{inact}[I]}{K_I + [I]} \quad (4)$$

where  $k_{obs}$  is the observed rate constant of inhibition at inhibitor concentration  $[I]$ ,  $k_{inact}$  is the maximum rate of inhibition given an infinite concentration of inhibitor, and  $K_I$  is the concentration of inhibitor that yields a half-maximum rate of inhibition.<sup>[48]</sup>

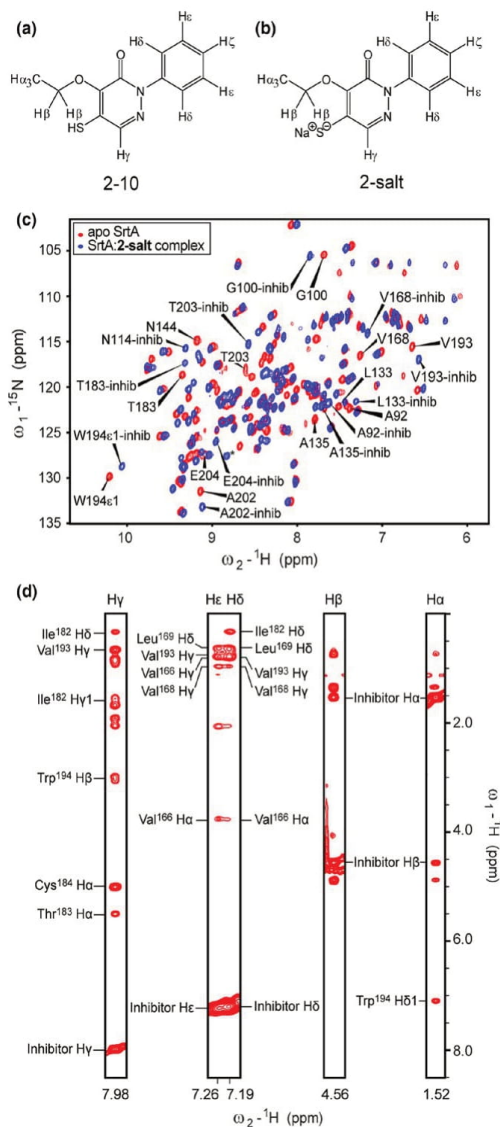
## 2.5 | Cell wall surface Protein A display assay

Protocol for the cell-based assay was adapted from Zhang *et al.*<sup>[49]</sup> Overnight cultures of *S. aureus* were diluted 1:100 into tryptic soy broth (TSB) with or without inhibitor supplement and grown at 37°C with rotation to an A<sub>600</sub> of 0.5. Molecules 2–62 and 2–17 inhibited bacterial growth at 100  $\mu$ M, and therefore, these cultures were only diluted back 1:10 into TSB; data are normalized to account for background; 600  $\mu$ l aliquots were removed, and *S. aureus* were sedimented by centrifugation (12,000 $\times$ g for 5 min). Bacteria were suspended in 600  $\mu$ l PBS, 2  $\mu$ l FITC-labeled IgG (Sigma, St. Louis, MO, USA) was added, and samples were incubated at room temperature for 1 hr. The cells were again sedimented by centrifugation (12,000 $\times$ g for 5 min) and washed twice with PBS. Bacterial fluorescence intensity was monitored using the Flex Station (Molecular Devices, Sunnyvale, CA, USA) at 495-nm excitation and 520-nm emission. Sample aliquots were spread on TSB agar plates, and CFU were determined to derive relative fluorescence units per A<sub>600</sub> unit. Triplicate measurements were taken for each trial, and mean and standard deviation were calculated.

## 2.6 | Cytotoxicity test

The cytotoxicity CC<sub>50</sub> of each compound was determined using Promega CellTiter-Glo Luminescent Cell Viability Assay kit (Promega Corp., Madison, WI, USA) by measuring the total ATP levels to quantify the number of metabolically active cells upon drug treatment as described<sup>[50]</sup> Briefly, the compounds were diluted in 384 plates (20  $\mu$ l/well) in

**FIGURE 1** NMR spectra of SrtA:2-salt complex. (a) Structure of compound **2-10** (4-ethoxy-5-mercapto-2-phenyl-3(2H)-pyridazinone), a previously discovered Sa-SrtA inhibitor with  $IC_{50}$  of 13 or 1.5  $\mu\text{M}$  when dimerized through disulfide bond. (b) Structure of the sodium thiolate version of **2-10** (**2-salt**) used in NMR and LC-MS studies. (c)  $^1\text{H}$ - $^{15}\text{N}$  HSQC of apo (red)- and inhibitor-bound (blue) SrtA after 24 hr of incubation. Peaks that had significant chemical shifts ( $\Delta\delta$ ) upon binding of the inhibitor are labeled. Significant chemical shift changes are defined as greater than the average  $\Delta\delta +$  one standard deviation, where  $\Delta\delta = \sqrt{(\Delta\delta_H)^2 + (\Delta\delta_N/6.49)^2}$ , and  $\Delta\delta_H$  and  $\Delta\delta_N$  are the chemical shift changes in Hz for amide proton and amide nitrogen, respectively. (d) Selected panels showing intermolecular NOEs between SrtA and **2-salt**. Identity and chemical shifts for each inhibitor proton are shown at the top and bottom of each panel, respectively. Each assigned cross-peak is labeled with the corresponding proximal SrtA proton. Each assignment was verified in a 3D  $^{13}\text{C}$ -edited NOESY by identifying a corresponding inhibitor cross-peak. Not all cross-peaks could be identified due to chemical shift ambiguities, including all NOESY cross-peaks from the inhibitor H $\beta$  protons [Colour figure can be viewed at wileyonlinelibrary.com]



triplicate by a 10-point titration (195 nM to 100  $\mu\text{M}$ ) followed by the addition of 30  $\mu\text{l}$  HeLa cells (2500 cells/well). The plates were incubated at 37°C. Three days later, 50  $\mu\text{l}$  of CellTiter-Glo reagent was added to each well followed by a 2-min shaking and a 10-min incubation to lyse the cells. The relative luminescent intensity units (RLU) of each well were measured using an Infinite® M1000 PRO (Tecan US Inc., Morrisville, NC, USA) with its green filter and 1 s integration time.  $CC_{50}$  values were calculated by fitting the three independent sets of data to Equation 1 using SIGMAPLOT 6.0 (SPSS, Inc., Chicago, IL, USA), by replacing  $IC_{50}$  with  $CC_{50}$ .

### 3 | RESULTS

#### 3.1 | NMR structure of SrtA bound to a soluble pyridazinone analogue

Using high-throughput screening methods, we previously discovered that pyridazinone-based molecules effectively inhibit SrtA.<sup>[19]</sup> One of the most potent molecules that was discovered was 4-ethoxy-5-mercapto-2-phenylpyridazin-3(2H)-one (hereafter referred as compound **2-10**) (Figure 1a), which inhibits SrtA's in vitro activity with an  $IC_{50}$  of  $13 \pm 1 \mu\text{M}$ . However, **2-10** and other pyridazinone-based molecules produced from the screen are poorly soluble in aqueous solvent, and thus could not be optimized using structure-based methods. To overcome this problem, we synthesized a sodium thiolate analogue of **2-10**, 5-ethoxy-6-oxo-1-phenyl-1,6-dihydropyridazine-4-thiolate (**2-salt**) (Figure 1b). **2-salt** exhibits substantially improved solubility (up to 75 mM in aqueous buffer) and inhibits the enzyme with an  $IC_{50}$  of  $35 \pm 7.4 \mu\text{M}$ . Importantly, its increased solubility enables the productive formation of a SrtA:2-salt complex

in which the inhibitor is bound to the catalytic domain of sortase (residues Gln<sup>60</sup>-Lys<sup>206</sup> of SrtA). The NMR spectra of the complex are well resolved and differ substantially from apo-form of SrtA (Figure 1c). Triple resonance methods were used to assign the protein's  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  resonances in the SrtA:2-salt complex. A comparison with the previously published chemical shift assignments of apo-SrtA reveals large inhibitor-dependent chemical shift changes in atoms that are located within the enzyme's active site, and only small shift changes for atoms located elsewhere in the protein (Figure 1c). A comparison of the NOESY spectra of the complex and apo-SrtA reveals that residues located distal

to the active site exhibit similar NOE cross-peak patterns. We conclude from these data that the inhibitor binds to the enzyme's active site where it causes only localized structural changes.

NMR spectroscopy was used to define the structure of the SrtA:2-salt complex. Near-complete chemical shift assignments were obtained for both the protein and bound inhibitor. Because the inhibitor causes only localized changes in the structure of the enzyme, we employed a hybrid approach to determine the structure of the complex. NMR data were used to define the molecular basis of inhibitor binding in the SrtA:2-salt complex (see Section 2). A total of 229 experimental restraints define the structure of the enzyme's active site, including 156 intramolecular protein-protein distance, 20 intermolecular inhibitor-protein NOE distance, 43  $\phi$  and  $\psi$  dihedral angle, and 10  $^3J_{\text{HN-H}\alpha}$  coupling constant restraints. An ensemble containing 20 conformers representing the structure of the complex exhibit good covalent geometries and have no NOE, dihedral angle, or scalar coupling violations greater than 0.5 Å, 5°, or 2 Hz, respectively. Enzyme interactions with the inhibitor are well defined by

the NMR data, as the co-ordinates of the backbone and heavy atoms within the active site, as well as the inhibitor molecule have a root mean square deviation (RMSD) to the mean structure of  $0.20 \pm 0.07$  and  $0.69 \pm 0.06$  Å, respectively. Complete structure and restraint statistics are presented in Table 1.

The structure of the complex reveals that the pyridazinone inhibitor binds to a groove adjacent to the enzyme's active site (Figure 2a-d). The base of the group was formed by residues in strands  $\beta 4$  and  $\beta 7$ , while residues located in the  $\beta 2/\text{H}1$ ,  $\beta 3/\beta 4$ ,  $\beta 6/\beta 7$ , and  $\beta 7/\beta 8$  loops form the sides of the groove and partially shield the inhibitor from solvent. Numerous hydrophobic contacts are made to the 2-phenyl group of the inhibitor from the side chains of Val<sup>166</sup>, Val<sup>168</sup>, and Leu<sup>169</sup> within the  $\beta 6/\beta 7$  loop, Val<sup>193</sup> within the  $\beta 7/\beta 8$  loop, and Ile<sup>182</sup> on strand  $\beta 6$  (Figure 2c). These interactions are supported by intermolecular NOEs between the Val<sup>166</sup> H $\alpha$ , Val<sup>166</sup> H $\gamma$ , Val<sup>168</sup> H $\gamma$ , Leu<sup>169</sup> H $\delta$ , and Val<sup>193</sup> H $\gamma$  protons of the enzyme, and the H $\epsilon$  and H $\delta$  protons of the inhibitor (Figure 1d). Interestingly, an NOE was identified between the H $\delta$  proton of Ile<sup>182</sup> and the H $\delta$  proton of the inhibitor,

**TABLE 1** Statistics for the NMR modeled structure of SrtA bound to a pyridazinone inhibitor 2-salt

|   | SA <sup>a</sup>     | $\overline{\text{SA}}$ <sup>a</sup> |
|---|---------------------|-------------------------------------|
| RMS deviations from NOE interproton distance restraints (Å)                             |                     |                                     |
| Intramolecular <sup>b</sup> (156)   | $0.075 \pm 0.003$   | 0.087                               |
| Intermolecular (20)   | $0.104 \pm 0.007$   | 0.121                               |
| RMS deviations from dihedral angle restraints (°) <sup>c</sup> (43)                     |                     |                                     |
|   | $0.824 \pm 0.068$   | 0.715                               |
| RMS deviations from $^3J_{\text{HN-H}\alpha}$ coupling constants (Hz) <sup>c</sup> (10) |                     |                                     |
|   | $0.451 \pm 0.027$   | 0.772                               |
| Deviations from idealized covalent geometry   |                     |                                     |
| Bonds (Å)   | $0.0040 \pm 0.0000$ | 0.00574                             |
| Angles (°)  | $1.048 \pm 0.002$   | 1.0697                              |
| Impropers (°)   | $1.545 \pm 0.026$   | 1.5550                              |
| PROCHECK-NMR <sup>d</sup>   |                     |                                     |
| Most favorable region (%)   | $66.5 \pm 2.9$      | 66.7                                |
| Additionally allowed region (%)   | $29.3 \pm 3.2$      | 28.6                                |
| Generously allowed region (%)   | $4.3 \pm 1.5$       | 4.8                                 |
| Disallowed region (%)   | $0.0 \pm 0.0$       | 0                                   |
| Co-ordinate precision <sup>e</sup>  |                     |                                     |
| Protein backbone (Å)  | $0.20 \pm 0.07$     |                                     |
| Protein heavy atoms (Å)   | $0.69 \pm 0.06$     |                                     |

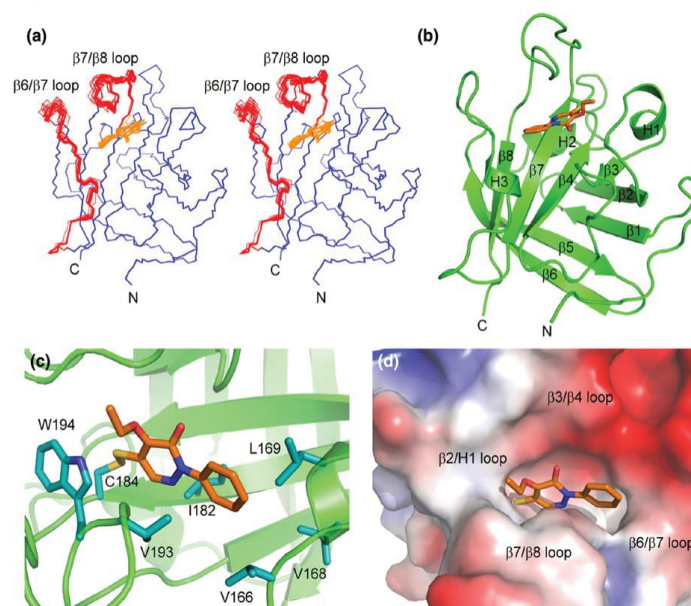
<sup>a</sup>The notation of the NMR structures is as follows: SA represent an ensemble of 20 best structures calculated by simulated annealing.  $\overline{\text{SA}}$  is the average energy-minimized structure. The number of terms for each restraint is given in parentheses. None of the structures exhibited distance violations greater than 0.5 Å, dihedral angle violations greater than 5°, or coupling constant violations greater than 2 Hz.

<sup>b</sup>Intramolecular NOE distance restraints include any NOE signals from any residue in the Sa-SrtA enzyme to residues Val<sup>166</sup>-Lys<sup>175</sup>, Thr<sup>183</sup>-Lys<sup>196</sup>, and the side chains from residues Leu<sup>97</sup>, Ser<sup>116</sup>, His<sup>120</sup>, Thr<sup>180</sup>, Ile<sup>182</sup>, Arg<sup>197</sup>, Ile<sup>199</sup>, and Val<sup>201</sup>.

<sup>c</sup>Experimental backbone dihedral angle restraints comprised 21  $\phi$  and 22  $\psi$  angles within residues Val<sup>166</sup>-Lys<sup>175</sup> and Thr<sup>183</sup>-Lys<sup>196</sup>.

<sup>d</sup>PROCHECK-NMR data includes residues Val<sup>166</sup>-Lys<sup>175</sup> and Thr<sup>183</sup>-Lys<sup>196</sup> of the Sa-SrtA:inhibitor complex.

<sup>e</sup>The co-ordinate precision is defined as the average atomic RMS deviation of the 20 individual simulated annealing structures and their mean co-ordinates. The reported values are for residues Val<sup>166</sup>-Lys<sup>175</sup> and Thr<sup>183</sup>-Lys<sup>196</sup> of the Sa-SrtA:inhibitor complex for the protein backbone RMS deviation. In addition, residues Leu<sup>97</sup>, Ser<sup>116</sup>, His<sup>120</sup>, Thr<sup>180</sup>, Ile<sup>182</sup>, Arg<sup>197</sup>, Ile<sup>199</sup>, and Val<sup>201</sup> were included for calculating protein heavy atom RMS deviation.



**FIGURE 2** NMR solution structure of the SrtA:2-salt complex. (a) Cross-eyed stereo image showing the ensemble of 20 lowest energy structures of the SrtA:2-salt complex. The bound compound 2-salt is shown in orange, and protein backbone atoms allowed to move during molecular dynamics simulations are shown in red (Val<sup>161</sup>-Lys<sup>175</sup> and Thr<sup>183</sup>-Lys<sup>196</sup>). The bundle is aligned using all of the protein backbone heavy atoms and the heavy atoms from the inhibitor. (b) Ribbon structure of the SrtA:2-salt complex. The covalently bound inhibitor is shown in stick representation and colored orange. (c) Expanded view of the SrtA active site with the inhibitor bound. Side chains of residues exhibiting intermolecular NOEs to the inhibitor are shown in cyan sticks. A hydrophobic pocket formed by Val<sup>166</sup>, Val<sup>168</sup>, Leu<sup>169</sup>, and Ile<sup>182</sup> bind the 2-phenyl group of the inhibitor, while the remainder of the inhibitor is positioned over the  $\beta 7$  and  $\beta 8$  strands. (d) Expanded view of the SrtA active site with the protein represented by its solvent accessible surface. The surface is colored by its electrostatic properties from acidic (red) to basic (blue) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

whereas no NOEs were observed to the inhibitor's He proton. This is consistent with the structure complex, as the 2-phenyl group is only partially buried within a hydrophobic pocket; it is sandwiched between the side chains of Val<sup>193</sup> on the  $\beta 7/\beta 8$  loop and Leu<sup>169</sup> on the  $\beta 6/\beta 7$  loop. The positioning of the central pyridazinone ring of the inhibitor is also well defined by the NMR data as NOEs between its H $\gamma$  atom and the Thr<sup>183</sup> H $\alpha$  (strand  $\beta 7$ ), Cys<sup>184</sup> H $\alpha$  (strand  $\beta 7$ ), Trp<sup>194</sup> H $\beta$  ( $\beta 7/\beta 8$  loop), Ile<sup>182</sup> H $\gamma 1$  (strand  $\beta 7$ ), Ile<sup>182</sup> H $\delta$  (strand  $\beta 7$ ), and Val<sup>193</sup> H $\gamma$  ( $\beta 7/\beta 8$  loop) protons are observed (Figure 1d). This orients the central ring such that the carbonyl group is directed toward the solvent, while the H $\gamma$  atom faces strand  $\beta 7$ . The indole ring of Trp<sup>194</sup> on the  $\beta 7/\beta 8$  loop closes the inhibitor recognition groove, shielding the disulfide bond between the inhibitor and Cys<sup>184</sup> from the solvent. This is evidenced by NOEs between the He1, H $\zeta 2$ , and H $\zeta 3$  atoms of the indole ring of Trp<sup>194</sup>, and the H $\delta$  methyl protons on Leu<sup>97</sup> within helix H1. This demonstrates the  $\beta 7/\beta 8$  loop of the 2-salt bound enzyme is positioned in a manner more similar to the apo-SrtA enzyme with the loop pressed against H1, and differs from the substrate-bound

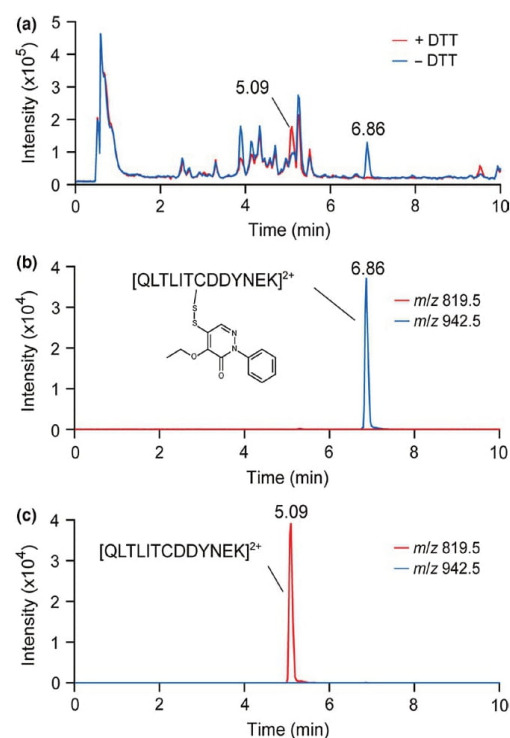
form of the enzyme, as these elements are separated by  $\sim 13$  Å (see Section 4).<sup>[20,32]</sup> The precise location of 4-ethoxy moiety is not well defined by the NMR data, as only a single NOE defines its positioning (between Trp<sup>194</sup> H $\delta 1$  and H $\alpha$  methyl of the inhibitor). However, in nearly all conformers within the ensemble it is positioned within a narrow groove formed between the  $\beta 7/\beta 8$  and  $\beta 2/H1$  loops.

