UCLA
UCLA Electronic Theses and Dissertations

Title

Elucidating the Molecular Basis of Protein and Polymer Display in Gram-Positive Bacteria for Novel Antibiotic Development

Permalink

https://escholarship.org/uc/item/3w30p3hn

Author

Kattke, Michele

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Elucidating the Molecular Basis of Protein and Polymer Display

in Gram-Positive Bacteria for Novel Antibiotic Development

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Molecular Biology

by

Michele Diedre Kattke

2017

ABSTRACT OF THE DISSERTATION

Elucidating the Molecular Basis of Protein and Polymer Display

in Gram-Positive Bacteria for Novel Antibiotic Development

by

Michele Diedre Kattke

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2017

Professor Robert Thompson Clubb, Chair

The emergence of multi-drug resistant bacteria has prompted novel antibiotic development by targeting non-essential pathways, such as virulence factor production and display during cell wall biosynthesis. Within Gram-positive bacteria, sortase transpeptidases covalently attach proteins to the cell wall or assemble pili using class A-F enzymes. Interestingly, class E sortases display proteins via recognition of a non-canonical LAXTG motif. We have determined the first crystal structure of a class E sortase, the 1.93 Å resolution structure of SrtE1 from *Streptomyces coelicolor*. The SrtE1 enzyme possesses structurally distinct β3/β4 and β6/β7 active site loops that contact the LAXTG substrate. Furthermore, molecular dynamics studies have identified a conserved tyrosine residue that likely confers substrate specificity for class E sortases. A second anti-virulence target, the TarA glycosyltransferase (GT), is highly conserved among Gram-positive bacteria and produces surface-anchored wall teichoic acid (WTA) polymers. The WTA biosynthetic mechanism involving TarA and other membrane-associated, enzymes is poorly understood due to a lack of structural characterization. We have

determined the 2.0 Å resolution crystal structure of the TarA enzyme from *Thermobacter italicus*, which adopts a structurally novel protein fold, termed class GT-E, and represents the first structurally characterized member of the WecB-TagA-CpsF GT family. Sequence conservation mapping onto experimentally observed TarA oligomer structures has identified putative functional residues and suggests formation of a competent active site through oligomerization, which will guide studies of substrate binding and catalysis. Furthermore, we describe two target-specific, cell-based assays for the discovery of sortase and TarA inhibitors. The first assay monitors sortase-dependent growth inhibition of wild-type and sortase-deficient *Actinomyces oris* strains in the presence of small molecule inhibitors. A pilot screen of 1280 compounds returned a hit rate of 0.3%, which has prompted large-scale high-throughput screening. The second assay utilizes a TarA-dependent morphological transition of a mutant *B. subtilis* strain complemented with the TarA enzyme from *S. aureus* (TarA+) to replace the endogenous enzyme activity. The drastic rod-shape to spherical morphological transition provides a robust HTS platform with a Z-prime score of 0.76. Ultimately, these structural and cell-based studies will promote the development of anti-virulence inhibitors to combat bacterial resistance.

The dissertation of Michele Diedre Kattke is approved.

Jeffrey Floyd Miller

Stephen Lawrence Zipursky

Robert Thompson Clubb, Committee Chair

University of California, Los Angeles

2017

DEDICATION

To my family, for inspiring me to do great things in life and make the world a better place.

To my grandfather, Jerry, for guiding me with his wisdom and his example.

To my mother, Sharon, for guiding me with her volition and positivity.

To my father, John, for inspiring me with his creativity and tenacity.

To my sister, Melissa, for supporting me with her love and companionship.

TABLE OF CONTENTS

Chapter 3. Crystal Structure of the *Streptomyces coelicolor* Sortase E1 Transpeptidase

Provides Insight into the Binding Mode of the Novel Class E Sorting Signal………………95

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

VITA

2005 – 2009   Bachelors of Science, Biochemistry with Distinction

University of Illinois at Urbana-Champaign

2008 – 2009   Undergraduate Researcher, Kranz Laboratory

University of Illinois at Urbana-Champaign

2009 – 2011   Post-Baccalaureate ORISE Research Fellow

U.S. Army Construction Engineering Research Laboratory

2011 – 2017   Doctoral Researcher, Clubb Laboratory

University of California, Los Angeles

2016   gRED Intern, Department of Protein Chemistry and Structural Biology

Genentech, Inc.

AWARDS AND PUBLICATIONS

2011   UCLA Graduate Division Eugene V. Cota-Robles Fellowship

2015   Molecular Biology Institute Paul D. Boyer Outstanding Teaching Award

2015   Philip Whitcome Pre-Doctoral Fellowship in Molecular Biology

2016   American Society for Biochemistry and Molecular Biology Travel Awardee

            "Elucidating the molecular basis of protein and polymer display in Gram-positive

            bacteria for novel antibiotic development"

2017   American Society for Biochemistry and Molecular Biology Travel Awardee

            "Studies of TarA, the Novel Wall Teichoic Acid Glycosyltransferase, for the

            Discovery of Gram-positive Bacterial Inhibitors"

Kattke M.D., Chan A.H., Duong A., Sexton D.L., Sawaya M.R., Cascio D., Elliot M.A., and

Clubb R.T. Crystal Structure of the *Streptomyces coelicolor* Sortase E1 Transpeptidase Provides

Insight into the Binding Mode of the Novel Class E Sorting Signal. *PLoS ONE* (2016).

Jacobitz, A.W., **Kattke, M.D.**, Wereszczynski, J., and Clubb, R.T. Sortase Transpeptidases:

Structural Biology and Catalytic Mechanism. *Adv. Protein Chem. Struct. Biol.* (2017). In press.

Chapter 1

Introduction to the Gram-positive Bacterial Cell Wall Architecture, Assembly, and Inhibition by Antibiotics

1.1   Overview

As multi-drug resistant bacteria rapidly emerge, the need for novel antibiotic treatments has become exceedingly urgent. Traditional antibiotics were highly effective for the treatment of bacterial infections as they inhibit essential processes, including cell wall biosynthesis, DNA/RNA synthesis, and protein synthesis. However, bacteria are developing resistance to circumvent these mechanisms of action, which has prompted novel strategies for antibiotic development, such as preventing virulence factor display on the bacterial cell surface. My thesis work focuses on structural characterization of protein targets that produce and display protein and polymer virulence factors during cell wall biosynthesis in Gram-positive bacteria, as well as cell-based, high-throughput screen assay development to facilitate novel inhibitor discovery. This chapter provides a broad overview of the structure and assembly of the Gram-positive bacterial cell wall, followed by detailed descriptions of the covalent surface display of proteins by sortase transpeptidases and wall teichoic acid biosynthesis. Finally, an overview of antibiotic discovery, current classes of antibiotics and their targets, antibiotic resistance concerns, and current antibiotic development trends is provided.

1.2   Structure and assembly of the Gram-positive bacterial cell wall

Gram-positive bacteria are monoderm organisms that have cell walls consisting of a single, thick peptidoglycan layer. The cell wall is a critical cellular structure that maintains proper cell morphology, provides mechanical stability, and serves as a barrier from the surrounding environment [1]. The Gram-positive cell wall also acts as a scaffold to display a variety of proteins and polymers on the cell surface, including capsular polysaccharides, wall teichoic acids (WTA), lipoteichoic acids (LTA), surface proteins, and pili. These surface-exposed macromolecules allow Gram-positive bacteria to productively interact with their

environment, promoting cell adhesion, nutrient acquisition, and host immune system evasion, among other functions.

In *Staphylococcus aureus*, assembly of the peptidoglycan precursor can be divided into three stages: 1) synthesis of Park's nucleotide in the cytoplasm, 2) generation of lipid II in the membrane, and 3) assembly of the muropeptide meshwork [2–4]. First, Park's nucleotide (UDP-MurNAc-L-Ala-D-iGlu-L-Lys-D-Ala-D-Ala) is synthesized in the cytoplasm through the sequential action of the MurABCDEF synthases. Park's nucleotide is then attached to the membrane-embedded carrier molecule, undecaprenyl-pyrophosphate (C55-PP), by MraY to form lipid I. The glycan-linked lipid is modified by MurG with an N-acetylglucosamine (GlcNAc) moiety to produce lipid II (C55-PP-MurNAc-GlcNAc-pentapeptide). The pentapeptide stem within lipid II is extended at position three with a penta-glycyl peptide branch by the peptidyl transferases, FemA, FemB and FemX [5,6]. The final lipid II-pentaglycyl product is flipped to the outer leaflet of the cell membrane by the MurJ flippase for peptidoglycan polymerization.

The peptidoglycan meshwork is assembled largely by penicillin-binding proteins (PBPs) at the extracellular face of the cell membrane. PBPs synthesize long glycan chains by utilizing their glycosyltransferase activity to covalently link GlcNAc and N-acetyl-muramic acid (MurNAc) moieties via $\beta(1\rightarrow4)$ glycosidic bonds [7]. The glycan strands are then crosslinked via their pentapeptide stems by PBP transpeptidation activity. Although the chemical composition of the pentapeptide stem varies among bacterial species, the terminal D-Ala-D-Ala motif is highly conserved, as it is recognized by PBPs. Additionally, some bacteria complement or replace bi-functional PBPs with shape, elongation, division and sporulation (SEDS) enzymes, which are glycosyltransferases that interact with mono-functional PBPs containing transpeptidase activity [8]. The resulting peptidoglycan layer is extensive in Gram-positive bacteria, spanning ~10-40

nm in contrast to ~5-10 nm in Gram-negative species, and acts as a molecular scaffold for anionic polymers (i.e. teichoic acids and capsular polysaccharides) and surface proteins (i.e. S-layer proteins and sortase-anchored proteins) [9].

1.2.1 Production and surface display of polymers in Gram-positive bacteria

Gram-positive bacteria decorate their peptidoglycan layer with diverse, anionic glyco-polymers. Teichoic acid is a highly abundant cell wall component that is unique to Gram-positive species. Two variations of teichoic acid exist: 1) wall teichoic acid (WTA), which is covalently attached to the peptidoglycan, and 2) lipoteichoic acid (LTA), which is anchored to the cell membrane. WTA represents the major species of cell wall polymer, accounting for >50% of the total cell wall content [10]. WTA contains a conserved disaccharide linkage unit that is extended with phosphodiester-linked polyol repeats of ribitol-phosphate (RboP) or glycerol-phosphate (GroP); WTA polymer biosynthesis and export by TarOABDFGH in *S. aureus* is described in detail below. WTA polymers are displayed by LCP ligases, which catalyze a phosphodiester bond between the conserved linkage unit and the C6 hydroxyl of MurNAc within the peptidoglycan [10]. WTA polymers are implicated in a range of bacterial functions, including host cell attachment, cation homeostasis, cell morphology maintenance, and antibiotic resistance, among others [10].

Conversely, LTAs are chemically diverse, glycosyl-phosphate polymers that are tethered to the cell membrane via diacylglycerol (DAG) lipid. LTAs are classified into different types based on their chemical complexity, ranging from unbranched, poly-GroP Type I polymers to complex structures of Type II-IV LTA [11]. During synthesis of the best-characterized Type I LTA in *S. aureus*, the YpfP glycosyltransferase first synthesizes a disaccharide glycolipid (Glc2-

DAG) anchor in the cytoplasm [12,13]. The glycolipid anchor is then transported to the extracellular face of the membrane, presumably by the LtaA flippase, where it is extended with GroP repeats generated from the head group of phosphatidylglycerol by LTA synthase (LtaS) [14]. The LTA main chain can acquire varying D-alanyl or glycosyl modifications at the C2 hydroxyl of the GroP repeats [11]. LTA polymers play important roles in bacterial physiology, specifically membrane homeostasis and virulence.

In most species, additional capsular polysaccharide polymers are covalently attached to the cell wall to create an outer capsule. Capsular polysaccharides are chemically diverse, containing multiple different sugars, glycosidic linkages and chain branching, and are synthesized through either Wzy-dependent or synthase-dependent mechanisms (as described for *Streptococcus pneumoniae* below) [15]. Wzy-dependent synthesis mechanistically resembles peptidoglycan biosynthesis. First, a sugar-phosphate moiety is transferred from a UDP-activated glycan to the undecaprenyl-phosphate (C55-P) lipid carrier at the cytoplasmic face of the membrane by a serotype-dependent initiating glycosyltransferase (CpsE for glucose-utilizing serotypes vs. WciI, WcjG, or WcjH in serotypes that lack glucose) [15,16]. Additional glycosyltransferases (CpsT, CpsF, CpsG, and CspI) add subsequent sugars to the lipid-glycan precursor to complete synthesis of the repeat unit, which is then transported to the extracellular face of the membrane by the Wzx flippase [15]. The capsular polysaccharide chain is polymerized in a non-processive manner by the Wzy polymerase, followed by covalent anchoring to GlcNAc in the peptidoglycan or an alternate membrane acceptor by an unknown enzyme. In synthase-dependent synthesis, a single enzyme catalyzes initiation, polymerization, and transport of the polymer [15]. Upon addition of eight sugars, the lipid-linked oligosaccharide becomes tightly associated with the synthase via a carbohydrate binding site to facilitate highly

processive synthesis of the polymer chain and transport across the cytoplasmic membrane [15]. Other Gram-positive bacteria use similar pathways to construct structurally diverse synthase-dependent and Wzy-dependent capsular polysaccharides to mask the cell surface and prevent complement-mediated opsonophagocytosis.

1.2.2 Surface display of proteins in Gram-positive bacteria

Proteins are secreted to the extracellular surface of the bacterial cell via the twin-arginine translocation (TAT) system or the Sec secretion system. The TAT system exports folded proteins (~25-70 Å in diameter) across the cell membrane by recognizing an N-terminal signal peptide containing a pair of adjacent arginine residues, termed "twin arginines" [17,18] A multi-subunit complex (TatBC) located at the cell membrane binds the signal peptide, which triggers recruitment and oligomerization of TatA protomers to form an active translocation site. The folded protein substrate is actively transported using proton motive force, released from the translocation site, and cleaved into its mature form by a signal peptidase. Conversely, unfolded precursor proteins (~12 Å in diameter) with an N-terminal signal peptide are secreted through the Sec translocon [19]. The Sec translocon consists of a membrane-embedded translocation pore (SecYEG) that recognizes the N-terminal signal peptide, as well as an ATPase (SecA) that pushes substrates through the hydrophilic channel [20,21]. Protein substrates are maintained in a secretion-competent state via either Sec- and signal recognition particle (SRP)-dependent mechanisms [20]. In the SRP-dependent mechanism, SRP binds the signal peptide, which temporarily arrests translation by the ribosome. The SRP-ribosome complex docks onto a membrane receptor (FtsY), upon which the nascent polypeptide is transferred to the Sec translocon for co-translational secretion [22]. In the Sec-dependent pathway, post-translational

secretion of fully synthesized proteins is mediated by secretion chaperone binding in the cytoplasm.

Surface proteins, including S-layer proteins (SLPs) and sortase-attached proteins, are anchored to the Gram-positive cell wall via non-covalent or covalent mechanisms. SLPs are highly abundant glycoproteins that self-associate into two-dimensional (2D) crystalline arrays that coat the bacterial cell, called S-layers [23]. SLPs non-covalently associate with cell wall polysaccharides, such as secondary cell wall polysaccharide in *Bacillus anthracis*, through surface layer homology (SLH) or cell wall binding 2 (CWB2) anchoring domains [23]. An additional crystallization domain facilitates self-assembly of SLPs into a continuous 2D array. SLPs can be O-glycosylated with glycan chains of diverse chemical composition. Construction of SLP glycan chains resembles WTA polymer production, where glycan moieties are transferred from nucleotide-activated sugars to a lipid carrier at the cytoplasmic membrane to form a linkage unit, which is appended with glycan repeats to produce the main chain [23]. The glycan chain is then transported to the extracellular surface by an ABC-transporter and appended to SLPs by a ligase at the cell surface. SLPs and associated proteins perform many functions, ranging from adhesion activity, binding defined ligands, maintenance of cell envelope integrity, and construction of a permeability barrier, among others [23].

Surface protein precursors contain an N-terminal signal peptide and a C-terminal cell wall sorting signal (CWSS), which consists of a semi-conserved LPXTG motif (where X is any amino acid), followed by a hydrophobic stretch and cluster of positively charged residues. The CWSS of surface proteins is recognized by sortase transpeptidase, which cleaves the pentapeptide motif between the threonine and glycine residues and transfers the cleaved protein to the amino group of the pentaglycl branch within lipid II (described in detail below) [19]. The

7

lipid II-linked protein product is then incorporated into the peptidoglycan via PBP transglycosylation and transpeptidation activity. Several classes of sortase anchor collections of protein substrates that contain distinct sorting signal motifs; these substrates promoter a variety of functions, including heme acquisition (class B), cell adhesion (class A), aerial hyphae formation (class E), and sporulation (class D), among others [24]. Additionally, sortases (class C) can also polymerize pili polymers, a process that is unique to Gram-positive bacteria [24].

1.3   Covalent surface protein display via sortase transpeptidases

Proteins displayed on the surface of bacterial pathogens play critical roles in the infection process by promoting bacterial adhesion to host tissues, acquisition of essential nutrients, evasion and suppression of the immune response and host cell entry [16, 17]. *S. aureus* and other Gram-positive bacteria display virulence factors using sortase cysteine transpeptidase enzymes (Figure 1.9.1) [13-15, 18-22]. Sortases reside on the extracellular membrane where they covalently attach proteins to peptidoglycan by catalyzing a transpeptidation that joins a C-terminal cell wall sorting signal (CWSS) within their protein substrate to the cross-bridge peptide. This process is best understood for the *S. aureus* sortase A protein and begins when a full-length, precursor protein containing an amino terminal leader peptide is exported from the cytoplasm through the secretory (Sec) pathway (Figure 1.9.1). The CWSS in sortase A substrates consists of a LPXTG-motif (where X denotes any amino acid), followed by a segment of hydrophobic amino acids and a tail composed primarily of positively charged residues. The C-terminal charged tail presumably prevents export, positioning the protein for processing by the extracellular membrane associated sortase enzyme. Catalysis occurs through a ping-pong mechanism that is initiated when the active site cysteine residue nucleophilically attacks the backbone carbonyl carbon of the threonine

residue within the LPXTG-motif, breaking the threonine-glycine peptide bond to create a sortase-protein complex in which the components are linked via a thioacyl bond [23, 24]. The protein is then transferred by sortase to the cell wall precursor lipid II, when the amino group in this molecule nucleophilically attacks the thioacyl linkage to create a peptide bond-linked, protein-lipid II product. Transglycosylation and transpeptidation reactions that synthesize the cell wall then incorporate this product into the peptidoglycan.

Alternatively, a second type of sortase, frequently called 'pilin polymerases,' construct surface pili by polymerizing pilin protein subunits. These pilin-assembling enzymes employ a similar transpeptidation reaction as sortase A, but instead of using lipid II as a nucleophile to attach proteins to the cell wall, a lysine amino group located within a protein pilin subunit is used as a secondary substrate to attack the sortase-protein thioacyl intermediate [25–27]. Bacteria often utilize multiple sortase enzymes with unique substrate specificities that enable microbes to non-redundantly display or assemble distinct proteins on the cell surface. Sortase-mediated surface protein display is reviewed in detail in Chapter 2.

1.4 The wall teichoic acid biosynthetic pathway

WTA is the major species of anionic glyco-polymer displayed on the peptidoglycan in Gram-positive bacteria, accounting for >50% of the cell wall content. WTA promotes a range of functions within cell physiology, including cation homeostasis, maintenance of cell morphology, resistance to antimicrobial peptides, biofilm formation, cell adhesion and pathogenicity [10,28]. In fact, WTA is thought to directly promote pathogenesis of Gram-positive organisms, as disrupting WTA production renders *S. aureus* avirulent and re-sensitizes methicillin-resistant *S. aureus* (MRSA) strains to β-lactam antibiotics [29,30]. Furthermore, *S. aureus* strains that are

devoid of WTA show significant defects in cell adhesion and virulence, indicating their promise as an antibiotic target [29].

The WTA pathway is best characterized in *B. subtilis 168*, *B. subtilis W23*, and *S. aureus*, which produce GroP polymers with Tag enzymes and RboP polymers with Tar enzymes. Synthesis of the WTA polymer occurs through sequential action of several enzymes at the cytoplasmic face of the cell membrane [31–33]. In *S. aureus*, a GlcNAc-ManNAc-GroP linkage unit is first constructed through the activity of TarO, TarA, and TarB, a process that is highly conserved across all Gram-positive bacteria. Specifically, the TarO glycosyltransferase catalyzes transfer of GlcNAc from the UDP-activated sugar to the C55-PP carrier, lipid α, at the cytoplasmic membrane [34]. Next, the TarA glycosyltransferase transfers ManNAc from UDP-ManNAc to C55-PP-GlcNAc, or lipid β, to form a disaccharide-lipid product [35]. The TarD cytidylyltransferase catalyzes transfer of L-alpha-Gro-3-P to CTP to produce CDP-Gro, which is utilized as a substrate by TarB to transfer a single GroP to lipid β to produce C55-PP-GlcNAc-ManNAc-GroP [36,37].

The conserved linkage unit is subsequently utilized as a scaffold for the polymerization of GroP or RboP repeats and additional modifications in a process that diverges among bacterial strains [29,34,37]. In *S. aureus*, TarF performs primase activity to transfer a single GroP unit to the TarB product. Next, the combined activity of the TarI cytidylyltransferase and TarJ alcohol dehydrogenase produce CDP-Rbo, which is utilized as a substrate by the bifunctional primase and polymerase, TarL, to transfer 40-60 RboP units to the growing main chain. An additional bifunctional primase and polymerase, TarK, is present in *S. aureus*, which produces electrophoretically distinct WTA polymer, called K-WTA. The resulting poly-RboP main chain

polymers can be further modified by the TarM glycosyltransferase, which catalyzes addition of alpha-O-GlcNAc, and the TarS glycosyltransferase, which transfers beta-O-GlcNAc [38–41].

In *B. subtilis W23*, TarK acts directly on the TarB product as a primase, making TarF dispensable [32]. TarL then acts as a polymerase to complete the poly-RboP main chain, which is modified with β-Glc units by the TarQ glycosyltransferase [41,42]. In *B. subtilis 168*, TagF utilizes polymerase activity to transfer 45-60 GroP units to the TagB product (homologous to TarB in *S. aureus* and *B. subtilis W23*) and complete the polymer. The TagE glycosyltransferase catalyzes transfer of α-Glc to the main chain polymer using UDP-Glc as a donor substrate [41,43]. Glycosylation of WTA affects its polymeric structure and has been shown to influence susceptibility to antibiotics, likely through affecting interactions with other cell wall components [42].

The completed WTA polymer, regardless of the chemical signature of the main chain, is then transported across the cell membrane by the ABC transporter, TarGH, through recognition of the conserved linkage unit [44]. Once outside the cell, the main chain polymer can be further modified with D-alanine to tune its electrostatic properties [40,42,45,46]. First, DltA activates D-alanine as an AMP ester and transfers the aminoacyl adenylate to DltC. A pantothenate cofactor of DltC forms a thioester with D-alanine, after which both DltB presumably transports D-alanine-charged DltC across the membrane where it can serve as a D-alanylation donor. DltD is thought to provide editing function to remove D-alanine from mischarged acyl carriers through hydrolase activity. Finally, the decorated WTA polymer is transferred to the cell wall by an LCP ligase, which covalently attaches the anomeric carbon of GlcNAc within the linkage unit to the C6 hydroxyl of the MurNAc within the peptidoglycan [47].

