Title

Structural and Mechanistic Insights into Pilus Assembly via Isopeptide-Bond Forming Sortase Enzymes and Translational Applications to Bioconjugation Technologies

Author

McConnell, Scott A

Publication Date

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Structural and Mechanistic Insights into Pilus Assembly via Isopeptide-Bond Forming Sortase Enzymes and Translational Applications to Bioconjugation Technologies

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Biochemistry and Molecular Biology

by

Scott McConnell

2020

ABSTRACT OF THE DISSERTATION

Structural and Mechanistic Insights into Pilus Assembly via Isopeptide-Bond Forming Sortase Enzymes and Translational Applications to Bioconjugation Technologies

by

Scott Andrew McConnell

Doctor of Philosophy in Biochemistry and Molecular Biology

University of California, Los Angeles, 2020

Professor Robert Thompson Clubb, Chair

Pathogenic Gram-positive bacteria cause a range of serious infections in humans and represent a significant threat to global health. The rising emergence of virulent strains which are resistant to our current arsenal of antibiotics escalates these dangers. Many Gram-positive bacteria display an array of proteins on their cell surface that enable them to interact with their environment during infections. Among the most important extracellular virulence factors are bacterial pili, which are adhesive filaments that are constructed by specialized cysteine transpeptidases through isopeptide linkages. This dissertation describes my efforts to elucidate the assembly mechanism of the archetypal SpaA-pilus from *Corynebacterium diphtheriae*.

Further, it describes parallel experiments aimed at repurposing the enzymes used in pilus biogenesis for protein engineering applications.

This thesis is divided into two major sections: Pilus Biology (Chapters 2 - 4) and Sortase Bioconjugation (Chapters 5 - 7). The Pilus Biology section describes efforts to elucidate the mechanism of Gram-positive pilus biogenesis using kinetic, structural and cellular experiments. Chapter 2 describes the biochemical reconstitution and characterization of the assembly reaction that builds the *C. diphtheriae* SpaA-pilus. SpaA-type pili are assembled by a Class C pilin polymerase sortase, ^{Cd}SrtA. Examination of the structure of this enzyme revealed that ^{Cd}SrtA is held in an inactive state *in vitro* by an autoinhibitory "lid" structure. We discovered that amino acid substitutions introduced into the "lid" activate the enzyme and permit biochemical characterization of the polymerization reaction. Chapter 3 characterizes the mechanism and *in vitro* kinetics of the lysine-isopeptide transpeptidation reaction that builds the SpaA pilus. We identify the rate limiting step in the mechanism and offer a kinetic explanation that explains why "lid" alterations activate ^{Cd}SrtA. Chapter 4 describes the solution structure of the pilin-pilin linkage which joins successive protomers in the SpaA pilus. Data from NMR dynamics, SAXS and biophysical measurements of protein stability reveal the mechanism by which Gram-positive pili are stabilized at each linkage site throughout the elongated pilus fiber.

The Bioconjugation section of this thesis describes our efforts to develop sortase enzymes into useful bioconjugation tools. Chapter 5 describes a versatile sortase-mediated protein nanocage ligation platform that facilitates enzymatic synergy by enhancing pathway flux between enzymes with complementary activities. As a proof of principle, the nanocage scaffold was functionalized with cellulolytic enzymes, demonstrating a marked enhancement in degradative synergy of cellulose substrate compared to unbound cellulases. Chapter 6 builds upon our understanding of the pilus biogenesis reaction to engineer the ^{Cd}SrtA sortase into a viable bioconjugation tool that can be used to attach peptide fluorophores to proteins via isopeptide bonds. Chapter 7 describes ongoing efforts to further optimize the ^{Cd}SrtA

bioconjugation tool using a directed evolution approach, with the goal of altering its substrate specificity and increasing its thermostability. Together, the research described in this thesis provides new insight into the biogenesis mechanism that is used by Gram-positive bacteria to produce adhesive pili and describes the development of a promising molecular tool for producing novel bioconjugates.

The dissertation of Scott McConnell is approved.

Todd O. Yeates

Joseph A. Loo

Megan M. McEvoy

Robert T. Clubb, Committee Chair

University of California, Los Angeles

2020

To my mother, who encouraged my scientific passions from an early age; my father, for his invaluable counsel and life advice; my two brothers, for their unconditional friendship; and my girlfriend, who I always turn to for affection and inspiration.

"I would rather have questions that can't be answered than answers that can't be questioned."

Richard Feynman

Table of Contents

List of Figures

Chapter 1: Introduction: The Roles of Sortases in Surface Protein Display and Bioconjugation

Chapter 2: *In vitro* reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking

List of Tables

Acknowledgements

I would like to thank my advisor, Professor Robert T. Clubb, for his incredible mentorship, encouragement and innovative ideas which contributed greatly to my development as a scientist. Additionally, I would like to thank my collaborators, Professors Joseph Loo, Hung Ton-That, Todd Yeates, and Jeff Wereszczynski and members of their lab for providing advice, resources, direction, and data for many of my projects. I thank members of my dissertation committee, Professors Todd Yeates, Joseph Loo, and Megan McEvoy for providing insightful feedback, encouragement and advocacy throughout my graduate career, as well as reviewing this dissertation. I would also like to thank Dr. Robert Peterson for sharing his incredible knowledge about biomolecular NMR experimentation and theory. Special thanks to Alex Lisker and Duilio Cascio for their valuable advice with computation and IT assistance. I would like to thank the members of the Clubb lab for their experimental expertise, brainstorming and general comradery over the years including, Dr. Brendan Mahoney, Dr. Ken Ellis-Guardiola, Orlando Martinez, Chris Sue, Dr. Brendan Amer, Dr. Grace Huang, Dr. Ramsay Macdonald, Jason Gosschalk, Dr. Alex Jacobitz, Dr. Megan Sjodt and Dr. Michele Kattke. Special thanks to my undergraduate research assistant Rachel McAllister for her years of dedication. Most importantly, I would like to thank my family and friends for this support and understanding. Finally, I would like to thank the Cellular and Molecular Biology Training Grant, the Philip Whitcome, Audree Fowler, and Dissertation Year Fellowships for their financial support.

Chapter Two of this dissertation is reformatted from a version of published manuscript: *In vitro* reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking. Chungyu Chang*, Brendan R. Amer*, Jerzy Osipiuk, Scott A. McConnell, I-Hsiu Huang, Van Hsieh, Janine Fu, Hong H. Nguyen, John Muroski, Erika Flores, Rachel R. Ogorzalek Loo, Joseph A. Loo, John A. Putkey, Andrzej Joachimiak, Asis Das, Robert T. Clubb, Hung Ton-That. Proceedings of the National Academy of Sciences. June

2018. Reproduced with permission from the Proceedings of the National Academy of Sciences. Asterisks denotes these authors contributed equally.

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., Alverez, 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 assistance.

Chapter Four of this dissertation is a version of a manuscript in preparation: Sortase-assembled pili in *Corynebacterium diphtheriae* are built using 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. (2020). We express appreciation to the Ton-That laboratory for valuable collaborative studies which contributed to this manuscript.

Chapter Five of this dissertation is a version of a published manuscript: Designed Protein Cages as Scaffolds for Building Multienzyme Materials. Scott A. McConnell*, Kevin A. Cannon*, Christian Morgan, Rachel McAllister, Brendan R. Amer, Robert T. Clubb, and Todd O. Yeates. ACS Synthetic Biology. January 2020. Reproduced with permission from the Proceedings of ACS Synthetic Biology. Asterisks denotes these authors contributed equally. We express appreciation to the Yeates laboratory for valuable collaborative studies which contributed to this manuscript.

Chapter Six of this dissertation is a version of a published manuscript: Protein Labeling via a Specific Lysine-Isopeptide Bond using the Pilin Polymerizing Sortase from *Corynebacterium diptheriae*. Scott A. McConnell, Brendan R. Amer, John Muroski, Janine Fu,

Chungyu Chang, Rachel R. Ogorzalek Loo, Joseph A. Loo, Jerzy Osipiuk, Hung Ton-That, Robert T. Clubb. Journal of the American Chemical Society. June 2018. Reproduced with permission from the Journal of the American Chemical Society. We express appreciation to the Loo laboratory for valuable collaborative studies which contributed to this manuscript.

Chapter 7 describes ongoing research and optimization for directed evolution approaches designed to identify improved variants of the ^{Cd}SrtA pilin polymerase.

Vita

Scott Andrew McConnell received primary and secondary education in Orinda, California, completing high school with honors in 2011. Scott enrolled at the University of California, Los Angeles in the fall of 2011. In 2015, he earned a Bachelor of Sciences degree in Biochemistry in June 2015 and received the HyperCube award in recognition of outstanding undergraduate research. In September 2015, Scott elected to continue his graduate studies at UCLA in the Biochemistry, Molecular and Structural Biology Program. During his graduate studies, Scott was supported by several funding sources, including the Dissertation Year Fellowship, the Philip Whitcome Fellowship, the Audree Fowler Fellowship, and the Cellular and Molecular Biology Training Grant. Scott is an author on six papers and several posters presented at conferences across the country.

Publications resulting from this Dissertation Research

Sortase-assembled pili in *Corynebacterium diphtheriae* are built using 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. (In prep.)

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.

Designed Protein Cages as Scaffolds for Building Multienzyme Materials. ACS Synthetic Biology. Scott A. McConnell*, Kevin A. Cannon, Christian Morgan, Rachel A. McAllister, Brendan R. Amer, Robert T. Clubb, and Todd O. Yeates. ACS Synthetic Biology. January 2020

Protein Labeling via a Specific Lysine-Isopeptide Bond using the Pilin Polymerizing Sortase from *Corynebacterium diptheriae*. Scott A. McConnell, Brendan R. Amer, John Muroski, Janine Fu, Chungyu Chang, Rachel R. Ogorzalek Loo, Joseph A. Loo, Jerzy Osipiuk, Hung Ton-That, Robert T. Clubb. Journal of the American Chemical Society. June 2018

In vitro reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking. Chungyu Chang, Brendan R. Amer, Jerzy Osipiuk, Scott A. McConnell, I-Hsiu Huang, Van Hsieh, Janine Fu, Hong H. Nguyen, John Muroski, Erika Flores, Rachel R. Ogorzalek Loo, Joseph A. Loo, John A. Putkey, Andrzej Joachimiak, Asis Das, Robert T. Clubb, Hung Ton-That. Proceedings of the National Academy of Sciences. June 2018

Chapter 1

Introduction: The Roles of Sortases in Surface Protein Display and Bioconjugation

1.1 Overview

Throughout much of human history, bacterial infections were serious medical events which resulted in rampant infectious diseases such as smallpox, cholera, diphtheria and pneumonia[1]. Beginning with the discovery of penicillin in 1928, a revolution in antibiotic discovery during the mid-twentieth century yielded a vast arsenal of broad-spectrum drugs capable of efficiently controlling bacterial infections[2,3]. This paradigm shift enabled a transformation of medicine: risky procedures such as invasive surgery and the deliberate attenuation of the immune system following organ transplants or as a result of chemotherapies became routine[2]. However, it was soon discovered that antibiotic resistance arises rapidly after widespread deployment of bactericidal drugs, and to date, strains resistant to most common antibiotics have been identified[4]. Indeed, the CDC reported in 2019 at least 2.8 million serious infections by bacterial strains resistant to our current antibiotics, resulting in 35,000 deaths in the United States[5]. Whereas first generation antibiotics were developed from the chemical scaffolds of natural metabolites which inhibit pathways that are critical for bacterial viability, modern understanding of molecular mechanisms of bacterial virulence afford us the opportunity to develop more carefully targeted drugs which will not precipitate resistant strains.

Bacteriology is broadly divided into Gram-positive and Gram-negative classifications. Gram-positive bacteria are thusly named because their thick cell wall retains crystal violet stain after washing in the Gram stain procedure[6]. Many of the most important bacterial pathogens are Gram-positive, such as *Staphyloccocus*, *Streptococcus*, *Clostridia* and *Corynebacteria* spp[7]. The cell wall of Gram-positive bacteria is a complex protective layer consisting of crosslinked peptidoglycan and wall techoic acid polymers which provides a scaffold for display of surface proteins with specific functions[8]. Many proteins adorning the cell wall are virulence factors which interact with the extracellular environment. Due to the thickness of the cell wall, simply targeting extracellular proteins to the membrane is generally insufficient for display. Thus, Gram-positive

bacteria evolved sortase-dependent pathways to elaborate their surface with virulence factors. Sortases are cysteine transpeptidases which affix proteins directly to the peptidoglycan cell wall in a “sorting reaction”. Typically, sortase substrates are monomeric proteins with roles ranging from immune system modulation, to nutrient acquisition, to spore formation[9]. In contrast, one distinct sortase class, Class C sortases, polymerize linear pilus filaments which are responsible for bacterial adhesion to host cells and tissues. Pilin polymerizing sortases are attractive drug targets due to their importance in bacterial adhesion and biofilm formation. Additionally, they have intriguing potential as bioconjugation reagents due to their unique specificity and bond forming mechanism.

1.2 The Sortase Transpeptidase Enzyme Family

One distinguishing feature of Gram-positive bacteria is a thick peptidoglycan cell wall that encircles the entire cell. The cell wall is composed of a peptidoglycan matrix consisting of repeated N-acetyl glucosamine (NAG) and N-acetylmuramic acid (NAM) subunits which are crosslinked via their peptide stems into glycan strands[10]. The cell wall maintains cell shape, prevents osmotic lysis and forms a scaffold for the attachment of extracellular proteins and secondary cell wall polymers[8]. Whereas Gram-negative bacteria have several secretion mechanisms mediated by membrane proteins across both bilayers[11], the thickness of the Gram-positive cell wall necessitates a distinct sortase-dependent pathway for the display of surface proteins. Sortases are nearly ubiquitous in Gram-positive bacteria and are also occasionally observed in Gram-negatives (e.g. *Shewanella putrefaciens*) and even Archaea (e.g. *Methanobacterium thermoautotrophicum*)[12].

Members of the sortase superfamily are binned into six different classes (A to F) based on their primary sequence similarity[13–16]. The genomes of most Gram-positive bacteria encode at least one sortase gene, while others contain several sortase homologs that attach different proteins to the cell wall or assemble pili. Sortase-mediated surface display regulates many important functions including sporulation, iron acquisition, pilus assembly and generalized cell wall housekeeping **(Figure 1.1)**[17]. Each sortase class recognizes a characteristic substrate type, with the exception of Class A, which exhibits a more promiscuous substrate profile and is capable of sorting myriad protein substrates to the cell wall. The most industrious housekeeping sortase gene is found in *Listeria monocytogenes*, which is responsible for mounting 43 separate surface proteins[16,18]. Due to their generalized function, Class A sortases are designated as “housekeeping” enzymes and are present in all annotated Gram-positive genomes, with the exception of Mycobacterium and Microplasma[13,14,19]. Housekeeping sortases are critical for bacterial virulence, as genetic knockouts yield bacteria with significantly attenuated virulence in

mouse models[20–24]. Sortase depletion also sometimes results in the accumulation of substrates in the membrane, which may sufficiently alter cellular physiology and promote more efficient immune clearance[25]. In *Actinomyces oris*, sortase depletion directly results cellular toxicity due to the toxic accumulation of glycosylated AcaC/GspA substrates in the cell membrane[26]. However, as a general rule, elimination of sortase genes does not seem to affect bacterial viability. This suggests that sortase inhibitors could serve as particularly efficient anti-infective agents that limit microbial infectivity, but otherwise do not affect bacterial growth outside the host. Thus, targeted sortase inhibition would generate minimal selective pressure for resistant strains to develop, which is a major concern with currently available bactericidal antibiotics[27]. Additionally, there are no sortase homologs in eukaryotes, mitigating the possibility of cross-reactivity of sortase inhibitors. Taken together, sortase inhibition represents an exciting new avenue for antibiotic development in an era of heightened need for novel anti-infective agents[28].

The Class A and C sortase types will be discussed in depth here. Class A sortases are expressed constitutively and mediate display of a wide range of substrates. All other sortase classes are found in distinct operon clusters along with their cognate substrates and conditionally expressed only in specific environmental conditions. Specifically, Class C sortases are harbored in dedicated genomic islands along with their pilin substrates: Typically one primary substrate which comprises the backbone of the pilus fiber, and one or two accessory pilins which are found at the pilus base to anchor the fiber to the cell wall, at the pilus tip to adhere to specific host tissues, or elaborated throughout the shaft as additional adhesins[29]. The class A sortase from *Streptococcus aureus* (^{Sa}SrtA) was the first discovered and best understood sortase from biochemical, structural and mechanistic standpoints[30,31]. Thus, ^{Sa}SrtA is the canonical founding member of the sortase family and forms the basis for our current understanding of sortase biology.

Sortase substrates are first secreted through the Sec translocon and membrane-tethered by virtue of a signal peptide and cell wall sorting signal peptide, respectively. Sortase substrates are targeted for secretion by a signal peptide consisting of 15 to 20 hydrophobic residues encoded at the N-terminus of the polypeptide[32]. Binding of the signal by a secretion chaperone maintains the precursor substrate in an unfolded state and facilitates transport of the complex to the Sec secretion machinery, where it is translocated across the membrane[11]. After the substrate is transported to the extracellular side of the membrane, signal peptidases cleave the signal peptide from the precursor protein and the protein folds into its mature conformation, which sometimes requires additional chaperones to aid in folding[33]. In the mature state, the substrate is tethered to the membrane at its C-terminus, which exposes a conserved LPXTG sorting signal motif for sortase recognition. Sortases and their substrates are known to accumulate in foci located close to division septa, where the cell wall is relatively thin and immature such that the energetic barrier to secretion and processing is the lowest[34]. Sortases are anchored to the bilayer by transmembrane helices and operate in the Gram-positive "periplasmic space" between the cell wall and membrane. Cryo-transmission electron microscopy of frozen hydrated slices of bacterial cells reveal the presence of a ~20 nm low-density zone surrounded by a thicker (20-80 nm) high density zone, representing the periplasmic space and cell wall, respectively[11,35,36]. This low-density space provides an optimal environment for sortases to process their substrates and avoid steric clash with the dense cell wall matrix, while remaining in close proximity to the peptidoglycan for subsequent anchoring.

A general two-step mechanism is well conserved across all sortase classes. An initial acylation step, where a 5-residue motif within the cell wall sorting signal (CWSS) at the substrate C-terminus is recognized and cleaved, is followed by a transpeptidation step where the sorting signal is transferred to a nucleophilic substrate, resulting in a new covalent linkage. In the first step of catalysis, sortase performs nucleophilic attack with its catalytic Cys residue on

the peptide bond between threonine and glycine residues within the CWSS. This results in the formation of a thioacyl intermediate between sortase its substrate **(Figure 1.2)**. With the exception of Class C sortases, the sortase-substrate intermediate complex is then resolved by another nucleophilic attack on the acyl linkage by a second substrate, the amino-terminus of a peptide crossbridge on Lipid II cell wall precursor[31]. This final transpeptidation step results in covalent transfer of the substrate from its initial membrane anchor to Lipid II which is then incorporated into the mature cell wall by the transglycosylation and transpeptidation processes of peptidoglycan biosynthesis. Thus, the sortase substrate is eventually fully incorporated into the mature cell wall matrix for display on the cell surface[11].

Detailed molecular models of substrate binding in this first intermediate are based on structural comparisons of the apo-enzyme and acyl enzyme-substrate complexes[37,38]. The globular core of ^{Sa}SrtA is an eight-stranded β-barrel fold, which is observed in many subsequent structures of homologous sortase enzymes[39–46] **(Figure 1.3)**. In the ^{Sa}SrtA -CWSS peptide complex structure, the LPXT sorting signal peptide is docked into the binding cleft, which is defined by a floor consisting of residues in the β4 and β7 strands, and four loops which form the "walls" of the cleft (β6/β7 loop, β7/β8 loop, β3/β4 loop and β2/α1 loop). The sorting signal ligand is arranged in an L-shaped pose in this binding pocket, with a 90° kink at the AP peptide bond. The β6/β7 loop is unstructured in the apo enzyme, but adopts a 3_{10} helix structure when the substrate is bound, shifted 10 Å into a "closed" conformation in an induced fit mechanism which allows extensive contacts with sorting signal residues upon binding[37]. Additionally, signal peptide binding wedges the β7/β8 loop away from the H1 helix as a rigid unit by 13 Å, resulting in the formation of a new groove with the catalytic His residue at the center, which is predicted to accommodate the second substrate[37].

Although not a general feature in all sortases, biochemical experiments have demonstrated that the canonical ^{Sa}SrtA enzyme has a strong calcium dependence. In reaction

conditions devoid of calcium ions, ^{Sa}SrtA exhibits a 5-fold decrease in activity[39,47]. A structural explanation for this phenomenon is readily apparent: A single calcium ion is coordinated by three glutamate residues, which stabilizes the β6/β7 loop into the closed conformation that is required for effective CWSS binding[37]. In homologous sortase enzymes, the binding pocket is sometimes "pre-formed" by alternative stabilizing bonds, negating the dependence on a divalent cation in these enzymes[44,48].

The sortase active site consists of catalytic triad of Arg197, Cys184, His120 (^{Sa}SrtA numbering) residues, which are located on adjacent β-strands **(Figure 1.3)**. There is some controversy over the exact role of the active site residues in catalysis. The catalytic cysteine performs a nucleophilic attack on the sorting signal, as described previously[27]. In order for this reaction to be chemically favorable, the active site cysteine and histidine residues must exist as a thiolate/imidazolium pair, which would be populated at 0.06% at physiological pH[49]. The low population of catalytically competent enzymes may explain the differences in reactivity observed *in vitro* compared to *in vivo* reaction rates. For this reason, a proposed role for the invariant arginine is to stabilize the deprotonated thiolate form of cysteine. However, it has also been proposed to have a role in stabilization of the oxyanion tetrahedral intermediate and positioning of the sorting signal by direct interaction with backbone residues[40]. Meanwhile, the invariant histidine is proposed to perform dual functions: as a general acid to protonate the amide leaving group after the scissile bond of the sorting signal is cleaved in the first step of the reaction, and then, once deprotonated, as a general base to activate the incoming terminal amine substrate for nucleophilic attack[49,50].

Housekeeping Class A sortases are found in nearly all Gram-positive bacteria from the Firmicutes phylum, but some species also encode additional classes of sortases with specialized functions. These sortases often recognize divergent sorting signal motifs, offering a possible explanation for substrate discrimination between classes[17]. Other than sorting signal

specificity, all homologs operate in a manner that is mechanistically similar to the canonical *S. aureus* Class A sortase. The Class C pilin polymerizing sortases are distinguished from all other classes by their unique nucleophile selectivity and reaction product. Instead of catalyzing single transpeptidation reactions where the sorting signal peptide bonds are cleaved and transferred to the cell wall via backbone peptide bond, pilin polymerases catalyze linkages between sorting signal peptides on one substrate to a lysine sidechain on a second substrate. They repeat this reaction to form linear polymers of proteins that are called pili. This enzyme class is discussed in the following section.

1.3 Pilus Biology

Pili (or fimbrae) are long, hair-like appendages which extend from bacterial cell surfaces to mediate two major types of interactions with the extracellular milieu. First, they mediate important homophilic interactions with other microbes which result in biofilm formation[51–53]. Second, pili form specific and strong attachments to host cells in order to overcome the general problem of the net repulsive effect caused by negative charges between bacterial and host cells[54,55]. As such, these filaments are important virulence factors for pathogenic bacteria.

Surface pili presented by Gram-negative and Gram-positive bacteria are structurally and functionally distinct. Gram-negative pili are relatively thick structures (5-12 nm in diameter and up to 20 μm in length) whose cross-sections consist of 3-6 protein subunits that are associated with one another by noncovalent forces[56]. Gram-negative pili have additional capabilities enabled by adaptive ATPase-driven retraction and elongation, including evasion of host immune response, transfer of genetic material and twitching motility[55,57,58]. The Gram-positive pilus polymer (1-5 nm x 3-5 μm) consists of repeating subunits that are covalently linked into a linear chain by intermolecular isopeptide bonds, which are installed by specialized Class C pilin polymerizing sortases[59]. The rest of this section describes the Gram-positive pilus.

During infections by pathogenic bacteria, pili promote adhesion to host cells and provoke strong immune responses. Nonpiliated mutant strains are significantly attenuated in virulence in mouse infection models and result in much lower levels of tumor necrosis factor- α and interleukin-6 (used as proxies for host inflammatory response) as compared to wild-type, piliated strains[60]. In another study, pregnant mice were immunized with hundreds of recombinant bacterial proteins from group B Streptococcus and their progeny were challenged with a lethal dose of wild-type streptococci. Three quarters of the antigens which conferred protection against this challenge are pilus proteins[61]. Due to their high immunogenicity, extracellular

presentation, repetitive structure, and high levels of expression during the early stages of infections, pili are considered excellent vaccine targets[62–64].

Pili are also important factors in biofilm formation. Biofilms are diverse bacterial communities encased in a protective matrix consisting of polysaccharides, proteins, membrane vesicles and DNA, which provide a barrier to resist antibiotic treatment and immune clearance[53,65]. Biofilm production is a unique bacterial state which involves differential expression of 30% of the proteome, resulting in *de novo* expression of around 200 dedicated proteins in *S. pnuemoniae*[66]. One of the most important factors expressed during biofilm production are pili. Specifically, the adhesive tip components mediate homophilic interactions between bacteria which promote the formation of three-dimensional microcolonies. Additionally, some co-aggregation proteins may "hijack" pilus machinery, replacing tip pilins at distal end of pilus fibers. In the case of *A. oris*, the coaggregation factor, CafA, can displace the tip pilin in Type 2 pili in order to mediate coaggregation with other oral bacteria[67]. Strong evidence for specific interactions by tip pilins comes from studies of group A Streptococcus, where it was demonstrated that tip pilin knockout strains are less virulent in mice and competition with exogenous recombinant tip protein reduces bacterial aggregation[68]. Additionally, several tip pilin proteins were identified in broad transposon screens of biofilm formers, highlighting the importance of pilus-mediated interactions with other bacteria in the initial phases of biofilm development[65]. From a biophysical perspective, single cell force microscopy studies found that piliated *Lactococcus lactis* adhere to other bacterial cells with forces roughly twice that of pilus-devoid strains[69].

The second function of pili is host cell adherence, which occurs via covalent or non-covalent mechanisms. The first class of noncovalent pilin interactions with host cells involve adhesive proteins with affinity for common elements of the eukaryotic extracellular matrix. Some tip pilin domains exhibit significant homology to the van Willebrand factor A fold (VWA), suggesting that they interact with extracellular components such as collagen, fibronectin and

laminin to mediate heterophilic adherence[70,71]. VWA domains do not appear to be involved in homotypic interactions with other bacteria, as they are dispensable for biofilm formation[72]. Another class of pilus adhesion involves tip pilins which exhibit very strong attachments through noncovalent catch bond mechanisms. Through high affinity “dock, lock, latch” mechanisms, this class of adhesins immobilize their ligands under high forces[73–75]. Unlike typical noncovalent slip bonds, which become weaker or have shorter lifetimes when subjected to increasing mechanical forces, receptor-pilin catch bonds are strengthened as increasing shearing forces are applied through mechanisms governed by mechanical forces[76].

The discovery of thioester domains (TEDs) in several tip pilins suggests a distinct covalent mechanism for pilus-mediated adhesion[77–80]. TED domains harbor intramolecular thioester bonds between Cys and Gln residues in a thiolactone ring composed of a Cys-Gly-Glu-Gln motif and appear to form autocatalytically in the correct structural environment[81]. These linkages were originally predicted to impart structural stability of the domain akin to the ubiquitous isopeptides (described in Section 1.4). However, a structural role is unlikely because these linkages are not positioned at domain boundaries (as are stabilizing isopeptide bonds) and further, their surface exposure renders them susceptible to nucleophilic attack. Instead, pilin thioester domain are involved in an adhesion mechanism akin to the human complement system, which employs highly reactive thioester bonds to bind hydroxyl groups of carbohydrates on bacterial cell wall as part of the innate immune response[81]. Once activated, complement proteins either react rapidly with bacterial proteins or are neutralized by hydrolysis to limit damaging side reactions to human proteins. In a similar manner, tip pilin TEDs likely undergo conformational rearrangements to expose the reactive acyl moiety of the thioester upon docking to specific host factors, forming covalent linkages which mediate strong attachments. One such example is the Cpa adhesin from *S. pyogenes*. This adhesin reacts via a thioester linkage with extracellular matrix components such as collagen to adhere to host tissues. In low force conditions, the thioester bond can be

reversed by reactive amines, such as histamine, which are common at inflammation sites. The intramolecular thioester can then reform at different attachment sites, enabling nomadic bacterial rolling. However, when large forces are applied the domain is extended such that the bond can no longer be reversed and the thioester is locked and able to survive nanoNewton perturbations. This mechanism allows the microbe to explore different environments and facilitates cell migration which is important for colonization under low force conditions, while also maintaining robust adhesion during high force. Thus, these “smart bonds” exhibit mechanical allostery which permits both nomadic and locked phases during adhesion[82].

The energetic cost of pilus biogenesis combined with heightened immunogenicity necessitate careful regulation of pilin expression and display. The idea that pilus expression is bistable is supported by multiple studies of piliated Gram-positive bacteria indicating that less than half of cells express pili[60,83,84]. In the *S. pneumoniae* PI-1 pilus assembly system, this regulation occurs at the transcriptional level, and notably, is not regulated by any additional genes outside of the pilus island gene cluster[84]. Additionally, regulation is highly adaptable to environmental cues. For example, *S. pyogenes* ramps up pilus production as temperatures approach human skin temperature[83]. As of yet, the mechanistic triggers for pilus upregulation are unknown. Evolutionarily, it may be advantageous to maintain two subsets of the population: one population that is well suited to colonize host tissues, and another that can easily escape host immune defenses that are aimed at pili.

1.4 Pilus Biogenesis by Class C Sortases

Class C pilin polymerizing sortases are found across Firmicutes and Actinobacteria[17]. Much of our current understanding of pilus biogenesis in Gram-positive bacteria is derived from the archetypal system found in *Corynebacterium diphtheriae.* Pathogenic Corynebacterium are causal factors in several important diseases in humans including diphtheria, endocarditis and urinary tract infections[85]. Analysis of its genome reveals 17 putative sortase substrates genes (containing LPxTG motifs), nine of which are pilin proteins[18]. *C. diphtheriae* encodes three separate pilus systems: SpaABC, SpaDEF and SpaHIG. SpaA-type pili preferentially adhere to host pharyngeal cells, while SpaD- and SpaH-type pili adhere to laryngeal and lung epithelial tissues, demonstrating the role of pili in dictating the tissue tropism of a given microbe[51,86]. The organization of each of these sets of genes into operons under the control of a single protomer ensures that the expression of all pilin precursors and cognate pilin polymerases are expressed cotemporally, presumably in response to certain environmental cues.

Extensive studies of the SpaABC pilus have made it paradigmatic. It consists of three subunits: the major pilin, SpaA (the shaft pilin), and two ancillary pilins: SpaC (the tip pilin) and SpaB (the basal pilin). As with other sortase substrates, pilin subunits are expressed in the cytoplasm and subsequently targeted to the Sec translocon by an N-terminal signal peptide. The precursor polypeptides are then partially secreted, but retained in the membrane by a C-terminal cell wall sorting signal (CWSS). The CWSS is a conserved tripartite motif consisting of a five residue sortase recognition site (LPLTG), flanked by a hydrophobic transmembrane region and a stretch of basic residues, which anchor the protein in the membrane. Pilin sortases and their substrates are known to cluster into distinct pilus assembly centers (or “pilusosomes”) on the cell membrane[34,87]. This organization results in a high local concentration of ^{Cd}SrtA-SpaA acyl intermediates at the assembly foci which accelerates polymerization.

SpaA pilus biogenesis is a biphasic process: polymerization and pilus elongation are carried out by the pilus-specific sortase, followed by transfer of the completed pilus to a housekeeping sortase, which then anchors the entire fiber to the cell wall for display. SpaC is the first pilin subunit to be incorporated. Located at the tip of each pilus, SpaC mediates interactions between microbe and host pharyngeal epithelial cells via a predicted VWA domain[85,86]. SpaC is ligated to a backbone SpaA pilin, which is in turn ligated to several hundred additional SpaA pilins during the polymerization reaction. Pilus length appears to be primarily correlated with the availability of the major pilin component which comprises the shaft, as overproduction of this component results in increased pilus length relative to wild-type strains[94]. However, the activity of housekeeping sortases may also be an important factor in modulating pilus length and spatial positioning[95]. Incorporation of SpaB into the polymerization machinery flips a "molecular switch" which terminates polymerization by handing the pilus off to ^{Cd}SrtF, a Class E housekeeping sortase[15,96]. *In vivo* studies have shown that *C. diphtheriae* strains with SpaB knocked out produce pilin fibers much longer than wild-type strains and release elongated pili to the culture medium, while overproduction of SpaB results in shortened pili[96]. This data suggests that SpaB acts as a signal to promote the transition from polymerization to the cell wall anchoring phase. ^{Cd}SrtF catalyzes the covalent attachment of the LAFTG sorting signal of SpaB to a cell wall precursor, Lipid II. The pilus is then embedded into the mature cell wall envelope when Lipid II is incorporated into the crosslinked peptidoglycan matrix.

The molecular details of the polymerization reaction have been well-studied. ^{Cd}SrtA recognizes the ^{491}LPLTG^{495} sorting signal and nucleophilically attacks the carbonyl carbon between Thr494 and Gly495 via its active site cysteine **(Figure 1.4)**. That reaction results in proteolytic cleavage of the peptide backbone, liberating SpaA from the membrane and forming a ^{Cd}SrtA-SpaA thioester intermediate[87]. Next, a ternary complex forms between the acyl ^{Cd}SrtA-SpaA intermediate and the second backbone SpaA substrate, known as the "attack complex"[88].

The ε-amine group of Lys190 (bolded) in the pilin motif (WxxxVxVYP**K**) of the second SpaA molecule attacks the newly formed thioester bond. This resolves the ^{Cd}SrtA-SpaA acyl adduct and forms a new SpaA-SpaA linkage through an intermolecular isopeptide bond between the carboxyl-terminus of Thr494 and the sidechain amine of Lys190.

Pilin sortases contain several unique structural features. The canonical sortase β-barrel architecture is conserved, but most members of this class also contain an additional N-terminal region that is inhibitory because a portion of it, called the "lid", blocks the active site **(Figure 1.3)**. This lid includes a DP(F/W/Y) motif, which interacts via its conserved aspartate and aromatic residue to anchor the lid in the active site[88–93]. NMR dynamics experiments demonstrated that the lid in the wild-type *S. pneumoniae* pilin sortase adopts a rigid, closed conformation in >99% of wild-type enzyme molecules in solution[89]. This N-terminal "lid" structure docks in the active site as pseudo-substrate, occluding pilin substrate access. Stopped-flow experiments demonstrate lid anchor mutants significantly enhance solvent accessibility of the catalytic cysteine *in vitro*[88]. Thus, sorting signal recognition in this enzyme class requires a conformational change that results in the disengagement of the lid structure from the active site, which is presumably initiated by interactions with the enzyme's substrates or other factors on the cell surface.

The minor pilins, SpaB and SpaC, are responsible for the adhesive properties of the pilus. When either is deleted, pharyngeal adherence is greatly compromised. The minor pilins are displayed both as part of the SpaA-pilus, and as monomeric components of the cell wall[86]. SpaC is located exclusively at the pilus tip, while SpaB is located at base and sometimes interspersed throughout the shaft[96]. Interestingly, deletion of SpaA does not have a significant effect on adherence[86], suggesting that its role is purely structural.

Most Gram-positive pili mirror the three-component architecture in the *C. diphtheriae* SpaA pilus, but less complex pili that contain only two pilin types are also common. These two component pili contain only tip and shaft pilins and are typified by pili in *B. cereus*, *Actinomyces*

spp. and the FCT-1 pilus in *S. pyogenes*[97–99]. In these systems, the pilus polymerizing sortase crosslinks the tip and to a backbone pilin, then polymerizes the shaft. In the absence of the basal pilin which normally promotes cell wall incorporation, the mechanism of polymerization termination is slightly less clear. However, housekeeping sortases are still required for anchoring, as the deletion of these genes prevents attachment to the cell wall[98]. In other cases, pili are assembled using multiple Class C sortases with partially overlapping functions. In which case, both enzymes can polymerize the pilus backbone, but one is sometimes dedicated to incorporating the basal pilin subunit within the pilus shaft or adding the tip pilin[44,51,94,100–102].

Bacterial adherence to host cells begins with initial pilus-mediated attachments followed by subsequent short-range interactions. Long pili can form strong attachments to specific receptors on host cells over long distances. The identity of these receptors have yet to be discovered for many pathogens, but in *Actinomyces oris*, which colonize tooth and mucosal niches, the Type I pilus is known to bind to tooth enamel via salivary acidic proline-rich proteins (PRPs) and statherines[103], while the Type II pilus in this microbe mediates interactions with both host cell surface glycoconjugates and specific polysaccharides on Streptococci in the oral microbiome[104,105]. Preliminary long-range attachment then enables the subsequent formation of “intimate zones of adhesion” by shorter adhesive pili or by non-pilin surface adhesins, as well as between bacterial surface carbohydrates and host lectins[85,86]. The close proximity afforded by this second set of interactions also permits introduction of toxins to the host cell, analogous to type III secretion systems in Gram-negative pathogens[106]. Additionally, in the case of Streptococci, this would provide an opportunity for intracellular invasion. In all, the bacterial adhesion pathway, which begins with pilus anchoring, is a critical step in pathogenesis for many microbes.

1.5 Structural Basis of Pilus Stability

Bacterial pili are responsible for maintaining adhesion to host cells and tissues in the face of considerable physical insults. During infections on mucosal membranes, bacteria will experience drag forces proportional to the velocity of mucociliary flow and the size of the bacterium. Propelled by cilia on epithelial cells lining the respiratory tract, the mucus lining of the lungs is replaced in minutes to hours[107]. As the mucociliary flow can approach 100 μm/sec[108], the corresponding forces can exceed several nanonewtons. Covalent bonds can rupture under forces of this magnitude, so bacterial pili use sophisticated force bearing mechanisms to maintain adherence. While Gram-negative pili are thick fibers with extensive quaternary structure and multi-subunit cross sections, Gram-positive pili are only a single subunit in thickness. Gram-negative bacteria remove stress at critical junctions under force by partially uncoiling their multi-subunit pilin helices[109]. In Gram-positive bacteria, which are constructed with a single subunit cross-section, both intrinsic stability of pilin protomers and inter-pilin linkages contribute to the overall stability of the pilus polymer. Interestingly, knockout of the shaft pilin does not abrogate pilus-mediated bacterial adhesion under static conditions, presumably because monomeric adhesive pilins on the cell surface are still capable of adhering to host receptors. However, under flow conditions which mimic mucus clearance, reduced adhesion is observed,[72] suggesting that the pilus shaft is critical for dissipating the extreme tensile forces encountered by adherent bacteria in their host niches.

The archetypal Gram-positive pilus, the SpaA pilus in *Corynebacterium diphtheriae,* mediates strong attachments to human pharyngeal cells which are structurally reinforced throughout the fiber. At its distal tip, the SpaA pilus features a von Willebrand adhesion (VWA) domain within a specialized tip pilin[86]. The VWA domain mediates noncovalent interactions with receptors on the target host cell that are stable and selective. As discussed in the previous section, other pili employ highly stable thioester bonds harbored in thioester domains (TEDs) as

“chemical harpoons” which form robust covalent linkages to target receptors[110]. Adhesive tip pilins are extended several micrometers from the bacterial surface by a pilus stalk composed of ~100-250 SpaA monomers[111]. The pilus shaft is composed of pilin subunits crosslinked together via isopeptide linkages to result in a single, continuous covalent path for the propagation of axial forces. At a total molecular weight of ~5-12 MDa, these fibers are among the largest known polypeptides[112]. The attachment site of the pilus to the bacterial cell is similarly robust, as the entire pilus is covalently incorporated into the crosslinked peptidoglycan matrix of the cell wall.

Intramolecular bonds in pilin domains form between lysine and asparagine or aspartate side chains which are auto-catalyzed by a proximal acidic residue during protein folding[113]. Isopeptide bonds are robust and unreactive: Unlike disulfide bonds, they are unaffected by changing redox conditions[114]. When assessed by atomic force microscopy, Gram-positive pili consisting of Cna-type domains are classified as “ultramechanically stable,” as the forces required to unfold them are the largest reported for single globular proteins (SpaA requires pulling forces of ~525 pN)[112]. Cna domains are immunoglobulin-like folds, which are intrinsically highly stable even in the absence of intradomain crosslinking[115]. However, additional stabilization in the form of isopeptide crosslinks is a conserved feature in Gram-positive pili subunit domains, indicating that the forces encountered by pili exceed the range of even such mechanically stable immunoglobulin domains as are found in titin or fibronectin[115–117].

The stability of each constituent backbone pilin is critical to the overall mechanostability of the pilus fiber. SpaA protomers are composed of three linearly arranged Cna-type domains[118]. The N-terminal domain (^{N}SpaA) and the C-terminal domain (^{C}SpaA), are CnaB type folds which house the reactive pilin lysine and sorting signal peptide, respectively. ^{C}SpaA and the middle domain share interlocking secondary structural elements and comprise a rigid linear unit. ^{C}SpaA is a force-bearing CnaB domain which features an isopeptide linkage located between the first and last β-strands of the fold[113]. This positioning allows axial force to travel

from the C-terminal signal peptide on the final strand (which is in turn crosslinked to the pilin lysine in the subsequent protomer) directly to the first strand through the internal bond, rendering the ^{C}SpaA domain mechanically inextensible[112]. The path of the axial force would then proceed to the final strand of the middle domain, which is a CnaA-type fold. This domain-type contains a slightly different internal isopeptide which forms a bridge between the first and penultimate β-strands. As a result, there are 29 residues in the intervening loop between internal linkage and the N-terminus of ^{C}SpaA, called the isopeptide delimited loop (IDL) **(Figure 1.5)**. The IDL is part of the force pathway and can be extended at large forces, providing an efficient mechanism for energy dissipation and shock absorption[112]. After the mechanical shock subsides, the CnaA domains rapidly refold into the stable ground conformation, primed for subsequent mechanical shocks. The length of the pilus, and thus the IDL reservoir, scales linearly with the forces that can be tolerated, suggesting that longer fibers can handle larger shocks. Thus, it has been suggested that pilus biogenesis may be upregulated to produce longer fibers which can accommodate increased external forces. The vast majority of shaft pilins contain a CnaA-type domain, suggesting that this strategy of force dissipation is widespread. Conversely, pili that are composed entirely of CnaB-type domains are predicted to be essentially inextensible.

In the case of the *C. diphtheriae* SpaA pilus, ^{N}SpaA lacks an internal isopeptide bond and is relatively flexible with respect to the rest of the protein, as evidenced by disorder observed in crystal structures[119–126]. However, the inter-pilin linkage is located at the C-terminal boundary of the N-terminal domain, such that the “load bearing spine” of force transduction through the pilus fiber bypasses the entirety of ^{N}SpaA, rendering the additional stabilization inconsequential to the overall stability of the pilus[121]. From an evolutionary standpoint, there is little selective pressure to maintain the internal isopeptide linkages in domains which are not force-bearing. However, there are several examples of N-terminal domains which do contain

intradomain isopeptide bonds, but these are typically slow-forming relative to the other Cna domains[43,121,126]. Whereas internal isopeptides in most Cna-type domains form autocatalytically during protein folding, the corresponding intradomain linkage in the pilin NTD apparently has a higher energy barrier to overcome. In the pilin motif (YPKN), the pilin Lys (the site of the intermolecular bond) is positioned directly adjacent to an Asn residue (the site of the intramolecular bond). As such, formation of the inter-pilin linkage may influence the subsequent formation of an additional inter-domain linkage by reordering the catalytic Asn, Glu, and Lys residues into a more optimal orientation for auto-catalytic proximity-based bond formation to occur. Indeed, in several pilins characterized to date, the ligation event also results in efficient formation of the intradomain isopeptide bond, which may effectively lock the NTD into a rigid conformation[43,121,127]. For example, in crystal structures of the full-length RrgB pilin in *S. pneumoniae,* docking of the signal peptide to the binding cleft of RrgB initiates the formation of an isopeptide bond within the N-terminal domain, but the bond is not observed in the absence of signal peptide docking[43]. In the BcpA pilus from *Bacillus anthracis*, the internal isopeptide bond is detected in native pili, but not in the recombinant forms, indicating that the inter-subunit linkage must be formed prior to formation of the internal isopeptide in the pilin lysine-bearing domain[29].

The crystal structure of GG-SpaA from *Lactibacillus rhamnosus* features several crystal forms in which the NTD exhibits a range of bent conformations relative to the rest of the protein. This intriguing structural heterogeneity led the authors to propose an “expose, ligate, seal” mechanism in which the nucleophilic N-terminal domain exists in a bent, unliganded form which favors nucleophilic attack. Following ligation, the NTD assumes a closed form in which it become locked into a rigid, linear conformation which propagates throughout the pilus fiber[126]. Pilin NTD flexibility in the apo-state may also be required for steric reasons. The pilin sortase is tethered to the membrane and may require a degree of flexibility in order to bridge the gap

between the two elongated substrates which are also membrane-bound. As such, it is likely that flexible the N-terminal domain functions primarily to mediate the interaction with the pilin sortase during polymerization. Thus, the importance of the NTD is primarily in mediating sortase recognition and not in structural integrity.

While the tensile strength and structural organization of individual pilin proteins is well understood, little is known about the junction between pilin protomers in the fiber. In Chapter 4 of this thesis, I describe a detailed structural and biophysical investigation of the inter-pilin linkage in *C. diphtheriae* SpaA pilus, which yields several important insights about stabilizing features present at linkage sites within the pilus fiber.

1.6 *In vitro* Bioconjugation Approaches using Sortase

Protein bioconjugates have proven potential for transformative applications from basic biological research to medical therapy. Specific uses include creation of specific antibody-drug conjugates, small molecule- or fluorophore-labeling of proteins in biophysical experiments, orientation-specific protein immobilization, cell-specific labeling, and the preparation of multifunctional protein nanoparticles and complexes for industrial purposes. However, specific and efficient bioconjugation remains a difficult problem.

Early approaches to protein modification aimed to leverage the differential reactivity profiles of specific amino acid side chains, including cysteine, lysine, aspartate, glutamate, tyrosine, tryptophan and the N- or C-termini of proteins[128,129]. Recently, non-natural amino acid incorporation techniques enabled the expansion of functionalities available on the protein surface, allowing for highly specific "bio-orthogonal" conjugation chemistries[130–132]. This class of reactions is suitable in certain circumstances, but if higher selectivity, milder reaction conditions or enzymatic control are required, alternative methodologies are necessary.

Several enzymatic approaches involving the development of enzymes which natively catalyze ligation reactions have yielded successful bioconjugation reagents. However, most are still limited by issues surrounding incomplete yield, reversibility, ligation-site promiscuity and long reaction times. Specifically, the use of naturally occurring transferase-type enzymes have been described, but the specificity of such reactions requires non-protein synthetic substrate analogs and are thus limited to niche applications. Examples of these approaches include farnesyltransferases[133], N-myristoyltransferases[134] and biotin ligases[135], which all require specifically engineered substrates for recognition that harbor isoprenoid, myristic acid or biotin analog functional groups, respectively.

Currently, the sortase A enzyme from *Staphylococcus aureus* (^{Sa}SrtA) is one of the most widely used bioconjugation enzymes, owing to its superior specificity and genetically-encoded

recognition tags. ^{Sa}SrtA catalyzes a transpeptidation reaction which involves recognition and cleavage of a specific peptide sequence followed by formation of a new amide linkage to the N-terminus of a second acceptor substrate **(Figure 1.6A)**. The first demonstration of efficient *in vitro* ligation using ^{Sa}SrtA was in 2004. In this study, Mao and coworkers showed that ^{Sa}SrtA is capable of ligating two recombinant protein substrates with minimal peptide tags: one with a C-terminal LPXTG signal and the other with a N-terminal oligoglycine[136]. Following this discovery, this general "sort-tagging" approach has been successfully deployed in protein-protein ligation[137], backbone cyclization[138], as well as site-specific fusions of target proteins to peptides, lipids, sugars, and small molecules[136,139,140]. ^{Sa}SrtA ligation chemistry has also been harnessed to enable immobilization of target proteins to coverslips for microarray experiments[141]. Recently, its utility in enabling biophysical studies of previously inaccessible systems in solution has been demonstrated. In this approach, ^{Sa}SrtA is used to append NMR-silent solubility tags to unstable protein targets to generate solubilized segmentally isotope-labeled domains for NMR studies[142,143]. Further advancements involved developing ^{Sa}SrtA variants which lack divalent metal cofactor dependence, which are capable of labeling living cells under low Ca^{2+} conditions[144,145].

^{Sa}SrtA bioconjugation is not without limitations. Reversibility is one major concern, as the ligation product itself contains an LPXTG motif which is a substrate for subsequent hydrolysis. Additionally, a peptide fragment containing an N-terminal Gly is generated as a product of the hydrolysis reaction, which may compete with the intended nucleophile in the reaction. Another obstacle is the relatively slow *in vitro* kinetics of wild-type ^{Sa}SrtA. Kinetic analysis of *in vitro* transpeptidation determined relatively weak binding affinities for the sorting signal (K_m = 7.33 mM) and Gly_5 substrates (K_m = 196 μM), as well as a low turnover number (k_{cat} = 0.28 s^{-1})[49]. These issues can be partially ameliorated by using large molar excesses of enzyme and sorting signal. Several creative approaches also address this limitation by using affinity capture

methods to increase the local concentration of substrate and purify a homogenous reaction product in a single step[146,147] or removing of the glycine leaving group by dialysis or centrifugation to prevent enzymatic reversal[143,148,149].

However, the same transpeptidation reaction occurs very rapidly *in vivo.* For example, the entire process from expression to secretion to cell wall attachment by ^{Sa}SrtA takes less than three minutes in the case of protein A anchoring[150]. As discussed in section 1.2, this is most likely due to the unfavorable protonation states of the active site residues in *in vitro* environments, although it could also be attributable to the decrease in optimal substrate access afforded by membrane anchors on both the enzyme and the substrate. As it became clear that this enzyme was vastly underperforming its kinetic potential in *in vitro* applications, several directed evolution campaigns aimed to improve activity. Chen and coworkers developed an evolved enzyme (dubbed ^{Sa}SrtA^{5M}) with an 140-fold improvement in coupling efficiency relative to wild-type[151]. Further improvements of ^{Sa}SrtA^{5M} yielded new variants with improved reaction parameters and optimized non-canonical conjugation techniques, including enzymes specialized in either N- or C-terminal modifications of antibodies[152].

Production of sophisticated bioconjugates sometimes requires simultaneous conjugation of single two distinct groups to a single protein target. Yet another directed evolution campaign developed variants with altered substrate profiles with exciting applications for orthogonal modification approaches[153]. This study resulted in the development of two additional transpeptidase variants which preferentially recognize LAXTG and LPXSG signal motifs which could allow simultaneous modification of the same target protein with two different substrates harboring the distinct signal motifs. Sortase homologs from other species with naturally divergent substrate preferences have also been used orthogonally[154]. Specifically, the sortase from *Streptococcus pyogenes* accepts nucleophile substrates with N-terminal alanine residues as opposed to the canonical glycine recognition motif, which paves the way for further

orthogonal modification workflows[154–156]. Nevertheless, conventional sortase bioconjugation is limited to C- or N-terminal backbone peptide bond attachments and is prone to enzymatic reversal of its own reaction product.

Enzymes capable of producing isopeptide linkages (peptide bonds where either the carboxyl or amino reactive group is part of a sidechain) between substrates have several advantages over backbone peptide bond ligases. First, most proteases target backbone peptide motifs, so isopeptide linkages are more resistant to proteolysis and thus more stable. Second, side chain ligations greatly expand the number of potential labeling sites on a given protein beyond the two termini. Finally, isopeptide bioconjugation reagents can be used in combination with existing backbone approaches to enable specific attachment of multiple molecules on a single protein.