In the structure of the complex, the sulphhydryl group of the active site cysteine residue (Cys<sup>184</sup>) is positioned adjacent to the 5-mercapto group of the inhibitor, suggesting that they may be joined via a disulfide bond. Although the mechanism of inhibition by the pyridazinone-based compounds is not known, the most potent pyridazinone molecules also contain 5-mercapto group consistent with them also covalently modifying the enzyme. To determine the mechanism of inhibition, we incubated 10-fold molar excess of 2-salt with SrtA, unbound 2-salt was then removed by dialysis and the 2-salt:SrtA complex was digested with trypsin. The digestion products were then analyzed using LC-MS before and after adding DTT. In the absence of DTT, a peak eluting at 6.86 min has a mass-to-charge ratio consistent

with it being a disulfide-linked QLTLITCDDYNEK:**2-salt** peptide, compatible with cysteine modification (Figure 3a,b). Moreover, when DTT is added this complex disappears and a new peak corresponding to the unmodified QLTLITCDDYNEK elutes at 5.09 min (Figure 3a,c). Similar LC-MS results were obtained when the parent compound **2-10** was reacted with the enzyme (data not shown). These data indicate pyridazinone-containing inhibitors such as **2-salt** and **2-10** inactivate SrtA by forming disulfide bond to Cys<sup>184</sup>.

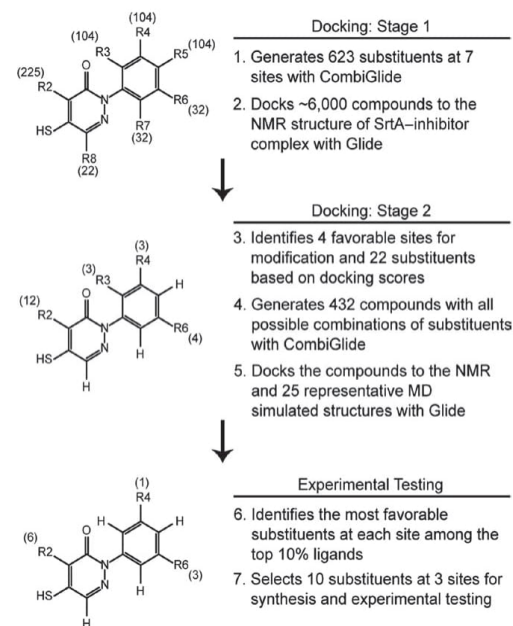
### 3.2 | Structure-based inhibitor optimization

Using the atomic co-ordinates of the SrtA:**2-salt** complex, computational approaches were used to quantitatively predict



**FIGURE 3** LC-MS traces of the SrtA:**2-salt** complex. (a) Overlay of the total ion traces of trypsin-digested 1:1 mixture of SrtA and **2-salt** before (blue) and after (red) addition of 5 mM DTT. Major differences occur at retention times 5.09 min and 6.86 min. (b) Chromatogram of trypsin-digested sample without addition of DTT. Only the traces of ions with  $m/z$  of 819.5 (red) and 942.5 (blue) are shown, which correspond to the  $[M + 2H]^{2+}$  ions of the unmodified Cys<sup>184</sup>-containing peptide and **2-salt** modified version of the peptide, respectively. (c) Similar to (b), but the chromatogram shown is the trypsin-digested sample with the addition of DTT [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

inhibitor modifications that could increase binding affinity and selectivity. In this procedure, analogues of the **2-10** pyridazinone compound were constructed in silico using the program CombiGlide and their binding poses and energies evaluated using the molecular docking program Glide. Briefly, the 5-mercapto-2-phenylpyridazin-3(2*H*)-one moiety of **2-10** was defined as the core in CombiGlide, and substituents derived from the program's fragment library were then systematically added to eight sites (Figure 4, sites named R2 to R8). The substituents were chosen based on their degree of solvent exposure, and for buried substituents, the accessible volume available to them within the enzyme's subsite. In particular, sites on the small molecule that are solvent exposed (R2–R5) were modeled with predominantly polar groups of various sizes, while sites located on the protein-facing side of the molecule (R6–R8) were modeled to contain mostly small hydrophobic groups (see Section 2). Initially, 6,047 analogues of **2-10** were computationally evaluated, which included 623 analogues containing a single alteration, and 5,424 analogues containing two alterations at sites R2 to R8 (Figure 4). Each analogue was then docked to the NMR structure and ranked based on their docking score. Analogues containing single-site modifications at R2, R3, R4, and R6 exhibited the largest



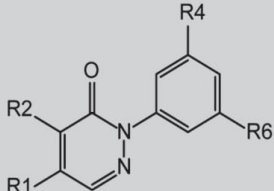
**FIGURE 4** Overview of the molecular docking and experimental testing process. Description of the docking, substituents selection, and experimental testing processes are shown on the right. The structure of the core of the pyridazinone compound is shown on the left. The number of substituents at each site is indicated in parentheses

improvements in binding (12, 3, 3, and 4 analogues containing unique substituents at these sites resulted in improved binding, respectively). Next, to discover small molecules with even greater binding affinity, docking experiments were performed for 432 analogues that each contained four alterations. The small molecules contained substituent changes at their R2, R3, R4, and R6 sites, which were chosen based on the aforementioned single-site docking studies. During this phase of the work, the docking calculations made use of the relaxed complex scheme approach to account for protein flexibility in the enzymes active site as previously described.<sup>[43,51,52]</sup> After completing the computational docking experiment, a total of 43 compounds were analyzed further, as their docking scores ranked in top 10% of the 432 analogues that were tested. These “top” molecules had docking scores ranging from -8.9 to -10.0, which is a significant improvement as compared to the **2-10** lead molecule (-4.08). An inspection of their chemical structures reveals that modification at sites R2, R4, and R6 are most beneficial (R2,  $-(\text{O}(\text{CH}_2)_2\text{OH})$ ,  $-(\text{O}(\text{CH}_2)_3\text{OH})$ ,  $-(\text{OCH}_2\text{COOH})$ ,  $-(\text{O}(\text{CH}_2)_2\text{COOH})$ ,  $-(\text{OCH}_2\text{CONH}_2)$ , and  $-(\text{O}(\text{CH}_2)_2\text{CONH}_2)$ ;

R4,  $-\text{C}(=\text{NH})\text{NH}_2$  (hereafter referred as amidine); R6,  $(-\text{CH}_3, -\text{F}, \text{ and } -\text{OCH}_3)$ .

Several promising pyridazinone analogues based on the docking analysis were synthesized and their inhibitory activity determined experimentally using a FRET-based assay (Table 2). Representative dose-response curves are shown in Fig. S1. Analogues were chosen for further study based on their synthetic feasibility and in order to maximize the chemical diversity of the molecular scaffolds that were tested. Initially, ten analogues of **2-10** were synthesized that contained single-site changes (**2-51** to **2-60**). As **2-10** oxidizes readily to form a more potent and stable symmetric disulfide dimer (**2-17**), all of the analogues were tested in their oxidized form and their potency compared to **2-17**. Improvements of twofold to 10-fold were obtained by making single-site alterations. In particular, non-polar modifications at site R6 are beneficial, presumably increasing van der Waals contacts to residues in the  $\beta 6/\beta 7$  loop of the enzyme that form a hydrophobic pocket (Figure 2c,d). In agreement with the structure, the R4 site on the opposite side of the benzene

**TABLE 2** SrtA inhibition of the pyridazinone compound derivatives



| Compound    | R1                | R2   | R4                    | R6               | Sa-SrtA IC <sub>50</sub> (μM) <sup>a</sup> | Ba-SrtA IC <sub>50</sub> (μM) <sup>b</sup> | CC <sub>50</sub> (μM) | cLogP |
|-------------|-------------------|--|-----------------------|------------------|--|--|-----------------------|-------|
| <b>2-10</b> | SH                | OCH <sub>2</sub> CH <sub>3</sub>                   | H                     | H                | 13 ± 1 <sup>d</sup>                        | 3.2 ± 1.7 <sup>d</sup>                     | n.d. <sup>c</sup>     | 3.03  |
| <b>2-17</b> | S-SR <sup>c</sup> | OCH <sub>2</sub> CH <sub>3</sub>                   | H                     | H                | 1.5 ± 0.4 <sup>d</sup>                     | 1.2 ± 0.4 <sup>d</sup>                     | ~100                  | 4.37  |
| <b>2-51</b> | S-SR <sup>c</sup> | OCH <sub>2</sub> CH <sub>3</sub>                   | H                     | F                | 0.16 ± 0.04                                | 1.34 ± 0.62                                | ~100                  | 4.55  |
| <b>2-52</b> | S-SR <sup>c</sup> | OCH <sub>2</sub> CH <sub>3</sub>                   | H                     | CH <sub>3</sub>  | 1.40 ± 0.30                                | 0.49 ± 0.11                                | n.d. <sup>c</sup>     | 4.58  |
| <b>2-53</b> | S-SR <sup>c</sup> | OCH <sub>2</sub> CH <sub>3</sub>                   | H                     | OCH <sub>3</sub> | 0.62 ± 0.18                                | 0.15 ± 0.05                                | ~100                  | 4.46  |
| <b>2-54</b> | S-SR <sup>c</sup> | OCH <sub>2</sub> CH <sub>3</sub>                   | C(=NH)NH <sub>2</sub> | H                | 0.54 ± 0.14                                | 2.08 ± 0.55                                | >100                  | 2.72  |
| <b>2-55</b> | S-SR <sup>c</sup> | O(CH <sub>2</sub> ) <sub>2</sub> OH                | H                     | H                | 0.22 ± 0.04                                | 0.33 ± 0.14                                | >100                  | 2.35  |
| <b>2-56</b> | S-SR <sup>c</sup> | O(CH <sub>2</sub> ) <sub>3</sub> OH                | H                     | H                | 0.84 ± 0.06                                | 1.86 ± 0.59                                | ~100                  | 2.80  |
| <b>2-57</b> | S-SR <sup>c</sup> | OCH <sub>2</sub> COOH                              | H                     | H                | 2.20 ± 0.20                                | 0.49 ± 0.19                                | n.d. <sup>c</sup>     | 2.02  |
| <b>2-58</b> | S-SR <sup>c</sup> | O(CH <sub>2</sub> ) <sub>2</sub> COOH              | H                     | H                | 6.10 ± 0.50                                | 9.98 ± 0.37                                | >100                  | 2.32  |
| <b>2-59</b> | S-SR <sup>c</sup> | OCH <sub>2</sub> CONH <sub>2</sub>                 | H                     | H                | 22 ± 2                                     | 94 ± 23                                    | n.d. <sup>c</sup>     | 2.11  |
| <b>2-60</b> | S-SR <sup>c</sup> | O(CH <sub>2</sub> ) <sub>2</sub> CONH <sub>2</sub> | H                     | H                | 3.25 ± 0.34                                | 11.90 ± 0.70                               | >100                  | 2.06  |
| <b>2-61</b> | S-SR <sup>c</sup> | O(CH <sub>2</sub> ) <sub>2</sub> OH                | H                     | F                | 0.41 ± 0.12                                | 0.60 ± 0.23                                | >100                  | 2.83  |
| <b>2-62</b> | S-SR <sup>c</sup> | O(CH <sub>2</sub> ) <sub>3</sub> OH                | H                     | F                | 0.02 ± 0.01                                | 0.45 ± 0.29                                | ~100                  | 3.43  |

<sup>a</sup>Or  $K_i^{\text{app}}$  for values that are lower than 0.5 μM as determined by Morrison's equation.

<sup>b</sup>Or  $K_i^{\text{app}}$  for values that are lower than 5 μM as determined by Morrison's equation.

<sup>c</sup>S-SR indicates that the compound is a disulfide-bonded dimer.

<sup>d</sup>Values taken from reference.<sup>19</sup>

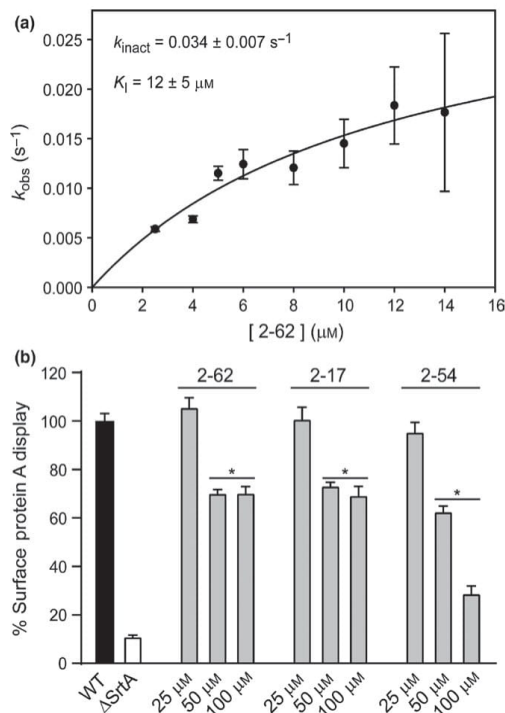
<sup>e</sup>n.d., no data.

ring is also a good site for the addition of a polar amidine group, which presumably enables contacts to Glu<sup>105</sup> located in the  $\beta$ 3/ $\beta$ 4 loop. Altering the length and polar character of the substituent attached to the R2 site is also advantageous, presumably because it enables the formation of additional hydrogen bonds to the residues located within the  $\beta$ 7/ $\beta$ 8 loop (e.g., Tyr<sup>187</sup>, Gly<sup>192</sup>, and Trp<sup>194</sup>). Two analogues, **2-61** and **2-62**, were synthesized that simultaneously alter substituents at sites R2 and R6. Compound **2-61** combines favorable O(CH<sub>2</sub>)<sub>2</sub>OH (R2) and F (R6) modifications, but surprisingly did not show a substantial improvement in activity. In contrast, significant improvements in activity are observed in compound **2-62** that combines O(CH<sub>2</sub>)<sub>3</sub>OH (R2) and F (R6) alterations (IC<sub>50</sub> value of 0.02 ± 0.01 μM). A more complete description of the structure–activity relationship analysis is provided in Section 4.

In order to determine whether the compounds had broad-spectrum activity against sortase enzymes, we tested their ability to inhibit the *Bacillus anthracis* class A sortase enzyme (Ba-SrtA), which similar to *S. aureus* SrtA (Sa-SrtA) anchors proteins to the cell wall that contain an LPXTG sorting signal. In general, all of the analogues had good inhibitory activity against Ba-SrtA, with IC<sub>50</sub> values in the low micromolar to high nanomolar range. However, distinct species-specific trends in inhibitory activity are apparent. For example, the most effective SrtA inhibitor (**2-62**) is 20-fold less active against Ba-SrtA (IC<sub>50</sub> = 0.45 ± 0.29 μM against Ba-SrtA as compared to IC<sub>50</sub> = 0.02 ± 0.01 μM against Sa-SrtA), whereas the best Ba-SrtA inhibitor compound **2-53** containing a methoxy group at the R3 site is fourfold more active against Ba-SrtA as compared to SrtA (Ba-SrtA IC<sub>50</sub> = 0.15 ± 0.05 μM versus Sa-SrtA IC<sub>50</sub> = 0.62 ± 0.18 μM). These subtle differences in potency likely arise from structural differences between the active sites of each enzyme, but nevertheless indicate that pyridazinone-based inhibitors are capable of inhibiting class A type sortases that recognize LPXTG sorting signals.

### 3.3 | Inhibitor inactivation kinetics and effect on protein display in intact cells

For the parent compound (**2-17**), positively charged compound (**2-54**), and the most potent compound based on its IC<sub>50</sub> value (**2-62**), we determined the rate at which they inactivate SrtA. The substrate and inhibitor were simultaneously added to SrtA and the observed rate constant of inhibition ( $k_{\text{obs}}$ ) was determined by fitting the progress curve to Equation 3 (see Section 2). Measurement of  $k_{\text{obs}}$  at different inhibitor concentrations yields  $k_{\text{inact}}$ , the rate constant describing inhibitor covalent modification of the enzyme, and  $K_1$ , which carries the same intrinsic kinetic significance as  $K_m$  and approximates the dissociation constant of the non-covalent E–I complex. The overall inactivation efficiency is a function



**FIGURE 5** Inhibitor inactivation kinetics and effect on protein display. (a) The rate of SrtA inhibition by compound **2-62** was determined by calculating  $k_{\text{obs}}$  for each inhibitor concentration using Equation 3 and subsequently calculating  $k_{\text{inact}}$  and  $K_1$  using Equation 4 listed in Section 2. This graph shows a representative curve of fitting the  $k_{\text{obs}}$  data with Equation 4. (b) The abundance of surface protein A displayed on *S. aureus* cell wall in the absence (control) or presence of pyridazinone inhibitor was quantified with a FITC-labeled IgG binding assay. Statistical significance (\* $p < .0007$ ) was determined using the unpaired, two-tailed Student's  $t$  test ( $n = 3$ , shown with standard deviation)

**TABLE 3** Inactivation kinetics of pyridazinone compound derivatives

| Compound    | $k_{\text{inact}}$ (s <sup>-1</sup> ) | $K_1$ (μM) | $k_{\text{inact}}/K_1$ (μM <sup>-1</sup> s <sup>-1</sup> ) |
|-------------|---------------------------------------|------------|--|
| <b>2-17</b> | 0.004 ± 0.001                         | 5 ± 3      | 0.0008 ± 0.0005  |
| <b>2-54</b> | 0.015 ± 0.003                         | 24 ± 9     | 0.0006 ± 0.0003  |
| <b>2-62</b> | 0.034 ± 0.007                         | 12 ± 5     | 0.003 ± 0.001  |

of covalent modification rate and binding affinity ( $k_{\text{inact}}/K_1$ ). Figure 5a shows a representative plot of  $k_{\text{obs}}$  measured as a function of inhibitor (**2-62**) concentration, which yields  $k_{\text{inact}}$  of 0.034 ± 0.007 s<sup>-1</sup>,  $K_1$  of 12 ± 5 μM and  $k_{\text{inact}}/K_1$  of 0.003 ± 0.001 μM<sup>-1</sup>s<sup>-1</sup>. Complete results are summarized in Table 3. Consistent with the IC<sub>50</sub> results, inactivation efficiency is significantly higher for **2-62** (0.003 ± 0.001 μM<sup>-1</sup>s<sup>-1</sup>),



as compared to **2-54** ( $0.0006 \pm 0.0003 \mu\text{M}^{-1}\text{s}^{-1}$ ) or **2-17** ( $0.0008 \pm 0.0005 \mu\text{M}^{-1}\text{s}^{-1}$ ).

The ability of the most promising inhibitors to disrupt sortase-mediated protein display in intact *S. aureus* cells was determined. In Newman strain of *S. aureus*, SrtA anchors 19 different surface proteins in the bacterial envelope,<sup>[53]</sup> including SpA, a molecule that binds the Fc $\gamma$  and Fab domains of host immunoglobulins.<sup>[54]</sup> The abundance of Ig binding to SpA in the bacterial cell wall envelope is therefore dependent upon the activity of *S. aureus* SrtA, enabling the effects of the pyridazinone molecules on SpA display to be measured. In the assay, cells are cultured with varying amounts of inhibitor and the amount of displayed SpA then quantified by adding fluorescein isothiocyanate (FITC)-labeled IgG. As expected, a  $\Delta$ srtA control strain displays significantly less SpA on its surface; protein display is reduced by ~80% in  $\Delta$ srtA as compared to WT *S. aureus* (Figure 5b). The addition of compounds **2-17**, **2-54**, or **2-62** to WT *S. aureus* cultures caused statistically significant reductions in SpA display. The largest effects are observed when **2-54** is added, which shows a dose-dependent decrease in SpA display. When 100  $\mu\text{M}$  is present in the cell culture, display levels are reduced to 35% of WT, approaching levels observed for the  $\Delta$ srtA strain.

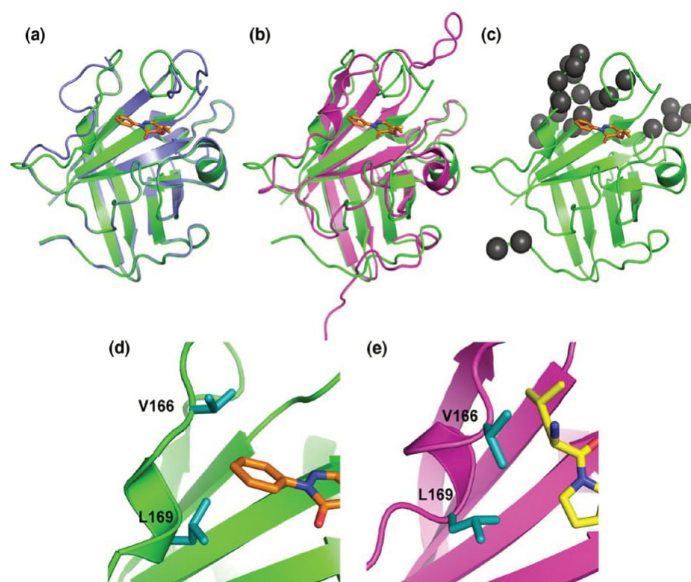
To gain insight into the therapeutic window at which the pyridazinone inhibitors could be dosed to treat *S. aureus* infections, their cytotoxicity against human cells was determined. In general, all of the compounds tested exhibit minimal cytotoxicity against human HeLa cells, as they have  $\text{CC}_{50}$  values in excess of 100  $\mu\text{M}$  (Table 2). Notably, the most potent inhibitor, **2-62**, has a  $\text{CC}_{50}$  to  $\text{IC}_{50}$  ratio as large as 5,000. Interestingly, an analysis of the data reveals that the  $\text{CC}_{50}$  values of the compounds are inversely correlated to their predicted octanol–water partition coefficients (cLogP), with molecules that have smaller (more negative) cLogP values exhibiting some of the highest  $\text{CC}_{50}$  values. This is not surprising, as a more negative cLogP value indicates that a compound is more soluble in water, and therefore less likely to be capable of crossing the cellular membrane where it could cause off-target effects that lead to cell death. As SrtA is located on the extracellular membrane of *S. aureus*, additional modification of the pyridazinone analogues to increase their solubility in water may further reduce off-target cellular toxicity.