1.5 Antibiotic development and the rise of resistance

The "Golden Age of Discovery" in the 1940s to 1960s initiated a revolution in the treatment of bacterial infections [48]. In this era, bacterial and fungal species were screened for production of metabolites with potent antimicrobial properties [49]. The prominent discovery strategy utilized the Waksman platform, in which cultures of soil-dwelling bacteria (i.e. streptomycetes) were systematically overlaid with susceptible test microorganisms and monitored for zones of growth inhibition [49,50]. The whole-cell screening approach relied entirely on phenotypic analysis and could be applied with no knowledge of the molecular target or mode of action [49,50]. Although simple, the discovery platform was highly effective at identifying a large arsenal of natural products (i.e. streptomycin) that possessed potent bacteriostatic or bactericidal effects coupled with minimal off-target side effects [49,50].

After 20 years of discovery with the Waksman platform, success rates for identifying of new scaffolds had diminished, and resistance towards natural products was apparent. The mid-1960s ushered in the "Medicinal Chemistry Era," in which scaffolds of existing natural products were chemically modified to produce new synthetic entities [48]. These chemical derivatives were analyzed for antimicrobial properties using whole cell screens and offered improvements, including lower dosages, expanded spectrum of susceptible microbes, and most importantly, the avoidance of resistance [48]. By the 1990s, several classes of antibiotics had been developed for therapeutic use. These therapeutics inhibited bacteria through common mechanisms of action (described in detail below): 1) cell wall biosynthesis, 2) protein synthesis, 3) RNA or DNA synthesis, 4) folate synthesis, or 5) disruption of the cell membrane [48]. However, due to the essential nature of the targeted molecular pathways, bacteria rapidly counterstriked with

destruction of the antibiotic, overproduction of the molecular target, target modification, and restriction of drug penetration or increased efflux [51–53].

The lapse in drug discovery during the 1960s to 1990s was met by a steady emergence of bacterial resistance, which has prompted innovative antibiotic discovery approaches [53]. In the current "Resistance Era," target-based screening platforms have been applied over traditional, whole-cell approaches [48,49]. Individual molecular targets are identified through high-throughput, bacterial genome-wide essentiality screens. Notably, growth conditions during the essentiality screen (i.e. nutrient levels, carbon source, or chemical perturbation, etc.) can influence the dispensability phenotype of the bacteria, emphasizing the need to mimic the nutrient-poor conditions encountered during host infection [48,54]. The identified essential genes comprise a repository of protein targets that are purified and screened against large, chemical libraries and monitored for *in vitro* inhibition. Further rational design of identified small molecules is facilitated by protein structure determination (i.e. target-ligand co-crystallization, *in silico* docking, etc.). However, two decades of target-based approaches have produced an underwhelming number of viable therapeutics, as their physical and chemical properties are frequently unsuitable (i.e. barrier impermeability, etc.) [53]; this is in contrast to natural product antibiotics, which had been favorably engineered during millions of years of co-evolution with bacterial species. Two notable drug successes using target-based approaches include the oxazolidinone-class compound, linezolid, and the lipopeptide class drug, daptomycin.

Commercially available antibiotics inhibit essential molecular targets; however, these once viable treatment options have been steady losing efficacy due to bacterial resistance mechanism. Cell wall biosynthesis inhibitors, including β-lactams, glycopeptides and β-lactamase inhibitors, modulate the ability of PBPs to crosslink peptidoglycan strands [48]. The

broad spectrum β-lactam class (i.e. penicillin, methicillin, etc.) irreversibly binds to PBPs to prevent the transpeptidation reaction that covalently joins adjacent peptidoglycan strands [49]. Resistance to β-lactams has been acquired through the *mecA* gene, which encodes the protein PBP2a that has low affinity for β-lactams, or by the presence of β-lactamases that hydrolyze the β-lactam ring and deactivate the antibiotic (i.e. AmpC, OXA, KPC, TEM/SHV, expanded spectrum β-lactamases (ESBL), etc.). Subgroups of the β-lactam class include cephalosporins (i.e. cefepime, cefoxitin, etc.) and carbapenems (i.e. imipenem), which are less susceptible to β-lactamases and display stability against ESBLs, respectively; however, the versatile hydrolytic capabilities of carbapenemases are rendering these compounds ineffective. Narrow spectrum glycopeptides antibiotics, such as vancomycin and teicoplanin, bind and sequester the terminal D-Ala-D-Ala dipeptide of the peptidoglycan peptide stem, preventing transpeptidation by PBPs. Modification of glycopeptides by addition of lipophilic side-chains has produced next generation lipoglycopeptide antibiotics (i.e. dalbavancin), which exhibit an extended half-life. Lastly, β-lactamase inhibitor and β-lactam combination therapies offer protection of the therapeutic compound against enzymatic hydrolysis.

Protein synthesis inhibitors prevent protein translation by binding the 50S or 30S subunits of bacterial ribosomes [49]. The 50S ribosome inhibitors either prevent formation of the initiation complex, as in the case of oxazolidinones (i.e. linezolid), or block translocation of peptidyl tRNAs, as demonstrated by macrolides (i.e. erythromycin), streptogramins (i.e. dalfopristin) and phenicols (i.e. chloramphenicol), which triggers release of the incomplete peptide chain. The 30S ribosome inhibitors block access of aminoacyl tRNAs to the ribosome (i.e. tetracycline) or promote mRNA-tRNA mismatching, as with aminoglycoside antibiotics (i.e. streptomycin). Resistance to protein synthesis inhibitors has been observed through several

mechanisms, including reduced membrane permeability, mutation of the 50S or 30S ribosomal subunits, increased drug efflux, and enzyme-catalyzed antibiotic modification.

DNA or RNA synthesis inhibitors prevent either DNA replication by binding DNA gyrase or mRNA transcription by binding RNA polymerase [49]. Specifically, the quinolone class of antibiotics and next generation fluoroquinolones (i.e. ciprofloxacin) bind to DNA gyrase or DNA topoisomerase IV to trap the DNA cleavage stage and prevent strand rejoining. Rifamycins (i.e. rifampicin) binds β-subunit of RNA polymerase at initiation stage to prevent RNA synthesis and downstream protein synthesis [55,56]. Resistance to quinolones, fluoroquinolones, and rifamycins has been achieved through mutations in the molecular target enzymes (DNA gyrase, DNA topoisomerase IV, or β-subunit of RNA polymerase), increased drug efflux through mutation in regulatory genes that control expression levels, or aminoglycoside-modifying enzymes.

Folate synthesis inhibitors and lipopeptide class antibiotics exploit unique mechanisms of action to target bacteria [57]. Folate synthesis inhibitors target a bacterial-specific pathway that is essential for DNA replication and cell division [49]. Inhibition of folate biosynthesis leads to bacteriostasis and eventual results in bacterial cell death. The sulfonamide drug class (i.e. sulfamethoxazole) competitively binds to dihydropteroate synthetase, which prevents addition of para-aminobenzoic acid (PABA) as a constituent into the folic acid molecule. Sulfonamide resistance is mainly achieved by increased production of the PABA substrate. In addition, lipopeptides (i.e. daptomycin) insert into and disrupt the bacterial cell membrane, which causes rapid depolarization and cell death. Resistance to lipopeptides has been achieved by altering the cell wall architecture (i.e. increased D-alanylation of teichoic acids, cell wall thickening) and cell membrane homeostasis.

Antibiotic discovery efforts are shifting towards the development of narrow-spectrum therapeutics and inhibition of *in vivo* essential targets [58]. The widespread use of broad spectrum antibiotics has driven resistance across a range of bacterial species. This selective pressure has produced multi-drug resistant pathogens, most notably, the ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebisiella pneumoniae*, *Acinetobacter baumanii*, *Pseudomonas aeruginosa*, *Enteobacteriaceae*), which have circumvented even our last resort antibiotics. Next generation discovery efforts will utilize unconventional, whole-cell screening approaches, such as target-based, whole-cell assays and cytological profiling, in contrast to traditional bacteriostatic and bactericidal monitoring [54]. Molecular targets will be re-prioritized to focus on *in vivo* essential targets, or genes that are required for growth during an infection, as well as virulence factors that are required for host colonization and pathogenicity. Furthermore, novel treatment strategies, specifically combinatorial approaches, will be powerful towards re-sensitizing bacteria to established antibiotics and slowing the development of antibiotic resistance in the future.

1.6   Scope of dissertation

The rapid evolution of multi-drug resistant bacterial strains has created an urgent need for development of novel therapeutics. In contrast to targeting essential processes, next generation antibiotics are shifting towards the inhibition of bacterial virulence factors. To aid rational design and screening efforts, we have structurally characterized members of two classes of Gram-positive virulence factors: 1) the first class E sortase transpeptidase, and 2) the structurally novel, TarA WTA glycosyltransferase. First, we determined the 1.93 angstrom resolution crystal structure of the soluble, extracytoplasmic domain of SrtE1 from *Streptomyces coelicolor*.

Activity assays of SrtE1 indicated a marked preference to hydrolyze a unique LAETG sorting signal motif over the canonical LPXTG motif. Molecular dynamics simulations of models of stable SrtE1 thioacyl intermediates bound to LAET or LPET peptides revealed stabilizing hydrogen bond contacts between a conserved tyrosine residue within an active site loop and the backbone amine of the alanine residue in the peptide substrate; this stabilizing interaction is disrupted by substrates containing a proline, which sterically clashes with the conserved tyrosine and deforms the substrate binding pocket (Chapter 3). Secondly, we determined the 1.8 Å crystal structure of the extracytoplasmic domain of the TarA glycosyltransferase from *Thermoanaerobacter italicus*. The TarA crystal structure revealed a novel glycosyltransferase fold, one of five that is currently known, which I have termed "GT-E." The enzyme forms dimer and trimer species that contains highly conserved residues at the interface according to Consurf sequence conservation analysis. Functional studies of the identified residues and C-terminal membrane anchor are currently being tested in a *B. subtilis* strain that we have complemented with the *S. aureus* TarA enzyme (Chapter 4).

To facilitate novel inhibitor discovery, we have developed whole-cell high-throughput screens towards two validated, Gram-positive molecular targets, sortase and the TarA WTA glycosyltransferase. The cell-based assays were scaled to 384-well microplates to validate robustness and high-throughput suitability by calculating Z-prime scores. The cell-based sortase assay measured differences in growth between a wild-type and mutant *Actinomyces oris* strain that lacks its house-keeping sortase, which produced a Z-prime score of 0.67 with controls. With a reliable dynamic range, the cell-based sortase assay was implemented in a pilot screen against the New Prestwick chemical library, which contains FDA-approved pharmacologically active compounds, and returned four hit compounds and a hit rate of 0.3%. Secondly, the cell-based

TarA assay measured morphological differences between a *S. aureus* TarA-complemented *B. subtilis* strain (rod-shape) and a mutant *B. subtilis* strain containing an endogenous *tagA* gene deletion (spherical-shape). The drastic morphological shift between the complemented, wild-type-like and mutant *B. subtilis* strains produced a Z-prime score of 0.76, indicating the robustness of this screening platform.

1.7 Figures
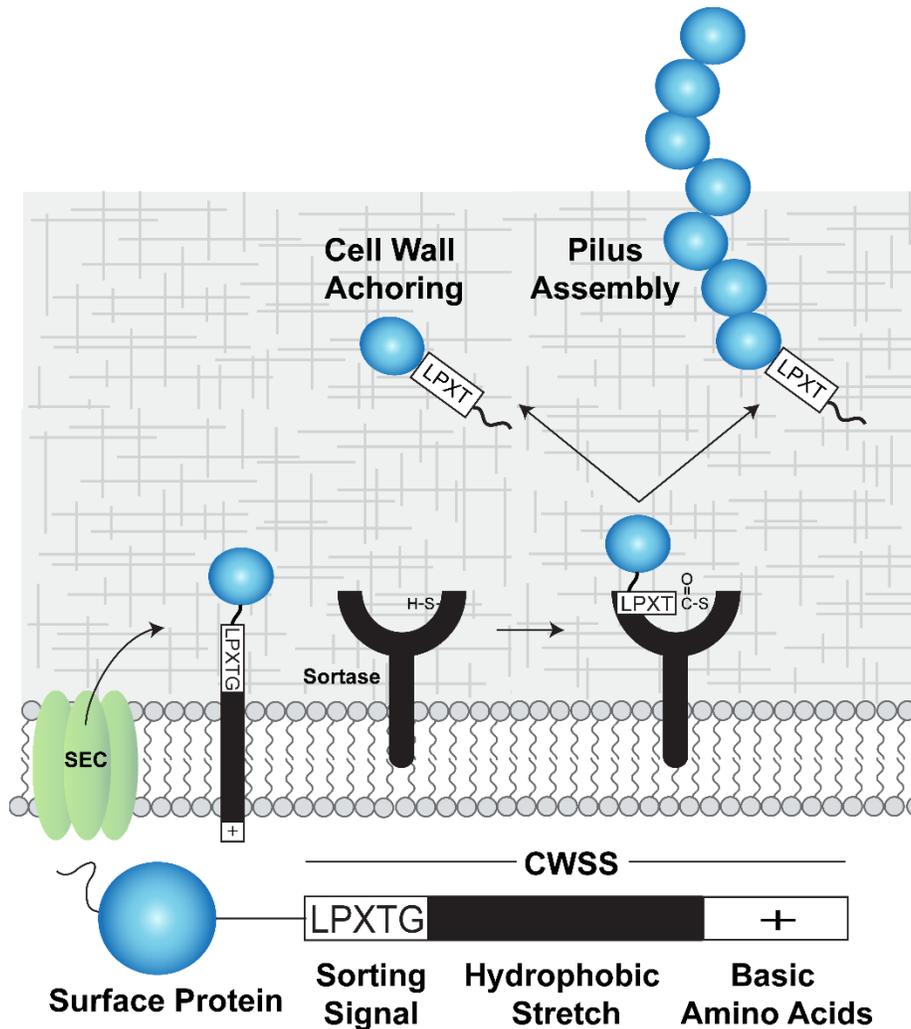
Figure 1.7.1. Sortase-mediated anchoring of surface proteins. An N-terminal signal peptide directs cell wall anchored proteins to the Sec pathway. The sortase substrate is retained in the membrane by the cell wall sorting signal (CWSS). Sortase cleaves the LPXTG sequence within the CWSS, and a thioester-linked acyl-enzyme intermediate is formed. Sortase catalyzes peptide bond formation between the CWSS and the free amine group on lipid II, and the protein-lipid II linked product is incorporated into the cell envelope via transglycosylation reactions that synthesize the peptidoglycan.
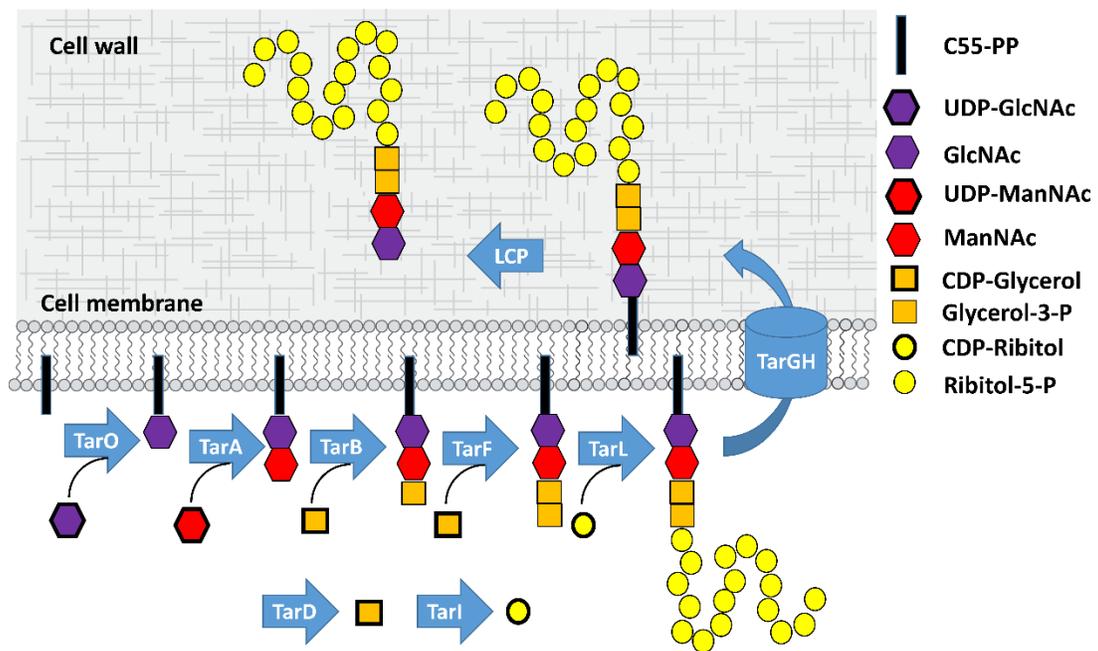
Figure 1.7.2. Biosynthesis of wall teichoic acid glyco-polymers in *S. aureus*.

Concerted action of conserved WTA biosynthetic enzymes, TarO, TarA, and TarB, results in formation of a linkage unit, followed by polymerization of the WTA chain by TagF family members (TarF, TarL) and transportation of the polymer to the exterior of the cell by the ABC transporter, TarGH, where it is covalently attached to the peptidoglycan by the LCP ligase.

1.8 References

1.   Malanovic N, Lohner K. Gram-positive bacterial cell envelopes: The impact on the activity of antimicrobial peptides. Biochim Biophys Acta - Biomembr. 2016;1858: 936–946. doi:10.1016/j.bbamem.2015.11.004

2.   Navarre WW, Schneewind O. Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. Microbiol Mol Biol Rev. 1999;63: 174–229. Available: http://www.ncbi.nlm.nih.gov/pubmed/10066836

3.   Bhavsar AP, Brown ED. Cell wall assembly in Bacillus subtilis: how spirals and spaces challenge paradigms. Mol Microbiol. 2006;60: 1077–1090. doi:10.1111/j.1365-2958.2006.05169.x

4.   Silhavy TJ, Kahne D, Walker S. The bacterial cell envelope. Cold Spring Harb Perspect Biol. 2010;2: a000414. doi:10.1101/cshperspect.a000414

5.   Ton-That H, Labischinski H, Berger-Bächi B, Schneewind O. Anchor structure of staphylococcal surface proteins. III. Role of the FemA, FemB, and FemX factors in anchoring surface proteins to the bacterial cell wall. J Biol Chem. 1998;273: 29143–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/9786923

6.   Rohrer S, Berger-Bächi B. FemABX peptidyl transferases: a link between branched-chain cell wall peptide formation and beta-lactam resistance in gram-positive cocci. Antimicrob Agents Chemother. 2003;47: 837–46. Available: http://www.ncbi.nlm.nih.gov/pubmed/12604510

7.   Sauvage E, Kerff F, Terrak M, Ayala JA, Charlier P. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. FEMS Microbiol Rev. 2008;32: 234–

258. doi:10.1111/j.1574-6976.2008.00105.x

8.   Henrichfreise B, Brunke M, Viollier PH. Bacterial Surfaces: The Wall that SEDS Built.

      Curr Biol. 2016;26: R1158–R1160. doi:10.1016/j.cub.2016.09.028

9.   Turner RD, Vollmer W, Foster SJ. Different walls for rods and balls: the diversity of

      peptidoglycan. Mol Microbiol. 2014;91: 862–74. doi:10.1111/mmi.12513

10.   Brown S, Santa JP, Jr M, Walker S. Wall Teichoic Acids of Gram-Positive Bacteria.

      2013; doi:10.1146/annurev-micro-092412-155620

11.   Percy MG, Gründling A. Lipoteichoic acid synthesis and function in gram-positive

      bacteria. Annu Rev Microbiol. 2014;68: 81–100. doi:10.1146/annurev-micro-091213-

      112949

12.   Kiriukhin MY, Debabov D V, Shinabarger DL, Neuhaus FC. Biosynthesis of the

      glycolipid anchor in lipoteichoic acid of Staphylococcus aureus RN4220: role of YpfP, the

      diglucosyldiacylglycerol synthase. J Bacteriol. 2001;183: 3506–14.

      doi:10.1128/JB.183.11.3506-3514.2001

13.   Grundling A, Schneewind O. Genes Required for Glycolipid Synthesis and Lipoteichoic

      Acid Anchoring in Staphylococcus aureus. J Bacteriol. 2007;189: 2521–2530.

      doi:10.1128/JB.01683-06

14.   Grundling A, Schneewind O. Synthesis of glycerol phosphate lipoteichoic acid in

      Staphylococcus aureus. Proc Natl Acad Sci. 2007;104: 8478–8483.

      doi:10.1073/pnas.0701821104

15.   Yother J. Capsules of *Streptococcus pneumoniae* and Other Bacteria: Paradigms for

      Polysaccharide Biosynthesis and Regulation. Annu Rev Microbiol. 2011;65: 563–581.

      doi:10.1146/annurev.micro.62.081307.162944

16. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, et al. Genetic Analysis of the Capsular Biosynthetic Locus from All 90 Pneumococcal Serotypes. PLoS Genet. 2006;2: e31. doi:10.1371/journal.pgen.0020031

17. Goosens VJ, Monteferrante CG, van Dijl JM. The Tat system of Gram-positive bacteria. Biochim Biophys Acta - Mol Cell Res. 2014;1843: 1698–1706. doi:10.1016/j.bbamcr.2013.10.008

18. Goosens VJ, van Dijl JM. Twin-Arginine Protein Translocation. Current topics in microbiology and immunology. 2016. pp. 69–94. doi:10.1007/82_2016_7

19. Schneewind O, Missiakas D. Sec-secretion and sortase-mediated anchoring of proteins in Gram-positive bacteria. Biochim Biophys Acta - Mol Cell Res. 2014;1843: 1687–1697. doi:10.1016/j.bbamcr.2013.11.009

20. Denks K, Vogt A, Sachelaru I, Petriman N-A, Kudva R, Koch H-G. The Sec translocon mediated protein transport in prokaryotes and eukaryotes. Mol Membr Biol. 2014;31: 58–84. doi:10.3109/09687688.2014.907455

21. Veenendaal AKJ, van der Does C, Driessen AJM. The protein-conducting channel SecYEG. Biochim Biophys Acta. 2004;1694: 81–95. doi:10.1016/j.bbamcr.2004.02.009

22. Draycheva A, Bornemann T, Ryazanov S, Lakomek N-A, Wintermeyer W. The bacterial SRP receptor, FtsY, is activated on binding to the translocon. Mol Microbiol. 2016;102: 152–167. doi:10.1111/mmi.13452

23. Fagan RP, Fairweather NF. Biogenesis and functions of bacterial S-layers. Nat Rev Microbiol. 2014;12: 211–222. doi:10.1038/nrmicro3213

24. Spirig T, Weiner EM, Clubb RT. Sortase enzymes in Gram-positive bacteria. Mol Microbiol. 2011;82: 1044–1059. doi:10.1111/j.1365-2958.2011.07887.x

25. Spirig T, Weiner EM, Clubb RT. Sortase enzymes in Gram-positive bacteria. Mol Microbiol. 2011;82: 1044–59. doi:10.1111/j.1365-2958.2011.07887.x

26. Ton-That H, Schneewind O. Assembly of pili in Gram-positive bacteria. Trends Microbiol. 2004;12: 228–234. doi:10.1016/j.tim.2004.03.004

27. Gaspar AH, Ton-That H. Assembly of Distinct Pilus Structures on the Surface of Corynebacterium diphtheriae. J Bacteriol. 2006;188: 1526–1533. doi:10.1128/JB.188.4.1526-1533.2006

28. Sewell EW, Brown ED. Taking aim at wall teichoic acid synthesis: new biology and new leads for antibiotics. J Antibiot (Tokyo). Nature Publishing Group; 2013; 1–9. doi:10.1038/ja.2013.100

29. D'Elia MA, Henderson JA, Beveridge TJ, Heinrichs DE, Brown ED. The N-acetylmannosamine transferase catalyzes the first committed step of teichoic acid assembly in Bacillus subtilis and Staphylococcus aureus. J Bacteriol. American Society for Microbiology; 2009;191: 4030–4. doi:10.1128/JB.00611-08

30. Farha MA, Leung A, Sewell EW, D'Elia MA, Allison SE, Ejim L, et al. Inhibition of WTA Synthesis Blocks the Cooperative Action of PBPs and Sensitizes MRSA to β-Lactams. ACS Chem Biol. 2013;8: 226–233. doi:10.1021/cb300413m

31. Lazarevic V, Mauël C, Möller SB, Abellan F-X, Karamata D. Comparison of ribitol and glycerol teichoic acid genes in Bacillus subtilis W23 and 168: identical function, similar divergent organization, but different regulation. Microbiology. Microbiology Society; 2002;148: 815–824. doi:10.1099/00221287-148-3-815

32. Brown S, Meredith T, Swoboda J, Walker S. Staphylococcus aureus and Bacillus subtilis W23 Make Polyribitol Wall Teichoic Acids Using Different Enzymatic Pathways. Chem

Biol. 2010;17: 1101–1110. doi:10.1016/j.chembiol.2010.07.017

33. Formstone A, Carballido-López R, Noirot P, Errington J, Scheffers D-J. Localization and Interactions of Teichoic Acid Synthetic Enzymes in Bacillus subtilis. J Bacteriol. 2008;190: 1812–1821. doi:10.1128/JB.01394-07

34. Soldo B, Lazarevic V, Karamata D. tagO is involved in the synthesis of all anionic cell-wall polymers in Bacillus subtilis 168. Microbiology. 2016;2953: 54–2079.