Several enzymatically-catalyzed and auto-catalyzed approaches to promote intermolecular isopeptide conjugation have been described. ^{Sa}SrtA is normally specific to N-terminal glycine nucleophiles, but under certain conditions, it can catalyze a noncanonical isopeptide bond formation between LPXTG and pilin motif peptides. However, this promiscuous activity is a side reaction which is significantly less specific and efficient[157–159]. Transglutaminase enzymes catalyze isopeptide crosslinks between lysine and glutamine sidechains, but this reaction is also highly nonspecific and primarily limited to bulk crosslinking applications such as in the food and textile industries[160,161].

Recently, by leveraging the autocatalytic internal isopeptide bond formation found within bacterial adhesin domains, an ingenious nonenzymatic approach for isopeptide ligation was described. This system was originally developed through dissection of a CnaB domain from the fibronectin binding adhesin in *Streptococcus pyogenes*, FbaB, into two components: a short peptide tag (SpyTag) and the complementary domain (SpyCatcher)[162]. The resultant fragments were then carefully engineered to recapitulate the spontaneous isopeptide formation observed

in the native domains, yielding a robust tagging system which is irreversible, highly stable, and complete within minutes. Further optimization resulted in a three part system, consisting of two short peptide tags (KTag and SpyTag) which are ligated in a similar manner when docked into the remainder of the FbaB domain (SpyLigase)[163]. This system offers a significant advantage by decreasing the size of the fusion required on each target protein, as the SpyLigase component dissociates following bond formation. A similar three-part system has also been developed with the tip pilin RrgA from Streptococcus pnuemoniae, called SnoopLigase[164].

Isopeptide linkages are also enzymatically catalyzed by pilin sortases, suggesting that these enzymes could be developed into a novel class of bioconjugation tools. The Class C sortase in *C. diphtheriae* polymerizes pili through the repeated catalysis of isopeptide linkages between the constituent subunits. In a mechanism similar to ^{Sa}SrtA, the first step of the reaction involves proteolytic cleavage of a C-terminal signal motif, resulting in a thioacyl sortase-pilin intermediate. The second step involves nucleophilic attack on the thioester bond by the sidechain amine group from a reactive lysine. As a result, a new intermolecular isopeptide bond linking two pilin protomers is formed **(Figure 1.6B)**. This enzyme serves as an excellent starting point for an isopeptide bioconjugation tool. The atomic structure of ^{Cd}SrtA revealed that its active site is masked by a polypeptide appendage called a “lid,” which is hypothesized to preventing substrate binding. Indeed, transpeptidation activity is severely inhibited *in vitro*, as the wild-type enzyme has no observable activity when expressed recombinantly[91]. As described in Chapter 2 of this thesis, a major breakthrough occurred when it was demonstrated that pilus biogenesis could be recapitulated *in vitro* by an enzyme variant harboring targeted mutations to anchoring residues within the lid[91]. A follow-up study, which is described in detail in Chapter 6, developed the reaction conditions necessary for a second-generation enzyme variant, called ^{Cd}SrtA^{3M}, to create site-specific protein-protein and peptide ligations[165]. Importantly, this novel bioconjugation

tool can be combined with the canonical ^{Sa}SrtA to install orthogonal additions to proteins owing to the distinct substrate profiles of each biocatalyst.

Further optimization involved removal of the entire lid structure, which yielded an enzyme variant with 7-fold enhancement in *in vitro* catalytical turnover compared to the previously described activated enzyme variant, ^{Cd}SrtA$^{3M\ 166}$. A complete discussion of the kinetic parameters of these ^{Cd}SrtA variants is described in Chapter 3 of this thesis. Interestingly, while other sortase classes are agnostic to the amino acid in the X position of the LPXTG signal motif, ^{Cd}SrtA demonstrates selectively for the native leucine at the position, offering another level of enzymatic specificity for multi-step reactions involving several sortase additions[166]. While the pilin polymerase-catalyzed isopeptide ligation system described here involves a large fusion that may not be desirable in certain applications, the site selectivity of the reaction and temporal enzymatic control offer unique advantages not found in other bioconjugation approaches.

1.7 Degradation of Recalcitrant Biomass for Biofuels and Bioproducts

Expanding global energy requirements due to rapid industrialization and population growth are straining traditional fossil fuel sources and raising fundamental concerns about the sustainability of our current energy infrastructure. Further, accumulation of carbon in the atmosphere yields undesirable environmental effects such as global warming. Thus, development of alternative and renewable energy technologies is highly desirable. Conversion of cellulosic biomass is one such promising approach for the production of sustainable liquid fuels. Nonrenewable fuel sources release carbon to the atmosphere that was once sequestered in the Earth, leading to carbon debt. Conversely, biofuels are considered to be carbon-neutral energy sources due to their integration into the biomass cycle: Carbon released as a product of burning biofuels during heat or power generation is subsequently reabsorbed through photosynthesis by new biomass, which is again converted into new biofuels.

As the main structural component of plant cell walls, lignocellulose is highly abundant and renewable, representing an attractive feedstock from which valuable biofuels and other biocommodities can be extracted. However, lignocellulose is extremely recalcitrant to degradation because plants have evolved defenses to prevent the deterioration of their structure by the elements or microbial processes. Thus, innovative methods are needed to convert it into its component sugars for downstream utilization. Lignocellulose is comprised of three major components: cellulose (25-55%), hemicellulose (8-30%), and lignin (18-35%)[167]. Lignin is a highly complex polymer containing a mixture of phenolic compounds linked through radical coupling reactions[168]. Lignin is the most challenging component to degrade owing to its complex and variable structure, and requires the concerted action of several oxidative enzymes such as peroxidases and laccases[169]. Additionally, lignin slows enzymatic degradation of crude biomass because it sequesters direct access to more easily digestible cellulose and hemicellulose components. Innovations in pretreatment of biomass involving mechanical, chemical or

biological processes help alleviate recalcitrance by disrupting the structure of lignin surrounding the cellulose components[170,171]. Hemicellulose is a sugar polymer that is composed of a number of different types of pentose and hexose sugars[172]. Hemicellulose degradation requires by a range of enzymes, including: xylanases, arabinases, and mannases[173]. Cellulose, a linear glucose homopolymer joined by β-1,4-glucosidic linkages, is the least complex component of biomass. The polymers form amorphous and crystalline structures through hydrogen bonding between glycan strands[167,174]. Despite considerable effort, a truly cost-effective and practical method for cellulolytic degradation on a large scale has not been developed. The rest of this chapter will focus specifically on the degradation of cellulose.

Bacterial and fungal organisms have evolved the capacity to degrade cellulose through concerted action of cellulase enzymes that hydrolyze the constitutive β-1,4-glucosidic linkages[173]. *Clostridium thermocellum* is a well-studied anaerobe which efficiently degrades lignocellulose by a coordinated multicomponent system. Degradation is carried out by the synergistic action of by three major types of cellulases: endoglucanases, exoglucanases, and β-glucosidases[175]. Substrate recognition is based on bulk structural features and differs between enzyme families. Specifically, endoglucanases recognize amorphous cellulose regions and hydrolyze at random internal positions of cellulose polymers. In contrast, exoglucanases processively degrade polymeric cellulose strands at reducing ends **(Figure 1.7)**. These activities are synergistic: Random internal cleavage of cellulose chains by endoglucanases creates available chain ends for further degradation by exoglucanase action[176]. Short glycans (such as cellobiose) that are produced by endo- and exo- acting cellulases are further degraded into glucose monomers by β-glucosidases[175,177]. To enhance this synergy, cellulose-degrading microbes display large supramolecular structures called cellulosomes, consisting of a central scaffoldin protein with multiple binding sites harboring various cellulases with different specificities[178–180]. Scaffoldins contain multiple cohesin sites which bind with very high affinity to

dockerin domains on cellulases. Carbohydrate-binding domains (CBMs) found on cellulases or the scaffoldin itself target the enzymes to crystalline cellulose and anchor the entire complex to its solid substrate[180]. The organization of cellulolytic enzymes into co-localized clusters promotes significant synergy due to optimal enzyme spacing which allows efficient channeling of reaction products between enzymes with complementary activities[181–183]. When compared to secreted enzyme systems from *T. reesei* which lack the synergy promoted by cellulosome co-docking, the specific activity of crystalline cellulase degradation by cellulosome-anchored cellulases is 15-fold enhanced[184]. Moreover, sugar uptake by the microbe is promoted by the close association of the cellulosome to the microbe surface, which removes potential enzyme feedback inhibitors such as glucose and cellobiose[185,186]. Thus, extensive synergy potentiated by co-localization on cellulosomes yields a highly efficient cellulose degradation machine.

Inspired by naturally occurring multienzyme cellulosomes, successful synthetic approaches for cellulose degradation involve co-localization of existing microbial cellulolytic enzymes. One approach is to genetically encode multiple pathways from cellulose degradation to bioproduct metabolism into living microbial cells, called consolidated bioprocessing organisms (CBPs)[187]. The advantage of this approach is in negating the need for the expensive addition of exogenous enzymes and multiple processing steps. CBP organisms have potential, but thus far genetically engineering organisms capable of simultaneous cellulose breakdown and fermentation of the soluble products into biofuels is a major challenge[188–190]. Controlling the metabolism of CBPs such that cellulase degradation and bioproduct conversion is favored over other cellular processes is another challenging problem[187].

Another attractive avenue of development involves the generation of synthetic cellulosome scaffolds. Chimeric mini-cellulosomes consist of recombinantly expressed individual cellulases and scaffold proteins. Clever approaches using chimeric dockerin-cohesin constructs have been successfully employed to direct the attachment of specific cellulases to

chosen positions on the scaffold to test pairwise interactions between cellulase families[191]. Cellulolytic degradation efficiency also increases with enzyme diversity on reconstituted cellulosomes, as trifunctional particles are considerably more active than bifunctional designer cellulosomes, especially when targeting more complex biomass substrates[192]. Synergistic degradation can also be optimized through tuning of the stoichiometry of individual cellulase components[193]. However, while reconstituted chimeric cellulosomes display impressive synergy and activity enhancement compared to free enzymes, they are still less active than native cellulosomes[194]. Additionally, the scaffoldin proteins are notoriously difficult to express recombinantly, owing to their large molecular weight and repetitive genetic sequence[191,193–195].

Completely synthetic scaffolds, which remove the dependence on native scaffoldin proteins, have been designed to co-localize cellulases and recapitulate the synergistic effects observed in native systems. In one approach, a self-assembling ring structure fused to cohesion domains, called the rosettazyme, was employed[189]. Dockerin-containing cellulases were bound to the complex and resulted in ~2.4-fold greater cellulose degradation than for free enzymes. Another approach appended cellulases to polystyrene nanospheres and reported 50-130% enhancements in crystalline cellulose degradation compared to free enzymes[196]. Importantly, this effect is primarily attributable to enhanced binding of the insoluble substrate due to avidity effects, as immobilization of cellulases on nanospheres does not enhance relative activity on soluble cellulose substrates.

Recent advances in computational protein engineering have enabled creation of specifically designed protein nanocages which self-assemble into precise architectures[197]. These synthetic scaffolds can accommodate multiple enzyme placements with defined geometric relationships. The degree of synergy displayed by cellulases is dependent on spatial organization and stoichiometry. Affixation of cellulase enzymes to a protein nanocage with known geometrical relationships gives the researcher the ability to control and optimize the

distance between enzymes displayed on the surface of the cage. Development of synthetic enzyme-display systems which mimic natural cellulosomes represents a promising new direction for tailored cellulolytic degradation of target substrates and highly efficient biomass conversion. As described in Chapter 5, site-specific bioconjugation of enzymes with via sortase chemistry is an especially facile and robust platform for producing these synergistic systems synthetically.

1.8 Figures

Figure 1.1 – Sortase classification. Sortases are classified by phylogenic characterization. Class A sortases are generalized “housekeeping” enzymes which affix a wide array of substrates to the cell wall. This enzyme class is found in most Gram-positive bacteria. Conversely, the following sortase classes are found only in select bacteria: Class B sortases are affix substrates involved in nutrient acquisition, such as heme capture. Class C sortases are pilin polymerases which catalyze isopeptide linkages between pilin shaft protomers to create the pilus shaft. Class D sortases affix substrates involved in spore formation.

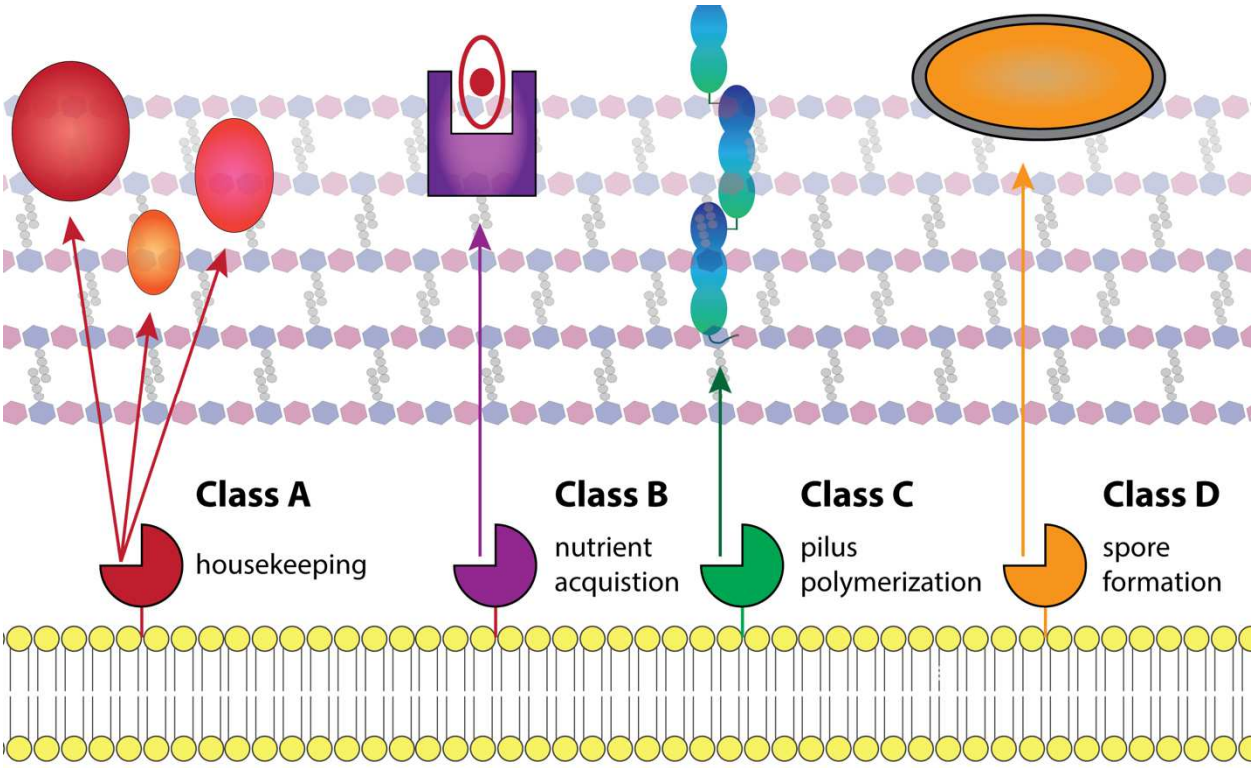


Figure 1.2 – Class A and C sortase mechanisms. The preliminary acylation step of the sortase mechanism is conserved in both classes. Both sortases recognize the LPXTG sorting signal motif on its substrates and the active site cysteine nucleophilically attacks the Thr-Gly bond within the CWSS, resulting in an acyl-enzyme intermediate (1). After which, the mechanisms diverge for the second step monomeric substrate anchoring sortase classes ("a" steps, top) and pilin polymerizing sortase classes ("b" steps, bottom). Cell wall anchoring (2a) requires a second nucleophilic attack by the pentaglycine crossbridge within a Lipid II cell wall precursor (2b), resulting in a new backbone peptide linkage between the extracellular protein and Lipid II, which is then incorporated in the mature peptidoglycan. Pilus biogenesis (2b) requires a second nucleophilic attack by a reactive pilin lysine within the pilin motif a second pilin protomer, resulting in an isopeptide linkage between pilins. The pilus is either elongated by repetition of this process, or anchored to the cell wall by separate housekeeping sortases in a process similar to (2a).

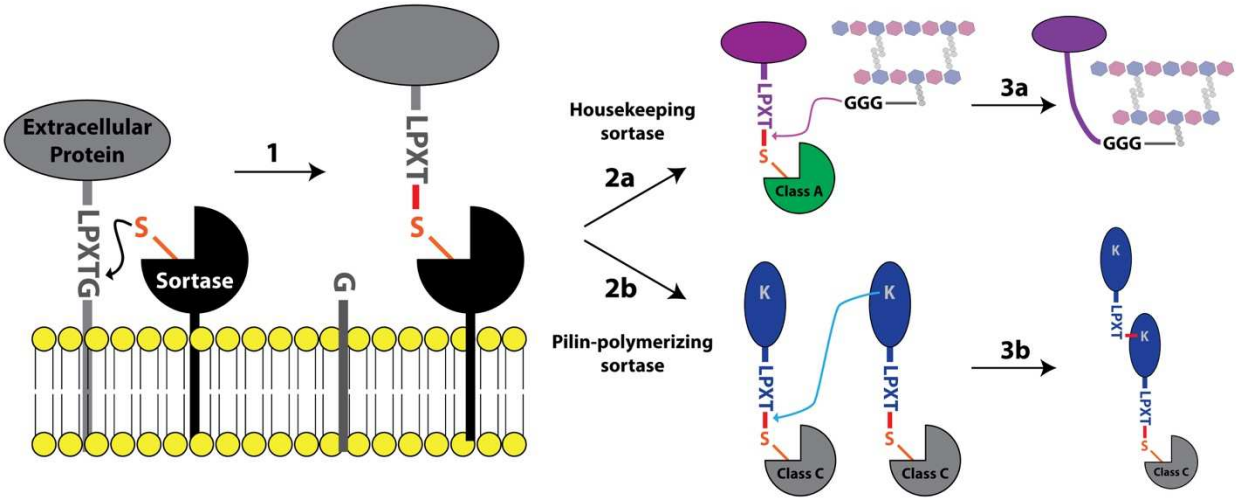


Figure 1.3 – Structural comparison of Class A and C sortases. Topology of the Class A sortase from *Staphylococcus aureus,* ^{Sa}SrtA (PDB 1IJA) (A-C) and the Class C sortase from *Corynebacterium diphtheriae,* ^{Cd}SrtA (PDB 5K9A) (D-F). The β6/β7 loop and β7/β8 loops are colored green and blue. A/D) Surface rendering with important loops show as cartoon representation. B/E) Cartoon rendering of both enzymes to highlight secondary structural elements and active site residues (yellow). C/F) Expanded view of the arrangement of active site residues. For ^{Cd}SrtA, the interaction between the lid and active site residues is also highlighted, with conserved lid residues W83 and D81 represented as sticks.

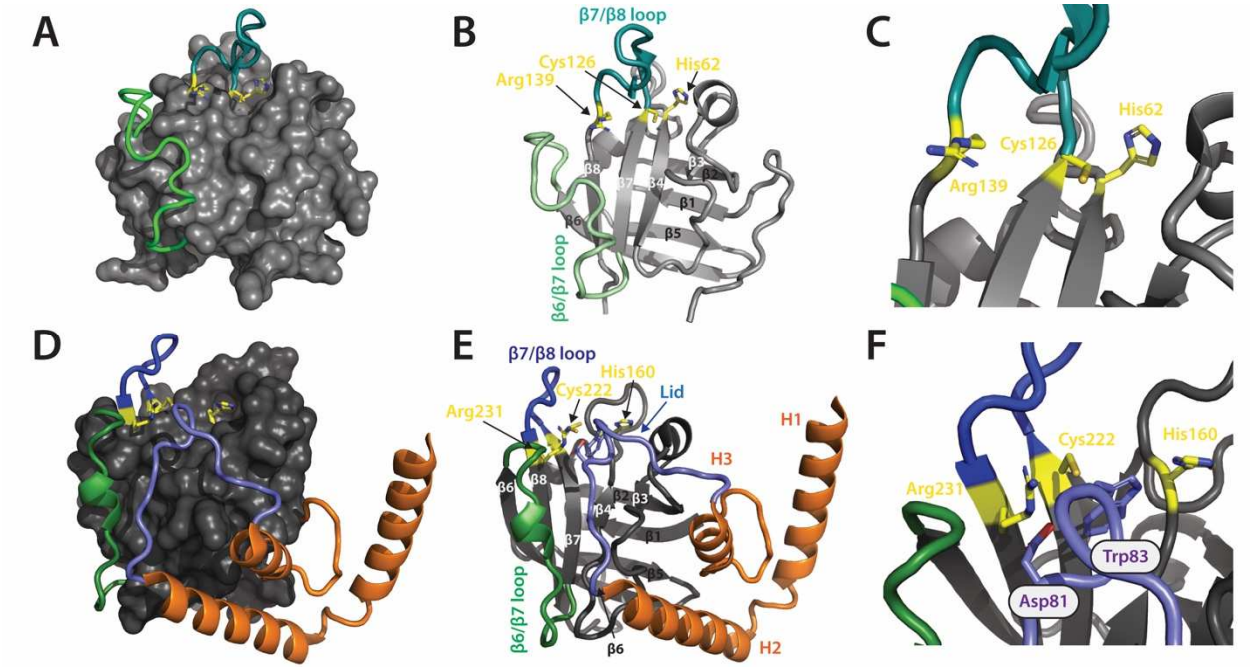


Figure 1.4 – Sortase-catalyzed pilus biogenesis in *Corynebacterium diphtheriae*. (1) Pilin precursors (SpaA) harboring an N-terminal signal peptide are directed to the Sec translocon for secretion. (2) The N-terminal signal peptide is cleaved by signal peptidases and the pre-pilin is partially secreted, but remains tethered to the cell membrane by a single pass transmembrane domain composed of a hydrophobic string of residues (h) and a cytoplasmic anchor of positively charged basic residues (+). The pilin sortase (SrtA), likewise tethered to the membrane, recognizes its substrates via an extracellular 5-residue LPXTG motif. The sequence is cleaved, resulting in a SrtA-SpaA thio-acyl intermediate. (3) An adjacent SpaA molecule resolves the acyl intermediate through nucleophilic attack by the ε-amine group on a reactive Lys in the N-terminal domain of SpaA, resulting in an isopeptide linkage between successive SpaA protomers. (4) The pilus fiber is eventually ligated to the cell wall by a housekeeping sortase. Schematic of the threonine-lysine isopeptide is shown (inset).

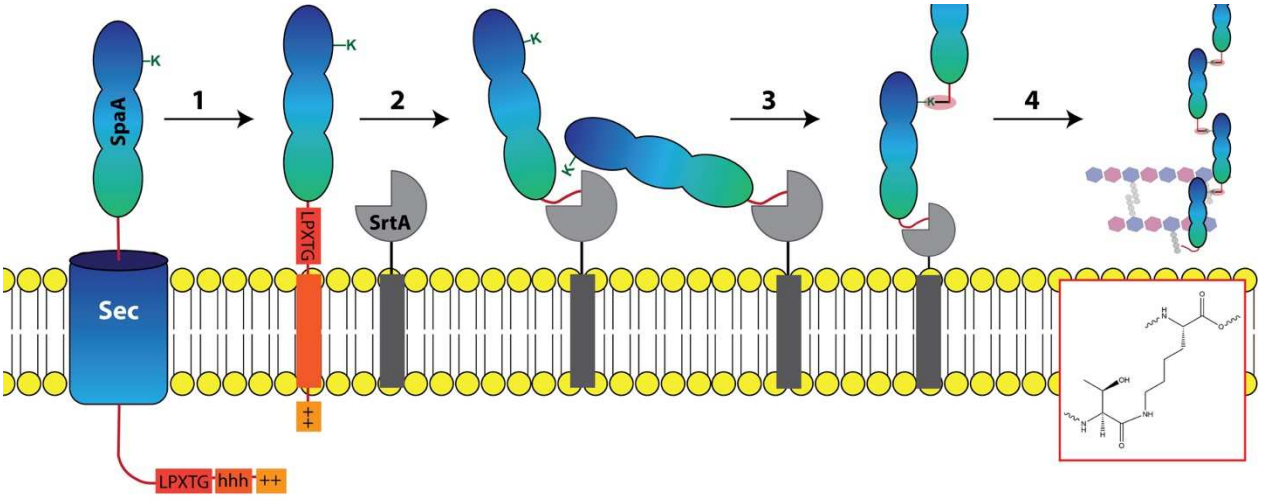


Figure 1.5 – Force transduction through the SpaA pilin subunit. A) The topology of the SpaA molecule is shown, with ^{N}SpaA colored blue and the middle and C-terminal domains colored green. Orange and pink arrows indicate the positions of isopeptide and disulfide bonds, respectively. B) The path of force transduction through the SpaA molecule is depicted in cartoon representation, with the rest of the molecule hidden. The isopeptide delimited loop (IDL) is indicated. Black arrows indicate pulling forces. Figure adapted from [112]. An expanded view of the internal disulfide and isopeptide bonds is shown for ^{C}SpaA (C) and ^{M}SpaA (D). E) Crystal lattice arrangement of the SpaA pilus is shown as surface and cartoon representations, with the path of force transduction highlighted in red.

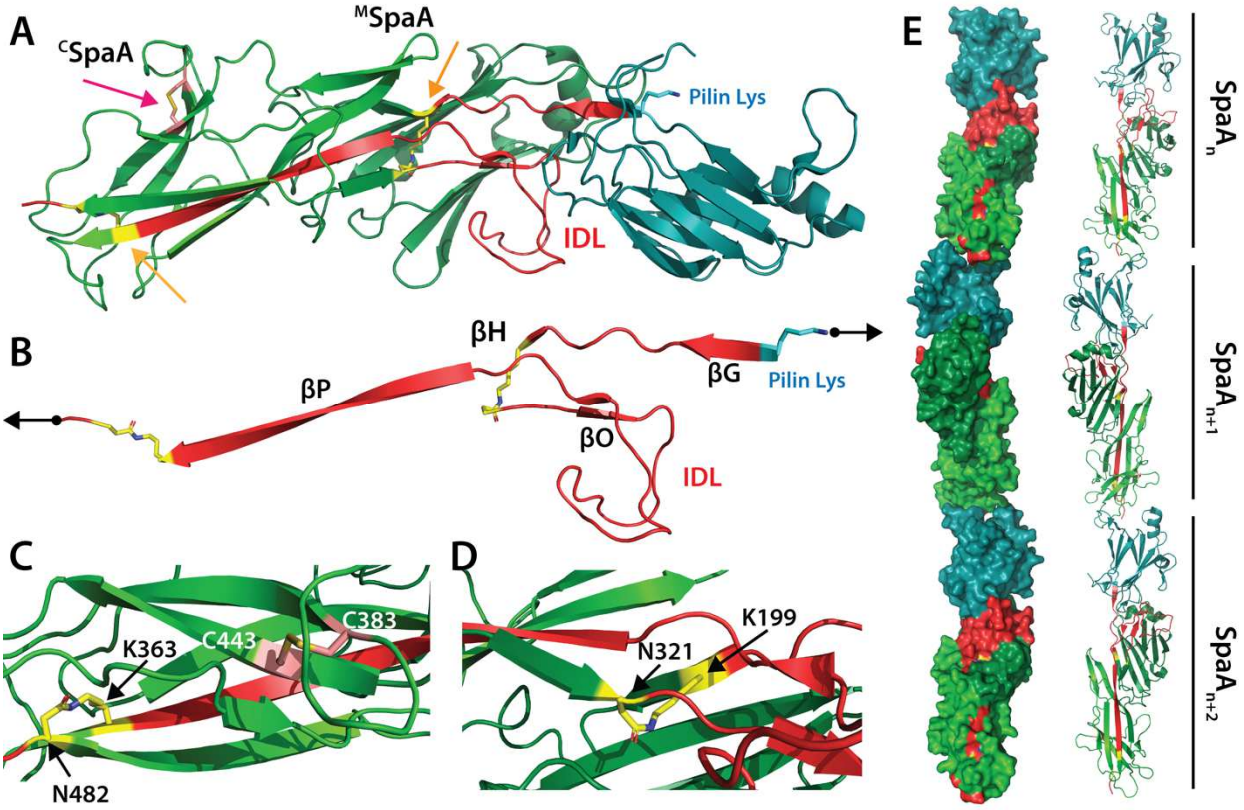


Figure 1.6 – Sortase bioconjugation. A) Class A sortases catalyze bioconjugation of recombinant proteins and peptides with N- and C-terminal oligoglycine or LPxTG sorting signal tags, respectively. Class A sortase hydrolyzes the LPxTG motif at the Thr and Gly peptide bond and catalyzes a transpeptidation reaction to join the sorting signal motif to the N-terminal oligoglycine via a backbone peptide bond. B) Class C sortases catalyze bioconjugation of recombinant proteins and peptides with pilin motif domain (PM) fusions or LPLTG sorting signal tags, respectively. Class C sortases hydrolyze the LPLTG motif at the Thr and Gly peptide bond and catalyzes a transpeptidation reaction to join the sorting signal motif to the reactive lysine side chain harbored in the pilin motif domain via an internal isopeptide bond. N- and C- termini of the target protein are indicated with Nt and Ct, respectively.

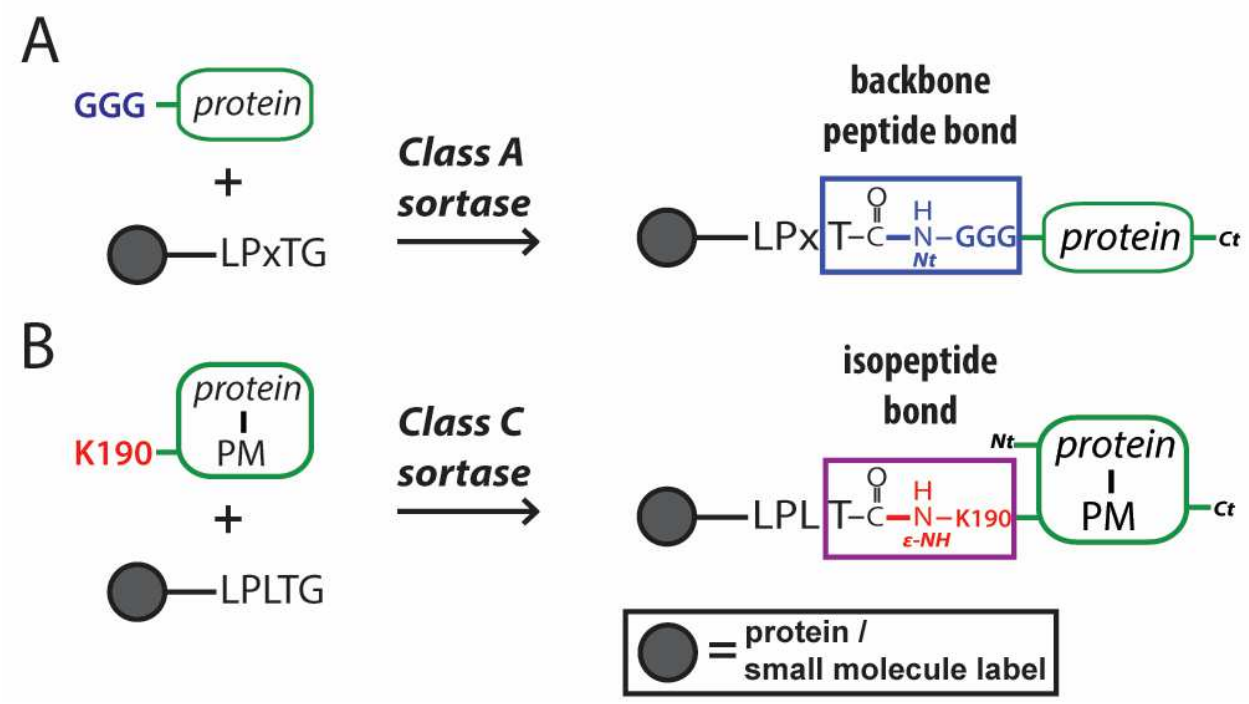
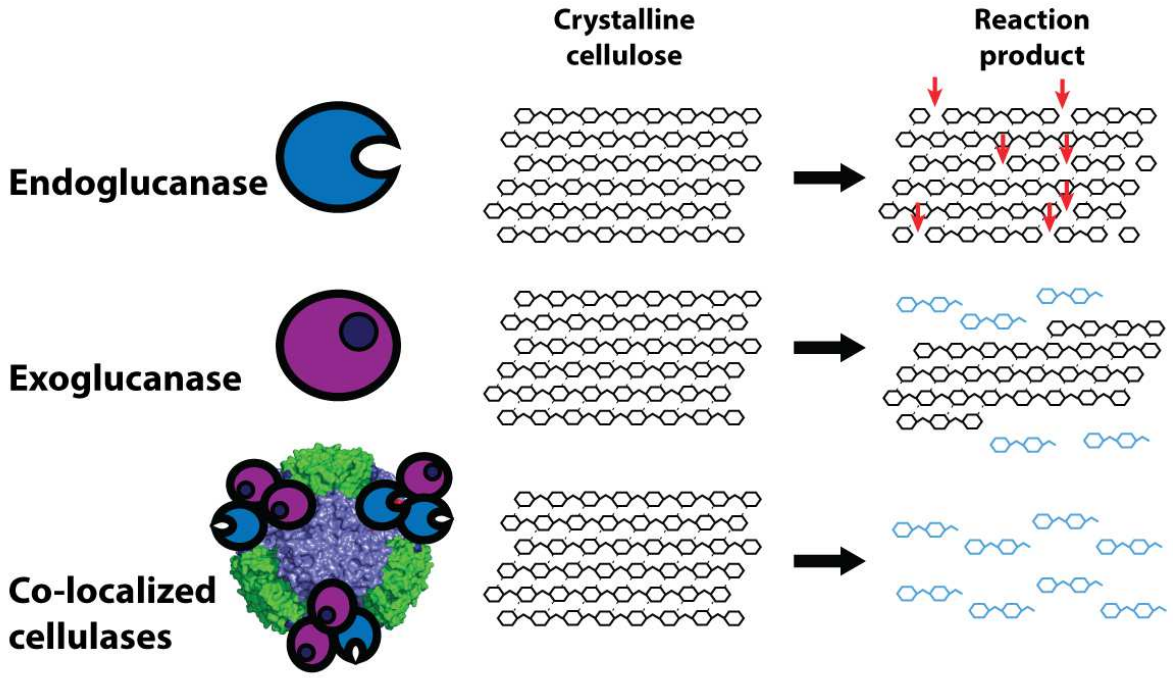


Figure 1.7 – Cellulose degradation by cellulases. Enzymatic degradation of crystalline cellulose by bacterial cellulases is accomplished by the concerted action of endo- and exo-glucanases. Endoglucanases are responsible for hydrolyzing internal by β-1,4-glucosidic linkages within crystalline regions of cellulose substrates. The reaction results in internal cleavages introduced within the glycan strands (indicated by red arrows). Exoglucanases (purple) degrade cellulose polymers at reducing ends of cellulose produced by endoglucanases in a processive manner. The reaction results in hydrolysis of glycan strands to product disaccharides (light blue), which are later cleaved into glucose sugars by β-glucosidases (not shown). When co-localized on solid scaffold (shown here as a protein nanoparticle) the intrinsic synergy of these activities is greatly enhanced.



1.9 References

1. Adedeji, W. A. The Treasure Called Antibiotics. *Ann. Ibadan Postgrad. Med.* **14**, 56–57 (2016).

2. Brown, E. D. & Wright, G. D. Antibacterial drug discovery in the resistance era. *Nature* **529**, 336–343 (2016).

3. Fleming, A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *Br. J. Exp. Pathol.* **10**, 226–236 (1929).

4. Ventola, C. L. The antibiotic resistance crisis: causes and threats. *P T J.* **40**, 277–283 (2015).

5. CDC. *Antibiotic resistance threats in the United States, 2019*. *Centers for Disease Control and Prevention, US Department of Health and Human Services* (2019).

6. Bartholomew, J. W. & Mittwer, T. The Gram stain. *Bacteriol. Rev.* **16**, 1–29 (1952).

7. Mainous III, A. G., Pomeroy, C., Stalder, T. S. & Preheim, L. C. Gram-Positive Bacteria. in *Management of Antimicrobials in Infectious Diseases* 29–41 (Humana Press, 2003). doi:10.1385/1-59259-036-5:29

8. 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**, 174–229 (1999).

9. Jacobitz, A. W., Kattke, M. D., Wereszczynski, J. & Clubb, R. T. Sortase Transpeptidases: Structural Biology and Catalytic Mechanism. in *Advances in Protein Chemistry and Structural Biology* **109**, 223–264 (Academic Press Inc., 2017).

10. Vollmer, W., Blanot, D. & De Pedro, M. A. Peptidoglycan structure and architecture. *FEMS Microbiology Reviews* **32**, 149–167 (2008).

11. Schneewind, O. & Missiakas, D. Sec-secretion and sortase-mediated anchoring of proteins in Gram-positive bacteria. *Biochim. Biophys. Acta - Mol. Cell Res.* **1843**, 1687–1697 (2014).

12. Pallen, M. J., Lam, A. C., Antonio, M. & Dunbar, K. An embarrassment of sortases- A richness of substrates? *Trends in Microbiology* **9**, 97–101 (2001).

13. Comfort, D. & Clubb, R. T. A Comparative Genome Analysis Identifies Distinct Sorting Pathways in Gram-Positive Bacteria. *Infect. Immun.* **72**, 2710–2722 (2004).

14. Dramsi, S., Trieu-Cuot, P. & Bierne, H. Sorting sortases: A nomenclature proposal for the various sortases of Gram-positive bacteria. *Research in Microbiology* **156**, 289–297 (2005).

15. Malik, A. & Kim, S. B. A comprehensive in silico analysis of sortase superfamily. *J. Microbiol.* **57**, 431–443 (2019).

16. Spirig, T., Weiner, E. M. & Clubb, R. T. Sortase enzymes in Gram-positive bacteria. *Mol. Microbiol.* **82**, 1044–1059 (2011).

17. Bradshaw, W. J. *et al.* Molecular features of the sortase enzyme family. *FEBS J.* **282**, 2097–2114 (2015).

18. Boekhorst, J., De Been, M. W. H. J., Kleerebezem, M. & Siezen, R. J. Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. *J. Bacteriol.* **187**, 4928–4934 (2005).

19. Scott, J. R. & Zähner, D. Pili with strong attachments: Gram-positive bacteria do it differently. *Mol. Microbiol.* **62**, 320–330 (2006).

20. Mazmanian, S. K., Liu, G., Jensen, E. R., Lenoy, E. & Schneewind, O. Staphylococcus aureus sortase mutants defective in the display of surface proteins and in the

pathogenesis of animal infections. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5510–5515 (2000).

21. Zink, S. D. & Burns, D. L. Importance of srtA and srtB for growth of Bacillus anthracis in macrophages. *Infect. Immun.* **73**, 5222–5228 (2005).

22. Bierne, H. *et al.* Inactivation of the srtA gene in Listeria monocytogenes inhibits anchoring of surface proteins and affects virulence. *Mol. Microbiol.* **43**, 869–881 (2002).

23. Majumdar, D. S. *et al.* Single-molecule FRET reveals sugar-induced conformational dynamics in LacY. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12640–5 (2007).

24. Weiss, W. J. *et al.* Effect of srtA and srtB gene expression on the virulence of Staphylococcus aureus in animal models of infection. *J. Antimicrob. Chemother.* **53**, 480–486 (2004).

25. Raz, A. *et al.* Streptococcus pyogenes sortase mutants are highly susceptible to killing by host factors due to aberrant envelope physiology. *PLoS One* **10**, e0140784 (2015).

26. Amer, B. R. & Clubb, R. T. A sweet new role for LCP enzymes in protein glycosylation. *Mol. Microbiol.* **94**, 1197–1200 (2014).

27. Clancy, K. W., Melvin, J. A. & McCafferty, D. G. Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. *Biopolymers* **94**, 385–396 (2010).

28. Cascioferro, S., Totsika, M. & Schillaci, D. Sortase A: An ideal target for anti-virulence drug development. *Microb. Pathog.* **77**, 105–112 (2014).

29. Hendrickx, A. P. A., Budzik, J. M., Oh, S.-Y. & Schneewind, O. Architects at the bacterial surface — sortases and the assembly of pili with isopeptide bonds. *Nat. Rev. Microbiol.* **9**, 166–176 (2011).

30. 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-.).* **285**, 760–763 (1999).

31. 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.* **96**, 12424–12429 (1999).

32. Emr, S. D., Hedgpeth, J., Clément, J. M., Silhavy, T. J. & Hofnung, M. Sequence analysis of mutations that prevent export of λ receptor, an Escherichia coli outer membrane protein. *Nature* **285**, 82–85 (1980).

33. 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* **8**, (2017).

34. Kline, K. A. *et al.* Mechanism for sortase localization and the role of sortase localization in efficient pilus assembly in Enterococcus faecalis. *J. Bacteriol.* **191**, 3237–3247 (2009).

35. Matias, V. R. F. & Beveridge, T. J. Cryo-electron microscopy reveals native polymeric cell wall structure in Bacillus subtilis 168 and the existence of a periplasmic space. *Mol. Microbiol.* **56**, 240–251 (2005).

36. Matias, V. R. F. & Beveridge, T. J. Native cell wall organization shown by cryo-electron microscopy confirms the existence of a periplasmic space in Staphylococcus aureus. *J. Bacteriol.* **188**, 1011–1021 (2006).

37. Suree, N. *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).

38. Chan, A. H. *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).

39. 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.* **98**, 6056–6061 (2001).

40. Jacobitz, A. W. *et al.* Structural and computational studies of the Staphylococcus aureus sortase B-substrate complex reveal a substrate-stabilized oxyanion hole. *J. Biol. Chem.* **289**, 8891–902 (2014).

41. Das, S. *et al.* Structure and specificity of a new class of Ca2+-independent housekeeping sortase from Streptomyces avermitilis provide insights into its non-canonical substrate preference. *J. Biol. Chem.* **292**, 7244–7257 (2017).

42. Robson, S. A., Jacobitz, A. W., Phillips, M. L. & Clubb, R. T. Solution structure of the sortase required for efficient production of infectious bacillus anthracis spores. *Biochemistry* **51**, 7953–7963 (2012).

43. El Mortaji, L. *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.* **841**, 833–841 (2011).

44. Manzano, C. *et al.* Sortase-Mediated Pilus Fiber Biogenesis in Streptococcus pneumoniae. *Structure* **16**, 1838–1848 (2008).

45. Cozzi, R. *et al.* Structure analysis and site-directed mutagenesis of defined key residues and motives for pilus-related sortase C1 in group B Streptococcus. *FASEB J.* **25**, 1874–1886 (2011).

46. Khare, B. *et al.* Structural differences between the Streptococcus agalactiae housekeeping and pilus-specific sortases: SrtA and SrtC1. *PLoS One* **6**, (2011).

47. Naik, M. T. *et al.* Staphylococcus aureus sortase a transpeptidase: Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop. *J. Biol. Chem.* **281**, 1817–1826 (2006).

48. Race, P. R. *et al.* Crystal Structure of streptococcus pyogenes Sortase A: Implications for sortase mechanism. *J. Biol. Chem.* **284**, 6924–6933 (2009).

49. 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**, 11188–11200 (2005).

50. 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* **46**, 7269–7278 (2007).

51. Danne, C. & Dramsi, S. Pili of Gram-positive bacteria: roles in host colonization. *Res. Microbiol.* **163**, 645–658 (2012).

52. Telford, J. L., Barocchi, M. A., Margarit, I., Rappuoli, R. & Grandi, G. Pili in Gram-positive pathogens. *Nat. Rev. Microbiol.* **4**, 509–519 (2006).

53. Mandlik, A., Swierczynski, A., Das, A. & Ton-That, H. Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends Microbiol.* **16**, 33–40 (2008).

54. Cegelski, L., Smith, C. L. & Hultgren, S. J. Microbial adhesion. in *Encyclopedia of Microbiology* 93–102 (2019). doi:10.1016/B978-0-12-801238-3.02317-5

55. Proft, T. & Baker, E. N. Pili in Gram-negative and Gram-positive bacteria - Structure, assembly and their role in disease. *Cell. Mol. Life Sci.* **66**, 613–635 (2009).

56. Fronzes, R., Remaut, H. & Waksman, G. Architectures and biogenesis of non-flagellar protein appendages in Gram-negative bacteria. *EMBO J.* **27**, 2271–2280 (2008).

57. 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.* **18**, 224–232 (2010).

58. Mulvey, M. A. *et al.* Induction and evasion of host defenses by type 1-piliated uropathogenic Escherichia coli. *Science (80-.).* **282**, 1494–1497 (1998).

59. Ton-That, H. & Schneewind, O. Assembly of pili in Gram-positive bacteria. *Trends Microbiol.* **12**, 228–234 (2004).

60. Barocchi, M. A. *et al.* A pneumococcal pilus influences virulence and host inflammatory responses. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2857–2862 (2006).

61. Maione, D. *et al.* Identification of a universal group B Streptococcus vaccine by multiple genome screen. *Science (80-.).* **309**, 148–150 (2005).

62. Maldarelli, G. A., De Masi, L., von Rosenvinge, E. C., Carter, M. & Donnenberg, M. S. Identification, immunogenicity, and cross-reactivity of type IV pilin and pilin-like proteins from Clostridium difficile. *Pathog. Dis.* **71**, 302–314 (2014).

63. Bermúdez-Humarán, L. G. *et al.* An inducible surface presentation system improves cellular immunity against human papillomavirus type 16 E7 antigen in mice after nasal administration with recombinant lactococci. *J. Med. Microbiol.* **53**, 427–433 (2004).

64. Wagachchi, D. *et al.* PilVax - A novel peptide delivery platform for the development of mucosal vaccines. *Sci. Rep.* **8**, 1–12 (2018).

65. Muñoz-Elías, E. J., Marcano, J. & Camilli, A. Isolation of Streptococcus pneumoniae biofilm mutants and their characterization during nasopharyngeal colonization. *Infect. Immun.* **76**, 5049–5061 (2008).

66. Allegrucci, M. *et al.* Phenotypic characterization of streptococcus pneumoniae biofilm development. *J. Bacteriol.* **188**, 2325–2335 (2006).

67. Reardon-Robinson, M. E. *et al.* Pilus hijacking by a bacterial coaggregation factor critical for oral biofilm development. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3835–3840 (2014).

68. Becherelli, M. *et al.* The ancillary protein 1 of Streptococcus pyogenes FCT-1 pili mediates cell adhesion and biofilm formation through heterophilic as well as homophilic interactions. *Mol. Microbiol.* **83**, 1035–1047 (2012).

69. Dramé, I. *et al.* Analysis of Homotypic Interactions of Lactococcus lactis Pili Using Single-Cell Force Spectroscopy. *ACS Appl. Mater. Interfaces* **12**, 21411–21423 (2020).

70. Izoré, T. *et al.* Structural Basis of Host Cell Recognition by the Pilus Adhesin from Streptococcus pneumoniae. *Structure* **18**, 106–115 (2010).

71. Krishnan, V. *et al.* Structure of Streptococcus agalactiae tip pilin GBS104: A model for GBS pili assembly and host interactions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1073–1089 (2013).

72. Konto-Ghiorghi, Y. *et al.* Dual role for pilus in adherence to epithelial cells and biofilm formation in Streptococcus agalactiae. *PLoS Pathog.* **5**, (2009).

73. Vitry, P. *et al.* Force-induced strengthening of the interaction between Staphylococcus aureus clumping factor B and loricrin. *MBio* **8**, (2017).

74. Becke, T. D. *et al.* Pilus-1 Backbone Protein RrgB of Streptococcus pneumoniae Binds Collagen i in a Force-Dependent Way. *ACS Nano* **13**, (2019).

75. Herman-Bausier, P. *et al.* Staphylococcus aureus clumping factor A is a force-sensitive molecular switch that activates bacterial adhesion. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 5564–5569 (2018).

76. Thomas, W. E., Vogel, V. & Sokurenko, E. Biophysics of Catch Bonds. *Annu. Rev. Biophys.* **37**, 399–416 (2008).

77. Linke-Winnebeck, C. *et al.* Structural model for covalent adhesion of the Streptococcus pyogenes pilus through a thioester bond. *J. Biol. Chem.* **289**, 177–189 (2014).

78. Walden, M., Crow, A., Nelson, M. D. & Banfield, M. J. Intramolecular isopeptide but not internal thioester bonds confer proteolytic and significant thermal stability to the S. pyogenes pilus adhesin Spy0125. *Proteins Struct. Funct. Bioinforma.* **82**, 517–527 (2014).

79. Pointon, J. A. *et al.* A highly unusual thioester bond in a pilus adhesin is required for efficient host cell interaction. *J. Biol. Chem.* **285**, 33858–33866 (2010).

80. Miller, O. K., Banfield, M. J. & Schwarz-Linek, U. A new structural class of bacterial thioester domains reveals a slipknot topology. *Protein Sci.* **27**, 1651–1660 (2018).

81. Law, S. K. A. & Dodds, A. W. The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci.* **6**, 263–274 (2008).

82. Alonso-Caballero, A. *et al.* Protein folding modulates the adhesion strategy of Gram positive pathogens. *bioRxiv* 743393 (2019). doi:10.1101/743393

83. Nakata, M. *et al.* Mode of expression and functional characterization of FCT-3 pilus region-encoded proteins in Streptococcus pyogenes serotype M49. *Infect. Immun.* **77**, 32–44 (2009).

84. de Angelis, G. *et al.* The streptococcus pneumoniae Pilus-1 displays a biphasic

expression pattern. *PLoS One* **6**, (2011).

85. Rogers, E. A., Das, A. & Ton-That, H. Adhesion by pathogenic corynebacteria. *Adv. Exp. Med. Biol.* **715**, 91–103 (2011).

86. Mandlik, A., Swierczynski, A., Das, A. & Ton-That, H. Corynebacterium diphtheriae employs specific minor pilins to target human pharyngeal epithelial cells. *Mol. Microbiol.* **64**, 111–124 (2007).

87. Guttilla, I. K. *et al.* Acyl Enzyme Intermediates in Sortase-Catalyzed Pilus Morphogenesis in Gram-Positive Bacteria. *J. Bacteriol.* **191**, 5603–5612 (2009).

88. Swierczynski, A. & Ton-That, H. Type III pilus of corynebacteria: Pilus length is determined by the level of its major pilin subunit. *J. Bacteriol.* **188**, 6318–6325 (2006).

89. Chang, C. *et al.* Cell-to-cell interaction requires optimal positioning of a pilus tip adhesin modulated by gram-positive transpeptidase enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 18041–18049 (2019).

90. 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.* **105**, 14147–52 (2008).

91. 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**, E5477–E5486 (2018).

92. Jacobitz, A. W. *et al.* 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**, 8302–8312 (2016).

93. Manzano, C., Izoré, T., Job, V., Di Guilmi, A. M. & Dessen, A. Sortase activity is

controlled by a flexible lid in the pilus biogenesis mechanism of Gram-positive pathogens. *Biochemistry* **48**, 10549–10557 (2009).

94. Cozzi, R. *et al.* Structure analysis and site-directed mutagenesis of defined key residues and motives for pilus-related sortase C1 in group B Streptococcus . *FASEB J.* **25**, 1874–1886 (2011).

95. Khare, B., Fu, Z. Q., Huang, I. H., Hung, T. T. & Narayana, S. V. L. The crystal structure analysis of group B streptococcus sortase C1: A model for the 'lid' movement upon substrate binding. *J. Mol. Biol.* **414**, 563–577 (2011).

96. Persson, K. Structure of the sortase AcSrtC-1 from Actinomyces oris. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 212–217 (2011).