## 4 | DISCUSSION

The emergence of methicillin-resistant *S. aureus* (MRSA) and other multidrug-resistant bacterial pathogens has created an urgent need for new antibiotics. The *S. aureus* SrtA enzyme is an attractive drug target as it covalently attaches virulence factors to the microbial surface that have important functions in the infection process, including mediating

bacterial adhesion to host tissues, nutrient acquisition, and the suppression and evasion of the immune system. SrtA's importance has been validated in animal models of infection that demonstrate that *S. aureus* srtA<sup>-</sup> mutants have attenuated virulence. To date, several synthetic, peptide-based, and natural product inhibitors of SrtA have been reported,<sup>[55]</sup> but only a few investigators have been able to rationally optimize these molecules using structural biology approaches. This is because it has been problematic to study the enzyme using X-ray crystallography, as its active site exhibits a high degree of conformational disorder that hinders crystallization. Moreover, the conformational disorder makes accurate computational modeling of enzyme–inhibitor interactions challenging. To overcome these problems, we employed a novel soluble pyridazinone analogue that enabled the NMR structure of enzyme pyridazinone to be determined. Using the structure of the complex, we rationally optimized several pyridazinone-based inhibitors of SrtA, leading to new second-generation molecules that have improved inhibitory activity and lower cellular toxicity.

Our structural studies of the water soluble pyridazinone inhibitor, **2-salt**, reveal that it causes a disordered to ordered structural change in the enzyme's active site that is reminiscent of the previously documented changes caused by binding the sorting signal substrate. Our previous NMR studies of SrtA in its apo-state have revealed that it contains a large active site loop that connects strands  $\beta_6$  to  $\beta_7$  (the  $\beta_6/\beta_7$  loop). In the absence of the substrate, the loop is structurally disordered and undergoes motions on the micro- to millisecond timescale.<sup>[20]</sup> However, upon binding the LPXTG sorting signal, motions in the loop are quenched, as a result of a disordered to ordered conformational change that enables it to partially encapsulate the leucine side chain within the bound peptide.<sup>[32]</sup> Interestingly, in the structure of the SrtA:**2-salt** complex, the inhibitor binds to the same site on SrtA as the LPXTG sorting signal, and similar to the substrate, it alters the structure and dynamics of the  $\beta_6/\beta_7$  active site loop (Figure 6). In the SrtA-sorting signal substrate complex, residues Val<sup>166</sup>-Leu<sup>169</sup> the  $\beta_6/\beta_7$  loop form a  $3_{10}$ -helix that facilitates closure over the substrate by forming a non-polar surface that contacts the leucine side chain. In the SrtA:**2-salt** complex, the 2-phenyl group in the inhibitor mimics the leucyl side chain, as contacts to it drive loop closure and formation of a  $3_{10}$ -helix. However, as compared to the substrate complex, the residues in the newly formed  $3_{10}$ -helix are shifted by one amino acid in the primary sequence; in the SrtA:**2-salt** and SrtA–substrate complexes, the newly formed active site helix is comprised of residues Gly<sup>167</sup>-Asp<sup>170</sup> and residues Val<sup>166</sup>-Leu<sup>169</sup>, respectively. This subtle shift is presumably caused by inhibitor contacts originating from side chains of Val<sup>166</sup> and Leu<sup>169</sup>, which in the SrtA:**2-salt** complex pack against the 2-phenyl group of the inhibitor necessitating that the backbone atoms of Val<sup>166</sup> adopt an extended



**FIGURE 6** Comparison of the SrtA:2-salt structure with apo- and holo-SrtA structures. SrtA:2-salt complex (green) superimposed with (a) apo-SrtA (blue, PDB code 1HJA) and (b) SrtA:LPAT\* complex with the substrate analogue LPAT\* removed (pink, PDB code 2KID). In apo-SrtA, residues Ser<sup>157</sup>-Lys<sup>175</sup> are unstructured in the  $\beta 6/\beta 7$  loop. In inhibitor complex, a portion of this loop is immobilized. The  $\beta 6/\beta 7$  loop in the SrtA:2-salt complex is unstructured from residues Thr<sup>156</sup>-Asp<sup>165</sup>, similar to apo-SrtA, and adopts a single conformation from residues Val<sup>166</sup>-Lys<sup>175</sup>, as seen in the SrtA:LPAT\* structure. The  $\beta 7/\beta 8$  loop adopts a conformation more closely resembling the apo-SrtA structure. (c) Non-proline amide cross-peaks that were not detected in the <sup>1</sup>H-<sup>15</sup>N HSQC spectra are shown in gray spheres. These residues include Met<sup>59</sup>-Gln<sup>60</sup>, Thr<sup>121</sup>-Asp<sup>124</sup>, Lys<sup>137</sup>, Thr<sup>156</sup>-Val<sup>166</sup>, and Lys<sup>196</sup>-Phe<sup>200</sup>. (d) Expanded view of the  $\beta 6/\beta 7$  loop  $3_{10}$ -helix in the SrtA:2-salt complex with 2-salt represented as orange sticks. (e), expanded view of the  $\beta 6/\beta 7$  loop and  $3_{10}$ -helix in the SrtA:LPAT\* complex with LPAT\* represented as yellow sticks [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

conformation (Figure 6d). In contrast, in the SrtA:substrate complex, Val<sup>166</sup> adopts a helical conformation such that its side chain is rotated away from the active site so as to accommodate the leucine side chain of the sorting signal substrate (Figure 6e).

The pyridazinone inhibitor does not fully mimic the sorting signal substrate as its binding only partially closes the  $\beta 6/\beta 7$  loop and it fails to substantially reposition the  $\beta 7/\beta 8$  loop active site loop. In our previously reported structure of the SrtA-substrate complex, in addition to inducing a  $3_{10}$ -helix formation within the  $\beta 6/\beta 7$  loop, substrate binding also caused a disordered to ordered transition in residues that immediately preceded the helix in the primary sequence. These changes fully closed the loop, and presumably constructed a catalytically competent active site in which the side chain guanidino group of Arg<sup>197</sup> is positioned to stabilize high-energy anionic reaction intermediates. However, in the structure of the inhibitor complex, the active site does not fully form as this region of the  $\beta 6/\beta 7$  loop remains flexible and disordered. This is evidenced by the absence of NMR resonances for residues Thr<sup>156</sup>-Asp<sup>165</sup> in the  $\beta 6/\beta 7$  loop in both the apo- and inhibitor-bound forms of the enzyme, which are

presumably broadened beyond detection because of motions on micro- to millisecond timescale (Figure 6c). This notion is also substantiated by the absence of signals for residues Arg<sup>197</sup>-Phe<sup>200</sup> in strand  $\beta 8$  in both forms of the enzyme, which are presumably rigid, but broadened beyond detection because their magnetic environments fluctuate as a result of motions in the proximal  $\beta 6/\beta 7$  loop. A comparison of the substrate- and inhibitor-bound forms of the enzyme also reveals other key differences. Previously, we have shown that upon binding the signal analogue the  $\beta 7/\beta 8$  loop in SrtA is displaced by  $\sim 13$  Å. This rearrangement results in the formation of a new groove between the  $\beta 7/\beta 8$  loop and helix H1, which has been hypothesized to form the binding site of the secondary substrate, the Gly<sub>5</sub> cross-bridge of lipid II. Interestingly, binding of the inhibitor displaces the loop by only  $\sim 3.5$  Å, such that the groove is not exposed. This is substantiated by the presence of NOEs in the spectra of the inhibitor complex between H $\epsilon$ 1, H $\zeta$ 2, and H $\zeta$ 3 of the indole ring of Trp<sup>194</sup> located in the  $\beta 7/\beta 8$  loop and H $\delta$  methyl protons on Leu<sup>97</sup> within helix H1 (similar NOEs are present in the NMR spectra of the apo-form of the enzyme, but absent in the SrtA-substrate complex). Thus, although the inhibitor

triggers partial closure of the  $\beta 6/\beta 7$  loop and the  $3_{10}$ -helix formation similar to the sorting signal substrate, the changes are less substantial and the conformation and dynamics of residues in the  $\beta 6/\beta 7$  loop nearest the active site, as well as the  $\beta 7/\beta 8$  loop, are not significantly affected by inhibitor binding.

The structure of the SrtA:**2-salt** complex is the first high-resolution structure of a SrtA-type enzyme bound to a small-molecule inhibitor. Previously, Zhulenkovs and colleagues solved the NMR structure of *S. aureus* SrtA in complex with a benzisothiazolinone-based inhibitor.<sup>[16]</sup> However, the coordinates of the inhibitor were not well defined, as only nine intermolecular enzyme-inhibitor NOEs defined its positioning. The authors speculated that both the  $\beta 6/\beta 7$  and  $\beta 7/\beta 8$  loops in the SrtA-benzisothiazolinone complex are partially disordered such that the bound inhibitor can adopt different conformations. This is distinct from pyridazinone-based molecules, as in the SrtA:**2-salt** complex a portion of the  $\beta 6/\beta 7$  loop near the inhibitor becomes structurally ordered and the  $\beta 7/\beta 8$  loop is structurally ordered and adopts a closed conformation that is similar to its conformation in the apo-form of the protein. Nevertheless, their binding mechanisms share two common features. First, both pyridazinone and benzisothiazolinone inhibitors covalently modify SrtA by forming a disulfide bond with Cys<sup>184</sup>. Second, despite the flexibility of the benzisothiazolinone compound and the partially disordered active site loops, in both inhibitor complexes the side chains of Val<sup>166</sup>, Val<sup>168</sup>, and Leu<sup>169</sup> in the  $\beta 6/\beta 7$  loop and Trp<sup>194</sup> in the  $\beta 7/\beta 8$  loop consistently interact with the bound inhibitor. Structures of the SrtB sortase have also been determined in complex with small-molecule inhibitors. To aid development of aryl ( $\beta$ -amino)ethyl ketone (AAEK)-based inhibitors, Maresco and colleagues solved the crystal structures of *B. anthracis* SrtB bound to two different AAEK compounds.<sup>[17]</sup> Similar to **2-salt** and the benzisothiazolinone compound, AAEK covalently modifies the active site cysteine residue (Cys<sup>233</sup> in SrtB). Interestingly, while AAEK does not make specific contacts with residues within the active site loops, the compound nevertheless induces ordering of the  $\beta 7/\beta 8$  loop and partial disordering of the  $\beta 6/\alpha 5$  loop (SrtB equivalent of the  $\beta 6/\beta 7$  loop). Combined, structural studies of sortase-inhibitor complexes reveal that sortase inhibitors could modulate the active site loops dynamics, and high-affinity binding might be achieved by maximizing contacts with the residues within the active site loops.

Rational approaches were employed to identify pyridazinone analogues with improved inhibitory activity. Concurrently, we also sought to increase the aqueous solubility of the inhibitors by adding polar substituents to sites on the inhibitor that based on the structure of the complex are surface-exposed. This was done to diminish the capacity of the inhibitor to cross the host cell membrane where it could cause undesirable off-target effects. As described in the section 3, analogues of **2-10** containing a single alteration at site

R2 to R6 were computationally evaluated for their ability to bind SrtA using the structure of the complex as a template. This identified substituents at each site that improved binding affinity, which were then tested in combination using docking calculations that employed the relaxed complex scheme approach to account for protein flexibility. Analogues that scored the highest in this analysis were then synthesized and experimentally tested. This led to a total of seven new analogues that have improved inhibitory activity against SrtA. Modification of the ortho-positions of the 2-phenyl ring have a large impact on activity and solubility. Based on the structure of the complex, substituents at these sites either project into the solvent (R4) or contact hydrophobic residues in the active site loop (R6). Compounds **2-51** to **2-53** selectively modify position R6 on the scaffold, and show improved activity when bulkier fluoro (**2-51**) or O-methyl groups (**2-53**) are added, resulting in nine- and twofold improvements in inhibitory activity, respectively. Based on the structure, these modifications fill a hydrophobic pocket on the enzyme that is formed by non-polar side chains originating from the  $\beta 6/\beta 7$  loop (Val<sup>166</sup>, Val<sup>168</sup>, and Leu<sup>169</sup>), the  $\beta 7/\beta 8$  loop (Val<sup>193</sup>), and the underlying beta sheet (Ile<sup>182</sup> in strand  $\beta 7$ ). Interestingly, only small improvements are observed for **2-52**, presumably because this analogue adds a methyl group at site R6 that may be too small to adequately fill the pocket. In contrast, while a fluoro group is slightly smaller than a methyl group, it is capable of forming multipolar contacts with the protein in addition to hydrophobic interactions,<sup>[56]</sup> which might explain the bigger improvement in activity of **2-51** over **2-52**. Similar to site R6, the R4 site is in an ortho-position, but it is located on the opposite side of the phenyl ring where it is surface-exposed and projected toward an anionic patch on the enzyme that is formed by the side chains of Glu<sup>105</sup> and Asp<sup>170</sup>. The positioning of the R4 substituent in the structure of the complex is compatible with the inhibitory properties of compound **2-54**, which exhibits a ~threefold reduction in its IC<sub>50</sub> relative to the lead molecule. This is because this analogue contains an amidine substituent at site R4 which can presumably form favorable electrostatic interactions. Notably, the addition of a polar amidine substituent at this site substantially decreases the cLogP score as compared to the parent molecule, suggesting that polar substituent addition to site R4 is a viable strategy for improving inhibitor solubility.

Based on the docking studies, a total of 12 analogues containing alterations at site R2 were synthesized and tested experimentally. Based on the structure of the complex, R2 substituents rest on a surface-exposed groove that is located in between the  $\beta 7/\beta 8$  and  $\beta 2/H1$  loops. The addition of hydroxyl bearing 2-hydroxyethoxy (**2-55**) and 3-hydroxypropoxy (**2-56**) at site R2 led to 6.8- and 1.8-fold improved activity, respectively. Although the co-ordinates of the R2 substituent (4-ethoxy) are poorly defined in the structure of SrtA:**2-salt**, the improved binding of these molecules may result from

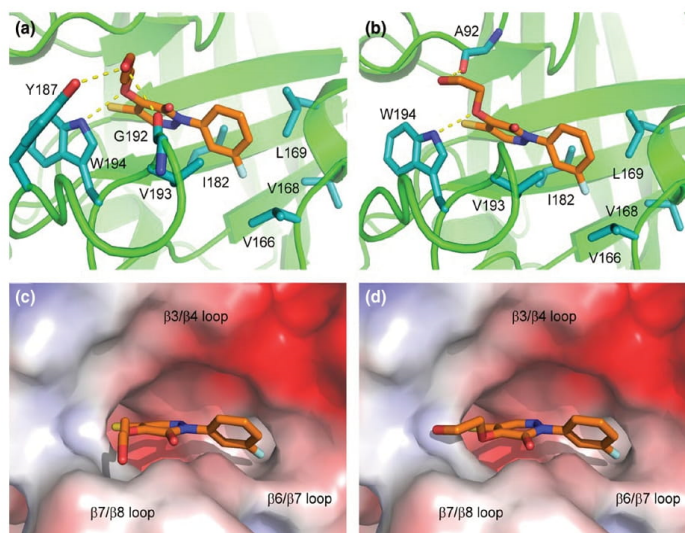
their ability to form additional hydrogen bonds via their hydroxyl group to residues within the  $\beta$ 2/H1 or  $\beta$ 7/ $\beta$ 8 loops. Interestingly, this poorly defined subsite on SrtA does not accommodate large carboxylic acid and amide substituents at site R2, as analogues containing these modifications (**2-57** to **2-60**) have increased  $IC_{50}$  values. In addition to *S. aureus* SrtA, we also tested each compound's ability to inhibit the activity of the SrtA sortase enzyme from *B. anthracis* (Ba-SrtA), another important human pathogen that causes anthrax disease.<sup>[57]</sup> SrtA and Ba-SrtA recognize related LPXTG sorting signals and their active sites adopt similar, but non-identical atomic structures.<sup>[47]</sup> As with SrtA, modifications at sites R2 and R6 in the lead molecule led to improved inhibitory activity. The greatest gains in activity occur when the R6 site is modified with a methoxy group (**2-53**) (eightfold improved activity as compared to the lead molecule). However, distinct trends in the SAR data for SrtA and Ba-SrtA are evident. For example, in contrast to SrtA, modifying the R4 site with amidine led to an increase in  $IC_{50}$ . This is presumably because the *bacillus* enzyme lacks the aforementioned acidic residues that are presumed to stabilize binding to SrtA. Combined, experimental testing of the computationally optimized analogues of **2-10** substantiates the importance of three sites on the pyridazinone scaffold (R2, R4 and R6), which when changed individually improve inhibitory activity against SrtA up to ninefold.

Modifying two sites on the pyridazinone scaffold yielded the most potent sortase inhibitor, compound **2-62** (2-(3-fluorophenyl)-4-(3-hydroxypropoxy)-5-mercaptopyridazin-3(*H*)-one). It is 70 times more active than the lead compound ( $IC_{50}$  of 21 nM) and contains 3-hydroxypropoxy and fluoro groups at sites R2 and R6, respectively. Its activity indicates that modification of sites R2 and R6 can have an additive effect, as individually these alterations are predicted to improve

binding 63-fold (6.8-fold for site R2 and ninefold for site R6). The docking pose of **2-62** suggests that its improved binding can be attributed to an increase in the van der Waals contact surface with the enzyme afforded by addition of a fluoro group at position R6 ( $\sim 10 \text{ \AA}^2$  increase in buried surface area) (Figure 7a,c). In addition, binding may be aided by new favorable interactions to the  $\beta$ 7/ $\beta$ 8 loop formed from the terminal hydroxyl on the  $O(\text{CH}_2)_3\text{OH}$  group, as it is positioned to donate and accept a hydrogen bond from the backbone carbonyl oxygen of Gly<sup>192</sup> and side chain hydroxyl group of Tyr<sup>187</sup>, respectively. Moreover, in the **2-62** docking pose the oxygen at the R2 site closest to the pyridazinone ring is positioned to accept a hydrogen bond from the  $\epsilon\text{NH}$  group in the indole ring of Trp<sup>194</sup>. Interestingly, a close analogue of **2-62** that contains  $O(\text{CH}_2)_2\text{OH}$  at position R2 instead of  $O(\text{CH}_2)_3\text{OH}$  (compound **2-61**) is  $\sim 20$  times less potent than **2-62**. Insight into its reduced potency is provided by its docking pose, which reveals that its shorter R2 side chain does not favorably hydrogen bond with Tyr<sup>187</sup> and Gly<sup>192</sup>, but instead forms a single hydrogen bond to the backbone oxygen of Ala<sup>92</sup> located in the  $\beta$ 2/H1 loop (Figure 7b,d).