35. Yu-Hui Zhang ‡, Cynthia Ginsberg ‡,§, Yanqiu Yuan ‡,§ and, Suzanne Walker* ‡. Acceptor Substrate Selectivity and Kinetic Mechanism of Bacillus subtilis TagA†. American Chemical Society ; 2006; doi:10.1021/BI060872Z

36. Fong DH, Yim VC-N, D'Elia MA, Brown ED, Berghuis AM. Crystal structure of CTP:glycerol-3-phosphate cytidylyltransferase from Staphylococcus aureus: Examination of structural basis for kinetic mechanism. Biochim Biophys Acta - Proteins Proteomics. 2006;1764: 63–69. doi:10.1016/j.bbapap.2005.10.015

37. Bhavsar AP, Truant R, Brown ED. The TagB protein in Bacillus subtilis 168 is an intracellular peripheral membrane protein that can incorporate glycerol phosphate onto a membrane-bound acceptor in vitro. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2005;280: 36691–700. doi:10.1074/jbc.M507154200

38. Koç C, Gerlach D, Beck S, Peschel A, Xia G, Stehle T. Structural and Enzymatic Analysis of TarM Glycosyltransferase from *Staphylococcus aureus* Reveals an Oligomeric Protein Specific for the Glycosylation of Wall Teichoic Acid. J Biol Chem. 2015;290: 9874–9885. doi:10.1074/jbc.M114.619924

39. Sobhanifar S, Worrall LJ, King DT, Wasney GA, Baumann L, Gale RT, et al. Structure and Mechanism of Staphylococcus aureus TarS, the Wall Teichoic Acid β-

glycosyltransferase Involved in Methicillin Resistance. Zhang G, editor. PLOS Pathog. 2016;12: e1006067. doi:10.1371/journal.ppat.1006067

40. Xia G, Maier L, Sanchez-Carballo P, Li M, Otto M, Holst O, et al. Glycosylation of wall teichoic acid in Staphylococcus aureus by TarM. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2010;285: 13405–15. doi:10.1074/jbc.M109.096172

41. Winstel V, Xia G, Peschel A. Pathways and roles of wall teichoic acid glycosylation in Staphylococcus aureus. Int J Med Microbiol. 2014;304: 215–221. doi:10.1016/j.ijmm.2013.10.009

42. Brown S, Xia G, Luhachack LG, Campbell J, Meredith TC, Chen C, et al. Methicillin resistance in Staphylococcus aureus requires glycosylated wall teichoic acids. Proc Natl Acad Sci U S A. 2012;109: 18909–14. doi:10.1073/pnas.1209126109

43. Allison SE, D'Elia MA, Arar S, Monteiro MA, Brown ED. Studies of the genetics, function, and kinetic mechanism of TagE, the wall teichoic acid glycosyltransferase in Bacillus subtilis 168. J Biol Chem. 2011;286: 23708–16. doi:10.1074/jbc.M111.241265

44. Schirner K, Stone LK, Walker S. ABC Transporters Required for Export of Wall Teichoic Acids Do Not Discriminate between Different Main Chain Polymers. ACS Chem Biol. 2011;6: 407–412. doi:10.1021/cb100390w

45. Peschel A, Vuong C, Otto M, Götz F. The D-alanine residues of Staphylococcus aureus teichoic acids alter the susceptibility to vancomycin and the activity of autolytic enzymes. Antimicrob Agents Chemother. American Society for Microbiology; 2000;44: 2845–7. doi:10.1128/AAC.44.10.2845-2847.2000

46. Neuhaus FC, Baddiley J. A continuum of anionic charge: structures and functions of D-

alanyl-teichoic acids in gram-positive bacteria. Microbiol Mol Biol Rev. American Society for Microbiology; 2003;67: 686–723. doi:10.1128/MMBR.67.4.686-723.2003

47.   Dengler V, Meier PS, Heusser R, Kupferschmied P, Fazekas J, Friebe S, et al. Deletion of hypothetical wall teichoic acid ligases in Staphylococcus aureus activates the cell wall stress response. FEMS Microbiol Lett. 2012;333: 109–20. doi:10.1111/j.1574-6968.2012.02603.x

48.   Brown ED, Wright GD. Antibacterial drug discovery in the resistance era. Nature. Nature Research; 2016;529: 336–343. doi:10.1038/nature17042

49.   Lewis K. Platforms for antibiotic discovery. 2013; doi:10.1038/nrd3975

50.   Schatz A, Bugie E, Waksman SA. Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. 1944. Clin Orthop Relat Res. 2005; 3–6. Available: http://www.ncbi.nlm.nih.gov/pubmed/16056018

51.   Cox G, Wright GD. Intrinsic antibiotic resistance: mechanisms, origins, challenges and solutions. Int J Med Microbiol. 2013;303: 287–92. doi:10.1016/j.ijmm.2013.02.009

52.   Fajardo A, Martínez-Martín N, Mercadillo M, Galán JC, Ghysels B, Matthijs S, et al. The neglected intrinsic resistome of bacterial pathogens. Falagas M, editor. PLoS One. 2008;3: e1619. doi:10.1371/journal.pone.0001619

53.   Silver LL. Challenges of Antibacterial Discovery. Clin Microbiol Rev. 2011;24: 71–109. doi:10.1128/CMR.00030-10

54.   Farha MA, Brown ED. Unconventional screening approaches for antibiotic discovery. Ann N Y Acad Sci. 2015;1354: 54–66. doi:10.1111/nyas.12803

55.   Heeb S, Fletcher MP, Chhabra SR, Diggle SP, Williams P, Cámara M. Quinolones: from antibiotics to autoinducers. FEMS Microbiol Rev. Wiley-Blackwell; 2011;35: 247–74.

doi:10.1111/j.1574-6976.2010.00247.x

56.   Zhanel GG, Walkty A, Vercaigne L, Karlowsky JA, Embil J, Gin AS, et al. The new

        fluoroquinolones: A critical review. Can J Infect Dis. Hindawi Publishing Corporation;

        1999;10: 207–38. Available: http://www.ncbi.nlm.nih.gov/pubmed/22346384

57.   Bourne C. Utility of the Biosynthetic Folate Pathway for Targets in Antimicrobial

        Discovery. Antibiotics. 2014;3: 1–28. doi:10.3390/antibiotics3010001

58.   Tacconelli E, Autenrieth IB, Peschel A. Fighting the enemy within. Science. 2017;355:

        689–690. doi:10.1126/science.aam6372

Chapter 2

Sortase Transpeptidases:
Structural Biology and Catalytic Mechanism

2.1 Overview

Gram-positive bacteria use sortase cysteine transpeptidase enzymes to covalently attach proteins to their cell wall and to assemble pili. In pathogenic bacteria sortases are potential drug targets, as many of the proteins that they display on the microbial surface play key roles in the infection process. Moreover, the *Staphylococcus aureus* Sortase A (SaSrtA) enzyme has been developed into a valuable biochemical reagent because of its ability to ligate biomolecules together *in vitro* via a covalent peptide bond. Here we review what is known about the structures and catalytic mechanism of sortase enzymes. Based on their primary sequences, most sortase homologs can be classified into six distinct sub-families, called class A to F enzymes. Atomic structures reveal unique, class-specific variations that support alternate substrate specificities, while structures of sortase enzymes bound to sorting signal mimics shed light onto the molecular basis of substrate recognition. The results of computational studies are reviewed that provide insight into how key reaction intermediates are stabilized during catalysis, as well as the mechanism and dynamics of substrate recognition. Lastly, the reported *in vitro* activities of sortases are compared, revealing that the transpeptidation activity of SaSrtA is at least 20-fold faster than other sortases that have thus far been characterized. Together, the results of the structural, computational, and biochemical studies discussed in this review begin to reveal how sortases decorate the microbial surface with proteins and pili, and may facilitate ongoing efforts to discover therapeutically useful small molecule inhibitors.

2.2 Introduction

Bacteria display a variety of proteins on their surface that to enable them to effectively interact with their environment. Gram-positive bacteria use sortase cysteine transpeptidase

enzymes to covalently attach proteins to their cell wall, and to assemble pili. Sortases in pathogenic bacteria are frequently important virulence factors, as many of the proteins that they display have key roles in the infection process, such as promoting bacterial adhesion, nutrient acquisition, and the evasion and suppression of the immune response [1]–[5]. As a result, a significant amount of effort has been put forth to elucidate the mechanism of sortase-mediated catalysis and to discover small-molecule sortase inhibitors that could function as potent anti-infective agents [1], [6]–[9]. Moreover, sortases have been developed into valuable biochemical reagents to ligate distinct biomolecules together via a covalent peptide bond. This *in vitro* transpeptidation activity has been harnessed for a variety of useful applications, including among others, covalently attaching proteins to cells, attaching fluorophores or drugs to antibodies, cyclizing proteins, and immobilizing peptides on solid surfaces [10]–[15].

Sortases perform two distinct functions in bacteria: 1) attach proteins directly to the cell wall or, 2) assemble pili, long proteinaceous fibers that project from the microbial surface (Figure 1). Both reactions are mechanistically related and operate on secreted proteins that contain a C-terminal cell wall sorting signal (CWSS). The CWSS contains a five residue sortase recognition motif, frequently LPXTG, that is followed by a hydrophobic domain and a positively charged cytoplasmic anchor that retains the protein substrate in the membrane [16]. The Sortase A enzyme from *Staphylococcus aureus* (SaSrtA) has been studied in detail and is archetypal [17], [18]. SaSrtA attaches surface proteins to the cell wall by recognizing an LPXTG motif within the CWSS of its protein substrate (Figure 2A). Catalysis begins when SaSrtA's active site cysteine residue nucleophilically attacks the carbonyl carbon in the peptide bond between the Thr and Gly residues in the sorting signal (Figure 2A, step 1) [7], [19]. This generates a tetrahedral intermediate that quickly collapses to form a semi-stable thioacyl intermediate in which sortase is covalently linked

via its cysteine residue to its protein substrate (Figure 2A, step 2). SaSrtA then recognizes a second substrate, the cell wall precursor, lipid II (Figure 2A, step 3) [20], [21] and catalyzes a reaction in which the N-terminal primary amine group within the cross-bridge peptide nucleophilically attacks the carbonyl carbon atom within the thioacyl bond. This second transient tetrahedral intermediate resolves into the protein-lipid II product in which the components are joined via a peptide bond (Figure 2A, step 4) [21]–[23]. The lipid II-linked protein is then incorporated into the peptidoglycan via the transpeptidation and glycosylation reactions that synthesize the cell wall. In contrast to attaching proteins to the cell wall, a second type of sortase, frequently called 'pilin polymerases,' construct bacterial pili by polymerizing pilin protein subunits (Figure 2B) [4], [5], [24]–[27]. These pilin-assembling enzymes employ a similar transpeptidation reaction as SaSrtA, but instead of using lipid II as a nucleophile to attach proteins to the cell wall, a lysine amino group located within a protein pilin subunit is used as a secondary substrate to attack the sortase-protein thioacyl intermediate (Figure 2B, steps 3 and 4). This reaction constructs pili by covalently linking protein subunits together via lysine-isopeptide bonds. Both types of sortase-catalyzed processes occur on the extracellular membrane, where the enzyme and its substrate are membrane-associated [4], [28], [29].

At present, 3,330 gene sequences encoding sortase enzymes have been identified within 1,098 species of bacteria [30]. Sortases are primarily found in Gram-positive bacteria, but are also present to a lesser extent in some species of Gram-negative and archaebacteria. Based on their primary sequences, most sortase homologs can be classified into six distinct sub-families, called class A to F enzyme [4], [31]. Class A, B and D enzymes are prevalent in bacteria within the Firmicutes phylum, while class E and F enzymes predominate in Actinobacteria. Class C enzymes are found in both Firmicutes and Actinobacteria. Similar to SaSrtA (a class A enzyme), all sortases

contain a His-Cys-Arg catalytic triad [32]–[34], and their primary sequences harbor a highly conserved TLXTC motif that contains the catalytically essential cysteine residue [7]. All sortases characterized to date catalyze a transpeptidation reaction that joins an LPXTG-like sorting signal within the CWSS of their protein substrate to an amino nucleophile. However, their sorting signal and nucleophile substrate specificities can vary substantially. These distinct specificities enable microbes to utilize more than one type of sortase to elaborate their surface, with each sortase operating non-redundantly to display or assemble distinct proteins on the cell surface [35], [36].

A number of excellent reviews have been written that describe the overall function of sortases in bacteria, their development as biochemical reagents, and efforts to discover therapeutically useful sortase inhibitors [1]–[15]. In this chapter, we review what is known about their atomic structures and the molecular basis of substrate recognition and catalysis.

2.3 Structural biology: enzyme structure and class specific variations

2.3.1 The archetypal SaSrtA enzyme

The NMR structure of SaSrtA determined by the Clubb and Schneewind groups was the first reported structure of a sortase enzyme [34] (Figure 3A). The primary sequence of this class A enzyme exhibits features that are generally conserved in other sortases. It has three components: (i) an N-terminal signal sequence that enables it to be transported across the membrane through the Sec translocon, (ii) a non-polar segment of amino acids that embed the enzyme in the bilayer, (iii) and a conserved, water-soluble C-terminal catalytic domain that contains the His-Cys-Arg triad [32], [33]. The structure of the catalytic domain was determined, residues 60–206 of SaSrtA

(SaSrtA$_{\Delta 59}$). This structure revealed the now canonical "sortase fold" that contains a closed eight stranded β-barrel architecture (Figure 3A). The atomic coordinates of SaSrtA$_{\Delta 59}$ were precisely defined, with the exception of a 19 amino acid flexible loop that connects strands β6 to β7 (the β6/β7 loop). Subsequently, a crystal structure of SaSrtA$_{\Delta 59}$ was determined at 2.0 Å resolution that is very similar to the solution structure; the Cα coordinates in structurally ordered parts of the crystal and NMR structures have an RMSD of 1.97 Å [37]. In both structures, residues in the His-Cys-Arg triad are positioned adjacent to one another within the active site (His120, Cys184 and Arg197 in SaSrtA). As described in detail later, subsequent studies revealed that the LPXTG sorting signal substrate of SaSrtA binds to a pocket that is positioned adjacent to the active site cysteine (Figure 3B) [38]. The base of the pocket in SaSrtA is formed by residues in strands β4 and β7, while the walls are formed by surface loops that connect strand β6 to β7 (β6/β7 loop), strand β3 to β4 (β3/β4 loop), and strand β2 to helix H2 (β2/H2 loop). All sortases are thought to utilize similarly positioned sorting signal binding grooves. Interestingly, SaSrtA requires Ca^{2+} for efficient catalysis, as the removal of this ion reduces activity 5-fold [34], [39]; other divalent cations also affected enzyme activity, but to a lesser extent. NMR chemical shift mapping and biochemical studies revealed that Ca^{2+} binds to a pocket located between the β3/β4 and β6/β7 loops, where it is likely coordinated by the sidechains of Glu105, Glu108, Asp112, Glu171 and a backbone carbonyl from Asn114 [34], [39]. NMR dynamics experiments indicate that Ca^{2+} binding alters the mobility and structure of the β6/β7 active site loop, thereby allosterically regulating the enzyme's affinity for the sorting signal. Interestingly, this mechanism of metal-dependent regulation appears to be unique to SaSrtA, as no other structurally characterized member of the sortase-superfamily contains a structurally similar Ca^{2+} binding pocket.

2.3.2 Class-specific variations

Structures of class A, B, C, D and E sortases have been reported, revealing unique class-specific variations that likely impact function and modulate substrate specificity. Table 1 lists the structures of sortases that have thus far been determined, and Figure 4 displays representative class A to E structures for comparison. As expected, all types of sortases contain a conserved catalytic domain that adopts a sortase-fold (Figure 4, *blue*). However, there are significant class-specific variations that are localized to four distinct structural foci: (*i*) the N-terminal segment that precedes the catalytic domain (*red*), (*ii*) the loop between strands *β6* and *β7* (the *β6/β7* loop) (*green*), (*iii*) the loop between strands *β7* and *β8* (the *β7/β8* loop) (*yellow*), and (*iv*) the C-terminal polypeptide segment that follows the catalytic domain (Figure 4). Below, we discuss these differences and highlight their functional implications if they are known.

2.3.2.1 Class A housekeeping enzymes: variable active site loops and N-termini may modulate substrate recognition

These enzymes are typified by the aforementioned SaSrtA sortase (Figure 4A). They have been proposed to perform a housekeeping role in the cell by anchoring a large number of functionally distinct proteins to the cell wall [35]. Bioinformatics and biochemical analyses indicate that they recognize sorting signals that contain an LPXTG consensus, where X is any amino acid. In addition to SaSrtA, class A enzyme structures have been reported for *Streptococcus pyogenes* [40], *Bacillus anthracis* [41], *Streptococcus agalactiae* [42], and *Streptococcus mutans* [43]. A common feature is the presence of a short helix within the *β6/β7* loop, which has been shown to contact the sorting signal upon substrate binding [38], [44]. However, the dynamics of

the loop can vary substantially. In the apo-form of SaSrtA, the loop is structurally disordered and mobile, therefore, lacking the *β6/β7* helix [34], [39]. However, sorting signal binding triggers a disordered-to-ordered transition, resulting in helix formation and helix-substrate interactions in the holo-enzyme (described below) [38]. In contrast, all other class A enzyme structures contain a pre-formed binding pocket for the sorting signal in which the *β6/β7* loop adopts an ordered state that contains the short helix.

Interestingly, class A enzymes exhibit structural variations near their catalytic histidine residue, causing some structures to contain a second groove that leads into the active site. This structural variation was first highlighted in the crystal structure of *S. pyogenes* SrtA (SpySrtA) [40]. Its catalytic domain adopts the same canonical eight-stranded *β*-barrel sortase fold typified by SaSrtA, but the positioning of its Cys-S sulfhydryl group differs. In SpySrtA, this sulfhydryl group points towards the active site His-δN and is separated by a distance of 5.4 Å, whereas in SaSrtA it points away from His-δN such that they are separated by 6.5 Å. These differences are illustrated by Figures 3B and 3C, which show the structures of SaSrtA and SpySrtA, respectively. The side chain positioning in the SpySrtA structure results in the formation of a unique groove that leads into the active site; the walls are defined by residues in helix H1 and the *β7/β8* loop, and the base is defined by residues in the *β4/β5* loop (Figure 3C, *yellow*). This groove is positioned adjacent to the active site, opposite the sorting signal binding groove, providing a potential binding site for amino acids located C-terminal to the LPXTG motif in the protein substrate, or for the lipid II substrate. In contrast, this groove is partially masked in the structure of apo-SaSrtA because of interactions between residues in the *β7/β8* loop and helix H1. The structure of the *S. agalactiae* SrtA (SagSrtA) has also been determined, and like SpySrtA, contains the same secondary groove adjacent to the active site. The SagSrtA structure is unique for the conformation of its *β6/β7* loop;

however, it seems likely that this difference may be a byproduct of the buffer used to crystallize the protein, as residues in the loop coordinate a Zn^{2+} ion in conjunction with an adjacent protein in the asymmetric unit [42] .

Some class A enzymes contain a flexible, N-terminal appendage that may modulate substrate binding. In the NMR structure of the *B. anthracis* SrtA (BaSrtA) sortase, the appendage, formed by residues Asp^{57} to Val^{79}, wraps around the body of the protein to contact the enzyme's active site (Figure 4A, colored *red*) [41]. The first eight residues, Asp^{57} to Pro^{64}, adopt an extended conformation and partially shield the active site His^{126} residue from the solvent, while the remainder of the appendage contains a short alpha helix and wraps around the surface of the catalytic domain to contact helix H2 and the $\beta 2/H2$ loop. Recent studies suggest that the N-terminal appendage modulates substrate access to the enzyme, possibly increasing the efficiency of protein display by reducing unproductive hydrolytic cleavage of enzyme-protein covalent intermediates that form during the cell wall-anchoring reaction [44]. Conformational plasticity in a related N-terminal segment has also recently been observed in the structure of the class A sortase from *S. mutans* SrtA (SmSrtA) [43]. The SmSrtA crystal structure reveals a dimer in which an extended N-terminal helix preceding the catalytic domain interacts with residues in the active site of a symmetry-related molecule. Although these intermolecular interactions may be an artifact of crystallization, they highlight the proclivity of amino acids preceding the catalytic domain to interact with the enzyme's active site, which has now been observed in BaSrtA and several class C enzymes (described below). NMR studies also revealed that the SaSrtA and BaSrtA enzymes exhibit distinct active site conformational dynamics even though they recognize sorting signals that contain an LPXTG motif [34], [39], [41]. In BaSrtA, the $\beta 6/\beta 7$ and $\beta 7/\beta 8$ loops adopt rigid and mobile states prior to engaging the sorting signal, respectively; however, these loop dynamics

are reversed in SaSrtA (the $\beta6/\beta7$ and $\beta7/\beta8$ loops are mobile and structured, respectively). The coordinates for the class A enzyme from *Streptococcus pneumoniae* (SpnSrtA) have also been deposited in the PDB, but its biological significance is unclear as the protein adopts an unusual β-strand-swapped dimer and a paper describing this structure has not been published.