97. Mora, M. *et al.* Group A Streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15641–15646 (2005).

98. Budzik, J. M., Marraffini, L. A. & Schneewind, O. Assembly of pili on the surface of Bacillus cereus vegetative cells. *Mol. Microbiol.* **66**, 495–510 (2007).

99. Mishra, A., Das, A., Cisar, J. O. & Ton-That, H. Sortase-catalyzed assembly of distinct heteromeric fimbriae in Actinomyces naeslundii. *J. Bacteriol.* **189**, 3156–3165 (2007).

100. Gaspar, A. H. & Ton-That, H. Assembly of distinct pilus structures on the surface of Corynebacterium diphtheriae. *J. Bacteriol.* **188**, 1526–1533 (2006).

101. Dramsi, S. *et al.* Assembly and role of pili in group B streptococci. *Mol. Microbiol.* **60**, 1401–1413 (2006).

102. Lemieux, J., Woody, S. & Camilli, A. Roles of the sortases of Streptococcus pneumoniae in assembly of the RlrA pilus. *J. Bacteriol.* **190**, 6002–6013 (2008).

103. Gibbons, R. J. & Hay, D. I. Human salivary acidic proline-rich proteins and statherin promote the attachment of Actinomyces viscosus LY7 to apatitic surfaces. *Infect. Immun.* **56**, 439–445 (1988).

104. Sandberg, A. L. *et al.* Putative glycoprotein and glycolipid polymorphonuclear leukocyte receptors for the Actinomyces naeslundii WVU45 fimbrial lectin. *Infect. Immun.* **63**, (1995).

105. Yeung, M. K. Molecular and genetic analyses of Actinomyces spp. *Crit. Rev. Oral Biol. Med.* **10**, 120–138 (1999).

106. Cornelis, G. R. The type III secretion injectisome. *Nature Reviews Microbiology* **4**, 811–825 (2006).

107. Knowles, M. R. & Boucher, R. C. Mucus clearance as a primary innate defense mechanism for mammalian airways. *J. Clin. Invest.* **109**, 571–577 (2002).

108. Cone, R. A. Barrier properties of mucus. *Advanced Drug Delivery Reviews* **61**, 75–85 (2009).

109. Zakrisson, J., Wiklund, K., Axner, O. & Andersson, M. Helix-like biopolymers can act as dampers of force for bacteria in flows. *Eur. Biophys. J.* **41**, 551–560 (2012).

110. Walden, M. *et al.* An internal thioester in a pathogen surface protein mediates covalent host binding. *Elife* **4**, 1–24 (2015).

111. Ton-That, H. & Schneewind, O. Assembly of pili on the surface of Corynebacterium diphtheriae. *Mol. Microbiol.* **50**, 1429–1438 (2003).

112. Echelman, D. J. *et al.* CnaA domains in bacterial pili are efficient dissipaters of large mechanical shocks. *Proc. Natl. Acad. Sci.* **113**, 2490–2495 (2016).

113. Kang, H. J., Coulibaly, F., Clow, F., Proft, T. & Baker, E. N. Stabilizing Isopeptide Bonds Revealed in Gram-Positive Bacterial Pilus Structure. *Science (80-.).* **318**, 1625–1628 (2007).

114. Kwon, H., Young, P. G., Squire, C. J. & Baker, E. N. Engineering a Lys-Asn isopeptide bond into an immunoglobulin-like protein domain enhances its stability. *Sci. Rep.* **7**, 42753 (2017).

115. Alegre-Cebollada, J., Badilla, C. L. & Fernández, J. M. Isopeptide bonds block the mechanical extension of pili in pathogenic Streptococcus pyogenes. *J. Biol. Chem.* **285**, 11235–11242 (2010).

116. Abolbashari, M. H. & Ameli, S. Mechanical unfolding of titin I27 domain: Nanoscale simulation of mechanical properties based on virial theorem via steered molecular dynamics technique. *Sci. Iran.* **19**, 1526–1533 (2012).

117. Eckels, E. C., Tapia-Rojo, R., Rivas-Pardo, J. A. & Fernández, J. M. The Work of Titin Protein Folding as a Major Driver in Muscle Contraction. *Annu. Rev. Physiol.* **80**, 327–351 (2018).

118. Kang, H. J., 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. *Proc. Natl. Acad. Sci.* **106**, 16967–16971 (2009).

119. Budzik, J. M. *et al.* Intramolecular amide bonds stabilize pili on the surface of bacilli. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19992–19997 (2009).

120. Cozzi, R. *et al.* Structure and assembly of group B streptococcus pilus 2b backbone protein. *PLoS One* **10**, (2015).

121. Kang, H. J. *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. Sect. D Biol. Crystallogr.* **70**, 1190–1201 (2014).

122. 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**, 1205–1220 (2011).

123. Spraggon, G. *et al.* Supramolecular Organization of the Repetitive Backbone Unit of the Streptococcus pneumoniae Pilus. *PLoS One* **5**, e10919 (2010).

124. Vengadesan, K., Ma, X., Dwivedi, P., Ton-That, H. & Narayana, S. V. L. A model for group B Streptococcus pilus type 1: The structure of a 35-kDa C-terminal fragment of the major pilin GBS80. *J. Mol. Biol.* **407**, 731–743 (2011).

125. Nuccitelli, A. *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**, (2011).

126. Chaurasia, P., Pratap, S., Palva, A., von Ossowski, I. & Krishnan, V. Bent conformation of a backbone pilin N-terminal domain supports a three-stage pilus assembly mechanism. *Commun. Biol.* **1**, (2018).

127. Hendrickx, A. P. A. *et al.* Isopeptide bonds of the major pilin protein BcpA influence pilus structure and bundle formation on the surface of Bacillus cereus. *Mol. Microbiol.* **85**, 152–163 (2012).

128. Krall, N., Da Cruz, F. P., Boutureira, O. & Bernardes, G. J. L. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **8**, 103–113 (2016).

129. Kim, Y. *et al.* Efficient site-specific labeling of proteins via cysteines. *Bioconjug. Chem.* **19**, 786–791 (2008).

130. Chen, X. & Wu, Y. W. Selective chemical labeling of proteins. *Organic and Biomolecular Chemistry* **14**, 5417–5439 (2016).

131. McKay, C. S. & Finn, M. G. Click chemistry in complex mixtures: Bioorthogonal bioconjugation. *Chemistry and Biology* **21**, 1075–1101 (2014).

132. Lang, K. & Chin, J. W. Bioorthogonal reactions for labeling proteins. *ACS Chem. Biol.* **9**, 16–20 (2014).

133. Rashidian, M., Song, J. M., Pricer, R. E. & Distefano, M. D. Chemoenzymatic reversible immobilization and labeling of proteins without prior purification. *J. Am. Chem. Soc.* **134**, 8455–8467 (2012).

134. Kulkarni, C., Kinzer-Ursem, T. L. & Tirrell, D. A. Selective functionalization of the protein N terminus with N-myristoyl transferase for bioconjugation in cell lysate. *ChemBioChem* **14**, 1958–1962 (2013).

135. Chapman-Smith, A., Mulhern, T. D., Whelan, F., Cronan, J. E. & Wallace, J. C. The C-terminal domain of biotin protein ligase from E. coli is required for catalytic activity. *Protein Sci.* **10**, 2608–2617 (2008).

136. Mao, H., Hart, S. A., Schink, A. & Pollok, B. A. Sortase-Mediated Protein Ligation: A New Method for Protein Engineering. *J. Am. Chem. Soc.* **126**, 2670–2671 (2004).

137. Policarpo, R. L. *et al.* Flow-Based Enzymatic Ligation by Sortase A. *Angew. Chemie Int. Ed.* **53**, 9203–9208 (2014).

138. Popp, M. W., Dougan, S. K., Chuang, T. Y., Spooner, E. & Ploegh, H. L. Sortase-catalyzed transformations that improve the properties of cytokines. *Proc. Natl. Acad. Sci.*

U. S. A. **108**, 3169–3174 (2011).

139. Parthasarathy, R., Subramanian, S. & Boder, E. T. Sortase A as a novel molecular 'stapler' for sequence-specific protein conjugation. *Bioconjug. Chem.* **18**, 469–476 (2007).

140. Rashidian, M., Dozier, J. K. & Distefano, M. D. Enzymatic labeling of proteins: Techniques and approaches. *Bioconjug. Chem.* **24**, 1277–1294 (2013).

141. Chan, L. *et al.* Covalent Attachment of Proteins to Solid Supports and Surfaces via Sortase-Mediated Ligation. *PLoS One* **2**, e1164 (2007).

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

143. Kobashigawa, Y., Kumeta, H., Ogura, K. & Inagaki, F. Attachment of an NMR-invisible solubility enhancement tag using a sortase-mediated protein ligation method. *J. Biomol. NMR* **43**, 145–150 (2009).

144. Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E. & Ploegh, H. L. Sortagging: A versatile method for protein labeling. *Nat. Chem. Biol.* **3**, 707–708 (2007).

145. Wu, Q., Ploegh, H. L. & Truttmann, M. C. Hepta-Mutant Staphylococcus aureus Sortase A (SrtA7m) as a Tool for in Vivo Protein Labeling in Caenorhabditis elegans. *ACS Chem. Biol.* **12**, 664–673 (2017).

146. Warden-Rothman, R., Caturegli, I., Popik, V. & Tsourkas, A. Sortase-tag expressed protein ligation: Combining protein purification and site-specific bioconjugation into a single step. *Anal. Chem.* **85**, 11090–11097 (2013).

147. Wang, H. H., Altun, B., Nwe, K. & Tsourkas, A. Proximity-Based Sortase-Mediated Ligation. *Angew. Chemie* **129**, 5433–5436 (2017).

148. Pritz, S. *et al.* Synthesis of biologically active peptide nucleic acid-peptide conjugates by sortase-mediated ligation. *J. Org. Chem.* **72**, 3909–3912 (2007).

149. Freiburger, L. *et al.* Efficient segmental isotope labeling of multi-domain proteins using Sortase A. *J. Biomol. NMR* **63**, 1–8 (2015).

150. Schneewind, O., Model, P. & Fischetti, V. A. Sorting of protein a to the staphylococcal cell wall. *Cell* **70**, 267–281 (1992).

151. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci.* **108**, 11399–11404 (2011).

152. Chen, L. *et al.* Improved variants of SrtA for site-specific conjugation on antibodies and proteins with high efficiency. *Sci. Rep.* **6**, 1–12 (2016).

153. Dorr, B. M., Ham, H. O., An, C., Chaikof, E. L. & Liu, D. R. Reprogramming the specificity of sortase enzymes. *Proc. Natl. Acad. Sci.* **111**, 13343–13348 (2014).

154. Antos, J. M. *et al.* Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J. Am. Chem. Soc.* **131**, 10800–10801 (2009).

155. Nikghalb, K. D. *et al.* Expanding the Scope of Sortase-Mediated Ligations by Using Sortase Homologues. *ChemBioChem* **19**, 185–195 (2018).

156. Schmohl, L. *et al.* Identification of sortase substrates by specificity profiling. *Bioorganic Med. Chem.* **25**, 5002–5007 (2017).

157. Bellucci, J. J., Bhattacharyya, J. & Chilkoti, A. A noncanonical function of sortase enables site-specific conjugation of small molecules to lysine residues in proteins. *Angew. Chemie - Int. Ed.* **54**, 441–445 (2015).

158. Dasgupta, S., Samantaray, S., Sahal, D. & Roy, R. P. Isopeptide ligation catalyzed by

quintessential sortase A: Mechanistic cues from cyclic and branched oligomers of indolicidin. *J. Biol. Chem.* **286**, 23996–24006 (2011).

159. Möhlmann, S., Mahlert, C., Greven, S., Scholz, P. & Harrenga, A. In vitro Sortagging of an Antibody Fab Fragment: Overcoming Unproductive Reactions of Sortase with Water and Lysine Side Chains. *ChemBioChem* **12**, 1774–1780 (2011).

160. Yokoyama, K., Nio, N. & Kikuchi, Y. Properties and applications of microbial transglutaminase. *Appl. Microbiol. Biotechnol.* **64**, 447–454 (2004).

161. Folk, J. E. Mechanism and basis for specificity of transglutaminase-catalyzed ε-(γ-glutamyl) lysine bond formation. *Adv. Enzymol. Relat. Areas Mol. Biol.* **54**, 1–56 (1983).

162. Veggiani, G., Zakeri, B. & Howarth, M. Superglue from bacteria: Unbreakable bridges for protein nanotechnology. *Trends Biotechnol.* **32**, 506–512 (2014).

163. Fierer, J. O., Veggiani, G. & Howarth, M. SpyLigase peptide-peptide ligation polymerizes affibodies to enhance magnetic cancer cell capture. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E1176 (2014).

164. Buldun, C. M., Jean, J. X., Bedford, M. R. & Howarth, M. SnoopLigase Catalyzes Peptide-Peptide Locking and Enables Solid-Phase Conjugate Isolation. *J. Am. Chem. Soc.* **140**, 3008–3018 (2018).

165. McConnell, S. A. S. A. *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).

166. 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).

167. Zhao, X., Zhang, L. & Liu, D. Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. *Biofuels, Bioprod. Biorefining* **6**, 465–482 (2012).

168. Boerjan, W., Ralph, J. & Baucher, M. Lignin Biosynthesis. *Annu. Rev. Plant Biol.* **54**, 519–546 (2003).

169. Bugg, T. D. H., Ahmad, M., Hardiman, E. M. & Rahmanpour, R. Pathways for degradation of lignin in bacteria and fungi. *Natural Product Reports* **28**, 1883–1896 (2011).

170. Wyman, C. E. *et al.* Coordinated development of leading biomass pretreatment technologies. *Bioresour. Technol.* **96**, 1959–1966 (2005).

171. Ghosh, P. & Singh, A. Physicochemical and Biological Treatments for Enzymatic/Microbial Conversion of Lignocellulosic Biomass. *Adv. Appl. Microbiol.* **39**, 295–333 (1993).

172. Scheller, H. V. & Ulvskov, P. Hemicelluloses. *Annu. Rev. Plant Biol.* **61**, 263–289 (2010).

173. Sethi, A. & Scharf, M. E. Biofuels: Fungal, Bacterial and Insect Degraders of Lignocellulose. in *eLS* (John Wiley & Sons, Ltd, 2013). doi:10.1002/9780470015902.a0020374

174. Zhao, X., Zhang, L. & Liu, D. Biomass recalcitrance. Part II: Fundamentals of different pre-treatments to increase the enzymatic digestibility of lignocellulose. *Biofuels, Bioproducts and Biorefining* **6**, 561–579 (2012).

175. Medie, F. M., Davies, G. J., Drancourt, M. & Henrissat, B. Genome analyses highlight the different biological roles of cellulases. *Nature Reviews Microbiology* **10**, 227–234 (2012).

176. Yang, B., Dai, Z., Ding, S. Y. & Wyman, C. E. Enzymatic hydrolysis of cellulosic biomass.

Biofuels **2**, 421–450 (2011).

177. Ghose, T. K. Cellulase biosynthesis and hydrolysis of cellulosic substances. in 39–76 (Springer, Berlin, Heidelberg, 1977). doi:10.1007/3-540-08363-4_2

178. Bayer, E. A., Morag, E. & Lamed, R. The cellulosome - A treasure-trove for biotechnology. *Trends in Biotechnology* **12**, 379–386 (1994).

179. Bayer, E. A., Shimon, L. J. W., Shoham, Y. & Lamed, R. Cellulosomes - Structure and ultrastructure. *J. Struct. Biol.* **124**, 221–234 (1998).

180. Fontes, C. M. G. A. & Gilbert, H. J. Cellulosomes: Highly Efficient Nanomachines Designed to Deconstruct Plant Cell Wall Complex Carbohydrates. *Annu. Rev. Biochem.* **79**, 655–681 (2010).

181. Mayer, F., Coughlan, M. P., Mori, Y. & Ljungdahl, L. G. Macromolecular Organization of the Cellulolytic Enzyme Complex of Clostridium thermocellum as Revealed by Electron Microscopy. *Appl. Environ. Microbiol.* **53**, (1987).

182. Cunha, E. S., Hatem, C. L. & Barrick, D. Synergistic enhancement of cellulase pairs linked by consensus ankyrin repeats: Determination of the roles of spacing, orientation, and enzyme identity. *Proteins Struct. Funct. Bioinforma.* **84**, 1043–1054 (2016).

183. Barth, A. *et al.* Dynamic interactions of type I cohesin modules fine-tune the structure of the cellulosome of Clostridium thermocellum. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E11274–E11283 (2018).

184. You, C., Zhang, X. Z., Sathitsuksanoh, N., Lynd, L. R. & Percival Zhang, Y. H. Enhanced microbial utilization of recalcitrant cellulose by an ex vivo cellulosome-microbe complex. *Appl. Environ. Microbiol.* **78**, 1437–1444 (2012).

185. Demain, A. L., Newcomb, M. & Wu, J. H. D. Cellulase, Clostridia, and Ethanol. *Microbiol.*

Mol. Biol. Rev. **69**, 124–154 (2005).

186. Lu, Y., Zhang, Y. H. P. & Lynd, L. R. Enzyme-microbe synergy during cellulose hydrolysis by Clostridium thermocellum. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16165–16169 (2006).

187. Olson, D. G., McBride, J. E., Joe Shaw, A. & Lynd, L. R. Recent progress in consolidated bioprocessing. *Current Opinion in Biotechnology* **23**, 396–405 (2012).

188. Tsai, S. L., Oh, J., Singh, S., Chen, R. & Chen, W. Functional assembly of minicellulosomes on the Saccharomyces cerevisiae cell surface for cellulose hydrolysis and ethanol production. *Appl. Environ. Microbiol.* **75**, 6087–6093 (2009).

189. Mitsuzawa, S. *et al.* The rosettazyme: A synthetic cellulosome. *J. Biotechnol.* **143**, 139–144 (2009).

190. Wen, F., Sun, J. & Zhao, H. Yeast surface display of trifunctional minicellulosomes for simultaneous Saccharification and fermentation of cellulose to ethanol. *Appl. Environ. Microbiol.* **76**, 1251–1260 (2010).

191. Fierobe, H.-P. *et al.* Degradation of Cellulose Substrates by Cellulosome Chimeras. *J. Biol. Chem.* **277**, 49621–49630 (2002).

192. Fierobe, H. P. *et al.* Action of designer cellulosomes on homogeneous Versus complex substrates: Controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin. *J. Biol. Chem.* **280**, 16325–16334 (2005).

193. Hirano, K., Nihei, S., Hasegawa, H., Haruki, M. & Hirano, N. Stoichiometric assembly of the cellulosome generates maximum synergy for the degradation of crystalline cellulose, as revealed by in vitro reconstitution of the Clostridium thermocellum cellulosome. *Appl. Environ. Microbiol.* **81**, 4756–4766 (2015).

194. Krauss, J., Zverlov, V. V. & Schwarz, W. H. In Vitro Reconstitution of the Complete

Clostridium thermocellum Cellulosome and Synergistic Activity on Crystalline Cellulose. *Appl. Environ. Microbiol.* **78**, 4301–4307 (2012).

195. Currie, D. H. *et al.* Functional heterologous expression of an engineered full length CipA from Clostridium thermocellum in Thermoanaerobacterium saccharolyticum. *Biotechnol. Biofuels* **6**, 32 (2013).

196. Blanchette, C., Lacayo, C. I., Fischer, N. O., Hwang, M. & Thelen, M. P. Enhanced cellulose degradation using cellulase-nanosphere complexes. *PLoS One* **7**, e42116 (2012).

197. King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103–108 (2014).

Chapter 2

In vitro reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking

2.1 Overview

Corynebacterium diphtheriae is an important Gram-positive pathogen which causes human infections in the upper respiratory tract. Adherence to human tissues and biofilm is mediated through extended proteinaceous polymers called pili, or fimbrae. These fibers are constructed through head-to-tail isopeptide linkages by pilin polymerizing sortases. Because of the important roles of pili in adhesion and biofilm formation, these structures are of intense interest, but complete elucidation of their mechanism of action is challenging because this reaction is difficult to reproduce biochemically. In this chapter, I will discuss our efforts toward reconstituting this system *in vitro*. I will discuss the biochemical, structural, and *in vivo* studies which broadened our understanding of pilus biogenesis and provide a useful system for further inquiry into the mechanism of this process (led by Drs. Brendan R. Amer and Chungyu Chang). My contributions to this manuscript included assisting with protein design for the activated enzyme variants and substrate proteins, modeling of the ternary "attack complex" and the structural, biochemical and phylogenetic analysis that led to the identification of a conserved sequence motif on Class C sortases which is implicated in nucleophilic substrate recognition.

This chapter is reprinted with permission from a peer-reviewed article "*In vitro* reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking." 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., Ton-That, H. *Proc. Nat. Acad. Sci. U.S.A.* **115**, E5477-E5486 (2018).

2.2 *In vitro* reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking

In vitro reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking

Chungyu Chang[a,1], Brendan R. Amer[b,c,1], Jerzy Osipiuk[d,e], Scott A. McConnell[b,c], I-Hsiu Huang[f], Van Hsieh[b,c], Janine Fu[b,c], Hong H. Nguyen[b,c], John Muroski[b,c], Erika Flores[a], Rachel R. Ogorzalek Loo[b,c], Joseph A. Loo[b,c], John A. Putkey[g], Andrzej Joachimiak[d,e], Asis Das[h], Robert T. Clubb[b,c,2], and Hung Ton-That[a,2]

[a]Department of Microbiology and Molecular Genetics, University of Texas Health Science Center, Houston, TX 77030; [b]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095; [c]University of California, Los Angeles-US Department of Energy Institute of Genomics and Proteomics, University of California, Los Angeles, CA 90095; [d]Center for Structural Genomics of Infectious Diseases, Argonne National Laboratory, Argonne, IL 60439; [e]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637; [f]Department of Microbiology and Immunology, College of Medicine, National Cheng Kung University, Tainan 701, Taiwan; [g]Department of Biochemistry and Molecular Biology, University of Texas Health Science Center, Houston, TX 77030; and [h]Department of Molecular Biology and Biophysics, University of Connecticut Health Center, Farmington, CT 06030

Edited by Ralph R. Isberg, Howard Hughes Medical Institute and Tufts University School of Medicine, Boston, MA, and approved May 2, 2018 (received for review January 18, 2018)

Covalently cross-linked pilus polymers displayed on the cell surface of Gram-positive bacteria are assembled by class C sortase enzymes. These pilus-specific transpeptidases located on the bacterial membrane catalyze a two-step protein ligation reaction, first cleaving the LPXTG motif of one pilin protomer to form an acyl-enzyme intermediate and then joining the terminal Thr to the nucleophilic Lys residue residing within the pilin motif of another pilin protomer. To date, the determinants of class C enzymes that uniquely enable them to construct pili remain unknown. Here, informed by high-resolution crystal structures of corynebacterial pilus-specific sortase (SrtA) and utilizing a structural variant of the enzyme (SrtA^{2M}), whose catalytic pocket has been unmasked by activating mutations, we successfully reconstituted in vitro polymerization of the cognate major pilin (SpaA). Mass spectrometry, electron microscopy, and biochemical experiments authenticated that SrtA^{2M} synthesizes pilus fibers with correct Lys–Thr isopeptide bonds linking individual pilins via a thioacyl intermediate. Structural modeling of the SpaA–SrtA–SpaA polymerization intermediate depicts SrtA^{2M} sandwiched between the N- and C-terminal domains of SpaA harboring the reactive pilin and LPXTG motifs, respectively. Remarkably, the model uncovered a conserved TP(Y/L)XIN(S/T)H signature sequence following the catalytic Cys, in which the alanine substitutions abrogated cross-linking activity but not cleavage of LPXTG. These insights and our evidence that SrtA^{2M} can terminate pilus polymerization by joining the terminal pilin SpaB to SpaA and catalyze ligation of isolated SpaA domains in vitro provide a facile and versatile platform for protein engineering and bio-conjugation that has major implications for biotechnology.

Corynebacterium diphtheriae | sortase | pilus polymerization | protein ligation | transpeptidation

Adhesive protein polymers, called "pili" or "fimbriae," are expressed on the cell envelope by many Gram-negative and Gram-positive bacteria, and they are critical for bacterial virulence (1). Many types of Gram-negative pili have been reported, including the well-studied retractable type IV, conjugative, and chaperone-assisted pili (2). These pili are formed by distinct pathways (2, 3); however, none of these pili are covalently linked polymers, unlike the sortase-catalyzed pili found in many Gram-positive bacteria, including *Actinomyces oris*, *Enterococcus faecalis*, *Bacillus cereus*, and numerous species of streptococci and lactobacilli (4–6).

One of the well-studied sortase-mediated pilus assembly systems involves *Corynebacterium diphtheriae* (7), the causative agent of pharyngeal diphtheria (8). *C. diphtheriae* produces three distinct pilus types (7, 9, 10), each comprised of a pilus tip adhesin, a pilus shaft made of the major pilin, and a base pilin that is covalently anchored to the cell wall (11). The archetypal SpaA-type pilus, which mediates adherence to the pharyngeal epithelium (12), consists of the tip pilin SpaC, shaft pilin SpaA, and pilus base SpaB (13). A pilus-specific sortase named "SrtA" is required for pilus polymerization (13), performing a repetitive, irreversible transpeptidation reaction that covalently links the pilin subunits via an isopeptide bond (14). Although each Spa pilin harbors a cell wall sorting signal (CWSS), which starts with a conserved LPXTG motif, followed by a stretch of hydrophobic amino acids and a positively charged tail (15), SpaA contains a pilin motif with the Lys residue K190 acting as a nucleophile for the aforementioned transpeptidation reaction (13). According to the current model (16), SrtA cleaves the LPXTG motif of Spa pilins between Thr and Gly, forming acyl-enzyme intermediates

Significance

Gram-positive sortase enzymes represent two broad functional categories—those that cross-link proteins to the cell wall and those that can catalyze this reaction and polymerize proteins to build adhesive pilus fibers. Here we report an in vitro reproduction of a robust pilus polymerization reaction using a variant of a corynebacterial pilus-specific sortase in which the catalytic center is unmasked. By molecular modeling, we uncovered a conserved structural element of pilus-specific sortases critical for protein ligation in vitro and further demonstrated that the activated sortase ligates the isolated domains of the pilin harboring the donor and acceptor motifs for ligation. Besides enabling future molecular studies and antibiotic development, our system provides a powerful platform for bioconjugation and protein engineering.

Author contributions: C.C., B.R.A., J.O., J.A.P., A.J., R.T.C., and H.T.-T. designed research; C.C., B.R.A., J.O., S.A.M., I.-H.H., V.H., J.F., H.H.N., J.M., E.F., R.R.O.L., and J.A.P. performed research; C.C., B.R.A., J.O., J.A.L., J.A.P., A.J., A.D., R.T.C., and H.T.-T. analyzed data; and C.C., B.R.A., J.O, A.D., R.T.C., and H.T.-T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID codes 5K9A and 6BWE).

[1]C.C. and B.R.A. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: rclubb@mbi.ucla.edu or ton-that.hung@uth.tmc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1800954115/-/DCSupplemental.

Published online May 29, 2018.

between the Thr residue and the SrtA catalytic Cys residue. This intermediate is then nucleophilically attacked by the reactive K190 of an incoming SpaA subunit. In pilus biogenesis, the SpaC–SrtA acyl-enzyme intermediate forms first, resulting in the joining of the ε-amine group of K190 to the Thr carbonyl carbon atom in the LPXT of SpaC. Pilus polymerization ensues when additional SpaA protomers are joined progressively to the pilus base by the SrtA enzyme via the same Lys-mediated transpeptidation reaction. Polymerization is terminated with the entry of SpaB into the pilus base (11), which is then anchored to the cell wall by the housekeeping sortase SrtF (17). This cell wall anchoring of pilus polymers is likely similar to that of surface proteins catalyzed by the prototype SrtA enzyme from *Staphylococcus aureus* (18, 19). While most of this biphasic model of pilus assembly in Gram-positive bacteria (6), i.e., pilus polymerization followed by cell wall anchoring, has been validated experimentally, the molecular determinants that make up a pilus-specific sortase and enable the enzyme to join proteins together remain unknown.

The SrtA enzyme is classified as a member of the class C sortase subgroup within the sortase superfamily that has the unique ability to cross-link proteins via Lys–Thr isopeptide bonds (20, 21). Although all sortases share a canonical β-barrel sortase superfamily fold (22, 23), class C enzymes are distinguished by the presence of a conserved N-terminal region that forms a "lid" that covers the active site structurally and functionally (24–26). In *Streptococcus pneumoniae*, X-ray crystallographic evidence originally suggested that the lid region was flexible, possibly modulating substrate binding; however, subsequent studies in solution utilizing NMR showed this region to be relatively rigid in the SrtC1 enzyme (24, 27–29). Mutations of the lid region in *A. oris* SrtC2 or *Streptococcus agalactiae* SrtC1 did not alter the pilus polymerizing activities in vivo (30, 31); nonetheless, the mutations caused enzyme instability and increased hydrolytic activity in *S. agalactiae* SrtC1 (30), supporting a regulatory role for the N-terminal lid. However, the unique structural properties that enable class C sortase enzymes to cross-link proteins have not been identified.

We report here the crystal structures of the *C. diphtheriae* class C sortase SrtA lacking the signal peptide and transmembrane domain (referred to as "SrtA^{WT}") and a mutant of this protein that has substitutions in the lid interface which normally masks

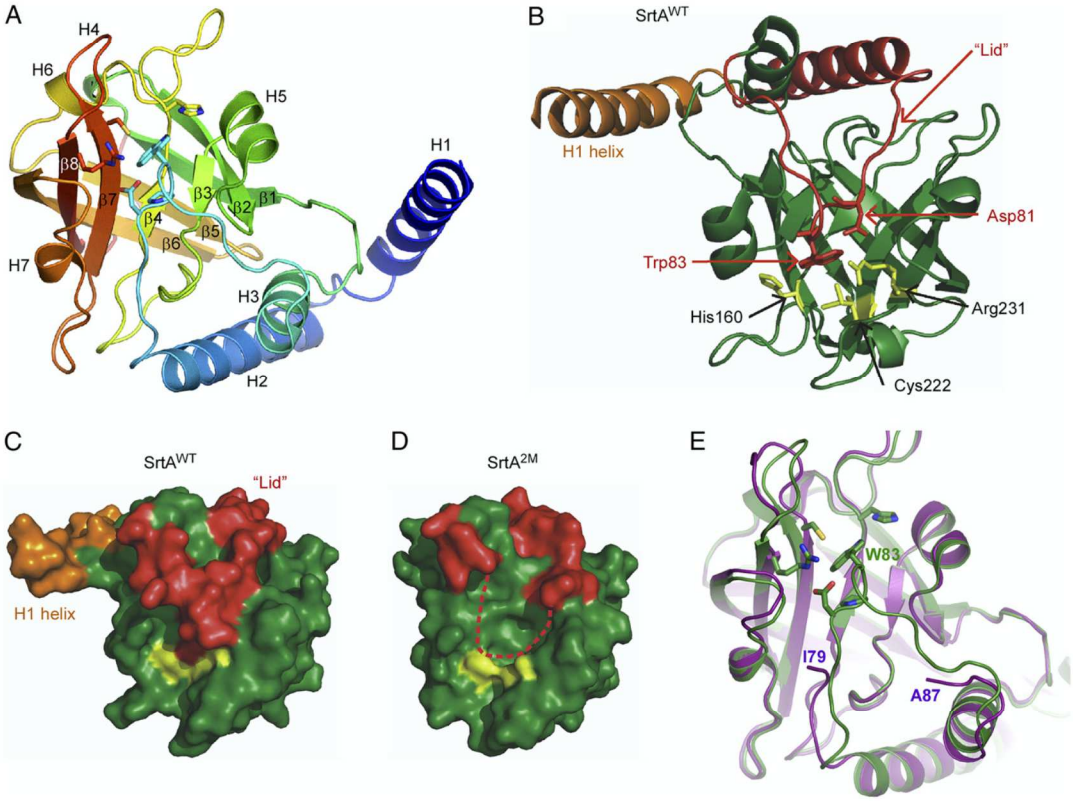


Fig. 1. Structural analysis of the *C. diphtheriae* pilus-specific sortase SrtA. (*A*) The crystal structure of SrtA was determined to 2.1-Å resolution, with the overall protein fold presented as rainbow coloring from blue to red corresponding to the N- to C-terminal positions. The helices are marked as H1–H7, and the β-strands are marked as β1–β8. (*B*) The lid region is marked in red with conserved lid residues D81 and W83 and catalytic residues C222, H160, and R231 in yellow. The C222 residue is shown only in the main conformation. (*C* and *D*) Hydrophobic surface renderings of WT SrtA (SrtA^{WT}) (*C*) and the lid mutant (SrtA^{2M}) structures (*D*) with the H1 helix and lid loop structures in red. The H1 helix of SrtA^{2M} is absent, and its lid structure is not visible, as indicated by a red dashed line. (*E*) Superposition of the SrtA^{WT} (green) and the lid mutant SrtA^{2M} (pink) structures was generated by PyMOL.

Table 1. Crystal data collection statistics

Statistics	$SrtA^{WT}$	$SrtA^{2M}$
X-ray wavelength, Å	0.9792	0.9792
Space group	P 6_122	P 2_1
Unit cell dimensions	a = b = 77.7 Å, c = 202.3 Å, $\alpha = \beta = 90°$, $\gamma = 120°$	a = 65.2 Å, b = 45.5 Å, c = 74.9 Å, $\alpha = \gamma = 90°$, $\beta = 96.4$
Resolution, Å	38.9–2.1 (2.14–2.1)	38.9–1.85 (1.88–1.85)
No. of unique reflections	21,872 (1,050)	36,705 (1,431)
Completeness, %	99.7 (100)	97.7 (77.4)
R-merge	0.142 (0.916)	0.087 (0.410)
CC1/2, $Å^2$	–(0.939)	–(0.741)
I/σ	30.7 (6.4)	16.5 (1.95)
Redundancy	15.4 (15.2)	3.5 (2.2)
Molecules per asymmetric unit	1	2
No. of protein residues	215	430

Numbers in parenthesis are shown for the highest-resolution shell.

the catalytic pocket ($SrtA^{2M}$). Using these recombinant enzymes and an SpaA substrate that lacks the signal peptide and transmembrane domain, we succeeded in reconstituting the SpaA pilus shaft polymerization reaction in vitro, demonstrating that the removal of SrtA's lid not only unmasks the catalytic center structurally but also enables the polymerizing activity in vitro. Subsequently, by structural modeling and phylogenic and mutational analyses, we identified two structural elements that enable SrtA to cross-link proteins. Importantly, we showed that the activated sortase can ligate the isolated pilin domains, thus defining the donor and acceptor motifs for the ligation reaction. The system we report provides a platform for in vitro mechanistic investigations of Gram-positive pilus assembly, antibiotic development, and biotechnological applications of protein modification and conjugation via a unique transpeptidation reaction.

Results and Discussions

Structure of the *C. diphtheriae* Pilus-Specific Sortase. The archetypal SpaA pilus polymer produced by corynebacteria is built by the dedicated pilus-specific sortase SrtA (7, 13). To gain insight into the mechanism of pilus polymerization, we determined the structure of SrtA by X-ray crystallography. We performed crystallization screens using a soluble fragment encompassing the catalytic domain of SrtA (residues 37–257, $SrtA^{WT}$), which was cloned, expressed, and purified from *Escherichia coli*. $SrtA^{WT}$ crystallized as a homodimer in the $P6_1$ 2 2 space group. Diffraction data were collected to 2.1-Å resolution and were phased by molecular replacement (Tables 1 and 2). The electron density for residues 37–248 was well defined, enabling their structure to be modeled, while density for the remaining C-terminal residues is missing, presumably due to a disordered state.

Table 2. Structure refinement statistics

Statistics	$SrtA^{WT}$	$SrtA^{2M}$
Resolution range, Å	38.9–2.1 (2.157–2.1)	38.9–1.85 (1.898–1.85)
Reflections	21,846 (1,533)	36,692 (2,195)
σ cutoff	None	None
R-value, all, %	16.22	17.26
R-value (R-work), %	16.05 (17.4)	17.12 (23.7)
Free R-value, %	19.56 (24.0)	20.08 (23.7)
Rms deviations from ideal geometry		
Bond length, Å	0.017	0.012
Angle, °	1.76	1.60
Chiral, Å	0.101	0.088
No. of atoms		
Protein	1,737	2,974
Sulfate	5	—
Water	167	258
Mean B-factor, $Å^2$		
All atoms	32.0	30.7
Protein atoms	31.2	30.3
Protein main chain	28.6	29.0
Protein side chain	33.9	31.6
Sulfate	66.0	—
Water	38.5	35.9
MolProbity summary		
Ramachandran outliers, %	0.0	0.0
Ramachandran favored, %	97.18	98.89
Rotamer outliers, %	2.02	0.89
C-beta deviations	0.0	0.0
Clash score	1.73	1.34
MolProbity score	1.31	0.86

BIOCHEMISTRY

The overall structure of $SrtA^{WT}$ conforms to the typical sortase fold described previously (22), containing an eight-stranded β-barrel core flanked by several 3_{10} and α-helices (Fig. 1 *A* and *B*). Three additional α-helices are located at the N terminus of $SrtA^{WT}$ (Fig. 1 *A* and *B*) and contain the distinguishing lid structure that occludes the enzyme's active site in a class C sortase (Fig. 1*B*). Interestingly, the H1 helix mediates homodimerization in the crystal structure and is generally removed from the body of the enzyme (Fig. 1 *A* and *B*), while helices H2 and H3 are positioned immediately adjacent to the active site and are connected by a loop that contains the highly conserved DPW lid motif that interacts with the active site (Fig. 1 *A* and *B*). W83 in the lid participates in aromatic stacking interactions with the active site C222 and nearby H160 residues. In addition, D81 within the motif interacts with the active site R231 residue, suggesting its regulatory role in lid positioning and pilin polymerization. Importantly, residues within the catalytic triad His–Cys–Arg are well resolved, and C222 can be modeled in two distinct positions with 50% occupancy, pointing both toward and away from the active site (Fig. 1 *A* and *B*).

To investigate the functional importance of the lid in polymerization, we next generated a recombinant SrtA mutant protein in which the DPW lid motif (residues 81–83) was mutated to GPG, hereafter referred to as "$SrtA^{2M}$." We succeeded in determining the crystal structure of $SrtA^{2M}$ at 1.85-Å resolution using crystallization conditions that differed from those used for the WT protein (*Materials and Methods*). In the electron density map for $SrtA^{2M}$, residues 80–86 that represent the lid were invisible. Presumably, the lid residue substitutions prevented contacts with the active site, causing the mutant lid to adopt a range of conformations. Remarkably, a second major difference between the two structures is the absence of interpretable electron density for the H1 helix in the $SrtA^{2M}$ lid mutant, which might be caused by flexibility around the hinge between helices H1 and H2 and by the absence of stabilizing interactions with neighboring molecules in the crystals of $SrtA^{2M}$.

To evaluate the involvement of the predicted catalytic residues and the lid in pilus assembly, corynebacterial cells harboring WT and its isogenic mutants were subjected to cell fractionation, and protein samples were immunoblotted with specific antibodies against SpaA (α-SpaA), the cognate substrate of SrtA that forms the pilus shaft (7, 13). As shown in Fig. 2*A*, SpaA polymers (P) were observed in cell wall fractions of the WT strain, but they were absent in the *srtA* deletion mutant, as previously reported (11). Ectopic expression of SrtA rescued the pilus assembly defect of the Δ*srtA* mutant (Fig. 2*A*, third lane), and Ala substitution of the catalytic residues C222, H160, and R231 abrogated pilus assembly (Fig. 2*A*, last three lanes). In control experiments, we demonstrated that none of these mutations affected the assembly of the SpaH-type pili, as expected (Fig. 2*B*) (9). Strikingly, the lid mutants are catalytically active in pilus polymerization. Like the WT and the complementing strains, strains expressing mutations in the DPW motif still produced pilus polymers (Fig. 2*C*), and immunoblotting analysis of the membrane fractions revealed no changes in the SrtA protein level when the lid was mutated (Fig. 2*D*).

To visualize these SpaA polymers, corynebacterial cells were immobilized on carbon-coated nickel grids, washed with water, and stained with 0.75% uranyl formate before viewing with an electron microscope. Because the parental strain NCTC 13129 produced short pili that were hardly detected (*SI Appendix*, Fig. S1*A*, WT), we constructed a multicopy vector expressing both SpaA and SrtA. Using this vector as a template, $SrtA^{2M}$ was generated by site-directed mutagenesis (*Materials and Methods*). The generated vectors were introduced into a corynebacterial double mutant lacking both *spaA* and *srtA* (Δ*spaA*/Δ*srtA*). Compared with the WT strain, overexpression of SpaA and SrtA resulted in increased production of long pili, as expected (*SI Appendix*, Fig. S1*A*, Δ*spaA*/Δ*srtA*/pSpaA-SrtA). Consistent with the above results, mutations in the DPW motif did not affect pilus assembly (*SI Appendix*, Fig. S1*A*, Δ*spaA*/Δ*srtA*/pSpaA-$SrtA^{2M}$). To confirm that these long pilus fibers are SpaA pili, the same set of strains was subjected to immunoelectron microscopy (IEM) (32), whereby immobilized cells were stained with α-SpaA, followed by staining with gold particles conjugated with IgG, before washing and staining with uranyl acetate. As shown in *SI Appendix*, Fig. S1*B*, SpaA-stained pili were detected in both strains producing WT SrtA and $SrtA^{2M}$, whereas short pili were observed in the WT strain, and no pili were observed in the Δ*spaA*/Δ*srtA* and Δ*spaA*/Δ*srtA*/pSpaA^{K190A}-$SrtA^{2M}$ strains.

Thus, the overall structure of the *C. diphtheriae* $SrtA^{WT}$ enzyme resembles class C sortases, or pilus-specific sortases, which possess a distinguishing feature of this class of enzymes, the lid region (16, 21). In agreement with previous studies (30, 31), the elimination of the lid's interaction at the catalytic pocket does not dramatically affect pilus assembly in vivo.

In Vitro Reconstitution of Archetypal *C. diphtheriae* SpaA Pilus Polymerization. Previous structural and biochemical studies of pilus-specific sortase enzymes in several streptococcal species indicate that the lid may modulate substrate entry into the active site (24, 25, 28, 30). We envisioned that a loss of lid closure might increase the accessibility of the active site. To test this hypothesis, we used the thiol-reactive reagent 4,4′-dithiodipyridine (DTDP) (33, 34) to probe the solvent accessibility of the catalytic Cys residue (C222). Disulfide exchange between thiol side chains of Cys residues and DTDP gives rise to 4-thiopyridone, which shows strong absorption at 324 nm (34). The recombinant proteins $SrtA^{WT}$ or $SrtA^{2M}$ (0.6 mg/mL) were rapidly mixed with 0.32 mM DTDP, and the rate of reaction between DTDP and C222 was monitored as an increase in absorbance at 324 nm. Time-dependent changes in absorbance were fit to single or double exponential equations to derive rates, as described in *Materials and Methods*. As shown in Fig. 3*A*, data for the $SrtA^{WT}$ enzyme best fit an equation with a single exponential rate of 2.17 ± 0.02/min. In contrast, data for the lid anchor mutant best fit an equation with two much faster exponential rates, 228 ± 7/min and 16 ± 3/min (Fig. 3*A*, *Inset*), which indicates that the catalytic C222 was readily accessible in this mutant. The two different rates may be due to slow exchange between two conformations in the mutant protein. If so, the conformation with the faster rate is the dominant form, since it represents 80% of the total change in absorbance.

The increased DTDP reactivity of the active site Cys residue in $SrtA^{2M}$ described above raises the possibility that the mutant enzyme may be able to assemble pili in vitro, which has been difficult to reconstitute so far for pilus-assembling sortases. We therefore sought to reconstitute pilus polymerization in vitro using various recombinant sortase enzymes and a soluble form of SpaA (residues 30–500), which is devoid of the N-terminal signal peptide and C-terminal membrane anchor domain (see diagram in Fig. 3*E*). Sortases were mixed with SpaA at a 1:3 molar ratio, and aliquots were removed for SDS/PAGE analysis and Coomassie staining at 0, 24, and 48 h. In the $SrtA^{WT}$ samples, a few new high molecular mass (HMM) bands were weakly observed after 24 and 48 h of incubation, one migrating between the 50 and 100 kDa markers and the others around 100 kDa (Fig. 3*B*, lanes $SrtA^{WT}$). Remarkably, with the $SrtA^{2M}$ enzyme, HMM SpaA polymers ($SpaA_P$) were abundantly formed within 24 h and increased further after 48 and 72 h (Fig. 3*B*, lanes $SrtA^{2M}$). Consistent with the results in Fig. 2*A*, the catalytically inactive enzyme in which C222 was replaced by Ala, $SrtA^{C222A}$, failed to produce any SpaA polymers (Fig. 3*B*, lanes $SrtA^{C222A}$). Intriguingly, removal of the H1 helix in the $SrtA^{2M}$ enzyme also abrogated pilus polymerization (Fig. 3*B*, lanes $^{\Delta}SrtA^{2M}$). The significance of this helix in the transpeptidation activity of sortase is discussed below.

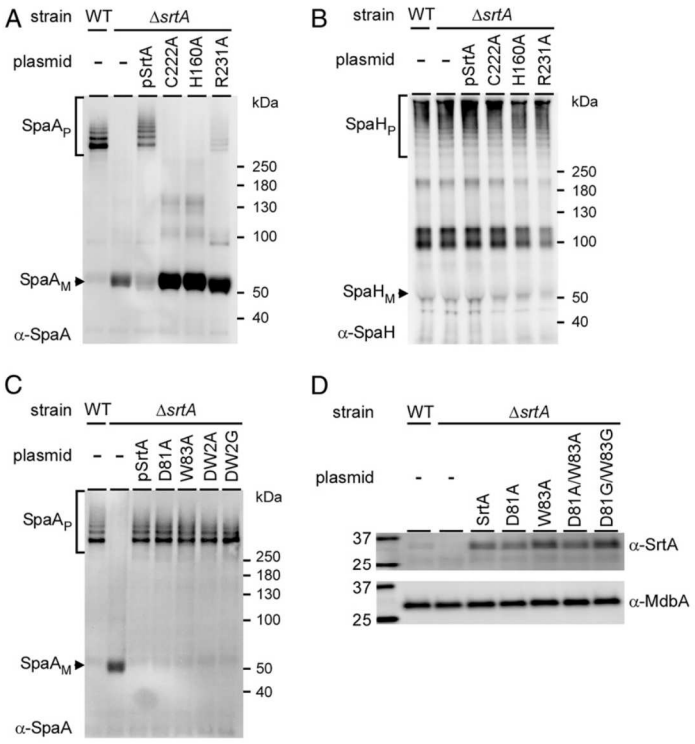


Fig. 2. Catalytic residues are required for pilus assembly in vivo. Cells of *C. diphtheriae* strains in equivalent numbers were subjected to cell fractionation. Protein samples from cell wall fractions (*A*–*C*) and protoplasts (*D*) were analyzed by immunoblotting with specific antibodies against SpaA (α-SpaA) (*A* and *C*), SpaH (α-SpaH) (*B*), and SrtA (α-SrtA) (*D*), with α-MdbA as a membrane-loading control. Pilus monomers (subscript M), polymers (subscript P), and molecular mass markers are indicated.

To visualize HMM SpaA polymers formed by $SrtA^{2M}$, the reaction mixtures after 72 h of incubation were subjected to electron microscopy, whereby aliquots were applied to nickel grids; bound proteins were washed and stained with 0.75% uranyl formate before viewing with an electron microscope. As shown in Fig. 3*C*, strands of SpaA polymers were observed in the reaction with the $SrtA^{2M}$ enzyme but not with an $SpaA^{K190A}$ mutant substrate defective in the nucleophilic attack, thus authenticating the visualization of Gram-positive pilus polymers synthesized in vitro. The synthesized pilus polymers had a width of ∼10 nm and a length ranging from ∼200–500 nm, equivalent to 25–62 subunits, with each protomer measuring about 8 nm (35). For comparison, pili produced by $SrtA^{2M}$ in vivo had widths ranging from 7.6–9.3 nm and lengths up to 2 μm (*SI Appendix*, Fig. S1*A*).

To determine if the recombinant $SrtA^{2M}$ enzyme faithfully catalyzes the pilus transpeptidation reaction, we determined whether the SpaA subunits in the HMM SpaA polymers were linked together via covalent Lys isopeptide bonds in which the Thr residue of the LPLT sorting signal was joined to the Lys residue within the pilin motif (13). Indeed, MS analysis of excised HMM SpaA polymer SDS/PAGE bands revealed the presence of an isopeptide bond between the carbonyl carbon of T494 and the sidechain amine of K190 (Fig. 3*D* and Table 3), as was observed in the native SpaA pili assembled in vivo (36). In line with the role of Lys190 in pilus polymerization, the SpaA mutant substrate in which K190 was replaced by Ala was unable to form polymers with the active $SrtA^{2M}$ enzyme (Fig. 3*B*, lanes $SpaA^{K190A}$). Remarkably, the MS data revealed the presence of $SrtA^{2M}$ and SpaA in the marked band migrating between the 50 and 100 kDa markers in both $SrtA^{WT}$ and $SrtA^{2M}$ samples (Fig. 3*B*, asterisks), suggesting that the enzyme is joined to the SpaA substrate via a labile thioacyl bond forming an acyl-enzyme heterodimer intermediate. To our astonishment, the MS analysis also revealed the presence of $SrtA^{2M}$ in the HMM SpaA polymer bands migrating at and above the 100-kDa marker (Fig. 3*B*, lanes $SrtA^{2M}$, bracket). The results are in agreement with our previous identification of the native acyl-enzyme intermediates formed between SrtA and SpaA polymers in vivo in *C. diphtheriae* as demonstrated by immunoblotting (37).