Because the pyridazinone molecules covalently inactivate sortase, the  $IC_{50}$  values reported in Table 2 may not accurately define each inhibitor's potency, as this parameter does not reveal the rate at which the small molecules modify and inactivate the enzyme. We therefore measured the  $k_{\text{inact}}$  and  $K_I$  parameters for **2-17**, **2-54**, and **2-62**, which were also tested using the cellular assay. As reported in Table 3, **2-62** has the highest inactivation efficiency ( $k_{\text{inact}}/K_I = 0.003 \mu\text{M}^{-1}\text{s}^{-1}$ ), which is consistent with it having the lowest measured  $IC_{50}$  value when it is pre-incubated with the enzyme for 1 hr (Table 2). Although **2-10** was not tested due to its tendency to oxidize into **2-17**, presumably

**FIGURE 7** Docking poses of compounds **2-61** and **2-62**. Expanded view of the active site showing interactions between the protein and compound **2-62** (a, c) and compound **2-61** (b, d). In (a) and (b), residues that contact the compound are shown as cyan sticks. Hydrogen bonds are indicated by yellow dotted lines. In (c) and (d), the protein solvent accessible surface is shown and colored by its electrostatic properties from acidic (red) to basic (blue) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



the ~10-fold lower  $IC_{50}$  of **2-17** as compared to **2-10** is caused by a difference in their  $k_{inact}$  values. This is because it is expected that **2-17** will have a higher  $k_{inact}$  value as it inactivates sortase's Cys<sup>184</sup> residue via a thiol-disulfide exchange mechanism, whereas **2-10** inactivates sortase via a slower thiol-thiol oxidation process. Notably, **2-62** inactivates SrtA significantly more efficiently than previously reported diazoketone or chloromethylketone containing peptidomimics, which exhibited  $k_{inact}/K_I$  values of  $0.0004 \mu M^{-1}s^{-1}$  and  $0.0009 \mu M^{-1}s^{-1}$ , respectively.<sup>[58]</sup> This is presumably a result of the higher reactivity of the sulphhydryl group, which we have shown modifies the active site cysteine residue (Figure 3). Importantly, the SrtA inhibitors are capable of reducing protein display in intact *S. aureus* cells, as incubating analogues **2-17**, **2-54**, and **2-62** with bacterial cell cultures dramatically reduces SpA reporter protein display (Figure 5b). Interestingly, although **2-62** is the most potent compound in vitro based on its  $k_{inact}/K_I$  and  $IC_{50}$  values, **2-54** is nonetheless more effective than **2-17** or **2-62** at inhibiting SpA display in intact cells. This could be due to the positive charge on the amidine group in **2-54**, which may better allow it to traverse the cell wall as compared to the less polar **2-17** and **2-62** molecules. This idea is consistent with the finding that **2-17** and **2-62** do not exhibit dose-dependent activity in the cellular assay when their concentration is increased from 50 to 100  $\mu M$ . Presumably, their higher cLogP values as compared to **2-54** cause them to be less soluble in the growth media used in cellular assay, thereby limiting the effective inhibitor concentration that can be achieved at the cell surface. In general, our inhibitors are less active in the cellular assay as compared to the in vitro assay. It is conceivable that their effective concentration in the cellular assay is reduced because they non-specifically interact with the bacterial cell wall, or because they interact with components that are uniquely present in the growth media (e.g., secreted proteins). In particular, the presence of low molecular weight thiol-containing compounds in the media (e.g., glutathione or coenzyme A) could inactivate the inhibitors by covalently modifying their thiol group. Importantly, although complete disruption of display was not achieved at 100  $\mu M$ , the in vitro and cellular potencies of our compounds are similar to, or better than, recently described SrtA inhibitors that are efficacious in treating *S. aureus* infections in a mouse model.<sup>[49]</sup>

## 5 | CONCLUSIONS

In summary, using NMR spectroscopy, mass spectrometry and a designed soluble analogue, we have determined the mechanism through which pyridazinone-based small molecules inhibit the *S. aureus* SrtA enzyme. We demonstrate

that these inhibitors partially mimic SrtA's natural substrate by partially inducing a disordered to ordered conformational change in the  $\beta 6/\beta 7$  active site loop. Using computational and synthetic chemistry approaches, several second-generation inhibitors have been produced that have increased inhibitory activity, both in vitro and on the bacterial cell surface. At present, two SrtA inhibitors have now been reported that are efficacious in treating potentially lethal *S. aureus* infections in animal models. These molecules include, (Z)-3-(2,5-dimethoxyphenyl)-2-(4-methoxyphenyl)acrylonitrile (DMMA) and 3-(4-pyridinyl)-6-(2-sodiumsulfonatephenyl)[1,2,4]triazolo[3,4-b][1,3,4]thiadiazole (triazolothiadiazole), which were efficacious when dosed at 20 and 40 mg of compound per kg of animal, respectively.<sup>[49,59]</sup> The **2-62** pyridazinone compound reported in this paper has a lower  $IC_{50}$  than these molecules; **2-62** has an  $IC_{50} = 0.02 \mu M$ , versus  $IC_{50}$  values of 9.2 and 9.3  $\mu M$  for DMMA and triazolothiadiazole, respectively. This suggests that **2-62** and related pyridazinone molecules are potential candidates for further development into anti-infective agents. Such new therapeutics are needed, as many microbial pathogens are increasingly becoming resistant to current antibiotic therapies.

## ACKNOWLEDGMENTS

This research was supported by funding from the National Institutes of Health (AI52217 to R.T.C. and M.E.J.; and GM31749 to J.A.M.). A.H.C. was supported by the UCLA Chemistry Biology Interface Training program (NIH T32GM008496). B.R.A. was supported by the UCLA-MBI Whitcome Pre-Doctoral Training Grant. Additional support at UCSD has been provided by the National Science Foundation, the Howard Hughes Medical Institute, the Center for Theoretical Biological Physics, the National Biomedical Computation Resource, and the NSF Supercomputer Centers. Mass Spectrometry Instrumentation was made available through the support of Greg Khitrov (University of California, Los Angeles Molecular Instrumentation Center—Mass Spectrometry Facility in the Department of Chemistry). The mass spectrometry data were collected through a project described and supported by Grant Number S10-RR025631 from the National Center for Research Resources.

## CONFLICT OF INTEREST

There are no conflict of interests.

## NOTE

<sup>a</sup> Center for Disease Control and Prevention. 2013. Antibiotic Resistance Threats in the United States, 2013. <http://www.cdc.gov/drugresistance/threat-report-2013/>.

## REFERENCES

- [1] F. D. Lowy, *N. Engl. J. Med.* **1998**, *339*, 520.
- [2] D. J. Pallin, D. J. Egan, A. J. Pelletier, J. A. Espinola, D. C. Hooper, C. A. Camargo Jr, *Ann. Emerg. Med.* **2008**, *51*, 291.
- [3] L. F. McCaig, L. C. McDonald, S. Mandal, D. B. Jernigan, *Emerg. Infect. Dis.* **2006**, *12*, 1715.
- [4] K. J. Welsh, A. N. Abbott, E. M. Lewis, J. M. Gardiner, M. C. Kruzel, C. T. Lewis, J. F. Mohr, A. Wanger, L. Y. Armitige, *J. Clin. Microbiol.* **2010**, *48*, 894.
- [5] F. M. Marty, W. W. Yeh, C. B. Wennersten, L. Venkataraman, E. Albano, E. P. Alyea, H. S. Gold, L. R. Baden, S. K. Pillai, *J. Clin. Microbiol.* **2006**, *44*, 595.
- [6] Y. Ikeda-Dantsuji, H. Hanaki, F. Sakai, K. Tomono, Y. Takesue, J. Honda, Y. Nonomiya, A. Suwabe, O. Nagura, K. Yanagihara, H. Mikamo, K. Fukuchi, M. Kaku, S. Kohno, C. Yanagisawa, T. Nakae, K. Yoshida, Y. Niki, *J. Infect. Chemother.* **2011**, *17*, 45.
- [7] W. W. Navarre, O. Schneewind, *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 174.
- [8] T. J. Foster, J. A. Geoghegan, V. K. Ganesh, M. Hook, *Nat. Rev. Microbiol.* **2014**, *12*, 49.
- [9] S. K. Mazmanian, G. Liu, E. R. Jensen, E. Lenoy, O. Schneewind, *Proc Natl Acad Sci U S A* **2000**, *97*, 5510.
- [10] S. Cascioferro, M. Totsika, D. Schillaci, *Microb. Pathog.* **2014**, *77*, 105.
- [11] T. Spirig, E. M. Weiner, R. T. Clubb, *Mol. Microbiol.* **2011**, *82*, 1044.
- [12] S. K. Mazmanian, G. Liu, H. Ton-That, O. Schneewind, *Science* **1999**, *285*, 760.
- [13] D. A. Rasko, V. Sperandio, *Nat. Rev. Drug Discov.* **2010**, *9*, 117.
- [14] S. Escaich, *Curr. Opin. Chem. Biol.* **2008**, *12*, 400.
- [15] N. Suree, M. E. Jung, R. T. Clubb, *Mini Rev. Med. Chem.* **2007**, *7*, 991.
- [16] D. Zhulenkova, Z. Rudevica, K. Jaudzems, M. Turks, A. Leonchiks, *Bioorg. Med. Chem.* **2014**, *22*, 5988.
- [17] A. W. Maresso, R. Wu, J. W. Kern, R. Zhang, D. Janik, D. M. Missiakas, M. E. Duban, A. Joachimiak, O. Schneewind, *J. Biol. Chem.* **2007**, *282*, 23129.
- [18] R. Zhang, R. Wu, G. Joachimiak, S. K. Mazmanian, D. M. Missiakas, P. Gornicki, O. Schneewind, A. Joachimiak, *Structure* **2004**, *12*, 1147.
- [19] N. Suree, S. W. Yi, W. Thieu, M. Marohn, R. Damoiseaux, A. Chan, M. E. Jung, R. T. Clubb, *Bioorg. Med. Chem.* **2009**, *17*, 7174.
- [20] U. Ilangovan, H. Ton-That, J. Iwahara, O. Schneewind, R. T. Clubb, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6056.
- [21] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277.
- [22] D. S. Garrett, R. Powers, A. M. Gronenborn, G. M. Clore, *J. Magn. Reson.* **1991**, *95*, 214.
- [23] R. Keller, *The Computer Aided Resonance Assignment Tutorial*, CATINA Verlag, Goldau, Switzerland **2004**.
- [24] J. Cavanagh, W. J. Fairbrother, A. G. Palmer, N. J. Skelton, *Protein NMR Spectroscopy*, 2nd ed., Elsevier Science and Technology, San Diego, CA **2006**.
- [25] Q. Teng, *Structural Biology: Practical NMR Applications*, Springer Verlag, New York **2005**.
- [26] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J. Biomol. NMR* **2009**, *44*, 213.
- [27] G. W. Vuister, A. Bax, *J. Am. Chem. Soc.* **1993**, *115*, 7772.
- [28] T. Herrmann, P. Guntert, K. Wuthrich, *J. Biomol. NMR* **2002**, *24*, 171.
- [29] T. Herrmann, P. Guntert, K. Wuthrich, *J. Mol. Biol.* **2002**, *319*, 209.
- [30] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. M. Clore, *J. Magn. Reson.* **2003**, *160*, 65.
- [31] Y. Zong, T. W. Bice, H. Ton-That, O. Schneewind, S. V. Narayana, *J. Biol. Chem.* **2004**, *279*, 31383.
- [32] N. Suree, C. K. Liew, V. A. Villareal, W. Thieu, E. A. Fadeev, J. J. Clemens, M. E. Jung, R. T. Clubb, *J. Biol. Chem.* **2009**, *284*, 24465.
- [33] R. Koradi, M. Billeter, K. Wuthrich, *J. Mol. Graph.* **1996**, *14*(51–5), 29.
- [34] W. L. DeLano, *The PyMOL Molecular Graphics System*. The PyMOL Molecular Graphics System, DeLano Scientific LLC, Palo Alto, CA **2006**.
- [35] G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, *J. Comput. Aided Mol. Des.* **2013**, *27*, 221.
- [36] T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, J. L. Banks, *J. Med. Chem.* **2004**, *47*, 1750.
- [37] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, *J. Med. Chem.* **2004**, *47*, 1739.
- [38] R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, D. T. Mainz, *J. Med. Chem.* **2006**, *49*, 6177.
- [39] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* **2006**, *25*, 247.
- [40] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157.
- [41] W. D. Cornell, P. Cieplak, C. I. Bayly, P. A. Kollman, *J. Am. Chem. Soc.* **1993**, *115*, 9620.
- [42] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, A. Rendell, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, *Gaussian 09*, Revision D.01, Gaussian, Inc., Wallingford, CT, **2013**.
- [43] A. H. Chan, J. Wereszczynski, B. R. Amer, S. W. Yi, M. E. Jung, J. A. McCammon, R. T. Clubb, *Chem. Biol. Drug Des.* **2013**, *82*, 418.
- [44] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, *Proteins* **2010**, *78*, 1950.
- [45] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781.

- [46] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, *Bioinformatics* **2013**, *29*, 845.
- [47] E. M. Weiner, S. Robson, M. Marohn, R. T. Clubb, *J. Biol. Chem.* **2010**, *285*, 23433.
- [48] R. A. Copeland, *Evaluation of Enzyme Inhibitors in Drug Discoveries: A Guide for Medicinal Chemists and Pharmacologists*, John Wiley & Sons, New Jersey **2005**.
- [49] J. Zhang, H. Liu, K. Zhu, S. Gong, S. Dramsi, Y. T. Wang, J. Li, F. Chen, R. Zhang, L. Zhou, L. Lan, H. Jiang, O. Schneewind, C. Luo, C. G. Yang, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13517.
- [50] S. P. Crouch, R. Kozłowski, K. J. Slater, J. Fletcher, *J. Immunol. Methods* **1993**, *160*, 81.
- [51] J. H. Lin, A. L. Perryman, J. R. Schames, J. A. McCammon, *Biopolymers* **2003**, *68*, 47.
- [52] R. E. Amaro, R. Baron, J. A. McCammon, *Aided Mol. Des.* **2008**, *22*, 693.
- [53] L. A. Marraffini, A. C. Dedent, O. Schneewind, *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 192.
- [54] F. Falugi, H. K. Kim, D. M. Missiakas, O. Schneewind, *MBio* **2013**, *4*, e00575.
- [55] S. Cascioferro, D. Raffa, B. Maggio, M. V. Raimondi, D. Schillaci, G. Daidone, *J. Med. Chem.* **2015**, *58*, 9108.
- [56] K. Muller, C. Faeh, F. Diederich, *Science* **2007**, *317*, 1881.
- [57] M. Mock, A. Fouet, *Annu. Rev. Microbiol.* **2001**, *55*, 647.
- [58] C. J. Scott, A. McDowell, S. L. Martin, J. F. Lynas, K. Vandenbroeck, B. Walker, *Biochem J.* **2002**, *366*, 953.
- [59] K. B. Oh, K. W. Nam, H. Ahn, J. Shin, S. Kim, W. Mar, *Biochem. Biophys. Res. Commun.* **2010**, *396*, 440.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Chan AH, Yi SW, Weiner EM, et al. NMR structure-based optimization of *Staphylococcus aureus* sortase A pyridazinone inhibitors. *Chem Biol Drug Des.* 2017;90:327–344. <https://doi.org/10.1111/cbdd.12962>

## 5.2.2 Supplementary Information

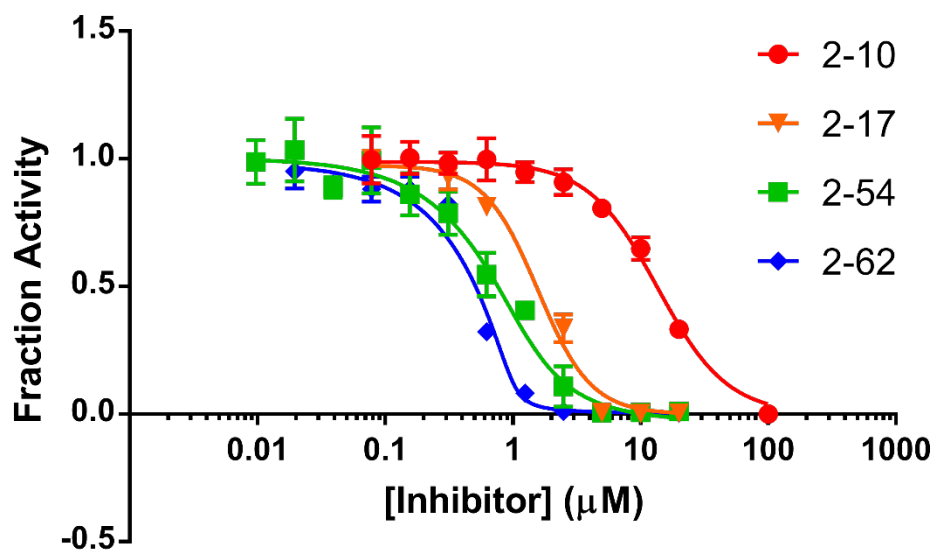


Figure 5.2.S1 Overlay of the *in vitro* dose-response curves of 2-10 (red), 2-17 (orange), 2-54 (green) and 2-62 (blue).



## Synthetic procedures of pyridazinone analogues

### 1. Chemistry

Materials were obtained from commercial suppliers and were used without purification. All moisture sensitive reactions were conducted under an argon atmosphere using oven-dried glassware and standard syringe/septa techniques. Most of reactions were monitored by silica gel TLC using UV light followed by visualization with a *p*-anisaldehyde or ninhydrin staining solution. Some reactions were monitored by the crude <sup>1</sup>H NMR spectrum. <sup>1</sup>H NMR spectra were measured at 400 MHz in CDCl<sub>3</sub> unless stated otherwise and data were reported as follows in ppm (δ) from the internal standard (TMS, 0.0 ppm): chemical shift (multiplicity, integration, coupling constant in Hz.). 2D-NMR experiments (NOESY, COSY and TOCSY) were performed at 500 MHz to confirm the regioselectivity of the substitution reactions. Melting points of solid compounds were determined on a Thomas Hoover capillary melting point apparatus. The purity of the new compounds was assessed by several methods: high-field proton and carbon NMR (lack of significant impurities), R<sub>f</sub> values on TLC (lack of obvious impurities), melting point, and mass spectrometry.

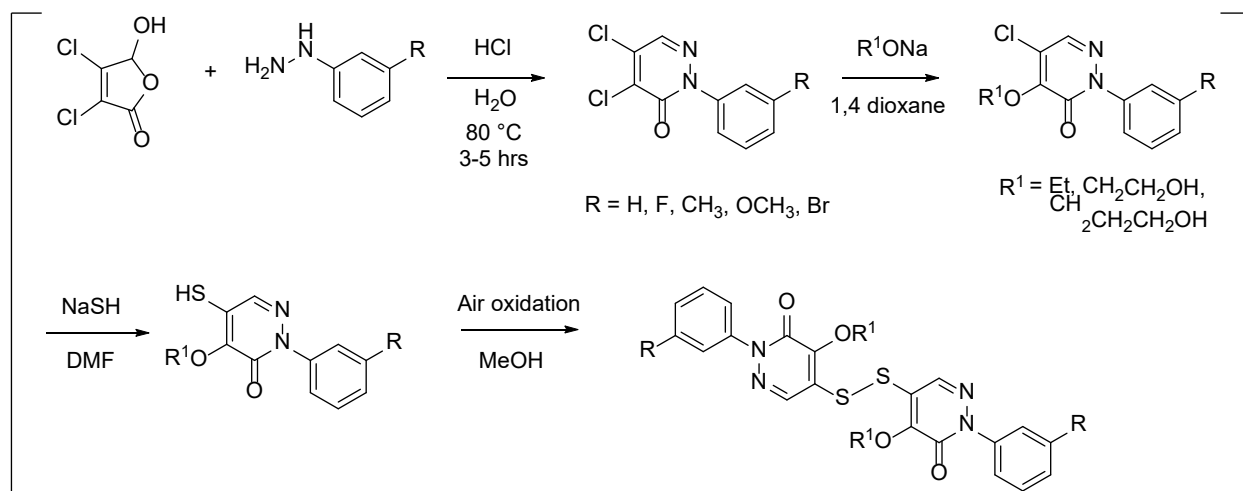
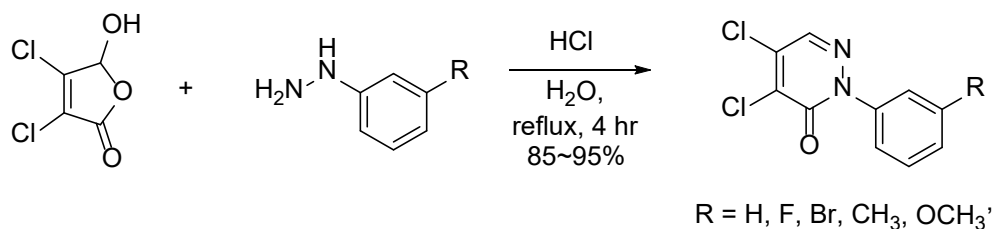


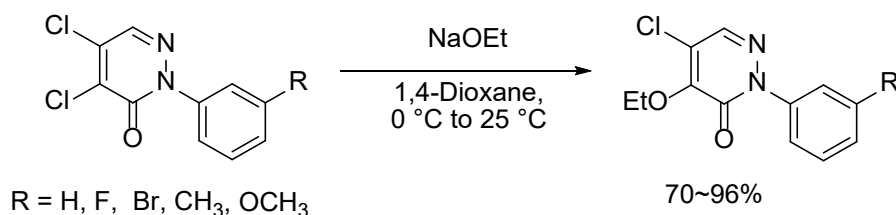
Figure 5.2.S2 General scheme

### 1.1 General Procedure for the Synthesis of 2-substituted-4,5-dichloropyridazin-3-ones, e.g., 2-Phenyl-4,5-dichloropyridazin-3-one.



To a solution of phenylhydrazine (2.9 mL, 30 mmol) in dilute HCl (4 M, 60 mL) was added mucochloric acid (5 g, 30 mmol) at 25 °C. The solution was refluxed for 3 h. The resulting suspension was filtered and washed three times with water. The solids were dried under high vacuum to give 7 g of the yellowish white solid, 94%. R = H mp 158 °C. <sup>1</sup>H NMR 7.91 (1H, s), 7.57 (2H, m), 7.48 (2H, m), 7.42 (1H, m); <sup>13</sup>C NMR 156.15, 140.86, 136.39, 136.14, 135.33, 128.95, 128.89, 125.17.