2.3.2.2 Mixed function class B enzymes: an extended β6/β7 loop is used to recognize non-canonical sorting signals

Class B enzymes have diverse functions, with members of this sub-family either anchoring proteins to the cell wall or acting as pilin polymerases that assemble pili [45]–[47]. The sorting signals recognized by class B enzymes vary, but are predicted to have a NPX[T/S][N/G/S] consensus instead of LPXTG [35]. Five crystal structures of class B sortases have been reported. When compared to the canonical SaSrtA structure, two major differences are apparent as class B enzymes contain: (i) additional helices located N-terminal to the catalytic domain (Figure 4B, colored *red*), and (ii) a much longer $\beta6/\beta7$ loop that contains an additional α-helix (Figure 4B, colored *green*). Structures of class B enzymes that attach proteins to the cell wall were determined first. In 2004, structures of the *S. aureus* SrtB (SaSrtB) enzyme bound to several non-specific sulfhydryl modifiers were elucidated [48]. As described below, subsequent studies of SaSrtB bound to its signal peptide revealed that the extended $\beta6/\beta7$ loop is involved in recognizing its distinct NPQTN sorting signal substrate [49]. The role of the additional N-terminal helices remains unknown, but they may be important for dictating this class B enzyme's preference for anchoring substrates to buried, uncrosslinked portions of the cell wall [50], [51]. The apo-structures of SaSrtB and *B. anthracis* SrtB (BaSrtB) were also reported in 2004 and are structurally similar (C_α coordinate RMSDs 3.2Å) [52]. These structures differ in that a portion of the $\beta7/\beta8$ loop in BaSrtB

appears to be dynamic, as electron density for this region is missing. Minor conformational variations also occur in the $\beta6/\beta7$ loop and the short loop that connects helices H1 and H2 in these enzymes. A structure of BaSrtB bound to a aryl (beta-amino)ethyl ketone inhibitor has also been determined, revealing only small structural differences with apo-form of the enzyme [53]. Very recently the atomic structure of the SrtB enzyme from *Clostridium difficile* (CdSrtB) was determined by X-ray crystallography [54] and is nearly identical in structure to SaSrtB (RMSD = 1.93 Å for all C_α atoms). Interestingly, unlike the SaSrtB and BaSrtB proteins that attach heme-binding proteins to the cell wall, CdSrtB appears to play a more generalized function, as a genetic analysis has predicted that it attaches seven proteins to the cell wall, none of which are thought to be involved in iron acquisition [54].

Some members of the class B sub-family function as pilin polymerases, instead of attaching proteins to the cell wall. Structures of two polymerizing class B sortases have been determined, *S. pyogenes* (SpySrtB) and *S. pneumoniae* (SpnSrtB) (also referred to as SrtG-1) [45], [55]. This work revealed that class B enzymes that attach proteins to the cell wall or assemble pili adopt generally similar tertiary structures. However, the polymerizing enzymes are unique because they contain an additional short helix and β-strand within the extended $\beta6/\beta7$ loop (Figure 4B, SpySrtB). The latter alteration adds an uncommon 9th β-strand to the protein that is not inserted into the conserved β-barrel core, but instead paired with a portion of strand $\beta6$ on the protein's surface. Understanding how class B sortases can have similar structures yet distinct functions is a major unresolved question.

2.3.2.3 Class C pilin polymerases: an N-terminal "lid" regulates sorting signal substrate access

Class C sortases can function as polymerases that link pilin proteins together via lysine-isopeptide bonds to construct pili (Figure 2) [24]. In some instances, a single class C enzyme can also perform double duty, acting as both a pilin polymerase and a cell wall anchoring sortase that attaches proteins to the peptidoglycan. Members of this class recognize proteins with sorting signals that contain the consensus [L/I]PXTG [35]. Because of their unique polymerizing function, class C enzymes have been actively studied with a total of 15 structures being reported to date. Interestingly, nearly all of these enzymes contain a "lid" structure, an elongated N-terminal region that occludes the active site (Figure 4C, *red*) [4], [26], [56], [57]. Members of this group also harbor a unique C-terminal nonpolar helix that is important for function and likely embedded in the membrane [28], [29].

The first class C enzymes to be structurally characterized were SrtC-1 and SrtC-3 from *S. pneumoniae* (SpnSrtC1 and SpnSrtC3) by Manzano and coworkers in 2008 [56]. This work revealed the presence of an N-terminal lid that contains a conserved DP(F/W/Y) motif. The aspartic acid residue in the lid motif favorably contacts the conserved active site arginine residue in the His-Cys-Arg triad. The lid is also bound to the active site via sulfur-aromatic interactions between the active site cysteine and aromatic residues in the lid (Phe, Trp, or Tyr depending upon the enzyme). Due to these key interactions, the Asp and aromatic residues in the lid motif are referred to as "anchors" [56], [57]. Interestingly, the B-factors for residues comprising the lid are elevated, suggesting that they are mobile in solution and transiently detach from the enzyme's active site. Subsequent structures of class C sortases from *S. agalactiae* [29], [58], [59], *A. oris* [60], and *S. pneumoniae* [61] supported this idea, revealing similar patterns of elevated B-factors

or stretches of missing electron density in regions flanking the lid anchor residues. Several other class C structures have been determined, which reveal only small shifts in lid positioning or subtle differences in the amount of electron density that define their lids.

It has been proposed that the lid regulates enzyme activity [56], [57] (add refs [47], [48]. When it adopts the closed state observed in nearly all crystal structures, the lid occludes the binding site for the sorting signal and holds the enzyme in an inactive state. The enzyme can then become activated by partial dislodgement of the lid, enabling binding of the sorting signal and formation of the enzyme-substrate thioacyl intermediate. This notion is compatible with modeling studies of the sortase-signal complex, as well as *in vitro* data that has demonstrated that mutants harboring alterations in the lid exhibit increased rates of sorting signal hydrolysis [29] and, in some instances, reduced stability [57]. However, the role of the lid in catalysis is not fully understood, since cellular studies of pilin polymerases containing mutations that should presumably dislodge the lid exhibit wild-type transpeptidation activity *in vivo* [28], [29]. Moreover, structural data has shown that the lid does not completely block access to the enzyme active site, since the crystal structure of *S. agalactiae* SrtC-1 (SagSrtC1) shows the enzyme bound to the non-specific sulfhydryl modifier (MTSET) despite adopting a closed-lid state [58].

Recent NMR and computational studies of SpnSrtC1 suggest that the lid in class C enzymes prefers to adopt a closed and rigid state [62]. This work revealed that the lid in SpnSrtC1 adopts a rigid conformation in solution that is devoid of large magnitude conformational excursions that occur on mechanistically relevant timescales. Additionally, point mutations in the lid were shown to induce dynamic behavior that correlates with increased hydrolytic activity and sorting signal substrate access to the active site cysteine. These results support the notion that the lid in this class C enzyme has a negative regulatory function, and imply that a significant energetic

barrier must be surmounted to dislodge it from the active site and initiate pilus biogenesis. Presumably, an as of yet unidentified factor(s) must pry the lid open to hold the enzyme in a catalytically active state that can assemble pili.

One structure of the SagSrtC1 enzyme appears to have captured the lid in an "open" conformation, providing insight into the mechanism of lid-opening that is expected to precede binding of the sorting signal (Figure 4C, compare left to right) [58]. The structure of SagSrtC1 was determined from multiple crystalline forms, and one of these structures, solved in space group C2, showcases the lid in an "open" conformation. The enzyme maintains the typical sortase fold and, excluding the N-terminal extension preceding the β-barrel core, is extremely similar to SagSrtC1 structures previously solved in space groups $P2_12_12_1$ and $P3_12$ with an average backbone RMSD of 0.72 Å [58]. However, in space group C2, residues A38-E71, which typically form the lid structure, instead form an extended helical structure with the aromatic lid anchor residue of the conserved DP(F/W/Y) motif (Y51) displaced from the active site by over 30 Å to a position where it stacks against the backbone of helix H2. An additional crystal structure of *S. suis* SrtC-1 (SsSrtC1) also maintains a similar "open" conformation of the lid, with the same extended helix replacing what was expected to be a closed lid [63]. As nearly all structures of class C enzymes possess a closed lid, which NMR and computational studies suggest is immobile, it is tempting to speculate that the unique open structure observed for SagSrtC1 and SsSrtC1 arose from the solvent conditions used to obtain this crystal form of the protein. Finally, when regions outside of the lid are compared in other class C structures, perhaps the most significant deviation from the norm is the addition of a short C-terminal α-helix opposite the active site in the structure of SpnSrtC3; however, its functional importance has not been determined [56].

2.3.2.4 Class D enzymes: specialized sortases that attach cell wall proteins that contain an LPXTA sorting signal

Class D sortases predominate in *Bacilli* species and recognize an unusual LPXT<u>A</u> motif consensus, in which an alanine (underlined) replaces the canonical glycine residue [35]. Currently, only two structures of class D enzymes have been reported. In 2012, the NMR structure of the *B. anthracis* SrtD enzyme was published (BaSrtD, and also previously referred to as SrtC) [64]. BaSrtD anchors proteins required for efficient sporulation to the surface of the cell wall. The catalytic domain of BaSrtD adopts the conserved eight-stranded β-barrel sortase fold (Figure 4D). Structurally, it is most similar to members of the class A subfamily as it contains an ordered 3_{10} helix within the $\beta6/\beta7$ loop and lacks the elongated $\beta6/\beta7$ loop and lid that are found in class B and C enzymes, respectively (Figure 4). Interestingly, ultracentrifugation studies indicate that isolated BaSrtD forms a dimeric structure with a K_D of 89 µM [64]. Based on resonance line broadening effects observed in its NMR spectrum, BaSrtD dimerization is potentially mediated by residues in the structurally disordered $\beta2/\beta3$ and $\beta4/H1$ surface loops that are positioned adjacent to the active site histidine residue. It has been suggested that this disordered surface may mediate interactions with lipid II or other factors on the cell surface, but this has not been demonstrated experimentally. Recently, the structure of the *Clostridium perfringens* SrtD (CpSrtD) sortase was determined at 1.99Å [65]. Similar to BaSrtD, CpSrtD adopts a β-barrel sortase fold that contains a short helix within the β6/ β7 loop. However, CpSrtD also contains two alpha helices at its N-terminus, and unlike BaSrtA, its $\beta2/\beta3$ and $\beta4/H1$ loops are structurally ordered (the $\beta2/\beta3$ loop also contains a two-turn alpha helix). Furthermore, CpSrtD exists as a monomer according to crystallographic and dynamic light scattering studies. Interestingly, CpSrtD exhibits catalytic activity *in vitro* that is enhanced in a magnesium-dependent manner, making it one of only two

known sortases (the other being SaSrtA) whose activity is modulated by metal ions. The origin of this stimulatory effect is not known in CpSrtD, but in SaSrtA, metal binding increases enzyme activity by modulating the structure and dynamics of the $\beta6/\beta7$ loop (described below).

2.3.2.5 Class E enzymes: a novel LAXTG sorting signal for anchoring of surface proteins in Actinobacteria

Class E enzymes predominate in soil and freshwater-dwelling Actinobacteria and have not been studied extensively. Bioinformatic predictions suggest that members of this group recognize an unusual L<u>A</u>XTG sorting signal motif in which the highly conserved proline residue is replaced with alanine (underlined) [35]. Two class E sortases from *Streptomyces coelicolor*, ScSrtE1 and ScSrtE2, have been shown to display chaplin proteins to promote aerial hyphae development [66]. *In vitro* studies indicate that ScSrtE1 and ScSrtE2 can hydrolyze LAETG- and LAHTG-containing peptides, cleaving the peptide bond after the threonine residue. The enzymes exhibit promiscuous activity, as they also cleave an LAETG peptide at a secondary site following the alanine [66]. Very recently, we reported the first crystal structure of a class E sortase, the 1.93 Å resolution structure of ScSrtE1 (Figure 4E) [67]. The structure is similar to class A enzymes, as its $\beta6/\beta7$ loop contains a single 3_{10} helix. However, variations in the conformation of its $\beta3/\beta4$ and $\beta6/\beta7$ loops are evident. ScSrtE1 contains a 21 amino acid insertion immediately following the 3_{10} helix in the $\beta6/\beta7$ loop. This long insertion is similar in length to that observed in class B sortases, but is distinctly devoid of secondary structure, whereas class B sortases contain an additional alpha helix. In the ScSrtE1 structure, the active site is bound to a tripeptide that is presumably a proteolytic protein fragment that was present in the crystallization buffer. The coordinates of the tripeptide and computational modeling with substrate mimics suggest that ScSrtE1 and other members of this group may use a

class E-specific tyrosine residue present within their $\beta3/\beta4$ loops to recognize the alanine backbone within the LAXTG substrate. It is possible that the tyrosine participates in a hydrogen-bond with the amide nitrogen of the alanine residue within the signal, an interaction that is not possible in substrates containing a proline at this position. However, the role of the conserved tyrosine in dictating substrate specificity was not experimentally determined because single amino acid mutants of ScSrtE1 that altered this position in the protein were unstable. Class F enzymes are also prevalent in Actinobacteria, but they have not been structurally or biochemically characterized.

2.4 Structural biology: molecular basis of substrate recognition

All sortases characterized to date catalyze a transpeptidation reaction that joins an LPXTG-like sorting signal within the CWSS of their protein substrate to an amino nucleophile (Figure 2). However, depending upon the type of sortase, the chemical structure of these substrates can vary substantially. For example, biochemical and bioinformatics analyses suggest that class A, B, C, D and E enzymes have evolved specificities for distinct LPXTG, NPX[T/S][N/G/S], [L/I]PXTG, LPXTA and LAXTG sorting signals, respectively (unambiguous differences from LPXTG are underlined) [35]. In addition, they can either recognize nucleophiles that originate from lipid II (sortases that anchor proteins to the cell wall), or a lysine residue located within another protein (sortases that function as pilin polymerases) (Figure 2). Typically, microbes encode genes for more than one sortase enzyme [36]. Their distinct substrate specificities enable multiple sortases to non-redundantly operate, with different types of sortases "sorting" distinct proteins to the cell surface or assembling pili. Below, we summarize what is currently known about the molecular basis of substrate recognition.

2.4.1 Sorting signal recognition

The sorting signal recognized by sortases is located within the CWSS of the protein substrate, which in turn is embedded in the bilayer via its hydrophobic domain (Figure 1) [16]. Our structural studies have shed light onto how class A and B enzymes recognize LPATG and NPQTN sorting signals, respectively [38], [44], [49]. Since sortases can hydrolyze their cognate sorting signals and bind to them weakly *in vitro*, these structural studies made use of a substrate analog developed by the Jung group in which the threonine moiety contains a sulfhydryl group in place of its carbonyl atom [68]. The analog (hereafter called T*), mimics the threonine residue in the native sorting signal substrate, but forms a disulfide bond with the active site cysteine residue. This leads to the production of a stable enzyme-substrate complex that is suitable for structural studies.

Using T*-containing sorting signal analogs the atomic structures of three sortase-substrate complexes have been determined by the Clubb group: the class A SaSrtA-LPXT* (Figure 5A), class B SaSrtB-NPQT* (Figure 5B), and class A BaSrtA-LPAT* (Figure 5C) complexes [38], [44], [49]. To facilitate a discussion of the binding interactions that govern signal recognition, we henceforth refer to residues in each sorting signal in relation to their positioning relative to the scissile bond. Residues in an L-P-X-T-G sorting signal are referred to as P4-P3-P2-P1-P1' and their corresponding binding sites on the enzyme as subsites S4-S3-S2-S1-S1', respectively.

2.4.1.1 Sorting signal recognition by class A enzymes

The NMR structure of the SaSrtA-LPAT* complex provided the first insight into the molecular basis of sorting signal recognition [38] (Figure 5A). In the structure of the complex, the

peptide substrate binds to a pocket adjacent to the active site cysteine whose base is formed by residues located in strands $\beta4$ and $\beta7$, and whose walls are formed by surface loops that connect strand $\beta6$ to $\beta7$ ($\beta6/\beta7$ loop), strand $\beta3$ to $\beta4$ ($\beta3/\beta4$ loop), and strand $\beta2$ to helix H2 ($\beta2/H2$ loop). The location of this binding site is consistent with chemical shift mapping studies [69]. Binding of the signal to SaSrtA causes a major reorganization of the active site, including a disordered-to-ordered transition of the $\beta6/\beta7$ loop to create a short 3_{10} helix that contacts the signal, as well as displacement of the $\beta7/\beta8$ loop. The latter change may have a regulatory role, exposing the active site histidine to the solvent and possibly facilitating the binding of the secondary lipid II substrate only after the sorting signal has first bound to the enzyme. Additional ^{15}N relaxation analyses revealed that the $\beta6/\beta7$ loop, which is highly dynamic in the unbound state, rigidifies upon peptide binding [38], [39]. The peptide binding mode in the SaSrtA-LPAT* complex differs substantially from a previously reported crystal structure of SaSrtA non-covalently bound to a LPETG peptide [37]. In the crystal structure, the peptide is presumably non-specifically bound, which is not surprising as the sorting signal substrate binds to SaSrtA with very weak affinity ($K_m = 7.33$ mM) [70].

The structure of the SaSrtA-LPAT* complex reveals how class A enzymes recognize the P4 and P3 residues within the LPXTG sorting signal, Leu and Pro, respectively [38]. The P4 leucine side chain is positioned within a large hydrophobic S4 pocket that is formed by residues V161, V166, V168, and L169 in the reordered $\beta6/\beta7$ loop, and I199 on strand $\beta8$. The P3 Pro residue plays a key architectural role, forming a kink in the sorting signal that enables it to adopt an 'L-shaped' structure in which the C-terminal end is positioned towards the active site cysteine. Numerous hydrophobic contacts to the proline are formed by the S3 site. Surprisingly, in the structure of the SaSrtA-LPAT* complex, the positioning of the P1 and P2 side chains is

incompatible with biochemical and bioinformatics data [35], [71], [72]. Subsequent structural and computational studies revealed the origin of this discrepancy and are described in section [2.1.C].

The recently determined structure of the class A BaSrtA-LPAT* complex revealed a generally similar binding mode for the LPXTG sorting signal, but also indicated fundamental differences in the conformational dynamics and structure of the active site [44]. Similar to SaSrtA, the peptide adopts an L-shaped conformation in the BaSrtA-LPAT* complex by virtue of a kink at position P3, and the $\beta6/\beta7$ loop is instrumental in building the S4 site that dictates specificity. However, four significant differences in the binding mechanism are apparent. First, in contrast to SaSrtA, the $\beta6/\beta7$ loop in BaSrtA is structured in the absence of substrate [41] and only experiences modest changes upon signal binding (the coordinates of the Cα-backbone atoms in the loop in the apo- and signal-bound forms of BaSrtA have an RMSD of 0.76 Å). Thus, the $\beta6/\beta7$ loop forms a pre-formed binding pocket for the P4 residue in BaSrtA, whereas in SaSrtA, the loop is flexible prior to signal binding [38], [39]. Second, the $\beta7/\beta8$ loop, which was disordered in the unmodified BaSrtA structure, undergoes a disordered-to-ordered transition as a result of binding to the substrate. A comparison of the apo- and bound-forms of BaSrtA reveals that the active site cysteine residue is displaced ~7 Å upon peptide binding [41], [44]. This movement allows the $\beta7/\beta8$ loop to form new intra-protein interactions with residues within the $\beta4/\beta5$ loop, which presumably stabilize the $\beta7/\beta8$ loop and cause it to become ordered. In contrast, the $\beta7/\beta8$ loop of SaSrtA is structurally ordered in the apo-state and is displaced when the signal binds. Third, unlike SaSrtA, the BaSrtA enzyme contains an N-terminal appendage that partially encapsulates the sorting signal [38], [44]. The N-terminal appendage precedes the catalytic domain in the primary sequence and wraps around the body of the protein to contact the active site. The side chain of Ile61 in the appendage forms extensive contacts with the active site histidine, while the hydroxyl

group of Ser59 is positioned to form a hydrogen bond with the backbone carbonyl oxygen of the proline residue in the bound peptide. Interestingly, NMR and *in vitro* kinetics data suggest that the appendage transiently detaches from the isolated enzyme when the sorting signal is bound. A model of the thioacyl intermediate constructed from the coordinates of the complex suggests that the N-terminal appendage may obstruct nucleophile access to the active site, potentially increasing the efficiency of protein display by reducing the unproductive hydrolytic cleavage of enzyme-protein covalent intermediates. A fourth difference between the BaSrtA-LPAT* and SaSrtA-LPAT* complexes is the positioning of the P1 and P2 side chains, which are described in section [2.1.C].

2.4.1.2 Sorting signal recognition by class B enzymes

The structure of the SaSrtB-NPQT* complex provided the first-ever insight into signal recognition by class B enzymes [49], which binds an NPQTN sorting signal in the case of SaSrtB (Figure 4B) [47]. SaSrtB uses a rigid pocket to bind to the 'L' shaped peptide. Specifically, the P4 asparagine residue is primarily recognized by contacts to the $\beta6/\beta7$ loop. However, as compared to class A enzymes, the P4 residue is more solvent exposed, concordant with the more hydrophilic nature of the asparagine sidechain. The binding mode of the sorting signal in class A and B enzymes is compatible with biochemical data from the McCafferty group. This data demonstrated that replacement of the $\beta6/\beta7$ loop in SaSrtA with the corresponding loop from SaSrtB resulted in a chimeric protein that could recognize the SaSrtB-specific NPQTN sorting signal motif [73]. The P3 proline residue also appears to have the same function as in the BaSrtA-LPAT* and SaSrtA-LPAT* structures, altering the trajectory of the peptide so that its C-terminal end points towards the active site. As with BaSrtA, the $\beta6/\beta7$ loop in SaSrtB forms a rigid, pre-formed binding pocket

for the sorting signal; the structures of BaSrtA in the free- and bound-states are nearly identical (the C_α coordinates have an RMSD = 0.44 Å). Moreover, within the BaSrtA-LPAT* and SaSrtB-NPQT* complexes, the P1 and P2 side chains in the bound substrates adopt similar positions relative to the enzyme active site that are distinct from their positioning observed in the structure of the SaSrtA-LPAT* complex (see section [2.1.C]).

2.4.1.3 Sorting signal conformational heterogeneity: Thr-in versus Thr-out

A comparison of the structures of the SaSrtA-LPAT* and SaSrtB-NPQT* complexes revealed that there were major differences in the conformation of the P1 and P2 residues of the bound sorting signals [49]. In particular, the P2 glutamine residue in the SaSrtB-NPQT* complex rests along the wall of the pocket and points out towards the solvent, while in the SaSrtA-LPAT* structure, the analogous P2 residue (alanine) points towards the base of the active site. The P1 threonine residues in these complexes also adopt fundamentally distinct conformations. In the SaSrtB-NPQT* structure, the threonine is buried in the active site (Thr-in position) where it forms two hydrogen bonds with the active site arginine residue, whereas in the SaSrtA-LPAT* structure, it is projected toward solvent and hydrogen bonds with the active site histidine residue (Thr-out position). Interestingly, an inspection of the recently determined structure of the BaSrtA-LPAT* complex reveals a P1 and P2 positioning that is similar to what is seen in the SaSrtB-NPQT* substrate complex; the P1 threonine side chain adopts a Thr-in conformation in which it is buried within the active site. Thus, the Thr-in conformation of the peptide can be adopted by signal peptides bound to both class A and B enzymes.