To further probe the mechanism of SpaA pilus assembly, we dissected the SpaA molecule into two components: the N-terminal domain (^{N}SpaA, residues 30–194, encompassing the pilin motif with the K190 nucleophile) and the C-terminal domain (^{C}SpaA, residues 350–500, containing the CWSS with the LPLTG motif) (Fig. 3*E*). The recombinant proteins were expressed in *E. coli* and purified (*SI Appendix*). When the two isolated domains were mixed at equal concentrations (300 μM) in a reaction with the lid-substituted $SrtA^{2M}$ enzyme (100 μM), a di-polypeptide conjugate was readily formed (Fig. 3*E*). Significantly, the presence of the expected Thr-Lys isopeptide in this conjugate was confirmed by MS (Table 3). Furthermore, control reactions demonstrated that SrtA containing the WT lid is markedly inactive in catalyzing cross-linking of the isolated domains (*SI Appendix*, Fig. S2). Together, our results

BIOCHEMISTRY

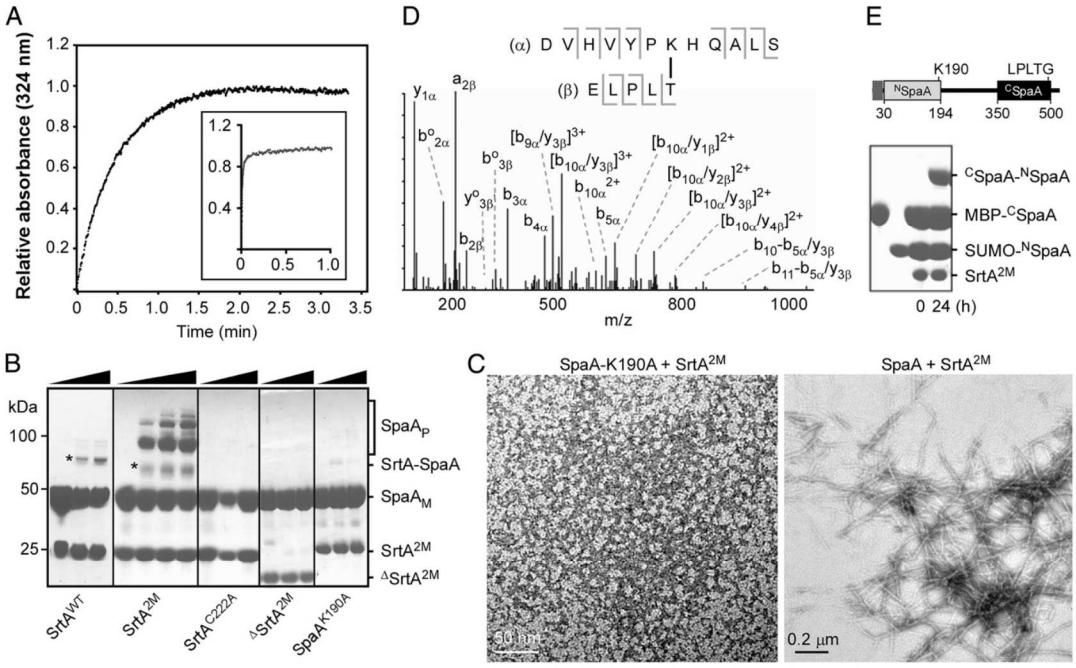


Fig. 3. Involvement of the lid in pilus polymerization in vitro. (*A*) Accessibility of the thiol group of the active site C222 residue in $SrtA^{WT}$ (main panel) or $SrtA^{2M}$ (*Inset*) enzymes was determined by stopped-flow experiments, whereby the reaction between the thiol-reactive reagent DTDP and C222 was monitored by absorbance at 324 nm. The experiments were performed in triplicate. (*B*) In vitro reconstitution of SpaA pilus polymerization was carried out at room temperature using various forms of recombinant SrtA and SpaA proteins at the molar ratio of 1:3. The reaction samples were analyzed by SDS/PAGE and Coomassie staining after 0, 24, and 48 h of incubation. Additional samples after 72 h of incubation were taken for the $SrtA^{2M}$ reactions (black triangles). SpaA monomers (M), polymers (P), and molecular mass markers are indicated. An SrtA–SpaA intermediate is marked with asterisks. (*C*) Protein samples from pilus polymerization reactions were analyzed by electron microscopy with negative staining using 0.75% uranyl formate. For comparison, recombinant $SrtA^{2M}$ and SpaA K190A proteins were included. (*D*) The isopeptide bond between residue T494 of the LPLTG motif and Lys residue K190 of the pilin motif in SpaA polymers in *B* was examined by MS/MS. Shown is the *m/z* tandem mass spectrum of the linked peptide (sequence shown in the *Inset*). (*E*) Fusion proteins between SUMO and the N-terminal SpaA domain (^{N}SpaA; residues 30–194) and between maltose-binding protein (MBP) and the C-terminal SpaA domain (^{C}SpaA; residues 350–500) were used with the $SrtA^{2M}$ enzyme in the in vitro pilus polymerization assay as described in *B*. The reaction samples were analyzed by SDS/PAGE and Coomassie staining after 24 h. The reactive Lys residue K190 and the LPXTG motif in the two domains are indicated.

support the concept that the lid in class C sortase functions in the molecular gating of substrate entry to the enzyme active site. Thus, we have demonstrated pilus polymerization in Gram-positive Actinobacteria in an in vitro reaction.

Structural Elements in a Sortase Required for Protein Polymerization. To gain insight into how SrtA joins the SpaA proteins together during polymerization, we performed molecular modeling of the ^{N}SpaA–SrtA–^{C}SpaA ternary complex in which the isopeptide bond is modeled using our previously determined crystal structures of SpaA [Protein Data Bank (PDB) ID code 3HR6] (36) and the isolated SrtA (PDB ID code 5K9A) proteins. We first generated a model of the SrtA–SpaA acyl-intermediate, juxtaposing the C terminus of the C-terminal SpaA domain with the active site C222 residue in SrtA. Because the crystal structure of SpaA lacks the CWSS that forms the acyl-intermediate with SrtA, we modeled the acyl-intermediate by placing the C-terminal domain of SpaA ~25 Å away from the active site Cys to accommodate the nine missing C-terminal residues that contain the CWSS. To construct the ternary complex, we then positioned the coordinates of the SpaA N-terminal domain near the acyl-intermediate to juxtapose the reactive Lys K190 of the pilin motif with the active site C222 residue (Fig. 4*A*). The resulting model of the ternary complex makes it readily obvious that the β7/β8 loop near the active site Cys residue and the N-terminal H1 helix in SrtA are in contact with the SpaA N-terminal domain, raising the possibility that these elements might play a role in recognizing the region of SpaA that houses the reactive Lys nucleophile. Strikingly, a primary sequence alignment of SrtA and other class C sortases indicates that they all contain a conserved TP(Y/L)XIN(S/T)H motif within the β7/β8 loop (*SI Appendix*, Fig. S3). This motif is clearly absent in other types of sortases that are known to attach proteins to the cell wall (class A, B, D, and E enzymes) but are unable to polymerize proteins. We thus postulated that the β7/β8 loop may play a role in conferring the polymerization activity in the class C enzymes.

In our model of the ternary reaction intermediate, the side chains of Y225, N228, and S229 within the TP(**Y**/L)XI**N**(**S**/T)H motif extend from the enzyme's surface in a position to contact ^{N}SpaA. To explore their possible roles in catalysis, we constructed a series of mutants of the lid-opened $SrtA^{2M}$ mutant enzyme in which each of these residues was individually replaced by Ala. The purified S229A and N228A mutant $SrtA^{2M}$ proteins were each defective in transpeptidation in vitro, as no isopeptide-linked SpaA–SpaA product was produced even after 48 h; the Y225A mutant protein had impaired transpeptidation activity as well, but

Table 3. MS analysis of synthetic SpaA pilus polymers

HPLC retention time, min	Calculated MH^+	Predicted peptides	Mass accuracy, ppm	Pilus assembly reaction
24.4	1947.0334	DVHVYPKHQALS :: ELPLT	4.0	Reaction with $SpaA_{30–500}$
26.5	1410.7627	DVHVYPK* :: ELPLT	2.8	Reaction with $SpaA_{30–500}$
22.9	1675.8802	DVHVYPKHQ* :: ELPLT	3.6	Reaction with $SpaA_{30–500}$
29.0	2336.2034	DVHVYPKHQALS :: NAGFELPLT	4.7	Reaction with $SpaA_{30–500}$
23.3	2400.2922	DVHVYPKHQALSEPVK :: ELPLT	2.9	Reaction with $SpaA_{30–500}$
ND	2485.2756	DQITLITCTPYAVNSHR :: ELPLT	ND	Reaction with $SpaA_{30–500}$
ND	2874.4455	DQITLITCTPYAVNSHR :: NAGFELPLT	ND	Reaction with $SpaA_{30–500}$
ND	1657.8717	DQITLITCTP* :: ELPLT	ND	Reaction with $SpaA_{30–500}$
ND	2047.0416	DQITLITCTP* :: NAGFELPLT	ND	Reaction with $SpaA_{30–500}$
24.5	1947.0334	DVHVYPKHQALS :: ELPLT	4.5	Reaction with ^{N}SpaA and ^{C}SpaA
26.5	1410.7627	DVHVYPK :: ELPLT	3.3	Reaction with ^{N}SpaA and ^{C}SpaA
22.9	1675.8802	DVHVYPKHQ :: ELPLT	3.5	Reaction with ^{N}SpaA and ^{C}SpaA
ND	2336.2034	DVHVYPKHQALS :: NAGFELPLT	ND	Reaction with ^{N}SpaA and ^{C}SpaA
ND	2400.2922	DVHVYPKHQALSEPVK :: ELPLT	ND	Reaction with ^{N}SpaA and ^{C}SpaA
ND	2485.2756	DQITLITCTPYAVNSHR :: ELPLT	ND	Reaction with ^{N}SpaA and ^{C}SpaA
ND	2874.4455	DQITLITCTPYAVNSHR :: NAGFELPLT	ND	Reaction with ^{N}SpaA and ^{C}SpaA
ND	1657.8717	DQITLITCTP* :: ELPLT	ND	Reaction with ^{N}SpaA and ^{C}SpaA
ND	2047.0416	DQITLITCTP* :: NAGFELPLT	ND	Reaction with ^{N}SpaA and ^{C}SpaA
36.5	1023.6448	PKLI :: ELPLT	11.9	Reaction with ^{C}SpaA and SpaB

MH^+, the mass of the singly protonated species; ND, not determined.
*Not expected cleavage sites.

to a lesser extent than the S229A and N228A mutants (Fig. 4*C*). Recall that the removal of the H1 helix in the $SrtA^{2M}$ enzyme also abrogates pilus polymerization (Fig. 3*B*, lanes $^{\Delta}SrtA^{2M}$). We have determined that the absence of the H1 helix does not cause the protein to unfold, since the ^{1}H-^{15}N heteronuclear single-quantum correlation (HSQC) spectra of $SrtA^{2M}$ and $^{\Delta}SrtA^{2M}$ are generally similar (*SI Appendix*, Fig. S4). We conclude that specific residues within the β7/β8 loop and the presence of the H1 helix form a functionally important contact surface with ^{N}SpaA. This is supported by experiments with an $SrtA^{2M}$ mutant harboring the N165A substitution in the proximal β4/β5 loop, which showed that this mutant retained nearly WT activity (Fig. 4*C*; lanes $SrtA^{N165A}$).

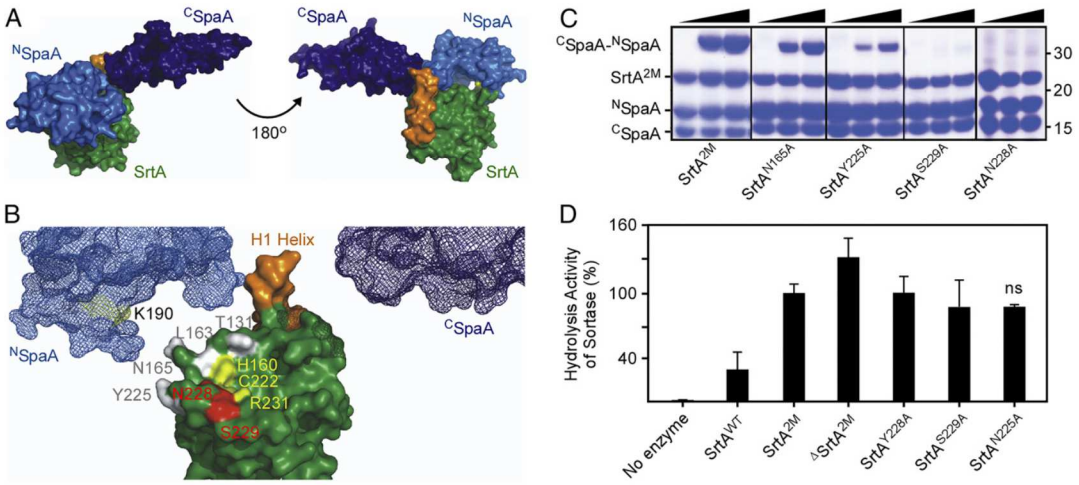


Fig. 4. Structural modeling reveals SrtA residues critical for transpeptidation activities. (*A*) An SrtA–SpaA pilin polymerase attack complex was visualized and assembled using PyMOL. Shown is an SpaA molecule splitting into its two domains, ^{N}SpaA (light blue) and ^{C}SpaA (dark blue). The SrtA enzyme is shown in green, and the H1 helix potentially bridging interactions between the two SpaA domains is seen in orange. (*B*) The detailed locations of the SrtA catalytic triad (marked in yellow) and the surrounding residues at the active site of the pilin polymerase attack complex are shown. The N-terminal H1 helix bridges the two reactive domains of SpaA and potentially facilitates interactions for the formation of the SrtA–SpaA polymerization attack complex. (*C*) Transpeptidation activity of $SrtA^{2M}$ and its variants (N165A, Y225A, N228A, and S229A) was determined in the pilus polymerization assay described in Fig. 3*E*, using domain substrates ^{N}SpaA and ^{C}SpaA. Protein samples were analyzed by SDS/PAGE and Coomassie staining after 24 h. The ligated product ^{C}SpaA–^{N}SpaA, sortase enzymes, substrates, and molecular markers are indicated. (*D*) Hydrolysis activity of SrtA enzymes was determined by an HPLC-based assay. WT or mutant SrtA (50 μM) was incubated with 500 μM KNAGFELPLTGGSGRI ($SpaA^{pep}$) in a 100-μL assay at 37 °C for 48 h. Reaction products were monitored and separated using HPLC at an absorbance of 215 nm. The peak fractions were collected and identified by MALDI-TOF-MS. The hydrolysis activity by $SrtA^{2M}$ is set as 100%. The results are presented as the average of three independent experiments; error bars indicate SDs; ns, not significant.

The model of the ternary complex raises the possibility that critical residues in the β7/β8 loop and the H1 helix may be required only for nucleophile recognition during the transpeptidation reaction but not for the other step of catalysis in which the LPXTG sorting motif is cleaved to form the thioacyl enzyme–substrate intermediate (4, 38). To test this hypothesis, we determined the importance of these structural elements in thioacyl-intermediate formation, using an established HPLC-based assay (39, 40) and an SpaA-derived peptide KNAGFELPLTGGSGRI ($SpaA^{pep}$) as the substrate. The enzymes and the $SpaA^{pep}$ substrate were mixed at a 1:10 molar ratio, and the loss of the intact peptide was monitored by HPLC with the hydrolysis activity of $SrtA^{2M}$ set as 100%. Consistent with a selective role in nucleophile recognition, none of the mutants exhibited any significant defect in cleaving the LPXTG motif (Fig. 4*D*). Importantly, the $^{\Delta}SrtA^{2M}$ enzyme, which was inactive in the pilus polymerization assay (Fig. 3*B*), cleaved the $SpaA^{pep}$ substrate with an efficiency comparable to that of the activated $SrtA^{2M}$ mutant enzyme (Fig. 4*D*). Under these conditions, the hydrolysis kinetics of the $SrtA^{2M}$ and $^{\Delta}SrtA^{2M}$ enzymes displayed a comparable V_{max} of 2.3 ± 0.2 and 3.3 ± 0.8 μM/h, respectively, unlike that of $SrtA^{WT}$, which was significantly reduced (0.6 ± 0.1 μM/h). These results prompt us to propose that the conserved TP(Y/L)XIN(S/T)H motif within the β7/β8 loop is a hallmark feature of the class C sortases that enables molecular recognition of the pilin motif Lys nucleophile in their cognate substrates. The H1 helix appears to play a similar role; however, it is not well conserved in class C sortases.

SrtA-Catalyzed Pilus Polymerization Is Terminated by SpaB. Our previous studies suggest that SpaB acts as a molecular switch that terminates pilus polymerization by incorporating into the pilus polymer as the terminal subunit, and this reaction requires the Lys residue K139 present on SpaB (11), which is then anchored to the cell wall by the housekeeping sortase SrtF (17). The SpaA polymer is presumed to be linked to the terminal SpaB via an isopeptide bond formed between the Thr residue of the SpaA LPXTG motif and K139 (Fig. 5*A*). To examine if this is the case, we produced a recombinant SpaB protein (residues 25–180), which lacked the N-terminal signal peptide and the hydrophobic domain and the C-terminal charged tail of the CWSS but contained the LAFTG motif. As a control, an SpaB mutant protein with the K139A substitution mutation was also generated. These recombinant proteins, expressed in and purified from *E. coli*, were then used in the pilus polymerization assay and analyzed by SDS/PAGE and Coomassie staining as described in Fig. 3*B*. In the presence of the $SrtA^{2M}$ enzyme, recombinant SpaA protein was polymerized into HMM species as expected (Fig. 5*B*, first two lanes). Remarkably, when SpaB was added into this reaction, the formation of SpaA polymers was significantly reduced, and the SpaA–SpaB dimer accumulated (Fig. 5*B*, third and fourth lanes, asterisk); it is important to note that while some trimeric forms of SpaA were observed, $SrtA^{2M}$ was unable to further polymerize SpaA pilins in the presence of WT SpaB, although SpaA substrates were abundantly available (Fig. 5*B*, lane 4). This suggests that the SpaB K139 may be more nucleophilic than the SpaA K190 or that SpaB K139 may have a higher affinity and the ability to outcompete SpaA for pilin cross-linking reactions. The fact that SpaA pilus polymerization catalyzed by $SrtA^{2M}$ was not affected in the presence of the SpaB K139A mutant protein is consistent with K139 as the nucleophile in the cross-linking reaction (Fig. 5*B*, last two lanes).

The observation that the SpaA–SpaB dimer is the predominant form of pilin conjugates produced by $SrtA^{2M}$ in the presence of SpaB (Fig. 5*B*, lane 4) prompted us to test whether the pilin motif is dispensable for SrtA-catalyzed SpaA–SpaB conjugation. To examine if this is the case, the recombinant protein ^{C}SpaA (see Fig. 4*C*), lacking the pilin motif, was used in place of recombinant SpaA. The reaction was performed as described in Fig. 4*C*. Indeed, after 24 h a band corresponding to the size of a truncated SpaA–SpaB conjugate was observed (Fig. 5*C*). To corroborate this and analyze the linkage between SpaA and SpaB, this band was excised for MS analysis (see Fig. 3*D* for methods). MS/MS data confirmed the isopeptide bond between the SpaA T494 residue and the SpaB K139 residue (Fig. 5*D* and Table 3). Clearly, our results demonstrate that SpaB is a termination factor for pilus polymerization.

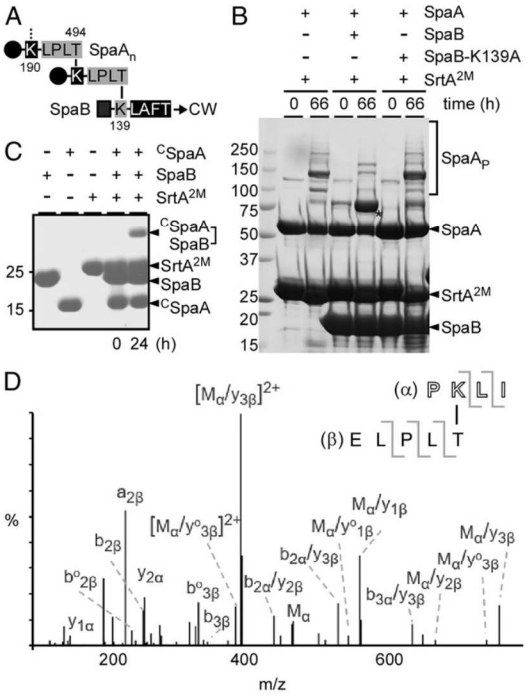


Fig. 5. SrtA-catalyzed pilus polymerization is terminated by the pilus base SpaB pilin. (*A*) Depicted is an SpaA polymer ($SpaA_n$) with individual subunits cross-linked by an isopeptide bond between K190 and T494. The SpaA polymer is linked to the terminal SpaB via an isopeptide bond between the T494 residue of SpaA and the K139 residue of SpaB, which in turn is covalently attached to the bacterial peptidoglycan (CW). (*B*) A pilus polymerization termination assay was performed with $SrtA^{2M}$ enzyme and SpaA substrate (2:1 molar ratio) in the presence (same concentration as SpaA) or absence of SpaB or SpaB with the K139A mutation. The reactions were stopped after 66 h by the addition of SDS-containing sample buffer, and protein samples were analyzed by SDS/PAGE and Coomassie staining. The asterisk indicates an SpaA–SpaB dimer. Molecular mass markers (kDa) are shown. (*C*) ^{C}SpaA (residues 350–500), SpaB, and $SrtA^{2M}$ were used in the pilus polymerization termination assay, with substrates and enzyme at a 3:1 molar ratio. Protein samples were analyzed by SDS/PAGE and Coomassie staining after 24-h incubation. (*D*) A gel band corresponding to a ^{C}SpaA–SpaC dimer was excised for tryptic digestion and MS/MS. Shown is the MS/MS spectrum, which revealed the presence of the isopeptide bond formed between T494 and K139.

In conclusion, we report here the high-resolution crystal structures of the *C. diphtheriae* pilus-specific sortase SrtA enzyme ($SrtA^{WT}$) and a mutant form of the enzyme with mutations in the lid region ($SrtA^{2M}$) and through these illuminate some of the basic features of the sortase that functions to polymerize pilus proteins in Gram-positive bacteria. The structure of the WT enzyme displayed a characteristic "closed" configuration of a class C sortase with its catalytic site occluded by a molecular lid (Fig. 1). By introducing specific amino acid substitutions within

the lid, we were able to generate an enzyme whose catalytic pocket displayed an open conformation with no other major perturbations detected in the atomic structure. The functional importance of these two states of the enzyme was demonstrated by our ability to reconstitute a robust pilus polymerization reaction in vitro using the cognate shaft pilin. While the WT form could not polymerize the shaft pilin in a reaction, the opened-lid version of the enzyme is highly active. We showed that the activated enzyme was able to recognize the sorting signal, form the relevant acyl-enzyme intermediate, recognize the pilin motif Lys residue, and catalyze isopeptide bond formation conjugating Lys of one pilin protomer with Thr of another protomer (Fig. 3). We then utilized structural modeling to identify specific structural elements conserved in a pilus-specific sortase which are important for catalyzing the transpeptidation reaction in vitro (Fig. 4). We also provide additional in vitro evidence in support of SpaB being a terminator of pilus polymerization (Fig. 5).

Given that the $SrtA^{2M}$ enzyme is able to catalyze pilus polymerization and SpaB incorporation in vitro, although pilus assembly in vivo is not apparently altered when the mutant enzyme is present, we surmise that the SrtA lid might play some form of a modulatory role in pilus polymerization and termination in vivo. For instance, as SpaB is a preferred substrate of the housekeeping sortase SrtF (17), charging SrtF with SpaB triggers an SrtF–SrtA interaction. This interaction could potentially alter the configuration of the SrtA lid, allowing SrtA-mediated entry of SpaB to the pilus base and subsequently transferring the pilus polymer to SrtF to complete the cell wall-anchoring step. Future experiments may be designed to examine the lid dynamics in the presence or absence of SpaB or SrtF.

Importantly, we have shown here that the separated domains of the pilin, one containing the pilin motif and the other containing the sorting motif, could be ligated efficiently to produce a di-polypeptide conjugate containing the Lys–Thr isopeptide bond. This provides a powerful protein ligation platform for engineering designer proteins that is mechanistically different from the "sortagging" technology developed with the archetypal *S. aureus* sortase which normally functions to cross-link surface proteins to the bacterial cell wall but does not polymerize proteins (41, 42). We envision that the surface display of protein polymers, protein labeling of living cells, and protein ligation are a few examples of many potential biotechnological and biological applications of this enzyme.

Materials and Methods

Bacterial Strains, Plasmids, and Media. Bacterial strains and plasmids used in this study are listed in *SI Appendix*, Table S1. *SI Appendix, SI Materials and Methods* contains information regarding recombinant plasmids, protein purification, protein crystallization, and structure determination. *C. diphtheriae* strains were grown in Heart Infusion (HI) broth (Becton Dickinson) or on HI agar plates at 37 °C. When needed, kanamycin was added at a concentration of 25 μg/mL. *E. coli* DH5α and BL21 (DE3), used for cloning and protein expression and purification, respectively, were grown in either Luria–Bertani or 2× YT broth (Sigma-Aldrich) at 37 °C in the presence of ampicillin at 100 μg/mL.

In Vitro Reconstitution of Pilus Polymerization. In vitro reactions were carried out at room temperature, and proteins were dissolved in assay buffer [50 mM Tris·HCl (pH 8.0), 300 mM NaCl, 1 mM DTT]. All reactions used a fixed 100-μM concertation of SrtA enzyme and 300-μM SpaA substrate (either full length or each individual domain). Reactions were stirred gently by continuous rotation. Aliquots were taken at 0 h, 24 h, 48 h, and 72 h, and reactions were quenched by the addition of two volumes of SDS loading dye. For a pilus termination assay, SpaB was used at the same concentration as SpaA, whereas the molar concentrations of $SrtA^{2M}$ enzyme and SpaA substrate followed 3:1 or 2:1 ratios. The reactions were performed in 24 or 66 h, respectively.

Probing Accessibility of the SrtA Active Site. Reaction rates of DTDP and SrtA proteins via the Cys C222 residue were determined by stopped-flow experiments, which were performed at 23 °C using an Applied Photophysics Ltd. Model SX.18 MV sequential stopped-flow spectrofluorimeter with a 150 W Xe/Hg lamp and a dead time of 1.7 ms. All triplicate reactions were carried out in reaction buffer [50 mM 3-(*N*-morpholino)propanesulfonic acid (Mops), 200 mM KCl, 1 mM EDTA, pH 7.5]. Absorbance was monitored at 324 nm after solutions were rapidly mixed in syringe A, containing 0.6 mg/mL protein, and syringe B, containing 0.32 mM DTDP. Reaction rates (k) were derived by fitting data to the following equations with one (Eq. **1**) or two (Eq. **2**) rates:

$$A = A_{max} * \left(1 - e^{-kt}\right) \quad [1]$$

or

$$A = A_{max1} * \left(1 - e^{-k1t}\right) + A_{max2} * \left(1 - e^{-k2t}\right), \quad [2]$$

where A is absorbance at 324 nm at time t, and A_{max} is the maximum absorbance.

Cell Fractionation and Western Blotting. Cell fractionation and Western blotting were performed according to published procedures with some modifications (43, 44). Briefly, midlog-phase cultures of *C. diphtheriae* strains grown at 37 °C were normalized to an OD_{600} of 1.0 and were subjected to cell wall protein extraction using mutanolysin (300 U/mL). Protein samples obtained from culture medium (S) and cell wall (W) were trichloroacetic acid precipitated and acetone washed. The protoplasts after the cell wall extraction were used for analysis of cell membrane-bound proteins. Protein samples were resuspended in SDS sample buffer containing 3% urea and were heated at 100 °C for 10 min before SDS/PAGE analysis using 3–12% or 3–20% Tris-Gly gradient gels. Detection of proteins was performed by immunoblotting with specific antibodies (1:20,000 for α-SpaA; 1:4,000 for α-SpaH; 1:5,000 for α-MdbA; and 1:4,000 for α-SrtA).

MS of Pilus Polymers. Protein digestion and isopeptide bond identification were performed according to previous protocols (36, 45). Specifically, proteins entrapped in gel bands were reduced with 10 mM DTT (Sigma) at 60 °C for 1 h and then were alkylated with 50 mM iodoacetamide (Sigma) at 45 °C for a few minutes in the dark. These reduction and alkylation steps were skipped for the acyl-intermediate samples. Samples were digested with 200 ng trypsin (Thermo Scientific) at 37 °C overnight. At the end of trypsin digestion, 200 ng of Asp-N endoproteinase (Thermo Scientific) was added for another overnight incubation. Digested peptides were extracted from the gel bands in 50% acetonitrile/49.9% water/0.1% TFA and were cleaned with C18 StageTip (46) before MS analysis.

Digested peptides were separated on an EASY-Spray column (25 cm × 75 μm i.d., PepMap RSLC C18, 2 μm; Thermo Scientific) connected to an EASY-nLC 1000 nUPLC (Thermo Scientific) using a gradient of 5–35% acetonitrile in 0.1% formic acid and a flow rate of 300 nL/min for 30 min. Tandem mass spectra were acquired in a data-dependent manner with an Orbitrap Q Exactive mass spectrometer (Thermo Fisher Scientific) interfaced to a nanoelectrospray ionization source.

The raw MS/MS data were converted into MGF format by Thermo Proteome Discoverer 1.4 (Thermo Scientific). In-house programs to search for the isopeptides were used for two different approaches. The first approach was performed as previously described (36) and was used to calculate the masses of predicted peptides containing the isopeptide linkage to guide the search. The second approach was based on published (36) and our own observations of the presence of ions specific for the fragments of ELPLT (*m/z* 215.138, 225.122, 243.132, 294.180, 312.190, 322.174, and 340.186). The in-house programs sifted through tens of thousands of mass spectra searching specifically for this information and extracted MS/MS spectra for further analyses.

Determination of SrtA Hydrolysis by an HPLC-Based Assay. In vitro hydrolysis reactions were performed based on the method developed by Kruger et al. (47). WT or mutant SrtA (50 μM) was incubated with 500 μM KNAGFELPLTGGSGRI ($SpaA^{pep}$) in 100-μL reactions at 37 °C for 24 h. The reactions were quenched by adding 50 μL of 1 M HCl and were injected onto a Waters XBridge Peptide BEH C18 reversed-phase HPLC column. Peptides were eluted by applying a gradient from 5–51% acetonitrile (in 0.1% TFA) over 25 min at a flow rate of 1 mL/min. Elution of the peptides was monitored by absorbance at 215 nm. Peak fractions were collected, and their identities were confirmed by MALDI-TOF-MS.

Electron Microscopy. For visualization of in vitro pilus polymers, pilus polymerization reactions were diluted in half with water and 7-μL aliquots were applied onto carbon-coated nickel grids, washed five times with distilled water, and stained with 0.75% uranyl formate for 2 min before viewing by a JEOL JEM-1400 electron microscope.

For visualization of pili produced by corynebacterial cells, electron microscopy and IEM were performed according to a published protocol (48).

Briefly, corynebacterial cells grown on HI agar plates were washed and suspended in PBS. Seven-microliter aliquots of the cell suspension were applied onto nickel grids, washed, and stained with 0.75% uranyl formate before viewing by an electron microscope. For IEM, cells were stained with α-SpaA (1:100 dilution), followed by staining with 12-nm gold particles conjugated to IgG, before staining with 1% uranyl acetate.

To estimate the dimension of pili, ImageJ (https://imagej.nih.gov/ij/) was employed. Twenty-five measurements were performed at different locations of pili for each strain. Statistical analysis was performed by GraphPad Prism.

ACKNOWLEDGMENTS. We thank members of the R.T.C. and H.T.-T. laboratories for critical review and discussion of the manuscript and members of the Structural Biology Center at Argonne National Laboratory for help in conducting X-ray diffraction data collection. This work was supported by Department of Health and Human Services, NIH National Institute of Allergy and Infectious Diseases Award AI52217 (to R.T.C.) and in part by Center of Structural Genomics of Infectious Diseases Contracts HHSN272201200026C and HHSN272201700060C (to A.J.), National Institute of General Medical Sciences Grant GM103479 (to J.A.L.), and National Institute of Dental and Craniofacial Research Awards DE017382 and DE025015 (to H.T.-T.).

1. Kline KA, Dodson KW, Caparon MG, Hultgren SJ (2010) A tale of two pili: Assembly and function of pili in bacteria. *Trends Microbiol* 18:224–232.
2. Hospenthal MK, Costa TRD, Waksman G (2017) A comprehensive guide to pilus biogenesis in Gram-negative bacteria. *Nat Rev Microbiol* 15:365–379.
3. Thanassi DG, Bliska JB, Christie PJ (2012) Surface organelles assembled by secretion systems of Gram-negative bacteria: Diversity in structure and function. *FEMS Microbiol Rev* 36:1046–1082.
4. Ton-That H, Schneewind O (2004) Assembly of pili in Gram-positive bacteria. *Trends Microbiol* 12:228–234.
5. Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G (2006) Pili in Gram-positive pathogens. *Nat Rev Microbiol* 4:509–519.
6. Mandlik A, Swierczynski A, Das A, Ton-That H (2008) Pili in Gram-positive bacteria: Assembly, involvement in colonization and biofilm development. *Trends Microbiol* 16: 33–40.
7. Ton-That H, Schneewind O (2003) Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Mol Microbiol* 50:1429–1438.
8. Rogers EA, Das A, Ton-That H (2011) Adhesion by pathogenic corynebacteria. *Adv Exp Med Biol* 715:91–103.
9. Swierczynski A, Ton-That H (2006) Type III pilus of corynebacteria: Pilus length is determined by the level of its major pilin subunit. *J Bacteriol* 188:6318–6325.
10. Gaspar AH, Ton-That H (2006) Assembly of distinct pilus structures on the surface of *Corynebacterium diphtheriae*. *J Bacteriol* 188:1526–1533.
11. Mandlik A, Das A, Ton-That H (2008) The molecular switch that activates the cell wall anchoring step of pilus assembly in Gram-positive bacteria. *Proc Natl Acad Sci USA* 105:14147–14152.
12. Mandlik A, Swierczynski A, Das A, Ton-That H (2007) *Corynebacterium diphtheriae* employs specific minor pilins to target human pharyngeal epithelial cells. *Mol Microbiol* 64:111–124.
13. Ton-That H, Marraffini LA, Schneewind O (2004) Sortases and pilin elements involved in pilus assembly of *Corynebacterium diphtheriae*. *Mol Microbiol* 53:251–261.
14. Kang HJ, Coulibaly F, Clow F, Proft T, Baker EN (2007) Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. *Science* 318:1625–1628.
15. Navarre WW, Schneewind O (1999) Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol Mol Biol Rev* 63: 174–229.
16. Siegel SD, Liu J, Ton-That H (2016) Biogenesis of the Gram-positive bacterial cell envelope. *Curr Opin Microbiol* 34:31–37.
17. Swaminathan A, et al. (2007) Housekeeping sortase facilitates the cell wall anchoring of pilus polymers in *Corynebacterium diphtheriae*. *Mol Microbiol* 66:961–974.
18. Mazmanian SK, Liu G, Ton-That H, Schneewind O (1999) *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science* 285:760–763.
19. Ton-That H, Liu G, Mazmanian SK, Faull KF, Schneewind O (1999) Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc Natl Acad Sci USA* 96:12424–12429.
20. Dramsi S, Trieu-Cuot P, Bierne H (2005) Sorting sortases: A nomenclature proposal for the various sortases of Gram-positive bacteria. *Res Microbiol* 156:289–297.
21. Spirig T, Weiner EM, Clubb RT (2011) Sortase enzymes in Gram-positive bacteria. *Mol Microbiol* 82:1044–1059.
22. Jacobitz AW, Kattke MD, Wereszczynski J, Clubb RT (2017) Sortase transpeptidases: Structural biology and catalytic mechanism. *Adv Protein Chem Struct Biol* 109: 223–264.
23. Khare B, V L Narayana S (2017) Pilus biogenesis of Gram-positive bacteria: Roles of sortases and implications for assembly. *Protein Sci* 26:1458–1473.
24. Manzano C, Izoré T, Job V, Di Guilmi AM, Dessen A (2009) Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of Gram-positive pathogens. *Biochemistry* 48:10549–10557.
25. Khare B, Fu ZQ, Huang IH, Ton-That H, Narayana SV (2011) The crystal structure analysis of group B *Streptococcus* sortase C1: A model for the "lid" movement upon substrate binding. *J Mol Biol* 414:563–577.
26. Persson K (2011) Structure of the sortase AcSrtC-1 from *Actinomyces oris*. *Acta Crystallogr D Biol Crystallogr* 67:212–217.
27. Neiers F, et al. (2009) Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity. *J Mol Biol* 393:704–716.
28. Jacobitz AW, et al. (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* 120:8302–8312.
29. Manzano C, et al. (2008) Sortase-mediated pilus fiber biogenesis in *Streptococcus pneumoniae*. *Structure* 16:1838–1848.
30. Cozzi R, et al. (2011) Structure analysis and site-directed mutagenesis of defined key residues and motives for pilus-related sortase C1 in group B *Streptococcus*. *FASEB J* 25: 1874–1886.
31. Wu C, et al. (2012) 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* 194:2531–2539.
32. Sanchez BC, Chang C, Wu C, Tran B, Ton-That H (2017) Electron transport chain is biochemically linked to pilus assembly required for polymicrobial interactions and biofilm formation in the Gram-positive actinobacterium *Actinomyces oris*. *MBio* 8: e00399-17.
33. Putkey JA, et al. (1997) Fluorescent probes attached to Cys 35 or Cys 84 in cardiac troponin C are differentially sensitive to Ca(2+)-dependent events in vitro and in situ. *Biochemistry* 36:970–978.
34. Epps DE, Vosters AF (2002) The essential role of a free sulfhydryl group in blocking the cholesteryl site of cholesteryl ester transfer protein (CETP). *Chem Phys Lipids* 114: 113–122.
35. Echelman DJ, et al. (2016) CnaA domains in bacterial pili are efficient dissipaters of large mechanical shocks. *Proc Natl Acad Sci USA* 113:2490–2495.
36. Kang HJ, Paterson NG, Gaspar AH, Ton-That H, Baker EN (2009) The *Corynebacterium diphtheriae* shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. *Proc Natl Acad Sci USA* 106:16967–16971.
37. Guttilla IK, et al. (2009) Acyl enzyme intermediates in sortase-catalyzed pilus morphogenesis in Gram-positive bacteria. *J Bacteriol* 191:5603–5612.
38. Ton-That H, Marraffini LA, Schneewind O (2004) Protein sorting to the cell wall envelope of Gram-positive bacteria. *Biochim Biophys Acta* 1694:269–278.
39. Aulabaugh A, et al. (2007) Development of an HPLC assay for *Staphylococcus aureus* sortase: Evidence for the formation of the kinetically competent acyl enzyme intermediate. *Anal Biochem* 360:14–22.
40. Ton-That H, Mazmanian SK, Faull KF, Schneewind O (2000) 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* 275:9876–9881.
41. Antos JM, et al. (2017) Site-specific protein labeling via sortase-mediated transpeptidation. *Curr Protoc Protein Sci* 89:15.13.11–15.13.19.
42. Proft T (2010) Sortase-mediated protein ligation: An emerging biotechnology tool for protein modification and immobilisation. *Biotechnol Lett* 32:1–10.
43. Chang C, Mandlik A, Das A, Ton-That H (2011) Cell surface display of minor pilin adhesins in the form of a simple heterodimeric assembly in *Corynebacterium diphtheriae*. *Mol Microbiol* 79:1236–1247.
44. Reardon-Robinson ME, et al. (2015) A thiol-disulfide oxidoreductase of the Gram-positive pathogen *Corynebacterium diphtheriae* is essential for viability, pilus assembly, toxin production and virulence. *Mol Microbiol* 98:1037–1050.
45. Thevis M, Ogorzalek Loo RR, Loo JA (2003) In-gel derivatization of proteins for cysteine-specific cleavages and their analysis by mass spectrometry. *J Proteome Res* 2: 163–172.
46. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2:1896–1906.
47. Kruger RG, Dostal P, McCafferty DG (2004) Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the *Staphylococcus aureus* sortase transpeptidase SrtA. *Anal Biochem* 326:42–48.
48. Chang C, Huang IH, Hendrickx AP, Ton-That H (2013) Visualization of Gram-positive bacterial pili. *Methods Mol Biol* 966:77–95.

Chapter 3

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

3.1 Overview

In Chapter 2, we reconstituted the pilus polymerization reaction of *C. diphtheriae in vitro*. However, this reaction is significantly slower than the activity observed in ^{Sa}SrtA, the archetypal sortase enzyme. To understand the understand the reasons underlying this disparity in reactivity, we performed a detailed kinetic analysis of the ^{Cd}SrtA protein labeling reaction using a newly developed HPLC assay. Chapter 3 describes this analysis, in which we demonstrated that the first step of the reaction, formation of the enzyme-sorting signal acyl intermediate, is the rate-limiting step. We also designed a third-generation variant, ^{Cd}SrtA$^{\Delta}$, in which the auto-inhibitory "lid" structure is completely deleted. Quantitative measurements revealed that this variant exhibits a catalytic turnover rate 7-fold faster than the second-generation variant *in vitro*. My contributions included study design, development of the underlying HPLC transpeptidation assay used for kinetic analysis, sortase engineering aimed at activity enhancement, and quantification of transpeptidation for a library of sorting signal 'X' position peptides.

This chapter is reformatted with permission from a peer-reviewed article "Kinetics and Optimization of the Lysine-Isopeptide Bond Forming Sortase Enzyme from Corynebacterium diphtheriae." Sue, C.K., McConnell, S.A., Ellis-Guardiola, K., Muroski, J.M., McAllister, R.A., Yu, J., Alvarez, A.I., Chang, C., Loo, R.R.O., Loo, J.A., Ton-That, H. and Clubb, R.T. *Bioconj. Chem.* **31** 1624-1634 (2020).

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

Bioconjugate Chemistry

pubs.acs.org/bc Article

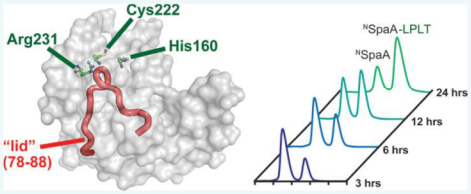
Kinetics and Optimization of the Lysine–Isopeptide Bond Forming Sortase Enzyme from *Corynebacterium diphtheriae*

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* (^{Cd}SrtA) has been reconstituted *in vitro*. A mutationally activated form of ^{Cd}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 ^{Cd}SrtA catalyzed protein labeling reaction. We show that the first step in catalysis is rate limiting, which is the formation of the ^{Cd}SrtA-peptide thioacyl intermediate that subsequently reacts with a lysine ε-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 (^{Cd}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 ^{Cd}SrtA$^{\Delta}$ generates its cross-linked product with a catalytic turnover number of 1.4 ± 0.004 h^{-1} and that it has apparent K_M values of 0.16 ± 0.04 and 1.6 ± 0.3 mM for its ^{N}SpaA and peptide substrates, respectively. ^{Cd}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 ^{Cd}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.[1−5] 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.[6−8] 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).[9,10] 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.[11,12] The sortase A enzyme from *Staphylococcus aureus* (^{Sa}SrtA) is one of the most widely used bioconjugating enzymes.[12−14] It has been successfully deployed to catalyze protein–protein ligations and backbone cyclization, and to modify proteins with peptides, lipids, sugars, and small molecules.[15−21] However, ^{Sa}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* (^{Cd}SrtA) can be used to

Received: March 25, 2020
Revised: May 7, 2020
Published: May 12, 2020

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. 2020, 31, 1624–1634

attach a peptide fluorophore via an isopeptide bond to a specific lysine residue within a protein.[22] Although ^{Cd}SrtA and ^{Sa}SrtA are both members of the sortase-superfamily of cysteine transpeptidases, they have distinct substrate specificities.[23] ^{Sa}SrtA is a class A sortase that catalyzes formation of backbone−backbone peptide bonds, whereas ^{Cd}SrtA is a class C sortase that joins molecules together via lysine−isopeptide bonds.[24,25] ^{Cd}SrtA assembles the SpaA pilus in *C. diphtheriae* by cross-linking SpaA "pilin" subunits via a lysine−isopeptide bond.[26,27] 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.[26] Repetition of this reaction forms the SpaA pilus, which is approximately 1−2 μm in length and further elaborated with unique tip and basal pilin proteins (Figure 1A).[27,28] *In vitro*, the native ^{Cd}SrtA enzyme is enzymatically inactive because it contains a polypeptide appendage that occludes its active site, called a "lid" (Figure 1C). However, ^{Cd}SrtA variants containing destabilizing amino acid substitutions in the lid exhibit *in vitro* activity.[28] The most active form of the enzyme thus far discovered is ^{Cd}SrtA^{3M}, which contains residues N37−Q257 in ^{Cd}SrtA and D81G/W83G/N85A substitutions in the lid.[22] ^{Cd}SrtA^{3M} is a promising bioconjugation tool, as it can be used to selectively modify proteins harboring the N-terminal domain of SpaA (^{N}SpaA) with peptide fluorophores. Modification via lysine−isopeptide bonds is attractive, as these linkages may be less susceptible to proteolysis and enzymatic reversibility.

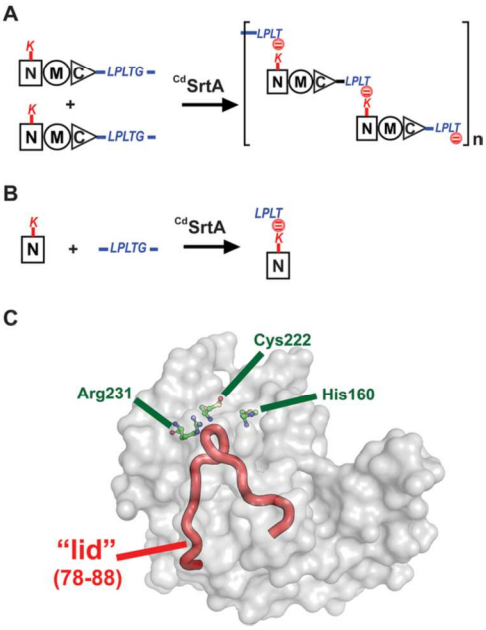


Figure 1. *C. diphtheriae* ^{Cd}SrtA pilin sortase catalyzes lysine isopeptide bond formation. (A) Schematic showing the pilin polymerization reaction catalyzed by ^{Cd}SrtA. The enzyme creates the SpaA pilus by polymerizing SpaA pilin proteins. In the reaction, it recognizes lysine (K190) side chain nucleophile within the N-terminal domain of SpaA (^{N}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 ^{Cd}SrtA enzyme ligates the isolated ^{N}SpaA domain to the peptide containing the LPLTG sorting signal (FELPLTGGSG). (C) Structure of ^{Cd}SrtA showing H160, C222, and R231 active site residues. The "Lid" is highlighted in red (residues P77 to S89).

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 ^{Cd}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 ^{Sa}SrtA, ^{Cd}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 ^{Cd}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 ^{Cd}SrtA^{3M} using SDS-PAGE.[22] 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 ^{Cd}SrtA^{3M} to join the N-terminal domain from SpaA (^{N}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}SpaA-LPLT product in which the K190 side chain in ^{N}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}SpaA) into the cross-linked protein−peptide product (^{N}SpaA-LPLT) (Figure 2B). The identity of the product and the location of its isopeptide linkage were previously confirmed by LC-MS/MS.[22,28] Initially, for each substrate (^{N}SpaA and the LPLTG peptide) the dependence of the reaction velocity on substrate concentration was determined (Figure 2C). This analysis reveals that ^{N}SpaA and peptide substrates do not saturate the enzyme even when they are present at concentrations of 500 μM (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}SpaA was varied from 62.5 to 500 μM 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}SpaA was held fixed at 500 μM. These measurements revealed that ^{Cd}SrtA^{3M} catalyzes isopeptide

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. 2020, 31, 1624−1634

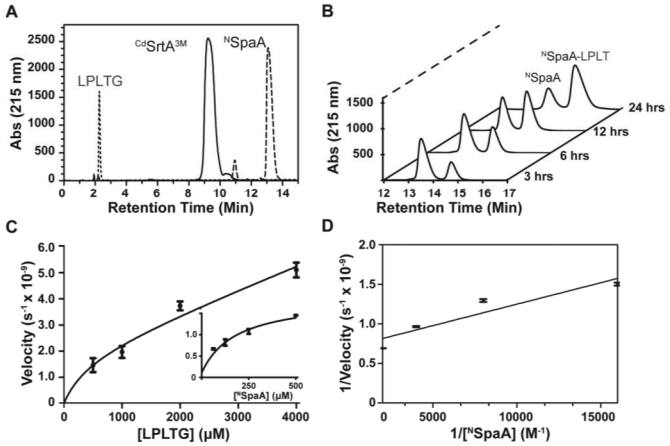


Figure 2. ^{Cd}SrtA transpeptidation assay: (A) Superimposed reversed-phase HPLC traces showing the separation of the sorting signal peptide (LPLTG, short dashes), ^{Cd}SrtA^{3M} (solid line), and ^{N}SpaA (long dashes). (B) Representative HPLC traces that track the progress of the reaction. Peaks corresponding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

Table 1. Kinetics of ^{Cd}SrtA Catalyzed Lysine–Isopeptide Formation[a]

	$k_{cat} \times 10^{-5}$ (s^{-1})[b]	$^{N}K_M \times 10^{-4}$ (M)	$^{S}K_M \times 10^{-4}$ (M)	$k_{cat}/^{N}K_M$ (s^{-1} M^{-1})
^{Cd}SrtA	n.d.[c]	n.d.	n.d.	n.d.
^{Cd}SrtA^{3M}	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.
^{Cd}SrtA$^{\Delta}$	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.
^{Sa}SrtA[d]	1600 ± 100	1.8 ± 0.1	73.3 ± 10.1[e]	86 ± 5

[a]All kinetics are approximations as saturating concentrations were not able to be measured. [b]Transpeptidation steady-state kinetic parameters for ^{Cd}SrtA were determine by the monitoring rate at which the enzyme ligated an FELPLTGGSG peptide to the ^{N}SpaA domain via a lysine–isopeptide bond. [c]n.d., not determined because an insufficient amount of product was detectable. [d]Transpeptidation steady-state kinetic parameters for ^{Sa}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. [e]Values are reported from Frankel et al. (2005) and measures reactions between a Abz-LPETG-Dap(Dnp)-NH_2 and pentaglycine.[29] ^{N}Refers to transpeptidation kinetics measure when ^{N}SpaA is varied and FELPLTGGSG concentration is held constant. ^{S}Refers to when FELPLTGGSG peptide is varied and ^{N}SpaA is held constant.

bond formation with apparent K_M values for ^{N}SpaA ($^{N}K_M$) and the LPLTG peptide ($^{S}K_M$) of 70 ± 10 μM 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}SpaA is held constant at 500 μM and the LPLTG peptide is varied).

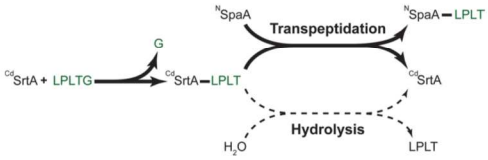
Formation of the ^{Cd}SrtA-LPLT Thioacyl Intermediate Is a Rate Limiting Step in Catalysis and Can Be Accelerated by Completely Removing the Lid. ^{Cd}SrtA^{3M} is the most active form of ^{Cd}SrtA thus far discovered and contains three substitutions in the inhibitory lid structure (D81G, W83G, N85A).[22] 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 ^{Cd}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.[28] Indeed, the measured steady-state kinetic parameters for a lid-deletion ^{Cd}SrtA$^{\Delta}$ (residues N37–Q257 of ^{Cd}SrtA in which the amino acids I78 to A88 are deleted) reveal that it is more active than ^{Cd}SrtA^{3M}; 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 ^{Cd}SrtA$^{\Delta}$ is catalytically more active than ^{Cd}SrtA^{3M}, we investigated how lid removal affected catalysis. By analogy to the prototypical ^{Sa}SrtA enzyme, ^{Cd}SrtA presumably catalyzes lysine–isopeptide bond formation via a two-step

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. 2020, 31, 1624–1634

process (Scheme 1).[23,29−31] In the transpeptidation mechanism, ^{Cd}SrtA's C222 thiol presumably functions as a nucleophile to attack the threonine carbonyl carbon in the LPLTG sorting signal to form a ^{Cd}SrtA-LPLT thioacyl intermediate. Then, the enzyme recognizes the K190 side chain amine group located within ^{N}SpaA, which resolves the thioacyl intermediate to form the lysine−isopeptide linked ^{N}SpaA-LPLT product. In ^{Sa}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.[29] 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. ^{Cd}SrtA has also been shown to proteolyze its LPLTG peptide substrate, but the kinetics and extent of proteolysis have not been rigorously measured.[22,28]

Scheme 1. Schematic Showing the Overall Mechanism of ^{Cd}SrtA-Catalyzed Isopeptide Formation (Top) and a Potential Hydrolytic Side Reaction (Bottom)



To determine if differences in the rate of the hydrolytic side reaction cause the ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M} 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 ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M} 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 μM enzyme and 500 μM 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 ^{Sa}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 ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M} 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 ^{Cd}SrtA$^{\Delta}$ originated from its ability to form the enzyme−LPLT thioacyl intermediate more rapidly than ^{Cd}SrtA^{3M} (Scheme 1). For both the ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M} 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

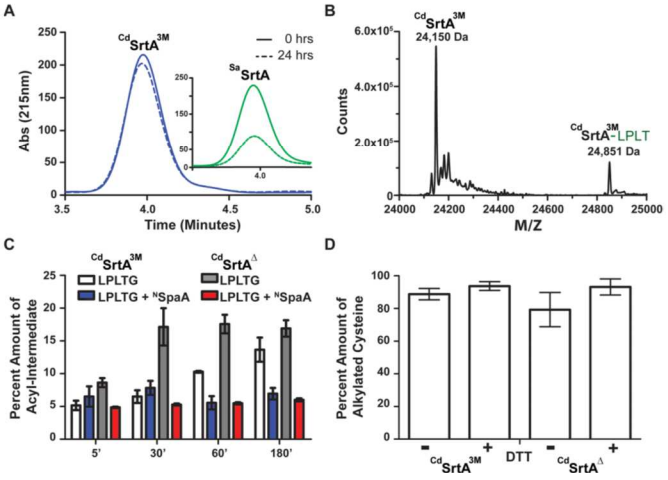


Figure 3. Characterization of ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$. (A) Representative reverse phase HPLC trace showing the change in the concentration of LPLTG peptide due to proteolysis by ^{Cd}SrtA^{3M} (blue) and ^{Sa}SrtA^{WT} (green) at 25 °C after 0 (solid lines) and 24 h (dashed lines). (B) Mass deconvolution of LC-MS data of ^{Cd}SrtA^{3M}. 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 ^{N}SpaA measured over a period of three hours. ^{Cd}SrtA^{3M} without ^{N}SpaA (white) in comparison to ^{Cd}SrtA$^{\Delta}$ without ^{N}SpaA (gray) shows the faster formation of the ^{Cd}SrtA$^{\Delta}$ acyl-intermediate. In contrast, both enzymes form these intermediates at similar rates when ^{N}SpaA is present (^{Cd}SrtA^{3M} (blue) and ^{Cd}SrtA$^{\Delta}$ (red)). (D) Data showing the amount of reduced active site C222 thiol in the ^{Cd}SrtA^{3M} and ^{Cd}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.