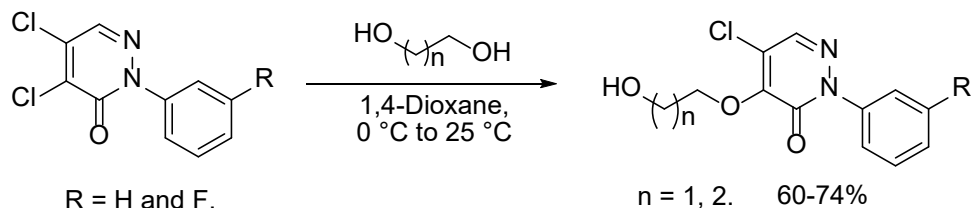
### 1.2 General Procedure for the Synthesis of 2-Substituted 4-Ethoxy-5-chloropyridazin-3-ones, e. g., 5-Chloro-4-methoxy-2-phenylpyridazin-3-one.



To a solution of 4,5-dichloro-2-phenylpyridazin-3(2H)-one (200 mg, 0.809 mmol) in 6 mL of 1,4-dioxane was added 1 mL of freshly generated NaOEt (0.8 M) in EtOH at 0 °C. The suspension was stirred for 2 h as the solution was slowly warmed to 25 °C. The suspension was concentrated and the mixture was subjected to flash column chromatography on silica gel to give 189 mg of the desired product, 92%. R = H mp 78 °C. <sup>1</sup>H NMR 7.84 (1H, s), 7.54 (2H, m),

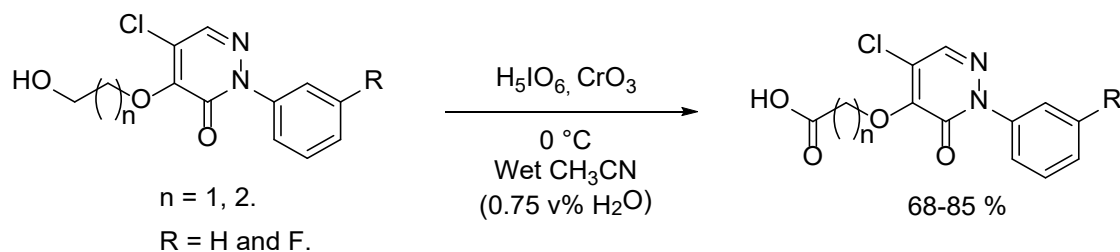
7.48 (2H, m), 7.41 (1H, m); 4.68 (2H, q,  $J = 7.2$  Hz), 1.42 (3H, t,  $J = 7.2$  Hz);  $^{13}\text{C}$  NMR 157.15, 151.36, 140.96, 138.17, 128.89, 128.56, 125.46, 123.62, 69.34, 15.94. For the other analogues, the yields varied from 70-96%.

### 1.3 General Procedure for the Synthesis of 2-Substituted 4-hydroxyalkoxy-5-chloropyridazin-3-ones, e.g., 5-chloro-4-(3-hydroxypropoxy)-2-phenylpyridazin-3(2H)-one.



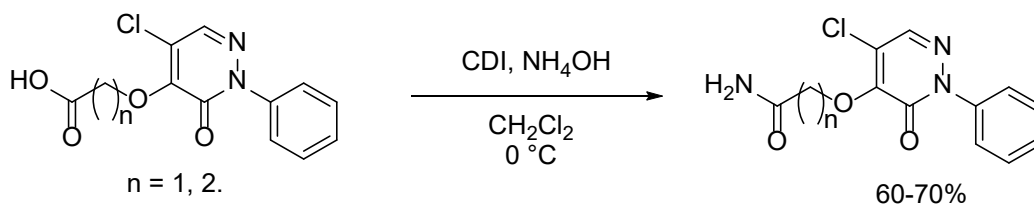
To a solution of **2-42** (3 g, 12.4 mmol) in 50 mL of dioxane was added 6.2 mL of sodium 3-hydroxypropan-1-olate (2 M) at 0 °C. The suspension was stirred for 2 h as it was allowed to warm to 25 °C. The suspension was concentrated and the mixture was subjected to flash column chromatography on silica gel to give 2.38 g of desired product, 65%. R = H, n = 2.  $^1\text{H}$  NMR 7.89 (1H, s), 7.52 (2H, m), 7.47 (2H, m), 7.42 (1H, m), 4.61 (2H, m), 3.88 (2H, m), 3.06 (1H, t,  $J = 6.3$  Hz), 2.03 (2H, m);  $^{13}\text{C}$  NMR 157.60, 151.62, 140.75, 138.29, 128.94, 128.78, 125.41, 125.11, 70.15, 58.79, 32.63. For the other analogues, the yields varied from 60 to 74%.

### 1.4 General Procedure for oxidation of primary alcohols to carboxylic acids, e.g., 3-((5-chloro-3-oxo-2-phenyl-2,3-dihydropyridazin-4-yl)oxy)propanoic acid.



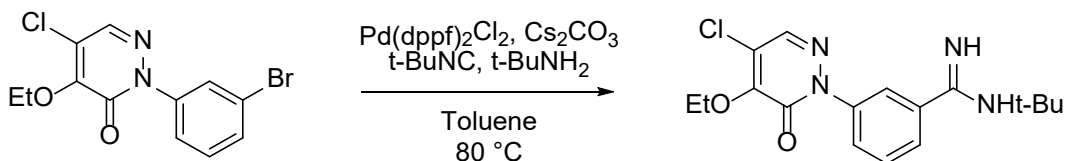
To a solution of 5-chloro-4-(3-hydroxypropoxy)-2-phenylpyridazin-3(2*H*)-one (80 mg, 0.28 mmol) in 5 mL of wet acetonitrile (0.75 % water by volume) was added 2 mL of H<sub>5</sub>IO<sub>6</sub>/CrO<sub>3</sub> solution at 0 °C (1). After stirring for 30 min, the suspension was concentrated and the mixture was subjected to an acid-base work up. After extracting with Et<sub>2</sub>O from aqueous layer and concentration, the crude was subjected to flash column chromatography on silica gel to give 55 mg of desired product, 65%. R = H, n = 2). <sup>1</sup>H NMR 7.86 (1H, s), 7.54 (2H, m), 7.48 (2H, m), 7.42 (1H, m), 4.86 (2H, t, *J* = 6.0 Hz), 2.87 (2H, t, *J* = 6.0 Hz), 2.17 (1H, bs); <sup>13</sup>C NMR 175.16, 157.12, 151.01, 140.77, 138.24, 128.95, 128.74, 125.44, 124.07, 68.33, 35.20.

### 1.5 General procedure for primary amide synthesis



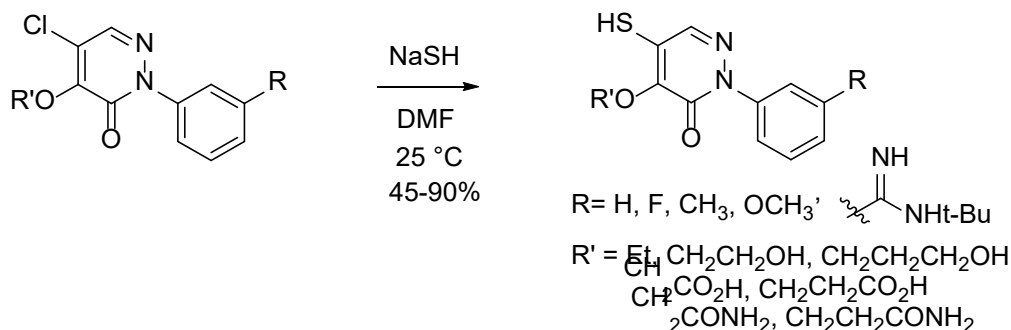
To a solution of 3-((5-chloro-3-oxo-2-phenyl-2,3-dihydropyridazin-4-yl)oxy)propanoic acid (100 mg, 0.34 mM) in dichloromethane (5 mL), was added CDI (1.2 eq) at 0 °C. The reaction mixture was stirred for 30 min and then NH<sub>4</sub>OH (1.1 eq) was added dropwise. The reaction mixture was stirred for an hour and then concentrated. The crude mixture was subjected to flash column chromatography on silica gel to give 70 mg of the desired amide product, 3-((5-chloro-3-oxo-2-phenyl-2,3-dihydropyridazin-4-yl)oxy)propanamide, 70%. n = 2. <sup>1</sup>H NMR 7.90 (1H, s), 7.54 (2H, m), 7.49 (2H, m), 7.44 (1H, m), 7.02 (1H, bs), 5.41 (1H, bs), 4.69 (2H, t, *J* = 5.5 Hz), 2.77 (2H, t, *J* = 5.5 Hz); <sup>13</sup>C NMR 172.28, 157.69, 151.11, 140.60, 138.24, 129.04, 128.94, 125.65, 125.34, 68.75, 36.79.

### 1.6 Procedure for Pd(II) catalyzed amidine synthesis. *N*-(*tert*-butyl)-3-(4-chloro-5-ethoxy-6-oxopyridazin-1(6*H*)-yl)benzimidamide



To a suspension of 2-(3-bromophenyl)-5-chloro-4-ethoxypyridazin-3(2H)-one (105 mg, 0.32 mmol), t-butylisocyanide (60  $\mu$ L), t-butylamine (100mg), and  $\text{Cs}_2\text{CO}_3$  (150 mg) in 3 mL of toluene was added 15 mg of  $\text{Pd}(\text{dppf})_2\text{Cl}_2$  (2). The resulting suspension was purged with argon and stirred for 5 h at 80  $^\circ\text{C}$ . The crude product was concentrated and subjected to flash column chromatography on silica gel to give 40 mg of desired product, 36%.  $^1\text{H}$  NMR 7.88 (1H, t,  $J$  = 2.0 Hz), 7.85 (1H, s), 7.77 (1H, m), 7.67 (1H, m), 7.51 (1H, t,  $J$  = 8.0 Hz), 5.97 (1H, bs), 4.66 (2H, q,  $J$  = 7.0 Hz), 1.46 (9H, s), 1.42 (3H, t,  $J$  = 7.0 Hz).  $^{13}\text{C}$  NMR 165.71, 157.11, 151.41, 141.05, 138.48, 136.95, 129.05, 127.96, 126.91, 123.88, 123.83, 69.46, 51.86, 28.83, 15.93.

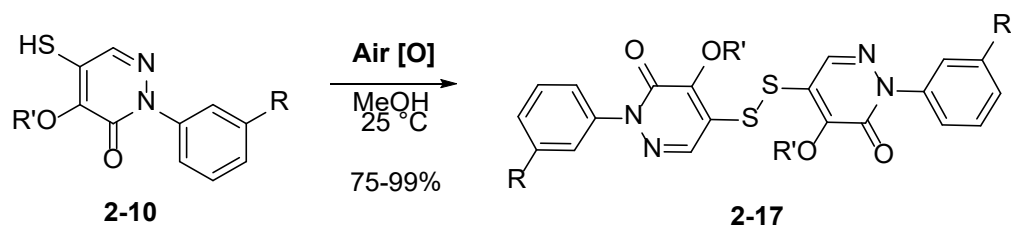
### 1.7 General Procedure for the Synthesis of 2-Substituted 4-Alkoxy-5-mercapto-pyridazin-3-ones, e.g., 4-ethoxy-5-mercapto-2-phenylpyridazin-3-one, 2-10.



To a solution of 5-chloro-4-ethoxy-2-phenylpyridazin-3(2H)-one (63 mg, 0.25 mmol) in 2 mL of DMF was added 70 mg of NaSH at 25  $^\circ\text{C}$ . After TLC showed complete consumption of starting material, the solution was concentrated under high vacuum and diluted with 10 mL of water. The aqueous layer was washed with ethyl acetate and then the pH of the aqueous layer was adjusted to 5~6 by addition of 1 M HCl (aq). Ethyl acetate (20 mL, two 10 mL portions) was added to the aqueous layer to extract the desired compounds. The organic layers were

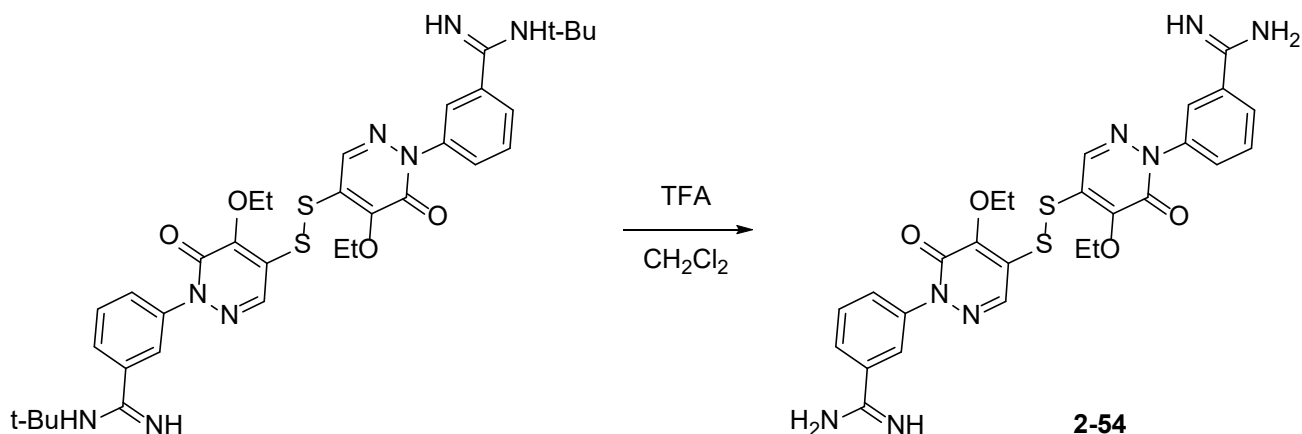
combined and dried over magnesium sulfate and concentrated to give the crude product, which was subjected to flash column chromatography on silica gel to give 45 mg of **2-10** as a white solid, 73%. R = H, R' = Et. mp 101 °C. <sup>1</sup>H NMR 7.72 (1H, s), 7.54 (2H, m), 7.46 (2H, m), 7.38 (1H, m), 4.63 (2H, q, *J* = 7.2 Hz), 4.04 (1H, s), 1.42 (3H, t, *J* = 7.2 Hz); <sup>13</sup>C NMR 155.76, 148.54, 141.16, 137.02, 128.80, 128.30, 125.51, 125.47, 68.73, 16.12. For the other analogues, the yields varied from 40-91%.

**1.8 General Procedure for the synthesis of disulfide dimers, e.g., Bis(4-ethoxy-2-phenyl-5-pyridazyl)disulfide, 2-17.**



A solution of **2-10** (10 mg, 0.028 mmol) in 2 mL of MeOH was stirred vigorously with the system open to the air. The progress of the reaction was monitored by TLC analysis. The solution was stirred for 3 h and the concentrated crude material was subjected to flash column chromatography on silica gel to give 11.9 mg of **2-17**, 85%. R = H, R' = Et. <sup>1</sup>H NMR 8.13 (1H, s), 7.55 (2H, m), 7.48 (2H, m), 7.39 (1H, m), 4.73 (2H, q, *J* = 7.2 Hz), 1.43 (3H, t, *J* = 7.2 Hz); <sup>13</sup>C NMR(DMSO) 155.36, 150.61, 141.44, 136.57, 128.97, 128.57, 126.09, 121.58, 68.81, 16.03.

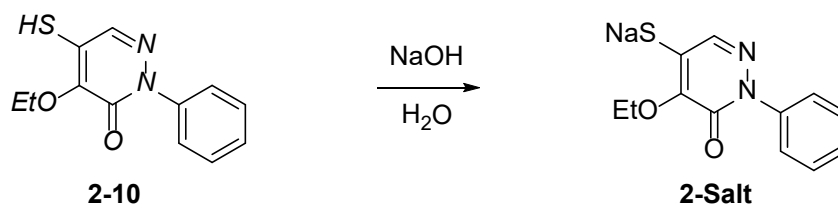
### 1.9 Procedure for the Synthesis of 3,3'-(disulfanediy)bis(5-ethoxy-6-oxopyridazine-4,1(6H)-diyl)dibenzimidamide, 2-54.



3,3'-(disulfanediy)bis(5-ethoxy-6-oxopyridazine-4,1(6H)-diyl)dibenzimidamide

To a solution of the *tert*-butyl protected amidine (110 mg, 0.16 mmol) in dichloromethane, was added trifluoroacetic acid (2 mL). The reaction mixture was heated to 50 °C (oil bath) and stirred for 12 h. The reaction mixture was concentrated and subjected to flash column chromatography on silica gel to give 50 mg of desired product, **2-54**, 54%. <sup>1</sup>H NMR 8.27 (1H, s), 8.06 (1H, t, *J* = 1.75 Hz), 7.93 (1H, dd, *J* = 7.5, 1.0 Hz), 7.74 (1H, dt, *J* = 7.5, 1.0 Hz), 7.59 (1H, t, *J* = 7.5 Hz), 4.65 (2H, q, *J* = 7.0 Hz), 3.35 (2H, bs) 1.41 (3H, t, *J* = 7.0 Hz); <sup>13</sup>C NMR 169.63, 155.84, 151.64, 141.39, 136.26, 134.54, 128.72, 128.67, 127.25, 126.42, 124.81, 69.23, 14.82

### 1.10 Procedure for the Synthesis of the salt of the mercaptan, 2-Salt



To a aqueous solution of NaOH (0.1 M, 1 mL) was added **2-10** (50 mg, 0.2 mmol) at 22 °C. The suspension was stirred for 1 min and EtOAc was added to extract the non-salt compounds from

the water-layer. The remaining aqueous solution containing the salt, **2-Salt** (~0.1 M), was used without further purification.

## References

1. Zhao MZ, Li J, Song ZG, Desmond R, Tschaen DM, Grabowski EJJ, Reider PJ (1998) A novel chromium trioxide catalyzed oxidation of primary alcohols to the carboxylic acids. *Tetrahedron Lett*; **39**: 5323-6.
2. Saluste CG, Whitby RJ, Furber M (2000) A palladium-catalyzed synthesis of amidines from aryl halides. *Angew Chem Int Edit*; **39**: 4156-8.



### **5.3.1 A Cell-based Screen in *Actinomyces oris* to Identify Sortase Inhibitors**



# A Cell-based Screen in *Actinomyces oris* to Identify Sortase Inhibitors

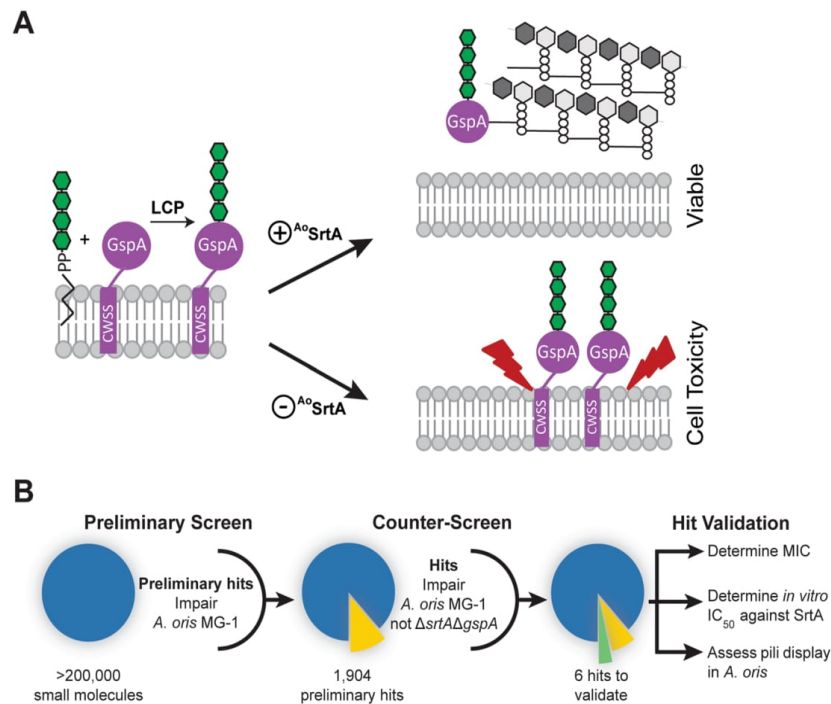
Jason E. Gosschalk<sup>1,2</sup>, Chungyu Chang<sup>3</sup>, Christopher K. Sue<sup>1,2</sup>, Sara D. Siegel<sup>4</sup>, Chenggang Wu<sup>4</sup>, Michele D. Kattke<sup>1,2</sup>, Sung Wook Yi<sup>1</sup>, Robert Damoiseaux<sup>5,6</sup>, Michael E. Jung<sup>1,7</sup>, Hung Ton-That<sup>3,5,✉</sup> & Robert T. Clubb<sup>1,2,7,✉</sup>

Sortase enzymes are attractive antivirulence drug targets that attach virulence factors to the surface of *Staphylococcus aureus* and other medically significant bacterial pathogens. Prior efforts to discover a useful sortase inhibitor have relied upon an *in vitro* activity assay in which the enzyme is removed from its native site on the bacterial surface and truncated to improve solubility. To discover inhibitors that are effective in inactivating sortases *in vivo*, we developed and implemented a novel cell-based screen using *Actinomyces oris*, a key colonizer in the development of oral biofilms. *A. oris* is unique because it exhibits sortase-dependent growth in cell culture, providing a robust phenotype for high throughput screening (HTS). Three molecules representing two unique scaffolds were discovered by HTS and disrupt surface protein display in intact cells and inhibit enzyme activity *in vitro*. This represents the first HTS for sortase inhibitors that relies on the simple metric of cellular growth and suggests that *A. oris* may be a useful platform for discovery efforts targeting sortase.