Several lines of evidence suggest that the Thr-in conformer observed in the structures of the BaSrtA-LPAT* and SaSrtB-NPQT* complexes represents the catalytically active form of the bound substrate. First, in these structures, the side chain of the P2 residue projects into the solvent and is not recognized. This makes sense as the P2 residue within the sorting signals, also called the 'X' position, is not recognized by sortases according to bioinformatics [35] and biochemical data [72]. This is in contrast to the SaSrtA-LPAT* complex in which the peptide adopts the Thr-out conformer and the side chain at the P2 'X' position projects into the active site where it contacts A118 and I182. Second, the high degree of sequence conservation at site P1, which contains a threonine residue in ~95% of predicted sorting signals, suggests functional relevance of the Thr-in sorting signal conformer observed in the BaSrtA-LPAT* and SaSrtB-NPQT* complexes. This is in marked contrast to the Thr-out conformer observed in the SaSrtA-LPAT* complex that projects into the solvent and is not recognized by the enzyme. Finally, as described in section [3], computational modeling of sortase reaction intermediates indicates that the Thr-in conformer likely facilitates the creation of an oxyanion hole for substrate stabilization during thioacyl intermediate formation.

2.4.2 Nucleophile recognition

Unlike the mechanism of sorting signal recognition, sortase binding to their secondary substrate, the amino nucleophile to which the sorting signal is joined, remains poorly understood. Depending upon the type of sortase, cell wall-anchoring versus pilin polymerizing, the nucleophile can either originate from a lipid II molecule or a lysine residue within a pilin protein substrate, respectively. Thus far, only lipid II recognition by sortases has been explored, but the mechanism

of binding still remains enigmatic. Experimental studies have investigated how the SaSrtA enzyme interacts with poly-glycine peptides that mimic the penta-glycine cross-bridge peptide moiety within the intact *S. aureus* lipid II molecule, as large quantities of the intact, water-soluble portion of lipid II are difficult to obtain. NMR chemical shift mapping of the SaSrtA-LPAT* complex using a tri-glycine peptide titrant revealed a low affinity, continuous interaction surface on the enzyme that contains portions of the $\beta7/\beta8$ loop, $\beta4$/H2 loop, and an N-terminal segment of helix H1 [38]. The mapping data coarsely defines the interaction surface, but does not provide specific details about tri-glycine-enzyme interactions. Interestingly, significant chemical shift changes are not observed when apo-SaSrtA is titrated with the tri-glycine peptide, but only when added to the SaSrtA-LPAT* complex. This observation is compatible with the proposed bi-bi mechanism of catalysis [70], [74] and suggests that LPXTG sorting signal binding may direct catalysis forward by causing alterations in the $\beta7/\beta8$ loop that unmask the binding surface for lipid II. Crystals of SaSrtB modified with a small molecule sulfhydryl modifier were soaked with a tri-glycine peptide, and the structure was determined by molecular replacement. In this complex, the N-terminal amine of the tri-glycine peptide is 6.4 Å from the active site histidine, which is occluded from solvent by a closed $\beta7/\beta8$ loop [48]. While the binding mode of the tri-glycine peptide to SaSrtB is generally compatible with the NMR chemical shift data from the SaSrtA-LPAT* complex, it does not fit the accepted view of the mechanism that would have the incoming nucleophile deprotonated by the active site histidine residue.

Interestingly, other class A sortases appear to contain a pre-formed binding site for the cross-bridge peptide in lipid II. In the crystal structures of the SpySrtA, SagSrtA, and SmSrtA sortases, their $\beta7/\beta8$ loops are displaced in the absence of the sorting signal, creating a surface-exposed groove adjacent to the active site histidine and cysteine residues (Figure 3C, yellow) [40],

[42], [43]. Recent docking studies of BaSrtA highlight a potential binding site for the amino component of lipid II in *B. anthracis* (diaminopimelic acid) [44]. This site is positioned adjacent to the active site histidine, between the *β*7/*β*8 and *β*4/H3 loops, and would position the nucleophile near the electrophilic carbonyl carbon atom in the thioacyl bond. Interestingly, exposure of this site requires partial displacement of the N-terminal appendage, which has been postulated to mask the acyl-linked reaction intermediate from hydrolysis. Conclusive determination of the molecular basis of amino nucleophile selectivity will require structure determination of sortases bound to their secondary substrates. The mechanism of pilin protein substrate recognition by pilin polymerases also remains to be determined.

2.5 Computational studies

2.5.1 The Thr-in conformation enables sortases to employ a substrate-stabilized active site

Computational studies have leveraged the experimentally determined structures of sortase-substrate complexes to gain insight into the mechanism of catalysis. Jacobitz et al. generated an energy minimized model of the thioacyl intermediate using the coordinates of the SaSrtB-NPQT* complex [49]. The model of the thioacyl intermediate is shown in Figure 6 and reveals that the Thr-in conformation adopted by the highly conserved P1 threonine residue (see section (2.1.C]) enables it to hydrogen bond to the active site arginine residue. This interaction positions the active site arginine proximal to the thioacyl bond, where it presumably can stabilize the two high-energy tetrahedral oxyanion intermediates that form during catalysis [49]. In this capacity, threonine-

53

arginine interactions facilitate the formation of an oxyanion hole in which the arginine neutralizes the negative charge of the oxyanion. Further stabilization of the oxyanion is achieved by interactions from the backbone amide of E224 that is located immediately C-terminal to the active site cysteine. This novel "substrate-stabilized oxyanion hole" presumably increases substrate specificity for a threonine residue at position P1 and is compatible with biochemical studies that have shown that the enzyme is unable to utilize sorting signals that contain conservative mutations at the P1 position (e.g. Ser or Val instead of Thr) [49]. A similar computational strategy was used to model the structure of the BaSrtA thioacyl intermediate using the experimentally determined coordinates of the BaSrtA-LPAT* complex [44]. This model displayed an analogous hydrogen bonding interaction between the P1 threonine residue and the active site arginine. As in the SaSrtB model, the P1 threonine carbonyl atom in the thioacyl bond is in close proximity to the active site arginine guanidino group, suggesting that both enzymes employ related substrate-stabilized oxyanion holes to facilitate catalysis [44]. However, it should be noted that the active site conformations of the two models differ slightly, since in the model of the BaSrtA thioacyl intermediate, the active site arginine adopts a more extended structure that allows it to form an additional hydrogen bond to the P3 proline residue.

2.5.2 Sorting signals bound to SaSrtA can interchange between Thr-out and Thr-in conformers

As described in section [2.1.C], a comparison of the experimentally determined structures of the SaSrtA-LPAT*, SaSrtB-NPQT* and BaSrtA-LPAT* revealed fundamental differences in the positioning of the P1 and P2 residues in the bound sorting signal [38], [44], [49].

To rectify the discrepancies between the SaSrtA-LPAT* Thr-out and SaSrtB-NPQT* Thr-in conformations, Jacobitz et al. performed molecular dynamics (MD) simulations for three thioacyl intermediate systems: SaSrtA-LPAT, SaSrtB-NPQT, and SaSrtB-NPAT [49]. Through the use of umbrella sampling calculations, the free energy landscape of transitions between the Thr-out and Thr-in states was mapped for each complex. These studies indicated that SaSrtA-LPAT could adopt both the Thr-in and Thr-out states with equal probability, while for both NPQT and NPAT substrates, SaSrtB could only sample the Thr-in state. Based on these results, it appears as if the Thr-in conformation is likely the more evolutionarily conserved state, and that the inherent flexibility of SaSrtA allows for the Thr-out conformation with an LPAT substrate that was captured by Suree et al. [38].

2.5.3 Dynamic sorting signal recognition by SaSrtA

MD studies of SaSrtA have provided additional insight into the sorting signal recognition process. By performing conventional and accelerated MD simulations of both the sorting signal free and bound states, Kappel et al. proposed that sorting signal binding is a mixture of conformational selection and induced fit mechanisms [75]. For example, the $\beta6/\beta7$ loop appears to follow the conformational selection paradigm: it sampled a range of stable conformations in the apo-state, some of which were relatively close to the bound configurations. In contrast, the $\beta7/\beta8$ "open" state from the NMR structure was only stable in the presence of a bound sorting signal, suggesting an induced fit mechanism. In addition, analysis of the sorting signal-bound conformations showed that an allosteric network runs throughout the protein, linking the calcium ion, sorting signal, and proposed lipid-II binding regions to one another. In a complementary study,

Moritsugu et al. used the multiscale enhanced sampling method to probe the allosteric effects of the calcium ion and sorting signal [76]. Simulations of each combination of bound states showed that binding of both molecules is required to stabilize the $\beta6/\beta7$ and $\beta7/\beta8$ loops in conformations observed in the NMR-LPAT* structure. Overall, these simulations point towards a mechanism in which calcium, sorting signal, and potentially lipid II binding are modulated by a dynamic network that includes the $\beta6/\beta7$ loop region in SaSrtA.

Other aspects of the SaSrtA recognition and catalytic processes have also been explored by computational studies. Biswas et al. used a hybrid MD and biochemical experimental approach to probe the roles of the conserved sorting signal Leu and Pro residues in substrate binding [77]. Comparative simulations with LPAT, APAT, and LAAT substrates demonstrated that contacts between the leucine sidechain and SaSrtA contribute to stabilize the $\beta6/\beta7$ loop, whereas the kink that is induced by the proline appears to be essential for recognition. In another study, Tian and Eriksson performed simulations in which His120 and Cys184 were in their zwitterionic and neutral forms [78]. Their results showed that Arg197 adopts distinct conformations based upon the charged state of the protein, which helps to stabilize the catalytically active form. It should be noted that each of these studies was performed with the sorting signal in the Thr-out state. Although the global effects of the Thr-in and Thr-out states on the induced fit/conformational selection process, allosteric networks, and recognition processes are likely similar, subtle differences may exist that influence some of the fine details that resulted from these simulations.

2.6 Catalytic mechanism

The current model of the molecular mechanism of the SaSrtA enzyme is presented in Figure 7. Kinetic studies indicate that catalysis occurs through a ping-pong mechanism that begins when the sortase recognizes the CWSS of a membrane anchored protein [70], [74]. The LPXTG-type sorting signal within the CWSS binds to a groove on the sortase whose base is formed by residues in the $\beta6/\beta7$ loop, strands $\beta4$, $\beta7$, $\beta8$, and whose walls are formed by residues in the $\beta2/\beta3$ and $\beta3/\beta4$ loops (Figure 4D). Here the sorting signal's L-shaped structure dictated by the highly conserved proline residue at P3 (>90% conserved) orients the residue P4 for recognition in subsite S4 on the $\beta6/\beta7$ loop, as the C-terminal end of the sorting signal is directed towards the active site [38]. In order for catalysis to proceed, the enzyme must contain a properly charged active site in which the cysteine and histidine residues are in their thiolate and imidazolium forms, respectively. Based on pKa measurements of their active site residues [70], [79], [41], when removed from the cell surface, in isolation less than 1% of the SaSrtA and BaSrtA sortases possess an appropriately charged active site that can perform catalysis. This explains why sortases in isolation catalyze reactions very slowly, as presumably only a small fraction of these enzymes are active (section [5]). The notion that the active site is predominantly dormant is substantiated by structural analyses of SaSrtA, which revealed that its active site cysteine and histidine residues are not close enough to form a thiolate-imidazolium ion-pair (the Cys-S-His-δN distance is 6.5 and 7.6 Å in the NMR and crystal structures, respectively) [34], [37]. In the rare instance that the cysteine and histidine residues are appropriately ionized, the cysteine thiolate nucleophilically attacks the carbonyl carbon on the P1 residue (Figure 7B). This leads to the formation of the first tetrahedral intermediate, which is likely stabilized by an oxanion hole that is formed by the active site arginine residue and a backbone amide in the $\beta7/\beta8$ loop [49]. In nearly all sorting signals a

57

threonine occurs at the P1 position, which based on structural and MD simulations helps to stabilize the oxyanion hole by forming a hydrogen bond to the active site arginine [49] (Figure 7C) The first oxyanion is a transient intermediate, and quickly collapses to form a semi-stable thioacyl intermediate as the scissile peptide bond is broken. In the thioacyl intermediate the substrate's P1-Thr and the active site cysteine are joined via a thioacyl bond (Fig. 7, d) [80]. This process is presumably assisted by the conserved active site histidine residue, which may act as a general acid to protonate the amino leaving group. Beyond stabilizing the oxyanion intermediate, the side chain of the active site arginine residue may orient the substrate in the active site by forming a hydrogen bond to the backbone carbonyl atom in the P2 residue [38], [44], [78]. In the next step of the reaction, a secondary substrate bearing an amine group enters the active site and is presumably deprotonated by the active site histidine residue to facilitate its nucleophilic attack on the thioacyl bond (Figure 7E). A second tetrahedral intermediate then forms which may also be stabilized by an oxyanion hole that is constructed with the assistance of the sorting signal's P1 residue (Figure 7F). The second tetrahedral intermediate then quickly collapses to form the peptide bond linked product. Sortases that anchor proteins to the cell wall join the protein substrate to the cell wall precursor lipid II, whereas pilin polymerases join the protein to a lysine amine located within another pilin protein. These enzymes are believed to use a similar mechanism to catalyze transpeptidation.

2.7 *In vitro* transpeptidation activity

Many sortase enzymes exhibit *in vitro* transpeptidation and/or proteolytic activity. *In vitro* activity was first demonstrated for the prototypical SaSrtA enzyme by Ton-That and coworkers

[17]. Their assay utilized FRET-based detection of activity using a reporter LPXTG sorting signal peptide that contained donor and quencher fluorophores at each end. Cleavage of this peptide by SaSrtA, liberates the donor from the quencher enabling enzyme activity to be detected as an increase in fluorescence. Initial studies demonstrated that SaSrtA catalyzes an *in vitro* transpeptidation that joins two peptides, one that contains an LPXTG sorting signal motif and a second peptide that contains N-terminal glycine residues [81], [74]. Studies using this assay also demonstrated that the active site histidine, cysteine and arginine residues are important for catalysis [32], [33]. Although the FRET-based assay is easy to employ, inner filter effects can occur at high substrate concentrations leading to inaccurate measurements of the enzyme's kinetic parameters [82]. Subsequent development of a medium-throughput HPLC enzyme assay enabled more accurate measurement of the kinetic parameters and revealed that transpeptidation occurs with a $k_{cat} = 0.28 \pm 0.02$ s^{-1}, and K_m values for its LPXTG and secondary Gly$_5$ peptide substrates of 7.33 ± 1.01 mM and 196 ± 64 µM, respectively [70]. In the absence of the Gly$_5$ peptide SaSrtA acts as a protease, hydrolytically cleaving the sorting signal between the threonine and glycine residues with a $k_{cat} = 0.086 \pm 0.015$ s^{-1}. As hydrolysis occurs much slower than transpeptidation, it can be largely avoided when the Gly$_5$ peptide is in excess. As described in section [4], when sortase is purified and removed from the cell surface only a small fraction (~0.06%) contains a properly ionized active site that can catalyze transpeptidation. This small subpopulation is much more enzymatically active with an estimated k_{cat}/Km greater than 10^5 M^{-1} [70]. On the cell surface the transpeptidation reaction catalyzed by membrane associated SaSrtA may occur at a rate that is faster than the rate of the reconstituted *in vitro* transpeptidation reaction that employs short peptide substrates. This is because pulse-chase labeling experiments using intact cells indicate that the reaction is complete in <3 min, and it is likely that individual sortase enzymes attach several

proteins to the cell wall during this time [16]. The increased *in vivo* rate may result from the fact that the sortase and both of its substrates are embedded in the membrane. It is also possible that there exists yet to be identified factors on the cell surface that facilitate sortase association with its substrates and/or the conversion of its active site into the properly ionized state.

Since the original work on SaSrtA, the *in vitro* enzymatic activities of several other sortases have been characterized (summarized in Table 2). Interestingly, none of these enzymes are as active as SaSrtA and in many instances only their proteolytic activity has been demonstrated. Generally, their activities have not been rigorously characterized as only the amount of product generated by the sortase has been measured after a specific incubation time. At present, only the SaSrtA, SpySrtA and SaSrtB sortases have been shown to catalyze a transpeptidation reaction that joins two peptides together via a peptide bond. The native SaSrtA enzyme catalyzes transpeptidation ~20-500 fold faster than the other enzymes making it a useful bioconjugation reagent. Moreover, the activity of SaSrtA has been improved using directed evolution approaches, resulting in tetramutant enzyme that is ~140 fold more active than the native SaSrtA [83]. Additional rate enhancements have been achieved by altering the reaction conditions and by fusing the nucleophile substrate to SaSrtA [84]. The reader is referred to a number of excellent reviews describing the use of sortase as a bioconjugation reagent [85]–[87].

2.8 Conclusions

Sortase enzymes are ubiquitous in Gram-positive bacteria where they attach proteins to the cell wall and construct pili. Their important role in displaying virulence factors makes them promising drug-targets [1], [6]–[9], while their ability catalyze *in vitro* transpeptidation has made

them useful bioconjugation reagents [10]–[15], [87]–[89]. Considerable effort has been put forth to elucidate the molecular mechanism of catalysis, class-specific structural features that dictate function, and the molecular basis of substrate recognition. Structural and computational studies of three sortase-peptide complexes have provided insight into the initial steps of catalysis -- binding of the sorting signal to the active site and formation of the first tetrahedral and thioacyl intermediates [38], [44], [49]. This work has shown that class A and B sortases recognize their sorting signal substrates in a similar manner. The side chains of the P1, P3 and P4 residues are recognized. The bound signal adopts an L-shaped conformation as a result of kink introduced at the proline P3 residue, positioning the side chain of the P4 residue within a pocket formed by the $\beta6/\beta7$ loop, and directing the P1 and P2 residues towards the active site cysteine. The positioning of P1 and P2 exhibit conformational heterogeneity, with substrate assisted catalysis occurring when the threonine P1 side chain contacts the active arginine residue to stabilize the oxyanion hole [49]. Other classes of sortases may bind their sorting signals in a generally similar manner, with class C pilin polymerases requiring unlatching of a lid structure to enable signal access [56], and class E enzymes using a unique surface to recognize alanine at position P3 instead of proline [67]. The second-half of the transpeptidation reaction remains poorly understood -- nucleophilic attack of the thioacyl enzyme-substrate intermediate by an amine group and peptide bond formation. No conclusive evidence has emerged to pinpoint the location of the secondary substrate binding site for lipid II or pilin proteins in cell wall anchoring or pilin assembling sortases, respectively. Deciphering how some sortases function as polymerases, while others attach proteins to the cell wall will require the development of robust biochemical assays to monitor pilus assembly and novel substrate analogs to visualize nucleophile recognition. Finally, although many small molecule sortase inhibitors have been identified, they have yet to be developed into a drug to treat

bacterial infections. Given the rising prevalence of antibiotic resistant bacteria, pressure to develop

viable sortase inhibitors as therapeutics is growing and will undoubtedly lead to the discovery and

characterization of more potent and specific compounds.

2.9 Figures
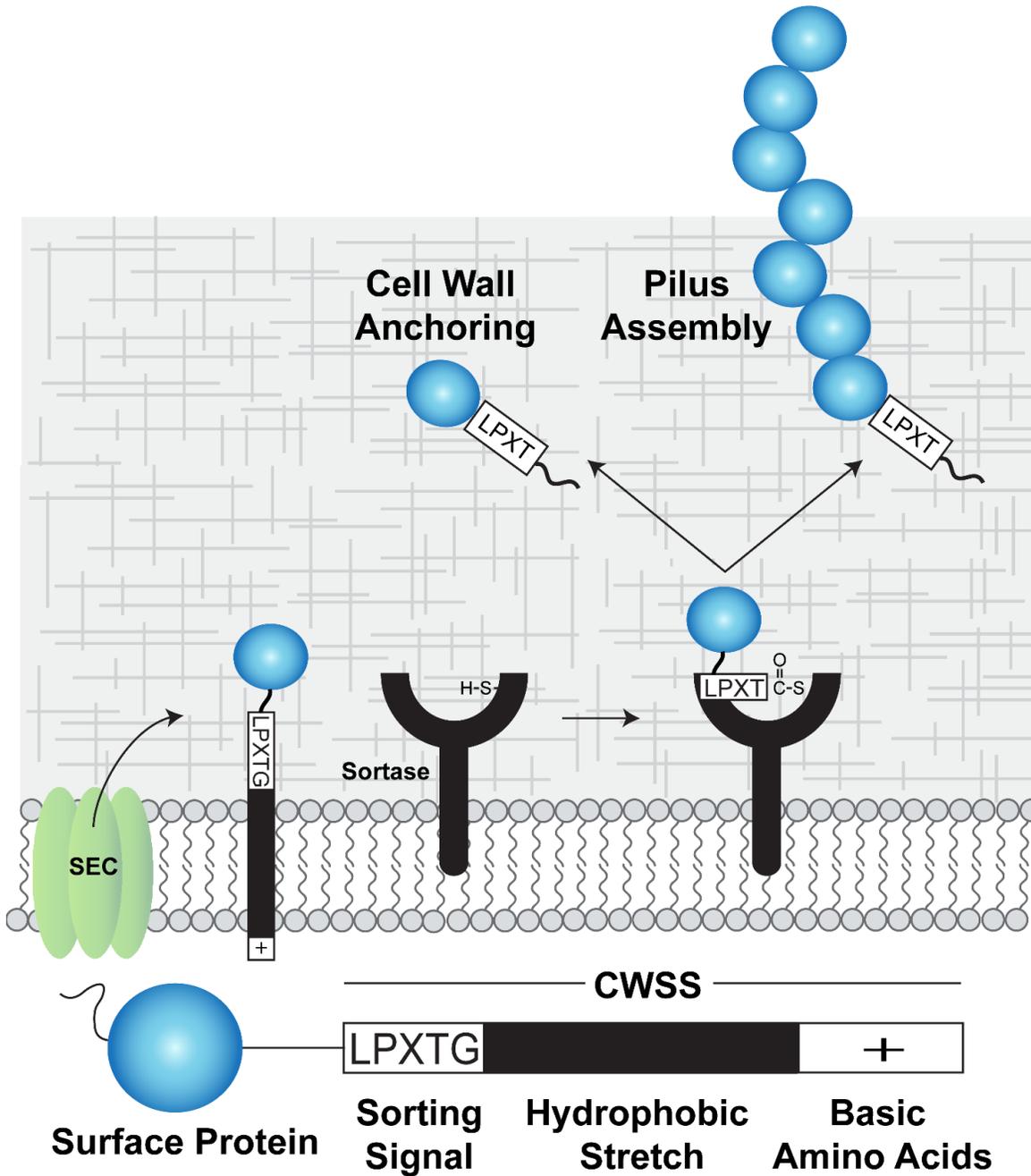
Figure 2.9.1 Sortase enzymes attach proteins to the cell wall and assemble pili.