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. 2020, 31, 1624−1634

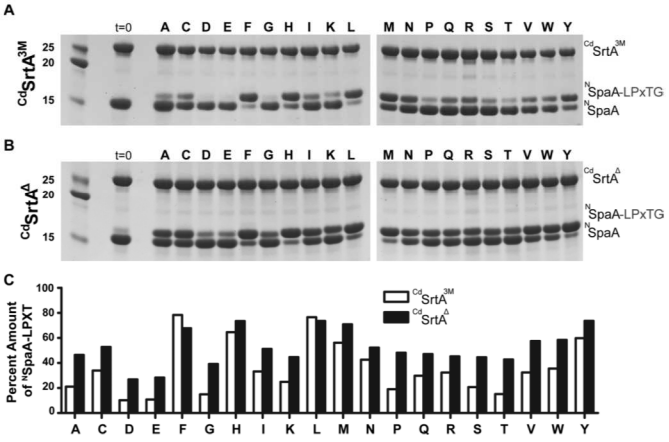


Figure 4. Role of the X residue in transpeptidation. (A) SDS-PAGE analysis of product formation when ^{Cd}SrtA^{3M} was reacted with ^{N}SpaA, and a series of LPXTG peptides in which the identity of the amino acid at the X position was systematically varied. Reactions (200 μM enzyme, 200 μM ^{N}SpaA, 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 ^{Cd}SrtA$^{\Delta}$ was used instead of ^{Cd}SrtA^{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 ^{N}SpaA converted to ^{N}SpaA-LPXT 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 ^{N}SpaA 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.[32] To determine if ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{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 μM 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 ^{Cd}SrtA$^{\Delta}$ forms the thioacyl intermediate more rapidly than ^{Cd}SrtA^{3M}, with ∼17 ± 2% of ^{Cd}SrtA$^{\Delta}$ joined to the peptide after 30 min, while it takes up to 3 h for ^{Cd}SrtA^{3M} to form similar levels of this reaction intermediate (Figure 3C). Interestingly, when similar time course experiments are performed in the presence of both ^{N}SpaA and LPLTG substrates, significantly lower amounts of thioacyl intermediate are observed for each type of enzyme (∼5% for both ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{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 ^{Cd}SrtA$^{\Delta}$ results from its ability to form the acyl-intermediate at a faster rate than ^{Cd}SrtA^{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.[23,33,34] The structures are consistent with detailed substrate specificity analyses of the class A ^{Sa}SrtA enzyme, which revealed that LPXTG sorting signals containing any amino acid at the X position can be used as substrates.[35] However, ^{Cd}SrtA 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).[33] Indeed, our prior studies of ^{Cd}SrtA^{3M} revealed that a leucine to alanine substitution at the X position of the LPLTG sorting signal slowed transpeptidation.[22] To investigate this issue in greater detail, the ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{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 ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M} enzymes were incubated with ^{N}SpaA 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, ^{Cd}SrtA$^{\Delta}$ is more active than ^{Cd}SrtA^{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 ^{Cd}SrtA is distinct from what has been observed for ^{Sa}SrtA, as similar peptide library studies have shown that after 24 h exposure all peptides in the library are processed to a similar extent by ^{Sa}SrtA.[35] 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 ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$.

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

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. 2020, 31, 1624−1634

the well-studied ^{Sa}SrtA sortase, three conserved residues in ^{Cd}SrtA are presumably required for catalysis: H160, C222, and R231 (Figure 1B).[36,37] 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 ^{Cd}SrtA^{3M} and ^{Cd}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, ^{Cd}SrtA^{3M} and ^{Cd}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 ^{Sa}SrtA transpeptidation reaction the analogous histidine, cysteine, and arginine residues are essential for catalysis *in vitro* (R197, C184, and H120 in the ^{Sa}SrtA sequence).[36] The cysteine thiol functions as a nucleophile, while the arginine guanidinium group (R197 in ^{Sa}SrtA) has been proposed to facilitate catalysis by stabilizing oxyanion tetrahedral intermediates.[23] Our finding that the C222A and R231A sortase variants are completely inactive is consistent with these residues having similar functions.[28] 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 ^{Cd}SrtA containing a H160A substitution retained some activity, since the analogous alteration in ^{Sa}SrtA disrupts transpeptidation *in vitro*.[36] In ^{Sa}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).[29,38] Whole cell studies have shown that ^{Cd}SrtA containing H160A are incapable of assembling surface pili.[28] The residual activity observed in the H160A enzyme suggests that residues in addition to H160 in ^{Cd}SrtA may facilitate these steps *in vitro*, albeit less efficiently.

We wondered whether the superior activity of ^{Cd}SrtA$^{\Delta}$ relative to ^{Cd}SrtA^{3M} 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.[39] 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 ^{Cd}SrtA$^{\Delta}$ and ^{Cd}SrtA^{3M} 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 ± 4 and 79 ± 10% of ^{Cd}SrtA^{3M} and ^{Cd}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 ± 3% and 93 ± 5% ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$ are reactive, respectively (Figure 3D). Thus, the C222 sulfhydryl in freshly purified ^{Cd}SrtA^{3M} and ^{Cd}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 +/−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 ^{Cd}SrtA$^{\Delta}$. Having defined substrate and reaction conditions that are optimal for activity, we directly compared the peptide labeling efficiencies of ^{Cd}SrtA^{3M} with the newer ^{Cd}SrtA$^{\Delta}$ enzyme. Consistent with our steady-state kinetic analyses, a temporal analysis using identical conditions clearly shows that ^{Cd}SrtA$^{\Delta}$ (blue trace) produces more cross-linked product than

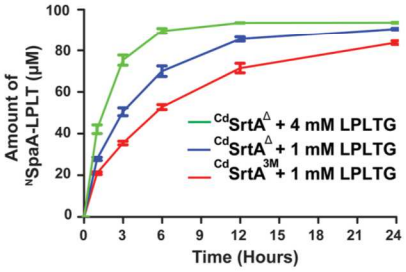


Figure 5. Labeling activity of ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$. Reactions containing each enzyme were performed using identical conditions: 100 μM enzyme, 100 μM ^{N}SpaA, and either 1 or 4 mM LPLTG peptide. ^{Cd}SrtA$^{\Delta}$ exhibits superior labeling activity at all measured time points.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

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

Comparison with the Prototypical ^{Sa}SrtA Sortase. ^{Sa}SrtA is the best-studied member of the sortase superfamily

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. 2020, 31, 1624−1634

and is routinely used to modify the N- and C-termini of proteins.[13,23] ^{Cd}SrtA and ^{Sa}SrtA catalyze mechanistically related transpeptidation reactions, but differ in the type of nucleophile that they use, i.e., a lysine ϵ-amine within ^{N}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 ^{Sa}SrtA joins the LPLTG peptide previously used in studies of ^{Cd}SrtA to its triglycine substrate (Gly_3), a reaction that creates a FELPTGGGG peptide product (Table 1).[29,40] The steady-state kinetic values measured from this analysis are generally consistent with published values that used a much shorter fluorogenic labeled peptide.[29] Our direct comparison reveals that ^{Sa}SrtA forms backbone peptide bonds ~40-times faster than ^{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 (^{Cd}SrtA$^{\Delta}$, 1.6 ± 0.3 mM; ^{Sa}SrtA, 7.33 ± 1.01 mM), and approximately ~10-fold lower K_M values for their respective amine nucleophiles (^{Cd}SrtA$^{\Delta}$, 160 ± 30 μM; ^{Sa}SrtA, 180 ± 10 μM).[29] In addition, based on prior studies by the McCafferty group and the studies presented here (Figure 3), for both ^{Sa}SrtA and ^{Cd}SrtA the first step of catalysis is rate limiting—the formation of the thioacyl intermediate (Scheme 1).[29,40] It is unclear why this step is so slow in both enzymes. However, it has been shown that ^{Sa}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.[30,41,42] Whether the active site of ^{Cd}SrtA also primarily exists in a catalytically dormant state remains to be determined. Startlingly, we found that ^{Cd}SrtA is quite inefficient at proteolyzing the LPLTG substrate, which is in stark contrast to previously reported studies of ^{Sa}SrtA that have shown that it catalyzes this reaction with a k_{cat} of 0.28 ± 0.02 s^{-1} (Figure 3A).[29] Presumably this difference originates from distinct active site features that enable ^{Sa}SrtA to use water as a nucleophile much more efficiently than ^{Cd}SrtA (Scheme 1). This idea is substantiated by our observation that the thioacyl intermediate in the ^{Cd}SrtA reaction forms to an appreciable extent in the absence of the nucleophile (Figure 3C). The reduced proteolytic activity of ^{Cd}SrtA is presumably advantageous, limiting the release of partially assembled pili from the bacterial cell surface. Similar to ^{Cd}SrtA, under certain conditions ^{Sa}SrtA can catalyze formation of a lysine isopeptide bond.[43−45] A thorough study by Dasgupta and colleagues demonstrated that ^{Sa}SrtA can join the sorting signal and lysine−containing peptides together via an isopeptide linkage.[44] 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 ^{Sa}SrtA does not discriminate between lysine and water nucleophiles. This is in contrast to ^{Cd}SrtA$^{\Delta}$ and ^{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 ^{Cd}SrtA-labeling system compared to ^{Sa}SrtA is that it requires the target protein to be expressed as a fusion to ^{N}SpaA to achieve maximum labeling. However, because ^{N}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 ^{Cd}SrtA, and discovered ^{Cd}SrtA$^{\Delta}$ which is ~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.[46] Although promising, these enzymes have not gained wide usage in site-specific protein labeling, presumably because of their penchant to catalyze spurious ligations.[46−49] In contrast, ^{Cd}SrtA exhibits a high level of specificity for its substrates, ligating peptides containing a LPXTG sequence to a specific lysine residue within the ^{N}SpaA domain. The molecular basis of specificity for the K190 side chain remains unknown, but presumably originates from protein−protein interactions between the ^{N}SpaA and ^{Cd}SrtA-LPLT thioacyl intermediate that function to properly position the nucleophile for catalysis. It may also arise from ^{N}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 ^{N}SpaA or the X residue within the LP(X)TG sorting signal. Further improvements in ^{Cd}SrtA$^{\Delta}$-mediated labeling activity may also be possible, as we estimate that current versions of the enzyme catalyze *in vitro* transpeptidations ~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 ^{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.[50−52] Finally, further improvements may also be achieved by obtaining a greater understanding of the process of substrate recognition and catalysis.

METHODS

Protein Reagents. Purified ^{Cd}SrtA^{3M} pilin sortase (residues N37−Q257 of SrtA from *C. diphtheriae*) and enzyme variants were expressed and purified as described previously.[22] Briefly, proteins were expressed from a pE-SUMO (Lifesensors) plasmid in *E. coli* BL21 (DE3) cells. The cells were grown up in LB supplemented with 500 μg/mL of Kanamycin at 37 °C until they reached an OD_{600} of ~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 8670*g* 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 22,720*g* for 50 min. Afterward, the cell lysate was purified via IMAC-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-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 ^{Cd}SrtA^{3M} variants were

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. 2020, 31, 1624−1634

created using standard molecular biology methods and confirmed by nucleotide sequencing. ^{N}SpaA (residues E30 to S195) and ^{Sa}SrtA (*S. aureus* Sortase A, residues Q60−K206) were purified as described previously.[28,53] All purified enzymes were stored at −20 °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 ^{Cd}SrtA catalyzed lysine−isopeptide bond formation. In this assay, the ^{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 μL volumes and contained 25 μM of pilin sortase (either ^{Cd}SrtA^{3M} or ^{Cd}SrtA$^{\Delta}$), 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 ^{N}SpaA (500 μM, 62.5 to 500 μM). All components were dissolved in buffer A. At these substrate concentrations, an estimated K_M for ^{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 °C and then flash-frozen with liquid N_2 and stored at −20 °C. The reactions were analyzed using a Phenomenex C4 column (5 μm, 4.6 × 150 mm) and with an initial dwell time of 3 min at 36% CH_3CN/0.1% TFA followed by a linear gradient from 36% to 46% CH_3CN/0.1% TFA for 10 min at 1 mL/min was applied. The column was subsequently flushed with high concentrations of CH_3CN/0.1% at 1 mL/min. ^{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 ^{Cd}SrtA was compared to ^{Sa}SrtA, which catalyzes a transpeptidation reaction that forms a backbone−backbone peptide bond between LPXTG and oligoglycine peptides. ^{Sa}SrtA transpeptidation activity was measured as described previously.[40] These reactions were performed in an identical manner to the ^{Cd}SrtA reaction described above, except that they contained 25 μM ^{Sa}SrtA instead of ^{Cd}SrtA and were supplemented with 10 mM calcium in Buffer A, and triglycine (Gly_3 peptide) (62.5 μM to 1 mM) instead of ^{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 μm, 4.6 × 150 mm) and a linear gradient from 26% to 30% CH_3CN/0.1% TFA over 8 min (1 mL/min). All HPLC experiments were performed on an Agilent 1100 HPLC. For both the ^{Sa}SrtA and ^{Cd}SrtA 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 ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$ 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 μL. The reactions contained: either ^{Cd}SrtA^{3M} or ^{Cd}SrtA$^{\Delta}$ (200 μM), DTT (5 mM), ^{N}SpaA (200 μM), and one of the FELPXTGGSG peptides (1 mM, Peptide 2.0). Transpeptidation reactions performed at 25 °C for 24 h, and then quenched by flash freezing in liquid N_2 and stored at −20 °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 ^{Sa}SrtA and ^{Cd}SrtA 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 μL: sortase (50 μM), LPLTG (500 μM), and DTT (5 mM). Reactions were incubated at 25 °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 μm, 4.6 × 150 mm) using a linear gradient from 26% to 30% CH_3CN/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. ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$ 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 °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.[54] Trypsin digestion was performed overnight at 37 °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 × 75 μm ID packed with PepMap RSLC C18, 2 μm 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 ^{N}SpaA were performed in buffer A and contained a total volume of 100 μL: sortase (25 μM), DTT (5 mM), LPLTG (1 mM), and ^{N}SpaA (250 μM) or buffer. All reactions were incubated at 25 °C and 10 μL time points were removed and frozen with liquid N_2 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 ^{Cd}SrtA-LPLT acyl intermediate cannot readily be separated from unmodified ^{Cd}SrtA by chromatography, its amount in the assay is estimated from the mass spectrometry data by assuming that the acyl-intermediate and unmodified enzyme

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. 2020, 31, 1624−1634

ionize to similar extents. Before being run on the LC-MS system, samples were then diluted with 90 μL of 200 mM L-tryptophan (internal standard) and then measured on a Zorbax 300SB-C3 (3.5 μm, 3.0 × 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 ^{N}SpaA has been verified, the mass of ^{N}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; Phone: (+1) 310 206 2334; Email: 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

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 AI52217 (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., Boutureira, 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) Samantaray, S., Marathe, U., Dasgupta, S., Nandicoori, V. K., and Roy, R. P. (2008) Peptideâ'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., Kattke, 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., Samantaray, 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.

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. 2020, 31, 1624−1634

(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.

https://dx.doi.org/10.1021/acs.bioconjchem.0c00163
Bioconjugate Chem. 2020, 31, 1624−1634

Chapter 4

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

Sortase-assembled pili in *Corynebacterium diphtheriae* are built via a latching mechanism

Scott A. McConnell[1], Rachel A. McAllister[1], Brendan R. Amer, Brendan Mahoney[1], Christopher K. Sue[1], Chungyu Chang[2], Hung Ton-That[2,3] and Robert T. Clubb[1,3†]

[1]Department of Chemistry and Biochemistry and the UCLA-DOE Institute of Genomics and Proteomics, University of California, Los Angeles, CA, USA

[2]Division of Oral Biology and Medicine, University of California, Los Angeles, Los Angeles, CA, USA

[3]Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA, USA

[†]To whom correspondence should be addressed:
Robert Clubb
Department of Chemistry and Biochemistry
University of California, Los Angeles
Los Angeles, CA 90095, USA.
E-mail: rclubb@mbi.ucla.edu;
Tel. (+1) 310-206-2334;

Running Title: Solution structure of the SpaA pilus

Keywords: *Corynebacterium diphtheriae*, pili, sortase, lysine isopeptide bond, Gram-positive bacteria, NMR, Small angle x-ray scattering
Author contributions: S.A.M., H.T-T. and R.T.C. designed research; S.A.M., R.A.M. and C.C. performed research; and all authors analyzed data; and S.A.M. and R.T.C. wrote the paper.

4.1 Overview

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 (1-5 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 inter-pilin crosslinks stabilize pilus structure is 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. The pilus structure and sortase-catalyzed assembly mechanism presented here are likely conserved in Gram-positive bacteria that display pili.

This chapter is written as a manuscript to be published. Much of the work described here is done in collaboration with Professor Hung Ton-That's lab at UCLA School of Dentistry.

4.2 Introduction

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[1,3,8,14]. Gram-positive bacteria display very thin (1-5 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[2,8,10–13].

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[8,11,13,16]. The SpaA pilus mediates adherence to the pharyngeal epithelium, and is formed from 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 (^{Cd}SrtA) assembles the pilus by catalyzing a repetitive, irreversible transpeptidation reaction that covalently links the pilin subunits together via an isopeptide bond. The shaft of the pilus is formed by ~100-250 crosslinked SpaA pilins[16]. ^{Cd}SrtA-catalyzed SpaA polymerization begins when SpaA pre-pilin proteins containing a 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). ^{Cd}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 ^{Cd}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 ^{Cd}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 ^{Cd}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 ^{Cd}SrtF[6]. Pilus biogenesis is thought to occur within "pilusosomes" on the cell surface, at which pilin substrates and pilus-specific sortases co-localize to facilitate rapid polymerization[18].

Despite their importance in bacterial physiology and pathogenesis, only structures of isolated, non-crosslinked 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 CryoEM 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 intra-domain 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 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 inter-pilin 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 inter-pilin crosslinking.

4.3 Results

4.3.1 NMR structure of the crosslinked ^{N}SpaA-signal complex.

To learn how Gram-positive pili are stabilized by inter-pilin lysine-isopeptide bonds, we examined how these crosslinks ligate SpaA pilins together to construct the shaft of the *C. diphtheriae* SpaA pilus[10,16]. The SpaA shaft pilin contains three autonomously folded domains, N-terminal (^{N}SpaA, residues 53-195), middle (^{M}SpaA, residues 195-349) and C-terminal (^{C}SpaA, residues 350-500) domains[23]. SpaA pilins are joined together via inter-pilin crosslink bonds that connect the ^{N}SpaA and ^{C}SpaA domains; a lysine-isopeptide bond links the side chain ε-amine group of K190 within ^{N}SpaA to the carbonyl group of the T494 residue present in a LPLTG sorting signal sequence that immediately follows ^{C}SpaA in the primary sequence (**Fig. 1A**). Previously, we demonstrated that a mutationally activated ^{Cd}SrtA enzyme covalently crosslinks peptides containing the LPLTG sorting signal sequence to ^{N}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 ^{N}SpaA-signal complex, in which the K190 side chain in ^{N}SpaA is joined via an isopeptide bond to the threonine residue in a sorting signal peptide (KNAGFELPLT peptide that corresponds to residue K485 to T494 in ^{C}SpaA) (**Fig. S1**). Heteronuclear multidimensional NMR spectroscopy was then used to determine the atomic structure of the complex using a total 2076 experimental restraints, including 66 intermolecular NOE distance restraints (**Table S1**, **Fig. S3**). The structure of the complex is well defined by the NMR data, as the backbone and heavy atom coordinates of residues T3-Q192 and G488-T494 in the ensemble can be superimposed to the average structure with a root mean square deviation (rmsd) of 0.47 ± 0.09 and 0.82 ± 0.07 Å, respectively (**Fig. 1C**).

^{N}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 a 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 ^{N}SpaA and bury ~630 Å^2 of solvent exposed surface area (**Fig. 1B**). A detailed summary of these interactions is provided in **Fig. S5**. The C-terminal T494 residue in the signal is joined via an isopeptide linkage to the sidechain 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 amongst proteins that contain the pilin motif (**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 sidechain 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 inter-pilin interface in the shaft of the SpaA pilus.

4.3.2 Crosslinking triggers the closure of a stabilizing latch over the inter-pilin linkage.

A comparison of the structure of the ^{N}SpaA-signal complex with a previously determined 1.6-Å crystal structure of unmodified SpaA protein reveals striking conformational differences **(Fig. 4.2D)**. While the apo- and complexed-forms of the ^{N}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 EF 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-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-^{N}SpaA and acquired ^{1}H-^{15}N steady-state NOE relaxation data for apo-^{N}SpaA and the ^{N}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. 4.2A, B)**. Interestingly, the ^{1}H-^{15}N steady-state NOE data reveal that signal attachment significantly retards motions in the AB loop, as residues M63-G86 in apo-^{N}SpaA exhibit small magnitude steady-state NOEs indicative of high mobility, whereas in the ^{N}SpaA-LPLT complex they are rigid with values of ~0.8 **(Fig. 4.2C)**. 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 contacts the N-terminal end of the sorting signal near the inter-pilin interface.

Limited proteolysis experiments of apo-^{N}SpaA and the ^{N}SpaA-LPLT complex indicate that the protein in the complex is ~42% more resistant to proteolytic degradation after 24 h **(Fig. S6A)**. The ^{N}SpaA-signal complex is also slightly more thermostable based on differential scanning fluorimetry experiments (its $\Delta_u G°$ increases by ~0.9 kJ) **(Fig. S6B)**[27,28]. 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 ^{N}SpaA.

4.3.3 Solution structure of the inter-pilin SpaA-SpaA interface.

To gain insight into the structure and dynamics of the inter-pilin interface that is repeated to build the SpaA pilus, we used the activated ^{Cd}SrtA enzyme to generate a crosslinked ^{C}SpaA-^{N}SpaA dimer (M.W. 31 kDa); in the dimer ^{N}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 ^{C}SpaA domain **(Fig. 4.1A)**. NMR spectra were acquired using samples of the dimer in which either the ^{C}SpaA or ^{N}SpaA domains were selectively labeled with nitrogen-15. A comparison with the

corresponding spectra of the isolated domains reveals that crosslinking causes substantial chemical shift changes, suggesting that the domains pack against one another in the dimer **(Fig. S7A, B)**. This is substantiated by molecular correlation time (τ_c) measurements using NMR ^{15}N relaxation data; as the τ_c of the ^{C}SpaA-^{N}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 re-orient (the τ_c values of the isolated ^{N}SpaA and ^{C}SpaA domains are 8.7 and 9.5 ns, respectively) **(Fig. S7D)**. Small angle X-ray scattering (SAXS) data of the crosslinked dimer also indicate that it is generally inflexible as evidenced by the distance distribution **(Fig. S8A)** and normalized Kratky plots of the data **(Fig. S8B)**. 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 ^{C}SpaA-^{N}SpaA dimer was determined using an integrated approach that employed SAXS, NMR and crystallographic data. An initial model of the ^{C}SpaA-^{N}SpaA complex was constructed using the crystal and NMR structures of ^{C}SpaA (PDB:3HR6) and ^{N}SpaA-LPLT (this work), respectively. SAXS data were then employed to drive multi-state rigid-body modeling of the complex using the MultiFOXS approach[29]. To account for potential domain-domain flexibility, 10,000 models were calculated and residues in the C-terminal tail of ^{C}SpaA (K483-A488) that bridge the globular domain and the sorting signal were allowed to move freely during the calculations (see Methods). A single state model of the ^{C}SpaA-^{N}SpaA dimer best fits the SAXS data, further indicating that the domains are arranged in a defined orientation **(Fig. S8C)**. In the structure, the FG loop in ^{N}SpaA is inserted between the UV and PQ loops in ^{C}SpaA, thereby positioning the sorting signal following ^{C}SpaA within ^{N}SpaA's peptide binding groove **(Fig. 4.3A)**. The protein-protein interface buries 1,270 Å^2 of solvent exposed surface area[30] and is further stabilized by interactions between the N-terminal portion of helix α1 in ^{N}SpaA and UV loop in ^{C}SpaA. This packing arrangement explains why the W181 residue within the WxxxVxVYPK motif is conserved, as its indole side chain appears to

play a key structural role in stabilizing SpaA-SpaA interface by packing into a hydrophobic surface formed by residues A450 and Y453 in ^{C}SpaA's UV loop. Stabilizing electrostatic interactions surround this nonpolar interface **(Fig. 4.3B**, top**)**. On one side, the negatively charged D179 side chain in ^{N}SpaA's FG loop is packed against a cationic surface on ^{C}SpaA formed by residues K485 (signal peptide) and R374 (PQ loop) **(Fig. 4.3B**, bottom left**)**, while on the other side, hydrophilic interactions occur between residues at the N-terminal end of α1 helix in ^{N}SpaA (T99, T100 and Q101) and residues located in ^{C}SpaA's UV loop (K483, K484, E454) **(Fig. 4.3B**, bottom right**)**. Notably, the SAXS-derived solution structure of the ^{C}SpaA-^{N}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 data (described in **Fig. S9**).

4.3.4 The sorting signal must be partially transferred to SpaA to initiate crosslinking.

Guided by the solution structures of the ^{N}SpaA-signal complex and the crosslinked ^{C}SpaA-^{N}SpaA dimer, we employed *in vitro* crosslinking and cellular assays to probe the mechanism of pilus biogenesis. Initially, ^{N}SpaA proteins containing amino acid substitutions at conserved sites within the SpaA-SpaA interface were tested for their ability to serve as substrates for ^{Cd}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, 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 HPLC-based assay that quantitatively measures the steady-state kinetics of crosslinking[26] (**Fig. 4B, S10**). 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 ε-amine group in ^{N}SpaA attacks the thioacyl linked ^{Cd}SrtA-SpaA intermediate. The V187D and Y188G variants exhibit the largest defects in transpeptidation and alter sidechains 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 form 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 reduce the rate of transpeptidation by increasing the K_M (A170S, W181A). This suggests that when the thioacyl ^{Cd}SrtA-SpaA intermediate formed in the first step of catalysis encounters ^{N}SpaA, the sorting signal bound to the ^{Cd}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 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 precursors. To corroborate the fractionation results, the same set of strains was analyzed by immuno-electron 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). 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 ^{Cd}SrtA can only crosslink sorting signals to K190 when it is housed in a structurally intact ^{N}SpaA domain[24].

4.4 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}SpaA-^{N}SpaA interface that is repeated to build the shaft of the *C. diphtheriae* SpaA pilus (**Fig. 1A**). The structure of ^{N}SpaA covalently attached to the ^{C}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}SpaA domain to model the solution structure of the inter-pilin interface that is formed by sortase crosslinking of the ^{C}SpaA and ^{N}SpaA domains. This work reveals that sortase crosslinking immobilizes the pilin subunits, triggering the formation of an extensive inter-pilin interface in which the sorting signal following ^{C}SpaA is inserted into a non-polar groove on ^{N}SpaA (**Fig. 3A**). Residues within the conserved WxxxVxVYPK 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}SpaA-^{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 inter-pilin isopeptide bonds positioned on opposite faces of the polymer and each crosslinked SpaA-SpaA unit forming a “S” shape because of a ~140° kink at the ^{C}SpaA-^{N}SpaA junction. The SpaA pilus and other Gram-positive pili are presumably flexible, as they appear as non-linear, hair-like structures in transmission electron micrographs (**Fig. 4D**)[16,22,33]. This flexibility likely originates from intra-pilin 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 allows 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 inter-pilin ^{C}SpaA-^{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}SpaA-^{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 (**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]. Our model of the pilus explains how it can withstand these forces, as the sortase-installed T494-K190 crosslink between adjacent SpaA pilins effectively bypasses the entirety of the ^{N}SpaA domain (**Figs. 1D, 3B**). Thus, as originally predicted by single molecule pulling experiments, the load bearing spine of the pilus only goes through the middle (^{M}SpaA) and C-terminal (^{C}SpaA) domains, which contain spontaneously forming intra-pilin isopeptide linkages that can bear large pulling forces of ~525 pN[21,23,45].

To assemble pili, ^{Cd}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. **Figure 5** shows a working model of lysine-isopeptide bond forming step catalyzed by ^{Cd}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 ^{C}SpaA domain (green) to the K190 amine group in ^{N}SpaA (blue) (**Fig. 5A**). Presumably, two thioacyl linked enzyme-substrate intermediates mediate this interaction and are tethered to the membrane via their respective ^{Cd}SrtA enzymes[10,18]. The growing $(SpaA)_n$ polymer is housed in a ^{Cd}SrtA-$(SpaA)_n$ 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 ^{Cd}SrtA's active site cysteine. New SpaA proteins enter the reaction as similarly bonded thioacyl enzyme-substrate intermediates (^{Cd}SrtA-SpaA)

after their sorting signals are nucleophically attacked by the enzyme's active site cysteine residue (not shown). A single protein is then added to the shaft when the reaction intermediates form a ternary complex that enables the K190 nucleophile on ^{Cd}SrtA-SpaA to resolve the thioacyl bond in the ^{Cd}SrtA-(SpaA)$_n$. During this process the growing pilus is transferred from one enzyme to another and is then poised to react with a new ^{Cd}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 ^{N}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 ^{N}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 ^{Cd}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 ^{N}SpaA in order to activate K190 for catalysis. This is because several ^{N}SpaA variants that alter contacts to sorting signal residues but are positioned distal to the site of crosslinking slow transpeptidation by increasing the enzyme's K_M for ^{N}SpaA (**Fig. 4B**, W181A/A170S). Partial signal transfer would act as a zipper, juxtaposing K190 and the ^{Cd}SrtA-(SpaA)$_n$ 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 ^{N}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, non-nucleophilic state at physiological pH (its calculated pKa 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. The latch mechanism is likely a conserved feature of sortase-catalyzed pilus biogenesis reactions, as structures of 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 novel antibiotics that work by inhibiting pilus assembly.

Acknowledgements:

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 National Institutes of Health 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 DOE (DE-FC03-02ER63421). SAXS data was collected at SIBYLS which is supported by the DOE-BER IDAT DE-AC02-05CH11231 and NIGMS ALS-ENABLE (P30 GM124169 and S10OD018483).

4.5 Methods

4.5.1 Production of the ^{N}SpaA -LPLT complex and ^{C}SpaA-^{N}SpaA dimer.

The amino-terminal domain of the SpaA protein from *C. diphtheriae* (^{N}SpaA, residues E53-S195) was produced and purified as described previously[24,25]. The ^{N}SpaA-LPLT complex was generated by enzymatic covalent ligation of synthetic peptide to [^{13}C,^{15}N] ^{N}SpaA using a previously described activated variant of ^{Cd}SrtA[24,26] (^{Cd}SrtA^{3M}, residues N37-Q257, containing D81G/W83G/N85A mutations). Complete modification of ^{N}SpaA to its cognate sorting signal peptide occurred after incubation of 100 μM ^{Cd}SrtA^{3M}-His_6, 100 μM ^{N}SpaA and 1 mM synthetic peptide derived from the SpaA sorting signal motif (KNAGFELPLTGGSGRI) (Peptide2.0) in modification buffer (50 mM Tris pH 8.0, 300 mM NaCl, 5 mM TCEP) for 24 h at room temperature. Sortase and unreacted peptide was removed from the reaction by HisPure Co^{2+} purification and subsequent concentration by Amicon spin filters with a 10 kD MWCO and complex formation was confirmed by MALDI-TOF MS as well as SDS-PAGE analysis. The sample was exchanged into NMR buffer (50 mM NaH_2PO_4 pH 6.0, 100 mM NaCl, 8% D2O, 0.01% NaN_3) and diluted to a concentration of 1.2 mM for NMR studies. Subsequently, the sample was lyophilized and re-dissolved into 100% deuterated NMR buffer for additional NMR studies. The crosslinked ^{C}SpaA-^{N}SpaA dimer was prepared as described for the ^{N}SpaA-LPLT complex, but employed 300 μM ^{C}SpaA (SpaA, residues R350-I500) instead of peptide. The dimer was purified from the reaction components using HisPure Co^{2+} IMAC (Thermo Scientific) and size exclusion chromatography (Superdex Increase 10/300 GL, GE Healthcare). For NMR samples of the dimer, either ^{N}SpaA or ^{C}SpaA was uniformly labeled with nitrogen-15 and the other component was expressed in natural abundance nitrogen-14 media.

4.5.2 Immuno-electron microscopy and cell-fractionation studies.

Cells of the *C. diphtheriae* Δ*spaA* mutant expressing wild-type SpaA or individual SpaA mutants from a plasmid were grown on HIB agar plates. A full loop of cells were collected, suspended in,

and washed with PBS buffer. A drop (7 µL) of bacterial suspension in PBS was placed on the carbon-coated nickel grids for immunogold labelling as previously reported[25,54]. 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_{600}=0.5-0.6). Cells were normalized to OD_{600} at 1 before harvest. The harvested cells were fractionated into medium (S) and cell-wall associated (W) fractions as described previously[7]. The samples were separated by 3-14% gradient SDS-PAGE and analyzed by immunoblotting with α-SpaA antibody.

4.5.3 NMR structural determination and relaxation measurements.

NMR spectra were collected at 298K on Bruker Avance III HD 600-MHz and Avance NEO 800-MHz spectrometers equipped with triple resonance probes. NMR data were processed with NMRPipe[55], and analyzed using CARA[56] (version 1.8.4), XIPP[57] (version 1.19.6 p0), and NMRFAM-Sparky[58]. ^{1}H, ^{13}C, and ^{15}N protein chemical shift 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. Chemical shifts of the unlabeled (natural abundance) sorting signal peptide were assigned using: 2D (F1,F2) ^{13}C-filtered NOESY and 2D (F1) ^{13}C,^{15}N-filtered TOCSY experiments[59].

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 3D (F1) ^{13}C, ^{15}N -filtered (F2) ^{13}C -edited NOESY-HSQC and (F1) ^{13}C, 15 N-filtered (F2) ^{15}N-edited NOESY-HSQC, and 2D (F1) ^{13}C -filtered NOESY spectra. ψ and φ dihedral restraints were obtained from secondary ^{13}C

chemical shifts as calculated by TALOS-N[60] and $^{3}J_{HN\alpha}$ measurements from the HNHA spectrum. Additional ψ angle restraints were obtained from analysis of ^{15}N-edited NOESY spectrum. Rotamer assignments and χ_1 angle restraints for β-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[61,62]. 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 UNIO'10[63,64]. The NOESY data were then manually inspected using he program Xipp[57] 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 110 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 ^{N}SpaA-LPLT and have been deposited in the Protein Data Bank. The programs MOL-MOL[65] and PyMOL[66] 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 ^{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[58] and included: ^{15}N longitudinal relaxation rates (R_1), transverse relaxation rates (R_2), and {^{1}H}-^{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 ^{N}SpaA-signal complexes, respectively. For inclusion in the calculations data from each residue must meet the following criteria: isolated ^{1}H-^{15}N cross peaks and {^{1}H}-^{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 ^{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:

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

Where M= molecular weight of the protein, ρ = the average density for proteins (1.37 g/cm^3), N_a= Avogadro's number and r_w= hydration radius (1.6-3.2 Å)[67].

After calculating the hydrodynamic radius of the protein of interest, and assuming a spherical approximation, the Stokes' law equation was used to calculation an expected rotational correlation time:

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

Where η= the viscosity of the solvent, r = hydrodynamic radius (calculated above), k = Boltzmann constant, and T = acquisition temperature. Experimental values of τ_c for the complex and isolated domains were estimated using the ratio of T_1 and T_2 NMR relaxation rates[68].

4.5.4 Differential scanning fluorimetry (DSF) and protease sensitivity measurements.

The melting temperature and thermodynamic parameters were extracted from DSF data by a method described previously[30]. Briefly, ^{N}SpaA proteins were diluted to 50 μM in assay buffer (50 mM Tris pH 8.0, 300 mM NaCl), supplemented with 15X SYPRO Orange (Sigma) at a total volume of 20 μL. Thermal denaturation reactions were run on a CFX Connect qPCR system (BioRad). A heating rate of 0.2°C/min was employed from 4 to 95 °C and fluorescence measurements (excitation at 525 ± 10 nm, detection at 570 ± 10 nm) were acquired after each 0.5° 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 ^{N}SpaA or the ^{N}SpaA-signal complex was dissolved in assay buffer (50 mM Tris pH 8.0, 300 mM NaCl) at a concentration of 20 μM in a volume of 100 μL. Trypsin protease stock solutions were created as described by the manufacturer (Sigma). 0-200 ng of trypsin protease was added to the reactions and incubated at 37°C. Samples from each reaction were taken after 6 and 24 hours, separated by SDS-PAGE, and analyzed by densitometry.

4.5.5 Quantitative transpeptidation measurements.

The gel-based fluorescence assays were performed at room temperature as previously described[24,26]. All reactions included 25 μM ^{Cd}SrtA$^{\Delta}$, 250 μM fluorescent sorting signal peptide (FITC-KNAGFELPLTGGSGRI), and 25 μM wild-type or variant ^{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_2O and fluorescence data was acquired with a Pharos FX gel imager (BioRad). Fluorescein isothiocyanate (FITC) was detected by excitation with a 488 nm laser line and detection with a 515-545 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 high performance liquid chromatography (RP-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 μM ^{Cd}SrtA Δ78-87, 1 mM sorting signal peptide (FELPLTGGSG), 5 mM DTT and 50-300 μM wild-type or mutant ^{N}SpaA. Different variants were incubated either 3.5 h (wild-type, L62D, ABΔ15, A170S) or 16 h (H60A, W181A, V187D, Y188G) depending on reactivity. The reactions were stopped by flash freezing in liquid nitrogen after incubation. Each reaction condition was run in duplicate. 25 μL of each reaction was injected onto a Water Symmetry 300 C4 HPLC column (4.6x150 mm, 5 μm particle size). Proteins were eluted by applying a gradient from 35 to 46% acetonitrile (with 0.1% trifluoroacetic acid) over 12 minutes 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 was plotted as Lineweaver-Burke in order to calculate kinetic parameters for each ^{N}SpaA variant.

4.5.6 Small-angle X-ray scattering (SAXS) analysis.

Scattering data were generated at the SIBYLS beamline (Advanced Light Source (ALS), Lawrence Berkeley National Laboratory)[69]. Purified ^{C}SpaA-^{N}SpaA complex (10 mg/mL) was dissolved in size exclusion chromatography (SEC) buffer (50 mM NaH_2PO_4, pH 6.5, 100 mM NaCl, 0.01% NaN_3) 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 of frames corresponding to the elution of the complex. Radius of gyration (R_g) and maximal particle dimension (D_{max}) were calculated by Guiner analysis (BioXTAS RAW[70]) 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 Å and 51.0 Å, respectively. From the distance distribution function, the D_{max} and Porod volume were calculated to be 8.7 nm and $3.72x10^4$ Å^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_{max}

of 19.8 Å and 5.1 nm, respectively ($R_g \approx 3\sqrt[3]{n}$ and $D_{max} \approx 2.6R_g$)[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[71] 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_{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 ^{C}SpaA-^{N}SpaA isopeptide dimer starting structure by merging the crystal and NMR coordinates of ^{C}SpaA (PDB:3HR6) and ^{N}SpaA-LPLT (this work), respectively, into a single coordinate file using PyMOL[66]. The coordinates were energy minimized in GROMACS[72] 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 (RRT) search algorithm in the Integrative Modeling Platform (IMP) software package[73]. A SAXS profile was then generated for each model using the FOXS method and the best scoring multistate models were enumerated. The MultiFOXS algorithm predicted two single state models, one with a significantly better fit to the experimental SAXS data (χ^2=0.89, 1.24). For subsequent analysis, we chose the model with the superior χ^2 value. No multi-state models were predicted, ruling out the possibility that the complex consists of an ensemble of conformations in solution.

4.6 Figures and Table

Figure 4.1 – An isopeptide bond between a C-terminal cell wall sorting signal peptide and reactive lysine in the N-terminal domain of pilin protomers forms the linkage SpaA molecules. 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, ^{C}SpaA-^{N}SpaA complex and ^{N}SpaA-signal, are boxed in grey dashed lines and solid black lines, respectively. B) The 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-peptide complex are displayed. The backbone of the NTD domain is represented by blue ribbons. The five C-terminal 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 in the energy minimized average NMR complex structure. E) An expanded view of the peptide binding interface shows how the peptide is bound in the cleft of ^{N}SpaA. Residues on SpaA exhibiting intermolecular NOEs to the peptide are shown as sticks. The AB loop is colored purple and slightly transparent. Interacting residues in the core of the domain and within the AB loop are colored yellow and pink, respectively.

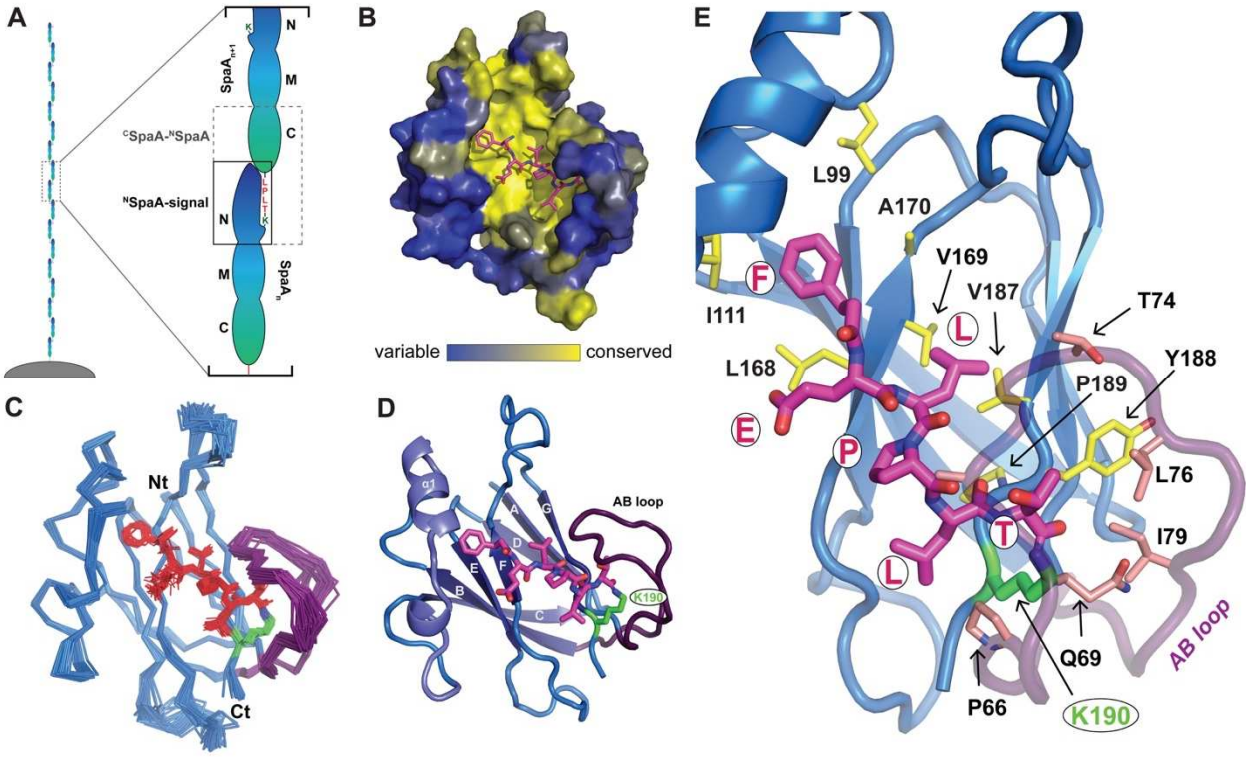

A
B
C
D
E
SpaA$_{n+1}$
^{C}SpaA-^{N}SpaA
^{N}SpaA-signal
SpaA$_n$
N
M
C
K
variable
conserved
Nt
Ct
AB loop
K190
L99
A170
V169
V187
T74
I111
L168
Y188
P189
L76
I79
Q69
P66
F
L
E
P
T

Table 4.1 – Structural statistics of the solution structure of ^{N}SpaA-signal complex.

	$\langle SA \rangle^a$	$\langle \overline{SA} \rangle^a$
Root mean square deviations		
NOE interproton distance restraints (Å) [b]		
All (2070)	0.027 ± 0.002	0.040
Intermolecular (66)	0.017 ± 0.006	0.019
Dihedral angles restraints (°)[c] (314)	0.528 ± 0.058	0.752
$^{3}J_{HN}\alpha$ coupling constants (Hz) (91)	0.534 ± 0.031	0.762
Secondary ^{13}C shifts (p.p.m.)		
$^{13}C_\alpha$ (125)	1.067 ± 0.184	1.041
$^{13}C_\beta$ (125)	1.414 ± 0.185	1.374
Deviation from idealized covalent geometry		
bonds (Å)	0.011 ± 0.00002	0.004
angles (°)	0.673 ± 0.006	0.530
impropers (°)	0.312 ± 0.009	0.517
PROCHECK results (%) [d]		
most favorable region	77.8 ± 1.4	81.1
additionally allowed region	20.7 ± 1.5	18.0
generously allowed region	1.5 ± 0.5	0.9
disallowed region	0.0 ± 0.0	0.0
Coordinate Precision (Å) [e]		
Protein backbone	0.47 ± 0.09	
Protein heavy atoms	0.82 ± 0.07	

[a] $\langle SA \rangle^a$ represents an ensemble of the 40 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

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

[c] The experimental dihedral angle restraints were as follows: 133 ϕ, 135 ψ, and 46 χ_1 angular restraints

[d] PROCHECK-NMR[1] data includes residues 52-195 of ^{N}SpaA and residues 488-494 of the signal peptide. For the structured regions of the protein and peptide, 98 ± 1 % of the residues were in the favored or allowed regions of the Ramachandran plot

[e] The coordinate precision is defined as the average atomic root mean square deviation (rmsd) of the 40 individual SA structures and their mean coordinates. These values are for residues 54-192 of ^{N}SpaA and residues 488-494 in the signal peptide. Backbone atoms refers to the N, C^α, and C′ atoms

Figure 4.2 – The AB loop undergoes a disordered-to-ordered transition during following crosslinking. A) ^{1}H-^{15}N HSQC correlation spectra of apo-^{N}SpaA (green) and ^{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 for each residue are plotted with respect to primary sequence. CSPs are binned into >0.75 ppm, 0.75-0.5 ppm, 0.5-0.25 ppm, and 0.1-0.25 ppm (indicated by red dotted lines of increasing transparency). C) Heteronuclear NOE data is 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 α1 helix (blue) highlighted with cartoon representations and the signal peptide shown as magenta sticks. The apo-^{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.

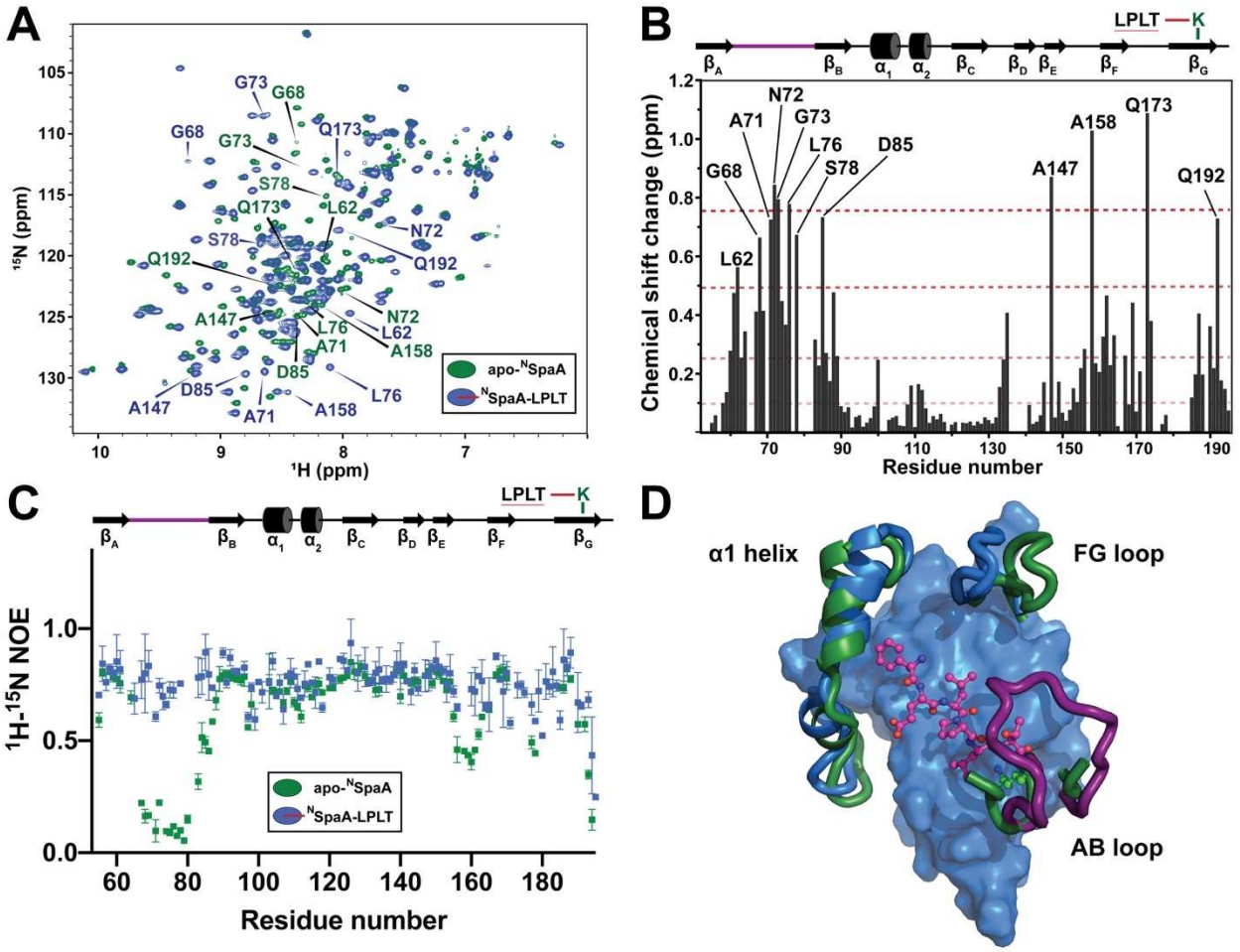

A
B
C
D
LPLT — K
apo-NSpaA
NSpaA-LPLT
Chemical shift change (ppm)
Residue number
1H-15N NOE
α1 helix
FG loop
AB loop

Figure 4.3 – SAXS structure of the SpaA-SpaA junction. A) The best fit single-conformation rigid body model from the MultiFOXS calculation (^{C}SpaA - green, ^{N}SpaA- blue). B) Details of the interface formed by the solution structure of the ^{C}SpaA-^{N}SpaA complex. In the top panel, the two domains and the interacting loops are shown as surface and cartoon representations, respectively. The two bottom panels depict expanded views of two interaction faces with ^{C}SpaA mediated by the FG loop and α1 helix of ^{N}SpaA. C) GASBOR *ab initio* model from solution scattering, shown as grey spheres. The rigid body model calculated by MultiFOXS is aligned with the GASBOR model and depicted in ribbon representation. Bottom and right shapes are the same GASBOR models rotated counterclockwise by 90 degrees 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).