Proteins displayed on the surface of bacterial pathogens play critical roles in the infection process by promoting bacterial adhesion to host tissues, acquisition of essential nutrients, evasion and suppression of the immune response and host-cell entry<sup>1,2</sup>. Gram-positive bacteria display virulence factors using sortases, cysteine transpeptidase enzymes that covalently attach proteins to peptidoglycan precursors or assemble pili<sup>3–6</sup>. Sortase enzymes are potential drug targets as they are required for the virulence of methicillin-resistant *Staphylococcus aureus* (MRSA), which causes a wide range of life-threatening diseases, such as pneumonia, meningitis, osteomyelitis, endocarditis, toxic shock syndrome, bacteremia, and sepsis<sup>7</sup>. These infections are major health concerns as they are estimated to cause 16,485 fatalities in the United States each year<sup>8</sup>. Sortase enzymes also contribute to the virulence of other clinically important pathogens, including among others: *Enterococcus faecalis*, *Listeria monocytogenes*, *Bacillus anthracis*, *Streptococcus pyogenes* and *Streptococcus pneumoniae*<sup>9</sup>. Thus, small-molecule inhibitors of sortase enzymes may prove to be useful antimicrobial agents to treat infections caused by MRSA and other bacterial pathogens.

The sortase enzyme in *S. aureus* (<sup>SrtA</sup>) has been characterized to the greatest extent<sup>3–6</sup>. It recognizes protein substrates that harbor a C-terminal cell wall sorting signal (CWSS), which is comprised of a conserved LPXTG-type motif (where X denotes any amino acid), followed by a hydrophobic transmembrane segment and positively charged residues. The membrane-bound <sup>SrtA</sup> covalently attaches protein substrates to the peptidoglycan via a transpeptidation reaction by cleaving the LPXTG motif between the Gly and Thr residues and joining the cleaved LPXT to the cross-bridge peptide of lipid II, a peptidoglycan precursor. Catalysis occurs through a ping-pong mechanism that is initiated when the active-site cysteine nucleophile attacks the backbone carbonyl carbon of the threonine residue within the LPXTG motif, breaking the threonine-glycine peptide bond to create a thioacyl-linked sortase-protein complex<sup>10–13</sup>. The protein-lipid II product of the sortase catalyzed reaction is formed when the thioacyl substrate-enzyme intermediate is resolved by the amino group within lipid II. Cell wall synthesis reactions then incorporate the protein-lipid II product into the peptidoglycan, displaying the protein on the microbial surface. Over 3,100 species of bacteria contain genes encoding for sortases related to <sup>SrtA</sup><sup>14,15</sup>,

<sup>1</sup>Department of Chemistry and Biochemistry, University of California, Los Angeles, USA. <sup>2</sup>UCLA-DOE Institute of Genomics and Proteomics, University of California, Los Angeles, USA. <sup>3</sup>Division of Oral Biology and Medicine, University of California, Los Angeles, USA. <sup>4</sup>Department of Microbiology and Molecular Genetics, University of Texas Health Science Center, Houston, TX, USA. <sup>5</sup>Department of Molecular and Medicinal Pharmacology, University of California, Los Angeles, USA. <sup>6</sup>California NanoSystems Institute, University of California, Los Angeles, USA. <sup>7</sup>Molecular Biology Institute, University of California, Los Angeles, 611 Charles Young Drive East, Los Angeles, CA, 90095, USA. ✉e-mail: [htonthat@dentistry.ucla.edu](mailto:htonthat@dentistry.ucla.edu); [rclubb@mbi.ucla.edu](mailto:rclubb@mbi.ucla.edu)



**Figure 1.** Design and overall work-flow of cell-based inhibitor screen. (A) Schematic showing how the activity of the *A. oris* SrtA ( $^{Ao}$ SrtA) enzyme is required for cell viability. A fully functioning  $^{Ao}$ SrtA enzyme is needed to attach the glycosylated GspA protein (colored purple with green glycosylation) to the cell wall (top). Reduced gene expression of the  $^{Ao}$ SrtA enzyme has been shown to be lethal, presumably because of GspA accumulation in the membrane (bottom). Lethality is dependent upon glycosylation of GspA by the LytR-CpsA-Psr enzymes (LCP). (B) Overview of the sortase inhibitor screen. The effects of small molecules on wild-type *A. oris* MG-1 were determined for several compound libraries (left). Preliminary hits (1,904) that impaired growth were counter-screened by determining their ability to affect the growth of a  $\Delta srtA/\Delta gspA$  strain, whose viability is dependent upon the activity of  $^{Ao}$ SrtA (middle). Strain-specific growth inhibitors were then validated for sortase inhibitor activity using biochemical and cellular approaches (right). Adobe Illustrator Version: 15.0.0 (<https://www.adobe.com/products/illustrator.html>).

but in some instances these enzymes perform lysine-isopeptide transpeptidation reactions that construct pili virulence factors<sup>16</sup>.

Given its potential as a drug target, because sortase mutants are attenuated in virulence, considerable effort has been put forth to discover small molecule  $^{Sa}$ SrtA inhibitors<sup>9,17–19</sup>. Previously reported studies have searched for inhibitors by monitoring the activity of the purified  $^{Sa}$ SrtA enzyme *in vitro* using a Förster resonance energy transfer (FRET) assay. The FRET assay has been used to screen small-molecule compound libraries<sup>20–23</sup> and to assess the potency of rationally designed peptidomimetics<sup>24–27</sup>, natural products<sup>28–38</sup>, and small molecules identified using virtual screening approaches<sup>39–42</sup>. While several inhibitors have been discovered, to the best of our knowledge, none have advanced into clinical trials. It is possible that some of these compounds are unable to effectively inhibit the enzyme in its natural context, the extra-cellular bacterial membrane where it may associate with components of the protein secretion and cell wall synthesis machinery. A cell-based assay for high-throughput screening (HTS) for sortase inhibitors could overcome this limitation, but has yet to be implemented in *S. aureus* because inhibiting  $^{Sa}$ SrtA activity does not significantly affect the growth or morphology of this microbe in cell culture<sup>6</sup>.  $^{Sa}$ SrtA activity can be detected in cells, but these methods are cumbersome and require antibody detection of sortase-displayed proteins<sup>41</sup>, cell adhesion assays<sup>43</sup> or incubation of cells with fluorogenic peptidyl sortase substrates that can be slow to label cells<sup>44</sup>.

Recently, Wu *et al.* made the surprising discovery that the viability of the oral bacterium *Actinomyces oris* MG-1 in cell culture depends on the activity of its sortase ( $^{Ao}$ SrtA)<sup>45,46</sup>.  $^{Ao}$ SrtA anchors the glycosylated surface protein A (GspA) to the cell wall. In this process, GspA is first glycosylated by the LCP enzyme and then attached to the cell wall by  $^{Ao}$ SrtA via lipid II (Fig. 1A, top). Interestingly, reducing  $^{Ao}$ SrtA expression causes cell arrest, presumably due to glycol-stress caused by accumulation of glycosylated GspA in the membrane (Fig. 1A, bottom). To the best of our knowledge, *A. oris* is the only known bacterium that exhibits a sortase-dependent growth

phenotype in cell culture. Here we report the development of a cell-based assay to screen for sortase inhibitors that takes advantage of this unique phenotype. High throughput implementation of the assay was used to screen compound libraries and led to the discovery of several small molecule sortase inhibitors that are validated using biochemical and cellular approaches.

## Results

**Development and implementation of the cell-based screen.** Our previous studies have shown that reducing *srtA* gene expression in *A. oris* leads to cell death and that this phenotype is conditionally dependent on *gspA* expression<sup>45,46</sup>. To exploit this unique dependence, we developed a cell-based assay to screen for sortase inhibitors that compares the growth-inhibitory effects of small molecules cultured with *A. oris* MG-1 (wild-type) versus a  $\Delta srtA/\Delta gspA$  strain. In the assay, it is presumed that a small molecule sortase inhibitor will selectively impair growth of *A. oris* MG-1 by causing GspA to accumulate in the membrane, whereas growth of the  $\Delta srtA/\Delta gspA$  strain should be unaffected. Prior to implementing the assay in HTS, we optimized conditions for *A. oris* growth in 384-well plate format. Both MG-1 and  $\Delta srtA/\Delta gspA$  strains had similar growth rates and final cell densities in brain-heart infusion broth. Control growth experiments using media containing 1% dimethyl sulfoxide (DMSO) confirmed that both strains are tolerant to this solvent at concentrations used for screening. The growth phenotype of both *A. oris* strains was sufficient for use in a HTS, as a *Z'* score of 0.81 (MG-1) and 0.71 ( $\Delta srtA/\Delta gspA$ ) was determined when positive (1  $\mu\text{g mL}^{-1}$  penicillin) and negative (media only) control experiments are performed<sup>47</sup>.

The overall workflow for the HTS is shown in Fig. 1B. Initially, small molecules were screened for their ability to kill the wild-type MG-1 strain in 384-well format. Molecules that reduced growth by more than 2.75 standard deviations from the average growth were considered preliminary hits. These molecules were then tested in high-throughput for their ability to impair  $\Delta srtA/\Delta gspA$  growth, which serves as a counter-screen. Small molecules that exhibited differential growth effects in MG-1 but not  $\Delta srtA/\Delta gspA$  strains were considered hits and carried forward for further analysis. These hits were validated by determining their minimum inhibitory concentrations (MICs) in cell culture, measuring the *in vitro* half maximal inhibitory concentration of each compound ( $\text{IC}_{50}$ ) against the isolated <sup>60</sup>SrtA and <sup>58</sup>SrtA enzymes, and their ability to affect protein display in *A. oris* cells.

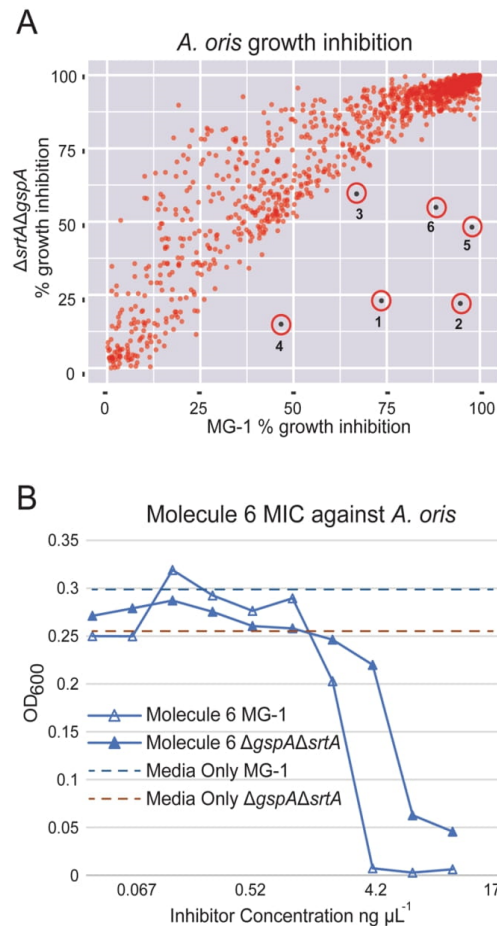
In the primary screen, a total of 200,834 small molecules were tested for their ability to impair *A. oris* MG-1 growth. Cells were added to an optical density ( $\text{OD}_{600}$ ) of 0.01 in media containing 10  $\mu\text{M}$  of each molecule, followed by a 15 hour incubation at 37 °C and end-point  $\text{OD}_{600}$  measurement. The growth effect of each molecule was normalized and expressed as a percent growth relative to the DMSO-only control (see Methods). A total of 1,904 molecules were classified as preliminary hits (0.95% of the molecules tested), as they reduced MG-1 growth after 15 hours by at least 34%, or 2.75 standard deviations below the average. The preliminary hits were then subjected to a counter-screen in which their growth effects on  $\Delta srtA/\Delta gspA$  were determined. In order to eliminate potentially erroneously identified preliminary hits, the counter-screen and preliminary screen was performed in duplicate for the 1,904 preliminary hit molecules. A molecule was deemed as a hit if it caused greater than 15% differential growth effect in duplicate when the MG-1 and  $\Delta srtA/\Delta gspA$  strains were compared. Figure 2 shows a plot of each preliminary hit molecule's growth effect against the MG-1 and  $\Delta srtA/\Delta gspA$  strains. Data is plotted as the percentage of growth inhibition for each molecule (see Methods).

A total of six compounds preferentially affected MG-1 growth in duplicate, constituting 0.3% of the preliminary hits identified from the primary screen and 0.003% of the total number of compounds tested. The six compounds were termed 1–6 and carried forward for further analysis (Fig. 3). R Version: 3.3.3 (<https://www.r-project.org/>)

**Evaluation of the sortase inhibitor candidates.** Two validation assays were performed for each hit molecule that measured: (i) the minimum inhibitory concentration (MIC) required to prevent growth of the MG-1 and  $\Delta srtA/\Delta gspA$  strains, and (ii) their *in vitro* inhibitory activity against the enzymatic activity of the *A. oris* and *S. aureus* sortase enzymes. Initially, the selective growth effect originally observed in the HTS that used 384-well plates was more rigorously defined by determining each molecule's MIC value for the MG-1 and  $\Delta srtA/\Delta gspA$  strains. Cells were cultured in 100  $\mu\text{L}$  media containing 2-fold dilutions of each small molecule between 320  $\mu\text{M}$  and 0.039  $\mu\text{M}$  (approximately 134  $\mu\text{g mL}^{-1}$  to 0.16  $\mu\text{g mL}^{-1}$ ). Each of the six preliminary hit molecules (1–6) has a lower MIC for the MG-1 strain (between 16.4  $\mu\text{g mL}^{-1}$  (40  $\mu\text{M}$ ) and 1.0  $\mu\text{g mL}^{-1}$  (2.6  $\mu\text{M}$ ) than for  $\Delta srtA/\Delta gspA$  (Table 1). Two additional molecules exhibited activity in only one replicate of the high-throughput counter-screen assay; however, we chose to interrogate their activity in proceeding assays to better appreciate the reliability of the HTS. Neither exhibited marked differences in their MIC values when tested in MG-1 and  $\Delta srtA/\Delta gspA$  strains, which affirms the strength of the duplicate experiment cutoff. Thus, the MIC data confirm the results of the HTS by demonstrating that only the duplicate hit molecules preferentially impair the growth of *A. oris* MG-1.

The ability of the hit molecules to inhibit sortase activity was measured *in vitro* using an established FRET-based assay<sup>13</sup> that measures the ability of each enzyme to cleave a fluorogenic peptide substrate; Abz-LAQTG-Dap(Dnp)-NH<sub>2</sub> and Abz-LPETG-Dap(Dnp)-NH<sub>2</sub>, substrates for <sup>60</sup>SrtA and <sup>58</sup>SrtA, respectively. Each enzyme cleaves between the threonine and glycine bond in these peptides, leading to a measurable increase in fluorescence (see Methods). The  $\text{IC}_{50}$  value of each compound was determined for both the <sup>60</sup>SrtA and <sup>58</sup>SrtA enzymes. Compounds 3, 4 and 6 are bona fide enzyme inhibitors, as each inhibits both the <sup>60</sup>SrtA and <sup>58</sup>SrtA enzymes with  $\text{IC}_{50}$  values ranging from 30–70  $\mu\text{g mL}^{-1}$  (73–170  $\mu\text{M}$ ) (Table 1). In contrast, molecules 1, 2 and 5 failed to inhibit <sup>60</sup>SrtA *in vitro*, suggesting that they selectively affect *A. oris* MG-1 growth through a mechanism not involving sortase inhibition.

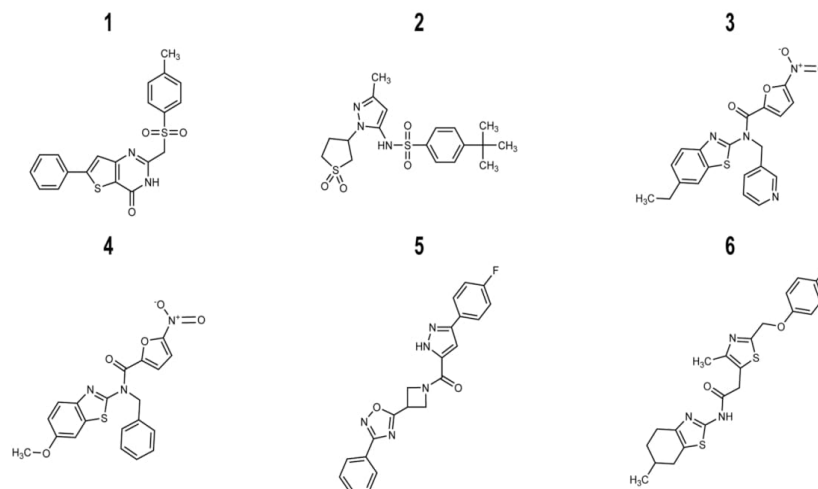
**Sortase inhibitors affect pilus assembly in *A. oris*.** The MIC and  $\text{IC}_{50}$  data suggest that at least three of the hit molecules are capable of inhibiting <sup>60</sup>SrtA either within the context of the cell or as an isolated



**Figure 2.** Growth effects of the screened molecules. **(A)** Scatter plot showing the growth effects of small molecules when cultured with either *A. oris* wild-type MG-1 or mutant strain  $\Delta srtA/\Delta gspA$ . The percent growth inhibition for each strain is relative to the parent strain cultured in the absence of the small molecule (Methods). Molecules (1–6) exhibited statistically significant strain-specific effects on growth, impairing growth of the wild-type more than the  $\Delta srtA/\Delta gspA$  strain (circled and labeled). **(B)** Plot showing the effects of molecule 6 on the growth *A. oris* MG-1 (open triangles) and  $\Delta srtA/\Delta gspA$  (filled triangles) strains. The corresponding strains grown in identical conditions in the absence of molecule exhibited  $OD_{600}$  values of 0.29 and 0.26, respectively. The data show that 6 has a lower MIC for MG-1 as compared to  $\Delta srtA/\Delta gspA$ , and is consistent with the HTS data shown in panel (A).

transpeptidase. To understand the physiological effects that these molecules have on the display of pili and surface proteins on the surface of *A. oris*, we grew cells with  $10\ \mu M$  of each hit molecule and assessed sortase activity by western blot and electron microscopy as previously reported<sup>45</sup>. To avoid a confounding problem that inhibiting  $A^o$ SrtA causes cell arrest, we performed the experiments in a mutant devoid of *gspA*; the aforementioned genetic suppressor of sortase lethality. Consistent with both MIC and  $IC_{50}$  data, when samples were immunoblotted with antibodies against the type 2 pilus shaft protein FimA, both molecules 3 and 4 decreased FimA polymers as compared to the control (Fig. 4; compare the first two lanes with lanes containing molecules #3 and #4).

To examine if these molecules inhibit cell wall anchoring of surface proteins (e.g. GspA), we similarly treated cells devoid of *lcpA*, i.e.  $\Delta lcpA$ , another suppressor of *srtA*<sup>46</sup>. When blotted with antibodies against GspA, no GspA polymers were observed in samples treated with molecules 3, 4 and 6 (Fig. 4B). To corroborate these observations, we subjected *A. oris* cells of  $\Delta gspA$  and  $\Delta lcpA$  strains to  $10\ \mu M$  of each molecule (3, 4 and 6) and studied pilus display with transmission electron microscopy (TEM), whereby *A. oris* cells were immobilized on nickel grids and stained with 1% uranyl acetate prior to viewing by an electron microscope. Intriguingly, compared to the untreated control, treatment of molecules 3, 4 and 6 caused significant reduction of pilus assembly at the cell



**Figure 3.** Chemical structures of preliminary hit molecules. Chemical structures of the hit molecules (1–6) from the *A. oris* screen. These molecules exhibit statistically significant strain-specific effects on growth, impairing the wild-type more than the  $\Delta srtA/\Delta gspA$  strain. Compounds 3, 4 and 6 were validated sortase inhibitors, whereas 1, 2 and 5 selectively impair growth through an unknown, non-sortase dependent mechanism. ChemDraw Professional Version: 19.0.1.28 (<https://www.perkinelmer.com/product/chemdraw-professional-chemdrawpro>).