Overview of anchoring and pilus assembly reactions. A protein that is to be displayed (blue)

contains an N-terminal secretion signal and a C-terminal cell wall sorting signal (CWSS). The CWSS contains an LPXTG-like sorting signal sequence that is processed by the sortase, a non-polar polypeptide segment (black), and a C-terminal segment of positively charged residues (+). After secretion through the Sec translocon, the protein remains embedded in the lipid bilayer via the non-polar segment within the CWSS. The sortase enzyme then cleaves between the threonine and glycine residues to form a sortase-protein thioacyl intermediate in which the active site cysteine is covalently linked to the carbonyl carbon atom of the threonine. There are two basic types of sortases: 1) cell wall anchoring sortases that attach protein to the cross-bridge peptide of the cell wall, and 2) pilin polymerase sortases that covalently link pilin subunits together via lysine isopeptide bonds. In both cases, the enzymes function as transpeptidases. Some sortases are capable of performing both functions, attaching proteins to the cell wall and polymerizing pili.
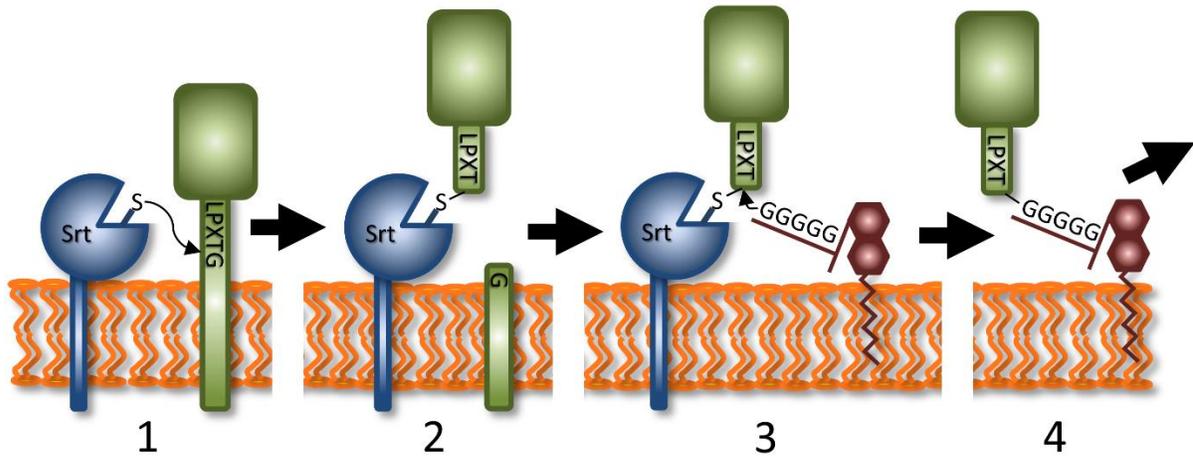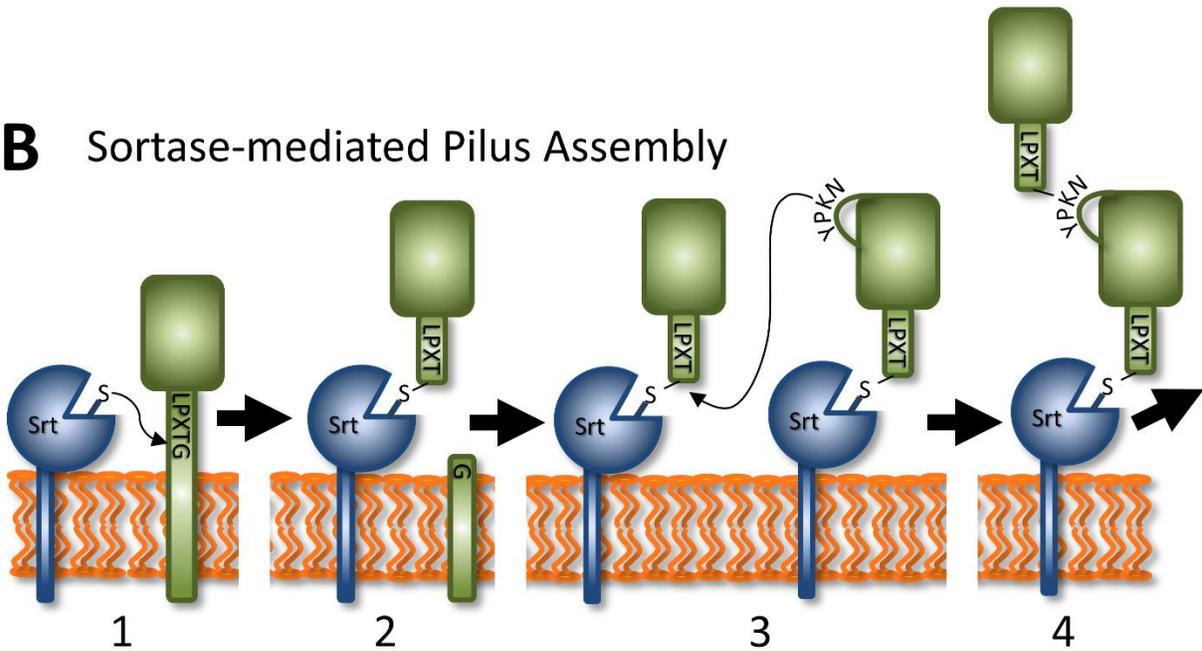
Figure 2.9.2 Mechanism of cell wall protein anchoring and pilus assembly.

Sortases perform two basic functions in bacteria: 1) attach proteins to the cell wall, and 2) join

proteins together to construct pili. (A) In the cell wall anchoring reaction, the sortase and

substrate are both membrane bound. The reaction occurs via four distinct steps. Sortase first

recognizes a sorting signal motif within the CWSS and nucleophilically attacks the threonine's

carbonyl carbon atom via its active site cysteine residue (for demonstration purposes the LPXTG sorting signal recognized by class-A type enzymes is shown, *step 1*). The LPXTG sorting signal is then cleaved to produce sortase-substrate thioacyl intermediate (*step 2*). Next, the cross-bridge peptide from a lipid II molecule nucleophilically attacks the thioacyl intermediate (*step 3*). Lastly, a new peptide bond is formed between the lipid II molecule and surface protein to produce a protein-lipid II intermediate that is incorporated into cell wall by the transglycosylation and transpeptidation reactions that synthesize the peptidoglycan (*step 4*). (B) In the pilus assembly reaction, steps 1-2 produce a sortase-substrate thioacyl intermediate, similar to the cell wall anchoring reaction. In this reaction, the sortase recognizes a pilin protein that contains a CWSS. However, a lysine residue within the pilin motif from an adjacent pilin protein performs the nucleophilic attack on thioacyl intermediate (*step 3*). A new protein-protein isopeptide bond is formed that covalently links the pilin proteins (*step 4*). This assembly process is repeated to build an isopeptide-linked pilus shaft that contains multiple pilin proteins. Depending on the type of pilus, distinct tip and base pilin proteins can be located at the ends of the pilus shaft, which are incorporated through a similar mechanism and involve covalent linkages via lysine-derived isopeptide bonds. Finally, the intact pilus is attached to the cell wall via sortase-catalyzed attachment of the pilus to lipid II, similar to cell wall protein display. Some sortases are capable of performing both functions, attaching proteins to the cell wall and functioning as pilin polymerases.

Figure 2.9.3 Structure and transpeptidation reaction of representative class A sortases. (A) *S. aureus* SrtA **(**SaSrtA) NMR structure (*cartoon*), showcasing an eight-stranded *β*-barrel with active site His120, Cys184 and Arg197 residues (*sticks*). (B) SaSrtA NMR structure (*green surface*) with active residues Arg **(***blue*) and Cys (*orange*). The active site His is occluded by a closed *β*7/*β*8 loop, and there is no obvious groove for a full length peptide to exit the active site. (C) *S. pyogenes* SrtA (SpySrtA) structure (*green surface*) with active site Arg **(***blue*), Cys (*orange*), and His (*cyan*) residues. An open *β*7/*β*8 loop creates a clear channel that can be seen running between active Cys and His residues, indicating the potential exit channel (*yellow*) for the full-length peptide substrate. (D) *In vitro*, SaSrtA catalyzes a reversible transpeptidation reaction

(*top,* indicated by solid arrows) in which it joins LPXTG and (Gly)$_3$ peptides. In the absence of glycine oligopeptide, SaSrtA acts a protease and cleaves the LPXTG peptide between its threonine and glycine residues (*bottom*, indicated by a dashed arrow). In this spurious pathway, a water molecule, instead of lipid II, performs the second nucleophilic attack to cleave the thioacyl bond between sortase and substrate, thereby hydrolyzing the peptide. On the cell surface, hydrolysis is presumably undesirable, as proteolysis separates the protein from its membrane anchor, releasing it from the microbe. Transpeptidation occurs faster than the rate of proteolysis *in vitro*, making SaSrtA a valuable bioconjugation reagent (kcat = 0.28 ± 0.02 and 0.086 ± 0.015 s^{-1}, respectively). Although all sortases are thought to catalyze transpeptidation reactions on the cell surface, this activity has only been reconstituted *in vitro* for a few sortases in addition to SaSrtA (listed in Table 2).
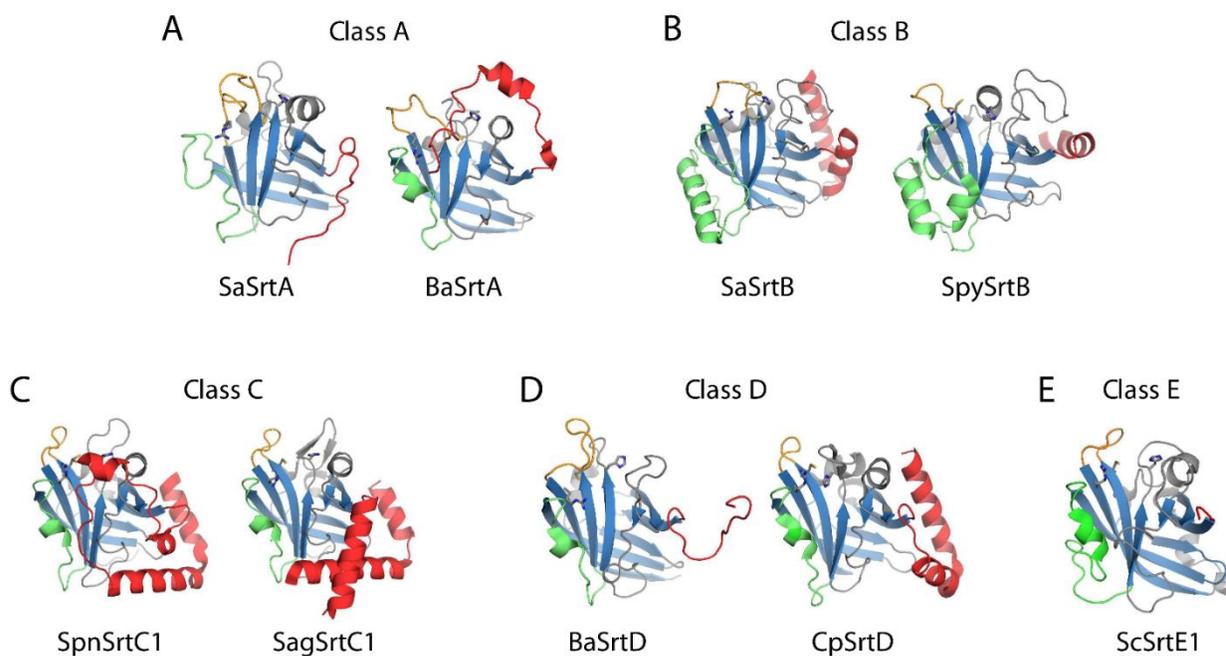
Figure 2.9.4 Structural variation by class of sortase. Sortases representative of the major themes seen for each class are displayed (*cartoon*) with active site residues (*sticks*). The hallmark sortase *β*-barrel (*blue*) and major sources of structural variability are highlighted, including N-terminus (*red*), *β*6/*β*7 loop (*orange*), and *β*7/*β*8 loop (*green*). Panels A to E show representative class A to E enzymes, respectively.
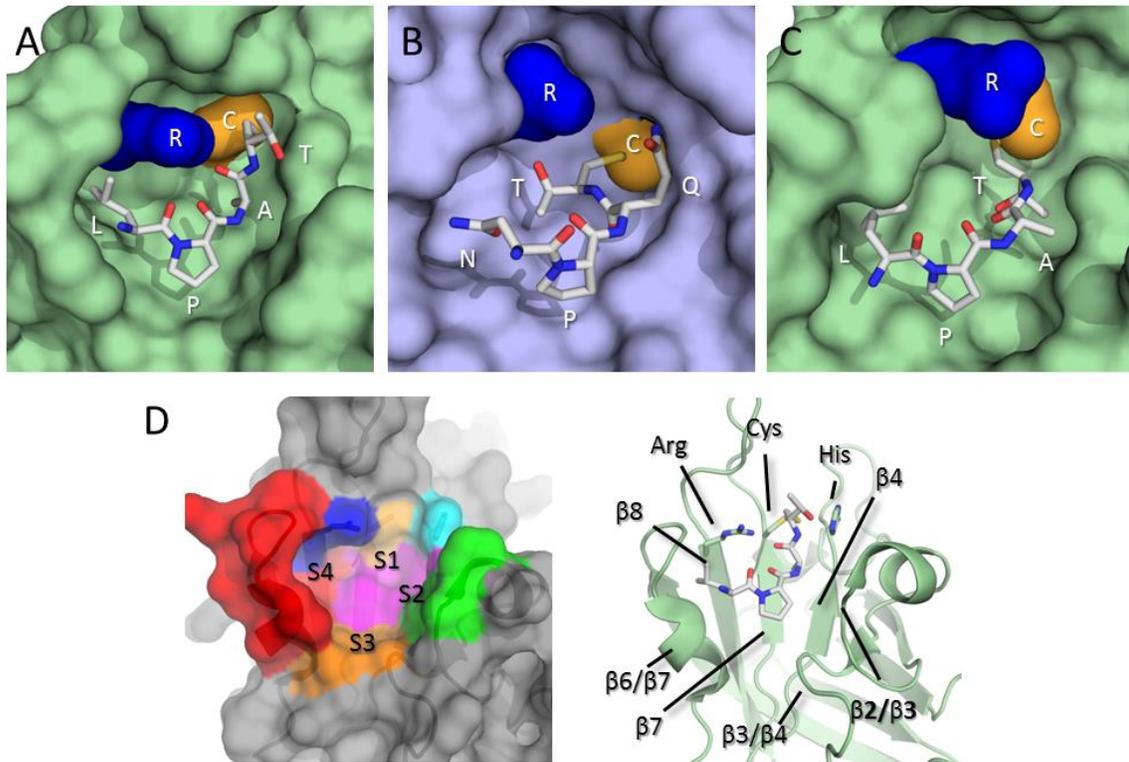
Figure 2.9.5 Sorting signal recognition. (A) The SaSrtA-LPAT* complex. (B) The SaSrtB-NPQT* complex. (C) The BaSrtA-LPAT* complex, shown with N-terminal appendage removed from view for clarity. Enzymes are shown as surface representations with SrtA types in *light green* and SaSrtB in *light blue*, Substrate mimics are shown as grey sticks. Active site Cys and Arg residues are shown as *gold* and *blue* surfaces, respectively. (D) Conserved recognition sites for sortase enzymes. Left, SaSrtA shown as a transparent surface representation with recognition subsites determined from the combination of sortase structures color coded as follows: S4 is shown in *red*, S3 in *orange*, S2 in *green*, and S1 in *magenta,* and active site Arg in *blue,* Cys in *gold*, and His in *cyan*. Right, Cartoon diagram of SaSrtA with secondary structure elements that contribute to substrate binding labeled for clarity.
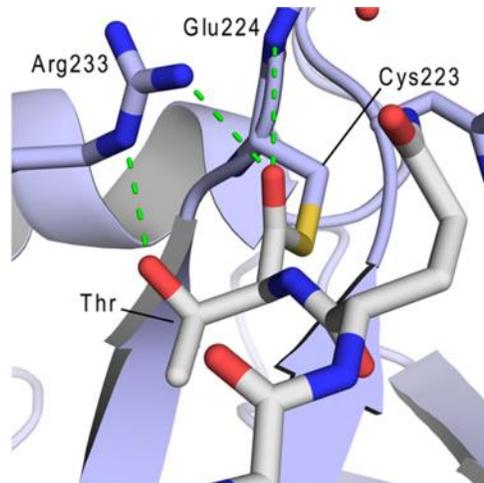
Figure 2.9.6 The substrate-stabilized oxyanion hole. The energy minimized model of the SaSrtB-NPQT thioacyl intermediate displayed with SaSrtB (*light blue cartoon*), residues in the active site and oxyanion hole (*sticks*), and NPQT substrate (*grey sticks*). The sidechain hydroxyl of the substrate's P1 Thr residue and backbone carbonyl participate in a hydrogen bonding network with the active site Arg, and the backbone amide of Glu224 that together build an oxyanion hole to stabilize the high energy tetrahedral reaction intermediates. Reproduced from Jacobitz *et* al. 2012.
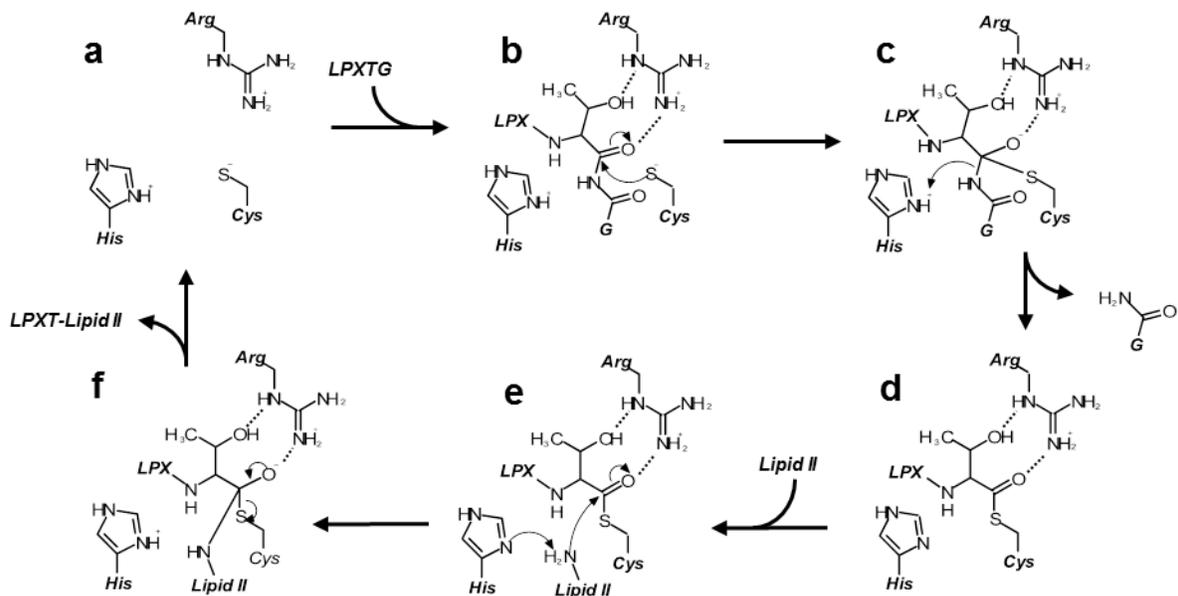
Figure 2.9.7 Molecular mechanism of sortase enzymes. The active site of sortase consists of a His-Cys-Arg triad, and in its active form, the His and Cys residues form a thiolate-imidazolium ion pair (a). The reaction begins with recognition of an appropriate sorting signal (here, the LPXTG sorting signal for SrtA types is shown), and the active site Cys residue performs nucleophilic attack on the carbonyl carbon at the substrate's P1 position (b). An oxyanion tetrahedral intermediate is stabilized by the nearby Arg residue that is likely oriented by interactions with the side chain of the substrate's P1 residue, which is a threonine in over 95% of all substrates (c). The active His residue concomitantly donates a proton to the leaving group, and the tetrahedral transition state then collapses to form a semi-stable, thioacyl intermediate between the substrate's P1 residue and the active site Cys (d). Next, the secondary substrate, (here, shown as lipid II used by cell wall anchoring sortases) enters the active site, where its terminal amine is

deprotonated by the active His residue before performing nucleophilic attack on the carbonyl carbon in the thioacyl bond (e); this second tetrahedral intermediate (f) collapses to form a peptide bond between the two substrates, and the product is finally released to leave the regenerated active site (a).

2.10 Tables

Table 2.10.1 Structurally characterized sortase enzymes.

Organism and Sortase	PDB	Bound Ligands or Substrates	Method
		Class A	
S. aureus SrtA (SaSrtA)	1IJA	N/A	NMR
S. aureus SrtA-C184A (SaSrtA)	1T2O	N/A	Xray
S. aureus SrtA (SaSrtA)	1T2P	N/A	Xray
S. aureus SrtA (SaSrtA)	1T2W	LPETG	Xray
S. aureus SrtA (SaSrtA)	2KID	Cbz-LPAT*, Ca^{2+}	NMR
S. aureus SrtA (SaSrtA)	2MLM	benzo[d]isothiazol-3-one based inhibitor	NMR
B. anthracis SrtA (BaSrtA)	2KW8	N/A	NMR
B. anthracis SrtA (BaSrtA)	unpub	Cbz-LPAT*	NMR
S. pyogenes SrtA	3FN5	N/A	Xray
S. pyogenes SrtA	3FN6	Cys in suflphenic acid form	Xray

S. pyogenes SrtA	3FN7	N/A	Xray
S. agalactiae SrtA	3RCC	Zn^{2+}	Xray
S. pneumoniae SrtA	4O8L	N/A	Xray
S. pneumoniae SrtA-C207A	4O8T	N/A	Xray
S. mutans SrtA	4TQX	chalcone	Xray

Class B

S. aureus SrtB (SaSrtB)	1NG5	N/A	Xray
S. aureus SrtB (SaSrtB)	1QWZ	MTSET	Xray
S. aureus SrtB (SaSrtB)	1QX6	E-64	Xray
S. aureus SrtB (SaSrtB)	1QXA	Gly_3	Xray
S. aureus SrtB (SaSrtB)	4FLD	Cbz-NPQT*	Xray
B. anthracis SrtB	1RZ2	N/A	Xray
B. anthracis SrtB	2OQW	AAEK1	Xray
B. anthracis SrtB	2OQZ	AAEK2	Xray
S. pyogenes SrtB	3PSQ	Zn^{2+}, Cl^-	Xray
C. difficile SrtB	4UX7	N/A	Xray

Class C

A. oris SrtC-1	2XWG	Ca^{2+}	Xray
S. pneumoniae SrtC-1	2W1J	glycerol	X-ray
S. *pneumoniae* SrtC-3	2W1K	N/A	Xray
S. pneumoniae SrtC-1-H131D	2WTS	alanine	Xray
S. pneumoniae SrtC-2	3G66	N/A	X-ray
S. pneumoniae SrtC-2	3G69	SO_4^{2-}	Xray
S. agalactiae SrtC-1Pilus Island-2a	3O0P	N/A	Xray
S. agalactiae SrtC-1 Pilus Island-1 "Type III"	3RBI	N/A	Xray
S. agalactiae SrtC1- Pilus Island-1 C184A; KDPYS to IPNTG	3RBJ	N/A	Xray
S. agalactiae SrtC-1 Pilus Island-1 "Type II"	3RBK	N/A	Xray
S. agalactiae SrtC1 Pilus	3TB7	N/A	Xray

Island-1

"Type I" - open lid

S. agalactiae SrtC-1 Pilus Island-1	3TBE	MTSET	Xray
S. agalactiae SrtC-2 Pilus Island-1	4G1H	Ca^{2+}	Xray
S. agalactiae SrtC-1 Pilus Island-1	4G1J	N/A	Xray
S. suis SrtC-1	3RE9	N/A	Xray

Class D

B. anthracis SrtD	2LN7	N/A	NMR
C. perfringens SrtD	4D70	none	Xray

Class E

S. coelicolor SrtE-1	5CUW	N/A	Xray

Table 2.10.2. Activity of sortase enzymes *in vitro*.