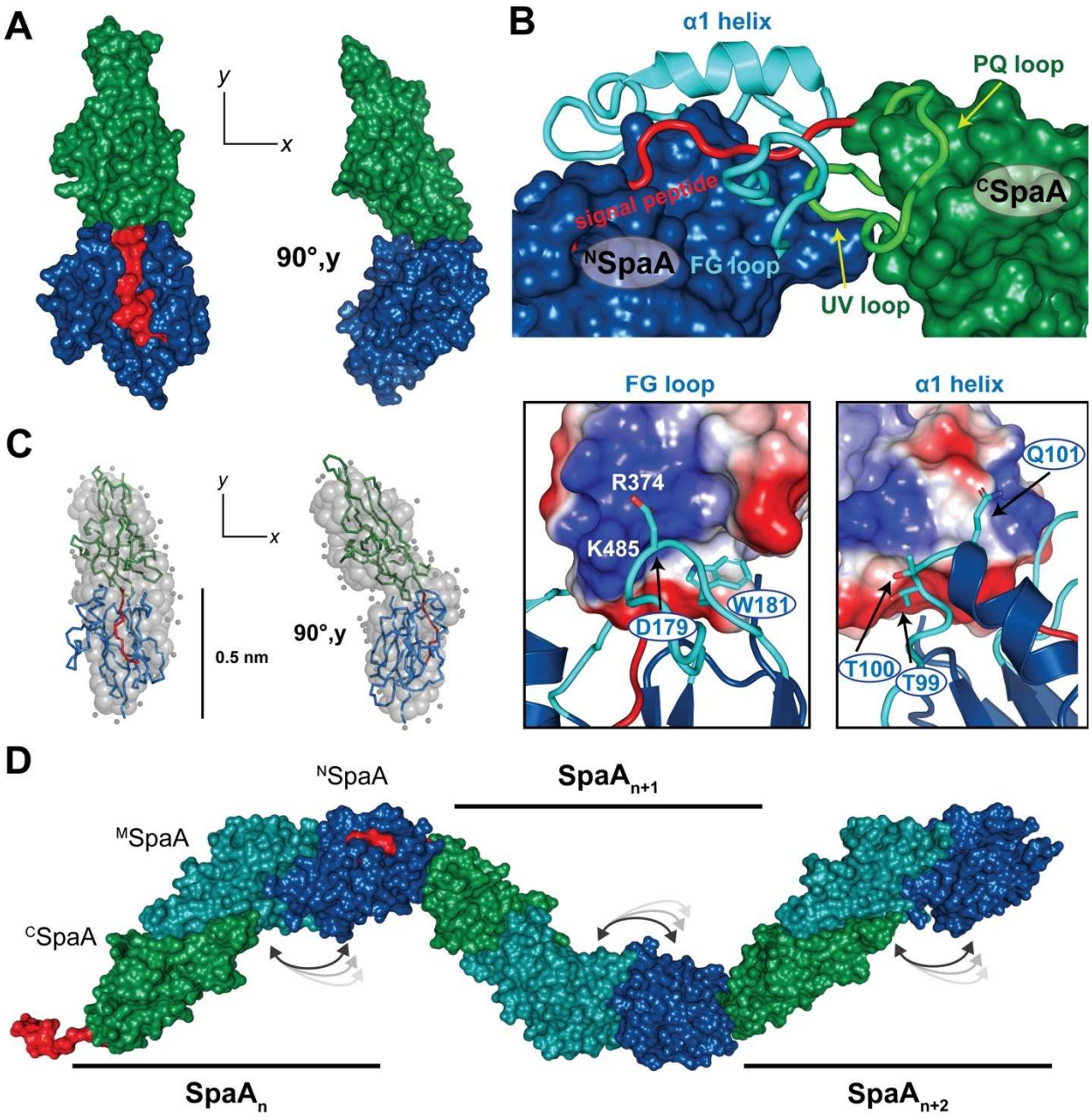

A
y
x
90°,y
B
α1 helix
PQ loop
signal peptide
NSpaA
FG loop
UV loop
CSpaA
FG loop
α1 helix
R374
K485
W181
D179
Q101
T100
T99
C
y
x
0.5 nm
90°,y
D
NSpaA
SpaAn+1
MSpaA
CSpaA
SpaAn
SpaAn+2

Figure 4.4 – *In vitro* and *in vivo* validation of key residues on the SpaA acceptor domain. A) Gel fluorescence assay to rapidly screen a library of ^{N}SpaA variants. The top panel shows an SDS-PAGE gel visualized by FITC fluorescence, indicating the presence of ^{Cd}SrtA (top band) or ^{N}SpaA (bottom band) conjugated to fluorescent peptide. The bottom panel shows the same SDS-PAGE gel visualized by Coomassie staining, in order to visualize the total protein composition of each lane. Peptide-labeled ^{N}SpaA variants typically have slightly lower electrophoretic mobility than the corresponding apo-^{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 ^{N}SpaA. C) Cells of the *C. diphtheriae* Δ*spaA* mutant expressing wild-type SpaA or individual SpaA mutants from a plasmid 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 (α-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, stained with α-SpaA, followed by IgG-conjugated 12-nm gold particles and 1% uranyl acetate prior to electron microscopy; scale bars of 0.5 μm.

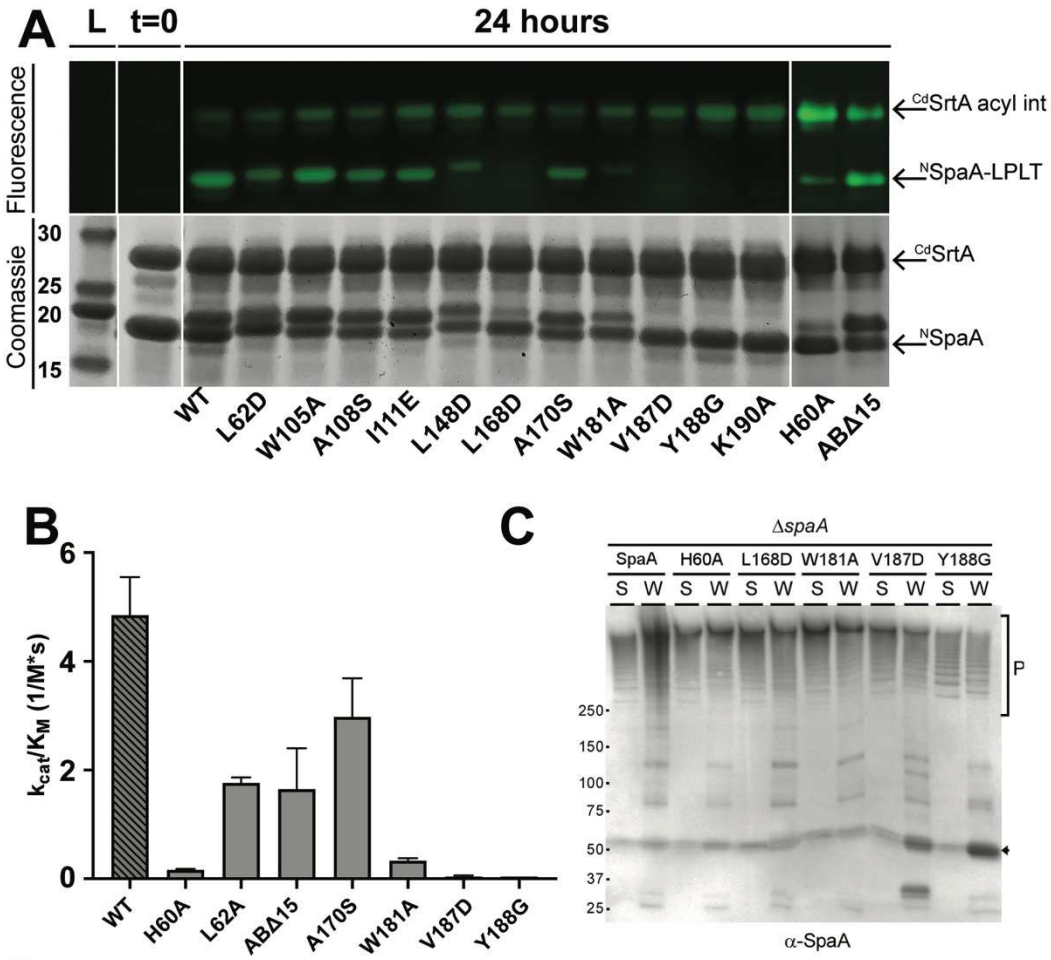
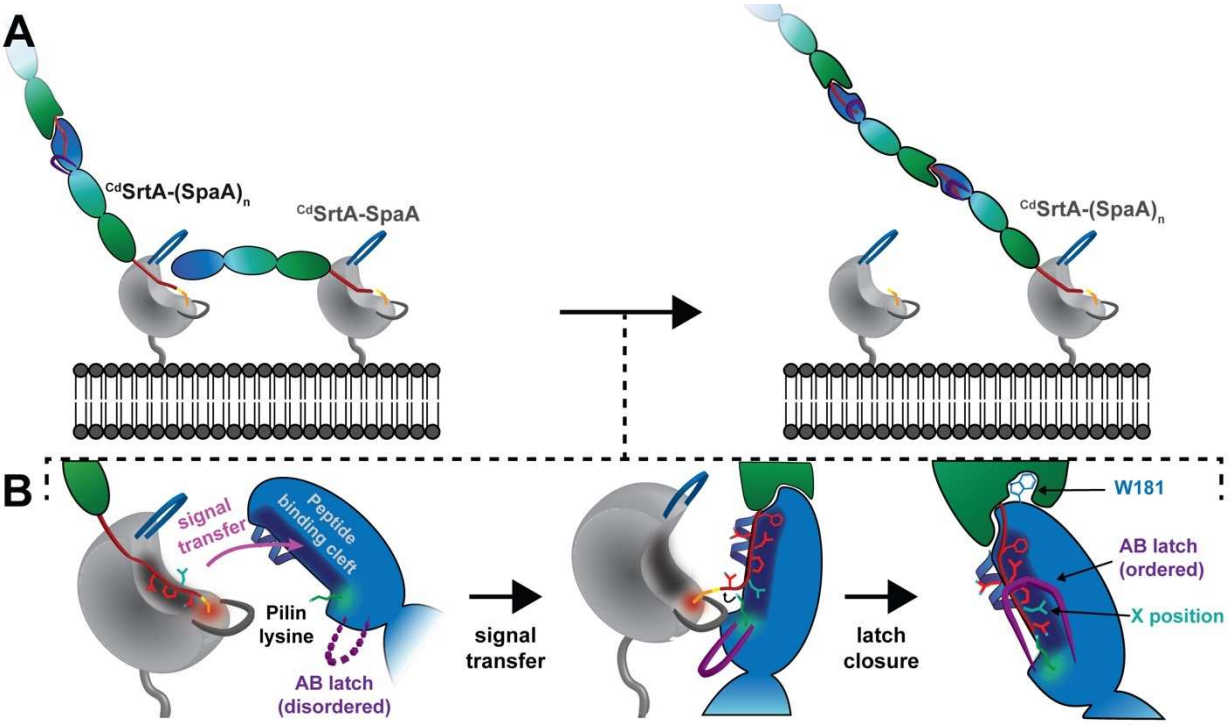

A
L
t=0
24 hours
Fluorescence
Coomassie
30
25
20
15
CdSrtA acyl int
NSpaA-LPLT
CdSrtA
NSpaA
WT
L62D
W105A
A108S
I111E
L148D
L168D
A170S
W181A
V187D
Y188G
K190A
H60A
ABΔ15
B
kcat/KM (1/M*s)
6
4
2
0
WT
H60A
L62A
ABΔ15
A170S
W181A
V187D
Y188G
C
ΔspaA
SpaA
H60A
L168D
W181A
V187D
Y188G
S
W
P
250
150
100
75
50
37
25
α-SpaA
D
ΔspaA/pSpaA
ΔspaA/pSpaAH60A
ΔspaA/pSpaAL168D
ΔspaA/pSpaAW181A
ΔspaA/pSpaAV187D
ΔspaA/pSpaAY188G

Figure 4.5 – The revised mechanism of pilus biogenesis. A) The general mechanism of sortase-mediated pilin ligation is illustrated. ^{Cd}SrtA has characteristic N-terminal "lid" appendage (blue) which occludes the catalytic cysteine residue, which must be opened to allow substrate entry to the active site, which allows for the formation of an enzyme-SpaA acyl intermediate. To build the SpaA pilus, two sortase-pilin acyl intermediates interact. The ^{Cd}SrtA-$(SpaA)_n$ species is tethered to the elongating pilus fiber and forms a transient ternary complex with ^{Cd}SrtA-SpaA (an acyl intermediate tethered to a single SpaA protomer). The reaction results in the transfer of the elongating pilus from the first to the second species via a new isopeptide linkage to the newly incorporated SpaA protomer. B) Molecular details of pilus biogenesis that occur during the above generalized reaction are depicted. 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. The sorting signal peptide (red) is bound in the ^{Cd}SrtA binding pocket (dark grey). 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 β7/β8 loop of ^{Cd}SrtA (grey). Following transpeptidation, the previously disordered AB loop collapses into a rigid latch-like conformation over the isopeptide linkage which provides additional stability to the linkage unit.


A
CdSrtA-(SpaA)n
CdSrtA-SpaA
CdSrtA-(SpaA)n
B
signal transfer
Peptide binding cleft
Pilin lysine
AB latch (disordered)
signal transfer
latch closure
W181
AB latch (ordered)
X position

4.7 References

1. Danne, C. & Dramsi, S. Pili of Gram-positive bacteria: roles in host colonization. *Res. Microbiol.* **163**, 645–658 (2012).

2. Ramirez, N. A., Das, A. & Ton-That, H. New Paradigms of Pilus Assembly Mechanisms in Gram-Positive Actinobacteria. *Trends Microbiol.* **0**, (2020).

3. 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.* **18**, 224–232 (2010).

4. Proft, T. & Baker, E. N. Pili in Gram-negative and Gram-positive bacteria - Structure, assembly and their role in disease. *Cell. Mol. Life Sci.* **66**, 613–635 (2009).

5. Fronzes, R., Remaut, H. & Waksman, G. Architectures and biogenesis of non-flagellar protein appendages in Gram-negative bacteria. *EMBO J.* **27**, 2271–2280 (2008).

6. 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.* **105**, 14147–52 (2008).

7. 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.* **79**, 1236–1247 (2011).

8. Mandlik, A., Swierczynski, A., Das, A. & Ton-That, H. Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends Microbiol.* **16**, 33–40 (2008).

9. Hospenthal, M. K., Costa, T. R. D. & Waksman, G. A comprehensive guide to pilus biogenesis in Gram-negative bacteria. *Nature Reviews Microbiology* **15**, 365–379 (2017).

10. Ton-That, H. & Schneewind, O. Assembly of pili in Gram-positive bacteria. *Trends*

Microbiol. **12**, 228–234 (2004).

11. Telford, J. L., Barocchi, M. A., Margarit, I., Rappuoli, R. & Grandi, G. Pili in Gram-positive pathogens. *Nat. Rev. Microbiol.* **4**, 509–519 (2006).

12. Scott, J. R. & Zähner, D. Pili with strong attachments: Gram-positive bacteria do it differently. *Mol. Microbiol.* **62**, 320–330 (2006).

13. Kang, H. J. & Baker, E. N. Structure and assembly of Gram-positive bacterial pili: Unique covalent polymers. *Current Opinion in Structural Biology* **22**, 200–207 (2012).

14. Abee, T., Kovács, Á. T., Kuipers, O. P. & van der Veen, S. Biofilm formation and dispersal in Gram-positive bacteria. *Current Opinion in Biotechnology* **22**, 172–179 (2011).

15. Van Gerven, N., Waksman, G. & Remaut, H. Pili and flagella: Biology, structure, and biotechnological applications. in *Progress in Molecular Biology and Translational Science* **103**, 21–72 (Elsevier B.V., 2011).

16. Ton-That, H. & Schneewind, O. Assembly of pili on the surface of Corynebacterium diphtheriae. *Mol. Microbiol.* **50**, 1429–1438 (2003).

17. Ton-That, H., Marraffini, L. A. & Schneewind, O. Sortases and pilin elements involved in pilus assembly of Corynebacterium diphtheriae. *Mol. Microbiol.* **53**, 251–261 (2004).

18. Guttilla, I. K. *et al.* Acyl Enzyme Intermediates in Sortase-Catalyzed Pilus Morphogenesis in Gram-Positive Bacteria. *J. Bacteriol.* **191**, 5603–5612 (2009).

19. Khare, B. & V. L. Narayana, S. Pilus biogenesis of Gram-positive bacteria: Roles of sortases and implications for assembly. *Protein Science* **26**, 1458–1473 (2017).

20. Kang, H. J., Coulibaly, F., Clow, F., Proft, T. & Baker, E. N. Stabilizing Isopeptide Bonds

Revealed in Gram-Positive Bacterial Pilus Structure. *Science (80-.).* **318**, 1625–1628 (2007).

21. Kang, H. J. & Baker, E. N. Intramolecular isopeptide bonds: Protein crosslinks built for stress? *Trends Biochem. Sci.* **36**, 229–237 (2011).

22. Hilleringmann, M. *et al.* Molecular architecture of Streptococcus pneumoniae TIGR4 pili. *EMBO J.* **28**, 3921–3930 (2009).

23. Kang, H. J., 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. *Proc. Natl. Acad. Sci.* **106**, 16967–16971 (2009).

24. McConnell, S. A. S. A. *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. 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**, E5477–E5486 (2018).

26. 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).

27. Chan, A. H. *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. Suree, N. *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. 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).

30. 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).

31. 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).

32. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).

33. Spraggon, G. *et al.* Supramolecular Organization of the Repetitive Backbone Unit of the Streptococcus pneumoniae Pilus. *PLoS One* **5**, e10919 (2010).

34. Budzik, J. M. *et al.* Intramolecular amide bonds stabilize pili on the surface of bacilli. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19992–19997 (2009).

35. Cozzi, R. *et al.* Structure and assembly of group B streptococcus pilus 2b backbone protein. *PLoS One* **10**, (2015).

36. Kang, H. J. *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. Sect. D Biol. Crystallogr.* **70**, 1190–1201 (2014).

37. 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**, 1205–1220 (2011).

38. Vengadesan, K., Ma, X., Dwivedi, P., Ton-That, H. & Narayana, S. V. L. A model for group B Streptococcus pilus type 1: The structure of a 35-kDa C-terminal fragment of the major pilin GBS80. *J. Mol. Biol.* **407**, 731–743 (2011).

39. Nuccitelli, A. *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**, (2011).

40. Chaurasia, P., Pratap, S., Palva, A., von Ossowski, I. & Krishnan, V. Bent conformation of a backbone pilin N-terminal domain supports a three-stage pilus assembly mechanism. *Commun. Biol.* **1**, (2018).

41. Durand, D. *et al.* NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **169**, 45–53 (2010).

42. Receveur-Brechot, V. & Durand, D. How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective. *Curr. Protein Pept. Sci.* **13**, 55–75 (2012).

43. Rogers, E. A., Das, A. & Ton-That, H. Adhesion by pathogenic corynebacteria. *Adv. Exp. Med. Biol.* **715**, 91–103 (2011).

44. Mandlik, A., Swierczynski, A., Das, A. & Ton-That, H. Corynebacterium diphtheriae employs specific minor pilins to target human pharyngeal epithelial cells. *Mol. Microbiol.* **64**, 111–124 (2007).

45. Echelman, D. J. *et al.* CnaA domains in bacterial pili are efficient dissipaters of large mechanical shocks. *Proc. Natl. Acad. Sci.* **113**, 2490–2495 (2016).

46. Kilambi, K. P. & Gray, J. J. Rapid calculation of protein pKa values using rosetta.

Biophys. J. **103**, 587–595 (2012).

47. Lyskov, S. *et al.* Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE). *PLoS One* **8**, e63906 (2013).

48. Nagarajan, R. *et al.* PDBparam: Online resource for computing structural parameters of proteins. *Bioinform. Biol. Insights* **10**, 73–80 (2016).

49. Persson, K., Esberg, A., Claesson, R. & Strömberg, N. The Pilin Protein FimP from Actinomyces oris: Crystal Structure and Sequence Analyses. *PLoS One* **7**, (2012).

50. Young, P. G. *et al.* Structural conservation, variability, and immunogenicity of the T6 backbone pilin of serotype M6 Streptococcus pyogenes. *Infect. Immun.* **82**, 2949–2957 (2014).

51. El Mortaji, L. *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.* **841**, 833–841 (2011).

52. Chaurasia, P., Pratap, S., Von Ossowski, I., Palva, A. & Krishnan, V. New insights about pilus formation in gut-adapted Lactobacillus rhamnosus GG from the crystal structure of the SpaA backbone-pilin subunit. *Sci. Rep.* **6**, (2016).

53. Shaik, M. M. *et al.* A structural snapshot of type II pilus formation in Streptococcus pneumoniae. *J. Biol. Chem.* **290**, 22581–22592 (2015).

54. Chang, C., Huang, I. H., Hendrickx, A. P. A. & Ton-That, H. Visualization of gram-positive bacterial pili. *Methods Mol. Biol.* **966**, 77–95 (2013).

55. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

56. Keller, R. L. J. *The Computer Aided Resonance Assignment Tutorial*. (2004).

57. Garrett, D. S., Cai, M. & Clore, G. M. XIPP: multi-dimensional NMR analysis software. *J. Biomol. NMR* **74**, 9–25 (2020).

58. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).

59. Clore, G. M., Bax, A. & Gronenborn, A. M. Stereospecific assignment of β-methylene protons in larger proteins using 3D15N-separated Hartmann-Hahn and13C-separated rotating frame Overhauser spectroscopy. *J. Biomol. NMR* **1**, 13–22 (1991).

60. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013).

61. Schwieters, C. D., Kuszewski, J. J. & Marius Clore, G. Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **48**, 47–62 (2006).

62. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73 (2003).

63. Herrmann, T., Güntert, P. & Wüthrich, K. 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).

64. Herrmann, T., Güntert, P. & Wüthrich, K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *Journal of Biomolecular NMR* **24**, 171–189 (2002).

65. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 (1996).

66. DeLano, W. L. The PyMOL Molecular Graphics System, Version 2.3. *Schrödinger LLC* http://www.pymol.org (2020). doi:10.1038/hr.2014.17

67. Cavanagh, J. *Protein NMR Spectroscopy : Principles and Practice*. (Academic Press, 2007).

68. Kay, L. E., Torchia, D. A. & Bax, A. Backbone Dynamics of Proteins As Studied by 15N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry* **28**, 8972–8979 (1989).

69. Classen, S. *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. Crystallogr.* **46**, 1–13 (2013).

70. Hopkins, J. B., Gillilan, R. E. & Skou, S. BioXTAS RAW: Improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J. Appl. Crystallogr.* **50**, 1545–1553 (2017).

71. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. Determination of domain structure of proteins from x-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001).

72. Hacker, S. M. *et al.* Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **9**, 1181–1190 (2017).

73. Russel, D. *et al.* Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* **10**, (2012).

Chapter 5

Designed Protein Cages as Scaffolds for Building Multienzyme Materials

5.1 Overview

Spatial arrangements of proteins can influence their functions and potentiate synergistic interactions. Cellulolytic degradation of biomass is an excellent example of an enzymatic pathway that is enhanced by colocalization of related enzymes through substrate channeling. In natural systems, these enzymes are often displayed together on a large protein array which is tethered to the bacterial surface to enable synergistic lignocellulose degradation. Recent advances in designed protein cages enabled us to reconstitute this higher order spatial arrangement in a controllable manner. Chapter 5 describes a collaborative effort with Dr. Todd Yeates' lab (UCLA) to design a platform for displaying multiple enzymes on protein cages which yield highly efficient cellulose degrading machines. My contributions to this study included sortase-mediated bioconjugation of cellulolytic enzymes to the engineered cage scaffolds, quantification of cellulase modification, and measurement of cellulolytic activity of the resultant cellulolytic nanoparticles.

5.2 Designed Protein Cages as Scaffolds for Building Multienzyme Materials

ACS SyntheticBiology

pubs.acs.org/synthbio Research Article

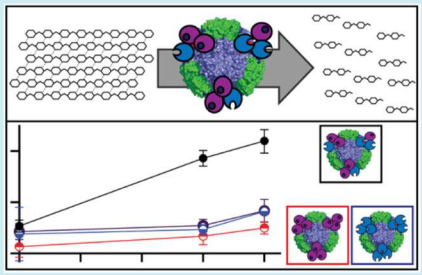
Designed Protein Cages as Scaffolds for Building Multienzyme Materials

Scott A. McConnell,[⊥] Kevin A. Cannon,[⊥] Christian Morgan, Rachel McAllister, Brendan R. Amer, Robert T. Clubb,* and Todd O. Yeates*

Cite This: *ACS Synth. Biol.* 2020, 9, 381−391 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The functions of enzymes can be strongly affected by their higher-order spatial arrangements. In this study we combine multiple new technologies—designer protein cages and sortase-based enzymatic attachments between proteins—as a novel platform for organizing multiple enzymes (of one or more types) in specified configurations. As a scaffold we employ a previously characterized 24-subunit designed protein cage whose termini are outwardly exposed for attachment. As a first-use case, we test the attachment of two cellulase enzymes known to act synergistically in cellulose degradation. We show that, after endowing the termini of the cage subunits with a short "sort-tag" sequence (LPXTG) and the opposing termini of the cellulase enzymes with a short polyglycine sequence tag, addition of sortase covalently attaches the enzymes to the cage with good reactivity and high copy number. The doubly modified cages show enhanced activity in a cellulose degradation assay compared to enzymes in solution, and compared to a combination of singly modified cages. These new engineering strategies could be broadly useful in the development of enzymatic material and synthetic biology applications.



KEYWORDS: *nanocages, sortase, cellulase, substrate channeling, bioconjugation*

Advances in protein engineering have created specifically designed nanocages that self-assemble into precise architectures.[1−9] The construction of symmetric nanocages from multiple copies of one (or a few) distinct protein subunits endows them with unique shapes, large sizes, a high degree of structural regularity, and polyvalency, among other features. Current efforts are focusing on developing these designer assemblies into useful materials.[10,11] Recently emerging applications include the creation of nanocages to encapsulate nucleic acids in their interior,[12,13] or to display specific protein components on their exterior. Developments of the latter type include the display of viral antigens to increase neutralizing antibody responses to infections,[14] antifreeze enzymes to increase ice-binding capacity,[15] and small proteins for structure determination by cryoelectron microscopy.[16,17]

In principle, designer protein nanocages could be ideal platforms for displaying enzymes with precisely defined spatial relationships. These structures could effectively mimic naturally occurring enzyme complexes that increase the efficiency of multistep reactions by channeling intermediates between sequentially acting enzymes. The utility of channeling substrates between sequentially acting enzyme sites has been exploited in diverse natural systems, from single molecule enzymes such as tryptophan synthase,[18] to giant compartmentalized metabolic complexes.[19−21] Such natural systems have motivated numerous synthetic biology efforts to improve enzyme performance and pathway flux by colocalizing enzymes in specific arrangements. Theoretical and computational studies have examined the structural and geometric regimes where channeling advantages might be significant,[22−24] and diverse combinations of enzymes and scaffolding frameworks have been explored for increasing pathway flux[25−37] (reviewed by Whitaker and Dueber,[38] as well as Quin *et al.*[39]). Notwithstanding recent studies employing very large natural protein cages (*i.e.*, bacterial microcompartments and encapsulins) to enclose multiple[29] or single enzymes or other proteins,[30,40−43] to the best of our knowledge, the exceptional spatial control afforded by designed (non-natural) nanocages has yet to be harnessed for such applications. We begin to explore those possibilities here.

Lignocellulosic plant biomass is a highly abundant and attractive renewable feedstock for producing biofuels and chemicals. However, its recalcitrance to hydrolysis limits its cost-effective usage on an industrial scale. A common approach

Received: October 5, 2019
Published: January 10, 2020

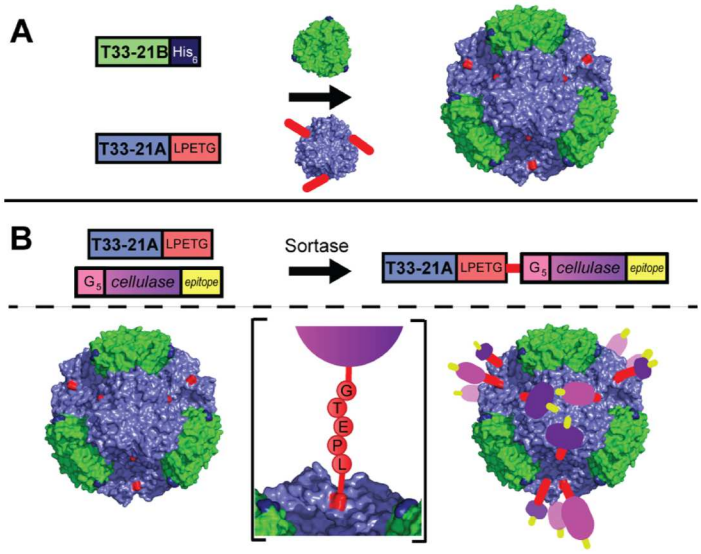


Figure 1. Assembly architecture of the T33-21-sort-tag cage design. (A) The self-assembling T33-21-SR protein nanocage consists of 4 copies of each of two trimeric protein components, T33-21A (blue) and T33-21B (green). The A and B components are engineered to include a C-terminal sort-tag (red) or His_6 tag, respectively. (B) Schematic diagram for attaching cellulase enzymes to the nanocage. (Top) A linear sequence diagram of the reaction catalyzed by sortase to add a cellulase enzyme to the B subunit of the designed cage. (Bottom) A structural depiction of the assembly produced. The recombinant cellulase enzymes are in shades of purple. The nucleophilic pentaglycine is pink. The peptide epitope for immune-detection is yellow.

to degrade lignocellulose into its component sugars is to employ a consortium of synergistically functioning cellulase enzymes that have distinct substrate specificities. Lignocellulose is comprised of varying amounts of cellulose (25−55%), hemicellulose (8−30%), and lignin (18−35%).[44] Cellulose is a polymer of β-1,4-linked glucose molecules that can hydrogen-bond with other cellulose polymers to form both amorphous and crystalline regions. It is synergistically degraded by three types of cellulases: endoglucanases, exoglucanases, and β-glucosidases.[45] Endoglucanases attack within a cellulose strand to hydrolyze the β-1,4-glucosidic bonds, producing new reducing and nonreducing ends that can be further broken down by exoglucanases. The shorter cellodextrin chains that are produced by these enzymes, including the disaccharide cellobiose, are then degraded into glucose monomers by β-glucosidases.[45,46] Hemicellulose is a sugar polymer that is composed of a number of different types of pentose and hexose sugars.[47] As compared to cellulose, it is more readily degraded by a range of enzymes, including among others: xylanases, arabinases, and mannases.[48,49] Finally, lignin, which surrounds and blocks enzyme access to cellulose and hemicellulose, is a complex polymer containing a mixture of phenolic compounds linked through radical coupling reactions.[48] A large number of enzymes are needed to degrade it, including peroxidases and laccases.[44,48−53]

Clostridium thermocellum and other species of anaerobic bacteria efficiently degrade lignocellulose using cellulosomes, large surface displayed protein complexes that house cellulases with complementary activities.[54−57] Clustering different types of cellulases within these structures promotes enzyme synergy, where the cellulolytic activities of the complexed enzymes are greater than that of individual enzymes due to their complementary activities and optimal enzyme spacing.[58−60] The presence of both hemicellulases and cellulases within the cellulosome also enables hemicellulose and cellulose fibers to be removed simultaneously, thereby overcoming potential physical hindrances. The benefits of the *C. thermocellum* cellulosomal system have been quantified: its specific activity against crystalline cellulose is 15-fold higher than the secreted enzyme system from *T. reesei*.[61] Moreover, placement of the cellulosome on the microbial surface increases the rate of hydrolysis by promoting cellulose-enzyme-microbe synergy. In this process, sugar uptake by the microbe presumably becomes more efficient by promoting import of the products into cells and by removing potential enzyme feedback inhibitors such as glucose and cellobiose from the environment.[62,63] The complex interplay of natural cellulases (*e.g.*, in the cellulosome) makes this system an attractive target for enzymatic engineering.

In this study, we sought to develop a modular platform to produce designer nanocages that display multiple enzymes in high copy number on their exterior. Diverse methods have been employed for attaching enzymes or other types of proteins to scaffolds. These include noncovalent methods based on protein or nucleic acid recognition motifs;[27,28,34,35,37] covalent connections by direct genetic fusion (*i.e.*, linear extension of two sequences);[14−17] and covalent connections of other types and origins, including isopeptide bonds. In the latter category, the SpyTag/SpyCatcher system and its variants have been widely exploited.[36,64−67] For modifying designed protein cages, previous modification methods have attached proteins *via* noncovalent interactions or by expressing

https://dx.doi.org/10.1021/acssynbio.9b00407
ACS Synth. Biol. 2020, 9, 381−391

externally displayed proteins as genetic fusions with a cage subunit. Here we develop a unique approach that harnesses the sequence specificity and robust ligation activity of the *S. aureus* sortase A (SrtA) enzyme, a widely used cysteine transpeptidase. We show that the surface of a designer nanocage can be elaborated with multiple cellulase enzymes using a sortase enzyme as the linking catalyst. Nanocage labeling is efficient, yielding cellulolytic protein nanoparticles whose component enzymes function synergistically. This system enables practically any protein or peptide-containing molecule with an exposed terminus to be grafted onto the exterior of a protein cage (or other oligomeric assembly). It also represents a step toward creating more elaborate cellulase-coated materials that could be useful in producing renewable biofuels and chemicals.

■ RESULTS AND DISCUSSION

We developed a general method to create enzyme-coated designer nanocages toward the goal of emulating naturally occurring multienzyme complexes that catalyze reactions with improved activity and fidelity. As a first-use case, we sought to coat a designed nanocage known as T33-21[4] with cellulase enzymes, which are known to exhibit synergistic activity against crystalline cellulose when clustered together within protein complexes such as cellulosomes. T33-21 nanocages are constructed from two types of protein subunits (T33-21A and -21B) that spontaneously assemble into a "dual tetrahedral architecture" that contains 12 copies of each subunit (Figure 1A).

In our procedure, enzymes are covalently attached to preassembled nanocages using the *S. aureus* sortase A transpeptidase, a powerful bioconjugation tool that joins biomolecules together *via* a backbone−backbone peptide bond.[74−76] Sortase links two proteins together when they contain LPXTG and oligoglycine sequences at their C- and N-termini, respectively. It catalyzes a transpeptidation in which the Thr-Gly bond within the LPXTG sequence is broken, and then replaced with a new Thr-Gly peptide linkage to the N-terminal amine group within the oligoglycine sequence (Figure 1B); in the present scenario this gives rise to a native polypeptide backbone connection. Among the suite of designed protein nanocages of known structure as candidates,[11] we reasoned that T33-21 nanocages would be good substrates for modification by sortase, as the C-termini of its subunits are outwardly disposed and thus potential sites for sortase modification, once modified to contain a LPXTG sequence. We therefore produced T33-21 nanocages in which the sequence of the T33-21A subunit is recombinantly extended to contain a C-terminal LPXTG sequence ("sort-tag").

After purification, protein cages bearing sort-tags at their termini were covalently modified with purified cellulases whose sequences were recombinantly extended to contain an N-terminal penta-glycine (Gly_5) sequence. To maximize their reactivity, the enzyme fusions were designed to unmask the terminal glycine during the purification procedure (see Methods). Short peptide epitope tags (myc and FLAG) were also appended to the C-termini of each cellulase for immunological probing. As a test of utility, cages were modified with the Cel48S exoglucanase and Cel8A endoglucanase from *C. thermocellum*, two enzymes that are well characterized and known to synergistically degrade cellulose.[56]

T33-21 nanocages harboring "sort-tags" (on 12 exposed termini) were efficiently modified with cellulases upon reaction with sortase. Initially we developed conditions to modify cages with a single type of enzyme, either Cel48S or Cel8A. In these reactions, 1 μM of assembled T33-21 nanocage containing a "sort-tag" was incubated with 1 μM of sortase and 150 μM cellulase possessing a N-terminal G5 sequence. In order to increase the rate of cage modification, we utilized a sortase variant that has 120-fold greater activity than the wild-type enzyme.[71] Sortase bioconjugation reactions are reversible, as the products contain N-terminal glycine and LPXTG sequences and can therefore react with sortase to regenerate the initial reactants. We mitigated this problem by conducting the reaction in dialysis conditions to remove the short peptide cleavage product and by restricting reaction times to 60 min.[77−80] SDS-PAGE analysis of the modification reactions confirmed that the cages were modified, as the T33-21A (21A-LPETG) protein component is progressively converted into higher molecular weight ligated species as the reaction proceeds (Figure 2A). In particular, 21A-LPETG was converted to 21A-Cel8A and 21A-Cel48S fusion polypeptides when the reactions were performed with appropriate cellulases harboring an N-terminal G_5 sequence (left and middle panels of Figure 2A, respectively). On the basis of the quantification of the band intensities, we estimate that ∼70−90% T33-21A is modified after 1 h.

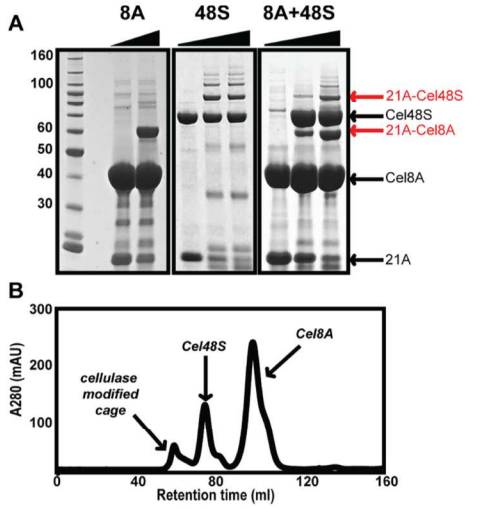


Figure 2. Sortase-catalyzed modification of cages. (A) The ligation reaction is tracked by SDS-PAGE separation of reaction components at different time points. Expected molecular weights for the products and substrates are indicated. Left panel: single-enzyme modification with Cel8A, middle panel: single-enzyme modification with Cel48S, right panel: double modification with both cellulases. For the double modification, a ratio Cel8A:Cel48S ratio of 2:1 was desired, so the cages were first reacted with Cel8A for 30 min, then Cel48S was added for an additional 30 min incubation. (B) Size exclusion was used to purify cellulolytic cages from monomeric cellulase components. Representative chromatogram from the two-enzyme modification reaction from gel filtration purifications.

We then generated cages modified with both types of cellulases by repeating the modification reaction using mixtures

https://dx.doi.org/10.1021/acssynbio.9b00407
ACS Synth. Biol. 2020, 9, 381−391

of the purified Cel48S and Cel8A enzymes. As both enzymes contain the same sortase-reactive N-terminal G_5 sequence, they can be expected to be stochastically ligated to the T33-21A subunit of the cage. Prior studies have shown that maximal synergy is observed when the endoglucanase is present in higher abundance relative to the exoglucanase.[81] Cages were therefore modified using a 2:1 ratio of Cel8A:Cel48S, leading to the expected ratio of the 21A-Cel8A and 21A-Cel48S polypeptides in the modified cages (Figure 2A, right panel). To estimate the number of cellulases attached to each cage, the cage complexes were purified using size exclusion chromatography (Figure 2B and Figure S1) and then analyzed using quantitative Western blotting. Purified singly (S and Cage-8A) and doubly (Cage-8A/48S) modified nanocages were subjected to SDS-PAGE followed by immunoblot detection of the myc- or FLAG-epitopes present on Cel48S and Cel8A, respectively (Figure 3). This analysis confirmed the presence

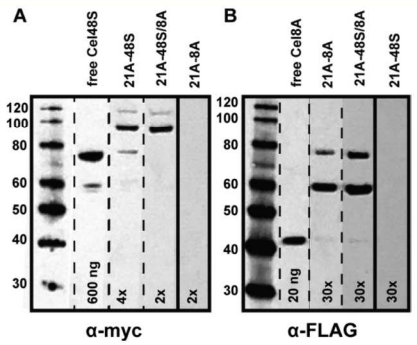


Figure 3. Detection of cellulase enzymes on purified cellulolytic cage by immunoblot analysis. Cellulase components are detected by α-FLAG antibody for the endoglucanase Cel8A or α-myc for the exoglucanase Cel48S, in combination with a secondary rabbit α-mouse IgG conjugated to horse radish peroxidase. For both immunoblots, a band corresponding to the ligated product T33-21A-cellulase is observed when incubated with the corresponding immunoblot. Negligible amounts of monomeric cellulase are visible in each purified stock. No cross reactivity between the myc and FLAG epitopes is observed in cages lacking the epitope of interest (*e.g.*, no signal is observed for Cage-8A in the α-myc immunoblot).

of the appropriate cellulase on each type of nanocage; probing gels with an antimyc antibody revealed the presence the 21A-Cel48S polypeptide in Cage-48S and Cage-8A/48S (Figure 3A), whereas a band with a molecular weight consistent with the 21A-Cel8A polypeptide is detected with an anti-FLAG antibody only in the Cage-Cel8A and Cage-Cel8A/Cel48S nanocages (Figure 3B). Absolute amounts of each type of enzyme attached to the cages were estimated by comparing band intensities of serial-diluted nanocages with corresponding standard curves obtained using known amounts of purified Cel8A or Cel48S (Figure S2). This analysis reveals that the single modified Cage-8A cages are coated with ∼5.8 Cel8A enzyme molecules and the Cage-48S cages are coated with ∼5.5 Cel48S enzyme molecules. On average, ∼48% and ∼46% of the available 21A subunits in the cages are modified with Cel8A and Cel48S, respectively, in the single modified cage experiments. As expected, a greater fraction of the 21A subunits are labeled with enzymes in the doubly modified Cage-8A/48S cages (∼71% of the 21A subunits) because nearly twice as much cellulase substrate was used in these labeling reactions. On the doubly modified Cage-8A/48S structure, ∼5.8 and ∼2.8 mol equiv of Cel8A and Cel48S, respectively, are present. Assuming that the attachment of enzymes is stochastic and the attachment sites are independent of each other (see Materials and Methods), about 96% of the cages are expected to carry at least one copy of the lower abundance Cel48S component, and 85% would carry at least two copies. More than 99% of the cages would carry at least 2 copies of the higher abundance Cel8A component.

The Cage-8A/48S complex was examined by negative stain electron microscopy (EM) to confirm that they retained their ordered structure after enzyme attachment (Figure 4). Before cellulase attachment with sortase, images of unmodified cages correspond closely to previously published EM images of T33-21 nanocages (Figure 4A).[4] This is expected and demonstrates that adding the "sort-tag" to the solvent exposed C-termini of the 21A subunits does not adversely affect cage assembly or stability. After fusing Cel8A and Cel48S to the cage *via* sortase, the symmetry of the cage assemblies is partially lost, but the scaffold core maintains its correct size and shape (Figure 4B). Notably, the presence of the cellulase enzymes is evident despite the flexible linkage to the T33-21A subunit, with the added components appearing as distinct puncta localized around the scaffold core. The EM results confirm that cellulases are attached to the exterior of the T33-21 nanocage.

Enzymes immobilized to the surface of the nanocage are active and function synergistically to degrade crystalline cellulose. The ability of the modified nanocages to degrade crystalline Avicel was determined using standard methods that monitor the production of reducing sugars (Figure 5). Singly modified cages containing Cel8A are more active than cages that only contain Cel48S, an observation that accords with the relative activities of the isolated enzymes (noncage attached). This behavior is consistent with the distinct substrate specificities of these enzymes, since crystalline cellulose is expected to contain significantly fewer nonreducing ends that are substrates for the Cel48S exoglucanase as compared to internal sites that can be cleaved by Cel8A.[50,51] Importantly, statistically significant gains in activity are achieved by colocalizing both types of enzymes together on Cage-Cel8A/Cel48S. The largest effects are observed within 8 h, with Cage-Cel8A/Cel48S activity 2.7-fold greater than the activity exhibited by a mixture of the singly modified cages with cellulase concentrations matched to the doubly modified experiment (Figure 5A). The difference in glucose production between divalent and monovalent cages mixtures with identical amounts of cellulases indicates that the origin of the rate enhancement is primarily due to synergistic effects afforded by this system, not intrinsic activity enhancement of individual enzymes derived from cage association. Colocalization of sequentially acting cellulases on a protein cage scaffold provides a mechanism for substrate channeling.

To quantify the effect of enzyme immobilization in the different types of cages, we calculated the stimulation factor (SF), as the fold increase of activity in the enzyme-coated cage compared to the corresponding free enzyme[82] (Table 1). Note that the free enzymes used for comparison here are the same recombinant constructs harboring N- and C-terminal peptide tags. Strikingly, after 8 h Cage-Cel8A/Cel48S exhibits a SF value of 3.9, whereas the corresponding mixture of singly modified cages has a SF value of 1.5. This indicates that

https://dx.doi.org/10.1021/acssynbio.9b00407
ACS Synth. Biol. 2020, 9, 381−391

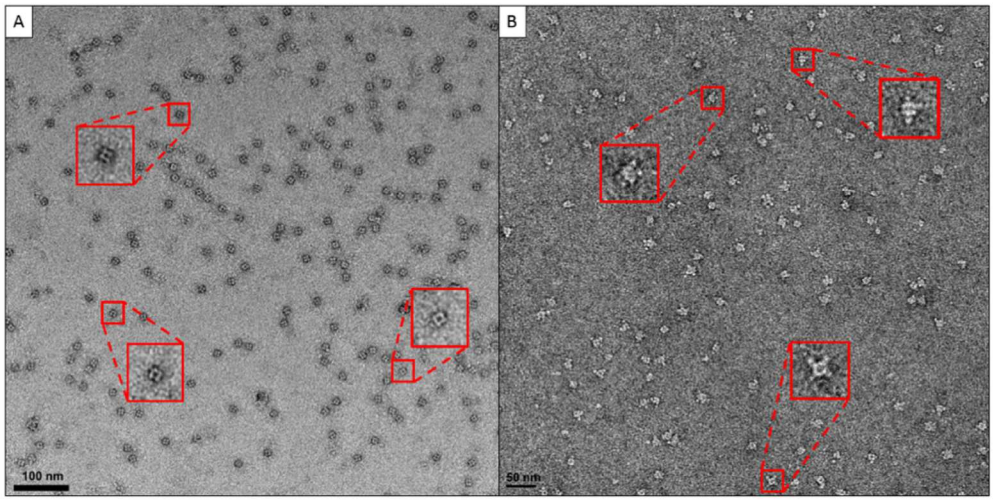


Figure 4. Representative negative stain EM images of the T33-21 cage before and after sortase ligation. (A) Individual T33-21 particles imaged before sortase ligation closely match previously published micrographs of the original cage construct. Highlighted particles show clear views along the 2-fold symmetry axis of the tetrahedron. (B) After fusing Cel8A and Cel48S to the cage *via* sortase ligation, distinct puncta localized around the cage particles, corresponding to the attached cellulase enzymes, are observed.

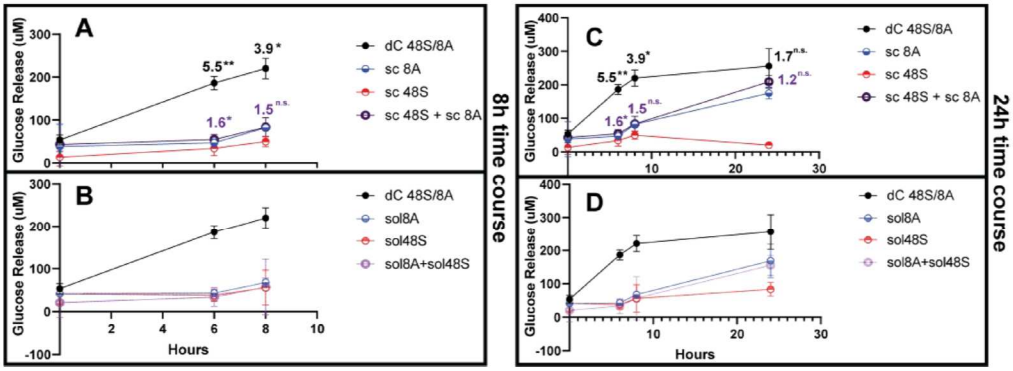


Figure 5. Degradation of cellulose substrate Avicel by cellulolytic cages. Enzyme-modified cages and corresponding free enzyme mixtures were assayed at several time points (6, 8, 24 h). Cellulose degradation was quantified by the release of reducing sugars, which were analyzed by incubation of reaction aliquots with dinitrosalicylic acid using established methods. (A,B) Initial reaction time course. (A) Cellulose degradation by doubly modified cages (*i.e.*, bearing both Cel8A and Cel48S enzymes) is compared at different time points to the activity of corresponding concentrations of the same enzymes on singly modified cages. (B) Cellulose degradation by double modified cages is compared at different time points to the activity of corresponding concentrations of the free enzymes in solution. (C,D) Extended time course of the same mixtures as shown in panels A and B, respectively, showing a leveling off of activity over time. Statistical significance of divalent cages (Cage-48S/8A) and the mixture of monovalent cages (sC-8A + sC-48S) was tested in Prism 8.2.1 using the Dunnett method of multiple comparisons. * represents $p < 0.05$, ** represents $p < 0.01$, n.s. represents not significant.

colocalization of Cel8A and Cel48S on the same nanocage scaffold provides synergistic enzyme activity, consistent with the complementary endo- and exoglucanase activities of the displayed enzymes. Further, synergy between the two enzymes occurs only when held in proximity on the same cage particle, and not when the enzymes are both dissolved freely in solution or present on separate cage scaffolds. Interestingly, the overall activity and degree of synergy declines at longer incubation times (24 h) (Figure 5B). The origin of this effect is presently unclear. Dynamic light scattering (DLS) and negative stain EM control experiments suggest that the cages remain stable after incubation at 37 °C for 24 h, with little to no disassembly of cage particles evident by either method (Figure S3). The leveling-off of activity over time likely reflects phenomena related to the complex cellulosic substrate, product inhibition,

https://dx.doi.org/10.1021/acssynbio.9b00407
ACS Synth. Biol. 2020, 9, 381−391

Table 1. Avicel Degradation by Cellulolytic Cages[a]

cellulase	released glucose (μM)	activity[b] (mU)	sp. activity (mU/mg)	SF[c]
Cage:8A/48S	220.3 ± 24	0.46	51.0	3.8
Cage:8A	82.2 ± 0.4	0.17	34.2	
Cage:48S	49.9 ± 12	0.10	26.0	
Cage:8A + Cage:48S	83.6 ± 23	0.17	19.4	1.5
Cel8A	67.9 ± 10	0.14	28.3	
Cel48S	55.9 ± 41	0.12	29.1	
Cel8A + Cel48S	57.1 ± 65	0.10	13.2	

[a]Released sugars from Avicel after 8h at 37C by mono singly modified- or (two-enzyme) divalent doubly modified cages and corresponding free enzymes. Cage-48S/8A = divalent two-enzyme, doubly modified cages (T33-21:Cel8A:Cel48S), mC Cage-8A or Cage-48S = monovalent singly modified cages (T33-21:Cel8A or T33-21:Cel48S); Cel8A or Cel48S= free cellulase enzymes. [b]1 U = amount of enzyme producing 1umol of reducing sugar per minute. [c]SF = (released soluble sugars by cellulolytic cages)/(released soluble sugars by the corresponding free enzyme pairs).

or perhaps other unanticipated molecular behavior that might be of interest for future analysis.