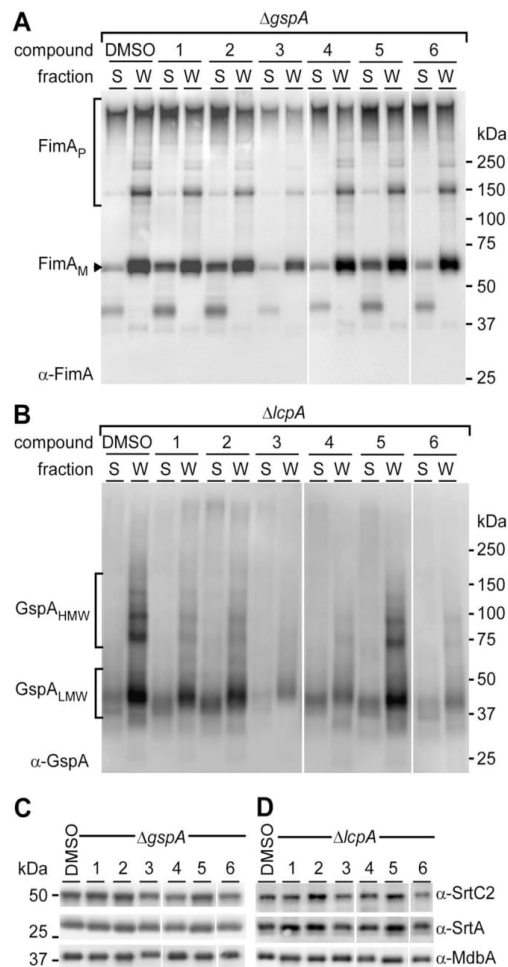
|  | Preliminary Hits |             |             |             |             |              |
|--|------------------|-------------|-------------|-------------|-------------|--------------|
|  | 1                | 2           | 3           | 4           | 5           | 6            |
| Differential Growth <sup>a</sup>                   | 65.7 ± 0.5%      | 96.2 ± 0.3% | 17.5 ± 3.5% | 38.5 ± 0.5% | 96.2 ± 0.3% | 71.8 ± 16.5% |
| MIC (MG-1) <sup>b</sup>                            | 2.0              | 0.5         | 8.2         | 16.4        | 1.0         | 2.1          |
| MIC ( $\Delta gspA/\Delta srtA$ ) <sup>b</sup>     | 4.0              | 2.1         | >130        | >130        | 2.0         | 8.4          |
| IC <sub>50</sub> ( <i>A. oris</i> ) <sup>c</sup>   | >190             | >200        | 30 ± 20     | 70 ± 50     | >190        | 60 ± 40      |
| IC <sub>50</sub> ( <i>S. aureus</i> ) <sup>c</sup> | 80 ± 20          | >200        | 90 ± 20     | 200 ± 100   | 110 ± 30    | 54 ± 8       |

**Table 1.** Growth and inhibitory properties of the preliminary hit molecules. <sup>a</sup>Differential effects of small molecule on wild-type MG-1 and  $\Delta gspA\Delta srtA$  *A. oris* strains measured in the high-throughput screen. See methods section complete definition. <sup>b</sup>Minimum inhibitor concentration (MIC) expressed in  $\mu\text{g ml}^{-1}$  for MG-1 and  $\Delta gspA\Delta srtA$  *A. oris* strains. <sup>c</sup>Half maximal inhibitory concentration for the *in vitro* enzyme activity of <sup>Sa</sup>SrtA (*S. aureus*) or <sup>Ao</sup>SrtA (*A. oris*) in  $\mu\text{g ml}^{-1}$ .

pole in either strain background, i.e.  $\Delta gspA$  or  $\Delta lcpA$  (Fig. 5). The results suggest that these molecules target the nascent pilus assembly machine.

## Discussion

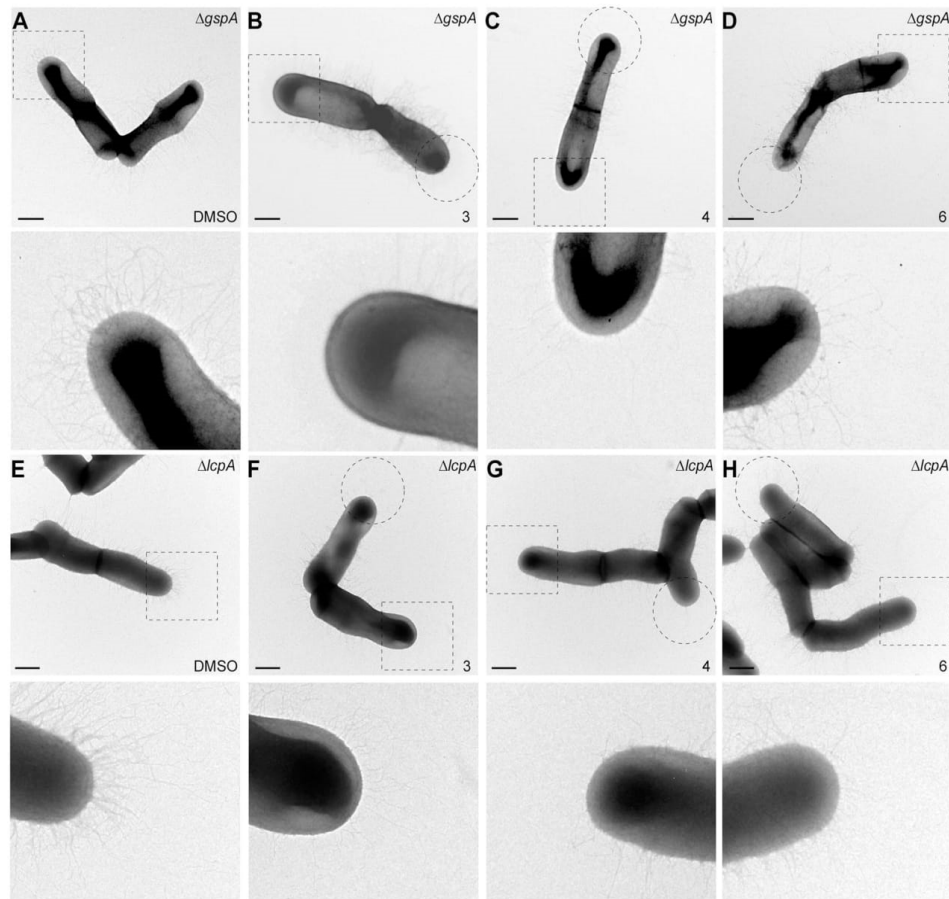
Sortase enzymes are promising drug targets as they mediate the display of important virulence factors in several medically significant bacteria. The *S. aureus* <sup>Sa</sup>SrtA enzyme is particularly interesting because it plays an important role in MRSA infections that are a major cause of human mortality. Despite over fifteen years of effort by a number of research groups, antivirulence agents that work by inhibiting <sup>Sa</sup>SrtA have yet to be discovered. This may be because nearly all screens for <sup>Sa</sup>SrtA inhibitors have used the isolated enzyme that is truncated at its N-terminus to increase solubility. Searching for inhibitors using a cell-based assay may yield better results as it could discover small molecules that are capable of traversing the cell wall and targeting the intact enzyme in the microbial membrane. Moreover, it could overcome screening limitations that are caused by the slow enzyme kinetics of the isolated protein. To the best of our knowledge, a robust cell-based sortase assay has yet to be developed, presumably because current methods used to monitor <sup>Sa</sup>SrtA activity in intact cells are laborious, requiring multi-step immunolabeling and fluorescence-detection experiments. To overcome this limitation, we developed and implemented in high throughput a cell-based assay to identify sortase inhibitors. Our approach exploits the unique growth dependence of *A. oris* on the activity of its SrtA enzyme, which is conditionally dependent on glycosylation of the GspA surface protein by an LCP enzyme (Fig. 1A). First, a primary screen was performed in which >200,000 small molecules were tested for their ability to impair the growth of wild-type *A. oris* MG-1 (Fig. 1B). Preliminary hit molecules were then counter-screened against the *A. oris*  $\Delta srtA/\Delta gspA$  mutant strain to eliminate small molecules that impaired *A. oris* growth via processes unrelated to sortase. In the primary screen,



**Figure 4.** Assessment of cell surface proteins. (A,B) Cell cultures of *A. oris* strains  $\Delta gspA$  (A) and *A. oris*  $\Delta lcpA$  (B) harvested after inhibitor treatment were subjected to cell fractionation. Samples in supernatant culture medium (S) and cell wall (W) fractions were immunoblotted with antibodies against the fimbrial shaft FimA ( $\alpha$ -FimA) and glycosylated GspA ( $\alpha$ -GspA). (C,D) The protoplast fractions were subjected to immunoblotting with antibodies against pilus-specific sortase SrtC2 ( $\alpha$ -SrtC2), housekeeping sortase SrtA ( $\alpha$ -SrtA), and the membrane protein MdbA ( $\alpha$ -MdbA) as control.

most molecules did not affect *A. oris* MG-1 growth when present in the cell culture at 10  $\mu$ M, as the mean and median percent growth of cultures containing small molecules relative cells grown in standard growth culture containing DMSO was 99.3% and 100.9%, respectively. The interquartile range (IQR) is 12.1%, meaning that 75% of molecules cells grow to 94.2% the average DMSO-only density or better. A total of 1,904 molecules in the screen reduced MG-1 growth to 66% growth (2.75 standard deviations below the average percent growth) and were subsequently counter-screened using the  $\Delta srtA/\Delta gspA$  mutant strain. For these molecules, growth effects on each strain were tested in duplicate on separate days to rigorously identify molecules that selectively impaired MG-1 growth. Six preliminary hit molecules were identified, 1–6 (0.32% of the 1,904 MG-1 growth effecting molecules). Each of the six molecules inhibit the growth of the MG-1 strain more than the  $\Delta gspA/\Delta srtA$  strain when tested in duplicate. The finding that few molecules cause selective growth effects is not surprising, as most molecules can be expected to be generally cytotoxic by inhibiting other essential non-sortase pathways within cell, e.g. division and respiration.

Repeating the counter-screen in duplicate significantly eliminated false positive results. This is evident by our finding that the six preliminary hits also exhibit strain-specific growth effects when they were re-purchased and tested in larger culture volumes (Table 1); each of the preliminary hits (1–6) exhibited lower MIC values against



**Figure 5.** Detection of pili by electron microscopy. Bacterial cultures of strains  $\Delta gspA$  (A–D) and  $\Delta lcpA$  (E–H) treated with compounds 3, 4 and 6 were subjected to negative staining with 1% uranyl acetate prior to electron microscopy; scale bar of 500 nm. Polar assembly of pili is heightened with dashed circles and rectangles. Enlarged areas in dashed rectangles are shown in below panels.

MG-1 than against  $\Delta gspA/\Delta srtA$ . The MIC values against MG-1 fall within a similar range, around  $10\ \mu\text{M}$  for all hits, which is expected, as the primary screen and counter-screen monitored growth effects at a single concentration. It is worth noting that the growth screen may have failed to detect potential sortase inhibitors with MIC values significantly greater than  $10\ \mu\text{M}$ , meaning higher concentrations of the small molecule are necessary for enzyme inhibition. In principle, this limitation could be overcome by repeating the HTS using higher concentrations of the small molecules in the screen, but this was not performed here because it was cost-prohibitive.

*In vitro* enzyme activity testing reveals that compounds 3, 4 and 6 inhibit both the  $^{\text{Ao}}\text{SrtA}$  and  $^{\text{Sa}}\text{SrtA}$  enzymes with  $\text{IC}_{50}$  values ranging from  $30$  to  $70\ \mu\text{g mL}^{-1}$  ( $73$ – $170\ \mu\text{M}$ ). Compounds 3 and 4 are bioisosteres and share related N-(1,3-benzothiazol-2-yl)-5-nitrofuranyl-2-carboxamide chemical scaffold. In compound 3 the scaffold is elaborated with N-3-pyridylmethyl and 6-ethyl-2-benzothiazole moieties, whereas 4 features benzyl and 5-methoxybenzothiazole. Effectively these compounds are bioisosteres, as the benzene is an isostere with pyridine, as are the ethyl and methoxy groups. Interestingly, molecules containing 5-nitrofuranyl functional groups are particularly potent at eliciting selective growth effects as they are enriched in our primary hits; 1% of the 1,904 hits from the cell based screen contained this functional group even though it is present in only 0.09% of 200,834 molecules that we screened. Compound 6 is unique, exhibiting a N-methyl-(4,5,6,7-tetrahydro-1,3-benzothiazol-2-yl)-1,3-thiazole-5-carboxamide scaffold. Importantly, compound 6 would not likely have been discovered using conventional FRET-based HTS as it is intrinsically fluorescent and presumably would have been disregarded as a hit. The  $\text{IC}_{50}$  values for the inhibitors against  $^{\text{Ao}}\text{SrtA}$  are higher than their measured MIC values. This suggests that they may also kill *A. oris* cells through non-sortase related mechanisms and/or that they more efficiently inhibit the *A. oris* enzyme when it is embedded in the microbial membrane. Preliminary hits 1, 2 and 5 also effect growth



in a strain-specific manner, but do not inhibit  $^{AoSrtA}$  *in vitro* (Table 1). The origins of their selective effects on growth remain to be determined. To the best of our knowledge, none of the hit molecules have been described as antimicrobials, however benzothiazole-based DNA gyrase B inhibitors have been shown to have modest growth inhibitor properties against Gram-positive *Enterococcus faecalis* when dosed at high concentrations (50  $\mu\text{M}$ )<sup>48</sup>.

Cellular studies indicate that sortase inhibitors 3, 4 and 6 alter the ability of *A. oris* to display GspA and to assemble pili on its surface. *A. oris* uses three distinct sortases to elaborate its cell envelope.  $^{AoSrtA}$  is a class E housekeeping sortase that attaches GspA and other proteins containing the LAQTG sorting signal to the cell wall<sup>49</sup>.  $^{AoSrtA}$  catalyzes a transpeptidation reaction that attaches GspA to lipid II, a cell wall precursor. *A. oris* also assembles surface pili (fimbriae) using two class C sortases, SrtC1 and SrtC2. These sortases catalyze transpeptidation by linking protein subunits of the pilus together via lysine-isopeptide bonds. Subsequent cell wall anchoring of pilus polymers to peptidoglycan requires  $^{AoSrtA}$ <sup>49</sup>. SrtC1 produces type 1 fimbriae that are comprised of the fimbrial shaft FimP protein and the tip fimbrellin FimQ protein. This structure mediates bacterial adherence to the tooth surface via FimQ interactions with salivary proline-rich protein deposits. The SrtC2 sortase produces type 2 fimbriae, made of the fimbrial shaft FimA protein and tip fimbrellin FimB; these fimbriae are required for bacterial adherence to host cells, biofilm formation, and bacterial coaggregation<sup>50,51</sup>. We performed cell fraction and immunoblot studies to gain insight into how inhibitors 3, 4 and 6 affect SrtC1 and  $^{AoSrtA}$  sortase activity in intact cells (Fig. 4). Molecule 6 inhibits  $^{AoSrtA}$  in cells as its presence significantly diminishes GspA display (Fig. 4B). In contrast, it has little effect on type 2 pilus production by the SrtC2 sortase as judged by immunoblots of the FimA shaft protein in fractionated cells. The diminished potency of molecule 6 against pili display indicates that it lacks inhibitory activity against class C sortases. This is substantiated by TEM images, which show that molecule 6 only modestly effects pili display in the  $\Delta\text{gspA}$  strain (Fig. 5A–D). Since  $^{AoSrtA}$  has been implicated in anchoring of the pilus to the peptidoglycan, one might still expect to see diminished pili display in the absence of the housekeeping sortase; however, it is not uncommon for class C enzymes to compensate for loss of the housekeeping sortase and independently display pili as SrtC2 is capable of catalyzing pilus polymerization and cell wall anchoring of pilus polymers<sup>49,52</sup>. Thus, the data supports molecule 6 as being more selective for  $^{AoSrtA}$  in intact cells.

Cellular studies suggest that compounds 3 and 4 inhibit both class E  $^{AoSrtA}$  and class C SrtC2 sortases. Our interrogation of type 2 pilus assembly reveals that 3 and 4 not only decrease the abundance of pili, but also reduce the amount of higher molecular weight FimA-containing polymers that are being formed (Fig. 4A, compare molecular weight bands  $\geq 150\text{kDa}$ ). Decreased pilin-polymerase activity suggests that SrtC2 is being inhibited. This is substantiated by TEM images of cells treated with these compounds 3 and 4, which show either completely absent or diminished pili at their poles, respectively (Fig. 5B,C). Inspection reveals that the pili in these cells become gradually shorter toward the poles, suggesting that the SrtC2 enzyme is being inhibited during the 3 hour period of growth. Moreover, cells treated with compounds 3 and 4 have decreased amounts of cell wall associated GspA, compatible with the fact that these molecules also inhibit the housekeeping  $^{AoSrtA}$  sortase (Fig. 4B). Consistent with the lower  $\text{IC}_{50}$  against  $^{AoSrtA}$  and  $^{SrtA}$  and the lower MIC against MG-1 relative to 4, we note that 3 appears to more severely limit GspA display than 4 at 10  $\mu\text{M}$  (Figs. 4B and 5B,C). This suggests that 3, an analog of 4, is the more potent molecule. It is important to note that these inhibitors were added to actively growing cell cultures that already produced pili. Given the presence of pili at the septal area and significantly reduced pili at the pole (Fig. 5), we surmise that these inhibitors target the new sortase machine at the pole, where the nascent peptidoglycan is synthesized and modified.

In conclusion, we have exploited the unique sortase-dependent growth phenotype of *A. oris* to screen small molecule compound libraries for sortase inhibitors. Three molecules, representing two unique scaffolds, inhibit sortase enzymes from both *A. oris* and *S. aureus* were discovered. This represents the first HTS for sortase inhibitors that relies on the simple metric of cellular growth and suggests that *A. oris* is a promising platform for sortase-targeted drug discovery. Future work will need to establish structure-activity relationships for the hit molecules to further increase their potency for potential use as novel anti-infectives, which are urgently needed to treat infections caused by MRSA and other drug-resistant bacterial pathogens.

## Materials and Methods

**Antibiotics, media and consumables.** Brain Heart Infusion Broth, modified (BHI) was purchased from Fisher Scientific and prepared as directed within one week of its use. Penicillin G and kanamycin were purchased from Fisher Scientific and stored as directed. 1000x antibiotic stocks were prepared in water, filtered with 0.2  $\mu\text{m}$  syringe filters, and stored at -20  $^{\circ}\text{C}$  until thawed immediately before use. Greiner 384-well plates (EK-30162) and universal lids (EK-2079) were purchased from E&K scientific and used once before disposal as medical waste.

**High-throughput screen.** A total of 200,834 small molecule compounds (Molecular Screening Shared Resource, UCLA) were dissolved in 100% Omnisolv methylsulfoxide (MilliporeSigma MX1456P-6) and stored in 384-well polypropylene plates. *Actinomyces oris* MG-1 and  $\Delta\text{gspA}\Delta\text{srtA}$  were streaked out on BHI agar plates containing 50  $\mu\text{g}/\text{mL}$  kanamycin and grown at 37  $^{\circ}\text{C}$  for 48 hours to allow single colonies to grow. Cultures were started from 2–3 colonies in BHI with 50  $\mu\text{g}/\text{mL}$  kanamycin and allowed to grow until an  $\text{OD}_{600}$  between 0.1–0.4. Meanwhile, 384-well plates were filled with 25  $\mu\text{L}$  BHI broth with 100  $\mu\text{g}/\text{mL}^{-1}$  kanamycin (columns 1–22) or 200  $\mu\text{g}/\text{mL}^{-1}$  Penicillin G (positive control, columns 23–24). Five hundred nanoliters of small molecules were transferred into the media (columns 3–22) using the Biomex FX<sup>3</sup> automated work station with a 384-well pin tool. Precultures were diluted to an  $\text{OD}_{600}$  of 0.02 in 1 L BHI broth without antibiotics. Twenty five microliters of cell culture was immediately aliquoted into plates containing 25  $\mu\text{L}$  media, antibiotics, and small-molecules using a Multidrop (Thermo LabSystems), resulting in 50  $\mu\text{g}/\text{mL}^{-1}$  kanamycin and 10  $\mu\text{M}$  small molecules with 1% DMSO or 100  $\mu\text{g}/\text{mL}^{-1}$  Penicillin G. Plates were immediately lidded and placed into a humidified Cytomat 6001 incubator at 37  $^{\circ}\text{C}$  and allowed to grow for 15.5 hours. Following incubation, plates were removed from the

Cytomat using a Thermo Spinnaker robotic arm on a rail, de-lidded, and placed into an EnVision high-speed plate reader for optical density measurement at 620 nm.