Sortase	Primary Substrate (K_m mM)	Secondary Substrate (K_m μM)	Cleavage k_{cat} (s^{-1})	Transpeptidation k_{cat} (s^{-1})
		Class A		
SaSrtA$_{Δ24}$[70]	Abz-**LPETG**-Dap(Dnp) (7.33 ± 1.01)	Gly$_5$ (196 ± 64)	0.086 ± 0.015	0.28 ± 0.02
SaSrtA$_{Δ24}$[82]	Abz-**LPETG**-Dap(Dnp) (5.5 mM)	Gly$_5$ (140)	NR	0.27
SaSrtA$_{Δ24}$[88]	Abz-**LPETGG**-Dap(Dnp) (8.76 ± 0.78)[a]	Gly$_5$ (NR)	NR	1.10 ± 0.06[a]
SaSrtA$_{Δ59}$[83]	Abz-**LPETG**K-(Dnp) (7.6 ± 0.5)	Gly$_3$ (140 ± 30)	NR	1.5 ± 0.2
SaSrtA$_{Δ59}$ Evolved tetramutant[83]	Abz-**LPETG**K-(Dnp) (0.17 ± 0.03)	Gly$_3$ (4800 ± 700)	NR	4.8 ± 0.8
SpySrtA$_{Δ81}$[40]	Abz-**LPETGG**-Dap(Dnp) (0.83 ± 0.11)	Ala$_2$ (NR)	NR	0.0136 ± 0.0011

BaSrtA$_{\Delta 56}$[41]	Abz-**LPET**G-Dap(Dnp) (0.038 ± 4)†	m-DAP*	0.0004 ± 0.0001†	*
BaSrtA$_{\Delta 56}$[44]	Abz-**LPET**G-Dap(Dnp) (0.306 ± 0.023)	m-DAP*	3.6 ± 0.2 x10^{-5}	*
BaSrtA$_{\Delta 64}$[44]	Abz-**LPET**G-Dap(Dnp) (0.173 ± 0.011)	m-DAP*	5.7 ± 0.2 x 10^{-5}	*
SmutSrtA$_{\Delta 40}$[43]	Dabcyl-QA**LPETG**EE-Edans (0.0904 ± 0.0047)†	NR	Yes	NR

Class B

SaSrtB$_{\Delta 21}$[73]	Abz-KVE**NPQTN**AGT-Dap(DNP) (7.8 ± 2)	Gly$_5$ (NR)	NR	5.4 ± 0.5 x 10^{-4}
SaSrtB$_{\Delta 31}$[49]	SNKDKVEN**PQTN**AGT (1.8)	Gly$_5$ (NR)	NR	1.010^{-4}
BaSrtB$_{\Delta 37}$[89]	Abz -KTD**NPKTG**DEA-Dap(DNP)	NR	Yes	NR
CdSrtB$_{\Delta 26}$[90]	KIVK**SPKTG**DETQLMK KPPV**PPKTG**DSTTIGK	Gly$_{4/5}$ or Ala-D-Glu-DAP	NR	NR

SpnSrtC1$_{17-228}$[56]	**IPQTG** in RrgB$_{30-633}$	YPKN in RrgB$_{30-633}$	No	Yes
SagSrtC1$_{43-254}$ **from PI-2a**[29]	Dabcyl-KKVT**IPQTG**GIGT-Edans (0.0138)	NR	Yes	NR
SagSrtC1$_{42-305}$ **from PI-1**[59]	Dabcyl-RPPGV**FPKTG**GIG-Edans (0.01358 ± 0.00063)	NR	1.16 ± 0.044 x10^{-3}	NR
	Dabcyl-RPS**IPNTG**GIG-Edans (0.03100 ± 0.00462)	NR	1.77 ± 0.101 x10^{-3}	NR
	Dabcyl-RGGL**IPKTG**EQQ-Edans (0.01639 ± 0.00250)	NR	0.77 ± 0.038 x10^{-3}	NR
SagSrtC2$_{42-283}$ **from PI-1**[59]	Dabcyl-RPPGV**FPKTG**GIG-Edans	NR	1.04 ± 0.058 x10^{-3}	NR

	(0.006385 ± 0.00142)			
	Dabcyl-RGGL**IPKTG**EQQ-Edans (0.02733 ± 0.00435)	NR	4.36 ± 0.256 x10^{-4}	NR
	Dabcyl-RPS**IPNTG**GIG-Edans (0.05715 ± 0.00354)	NR	5.56 ± 0.174 x10^{-3}	NR

Class D

BaSrtD$_{\Delta 55}$[64]	VQGEK**LPNTA**SNN	m-DAP (NR)	Yes	NR
BaSrtD$_{\Delta 55}$[91]	Abz-GEK**LPNTA**SNN-Dnp	m-DAP (NR)	Yes	NR
CpSrtD$_{\Delta 23\text{-}187}$[65]	A$\beta_{1\text{-}16}$-**LPQTG**S	NR	Yes	NR

Sorting signals for all substrates are highlighted in bold
Errors are reported where published
"Yes" indicates the reaction was performed *in vitro* but kinetics parameters were not reported
NR – Not Reported
m-DAP – meso-diaminopimelic acid
[a] These values calculated assuming a hydrolytic shunt mechanism
† Values reported from fluorescence assay and subject to inner filter effect and are likely underestimates of true parameters
* The enzyme reportedly does not perform this reaction *in vitro*

2.11 References

1. Cascioferro S, Totsika M, Schillaci D. Sortase A: An ideal target for anti-virulence drug development. Microb Pathog. 2014;77: 105–112. doi:10.1016/j.micpath.2014.10.007

2. Schneewind O, Missiakas D. Sec-secretion and sortase-mediated anchoring of proteins in Gram-positive bacteria. Biochim Biophys Acta - Mol Cell Res. 2014;1843: 1687–1697. doi:10.1016/j.bbamcr.2013.11.009

3. Schneewind O, Missiakas DM. Protein secretion and surface display in Gram-positive bacteria. Philos Trans R Soc London B Biol Sci. 2012;367. Available: http://rstb.royalsocietypublishing.org/content/367/1592/1123.long

4. Spirig T, Weiner EM, Clubb RT. Sortase enzymes in Gram-positive bacteria. Mol Microbiol. 2011;82: 1044–59. doi:10.1111/j.1365-2958.2011.07887.x

5. Siegel SD, Liu J, Ton-That H. Biogenesis of the Gram-positive bacterial cell envelope. Curr Opin Microbiol. 2016;34: 31–37. doi:10.1016/j.mib.2016.07.015

6. Bradshaw WJ, Davies AH, Chambers CJ, Roberts AK, Shone CC, Acharya KR. Molecular features of the sortase enzyme family. FEBS J. 2015;282: 2097–2114. doi:10.1111/febs.13288

7. Clancy KW, Melvin JA, McCafferty DG. Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. Biopolymers. 2010;94: 385–396. doi:10.1002/bip.21472

8. Maresso AW, Schneewind O. Sortase as a target of anti-infective therapy. Pharmacol Rev. 2008;60: 128–141. doi:10.1124/pr.107.07110

9. Suree N, Jung ME, Clubb RT. Recent advances towards new anti-infective agents that inhibit cell surface protein anchoring in Staphylococcus aureus and other gram-positive

pathogens. Mini Rev Med Chem. 2007;7: 991–1000. doi:10.2174/138955707782110097

10.   Popp MW-L, Ploegh HL. Making and Breaking Peptide Bonds: Protein Engineering
       Using Sortase. Angew Chemie Int Ed. 2011;50: 5024–5032. doi:10.1002/anie.201008267

11.   Schmidt M, Toplak A, Quaedflieg PJ, Nuijens T. Enzyme-mediated ligation technologies
       for peptides and proteins. Curr Opin Chem Biol. 2017;38: 1–7.
       doi:10.1016/j.cbpa.2017.01.017

12.   Antos JM, Truttmann MC, Ploegh HL. Recent advances in sortase-catalyzed ligation
       methodology. Curr Opin Struct Biol. 2016;38: 111–118. doi:10.1016/j.sbi.2016.05.021

13.   Voloshchuk N, Liang D, Liang J. Sortase A Mediated Protein Modifications and Peptide
       Conjugations. Curr Drug Discov Technol. 2016;12: 205–213.
       doi:10.2174/1570163812666150903115601

14.   Schmohl L, Schwarzer D. Sortase-mediated ligations for the site-specific modification of
       proteins. Curr Opin Chem Biol. 2014;22: 122–128. doi:10.1016/j.cbpa.2014.09.020

15.   Ritzefeld M. Sortagging: a robust and efficient chemoenzymatic ligation strategy.
       Chemistry. 2014;20: 8516–29. doi:10.1002/chem.201402072

16.   Schneewind O, Model P, Fischetti VA. Sorting of protein A to the staphylococcal cell
       wall. Cell. 1992;70: 267–281. Available: http://www.ncbi.nlm.nih.gov/pubmed/1638631

17.   Ton-That H, Liu G, Mazmanian SK, Faull KF, 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. National
       Academy of Sciences; 1999;96: 12424–9. doi:10.1073/PNAS.96.22.12424

18.   Mazmanian SK, Liu G, Ton-That H, Schneewind O. Staphylococcus aureus sortase, an
       enzyme that anchors surface proteins to the cell wall. Science. 1999;285: 760–763.

Available: http://www.ncbi.nlm.nih.gov/pubmed/10427003

19. Connolly K, Clubb R. No Title. In: Waksman G, Caparon M, Hultgren C, editors. Structural biology of bacterial pathogenesis. Washington DC: ASM Press; 2005. pp. 101–127.

20. Ruzin A, Severin A, Ritacco F, Tabei K, Singh G, Bradford PA, et al. Further evidence that a cell wall precursor [C(55)-MurNAc-(peptide)-GlcNAc] serves as an acceptor in a sorting reaction. J Bacteriol. 2002;184: 2141–2147. Available: http://www.ncbi.nlm.nih.gov/pubmed/11914345

21. Perry AM, Ton-That H, Mazmanian SK, Schneewind O. Anchoring of surface proteins to the cell wall of Staphylococcus aureus. III. Lipid II is an in vivo peptidoglycan substrate for sortase-catalyzed surface protein anchoring. J Biol Chem. 2002;277: 16241–16248. doi:10.1074/jbc.M109194200

22. Schneewind O, Fowler A, Faull KF. Structure of the cell wall anchor of surface proteins in Staphylococcus aureus. Science. 1995;268: 103–106. Available: http://www.ncbi.nlm.nih.gov/pubmed/7701329

23. Ton-That H, Faull KF, Schneewind O. Anchor structure of staphylococcal surface proteins. A branched peptide that links the carboxyl terminus of proteins to the cell wall. J Biol Chem. 1997;272: 22285–22292. Available: http://www.ncbi.nlm.nih.gov/pubmed/9268378

24. Ton-That H, Schneewind O. Assembly of pili on the surface of Corynebacterium diphtheriae. Mol Microbiol. 2003;50: 1429–38. Available: http://www.ncbi.nlm.nih.gov/pubmed/14622427

25. Hendrickx APA, Budzik JM, Oh S-Y, Schneewind O. Architects at the bacterial surface -

sortases and the assembly of pili with isopeptide bonds. Nat Rev Microbiol. 2011;9: 166–176. doi:10.1038/nrmicro2520

26. Mandlik A, Swierczynski A, Das A, Ton-That H. Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. Trends Microbiol. 2008;16: 33–40. doi:10.1016/j.tim.2007.10.010

27. Kline KA, Dodson KW, Caparon MG, Hultgren SJ. A tale of two pili: assembly and function of pili in bacteria. Trends Microbiol. 2010;18: 224–32. doi:10.1016/j.tim.2010.03.002

28. Wu C, Mishra A, Reardon ME, Huang I-H, Counts SC, Das A, et al. Structural determinants of Actinomyces sortase SrtC2 required for membrane localization and assembly of type 2 fimbriae for interbacterial coaggregation and oral biofilm formation. J Bacteriol. 2012;194: 2531–2539. doi:10.1128/JB.00093-12

29. Cozzi R, Malito E, Nuccitelli A, D'Onofrio M, Martinelli M, Ferlenghi I, 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. 2011;25: 1874–1886. doi:10.1096/fj.10-174797

30. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44: D279-85. doi:10.1093/nar/gkv1344

31. Dramsi S, Trieu-Cuot P, Bierne H. Sorting sortases: a nomenclature proposal for the various sortases of Gram-positive bacteria. Res Microbiol. 2005;156: 289–97. doi:10.1016/j.resmic.2004.10.011

32. Ton-That H, Mazmanian SK, Alksne L, Schneewind O. Anchoring of surface proteins to

the cell wall of Staphylococcus aureus. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis. J Biol Chem. 2002;277: 7447–7452. doi:10.1074/jbc.M109945200

33.   Marraffini LA, Ton-That H, Zong Y, Narayana SVL, Schneewind O. Anchoring of surface proteins to the cell wall of Staphylococcus aureus. A conserved arginine residue is required for efficient catalysis of sortase A. J Biol Chem. 2004;279: 37763–37770. doi:10.1074/jbc.M405282200

34.   Ilangovan U, Ton-That H, Iwahara J, Schneewind O, Clubb RT. Structure of sortase, the transpeptidase that anchors proteins to the cell wall of Staphylococcus aureus. Proc Natl Acad Sci. 2001;98: 6056–6061. doi:10.1073/pnas.101064198

35.   Comfort D, Clubb RT. A Comparative Genome Analysis Identifies Distinct Sorting Pathways in Gram-Positive Bacteria. Infect Immun. 2004;72: 2710–2722. doi:10.1128/IAI.72.5.2710-2722.2004

36.   Pallen MJ, Lam AC, Antonio M, Dunbar K. An embarrassment of sortases – a richness of substrates? Trends Microbiol. 2001;9: 97–101. doi:10.1016/S0966-842X(01)01956-4

37.   Zong Y, Bice TW, Ton-That H, Schneewind O, Narayana SVL. Crystal structures of Staphylococcus aureus sortase A and its substrate complex. J Biol Chem. 2004;279: 31383–31389. doi:10.1074/jbc.M401374200

38.   Suree N, Liew CK, Villareal VA, Thieu W, Fadeev EA, Clemens JJ, et al. The Structure of the Staphylococcus aureus Sortase-Substrate Complex Reveals How the Universally Conserved LPXTG Sorting Signal Is Recognized. J Biol Chem. 2009;284: 24465–24477. doi:10.1074/jbc.M109.022624

39.   Naik MT, Suree N, Ilangovan U, Liew CK, Thieu W, Campbell DO, 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. 2006;281: 1817–1826. doi:10.1074/jbc.M506123200

40.   Race PR, Bentley ML, Melvin JA, Crow A, Hughes RK, Smith WD, et al. Crystal Structure of Streptococcus pyogenes Sortase A. J Biol Chem. 2009;284: 6924–6933. doi:10.1074/jbc.M805406200

41.   Weiner EM, Robson S, Marohn M, Clubb RT. The Sortase A enzyme that attaches proteins to the cell wall of Bacillus anthracis contains an unusual active site architecture. J Biol Chem. 2010;285: 23433–23443. doi:10.1074/jbc.M110.135434

42.   Khare B, Krishnan V, Rajashankar KR, I-Hsiu H, Xin M, Ton-That H, et al. Structural differences between the Streptococcus agalactiae housekeeping and pilus-specific sortases: SrtA and SrtC1. PLoS One. 2011;6: e22995. doi:10.1371/journal.pone.0022995

43.   Wallock-Richards DJ, Marles-Wright J, Clarke D, Maitra A, Dodds M, Hanley B, et al. Molecular Basis of Streptococcus mutans Sortase A Inhibition by the Flavonoid Natural Product trans-Chalcone. Chem Commun. The Royal Society of Chemistry; 2015; doi:10.1039/C5CC01816A

44.   Chan AH, Yi SW, Terwilliger AL, Maresso AW, Jung ME, Clubb RT. Structure of the Bacillus anthracis Sortase A Enzyme Bound to Its Sorting Signal: A FLEXIBLE AMINO-TERMINAL APPENDAGE MODULATES SUBSTRATE ACCESS. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2015;290: 25461–74. doi:10.1074/jbc.M115.670984

45.   Kang HJ, Coulibaly F, Proft T, Baker EN. Crystal Structure of Spy0129, a Streptococcus pyogenes Class B Sortase Involved in Pilus Assembly. PLoS One. 2011;6: e15969.

doi:10.1371/journal.pone.0015969

46.   Lazzarin M, Cozzi R, Malito E, Martinelli M, D'Onofrio M, Maione D, et al. Noncanonical sortase-mediated assembly of pilus type 2b in group B Streptococcus. FASEB J. Federation of American Societies for Experimental Biology; 2015;29: 4629–40. doi:10.1096/fj.15-272500

47.   Mazmanian SK, Ton-That H, Su K, Schneewind O. An iron-regulated sortase anchors a class of surface protein during Staphylococcus aureus pathogenesis. Proc Natl Acad Sci U S A. 2002;99: 2293–2298. doi:10.2307/3057958

48.   Zong Y, Mazmanian SK, Schneewind O, Narayana SV. The Structure of Sortase B, a Cysteine Transpeptidase that Tethers Surface Protein to the Staphylococcus aureus Cell Wall. Structure. 2004;12: 105–112. doi:10.1016/j.str.2003.11.021

49.   Jacobitz AW, Wereszczynski J, Yi SW, Amer BR, Huang GL, Nguyen A V., et al. Structural and computational studies of the Staphylococcus aureus Sortase B-substrate complex reveal a substrate-stabilized oxyanion hole. J Biol Chem. 2014; jbc.M113.509273. doi:10.1074/jbc.M113.509273

50.   Mazmanian SK, Skaar EP, Gaspar AH, Humayun M, Gornicki P, Jelenska J, et al. Passage of Heme-Iron Across the Envelope of Staphylococcus aureus. Science (80-). 2003;299: 906–909. doi:10.1126/science.1081147

51.   Marraffini LA, Schneewind O. Anchor Structure of Staphylococcal Surface Proteins V. ANCHOR STRUCTURE OF THE SORTASE B SUBSTRATE IsdC. J Biol Chem. 2005;280: 16263–16271. doi:10.1074/jbc.M500071200

52.   Zhang R, Wu R, Joachimiak G, Mazmanian SK, Missiakas DM, Gornicki P, et al. Structures of sortase B from Staphylococcus aureus and Bacillus anthracis reveal catalytic

amino acid triad in the active site. Structure. 2004;12: 1147–56. doi:10.1016/j.str.2004.06.001

53. Maresso AW, Wu R, Kern JW, Zhang R, Janik D, Missiakas DM, et al. Activation of inhibitors by sortase triggers irreversible modification of the active site. J Biol Chem. 2007;282: 23129–39. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3366505&tool=pmcentrez&rendertype=abstract

54. Chambers CJ, Roberts AK, Shone CC, Acharya KR. Structure and function of a Clostridium difficile sortase enzyme. Sci Rep. Nature Publishing Group; 2015;5: 9449. doi:10.1038/srep09449

55. Shaik MM, Lombardi C, Maragno Trindade D, Fenel D, Schoehn G, Di Guilmi AM, et al. A structural snapshot of type II pilus formation in Streptococcus pneumoniae. J Biol Chem. 2015; M115.647834-. Available: http://www.jbc.org/content/early/2015/07/21/jbc.M115.647834.short

56. Manzano C, Contreras-Martel C, El Mortaji L, Izoré T, Fenel D, Vernet T, et al. Sortase-mediated pilus fiber biogenesis in Streptococcus pneumoniae. Struct (London, Engl 1993). 2008;16: 1838–1848. doi:10.1016/j.str.2008.10.007

57. Manzano C, Izoré T, Job V, Di Guilmi AM, Dessen A. Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of gram-positive pathogens. Biochemistry. 2009;48: 10549–10557. doi:10.1021/bi901261y

58. Khare B, Fu Z-Q, Huang I-H, Ton-That H, Narayana SVL. The crystal structure analysis of group B Streptococcus sortase C1: a model for the "lid" movement upon substrate binding. J Mol Biol. 2011;414: 563–577. doi:10.1016/j.jmb.2011.10.017

59.   Cozzi R, Prigozhin D, Rosini R, Abate F, Bottomley MJ, Grandi G, et al. Structural Basis
       for Group B Streptococcus Pilus 1 Sortases C Regulation and Specificity. PLoS One.
       2012;7: e49048. doi:10.1371/journal.pone.0049048

60.   Persson K. Structure of the sortase AcSrtC-1 from Actinomyces oris. Acta Crystallogr D
       Biol Crystallogr. 2011;67: 212–7. doi:10.1107/S0907444911004215

61.   Neiers F, Madhurantakam C, Fälker S, Manzano C, Dessen A, Normark S, et al. Two
       crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis
       and substrate specificity. J Mol Biol. 2009;393: 704–716. doi:10.1016/j.jmb.2009.08.058

62.   Jacobitz AW, Naziga EB, Yi SW, McConnell SA, Peterson R, Jung ME, 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. American Chemical Society;
       2016;120: 8302–8312. doi:10.1021/acs.jpcb.6b01930

63.   Lu G, Qi J, Gao F, Yan J, Tang J, Gao GF. A novel "open-form" structure of sortaseC
       from Streptococcus suis. Proteins. 2011;79: 2764–9. doi:10.1002/prot.23093

64.   Robson SA, Jacobitz AW, Phillips ML, Clubb RT. Solution Structure of the Sortase
       Required for Efficient Production of Infectious Bacillus anthracis Spores. Biochemistry.
       2012;51: 7953–7963. doi:10.1021/bi300867t

65.   Suryadinata R, Seabrook SA, Adams TE, Nuttall SD, Peat TS, IUCr, et al. Structural and
       biochemical analyses of a *Clostridium perfringens* sortase D transpeptidase. Acta
       Crystallogr Sect D Biol Crystallogr. International Union of Crystallography; 2015;71:
       1505–1513. doi:10.1107/S1399004715009219

66.   Duong A, Capstick DS, Di Berardo C, Findlay KC, Hesketh A, Hong H-J, et al. Aerial
       development in Streptomyces coelicolor requires sortase activity. Mol Microbiol.