Our system is distinct from previously reported protein complexes that have sought to harness cellulase synergies by colocalization. The most popular approach has been to construct miniaturized cellulosomes ("mini"-cellulosomes) in which cellulases are bound to a central scaffoldin polypeptide *via* complementary dockerin-cohesin binding modules. While significant increases in activity have been achieved, these structures are generally not as cellulolytic as other complex native cellulosomes.[82−86] Their production can also be laborious, as the large scaffoldin protein is notoriously difficult to express and purify. Noncellulosome based strategies have also been pursued, including functionalizing inorganic ferromagnetic nanoparticles with cellulases,[82] and more recently using naturally self-assembling protein scaffolds to display cellulases *via* dockerin-cohesin interactions.[87] In one such study, four different cellulases were engineered with corresponding dockerin modules and mixed in different ratios to create cellulolytic "Rosettazymes".[88] Direct activity comparisons with these systems are not possible at this time because distinct assay conditions and cellulose substrates have been employed. In addition, cellulose degradation is known to exhibit nonlinear kinetics, causing measured values to be heavily influenced by enzyme and substrate concentrations. Bearing these limitations in mind, our cellulolytic cages display a maximal specific activity of 0.05 U/mg and stimulation factor of 3.9, which are comparable to related cellulolytic assemblies that contained two or three cellulases.[82,83,86] Moreover, literature precedence suggests that straightforward additions to our platform could yield further gains in nanocage cellulolytic activity. These include adding a tethered β-glucosidase to eliminate the buildup of cellobiose that can cause product inhibition;[82,89,90] adding cellulose binding modules to promote targeting to cellulose;[82,84,85,91] and increasing cellulase diversity to expand the range of substrates that can be degraded.[82,83,91,92]

CONCLUSION

This study introduces a new sortase-based approach for attaching multiple enzyme molecules to an ordered protein scaffold. Here we chose a designed protein nanocage for the enzyme scaffold array, though the sortase-based strategy can be used as easily to attach enzymes to other types of protein arrays, including filaments and two or three-dimensional extended materials. Essentially the sole requirement for the substrate scaffold array is either exposed N- or C-termini that can be engineered to contain sortase-reactive polyglycine or LPXTG sequences for modification. The approach could also be generalized for attaching more diverse molecules to scaffolds. Indeed, prior to modifying our designed cage with cellulase enzymes, as a first test of reactivity we demonstrated the ability to attach a small SUMO protein domain (Figure S4). In principle any type of biomolecule can be attached if it contains a terminal glycine residue, which can be incorporated easily into synthetic peptide substrates as well as lipids, fluorophores, and other molecules using appropriate chemistries.

Wide applications are possible with the facile attachment of proteins (or other molecules) to protein scaffolds. In this study we demonstrated a potential use for multienzyme colocalization, with notable synergistic effects, but this represents one of many applications. Protein scaffolds have previously been modified with functional groups in various ways including by chemical reactions[9,93,94] to enable targeting to specific cells or tracking. Methods to encapsulate enzymes inside protein cages to promote specialized reactions, protect sensitive enzymes, or to provide controllable substrate release have also been reported.[29,30,40−42,95−99] The present study opens yet another new range of possibilities for protein and enzyme organization based on sortase and nanocage technologies.

MATERIALS AND METHODS

Recombinant Protein Expression and Purification. T33-21 protein cages were prepared by coexpressing genes encoding its T33-21A and -21B subunits from a pET-22b plasmid (Novagen). Genes were purchased from IDT and inserted *via* Gibson Assembly into the vector. The amino acid sequences were the same as previously reported for the T33-21 cage,[4] except that a 16-residue sortase recognition tag (QSKKSELPETGGEEST) was appended to the C-terminus of the A component. All cellulase proteins used in this study were expressed with pE-SUMO (LifeSensors) expression vector. The expression plasmids containing the individual cellulase enzymes were ligated into a pSUMO expression plasmid using Gibson Assembly. The assembly reaction was engineered to include a penta-glycine repeat at the junction between the SUMO and cellulase gene, such that cleavage by the Ulp1 protease yields a cellulase with a functional pentaglycine (G_5) nucleophile at its amino-terminus. This approach avoids potential problems associated with the incomplete removal of the N-terminal methionine that can arise when sortase substrates are expressed in bacteria with the glycine nucleophile directly following the start codon.[68−70] Additionally, each cellulase was appended at its carboxy-terminus with a peptide epitope to enable immunological identification of each of the cellulase components. An improved pentamutant variant of the Sortase A protein from *Staphyloccocus aureus* (SrtA) was used in this study, bearing the following mutations: P94R/D160N/D165A/K190E/K196T.[71]

For all protein constructs, *E. coli* BL21 (DE3) cells harboring the target expression plasmid were grown in LB supplemented with either ampicillin at 100 μg/mL (T33-21-SR) or kanamycin at 500 μg/mL (cellulase constructs) at 37 °C until OD_{600} of ∼0.6. Cells were induced with 1 mM IPTG and

https://dx.doi.org/10.1021/acssynbio.9b00407
ACS Synth. Biol. 2020, 9, 381−391

protein expression was allowed to proceed overnight at 17 °C. Cells were then harvested by centrifugation (7000*g* for 10 min). Cells containing the T33-21-SR construct were lysed by sonication in a buffer of 50 mM Tris pH 8.0, 250 mM NaCl, and 20 mM imidazole. The protein cage was then subjected to Ni-affinity purification on a HisTrap HP column (GE Life Sciences) with gradient elution in the range of 20−500 mM imidazole. Elution fractions containing the protein cage components (confirmed by SDS-PAGE) were then dialyzed into a buffer containing 25 mM Tris pH 8.0, 150 mM NaCl, and 1 mM DTT to be used in subsequent experiments. All cellulase proteins were purified as a His6-SUMO- fusion using HisPure Co^{2+} IMAC resin (Thermo Scientific) per the manufacturer's instructions. Briefly, cell pellets were resuspended in 50 mM Tris-HCl pH 8.0, 300 mM NaCl (lysis buffer) and lysed by sonication. The cell lysate was then fractionated by centrifugation (15 000*g* for 40 min) and the supernatant was loaded onto HisPure Co^{2+} IMAC resin. Proteins were then eluted from the resin using lysis buffer supplemented with 200 mM imidazole. The His6x-SUMO tag was removed by adding His6-Ulp1 protease, and subsequent HisPure Co^{2+} purification. All proteins were then loaded onto a SuperDex75 size exclusion column as a final purification step. Protein purity was determined by SDS-PAGE analysis.

Cellulase Labeling of Nanocages Using Sortase. Sortase bioconjugation reactions to covalently ligate the engineered cellulase components to synthetic protein cages were performed as 1 mL reactions at room temperature. All proteins were diluted into SrtA modification buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM DTT, 5 mM CaCl2). The following protein concentrations were used for the attachment reaction: 1 μM SrtA enzyme, 1 μM of assembled T33-21-SR protein cages and 50−150 μM G5-Cel8A-FLAG and/or 200 μM G5-Cel48S-Myc. Reactions were allowed to proceed for 1 h while being dialyzed against sortase modification buffer to remove the hydrolysis product and prevent reverse reaction. The reaction components were then separated using a precast NuPAGE 4−12% gradient Bis-Tris protein gel (Thermo Scientific). Gels were stained with Coomassie Blue G-250 for analysis of reaction progress. Assembled cages were separated from other reaction components, including monomeric cellulases and sortase, using size-exclusion chromatography with a SuperDex75 column (GE).

Functionalized Nanocages Quantification. The number of enzymes that were attached to the purified cages were quantified by Western blot using *α-myc* and *α*-FLAG antibodies. Protein standards were created for each of the cellulase constructs by dilution to known concentrations. On the same SDS-PAGE gel, serial dilutions of each type of cellulase-modified cage were run alongside the standards. After the SDS-PAGE was run, the separated proteins were transferred to a PVDF membrane using the iBlot2 device (20W, 7 min, Invitrogen). The membranes were then sequentially blocked, incubated with primary antibody (mouse *α-myc* or *α*-FLAG), washed, incubated with secondary antibody (rabbit *α*-mouse IgG-HRP), and washed again by capillary action using the iBind system (Invitrogen). Immunological detection of proteins with appropriate peptide tags was facilitated by incubation of the membranes with luminol substrate. Luminescent signal is detected by autoradiography film (Genessee Scientific, 30 s exposure). The intensity of signal arising from lanes with protein standards (free cellulases of known concentration) was analyzed to create a standard curve. Intensity arising from bands with serial dilutions of the stocks of cellulolytic cages were fit to the standard curve to yield a precise measurement of the amount of cellulase present on each of the modified cages.

Estimates of the percentage of cages that would carry multiple copies of the components enzymes were obtained by evaluating binomial distributions, under the assumption that attachment of enzymes to potential sites on the cage were statistically independent. For the lower abundance component, Cel48S, whose fractional attachment was 2.8/12 = 0.2333, the fraction of cages expected to have 0 Cel48S attachments would be $0.233^{12} = 0.041$, and the fraction expected to carry only one copy would be $12 \times 0.233^{11} \times (1 - 0.233) = 0.151$. Similar calculations for the more abundant component (Cel8A) indicate that less than 1% of the cages would carry fewer than 2 copies of that enzyme.

A cellulose degradation assay was used to determine enzyme activity. Avicel PH101 (Sigma) was prepared as described previously,[72] then diluted to the desired cellulose concentration in cellulase assay buffer (50 mM sodium acetate, pH 5.5). For the cellulose degradation assays, Cel8A (an endoglucanase) and Cel48S (an exoglucanase) were selected on the basis of their previously established synergistic interactions. All cellulases were produced heterologously in *E. coli*. As a preliminary confirmation of cellulolytic degradative activity of the engineered cellulase constructs, reducing sugars release was measured after incubation with a soluble amorphous cellulose substrate (CMC). In order to measure enzyme activity against more recalcitrant crystalline cellulose substrates, aliquots of 800 μL enzyme mixtures were added to 200 μL of 10% Avicel substrate to create 1 mL reactions at a final Avicel concentration of 2%. Total enzyme amounts of Cel8A, Cel48S and *β*-glucosidase were 5, 4, and 100 μg, respectively. The enzymatic activity assays were performed while incubating at 37 °C with shaking at 225 rpm for specified reaction periods. For each time point, 100 μL aliquots were centrifuged to pellet the insoluble Avicel and 50 μL of the supernatant was mixed with 75 μL 3,5-Dinitrosalicylic acid (DNS) reagent for quantification of reducing sugars released during the reaction, using glucose as a standard.[73]

Negative stain electron microscopy was used to characterize the structure and integrity of the modified nanocages. Cage constructs were imaged by electron microscopy to verify the correct assembly of cages before and after the sortase-mediated attachment of cellulase enzymes. Proteins were subjected to size exclusion chromatography using a Superose 6 Increase 10/300 (GE Life Sciences) column before application to EM grids. Five μL of purified cages in the concentration range of 0.005 mg/mL to 0.02 mg/mL were applied to glow discharged, Formvar/carbon-coated 300-mesh copper grids (Ted Pella, Inc.) and stained with 2% uranyl acetate. Cages were imaged on an FEI Tecnai T12 transmission electron microscope at 120 kV.

Dynamic light scattering of protein constructs was performed on a DynaPro Plate Reader II system (Wyatt Technology) to monitor the presence of cage assemblies of the expected size in solution (Figure S3). DLS measurements were made with protein concentrations of approximately 1 mg/mL. Preincubation samples were measured in triplicate, while postincubation samples were measured only in duplicate (due to loss of some sample during pipetting and incubation), with 5 acquisitions per replicate. Measurements that showed

https://dx.doi.org/10.1021/acssynbio.9b00407
ACS Synth. Biol. 2020, 9, 381−391

polydispersity values of 100% or more were considered unreliable and were not included in the data analysis (all other data points had a polydispersity of less than 34%). Additionally, a minor species of 30 nm or more in radius was detected in all sample readings. This also appeared in buffer-only control samples and was thus treated as artifactual and left out of the reported analysis.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.9b00407.

Detailed gel filtration chromatograms for each divalent and monovalent cage, quantitative immunoblot data, SDS-PAGE and EM data demonstrating versatility of the nanocage/sortase attachment system, dynamic light scattering and negative stain electron microscopy images illustrating thermal stability of cages (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Robert T. Clubb − *UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, California, University of California, Los Angeles, California;* orcid.org/0000-0001-5718-3985; Phone: (+1) 310 206 2334; Email: rclubb@mbi.ucla.edu; Fax: (+1) 310 206 4779

Todd O. Yeates − *UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, California, University of California, Los Angeles, California;* orcid.org/0000-0001-5709-9839; Phone: (+1) 310-206-4866; Email: yeates@mbi.ucla.edu; Fax: (+1) 310-206-3914

Other Authors

Scott A. McConnell − *UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, California, University of California, Los Angeles, California*

Kevin A. Cannon − *UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, California, and University of California, Los Angeles, California*

Christian Morgan − *University of California, Los Angeles, California*

Rachel McAllister − *University of California, Los Angeles, California*

Brendan R. Amer − *UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, California, University of California, Los Angeles, California*

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.9b00407

Author Contributions

⊥SAM and KAC contributed equally. SAM and RM performed the cage chemical ligation experiments and enzymatic assays and their analysis. KAC and CM designed and produced the protein cage, and KAC performed the electron microscopy. SAM and BRA performed initial experiments on ligation methods. The study was conceived by KAC, BRA, TOY, and RTC. The manuscript was written by SAM, KAC, TOY, and RTC, with input from the other authors.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Yuxi Liu for advice on electron microscopy and the staff at EICN in the UCLA California NanoSystems Institute for assistance with electron microscopy. This work was funded by the US Department of Energy Office of Science award DE-FC02-02ER63421.

■ ABBREVIATIONS

CCR2, CC chemokine receptor 2; CCL2, CC chemokine ligand 2; CCR5, CC chemokine receptor 5; TLC, thin layer chromatography.

■ REFERENCES

(1) Padilla, J. E., Colovos, C., and Yeates, T. O. (2001) Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci. U. S. A. 98*, 2217−2221.

(2) King, N. P., Sheffler, W., Sawaya, M. R., Vollmar, B. S., Sumida, J. P., Andre, I., Gonen, T., Yeates, T. O., and Baker, D. (2012) Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science 336*, 1171−1174.

(3) Lai, Y. T., Cascio, D., and Yeates, T. O. (2012) Structure of a 16-nm cage designed by using protein oligomers. *Science 336*, 1129−1129.

(4) King, N. P., Bale, J. B., Sheffler, W., McNamara, D. E., Gonen, S., Gonen, T., Yeates, T. O., and Baker, D. (2014) Accurate design of co-assembling multi-component protein nanomaterials. *Nature 510*, 103−8.

(5) Bale, J. B., Gonen, S., Liu, Y., Sheffler, W., Ellis, D., Thomas, C., Cascio, D., Yeates, T. O., Gonen, T., King, N. P., and Baker, D. (2016) Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science 353*, 389−395.

(6) Sciore, A., Su, M., Koldewey, P., Eschweiler, J. D., Diffley, K. A., Linhares, B. M., Ruotolo, B. T., Bardwell, J. C. A., Skiniotis, G., and Marsh, E. N. G. (2016) Flexible, symmetry-directed approach to assembling protein cages. *Proc. Natl. Acad. Sci. U. S. A. 113*, 8681−8686.

(7) Badieyan, S., Sciore, A., Eschweiler, J. D., Koldewey, P., Cristie-David, A. S., Ruotolo, B. T., Bardwell, J. C. A., Su, M., and Marsh, E. N. G. (2017) Symmetry-Directed Self-Assembly of a Tetrahedral Protein Cage Mediated by de Novo-Designed Coiled Coils. *ChemBioChem 18*, 1888−1892.

(8) Cristie-David, A. S., Chen, J., Nowak, D. B., Bondy, A. L., Sun, K., Park, S. I., Banaszak Holl, M. M., Su, M., and Marsh, E. N. G. (2019) Coiled-Coil-Mediated Assembly of an Icosahedral Protein Cage with Extremely High Thermal and Chemical Stability. *J. Am. Chem. Soc. 141*, 9207−9216.

(9) Heddle, J. G., Chakraborti, S., and Iwasaki, K. (2017) Natural and artificial protein cages: design, structure and therapeutic applications. *Curr. Opin. Struct. Biol. 43*, 148−155.

(10) Aumiller, W. M., Uchida, M., and Douglas, T. (2018) Protein cage assembly across multiple length scales. *Chem. Soc. Rev. 47*, 3433−3469.

(11) Cannon, K. A., Ochoa, J. M., and Yeates, T. O. (2019) High-symmetry protein assemblies: patterns and emerging applications. *Curr. Opin. Struct. Biol. 55*, 77−84.

(12) Butterfield, G. L., Lajoie, M. J., Gustafson, H. H., Sellers, D. L., Nattermann, U., Ellis, D., Bale, J. B., Ke, S., Lenz, G. H., Yehdego, A., Ravichandran, R., Pun, S. H., King, N. P., and Baker, D. (2017) Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature 552*, 415−420.

(13) Edwardson, T. G. W., Mori, T., and Hilvert, D. (2018) Rational Engineering of a Designed Protein Cage for siRNA Delivery. *J. Am. Chem. Soc. 140*, 10439−10442.

(14) Marcandalli, J., Fiala, B., Ols, S., Perotti, M., de van der Schueren, W., Snijder, J., Hodge, E., Benhaim, M., Ravichandran, R., Carter, L., Sheffler, W., Brunner, L., Lawrenz, M., Dubois, P., Lanzavecchia, A., Sallusto, F., Lee, K. K., Veesler, D., Correnti, C. E.,

Stewart, L. J., Baker, D., Loré, K., Perez, L., and King, N. P. (2019) Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus. *Cell 176*, 1420−1431.
(15) Phippen, S. W., Stevens, C. A., Vance, T. D. R., King, N. P., Baker, D., and Davies, P. L. (2016) Multivalent Display of Antifreeze Proteins by Fusion to Self-Assembling Protein Cages Enhances Ice-Binding Activities. *Biochemistry 55*, 6811−6820.
(16) Liu, Y., Gonen, S., Gonen, T., and Yeates, T. O. (2018) Near-atomic cryo-EM imaging of a small protein displayed on a designed scaffolding system. *Proc. Natl. Acad. Sci. U. S. A. 115*, 3362−3367.
(17) Liu, Y., Huynh, D. T., and Yeates, T. O. (2019) A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nat. Commun. 10*, 1864.
(18) Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W., and Davies, D. R. (1988) Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from Salmonella typhimurium. *J. Biol. Chem. 263*, 17857−17871.
(19) Bobik, T. A., Lehman, B. P., and Yeates, T. O. (2015) Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. *Mol. Microbiol. 98*, 193−207.
(20) Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F., and Sutter, M. (2018) Bacterial microcompartments. *Nat. Rev. Microbiol. 16*, 277−290.
(21) Jorda, J., Lopez, D., Wheatley, N. M., and Yeates, T. O. (2013) Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci. 22*, 179−195.
(22) Idan, O., and Hess, H. (2013) Origins of Activity Enhancement in Enzyme Cascades on Scaffolds. *ACS Nano 7*, 8658−8665.
(23) Castellana, M., Wilson, M. Z., Xu, Y., Joshi, P., Cristea, I. M., Rabinowitz, J. D., Gitai, Z., and Wingreen, N. S. (2014) Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat. Biotechnol. 32*, 1011−1018.
(24) Jakobson, C. M., Tullman-Ercek, D., and Mangan, N. M. (2018) Spatially organizing biochemistry: choosing a strategy to translate synthetic biology to the factory. *Sci. Rep. 8*, 1−13.
(25) Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V., Prather, K. L. J., and Keasling, J. D. (2009) Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol. 27*, 753−759.
(26) Moon, T. S., Dueber, J. E., Shiue, E., and Prather, K. L. J. (2010) Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli. *Metab. Eng. 12*, 298−305.
(27) Delebecque, C. J., Lindner, A. B., Silver, P. A., and Aldaye, F. A. (2011) Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science 333*, 470−474.
(28) Conrado, R. J., Wu, G. C., Boock, J. T., Xu, H., Chen, S. Y., Lebar, T., Turnšek, J., Tomšič, N., Avbelj, M., Gaber, R., Koprivnjak, T., Mori, J., Glavnik, V., Vovk, I., Benčina, M., Hodnik, V., Anderluh, G., Dueber, J. E., Jerala, R., and DeLisa, M. P. (2012) DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic Acids Res. 40*, 1879−1889.
(29) Lawrence, A. D., Frank, S., Newnham, S., Lee, M. J., Brown, I. R., Xue, W.-F., Rowe, M. L., Mulvihill, D. P., Prentice, M. B., Howard, M. J., and Warren, M. J. (2014) Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor. *ACS Synth. Biol. 3*, 454−465.
(30) Giessen, T. W., and Silver, P. A. (2016) A Catalytic Nanoreactor Based on in Vivo Encapsulation of Multiple Enzymes in an Engineered Protein Nanocompartment. *ChemBioChem 17*, 1931−1935.
(31) Price, J. V., Chen, L., Whitaker, W. B., Papoutsakis, E., and Chen, W. (2016) Scaffoldless engineered enzyme assembly for enhanced methanol utilization. *Proc. Natl. Acad. Sci. U. S. A. 113*, 12691−12696.
(32) Yang, Z., Gao, X., Xie, H., Wang, F., Ren, Y., and Wei, D. (2017) Enhanced itaconic acid production by self-assembly of two biosynthetic enzymes in Escherichia coli. *Biotechnol. Bioeng. 114*, 457−462.
(33) Wang, Y., Heermann, R., and Jung, K. (2017) CipA and CipB as Scaffolds To Organize Proteins into Crystalline Inclusions. *ACS Synth. Biol. 6*, 826−836.
(34) Li, T., Chen, X., Cai, Y., and Dai, J. (2018) Artificial Protein Scaffold System (AProSS): An efficient method to optimize exogenous metabolic pathways in Saccharomyces cerevisiae. *Metab. Eng. 49*, 13−20.
(35) Lee, M. J., Mantell, J., Hodgson, L., Alibhai, D., Fletcher, J. M., Brown, I. R., Frank, S., Xue, W.-F., Verkade, P., Woolfson, D. N., and Warren, M. J. (2018) Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm. *Nat. Chem. Biol. 14*, 142−147.
(36) Zhang, G., Quin, M. B., and Schmidt-Dannert, C. (2018) Self-Assembling Protein Scaffold System for Easy in Vitro Coimmobilization of Biocatalytic Cascade Enzymes. *ACS Catal. 8*, 5611−5620.
(37) Lim, S., Jung, G. A., Glover, D. J., and Clark, D. S. (2019) Enhanced Enzyme Activity through Scaffolding on Customizable Self-Assembling Protein Filaments. *Small 15*, 1805558.
(38) Whitaker, W. R., and Dueber, J. E. (2011) Metabolic Pathway Flux Enhancement by Synthetic Protein Scaffolding, in *Methods in Enzymology* (Voigt, C., Ed.), pp 447−468, Academic Press.
(39) Quin, M. B., Wallin, K. K., Zhang, G., and Schmidt-Dannert, C. (2017) Spatial organization of multi-enzyme biocatalytic cascades. *Org. Biomol. Chem. 15*, 4260−4271.
(40) Wagner, H. J., Capitain, C. C., Richter, K., Nessling, M., and Mampel, J. (2017) Engineering bacterial microcompartments with heterologous enzyme cargos. *Eng. Life Sci. 17*, 36−46.
(41) Liang, M., Frank, S., Lünsdorf, H., Warren, M. J., and Prentice, M. B. (2017) Bacterial microcompartment-directed polyphosphate kinase promotes stable polyphosphate accumulation in E. coli. *Biotechnol. J. 12*, 1600415.
(42) Yung, M. C., Bourguet, F. A., Carpenter, T. S., and Coleman, M. A. (2017) Re-directing bacterial microcompartment systems to enhance recombinant expression of lysis protein E from bacteriophage ϕX174 in Escherichia coli. *Microb. Cell Fact. 16*, 71.
(43) Hagen, A., Sutter, M., Sloan, N., and Kerfeld, C. A. (2018) Programmed loading and rapid purification of engineered bacterial microcompartment shells. *Nat. Commun. 9*, 1−10.
(44) Zhao, X., Zhang, L., and Liu, D. (2012) Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. *Biofuels, Bioprod. Biorefin. 6*, 465−482.
(45) Medie, F. M., Davies, G. J., Drancourt, M., and Henrissat, B. (2012) Genome analyses highlight the different biological roles of cellulases. *Nat. Rev. Microbiol. 10*, 227−234.
(46) Ghose, T. K. (1977) Cellulase biosynthesis and hydrolysis of cellulosic substances, in *Advances in Biochemical Engineering* (Ghose, T. K., Fiechter, A., and Blakebrough, N., Eds.), Vol. 6, pp 39−76, Springer, Berlin, Heidelberg.
(47) Scheller, H. V., and Ulvskov, P. (2010) Hemicelluloses. *Annu. Rev. Plant Biol. 61*, 263−289.
(48) Boerjan, W., Ralph, J., and Baucher, M. (2003) Lignin Biosynthesis. *Annu. Rev. Plant Biol. 54*, 519−546.
(49) Bugg, T. D. H., Ahmad, M., Hardiman, E. M., and Rahmanpour, R. (2011) Pathways for degradation of lignin in bacteria and fungi. *Nat. Prod. Rep. 28*, 1883−1896.
(50) Sethi, A., and Scharf, M. E. (2013) Biofuels: Fungal, Bacterial and Insect Degraders of Lignocellulose, in *eLS*, John Wiley & Sons, Ltd., Chichester, UK.
(51) Yang, B., Dai, Z., Ding, S. Y., and Wyman, C. E. (2011) Enzymatic hydrolysis of cellulosic biomass. *Biofuels 2*, 421−450.
(52) Zhao, X., Zhang, L., and Liu, D. (2012) Biomass recalcitrance. Part II: Fundamentals of different pre-treatments to increase the enzymatic digestibility of lignocellulose. *Biofuels, Bioprod. Biorefin. 6*, 561−579.
(53) Banerjee, G., Car, S., Scott-Craig, J. S., Borrusch, M. S., Aslam, N., and Walton, J. D. (2010) Synthetic enzyme mixtures for biomass deconstruction: production and optimization of a core set. *Biotechnol. Bioeng. 106*, 707−720.

(54) Bayer, E. A., Morag, E., and Lamed, R. (1994) The cellulosome — A treasure-trove for biotechnology. *Trends Biotechnol. 12*, 379–386.
(55) Bayer, E. A., Shimon, L. J., Shoham, Y., and Lamed, R. (1998) Cellulosomes-structure and ultrastructure. *J. Struct. Biol. 124*, 221–234.
(56) Fontes, C. M. G. A., and Gilbert, H. J. (2010) Cellulosomes: Highly Efficient Nanomachines Designed to Deconstruct Plant Cell Wall Complex Carbohydrates. *Annu. Rev. Biochem. 79*, 655–681.
(57) Wen, F., Sun, J., and Zhao, H. (2010) Yeast surface display of trifunctional minicellulosomes for simultaneous Saccharification and fermentation of cellulose to ethanol. *Appl. Environ. Microbiol. 76*, 1251–1260.
(58) Barth, A., Hendrix, J., Fried, D., Barak, Y., Bayer, E. A., and Lamb, D. C. (2018) Dynamic interactions of type I cohesin modules fine-tune the structure of the cellulosome of Clostridium thermocellum. *Proc. Natl. Acad. Sci. U. S. A. 115*, E11274–E11283.
(59) Cunha, E. S., Hatem, C. L., and Barrick, D. (2016) Synergistic enhancement of cellulase pairs linked by consensus ankyrin repeats: Determination of the roles of spacing, orientation, and enzyme identity. *Proteins: Struct., Funct., Genet. 84*, 1043–1054.
(60) Mayer, F., Coughlan, M. P., Mori, Y., and Ljungdahl, L. G. (1987) Macromolecular Organization of the Cellulolytic Enzyme Complex of Clostridium thermocellum as Revealed by Electron Microscopy. *Appl. Environ. Microbiol. 53*, 2785–2792.
(61) You, C., Zhang, X.-Z., Sathitsuksanoh, N., Lynd, L. R., and Zhang, Y.-H. P. (2012) Enhanced Microbial Utilization of Recalcitrant Cellulose by an Ex Vivo Cellulosome-Microbe Complex. *Appl. Environ. Microbiol. 78*, 1437–1444.
(62) Lu, Y., Zhang, Y.-H. P., and Lynd, L. R. (2006) Enzyme-microbe synergy during cellulose hydrolysis by Clostridium thermocellum. *Proc. Natl. Acad. Sci. U. S. A. 103*, 16165–16169.
(63) Demain, A. L., Newcomb, M., and Wu, J. H. D. (2005) Cellulase, Clostridia, and Ethanol. *Microbiol. Mol. Biol. Rev. 69*, 124–154.
(64) Zakeri, B., Fierer, J. O., Celik, E., Chittock, E. C., Schwarz-Linek, U., Moy, V. T., and Howarth, M. (2012) Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U. S. A. 109*, 4347–4348.
(65) Reddington, S. C., and Howarth, M. (2015) Secrets of a covalent interaction for biomaterials and biotechnology: SpyTag and SpyCatcher. *Curr. Opin. Chem. Biol. 29*, 94–99.
(66) Liu, Y., Liu, D., Yang, W., Wu, X.-L., Lai, L., and Zhang, W.-B. (2017) Tuning SpyTag–SpyCatcher mutant pairs toward orthogonal reactivity encryption. *Chem. Sci. 8*, 6577–6582.
(67) Da, X.-D., and Zhang, W.-B. (2019) Active Template Synthesis of Protein Heterocatenanes. *Angew. Chem., Int. Ed. 58*, 11097–11104.
(68) Liu, D., Xu, R., Dutta, K., and Cowburn, D. (2008) N-terminal cysteinyl proteins can be prepared using thrombin cleavage. *FEBS Lett. 582*, 1163–1167.
(69) Wingfield, P. (2017) N-Terminal Methionine Processing. *Curr. Protoc. Protein Sci. 88*, 6.14.1–6.14.3.
(70) 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*, 10800–10801.
(71) Chen, I., Dorr, B. M., and Liu, D. R. (2011) A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U. S. A. 108*, 11399–11404.
(72) Vazana, Y., Moraïs, S., Barak, Y., Lamed, R., and Bayer, E. A. (2012) Designer Cellulosomes for Enhanced Hydrolysis of Cellulosic Substrates, in *Methods in Enzymology* (Gilbert, H. J., Ed.), Chapter 23, pp 429–452, Academic Press.
(73) Miller, G. L. (1959) Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal. Chem. 31*, 426–428.
(74) Pishesha, N., Ingram, J. R., and Ploegh, H. L. (2018) Sortase A: A Model for Transpeptidation and Its Biological Applications. *Annu. Rev. Cell Dev. Biol. 34*, 163–188.
(75) Jacobitz, A. W., Kattke, M. D., Wereszczynski, J., and Clubb, R. T. (2017) Sortase Transpeptidases: Structural Biology and Catalytic Mechanism. *Adv. Protein Chem. Struct. Biol. 109*, 223–264.
(76) Marraffini, L. A., DeDent, A. C., and Schneewind, O. (2006) Sortases and the Art of Anchoring Proteins to the Envelopes of Gram-Positive Bacteria. *Microbiol. Mol. Biol. Rev. 70*, 192–221.
(77) Proft, T. (2010) Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilisation. *Biotechnol. Lett. 32* (1), 1.
(78) Kobashigawa, Y., Kumeta, H., Ogura, K., and Inagaki, F. (2009) Attachment of an NMR-invisible solubility enhancement tag using a sortase-mediated protein ligation method. *J. Biomol. NMR 43*, 145.
(79) Pritz, S., Wolf, Y., Kraetke, O., Klose, J., Bienert, M., and Beyermann, M. (2007) Synthesis of Biologically Active Peptide Nucleic Acid–Peptide Conjugates by Sortase-Mediated Ligation. *J. Org. Chem. 72*, 3909–3912.
(80) Refaei, M. A., Combs, A., Kojetin, D. J., Cavanagh, J., Caperelli, C., Rance, M., Sapitro, J., and Tsang, P. (2011) Observing selected domains in multi-domain proteins via sortase-mediated ligation and NMR spectroscopy. *J. Biomol. NMR 49*, 3–7.
(81) Yoav, S., Barak, Y., Shamshoum, M., Borovok, I., Lamed, R., Dassa, B., Hadar, Y., Morag, E., and Bayer, E. A. (2017) How does cellulosome composition influence deconstruction of lignocellulosic substrates in Clostridium (Ruminiclostridium) thermocellum DSM 1313? *Biotechnol. Biofuels 10*, 222.
(82) Fierobe, H.-P., Bayer, E. A., Tardif, C., Czjzek, M., Mechaly, A., Bélaïch, A., Lamed, R., Shoham, Y., and Bélaïch, J.-P. (2002) Degradation of Cellulose Substrates by Cellulosome Chimeras SUBSTRATE TARGETING VERSUS PROXIMITY OF ENZYME COMPONENTS. *J. Biol. Chem. 277*, 49621–49630.
(83) Hirano, K., Nihei, S., Hasegawa, H., Haruki, M., and Hirano, N. (2015) Stoichiometric Assembly of the Cellulosome Generates Maximum Synergy for the Degradation of Crystalline Cellulose, as Revealed by In Vitro Reconstitution of the Clostridium thermocellum Cellulosome. *Appl. Environ. Microbiol. 81*, 4756–4766.
(84) Fierobe, H.-P., Mechaly, A., Tardif, C., Belaich, A., Lamed, R., Shoham, Y., Belaich, J.-P., and Bayer, E. A. (2001) Design and Production of Active Cellulosome Chimeras SELECTIVE INCORPORATION OF DOCKERIN-CONTAINING ENZYMES INTO DEFINED FUNCTIONAL COMPLEXES. *J. Biol. Chem. 276*, 21257–21261.
(85) Fierobe, H.-P., Mingardon, F., Mechaly, A., Bélaïch, A., Rincon, M. T., Pagès, S., Lamed, R., Tardif, C., Bélaïch, J.-P., and Bayer, E. A. (2005) Action of Designer Cellulosomes on Homogeneous Versus Complex Substrates CONTROLLED INCORPORATION OF THREE DISTINCT ENZYMES INTO A DEFINED TRIFUNCTIONAL SCAFFOLDIN. *J. Biol. Chem. 280*, 16325–16334.
(86) Krauss, J., Zverlov, V. V., and Schwarz, W. H. (2012) In Vitro Reconstitution of the Complete Clostridium thermocellum Cellulosome and Synergistic Activity on Crystalline Cellulose. *Appl. Environ. Microbiol. 78*, 4301–4307.
(87) Moraïs, S., Heyman, A., Barak, Y., Caspi, J., Wilson, D. B., Lamed, R., Shoseyov, O., and Bayer, E. A. (2010) Enhanced cellulose degradation by nano-complexed enzymes: Synergism between a scaffold-linked exoglucanase and a free endoglucanase. *J. Biotechnol. 147*, 205–211.
(88) Mitsuzawa, S., Kagawa, H., Li, Y., Chan, S. L., Paavola, C. D., and Trent, J. D. (2009) The rosettazyme: A synthetic cellulosome. *J. Biotechnol. 143*, 139–144.
(89) Olson, D. G., Tripathi, S. A., Giannone, R. J., Lo, J., Caiazza, N. C., Hogsett, D. A., Hettich, R. L., Guss, A. M., Dubrovsky, G., and Lynd, L. R. (2010) Deletion of the Cel48S cellulase from Clostridium thermocellum. *Proc. Natl. Acad. Sci. U. S. A. 107*, 17727–17732.
(90) Gefen, G., Anbar, M., Morag, E., Lamed, R., and Bayer, E. A. (2012) Enhanced cellulose degradation by targeted integration of a cohesin-fused β-glucosidase into the Clostridium thermocellum cellulosome. *Proc. Natl. Acad. Sci. U. S. A. 109*, 10298–10303.

https://dx.doi.org/10.1021/acssynbio.9b00407
ACS Synth. Biol. 2020, 9, 381–391

(91) Carrard, G., Koivula, A., Söderlund, H., and Béguin, P. (2000) Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. *Proc. Natl. Acad. Sci. U. S. A. 97*, 10342−10347.
(92) Hirano, K., Kurosaki, M., Nihei, S., Hasegawa, H., Shinoda, S., Haruki, M., and Hirano, N. (2016) Enzymatic diversity of the Clostridium thermocellum cellulosome is crucial for the degradation of crystalline cellulose and plant biomass. *Sci. Rep. 6*, 35709.
(93) Carrico, Z. M., Romanini, D. W., Mehl, R. A., and Francis, M. B. (2008) Oxidative coupling of peptides to a virus capsid containing unnatural amino acids. *Chem. Commun.*, 1205−1207.
(94) Miller, R. A., Presley, A. D., and Francis, M. B. (2007) Self-Assembling Light-Harvesting Systems from Synthetically Modified Tobacco Mosaic Virus Coat Proteins. *J. Am. Chem. Soc. 129*, 3104−3109.
(95) Patterson, D. P., Prevelige, P. E., and Douglas, T. (2012) Nanoreactors by Programmed Enzyme Encapsulation Inside the Capsid of the Bacteriophage P22. *ACS Nano 6*, 5000−5009.
(96) Patterson, D. P., Schwarz, B., Waters, R. S., Gedeon, T., and Douglas, T. (2014) Encapsulation of an Enzyme Cascade within the Bacteriophage P22 Virus-Like Particle. *ACS Chem. Biol. 9*, 359−365.
(97) Glasgow, J. E., Asensio, M. A., Jakobson, C. M., Francis, M. B., and Tullman-Ercek, D. (2015) Influence of Electrostatics on Small Molecule Flux through a Protein Nanoreactor. *ACS Synth. Biol. 4*, 1011−1019.
(98) ElSohly, A. M., Netirojjanakul, C., Aanei, I. L., Jager, A., Bendall, S. C., Farkas, M. E., Nolan, G. P., and Francis, M. B. (2015) Synthetically Modified Viral Capsids as Versatile Carriers for Use in Antibody-Based Cell Targeting. *Bioconjugate Chem. 26*, 1590−1596.
(99) Wörsdörfer, B., Woycechowsky, K. J., and Hilvert, D. (2011) Directed Evolution of a Protein Container. *Science 331*, 589−592.

https://dx.doi.org/10.1021/acssynbio.9b00407
ACS Synth. Biol. 2020, 9, 381−391

Chapter 6

Protein Labeling via a Specific Lysine-Isopeptide Bond Using the Pilin Polymerizing Sortase from *Corynebacterium diphtheriae*

6.1 Overview

Methods to site-specifically label proteins with peptide conjugates or protein fusions have many applications across biomedical and research fields. In particular, methods capable of installing isopeptide linkages are especially useful as these linkages are proteolytically stable and amenable to insertions at internal sites within proteins. Building on our knowledge of *Corynebacterium diphtheriae* pilus biogenesis described in Chapter 2, we have developed a new bioconjugation tool using rational mutagenesis to further activate the enzyme. In Chapter 6, we demonstrate its utility as a tool to install fluorogenic peptides, create protein fusions, and carry out orthogonal modifications at two sites on a single protein when combined with other bioconjugation tools.

This chapter is reprinted with permission from a peer-reviewed article “Protein Labeling via a Specific Lysine-Isopeptide Bond Using the Pilin Polymerizing Sortase from *Corynebacterium diphtheriae*.” 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. *J. Am. Chem. Soc.* **140** 8420-8423 (2018).

6.2 Protein Labeling via a Specific Lysine-Isopeptide Bond Using the Pilin Polymerizing Sortase from *Corynebacterium diphtheriae*

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
Cite This: J. Am. Chem. Soc. 2018, 140, 8420−8423
pubs.acs.org/JACS

Protein Labeling via a Specific Lysine-Isopeptide Bond Using the Pilin Polymerizing Sortase from *Corynebacterium diphtheriae*

Scott A. McConnell,[†] Brendan R. Amer,[†] John Muroski,[†] Janine Fu,[†] Chungyu Chang,[§] Rachel R. Ogorzalek Loo,[†] Joseph A. Loo,[†] Jerzy Osipiuk,[‡] Hung Ton-That,[§] and Robert T. Clubb*[,†]

[†]Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics and the Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California 90095, United States

[‡]Structural Biology Center, Argonne National Laboratory, Argonne, Illinois 60439, United States

[§]Department of Microbiology & Molecular Genetics, University of Texas Health Science Center, Houston, Texas 77030, United States

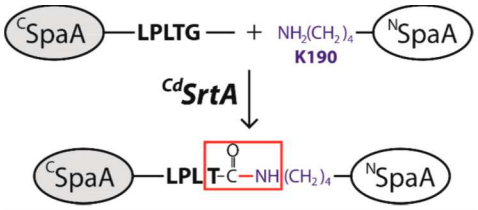
Supporting Information

ABSTRACT: Proteins that are site-specifically modified with peptides and chemicals can be used as novel therapeutics, imaging tools, diagnostic reagents and materials. However, there are few enzyme-catalyzed methods currently available to selectively conjugate peptides to internal sites within proteins. Here we show that a pilus-specific sortase enzyme from *Corynebacterium diphtheriae* (^{Cd}SrtA) can be used to attach a peptide to a protein via a specific lysine-isopeptide bond. Using rational mutagenesis we created ^{Cd}SrtA^{3M}, a highly activated cysteine transpeptidase that catalyzes *in vitro* isopeptide bond formation. ^{Cd}SrtA^{3M} mediates bioconjugation to a specific lysine residue within a fused domain derived from the corynebacterial SpaA protein. Peptide modification yields greater than >95% can be achieved. We demonstrate that ^{Cd}SrtA^{3M} can be used in concert with the *Staphylococcus aureus* SrtA enzyme, enabling dual, orthogonal protein labeling via lysine-isopeptide and backbone-peptide bonds.

Enzymatic methods that site-specifically functionalize proteins are of significant interest, as they can enable the creation of novel protein-conjugates for medical and research applications.[1−5] The *Staphylococcus aureus* sortase (^{Sa}SrtA) has been developed into a powerful protein engineering tool.[6−10] It catalyzes a transpeptidation reaction that covalently modifies the target protein via a backbone peptide bond, by joining peptide segments that contain a LPXTG "sort-tag" and an N-terminal oligoglycine amine group.[11,12] Several groups have now optimized this reaction to modify proteins with a range of molecules, including drugs, lipids, sugars, fluorophores, and peptides.[13−21] While ^{Sa}SrtA is potent tool, it is almost exclusively used to modify target proteins at their N- or C-termini, while it labels internal lysine side chains as a side reaction with low sequence specificity.[15,22,23] Here we show that a mutationally activated sortase enzyme from *Corynebacterium diphtheriae* (^{Cd}SrtA) can site-specifically install a peptide on a protein via a lysine-isopeptide bond. ^{Cd}SrtA and ^{Sa}SrtA have orthogonal activities, enabling dual peptide-fluorophore labeling of a protein via lysine isopeptide- and backbone peptide-bonds, respectively.

Gram-positive bacteria use specialized sortase enzymes to construct pili: long, thin fibers (0.2−3.0 μm × 2−10 nm) that project from the cell surface to mediate bacterial adherence to host tissues, biofilm formation and host immunity modulation.[24−26] These structures are distinct from pili produced by Gram-negative bacteria because their protein subunits (called pilins) are cross-linked by lysine-isopeptide bonds that confer enormous tensile strength.[27,28] Recently, we reconstituted *in vitro* the assembly reaction that builds the archetypal SpaA-pilus in *C. diphtheriae*, the causative agent of pharyngeal diphtheria.[29] ^{Cd}SrtA functions as a pilin polymerase, performing a repetitive transpeptidation reaction that covalently links adjacent SpaA pilin subunits together via lysine-isopeptide bonds. As shown in Scheme 1, ^{Cd}SrtA cross-links adjacent

Scheme 1. ^{Cd}SrtA-Catalyzed Isopeptide Bond Formation



SpaA proteins by connecting their N- (^{N}SpaA, residues 30−194) and C-terminal (^{C}SpaA, residues 350−500) domains, which contain a reactive WxxxVxVYPK pilin motif and LPLTG sorting signal sequences, respectively. In the reaction, ^{Cd}SrtA first cleaves the LPLTG sequence in ^{C}SpaA between the threonine and glycine, forming an acyl-enzyme intermediate in which the catalytic C222 residue in ^{Cd}SrtA is joined to ^{C}SpaA's threonine carbonyl atom. This transient intermediate is then nucleophilically attacked by the reactive K190 within ^{N}SpaA's pilin motif resulting in a T494-K190 isopeptide bond between

Received: May 17, 2018
Published: June 21, 2018

DOI: 10.1021/jacs.8b05200
J. Am. Chem. Soc. 2018, 140, 8420−8423

^{C}SpaA and ^{N}SpaA domains within adjacent pilin subunits. Previously, we demonstrated that wild-type ^{Cd}SrtA is catalytically inactive *in vitro* due to the presence of an N-terminal polypeptide segment, called a lid, that masks the enzyme's active site (Figure 1A).[30−34] Moreover, we showed that it was

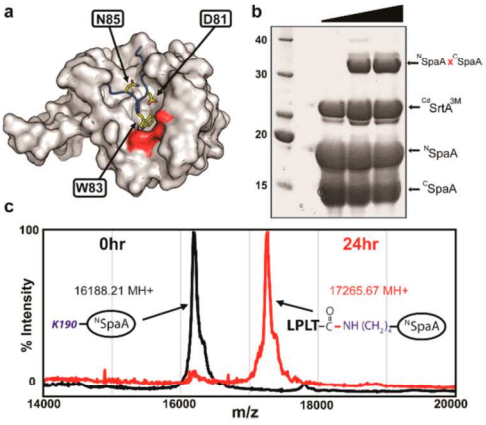


Figure 1. Mutationally activated ^{Cd}SrtA catalyzes lysine-isopeptide bond formation. (A) The structure of ^{Cd}SrtA^{WT} harbors an inhibitory "lid" structure (blue). Side chains that were mutated to activate the enzyme are shown as yellow sticks. The surface of the catalytic site is colored red. (B) Protein−protein ligation using the activated ^{Cd}SrtA^{3M} enzyme. SDS-PAGE analysis of the reaction demonstrating formation of the lysine-isopeptide linked ^{N}SpaAx^{C}SpaA product. The reaction (100 μM enzyme, 300 μM ^{C}SpaA and ^{N}SpaA) was sampled at 0, 24, and 48 h. (C) High yield protein-peptide labeling with ^{Cd}SrtA^{3M}. MALDI-MS data showing that >95% ^{N}SpaA is labeled with peptide containing the sort-tag, LPLTG$_{peptide}$. MALDI-MS spectra recorded at 0 (black) and 24 h (red) are overlaid.

possible to activate the enzyme by introducing D81G and W83G lid mutations and we demonstrated that a soluble catalytic domain harboring these mutations (^{Cd}SrtA^{2M}, residues 37−257 of ^{Cd}SrtA with D81G/W83G mutations) site-specifically ligates the isolated ^{N}SpaA and ^{C}SpaA domains *in vitro*.[29]

Toward the goal of creating a lysine modifying bioconjugation reagent we improved the ligation activity of ^{Cd}SrtA^{2M}, we defined substrate determinants that are required for catalysis. In addition to the aforementioned D81 and W83 mutations in ^{Cd}SrtA^{2M}, inspection of the crystal structure reveals three lid residues that may stabilize its positioning over the active site (I79, N85, K89). The ligation activities of triple mutants of ^{Cd}SrtA containing the D81G and W83G alterations, as well I79R, N85A or K89A substitutions were determined. A D81G/W83G/N85A triple mutant, hereafter called ^{Cd}SrtA^{3M}, has the highest level of ligation activity (Figures 1B and S1). After a 24 h incubation with the isolated ^{N}SpaA and ^{C}SpaA domains, ^{Cd}SrtA^{3M} produces 10.6-fold more cross-linked ^{N}SpaAx^{C}SpaA product than ^{Cd}SrtA WT and 35% more product than ^{Cd}SrtA^{2M} (Figure S1). The mutations in ^{Cd}SrtA^{3M} presumably further displace its lid, thereby facilitating enhanced binding of ^{C}SpaA's LPLTG sorting signal and subsequent acylation by C222. This is substantiated by our finding that the ^{Cd}SrtA^{3M} triple mutant exhibits the highest level of activity in a HPLC-based sorting signal cleavage assay that reports on formation of the acyl-enzyme intermediate (Figure S1) and previous studies that have shown that alterations in the lid increase C222 reactivity with 4,4′-dithiodipyridine.[29]

^{N}SpaA and ^{C}SpaA are joined by ^{Cd}SrtA^{3M} via their respective pilin motif and LPXTG sorting signal elements. To elucidate determinants required for recognition of the K190 nucleophile, ^{Cd}SrtA^{3M} was incubated with a peptide containing the pilin motif (DGWLQDVHVYPKHQALS) and either ^{C}SpaA or a peptide containing its C-terminal sorting signal (KNAGFELPLTGGSGRI) (Figure S2). In both instances, no detectable product was observed, indicating that ^{Cd}SrtA^{3M} requires additional tertiary elements within ^{N}SpaA to recognize K190. In contrast, when ^{Cd}SrtA^{3M} is incubated with ^{N}SpaA and the peptide containing the C-terminal sorting signal, > 95% of ^{N}SpaA is labeled with the peptide (Figure 1C). Moreover, LC-MS/MS analysis of the cross-linked species reveals that the components are joined via a site-specific isopeptide bond between the threonine within the sorting signal peptide and the Nε amine of K190 in ^{N}SpaA (Figure S3A).

We next demonstrated that ^{Cd}SrtA^{3M} can be used to label a target protein via an isopeptide bond with either a peptide fluorophore or another protein. In the labeling reaction a target protein is first expressed as a fusion with the ^{N}SpaA domain containing the pilin motif (hereafter called PM), and then reacted with a LPLTG-containing biomolecule and ^{Cd}SrtA^{3M} (Figure 2A). To demonstrate peptide fluorophore attachment using ^{Cd}SrtA^{3M}, we incubated the enzyme with ^{N}SpaA and a

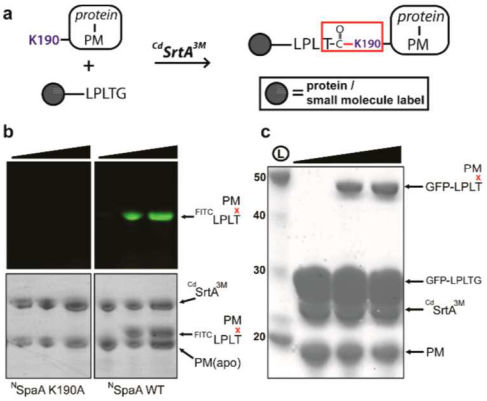


Figure 2. Labeling proteins via a lysine-isopeptide bond with ^{Cd}SrtA^{3M}. (A) Schematic showing ^{Cd}SrtA^{3M} catalyzed labeling of pilin motif (PM) fusion protein with a protein containing the LPLTG sorting signal or a LPLTG peptide with a functional label. (B) SDS-PAGE analysis of a fluoropeptide modification reaction containing ^{Cd}SrtA^{3M} (100 μM) and ^{FITC}LPLTG (1 mM) and either ^{N}SpaA (K190A) (lanes 1−3) or ^{N}SpaA WT (lanes 4−6) (both 100 μM). Top and bottom panels are the same gels visualized by fluorescence or by Coomassie staining, respectively. Reaction progress was measured at 0 (lanes 1,4), 24 (lanes 2,5) and 48 h (lanes 3,6). (C) Protein−protein ligation with ^{Cd}SrtA^{3M}. As in panel (B), except reactions contained GFP-LPLTG (300 μM) instead of the fluoropeptide. Reactants were visualized with Coomassie staining at 0, 24, and 48 h (lanes 1−3, respectively).