Raw data obtained from the screen was formatted in-house to upload to the Collaborative Drug Discovery Vault ([www.collaborativedrug.com](http://www.collaborativedrug.com)). Individual optical density readings were converted to a percent growth value  $(\text{OD}_{\text{sample}} - \text{OD}_{\text{Positive control}}) / (\text{OD}_{\text{Negative Control Average}} - \text{OD}_{\text{Positive control}})$  and percent growth inhibition (1-percent growth). The Z-factors for the individual molecules were determined. Molecules with a Z-factor less than or equal to -2.75 were considered preliminary hits to advance forward a generous collection of preliminary hit molecules. Five microliters of each of the 1904 preliminary hit molecules were re-arrayed into new 384-well polypropylene plates. Each of the 1904 preliminary hits were assayed against both the MG-1 and  $\Delta\text{gspA}\Delta\text{srtA}$  strains in duplicate, on two separate days. Percent difference of growth was calculated according to the equation  $100 - 100 \{ (\text{OD}_{\text{MG-1, Sample}} - \text{OD}_{\text{MG-1, Positive Control}}) / \text{OD}_{\Delta\text{gspA}\Delta\text{srtA, Sample}} - \text{OD}_{\Delta\text{gspA}\Delta\text{srtA, Positive Control}} \} / [ (\text{OD}_{\text{MG-1, Average}} - \text{OD}_{\text{MG-1, Positive Control}}) / \text{OD}_{\Delta\text{gspA}\Delta\text{srtA, Average}} - \text{OD}_{\Delta\text{gspA}\Delta\text{srtA, Positive Control}} ]$ . A score of 100 indicates that a molecule causes extreme difference in cell survivability between the two strains whereas a score of 0 indicates no differential effect on cellular growth. Molecules were ranked according to their relative growth.

**Hit validation assays.** Minimum inhibitory concentration (MIC) values were determined for molecules 1–6 according to the Clinical Laboratory Standard Institutes' (CLSI) Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria that Grow Aerobically, Approved standards-ninth edition; M07-A9 Vol. 32 No. 2 with the following alterations. Overnight cultures of MG-1 and  $\Delta\text{gspA}\Delta\text{srtA}$  were diluted to  $\text{OD}_{600}$  values of 0.01 before being added to BHI containing the appropriate molecule in 96-well plates. Mueller-Hinton broth was not usable because *A. oris* is a fastidious bacterium. Plates were sealed with Breatheasy seals and incubated for 18 hours, as *A. oris* has a doubling time of 2 hours. Plate seals were removed and  $\text{OD}_{600}$  values measured. Each plate was run with a positive control (cells, media, and Penicillin G), negative control (media and cells only), and contamination control (media only).

$\text{IC}_{50}$  values for the six hit molecules against the  $^{\text{A}0}\text{SrtA}$  and  $^{\text{S}0}\text{SrtA}$  enzymes were determined as previously described, with some modifications. Briefly, molecules 1–6 were serially diluted 2-fold (from 1.25 mM to 2.44  $\mu\text{M}$ , 500  $\mu\text{M}$  to 0.98  $\mu\text{M}$  final assay concentration) into 25  $\mu\text{M}$   $^{\text{A}0}\text{SrtA}$  or  $^{\text{S}0}\text{SrtA}$  (final assay concentration 10  $\mu\text{M}$ ) in buffer A (20 mM HEPES pH 7, 5 mM  $\text{CaCl}_2$ , 0.05% TWEEN, 30% DMSO). The final DMSO concentration was 18% in the assay. Samples were incubated 1 hour at room temperature and aliquoted into a 384-well plate (EK-30892). Thirty microliters of FRET peptide (Peptide 2.0) in buffer B (20 mM HEPES pH 7, 5 mM  $\text{CaCl}_2$ , 0.05% TWEEN), Abz-LPATG-Dap(Dnp)- $\text{NH}_2$  for  $^{\text{S}0}\text{SrtA}$  or Abz-LAQTG-Dap(Dnp)- $\text{NH}_2$  for  $^{\text{A}0}\text{SrtA}$  was added and the plate was immediately placed into a FlexStation II plate reader (Molecular Devices) and fluorescence was recorded at 335/420 nm excitation/emission wavelengths after 5 seconds of plate shaking to mix the reaction components. Plate fluorescence was measured over the course of 30 minutes.

**Cell fractionation and western blotting.** Cell fractionation and Western blotting were followed according to a published protocol<sup>53</sup>. Briefly, the cultures of *A. oris* strains grown in BHI at 37 °C until  $\text{OD}_{600}$  of 0.25 were aliquoted, and bacterial aliquots were treated with 10  $\mu\text{M}$  of individual sortase inhibitors for 3 hours. Bacterial cultures were then normalized to equal  $\text{OD}_{600}$ , and cells were fractionated into culture medium (S), cell wall (W), and protoplast fractions. Isolated fractions were subjected to protein precipitation by 7.5% trichloroacetic acid, followed by washing with cold acetone, except for the protoplast fractions. Protein samples were dissolved in hot sodium dodecyl sulfate (SDS)-containing sample buffer, separated by 3–12% Tris-glycine gradient gels, and subjected to immunoblotting with specific antisera ( $\alpha$ -FimA, 1:10,000 dilution;  $\alpha$ -GspA,  $\alpha$ -MdbA,  $\alpha$ -SrtC2 and  $\alpha$ -SrtA, 1:4000 dilution), followed by chemo-luminescence detection.

**Transmission electron microscopy.** To observe cell morphology by negative staining, *A. oris* cells harvested after inhibitor treatment were washed once with PBS and suspended in 0.1 M NaCl. A drop of 7  $\mu\text{L}$  of bacterial suspension was placed onto carbon-coated nickel grids and stained with 1% uranyl acetate. Samples were examined using a JEOL JEM1200.

## Data availability

All data from this study are available from the corresponding author.

Received: 23 December 2019; Accepted: 20 April 2020;

Published online: 22 May 2020

## References

- Fischetti, V. A. Surface Proteins on Gram-Positive Bacteria. *Microbiol Spectr* 7(4), GPP3–0012 (2019).
- Navarre, W. W. & Schneewind, O. Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol Mol Biol Rev* 63(1), 174–229 (1999).
- Siegel, S. D., Reardon, M. E. & Ton-That, H. Anchoring of LPXTG-Like Proteins to the Gram-Positive Cell Wall Envelope. *Curr Top Microbiol Immunol* 404, 159–175 (2017).
- Jacobitz, A. W., Kattke, M. D., Wereszczynski, J. & Clubb, R. T. Sortase Transpeptidases: Structural Biology and Catalytic Mechanism. *Adv Protein Chem Struct Biol* 109, 223–264 (2017).
- Spirig, T., Weiner, E. M. & Clubb, R. T. Sortase enzymes in Gram-positive bacteria. *Mol Microbiol* 82(5), 1044–1059 (2011).
- Mazmanian, S. K., Liu, G., Ton-That, H. & Schneewind, O. *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science* 285(5428), 760–763 (1999).
- Lee, A. S. *et al.* Methicillin-resistant *Staphylococcus aureus*. *Nat Rev Dis Primers* 4, 18033 (2018).
- Klein, E. Y. *et al.* National Costs Associated With Methicillin-Susceptible and Methicillin-Resistant *Staphylococcus aureus* Hospitalizations in the United States, 2010–2014. *Clin Infect Dis* 68(1), 22–28 (2019).
- Maresso, A. W. & Schneewind, O. Sortase as a target of anti-infective therapy. *Pharmacol Rev* 60(1), 128–141 (2008).

10. Frankel, B. A., Kruger, R. G., Robinson, D. E., Kelleher, N. L. & McCafferty, D. G. *Staphylococcus aureus* sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry* **44**(33), 11188–11200 (2005).
11. Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O. & Clubb, R. T. Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proc Natl Acad Sci U S A* **98**(11), 6056–6061 (2001).
12. Huang, X. *et al.* Kinetic mechanism of *Staphylococcus aureus* sortase SrtA. *Biochemistry* **42**(38), 11307–11315 (2003).
13. Ton-That, H., Mazmanian, S. K., Faull, K. F. & Schneewind, O. Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Sortase catalyzed *in vitro* transpeptidation reaction using LPXTG peptide and NH(2)-Gly(3) substrates. *J Biol Chem* **275**(13), 9876–9881 (2000).
14. Comfort, D. & Clubb, R. T. A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria. *Infect Immun* **72**(5), 2710–2722 (2004).
15. Dramsi, S., Trieu-Cuot, P. & Bienne, H. Sorting sortases: a nomenclature proposal for the various sortases of Gram-positive bacteria. *Res Microbiol* **156**(3), 289–297 (2005).
16. Ton-That, H. & Schneewind, O. Assembly of pili in Gram-positive bacteria. *Trends Microbiol* **12**(5), 228–234 (2004).
17. Cascioferro, S. *et al.* Sortase A Inhibitors: Recent Advances and Future Perspectives. *J Med Chem* **58**(23), 9108–9123 (2015).
18. Suree, N., Jung, M. E. & Clubb, R. T. Recent advances towards new anti-infective agents that inhibit cell surface protein anchoring in *Staphylococcus aureus* and other gram-positive pathogens. *Mini Rev Med Chem* **7**(10), 991–1000 (2007).
19. Cascioferro, S., Totsika, M. & Schillaci, D. Sortase A: An ideal target for anti-virulence drug development. *Microb Pathog* **77C**, 105–112 (2014).
20. Oh, K. B. *et al.* Discovery of diarylacrylonitriles as a novel series of small molecule sortase A inhibitors. *J Med Chem* **47**(10), 2418–2421 (2004).
21. Suree, N. *et al.* Discovery and structure-activity relationship analysis of *Staphylococcus aureus* sortase A inhibitors. *Bioorg Med Chem* **17**(20), 7174–7185 (2009).
22. Jaudzems, K. *et al.* Targeting Bacterial Sortase A with Covalent Inhibitors: 27 New Starting Points for Structure-Based Hit-to-Lead Optimization. *ACS Infect Dis* **6**(2), 186–194 (2020).
23. Wehrli, P. M. *et al.* Discovery and development of substituted thiazoles as inhibitors of *Staphylococcus aureus* Sortase A. *Bioorg Med Chem* **27**(19), 115043 (2019).
24. Kruger, R. G., Barkallah, S., Frankel, B. A. & McCafferty, D. G. Inhibition of the *Staphylococcus aureus* sortase transpeptidase SrtA by phosphinic peptidomimetics. *Bioorg Med Chem* **12**(13), 3723–3729 (2004).
25. Jung, M. E. *et al.* Synthesis of (2R,3S) 3-amino-4-mercapto-2-butanol, a threonine analogue for covalent inhibition of sortases. *Bioorg Med Chem Lett* **15**(22), 5076–5079 (2005).
26. Rentero Rebollo, I. *et al.* Development of Potent and Selective *S. aureus* Sortase A Inhibitors Based on Peptide Macrocycles. *ACS Med Chem Lett* **7**(6), 606–611 (2016).
27. Wang, J. *et al.* Oligopeptide Targeting Sortase A as Potential Anti-infective Therapy for *Staphylococcus aureus*. *Front Microbiol* **9**, 245 (2018).
28. Oh, I. *et al.* In vitro sortase A inhibitory and antimicrobial activity of flavonoids isolated from the roots of *Sophora flavescens*. *Arch. Pharm Res* **34**(2), 217–222 (2011).
29. Won, T. H. *et al.* Brominated aromatic furanones and related esters from the ascidian *Synoicum* sp. *J Nat Prod* **75**(12), 2055–2061 (2012).
30. Won, T. H. *et al.* Beta-carboline alkaloids derived from the ascidian *Synoicum* sp. *Bioorg Med Chem* **20**(13), 4082–4087 (2012).
31. Jeon, J. E. *et al.* Discorhabdins from the Korean marine sponge *Sceptrella* sp. *J Nat Prod* **73**(2), 258–262 (2010).
32. Kim, S. H. *et al.* Inhibition of sortase, a bacterial surface protein anchoring transpeptidase, by beta-sitosterol-3-O-glucopyranoside from *Fritillaria verticillata*. *Biosci Biotechnol Biochem* **67**(11), 2477–2479 (2003).
33. Kim, S. W., Chang, I. M. & Oh, K. B. Inhibition of the bacterial surface protein anchoring transpeptidase sortase by medicinal plants. *Biosci Biotechnol Biochem* **66**(12), 2751–2754 (2002).
34. Jang, K. H. *et al.* Aaptamines as sortase A inhibitors from the tropical sponge *Aaptos aaptos*. *Bioorg Med Chem Lett* **17**(19), 5366–5369 (2007).
35. Bae, J. *et al.* Sesterterpenes from the tropical sponge *Coscinoderma* sp. *J Nat Prod* **74**(8), 1805–1811 (2011).
36. Park, B. S. *et al.* Curcuma longa L. constituents inhibit sortase A and *Staphylococcus aureus* cell adhesion to fibronectin. *J Agric Food Chem* **53**(23), 9005–9009 (2005).
37. Younis, S., Taj, S. & Rashid, S. Structural studies of *Staphylococcus aureus* Sortase inhibitor via *Conus* venom peptides. *Arch Biochem Biophys* **671**, 87–102 (2019).
38. Nitulescu, G. *et al.* Discovery of natural naphthoquinones as sortase A inhibitors and potential anti-infective solutions against *Staphylococcus aureus*. *Drug Dev Res* **80**(8), 1136–1145 (2019).
39. Chan, A. H. *et al.* Discovery of *Staphylococcus aureus* Sortase A Inhibitors Using Virtual Screening and the Relaxed Complex Scheme. *Chem Biol Drug Des* **82**(4), 418–428 (2013).
40. Chenna, B. C. *et al.* Identification of novel inhibitors of bacterial surface enzyme *Staphylococcus aureus* Sortase A. *Bioorg Med Chem Lett* **18**(1), 380–385 (2008).
41. Zhang, J.; *et al.* Anti-infective therapy with a small molecule inhibitor of *Staphylococcus aureus* sortase. *Proc Natl Acad Sci U S A* **111**(37), 13517–13522.
42. Nitulescu, G. *et al.* Molecular Docking and Screening Studies of New Natural Sortase A Inhibitors. *Int J Mol Sci* **18**(10), 2217 (2017).
43. Elgalai, I. & Foster, H. A. Comparison of adhesion of wound isolates of *Staphylococcus aureus* to immobilized proteins. *J Appl Microbiol* **94**(3), 413–420 (2003).
44. Nelson, J. W. *et al.* A biosynthetic strategy for re-engineering the *Staphylococcus aureus* cell wall with non-native small molecules. *ACS Chem Biol* **5**(12), 1147–1155 (2010).
45. Siegel, S. D. *et al.* Structure and Mechanism of LcpA, a Phosphotransferase That Mediates Glycosylation of a Gram-Positive Bacterial Cell Wall-Anchored Protein. *MBio* **10**(1), E01580–18 (2019).
46. Wu, C. *et al.* Lethality of sortase depletion in *Actinomyces oris* caused by excessive membrane accumulation of a surface glycoprotein. *Mol Microbiol* **94**(6), 1227–1241 (2014).
47. Zhang, J. H., Chung, T. D. & Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen* **4**(2), 67–73 (1999).
48. Gjorgjieva, M. *et al.* Discovery of Benzothiazole Scaffold-Based DNA Gyrase B Inhibitors. *J Med Chem* **59**(19), 8941–8954 (2016).
49. Chang, C. *et al.* Cell-to-cell interaction requires optimal positioning of a pilus tip adhesion modulated by gram-positive transpeptidase enzymes. *Proc Natl Acad Sci U S A* **116**(36), 18041–18049 (2019).
50. Mishra, A. *et al.* Two autonomous structural modules in the fimbrial shaft adhesin FimA mediate *Actinomyces* interactions with streptococci and host cells during oral biofilm development. *Mol Microbiol* **81**(5), 1205–1220 (2011).
51. Mishra, A. *et al.* The *Actinomyces oris* type 2 fimbrial shaft FimA mediates co-aggregation with oral streptococci, adherence to red blood cells and biofilm development. *Mol Microbiol* **77**(4), 841–854 (2010).
52. Chang, C. *et al.* In vitro reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking. *Proc Natl Acad Sci U S A* **115**(24), E5477–E5486 (2018).
53. Wu, C. *et al.* Structural determinants of *Actinomyces* sortase SrtC2 required for membrane localization and assembly of type 2 fimbriae for interbacterial coaggregation and oral biofilm formation. *J Bacteriol* **194**(10), 2531–2539 (2012).

### Acknowledgements

We thank our laboratory members for critical review of the manuscript and discussions. This work was supported by grants from the National Institutes of Health/National Institute of Allergy and Infectious Diseases Grants R01 AI121360 and AI052217 (R.T.C) and National Institute of Dental & Craniofacial Research grant DE017382 (H.T.-T). C.K.S. was supported by a Cellular and Molecular Biology Training Grant (Ruth L. Kirschstein National Research Service Award GM007185). S.D.S was supported by NIDCR under Award F31-DE027295 and the Kopchick fellowship from MD Anderson University of Texas Health, Graduate School of Biomedical Sciences.

### Author contributions

J.E.G., M.D.K., H.T.-T. and R.T.C. conceptualized the study. M.D.K. did the preliminary work. J.E.G. and C.K.S. conducted the screen and analysis and followed up with the hit validation assay. C.Y.C., C.G.W. and S.D.S. did the western blotting and cell fractionation studies. S.W.Y. and M.E.J. helped to evaluate each of the hits. R.D. provided the facility for the screen and guidance for implementation. J.E.G., H.T.-T. and R.T.C. jointly wrote the article.

### Competing interests

The authors declare no competing interests.


### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-65256-x>.

**Correspondence** and requests for materials should be addressed to H.T.-T. or R.T.C.

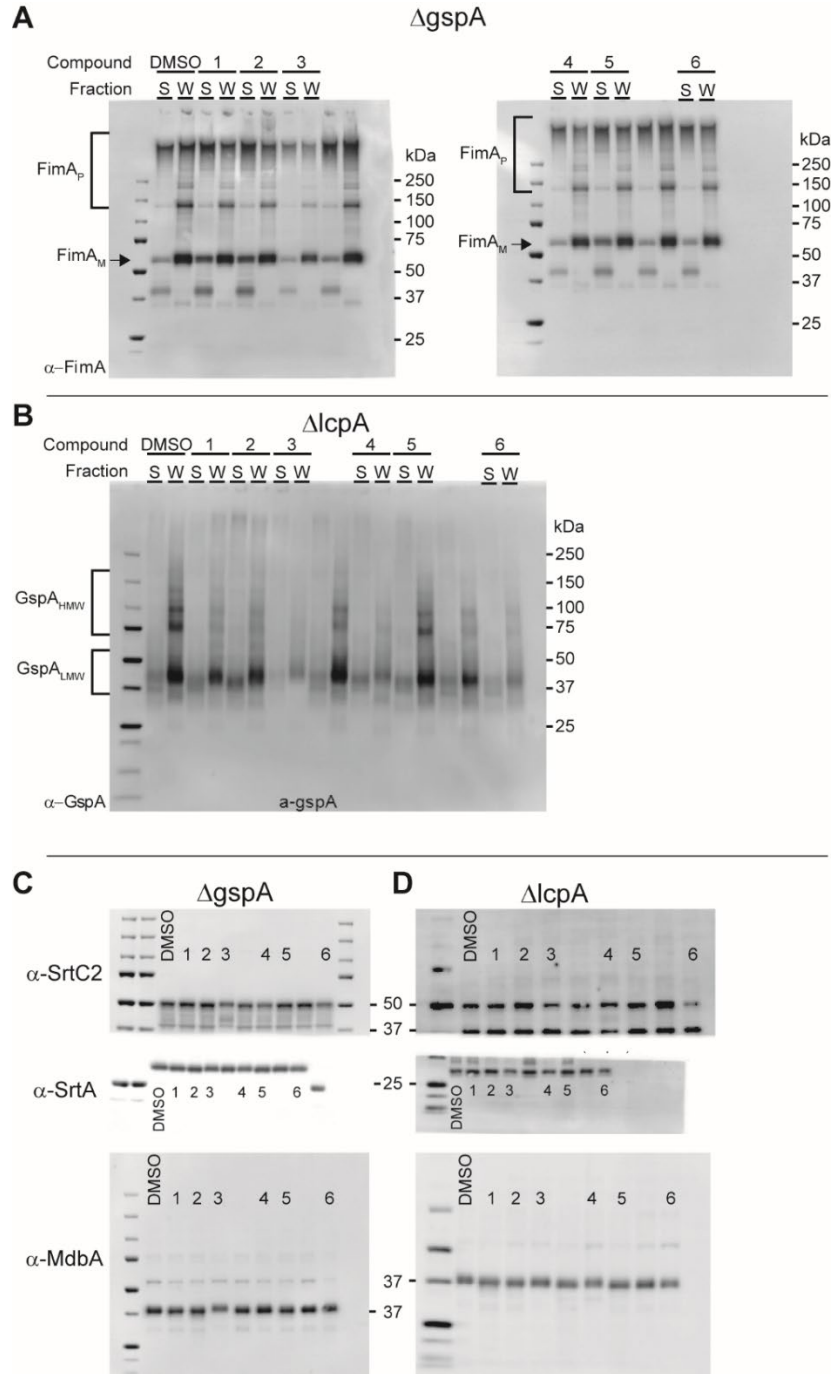
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

### 5.3.2 Supplementary Information



**Figure 5.3.S1 Full Length Gels for Figure 4.** Presented are the original, uncropped images used to generate Figure 4 presented in the main text.