DOI: 10.1021/jacs.8b05200
J. Am. Chem. Soc. 2018, 140, 8420−8423

fluorescent FITC-KNAGFELPLTGGSGRI peptide (^{FITC}LPLTG). After incubating for either 24 or 48 h the reaction components were separated by SDS-PAGE and visualized by either Coomassie staining or FITC fluorescence at 530 nm. ^{Cd}SrtA^{3M} labels ^{N}SpaA with the fluorescent peptide, yielding a ^{FITC}LPLTx^{N}SpaA cross-linked product (Figure 2B, right). Fluorophore labeling is specific, as ^{N}SpaA harboring a K190A mutation is unreactive in control experiments (Figure 2B, left).

To demonstrate that ^{Cd}SrtA^{3M} can also be used to join proteins together via an isopeptide bond, the isolated PM was reacted with green fluorescent protein engineered to contain a C-terminal LPLTGGSGRI sorting signal sequence (GFP-LPLTG). Incubation of these proteins with ^{Cd}SrtA^{3M} resulted in the appearance a higher molecular weight GFP-LPLTx^{N}SpaA cross-linked product (Figure 2C, S3B). Notably, the ^{Cd}SrtA^{3M} protein–protein ligation reaction is versatile, as labeling can be achieved with the PM fused to either the N- or C-terminus of the target protein.

The ^{Cd}SrtA and ^{Sa}SrtA enzymes recognize distinct nucleophiles, suggesting that they can be used orthogonally to selectively label a single target protein at different sites. To demonstrate orthogonal labeling we created a fusion protein that contained the Small Ubiquitin-like Modifier (SUMO) protein harboring a pentaglycine peptide and PM at its N- and C-termini, respectively (G_5-SUMO$_{PM}$). Our dual modification approach involves sequential reaction of the G_5-SUMO$_{PM}$ substrate with each sortase and peptide fluorophores containing the cognate sorting signal, as outlined in Figure 3A. To selectively modify G_5-SUMO$_{PM}$ (species 1), it was first incubated with ^{Cd}SrtA^{3M} and FITC-LPLTG$_{pep}$ to create at high yield G_5-SUMO$_{PM}$-FITC (species 2) (Figure 3B). After removal of excess FITC-LPLTG peptide using a desalting column, the target protein was then labeled at its N-terminus with AlexaFluor$_{546}$-LPATG using ^{Sa}SrtA. This was achieved by incubating species 2 with ^{Sa}SrtA and AlexaFluor$_{546}$-LPATG to produce the doubly labeled protein (species 3). Separation of the reaction products by SDS-PAGE confirms dual labeling, as the expected fluorescence for each probe is detected at ~33 kDa during the procedure (Figure 3B). In particular, FITC-labeled Gly$_5$-SUMO$_{PM}$ is produced after treatment with ^{Cd}SrtA^{3M} (488/530 nm excitation/emission), and persists after treatment with ^{Sa}SrtA that catalyzes the second conjugation step with AlexaFluor$_{546}$ (532/605 nm excitation/emission). We note that a similar labeling strategy can presumably be used for fusion proteins that contain the SpaB basal pilin instead of ^{N}SpaA, as we have recently shown that ^{Cd}SrtA^{3M} can also use SpaB as a nucleophile *in vitro*.[29] A strength of our approach is the distinct nucleophile and sorting signal substrate specificities of each sortase, which limits cross reactivity. In addition to recognizing distinct nucleophiles, our findings indicate that the sortases have unique sorting signal substrate specificities; ^{Cd}SrtA^{3M} is unable to hydrolyze or use as a transpeptidation substrate sorting signals containing the sequence LPATG that is readily used by ^{Sa}SrtA, but instead it is selective for peptides containing LPLTG (Figures S4, S5). Moreover, the isopeptide bond created by ^{Cd}SrtA^{3M} is not significantly hydrolyzed by ^{Sa}SrtA or ^{Cd}SrtA after 24 h (Figure S6). Thus, ^{Cd}SrtA acts preferentially on its LPLTG sorting signal substrate, preventing potential reversal of LPATG peptides installed by ^{Sa}SrtA. Similarly, the isopeptide linkages installed by ^{Cd}SrtA are not a substrate for reversal by ^{Sa}SrtA or ^{Cd}SrtA.

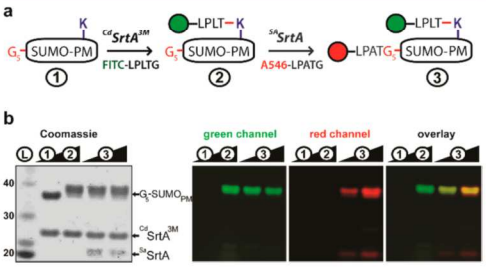


Figure 3. Orthogonal protein labeling using ^{Cd}SrtA^{3M} and ^{Sa}SrtA. (A) Sequential reaction scheme used to install fluorogenic peptides on a target protein via peptide- and isopeptide bonds. G_5-SUMO$_{PM}$ is a SUMO target protein that is fused to N- and C-terminal nucleophiles, pentaglycine (G_5) and the pilin motif (PM), respectively. (B) SDS-PAGE analysis of reaction mixture taken at different steps in the procedure: (1) prior to labeling, (2) after labeling with ^{FITC}LPLTG using ^{Cd}SrtA^{3M}, and (3) after labeling with A546-LPATG using ^{Sa}SrtA (0.25/2 h incubations). Panels show as indicated fluorescence gel imaging to detect FITC and A546 fluorophores using 488/530 (green channel) and 532/605 nm (red channel) wavelengths for excitation/emission, respectively, and the merged image of the gels demonstrating dual labeling. In the first panel, the same gel was visualized by coomassie staining.

The bioconjugation chemistry catalyzed by ^{Cd}SrtA^{3M} enables site-specific lysine labeling of a protein, creating an isopeptide linkage that may be less susceptible to proteolysis than conventional peptide bonds. An attractive feature of ^{Cd}SrtA^{3M} is its high degree of specificity for the ε-amine nucleophile within the pilin motif, which enables selective labeling. Transglutaminases can also modify protein lysine residues, but unlike ^{Cd}SrtA^{3M}, these enzymes exhibit minimal substrate specifity.[35,36] Similarly, ^{Sa}SrtA can modify lysines as a side reaction that occurs with minimal specificity and at low efficiency because the lysine ε-amine is not ^{Sa}SrtA's natural substrate.[15,22,23] Chemical methods that modify amino acid side chains have also been developed, but they often require cysteine or non-natural amino acid incorporation into the protein and in some instances harsh reaction conditions.[37] The bioconjugation chemistry catalyzed by ^{Cd}SrtA^{3M} is functionally similar to the nonenzymatic SpyTag/SpyCatcher system,[38,39] but its enzymatic activity affords greater control making ^{Cd}SrtA^{3M} an attractive new tool to engineer proteins.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b05200.

Additional data and procedures (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*rclubb@mbi.ucla.edu

ORCID

Robert T. Clubb: 0000-0001-5718-3985

Notes

The authors declare no competing financial interest.

DOI: 10.1021/jacs.8b05200
J. Am. Chem. Soc. 2018, 140, 8420–8423

■ 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 AI52217 (R.T.C. and H. T-T.), DE025015 (H. T-T), GM103479 (J.A.L.) and U.S. Department of Energy, Office of Biological and Environmental Research contract DE-AC02-06CH11357 (J.O.). S.A.M. was supported by a Cellular and Molecular Biology Training Grant (Ruth L. Kirschstein National Research Service Award GM007185). NMR equipment used in this research was purchased using funds from shared equipment grant NIH S10OD016336.

■ REFERENCES

(1) Proft, T. *Biotechnol. Lett.* **2010**, *32*, 1−10.
(2) Matsumoto, T.; Tanaka, T.; Kondo, A. *Biotechnol. J.* **2012**, *7*, 1137−1146.
(3) Walper, S. A.; Turner, K. B.; Medintz, I. L. *Curr. Opin. Biotechnol.* **2015**, *34*, 232−241.
(4) Krall, N.; Da Cruz, F. P.; Boutureira, O.; Bernardes, G. J. L. *Nat. Chem.* **2016**, *8*, 103−113.
(5) Rashidian, M.; Dozier, J. K.; Distefano, M. D. *Bioconjugate Chem.* **2013**, *24*, 1277−1294.
(6) Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O. *Science (Washington, DC, U. S.)* **1999**, *285*, 760−763.
(7) Antos, J. M.; Chew, G. L.; Guimaraes, C. P.; Yoder, N. C.; Grotenbreg, G. M.; Popp, M. W. L.; Ploegh, H. L. *J. Am. Chem. Soc.* **2009**, *131*, 10800−10801.
(8) Williamson, D. J.; Fascione, M. A.; Webb, M. E.; Turnbull, W. B. *Angew. Chem., Int. Ed.* **2012**, *51*, 9377−9380.
(9) Levary, D. A.; Parthasarathy, R.; Boder, E. T.; Ackerman, M. E. *PLoS One* **2011**, *6*, No. e18342.
(10) Tsukiji, S.; Nagamune, T. *ChemBioChem* **2009**, *10*, 787−798.
(11) Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. *Nat. Chem. Biol.* **2007**, *3*, 707−708.
(12) Mao, H.; Hart, S. A.; Schink, A.; Pollok, B. A. *J. Am. Chem. Soc.* **2004**, *126*, 2670−2671.
(13) Samantaray, S.; Marathe, U.; Dasgupta, S.; Nandicoori, V. K.; Roy, R. P. *J. Am. Chem. Soc.* **2008**, *130*, 2132−2133.
(14) Antos, J. M.; Miller, G. M.; Grotenbreg, G. M.; Ploegh, H. L. *J. Am. Chem. Soc.* **2008**, *130*, 16338−16343.
(15) Möhlmann, S.; Mahlert, C.; Greven, S.; Scholz, P.; Harrenga, A. *ChemBioChem* **2011**, *12*, 1774−1780.
(16) Wagner, K.; Kwakkenbos, M. J.; Claassen, Y. B.; Maijoor, K.; Böhne, M.; van der Sluijs, K. F.; Witte, M. D.; van Zoelen, D. J.; Cornelissen, L. A.; Beaumont, T.; Bakker, A. Q.; Ploegh, H. L.; Spits, H. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 16820−16825.
(17) Beerli, R. R.; Hell, T.; Merkel, A. S.; Grawunder, U. *PLoS One* **2015**, *10*, No. e0131177.
(18) Amer, B. R.; MacDonald, R.; Jacobitz, A. W.; Liauw, B.; Clubb, R. T. *J. Biomol. NMR* **2016**, *64*, 197−205.
(19) Dorr, B. M.; Ham, H. O.; An, C.; Chaikof, E. L.; Liu, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13343−13348.
(20) Chen, I.; Dorr, B. M.; Liu, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 11399−11404.
(21) Antos, J. M.; Truttmann, M. C.; Ploegh, H. L. *Curr. Opin. Struct. Biol.* **2016**, *38*, 111−118.
(22) Dasgupta, S.; Samantaray, S.; Sahal, D.; Roy, R. P. *J. Biol. Chem.* **2011**, *286*, 23996−24006.
(23) Bellucci, J. J.; Bhattacharyya, J.; Chilkoti, A. *Angew. Chem., Int. Ed.* **2014**, *54*, 441−445.
(24) Danne, C.; Dramsi, S. *Res. Microbiol.* **2012**, *163*, 645−658.
(25) Spirig, T.; Weiner, E. M.; Clubb, R. T. *Mol. Microbiol.* **2011**, *82*, 1044−1059.
(26) Ton-That, H.; Schneewind, O. *Trends Microbiol.* **2004**, *12*, 228−234.
(27) Echelman, D. J.; Alegre-Cebollada, J.; Badilla, C. L.; Chang, C.; Ton-That, H.; Fernández, J. M. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2490−2495.
(28) Yeates, T. O.; Clubb, R. T. *Science* **2007**, *318*, 1558−1559.
(29) 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.; Ton-That, H. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E5477−E5486.
(30) Jacobitz, A. W.; Naziga, E. B.; Yi, S. W.; McConnell, S. A.; Peterson, R.; Jung, M. E.; Clubb, R. T.; Wereszczynski, J. *J. Phys. Chem. B* **2016**, *120*, 8302−8312.
(31) Manzano, C.; Izoré, T.; Job, V.; Di Guilmi, A. M.; Dessen, A. *Biochemistry* **2009**, *48*, 10549−10557.
(32) Cozzi, R.; Zerbini, F.; Assfalg, M.; D'Onofrio, M.; Biagini, M.; Martinelli, M.; Nuccitelli, A.; Norais, N.; Telford, J. L.; Maione, D.; Rinaudo, C. D. *FASEB J.* **2013**, *27*, 3144−3154.
(33) Persson, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 212−217.
(34) Manzano, C.; Contreras-Martel, C.; El Mortaji, L.; Izoré, T.; Fenel, D.; Vernet, T.; Schoehn, G.; Di Guilmi, A. M.; Dessen, A. *Structure* **2008**, *16*, 1838−1848.
(35) Yokoyama, K.; Nio, N.; Kikuchi, Y. *Appl. Microbiol. Biotechnol.* **2004**, *64*, 447−454.
(36) Fontana, A.; Spolaore, B.; Mero, A.; Veronese, F. M. *Adv. Drug Delivery Rev.* **2008**, *60*, 13−28.
(37) Spicer, C. D.; Davis, B. G. *Nat. Commun.* **2014**, *5*, DOI: 10.1038/ncomms5740.
(38) Reddington, S. C.; Howarth, M. *Curr. Opin. Chem. Biol.* **2015**, *29*, 94−99.
(39) Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E690−E697.

DOI: 10.1021/jacs.8b05200
J. Am. Chem. Soc. 2018, 140, 8420−8423

Chapter 7

Selection Scheme and Discovery of Improved Variants of the ^{Cd}SrtA Polymerase Enzyme for Bioconjugation

7.1 Directed Evolution Overview

Over a very long time scale, natural evolution in every living organism on Earth has produced advantageous phenotypes which are well adapted for diverse environmental conditions. Random mutations in chromosomal DNA gives rise to aberrant protein variants which may exhibit slightly different phenotypes than the wild-type gene product. Most of the time, these replication mistakes are deleterious to the organism and result in decreased fitness levels and are thus selected against. Other time, the mutations have no functional effect on the phonotype. In rare circumstances, the genetic variation may result in a phenotype which is useful to the organism, i.e. it imparts enhanced fitness. Thus, the basis for natural selection is rare, but important, disfunction in the replication of genetic material.

For the specific applications in research, industry and therapeutic development, scientists have recently harnessed this basic concept to accelerate the generation of genetic variants and gene products with desired properties on a much shorter time scale. Deliberate alterations to genetic material are introduced to produce highly diverse protein libraries, from which members with desired properties are isolated by various selection methods. The process involves an iterative workflow which alternates between generation of diverse genetic libraries and screening or selection of functional variants. The top performing variants from previous rounds are typically used as templates for further randomization in later rounds, such that beneficial mutations accumulate throughout the evolutionary campaign. In this way, large regions of the possible sequence space of a given gene can be accessed rapidly by introduction of mutations to sites within the wild-type scaffold, resulting in impressive functional enhancements as compared to progenitor molecule. Laboratory evolution of biomolecules has proven to be a powerful strategy for improving or altering the activity of target biomolecules[1,2].

Genetic diversity is the critical starting point for any directed evolution approach. Researchers must first decide the method by which they will randomize their gene of interest. In

the most general case, completely random mutagenesis is applied to the gene of interest, often by an error-prone DNA replication step[3–5]. Treatment of template strands with mutagenizing chemicals can also increase mutation rates[6,7]. The protocol can be toggled to introduce more, or fewer, mutations to each library member on average. With this method, mutations are evenly distributed throughout the entire gene, at every single nucleotide position. The result is an evenly randomized library with mutations spanning the entire gene. Where structures of the gene product have been determined or extensive biochemical knowledge about the biomolecule is available, it is possible to focus randomization only on specific regions. In these situations, it is possible to avoid screening variants harboring mutations at positions with a low probability of affecting activity. It is also possible to explore the sequence space of single sites much more thoroughly (i.e. every possible amino acid substitution at a certain position can be tested)[8,9]. By simultaneously randomizing two or more amino acid positions which are known to interact with one another, cooperative effects are observed in favorable cases. Finally, homologous recombination is an approach which shuffles fragments from existing genes to combine useful variations from many different homologous sequences with natural diversity into a single evolved gene with novel characteristics[10,11]. Recently, advanced computational algorithms have been developed to maximize the probability of a high-quality, folded variant library by defining consensus sequences across many species to identify optimal crossover sites for recombination of homologous proteins fragments[12–14].

The second decision in a directed evolution campaign is the method of identifying protein variants with desired characteristics. In all approaches, it is critical to maintain genotype-phenotype coupling: the protein variant that exhibits a certain characteristic must be linked to the gene that encodes it. Screening approaches are the first category of directed evolution assays. In screens, each individual variant is assayed and ranked by some criteria. As a result, these experiments can be extremely data-rich, but also relatively time-consuming. A large

amount of information about the range of the phenotype gradient can be obtained, which is often useful for informing threshold limits for subsequent rounds of selection, but the intensive nature of these approaches often limits the library to a smaller size for practical reasons. One such approach involves spatial separation of cells expressing each variant in multi-well microtiter plates (MTPs)[15–17]. The resultant proteins can be purified and flexibly assayed by almost any assay that can be rapidly conducted. This approach is required if specialized assays are necessary, but is severely limited in its throughput. Directed evolution can also be coupled to fluorescence-activated cell sorting (FACS)[18,19]. This approach is especially well suited for enzymes for which a fluorescence reporter assay can be developed. FACS-based screen can screen libraries that are larger than MTP-based screens by ~4 orders of magnitude. Another advantage of FACS screening is that it enables counter-screens to deplete undesirable phenotypes from libraries[20]. *In vitro* compartmentalization (IVC) is another screening method where emulsions of single cells or cell-free expression systems are screened for activity by flow cytometry or other methods[21]. When the emulsion is broken, the DNA encoding the variants can be recovered.

The second category of phenotype selection in directed evolution are bulk selections. Selections remove the requirement to individually assess the phenotype of each variant by applying a selective pressure such that only variants with activity above a certain threshold will advance to the next round of the selection. Selection methods are massively high throughput (library sizes on the order of 10^{11}) and are limited only by the transformation bottleneck, which is unavoidable when relying bacterial expression of proteins. Selections based on binding affinity are common. In this approach, protein libraries are either displayed on the surface of cell or on the coat of phage and the corresponding genetic material is encased inside the cell or phage[21–23]. The libraries are then selected based on binding properties, while the nonbinding proteins are washed away and rejected. The linked genes are then easily isolated and sequenced for

further rounds of selection. The activity of the protein library can also be linked to organismal survival for purposes of screening. This is often accomplished by coupling enzyme activity with expression of an antibiotic resistance gene[24] or auxotroph complementation[25,26]. In both cases, the activity of protein or reporter of protein activity confers survival to the cell encoding that protein, while nonfunctional variants are rejected. Importantly, selections can often be progressively tuned to increase selective pressure in subsequent evolutionary rounds, such that the threshold for desired characteristic is continuously improved.

7.2 Directed Evolution of Sortases

Sortases are excellent candidates for directed evolution. First, they have impressive functional diversity (see Section 1.2). There are several families of sortase enzymes which all recognize different substrates and have different functions. Functional diversity in homologous proteins is considered a good predictor of successful in vitro evolution because theoretically, the ancestral protein progenitor underwent natural evolution in the past to produce the current assortment of homologs[27]. A degree of functional promiscuity is another promising trait, as small alterations to the parent protein can be expected to enhance this side reaction[28]. Indeed, sortase enzymes have a well-documented promiscuity in sorting signal and nucleophile substrate selection[29–31]. It has also been speculated that evolvable proteins often exist in a range of functionally diverse conformations and each separate conformation can be favored mutationally[32]. Again, sortases often have flexible loops which are implicated in reactivity[33–36]. It is not surprising then, that sortases have been the subject of many successful directed evolution approaches[15,16,18–20,37].

Despite the unique isopeptide-bond forming activity of pilin polymerizing sortases, no directed evolution approaches have addressed this enzyme class to date. The pilin polymerase from *Corynebacterium diphtheriae*, ^{Cd}SrtA, is the only enzyme of its class to be biochemically reconstituted with robust activity[38]. Based on the crystal structure, rational mutagenesis resulted in destabilization of the conserved “lid” structure which sterically blocks substrate access to catalytic residues, which activate this enzyme in vitro. Subsequently, additional mutations in the “lid” further activated this enzyme, paving the way for its development into a bioconjugation tool and full kinetic analysis of its isopeptide transpeptidation reaction in vitro[39,40] (see Section 1.5). Complete deletion of the “lid” yields an enzyme variant with significantly enhanced in vitro transpeptidation **(Figure 7.1)**.

However, this bioconjugation system still has serious drawbacks which will limit its widespread deployment. First, the transpeptidation reaction of ^{Cd}SrtA is still significantly slower than the canonical ^{Sa}SrtA sortase enzyme, requiring incubation times of up to 16 h to achieve complete modification of substrates, compared to minutes for improved variants of ^{Sa}SrtA[39]. Second, ^{Cd}SrtA is highly specific to a single lysine side chain, which is likely due to recognition of tertiary elements within the ^{N}SpaA acceptor domain. This bulky domain may be undesirable for many bioconjugation approaches, so minimization of this acceptor substrate is a priority. Reduced specificity would be advantageous because a minimal pilin motif peptide tag could be more easily engineered into proteins of interest in bioconjugation applications. Interestingly, ^{Cd}SrtA appears to recognize its acceptor domain through residues within a conserved motif within its β7/β8 loop[38], providing an attractive target for optimization. Finally, while destabilization of the inhibitory lid structure of ^{Cd}SrtA activates transpeptidation *in vitro*, it also causes a significant decrease in protein stability **(Figure 7.1C)**.

As such, future directed evolution efforts of ^{Cd}SrtA will follow a three-pronged approach: reprogramming of substrate recognition, improvement of thermal stability and acceleration of transpeptidation kinetics. For our purposes, thermal stability enhancements will likely require a random mutagenesis selection as stabilizing mutations are difficult to predict from the structure. The rest of this chapter describes efforts to develop and optimize directed evolution methods capable of achieving these goals.

7.3 Cell-Based DHFR Selection Approach

Directed evolution requires thorough exploration of the progenitor protein's sequence space to uncover variants with novel properties. In the absence of pre-existing knowledge to guide focused randomization, a random approach must be employed to search for mutations which impart enhanced reactivity. However, comprehensive randomization becomes unfeasible even for very short polypeptides (10^{13} unique combinations are possible for a 10-residue protein)[1]. Instead, library creation must focus on efficient sparse sampling of the sequence space. Random sampling of mutations throughout the sequence may identify positions that are important for catalysis or thermostability that are difficult to predict from the structure alone. However, a library of significant size would have to be generated to get sufficient sequence coverage, which would make manual screening onerous. Thus, we have developed a complementation-style cell viability selection to handle very large libraries.

The basic scheme for this selection involves co-transformation of ^{Cd}SrtA variants with a duet plasmid encoding two fragments of the murine dihydrofolate reductase (DHFR) gene based a previous successful selection scheme using ^{Sa}SrtA[37]. DHFR synthesizes tetrahydrofolate, an essential precursor for many metabolic processes in prokaryotic cells[41]. The C-terminal fragment of mDHFR expressed with an N-terminal fusion to the ^{N}SpaA acceptor domain and the N-terminal domain is expressed with a C-terminal signal peptide (LPLTG) **(Figure 7.1)**. The C-terminal DHFR fusion is encoded upstream of the N-terminal fragment to prevent stop codon readthrough which may result in a functional DHFR enzyme in the absence of sortase-mediated ligation. ^{Cd}SrtA variants are generated by error-prone PCR using a low-fidelity Taq polymerase with supplemented with 7 mM $MgCl_2$ and 0.25 mM $MnCl_2$ to increase error rates[42]. To minimize mutational bias, an unbalanced ratio of dGTP/dATP:dCTP/dTTP of 1:5 was employed. This protocol resulted in a mutational frequency which yielded 1-4 amino acid substitutions. The resultant variant genes were then introduced into expression plasmids

using a whole primer Megaprimer (MEGAWHOP) approach[43]. Finally, both plasmids were simultaneously transformed into *E. coli* BL21 cells, which were then cultured in minimal media lacking folates such that the cells would be dependent on DHFR-mediated folate production for viability. Endogenous bacterial DHFR was inhibited by trimethoprim, so that folate production is linked only to sortase-conjugated murine DHFR. In each cell, the metabolic defect will be resolved and the cell will be viable if the murine DHFR is rendered functional by timely transpeptidation via the ^{Cd}SrtA variant harbored in that cell. Conversely, if transpeptidation is too slow, the cell will be not be viable and is selected against. Selective pressure for variants with faster kinetics can be applied by shortening the transpeptidation window or increasing trimethoprim concentrations.

Preliminary data indicates that this approach is feasible for selection. Cells grown in trimethoprim have significantly slower growth as compared to uninhibited cultures. However, cells co-transformed with sortase and DHFR plasmids have significant growth advantages over cells with just DHFR plasmids, indicated that sortase is required for efficient ligation of the DHFR and that DHFR complementation yields a functional protein in the presence of the fusion tags added in this experiment **(Figure 7.2)**. Future experiments using this approach may select for ^{Cd}SrtA variants with enhanced bond forming kinetics or altered substrate specificity.

7.4 Fluorescence-Based Screen Approach

High-throughput fluorescence screening requires intensive evaluation of each variant, but also reduces the probability of missing top mutations. Screening is also compatible with focused randomization of specific sites. A popular approach combines initial identification of important positions through selection of random mutagenesis libraries, followed by focused mutagenesis of that position. Screens are lower throughput than selections, but provide more information about relative activities of each variant which is often lost in selections. Thus, we developed a screen along with our selection approach.

This screen is compatible with site-saturation mutagenesis library generation methods. Using established site saturation primers (NNK/NNS), 95% fractional library coverage can be achieved with a single 96-well MTP. Various approaches to reduce codon redundancy can also reduce screening effort dramatically. Specifically, the 22c-trick is one approach which uses a mixture of primers (NDT/VHG/TGG), which encode every single amino acid. Using this SSM approach, only 66 colonies need to be screened for the same fractional library coverage, and reductions in screening effort are even more pronounced when screen two or more sites by SSM simultaneously (50% and 300% less screening effort for two and three saturated sites, respectively)[8]. In our approach, we used three columns of the 96-well MTP for positive and negative controls and the remaining 72 wells contained variants randomized by the 22c-trick. Thus, we had greater than 95% fractional coverage of our library, meaning there is an excellent probablility that each plate contains every possible variant.

The basis of the screen is a fluorescence reporter for activity based on the peptide conjugation reaction catalyzed by ^{Cd}SrtA, as described in [40]. Briefly, ^{Cd}SrtA is capable of conjugating peptides encoding the LPLTG sorting signal motif to an acceptor domain (^{N}SpaA, residues 52-196) derived from its pilin substrate, SpaA. In this screen, a synthetic peptide is utilized which is conjugated to fluorescein group at its N-terminus. The recombinant ^{N}SpaA

acceptor is expressed with a C-terminal His_6 tag to facilitate purification. In the proof-of-principle study, we expressed and purified variants of ^{N}SpaA in an MTP format and added exogenous ^{Cd}SrtA and the fluorescent peptide after purification. It is important to note however, that this screen could also be carried out with randomized ^{Cd}SrtA expressed in the MTP and addition of exogenous wild-type ^{N}SpaA. After the peptide is crosslinked to the ^{N}SpaA acceptor during the incubation period, ^{Cd}SrtA and the signal peptide are removed from the wells by subsequent washes, while ^{N}SpaA is retained on the IMAC resin. The fluorescently labeled ^{N}SpaA acceptors in each well are then eluted to fresh 96-well plates for fluorescence readings **(Figure 7.3)**. Also note that the second round of wash steps can be omitted and the entire reaction can be analyzed by fluorescence anisotropy in a plate reader if the proper controls are used **(Figure 7.3B)**. The two techniques give directionally similar results, but small amounts of ^{Cd}SrtA-FITC-LPLT acyl intermediate in the unpurified approach contribute to noise in the anisotropy measurements. Based on fluorescence readings, variants are then ranked based on their degree of modification and top performers can be further assayed by orthogonal techniques to verify their transpeptidation kinetics. Importantly, because this screen involves an *in vitro* assay and is not dependent on cellular viability, heat shocks can be administered to purified proteins before screening to assess thermostability. Future applications of this screen could involve focused randomization of important sites on ^{Cd}SrtA or ^{N}SpaA from structural data or thermostability screens of ^{Cd}SrtA. Beyond that, future efforts may be directed to engineering the pilin motif into internal loops on test proteins and selecting for ^{Cd}SrtA variants with specificity for the pilin motif outside of the context of the ^{N}SpaA acceptor domain. Improved ^{Cd}SrtA variants with enhanced kinetic parameters, optimized nucleophile preference, and improved thermostability discovered by a tandem screening and selection scheme could be used as novel isopeptide ligation tools with important implications in bioconjugation.

7.5 Figures

Figure 7.1 – Comparison of ^{Cd}SrtA^{3M} and ^{Cd}SrtA$^{\Delta}$ variants. A) The transpeptidation activity of both ^{Cd}SrtA variants is tracked by a gel-based assay. ^{N}SpaA (harboring the reactive K190 lysine) is conjugated to ^{C}SpaA (harboring the signal peptide) and the isopeptide ^{N}SpaA-^{C}SpaA product is monitored by separation of the reaction by SDS-PAGE. B) Gel densitometry was used to quantify the product formation as a function of time. C) Differential scanning fluorimetry (DSF) data shows differences in thermal stability between the two ^{Cd}SrtA lid variants.

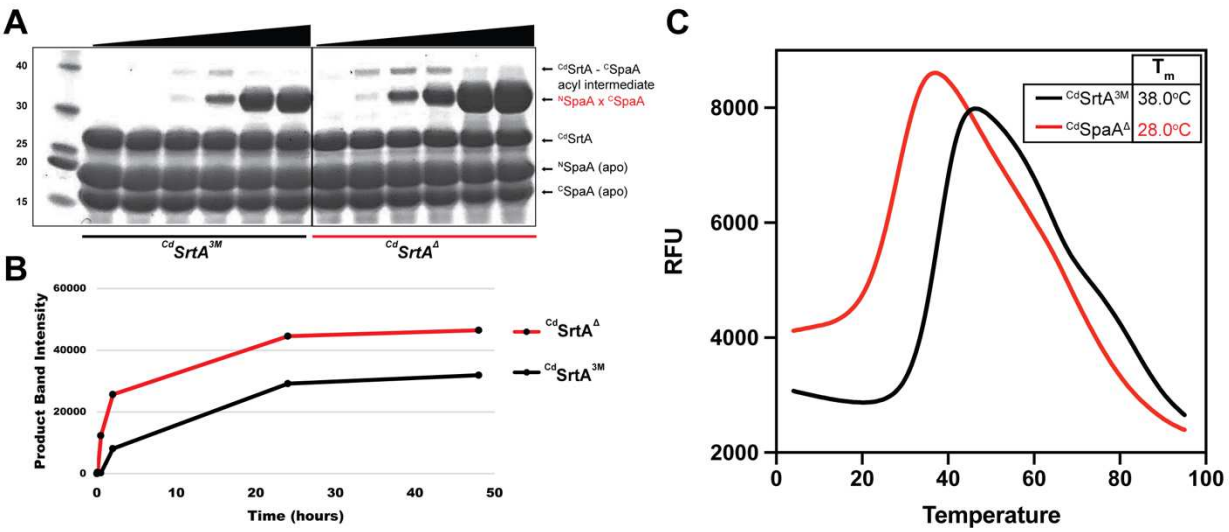


Figure 7.2 – DHFR fragment complementation selection. A) DHFR selection scheme. Randomized ^{Cd}SrtA variants are generated through error-prone PCR (epPCR) and co-transformed into *E. coli* BL21 cells. The DHFR fragments are expressed as fusions to the ^{N}SpaA acceptor and signal peptide donor substrates for the ^{Cd}SrtA-mediated transpeptidation reaction. The cells are then grown in the presence of trimethoprim, a bacterial DHFR inhibitor, such that efficient ligation of the two murine DHFR fragments is required for cell viability. During the course of the selection, ^{Cd}SrtA variants with improved kinetics will be more viable and predominate in the selection culture. After several days of selection, the cultures are plated and individual ^{Cd}SrtA variants are sequenced and their activity is quantified by *in vitro* assays. B) Crystal structure of murine DHFR (PDB 1U70)[43] with the C-terminal and N-terminal fragments colored red and blue, respectively. NADPH (green) and methotrexate (folate analog, cyan) are shown as sticks, localized to the active site. Both fragments of DHFR (red and blue) must be ligated together to form a competent active site. The position of the ^{N}SpaA and LPLTG signal peptide fusions in this approach are shown as red and blue spheres (fused to the DHFR-C and DHFR-N), respectively. C) Cell growth curves in media with different concentrations of trimethoprim inhibitor are displayed as various shades of grey or blue lines for strains with both ^{Cd}SrtA and the DHFR plasmids or just the DHFR plasmid, respectively. Top, trimethoprim growth curves are displayed along with uninhibited cultures as a line graph. Bottom, the same data is depicted as a bar graph.

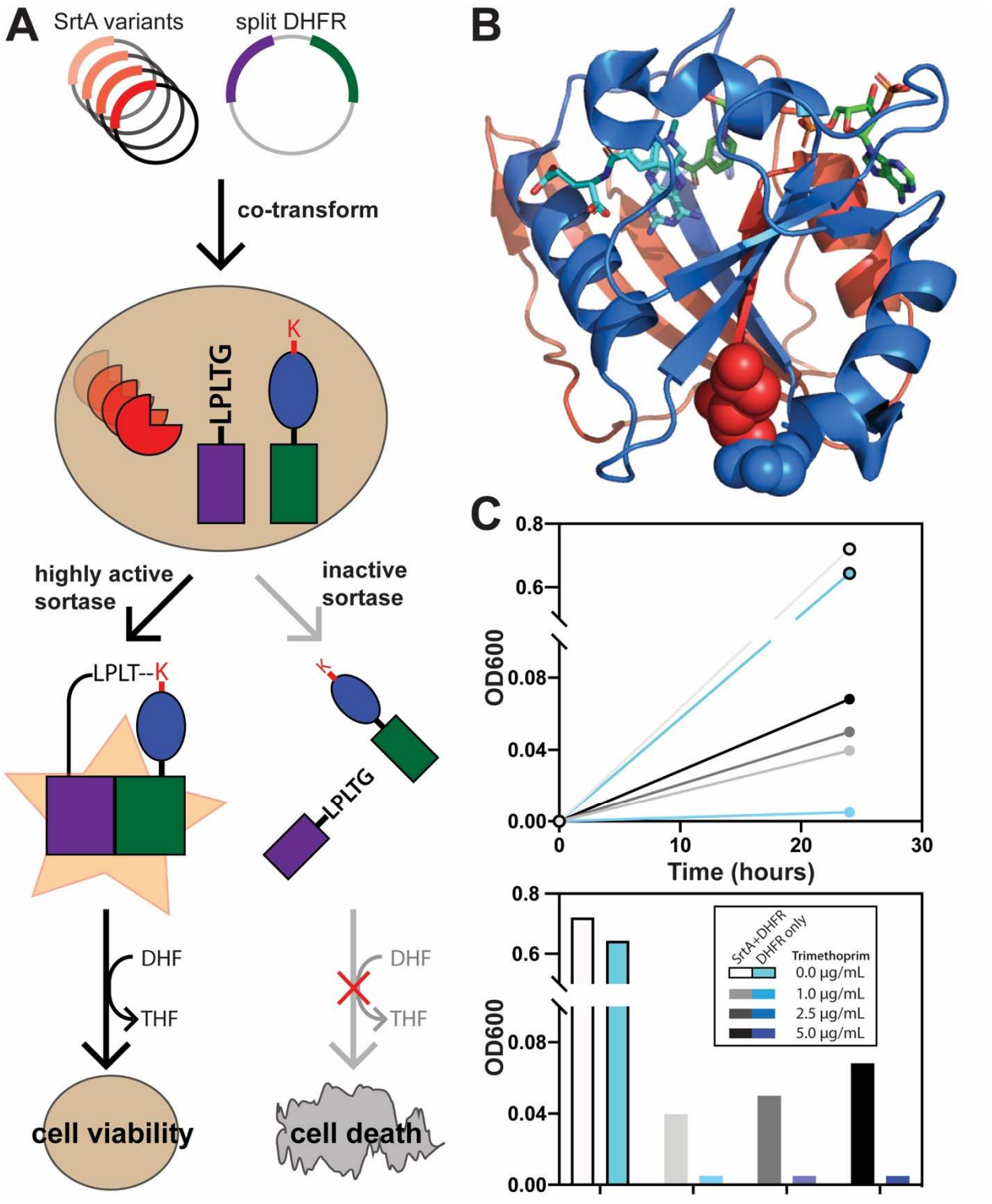
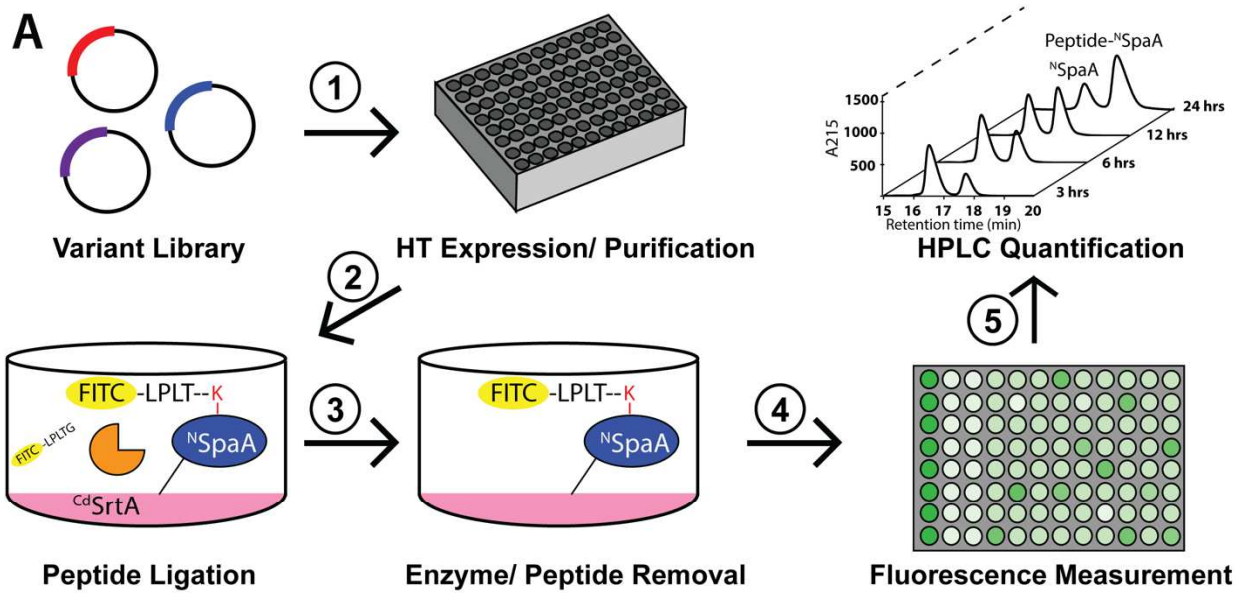
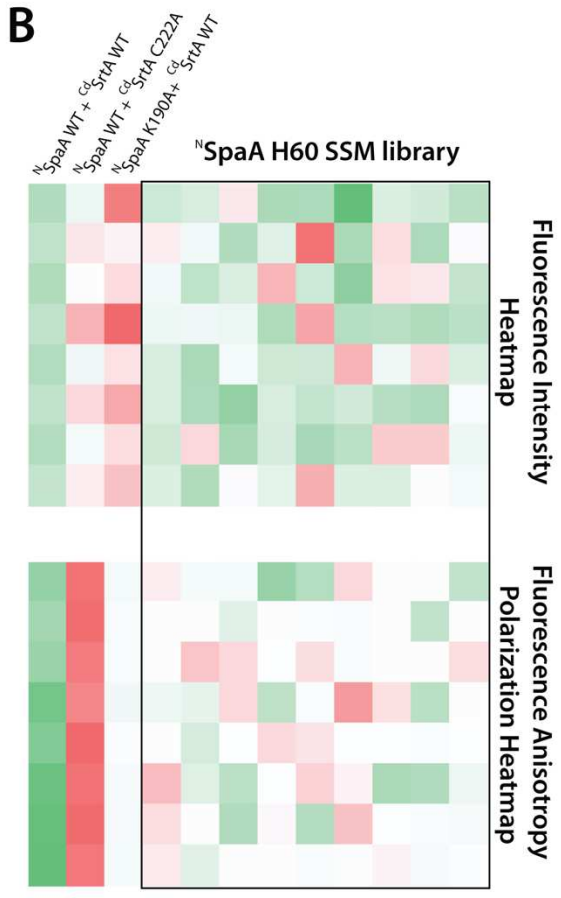
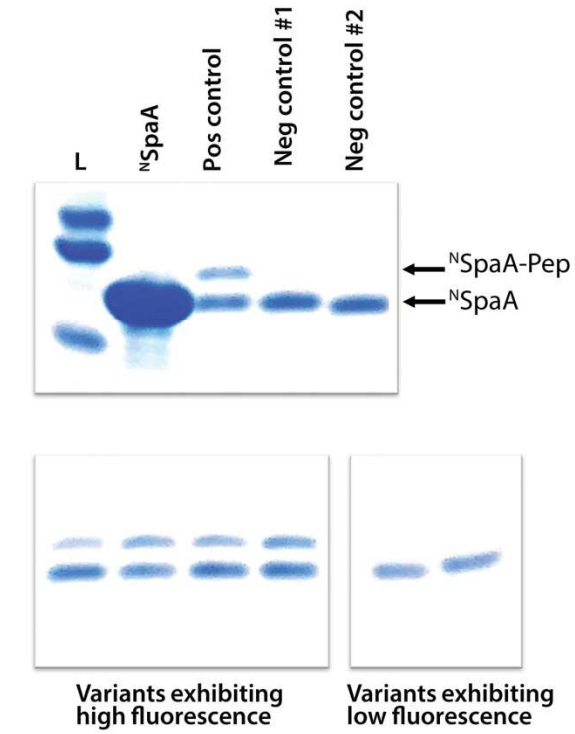

A
SrtA variants
split DHFR
co-transform
LPLTG
K
highly active sortase
inactive sortase
LPLT--K
LPLTG
DHF
THF
DHF
THF
cell viability
cell death
B
C
OD600
0.8
0.6
0.08
0.04
0.00
0
10
20
30
Time (hours)
SrtA+DHFR
DHFR only
Trimethoprim
0.0 µg/mL
1.0 µg/mL
2.5 µg/mL
5.0 µg/mL

Figure 7.3 – Fluorescence based screen. A) Schematic of screening procedure. ^{N}SpaA variants are generated by targeted site-saturation mutagenesis and produced in a 96-well high-throughput format. The variants are then simultaneously purified and retained on the cobalt resin. Exogenous ^{Cd}SrtA and FITC-LPLTG peptide are then supplemented to each well and the reaction is incubated at room temperature. During the incubation period, ^{Cd}SrtA covalently ligates the fluorescent peptide to the K190 sidechain on each ^{N}SpaA variant. The degree of labeling is dependent on the reactivity and kinetics of each variant. The reactions are again purified to remove excess peptide and enzyme and fluorescence intensity is measured with a plate reader. B) The fluorescence intensity in each well correlates with the degree of labeling. The fluorescence measurements for each variant is compared to corresponding wells containing positive controls (wild-type ^{N}SpaA and wild-type sortase) and negative controls (^{N}SpaA K190A or ^{Cd}SrtA C222A). Fluorescence intensity measurements (after purification, top) and fluorescence anisotropy (no purification, bottom) are displayed. C) The results of the fluorescence assay were then verified by orthogonal measurements of transpeptidation (SDS-PAGE and HPLC). SDS-PAGE data is shown for select wells with particularly high or low fluorescence to verify the results of the fluorescence data.


A
Variant Library
1
HT Expression/ Purification
Peptide-NSpaA
NSpaA
24 hrs
12 hrs
6 hrs
3 hrs
1500
1000
500
A215
15 16 17 18 19 20
Retention time (min)
HPLC Quantification
2
FITC -LPLT--K
FITC -LPLTG
NSpaA
CdSrtA
3
FITC -LPLT--K
NSpaA
4
5
Peptide Ligation
Enzyme/ Peptide Removal
Fluorescence Measurement


B
NSpaA WT + CdSrtA WT
NSpaA WT + CdSrtA C222A
NSpaA K190A+ CdSrtA WT
NSpaA H60 SSM library
Fluorescence Intensity Heatmap
Fluorescence Anisotropy Polarization Heatmap


C
L
NSpaA
Pos control
Neg control #1
Neg control #2
NSpaA-Pep
NSpaA
Variants exhibiting high fluorescence
Variants exhibiting low fluorescence

7.6 References

1. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nature Reviews Genetics* **16**, 379–394 (2015).

2. Xiao, H., Bao, Z. & Zhao, H. High throughput screening and selection methods for directed enzyme evolution. *Ind. Eng. Chem. Res.* **54**, 4011–4020 (2015).

3. Wilson, D. S. & Keefe, A. D. Random Mutagenesis by PCR. in *Current Protocols in Molecular Biology* **51**, 8.3.1-8.3.9 (John Wiley & Sons, Inc., 2001).

4. Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *Genome Res.* **2**, 28–33 (1992).

5. Vanhercke, T., Ampe, C., Tirry, L. & Denolf, P. Reducing mutational bias in random protein libraries. *Anal. Biochem.* **339**, 9–14 (2005).

6. Myers, R. M., Lerman, L. S. & Maniatis, T. A general method for saturation mutagenesis of cloned DNA fragments. *Science (80-.).* **229**, 242–247 (1985).

7. Lai, Y. P., Huang, J., Wang, L. F., Li, J. & Wu, Z. R. A new approach to random mutagenesis in vitro. *Biotechnol. Bioeng.* **86**, 622–627 (2004).

8. Kille, S. *et al.* Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2**, 83–92 (2013).

9. Reidhaar-Olson, J. F. & Sauer, R. T. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science (80-.).* **241**, 53–57 (1988).

10. Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391 (1994).

11. Crameri, A., Raillard, S. A., Bermudez, E. & Stemmer, W. P. C. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291 (1998).

12. HJ, W. *et al.* Computationally designed libraries for rapid enzyme stabilization. *Protein Eng. Des. Sel.* **27**, (2014).

13. Lehmann, M., Pasamontes, L., Lassen, S. F. & Wyss, M. The consensus concept for thermostability engineering of proteins. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology* **1543**, 408–415 (2000).

14. Voigt, C. A., Martinez, C., Wang, Z. G., Mayo, S. L. & Arnold, F. H. Protein building blocks preserved by recombination. *Nat. Struct. Biol.* **9**, 553–558 (2002).

15. Chen, L. *et al.* Improved variants of SrtA for site-specific conjugation on antibodies and proteins with high efficiency. *Sci. Rep.* **6**, 1–12 (2016).

16. Zou, Z., Mate, D. M., Rübsam, K., Jakob, F. & Schwaneberg, U. Sortase-mediated high-throughput screening platform for directed enzyme evolution. *ACS Comb. Sci.* **20**, 203–211 (2018).

17. Schatte, M. *et al.* Reporter Immobilization Assay (REIA) for Bioconjugating Reactions. *Bioconjug. Chem.* **27**, 1484–1492 (2016).

18. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming

enzymes using yeast display. *Proc. Natl. Acad. Sci.* **108**, 11399–11404 (2011).

19. Gianella, P., Snapp, E. L. & Levy, M. An in vitro compartmentalization-based method for the selection of bond-forming enzymes from large libraries. *Biotechnol. Bioeng.* **113**, 1647–1657 (2016).

20. Dorr, B. M., Ham, H. O., An, C., Chaikof, E. L. & Liu, D. R. Reprogramming the specificity of sortase enzymes. *Proc. Natl. Acad. Sci.* **111**, 13343–13348 (2014).

21. Piotukh, K. *et al.* Directed evolution of sortase A mutants with altered substrate selectivity profiles. *J. Am. Chem. Soc.* **133**, 17536–17539 (2011).

22. Park, H. Y. *et al.* Phage display screen for peptides that bind Bcl-2 protein. *J. Biomol. Screen.* **16**, 82–89 (2011).

23. PH, B., JJ, R. & PS, D. Rapid isolation of high-affinity protein binding peptides using bacterial display. *Protein Eng. Des. Sel.* **17**, (2004).

24. Liu, W., Hong, J., Bevan, D. R. & Zhang, Y. H. P. Fast identification of thermostable beta-glucosidase mutants on cellobiose by a novel combinatorial selection/screening approach. *Biotechnol. Bioeng.* **103**, 1087–1094 (2009).

25. Firestine, S. M., Salinas, F., Nixon, A. E., Baker, S. J. & Benkovic, S. J. Using an AraC-based three-hybrid system to detect biocatalysts in vivo. *Nat. Biotechnol.* **18**, 544–547 (2000).

26. Michnick, S. W., Ear, P. H., Landry, C., Malleshaiah, M. K. & Messier, V. Protein-Fragment Complementation Assays for Large-Scale Analysis, Functional Dissection and Dynamic Studies of Protein–Protein Interactions in Living Cells. in 395–425 (Humana Press, Totowa, NJ, 2011). doi:10.1007/978-1-61779-160-4_25

27. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology* **10**, 866–876 (2009).

28. Arnold, F. H. Directed evolution: Creating biocatalysts for the future. *Chem. Eng. Sci.* **51**, 5091–5102 (1996).

29. Bellucci, J. J., Bhattacharyya, J. & Chilkoti, A. A noncanonical function of sortase enables site-specific conjugation of small molecules to lysine residues in proteins. *Angew. Chemie - Int. Ed.* **54**, 441–445 (2015).

30. Bentley, M. L., Gaweska, H., Kielec, J. M. & McCafferty, D. G. Engineering the substrate specificity of Staphylococcus aureus sortase A: The β6/β7 loop from SrtB confers npqtn recognition to SrtA. *J. Biol. Chem.* **282**, 6571–6581 (2007).

31. Möhlmann, S., Mahlert, C., Greven, S., Scholz, P. & Harrenga, A. In vitro Sortagging of an Antibody Fab Fragment: Overcoming Unproductive Reactions of Sortase with Water and Lysine Side Chains. *ChemBioChem* **12**, 1774–1780 (2011).

32. Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).

33. Chan, A. H. *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).

34. Suree, N. *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).

35. Jacobitz, A. W. *et al.* 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**, (2016).

36. Moritsugu, K., Terada, T. & Kidera, A. Disorder-to-order transition of an intrinsically disordered region of sortase revealed by multiscale enhanced sampling. *J. Am. Chem. Soc.* **134**, 7094–7101 (2012).

37. Suliman, M. *et al.* Directed evolution provides insight into conformational substrate sampling by SrtA. *PLoS One* **12**, (2017).

38. 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**, E5477–E5486 (2018).

39. 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).

40. McConnell, S. A. S. A. *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).

41. Remy, I., Campbell-Valois, F. X. & Michnick, S. W. Detection of protein–protein interactions using a simple survival protein-fragment complementation assay based on the enzyme dihydrofolate reductase. *Nat. Protoc.* **2**, 2120–2125 (2007).

42. Leung, D. W., Chen, E. & Goeddel, D. V. A Method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1**, 11–15 (1989).

43. Miyazaki, K. MEGAWHOP cloning: A method of creating random mutagenesis libraries via megaprimer PCR of whole plasmids. in *Methods in Enzymology* **498**, 399–406 (Academic Press Inc., 2011).

44. Cody, V., Luft, J. R. & Pangborn, W. Understanding the role of Leu22 variants in methotrexate resistance: Comparison of wild-type and Leu22Arg variant mouse and human dihydrofolate reductase ternary crystal complexes with methotrexate and NADPH. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61**, 147–155 (2005).