Blackwell Publishing Ltd; 2012;83: 992–1005. doi:10.1111/j.1365-2958.2012.07983.x

67. Kattke MD, Chan AH, Duong A, Sexton DL, Sawaya MR, Cascio D, et al. Crystal Structure of the Streptomyces coelicolor Sortase E1 Transpeptidase Provides Insight into the Binding Mode of the Novel Class E Sorting Signal. Ton-That H, editor. PLoS One. Public Library of Science; 2016;11: e0167763. doi:10.1371/journal.pone.0167763

68. Jung ME, Clemens JJ, Suree N, Liew CK, Pilpa R, Campbell DO, et al. Synthesis of (2R,3S) 3-amino-4-mercapto-2-butanol, a threonine analogue for covalent inhibition of sortases. Bioorg Med Chem Lett. 2005;15: 5076–5079. doi:10.1016/j.bmcl.2005.07.073

69. Liew CK, Smith BT, Pilpa R, Suree N, Ilangovan U, Connolly KM, et al. Localization and mutagenesis of the sorting signal binding site on sortase A from Staphylococcus aureus. FEBS Lett. 2004;571: 221–226. doi:10.1016/j.febslet.2004.06.070

70. Frankel BA, Kruger RG, Robinson DE, Kelleher NL, McCafferty DG. Staphylococcus aureus sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. Biochemistry. 2005;44: 11188–11200. doi:10.1021/bi050141j

71. Fischetti V., Pancholi V, Schneewind O. Conservation of a hexapeptide sequence in the anchor region of surface proteins from Gram-positive cocci. Mol Microbiol. Blackwell Publishing Ltd; 1990;4: 1603–1605. doi:10.1111/j.1365-2958.1990.tb02072.x

72. Kruger RG, Otvos B, Frankel BA, Bentley M, Dostal P, McCafferty DG. Analysis of the Substrate Specificity of the Staphylococcus aureus Sortase Transpeptidase SrtA†. Biochemistry. 2004;43: 1541–1551. doi:10.1021/bi035920j

73. Bentley ML, Gaweska H, Kielec JM, McCafferty DG. Engineering the substrate specificity of Staphylococcus aureus Sortase A. The beta6/beta7 loop from SrtB confers

NPQTN recognition to SrtA. J Biol Chem. 2007;282: 6571–6581.

doi:10.1074/jbc.M610519200

74. Huang X, Aulabaugh A, Ding W, Kapoor B, Alksne L, Tabei K, et al. Kinetic mechanism of Staphylococcus aureus sortase SrtA. Biochemistry. 2003;42: 11307–11315. doi:10.1021/bi034391g

75. Kappel K, Wereszczynski J, Clubb RT, McCammon JA. The binding mechanism, multiple binding modes, and allosteric regulation of Staphylococcus aureus Sortase A probed by molecular dynamics simulations. Protein Sci. 2012;21: 1858–71. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3575916&tool=pmcentrez&rendertype=abstract

76. 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. American Chemical Society; 2012;134: 7094–101. Available: http://dx.doi.org/10.1021/ja3008402

77. Biswas T, Pawale VS, Choudhury D, Roy RP. Sorting of LPXTG peptides by archetypal sortase A: role of invariant substrate residues in modulating the enzyme dynamics and conformational signature of a productive substrate. Biochemistry. American Chemical Society; 2014;53: 2515–24. Available: http://dx.doi.org/10.1021/bi4016023

78. Tian B-X, Eriksson LA. Catalytic Mechanism and Roles of Arg197 and Thr183 in the Staphylococcus aureus Sortase A Enzyme. J Phys Chem B. 2011;115: 13003–13011. doi:10.1021/jp2058113

79. Connolly KM, Smith BT, Pilpa R, Ilangovan U, Jung ME, Clubb RT. Sortase from Staphylococcus aureus Does Not Contain a Thiolate-Imidazolium Ion Pair in Its Active

Site. J Biol Chem. 2003;278: 34061–34065. doi:10.1074/jbc.M305245200

80.   Aulabaugh A, Ding W, Kapoor B, Tabei K, Alksne L, Dushin R, et al. Development of an HPLC assay for Staphylococcus aureus sortase: evidence for the formation of the kinetically competent acyl enzyme intermediate. Anal Biochem. 2007;360: 14–22. doi:10.1016/j.ab.2006.10.021

81.   Ton-That, H., Mazmanian, S. K., Alksne, L., and Schneewind O. Anchoring of Surface Proteins to the Cell Wall of Staphylococcus aureus. SORTASE CATALYZED IN VITRO TRANSPEPTIDATION REACTION USING LPXTG PEPTIDE AND NH2-GLY3 SUBSTRATES. J Biol Chem. 2000;275: 9876–9881. doi:10.1074/jbc.275.13.9876

82.   Kruger RG, Dostal P, McCafferty DG. Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the Staphylococcus aureus sortase transpeptidase SrtA. Anal Biochem. 2004;326: 42–48. doi:10.1016/j.ab.2003.10.023

83.   Chen I, Dorr BM, Liu DR. A general strategy for the evolution of bond-forming enzymes using yeast display. Proc Natl Acad Sci U S A. 2011;108: 11399–404. doi:10.1073/pnas.1101046108

84.   Amer BR, Macdonald R, Jacobitz AW, Liauw B, Clubb RT. Rapid addition of unlabeled silent solubility tags to proteins using a new substrate-fused sortase reagent. J Biomol NMR. Springer Netherlands; 2016;64: 197–205. doi:10.1007/s10858-016-0019-z

85.   Wu Z, Guo Z. Sortase-Mediated Transpeptidation for Site-Specific Modification of Peptides, Glycopeptides, and Proteins. J Carbohydr Chem. 2012;31: 48–66. doi:10.1080/07328303.2011.635251

86.   Proft T. Sortase-mediated protein ligation: an emerging biotechnology tool for protein

modification and immobilisation. Biotechnol Lett. Springer Netherlands; 2010;32: 1–10. doi:10.1007/s10529-009-0116-0

87.   Tsukiji S, Nagamune T. Sortase-Mediated Ligation: A Gift from Gram-Positive Bacteria to Protein Engineering. ChemBioChem. 2009;10: 787–798. doi:10.1002/cbic.200800724

88.   Bentley ML, Lamb EC, McCafferty DG. Mutagenesis Studies of Substrate Recognition and Catalysis in the Sortase A Transpeptidase from Staphylococcus aureus. J Biol Chem. 2008;283: 14762–14771. doi:10.1074/jbc.M800974200

89.   Maresso AW, Chapa TJ, Schneewind O. Surface protein IsdC and sortase B are required for heme-iron scavenging of Bacillus anthracis. J Bacteriol. 2006;188: 8145–8152. doi:10.1128/JB.01011-06

90.   van Leeuwen HC, Klychnikov OI, Menks MAC, Kuijper EJ, Drijfhout JW, Hensbergen PJ. *Clostridium difficile* sortase recognizes a (S/P)PXTG sequence motif and can accommodate diaminopimelic acid as a substrate for transpeptidation. FEBS Lett. 2014;588: 4325–4333. doi:10.1016/j.febslet.2014.09.041

91.   Marraffini L a, Schneewind O. Targeting proteins to the cell wall of sporulating Bacillus anthracis. Mol Microbiol. 2006;62: 1402–17. doi:10.1111/j.1365-2958.2006.05469.x

Chapter 3

Crystal Structure of the *Streptomyces coelicolor* Sortase E1 Transpeptidase Provides Insight into the Binding Mode of the Novel Class E Sorting Signal

3.1   Overview

Many species of Gram-positive bacteria use sortase transpeptidases to covalently affix proteins to their cell wall or to assemble pili. Sortase-displayed proteins perform critical and diverse functions for cell survival, including cell adhesion, nutrient acquisition, and morphological development, among others. Based on their amino acid sequences, there are at least six types of sortases (class A to F enzymes); however, class E enzymes have not been extensively studied. Class E sortases are used by soil and freshwater-dwelling Actinobacteria to display proteins that contain a non-canonical LAXTG sorting signal, which differs from 90% of known sorting signals by substitution of alanine for proline. Here we report the first crystal structure of a class E sortase, the 1.93 Å resolution structure of the SrtE1 enzyme from Streptomyces coelicolor. The active site is bound to a tripeptide, providing insight into the mechanism of substrate binding. SrtE1 possesses β3/β4 and β6/β7 active site loops that contact the LAXTG substrate and are structurally distinct from other classes. We propose that SrtE1 and other class E sortases employ a conserved tyrosine residue within their β3/β4 loop to recognize the amide nitrogen of alanine at position P3 of the sorting signal through a hydrogen bond, as seen here. Incapability of hydrogen-bonding with canonical proline-containing sorting signals likely contributes to class E substrate specificity. Furthermore, we demonstrate that surface anchoring of proteins involved in aerial hyphae formation requires an N-terminal segment in SrtE1 that is presumably positioned within the cytoplasm. Combined, our results reveal unique features within class E enzymes that enable them to recognize distinct sorting signals, and could facilitate the development of substrate-based inhibitors of this important enzyme family.

3.2 Introduction

Gram-positive bacteria productively interact with their environment via surface displayed proteins anchored by sortase enzymes [1–10]. These cysteine transpeptidases modulate the functionality of bacterial surfaces by affixing proteins that perform a variety of functions, including cell adhesion, nutrient acquisition, immune evasion, aerial hyphae development and sporulation, among others [3,4,9,11]. Understanding the mechanism of catalysis and substrate recognition is of prime interest, as small molecule sortase inhibitors could have potent, anti-infective properties against pathogenic microbes by preventing them from displaying virulence factors [12–17]. Moreover, sortase-mediated protein ligation is an emerging biotechnology tool to modify and immobilize proteins, and a greater understanding of how these enzymes recognize their substrates could facilitate their rational engineering [18–27].

The catalytic mechanism of the *Staphylococcus aureus* sortase A enzyme (SaSrtA) has been characterized in detail and is paradigmatic [2]. SaSrtA, a class A enzyme, covalently anchors proteins to the cell wall by catalyzing a transpeptidation reaction that joins its protein substrate to the crossbridge peptide present in lipid II [28–30]. An N-terminal transmembrane (TM) segment positions SaSrtA at the cell membrane where it recognizes protein substrates via their C-terminal, cell wall sorting signal (CWSS). The CWSS consists of a LPXTG pentapeptide sorting signal motif (where X is any amino acid), followed by a hydrophobic segment that is embedded in the bilayer and a C-terminal cluster of positively-charged amino acids [28]. A conserved catalytic triad (His120, Cys184, Arg197) is required for transpeptidation activity in SaSrtA; this reaction is catalyzed through a ping-pong mechanism in which its active site cysteine residue nucleophilically attacks the backbone carbonyl carbon of the threonine residue within the LPXTG motif. Cleavage of the scissile T-G peptide bond forms a long-lived, sortase-protein thioacyl intermediate [31,32].

The thioacyl bond is then nucleophilically attacked by the amino group located in lipid II, creating a peptide bond-linked, protein-lipid II product [29,30,33]. The transpeptidation product is subsequently incorporated into the peptidoglycan via the conventional transglycosylation and transpeptidation reactions that synthesize the cell wall. All sortases are believed to catalyze transpeptidation reactions through a similar mechanism.

At present, over 1,800 gene sequences encoding sortase enzymes have been identified within ~600 species of bacteria [34]. Members of the sortase superfamily are predominantly found in Gram-positive bacteria and are grouped into distinct classes based on their amino acid sequences (class A to F enzymes) [3,35,36]. Biochemical and bioinformatics analyses suggest that class A, B, C, D and E enzymes have evolved specificities for LPXTG, NPXTN, LPXTG, LPXTA and LAXTG sorting signals, respectively (differences from LPXTG underlined). Most microbes express more than one type of sortase, which function non-redundantly to "sort" distinct proteins to the cell surface by recognizing their class-specific sorting signals. At present, atomic structures of class A, B, C, and D enzymes have been reported, revealing class-specific structural features [6]. Several studies using substrate analogues have also revealed how class A and B enzymes recognize their sorting signals [37–40]. However, the structure of a class E enzyme, or the mechanism though which it recognizes the unique LAXTG sorting signal substrate is not known.

Class E sortases are prevalent in soil-dwelling and aquatic actinobacteria (e.g. *Corynebacterium* and *Streptomyces* genera) [36]. *Streptomyces coelicolor* is one of the best-studied members of the Actinobacteria and uses two Class E enzymes to decorate its surface [11,41]. It exhibits a complex life cycle that has three morphologically distinct stages: vegetative hyphae, aerial hyphae, and spores. *S. coelicolor* is predicted to encode an astounding seven sortase enzymes: two class E and five class F enzymes. The class E enzymes in *S. coelicolor* (called SrtE1

and SrtE2) anchor chaplin proteins to the cell surface (ChpA, ChpB and ChpC) that function to promote the transition from vegetative growth to aerial hyphae formation. Strong evidence supports the notion that they recognize an unusual LAXTG sorting signal, as Duong et al. showed that SrtE1 and SrtE2 selectively cleave LAXTG-containing peptides *in vitro* and that they attach the ChpC protein bearing this motif to the cell wall [11,42]. These recognition events play a critical role in the lifecycle of this microbe as a *srtE1srtE2-* double mutant is delayed in aerial hyphae formation, is unable to sporulate, and fails to display chaplins on its aerial surfaces. Here we report the first atomic structure of a class E sortase, SrtE1 from *S. coelicolor*. The crystal structure, combined with biochemical, computational and cellular studies, provides insight in the mechanism of LAXTG sorting signal recognition.

3.3 Results and discussion

Actinobacteria display proteins on their cell surface using class E sortase enzymes that are distinct from other classes of sortase and have not been structurally characterized. To gain insight into their function, we investigated the class E enzymes from *S. coelicolor*, SrtE1 and SrtE2. We explored the importance of a unique cytoplasmic region in SrtE1 function and conducted structure-function studies of its catalytic domain to learn how it recognizes the novel LAXTG sorting signal.

3.3.1 Regions within the cytoplasmic tail of SrtE1 are important for its function *in vivo*

SrtE1 and SrtE2 contain a conserved C-terminal sortase catalytic domain (CAT) that is connected by ~30 non-polar amino acids (N) to a putative transmembrane (TM) helix (Fig 1A-B). The sequence of SrtE1 also contains an N-terminal extension that presumably resides within the

bacterium's cytoplasm (CE in Fig 1A). SrtE1 homologues of other actinobacteria in the streptomycetes genus also contain the cytoplasmic extension, which varies in length from ~100-200 amino acids (*e.g. Actinobacteria bacterium* and *Actinospica acidophila*) (S1 Fig). Notably, the CE segment is absent in nearly all other types of sortase enzymes, including the prototypical SaSrtA enzyme. While the sequence of the N-terminal CE segment in SrtE1 and its homologues varies considerably, two regions contain amino acids with similar physiochemical properties (S1 Fig). In SrtE1, the first conserved region corresponds to a short segment that is enriched for acidic amino acids (D50-E60) (enclosed in red box in S1 Fig), while a second region comprises a ~10 residue segment that is enriched with basic amino acids (R82-R92) (enclosed in blue box in S1 Fig). In addition, the ~25 residues that immediately precede the TM helix are also well conserved.

To probe the importance of the SrtE1 N-terminal cytoplasmic extension, we employed a mutagenic strategy to investigate its function. We first removed the nucleotides specifying residues G14-K113 of the *S. coelicolor* SrtE1-encoding gene (Δ14-113), which encompassed both the conserved acidic and basic patches (Fig 1A). This *srtE1* deletion variant was introduced into a strain lacking both *srtE1* and *srtE2* to test its ability to complement the sortase mutant phenotype (wild type *srtE1* alone can restore aerial development and sporulation). Interestingly, loss of this region results in a failure to promote sporulation in the sortase mutant, suggesting that it contains segments important for SrtE1 function (Fig 1C). As the basic patch is the most conserved sequence within this region, we set out to determine whether it is a critical functional determinant. We deleted residues R82-R90 (including five positively charged residues) and introduced the resulting construct into our sortase mutant (Δ82-90) (Fig 1A). This construct effectively restores aerial development and sporulation, suggesting this positively charged region is not important for function (Fig 1C). Finally, we created a third deletion variant lacking residues D15-G79, which

included the acidic, negatively charged region (Δ15-79) (Fig 1A). This SrtE1 variant is unable to function in place of wild type SrtE1 when introduced into the sortase mutant strain. Instead, it produces a phenotype that is indistinguishable from that of a complete null mutant, suggesting that segments within the first ~80 amino acids are important for SrtE1 function (Fig 1C). While we cannot rule out the possibility that this region merely contributes to the stability of SrtE1 (we were unable to generate a functional, tagged SrtE1 fusion protein *in vivo* to follow by immunoblotting), we speculate that this cytoplasmic tail, and particularly the negatively charged region, may promote interaction with other proteins in the membrane or cytoplasm.

3.3.2 The crystal structure of SrtE1 reveals unique class E features

We determined the 1.93Å resolution crystal structure of the C-terminal region of SrtE1 (SrtE1$^{\Delta N}$, residues T162-S352), which contains the conserved extracellular catalytic domain and a ~30 amino acid N-terminal linker segment that connects this domain to the putative TM anchor. Crystallization of SrtE1$^{\Delta N}$ was challenging as the protein irreversibly precipitated at low concentrations (~7 mg/ml) within a few days of storage at 4 ℃. However, for one condition in our screen, diffracting crystals formed after ~3-4 weeks from a dense precipitate and proteinaceous skin within the hanging drop. Attempts to reproduce crystal growth for structural characterization of enzyme-substrate complexes proved unsuccessful. SrtE1$^{\Delta N}$ crystallized in the C222$_1$ space group, with a single protein molecule in the asymmetric unit. The structure was determined by molecular replacement, and is well-defined by continuous electron density for residues forming the catalytic domain (P193-V351) (Fig 2). However, no density was observed for the ~30 amino acids in the N-terminal linker segment that precedes the catalytic domain (T162-E192), except for

residues A168-A170 (described in detail below). Complete data collection and structural statistics are provided in Table 1.

SrtE1$^{\Delta N}$ is monomeric and adopts a classical sortase-like β-barrel fold, but contains unusual β3/β4 and β6/β7 loops, as well as several accessory helices (Fig 2A). The catalytic domain starts with strand β1 (residues G196-I202), followed by a short helix H1 (residues P203-D206). The chain then forms a turn such that residues in the following strand β2 (residues V207-E213) interact with the β1 strand in an antiparallel manner. A 13-amino acid loop with an ordered helix H2 (residues V220-G224) then leads into strand β3 (residues V226-H228), which lays parallel to strand β2. A 16-amino acid loop precedes strand β4 (residues G244-A249), which interacts in an antiparallel manner with strand β3. The conserved active site histidine residue, His251, immediately follows strand β4 and is located on a long segment intervening between β4 and β5 that contained helix H3 (residues Y261-L265). Strand β5 (residues P270-E274) then aligns with strand β1 in an antiparallel manner, followed by a short turn that directs strand β6 (residues K278-T291) towards the active site. An extended, 26-residue loop containing an ordered 3_{10} helix H4 (residues N295-D300) leads into strand β7 (residues R313-T321), which interacts in a parallel manner with strand β4. The conserved catalytic cysteine residue (C320) is positioned at the end of the β7 strand, followed by a loop that reverses the direction of the polypeptide chain to lead into strand β8 (residues Y328-P341). The β8 strand contains a conserved active site arginine residue (R329) and is positioned antiparallel with respect to strands β6 and β7. Two short helices, H5 (residues S343-G345) and H6 (residues P347-V351), complete the structure.

During protein structure refinement, positive difference density within the active site clearly outlined the shape of a tripeptide. The sequence Ala-Gln-Ala (AQA) fit the density well (Fig 2B). This sequence assignment is consistent with residues Ala168-Ala170 in the N-terminal

linker of SrtE1$^{\Delta N}$ that precede the catalytic domain. Ala168 in the tripeptide is positioned farthest from the active site near the β3/β4 and β6/β7 loops (Fig 2C). The remainder of the tripeptide projects towards the active site with the side chain of Gln169 contacting hydrophobic residues in the β6/β7 loop (P293 and V298) and β7 strand (I331), and the side chain of the C-terminal Ala170 pointing towards the β2/H2 loop where it engages in hydrophobic interactions with Ala249 located in strand β4. Two potential enzyme-peptide hydrogen bonds are possible: the active site Arg329 side chain may form a hydrogen bond to the backbone carbonyl oxygen of Glu169 (N-O distance 3.4 Å), and the side chain hydroxyl oxygen of Tyr229 located in the β3/β4 loop may form a hydrogen bond to the backbone amine of Ala170 (O-N distance 3.4 Å). Unlike the AQA tripeptide, electron density for residues Gln171-Glu192 that separated Ala168-Ala170 from the structured catalytic domain was not observed. The notion that these residues in the N-terminal linker are structurally disordered or undergo exchange between two or more distinct conformational states, consistent with a "lid" function observed for class C enzymes, is supported by a two dimensional (2D) ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum of SrtE1$^{\Delta N}$, which lacks cross peaks for ~30 residues (S2 Fig). However, more conclusive NMR evidence could not be obtained as the protein construct was unstable, making it difficult to obtain sequence specific resonance assignments. The elevated B-factors of the tripeptide suggest that the peptide may not be bound at full occupancy throughout the crystal, possibly due to incomplete proteolytic cleavage to produce the tripeptide, or constraints of crystal packing. However, the contacts observed appear chemically reasonable, suggesting a degree of biological relevance. Interestingly, the amino acid sequence of the tripeptide is present in many SrtE1 enzymes (S1 Fig), but not conserved in SrtE2 or other class E enzymes (Fig 1B).

SrtE1 contains two active site loops that are structurally distinct from class A, B, C, and D sortases. Structures of representative class A, B, C and D enzymes have been determined previously [6]. As shown in Fig 3A, a phylogenetic comparison to these enzymes reveals that the sequences of the class E sortases from *S. coelicolor*, SrtE1 and SrtE2, differ substantially. Furthermore, a detailed structural comparison reveals that the conformation of the β3/β4 and β6/β7 loops in SrtE1 are distinct (Fig 3B). Based on our previously determined structures of class A and B enzymes bound to their substrates, these loops in SrtE1 are presumably part of the binding site for the LAXTG sorting signal [37–39,44]. In Fig 3B, the structures of these loops in SrtE1 are compared to the corresponding loops of the class A and B enzymes from *S. aureus*, which we have structurally characterized bound to their sorting signals [37,38]. It is readily apparent that the β6/β7 (blue) and β3/β4 (green) loops in SrtE1 are positioned closer to one another, resulting in a more confined binding site for the sorting signal (Fig 3B). Notably, the β3/β4 loop in SrtE1 is slightly longer than in other sortase enzymes, and as described below, contains a conserved tyrosine residue that may enable class E enzymes to recognize LAXTG sorting signals. Furthermore, the β6/β7 loop in SrtE1 contains a 21 amino acid insertion relative to class A enzymes that immediately follows a 3_{10} helix positioned adjacent to the active site cysteine (yellow) (Fig 3B). This long insertion is similar in length to that observed in class B sortases, but is distinctly devoid of secondary structure, whereas class B sortases contain an additional alpha helix.

3.3.3 A distinct subsite on SrtE1 enables it to accommodate an alanine residue at position P3 in the sorting signal

Previously, we demonstrated that SrtE1 attaches proteins to the cell wall that contain a novel LAXTG sorting signal [11]. All sortases characterized to date recognize protein sorting

signals that contain a conserved PXT motif at their core (i.e. LPXTG for class A enzymes). Following convention, residues in the sorting signal are numbered based on their positioning relative to the scissile Thr-Gly peptide bond, where residues in the sequence L-P-*X*-T-G sorting signal are referred to as P4-P3-P2-P1-P1', respectively (sites on the enzyme that recognize these residues are referred to as S4-S3-S2-S1-S1', respectively). The P3 site within the sorting signal is important for substrate recognition by other sortase enzymes, as biochemical studies have shown that mutation of this residue to alanine disrupts the ability of class A and B sortases to process their substrates [38,46]. Interestingly, the sorting signals recognized by SrtE1 are highly unusual as they contain an alanine at position P3, which contrasts ~90% of all known sorting signals that contain a proline. The distinct substrate specificity of SrtE1 appears to be a hallmark of class E enzymes, as comparative genome analyses predict that they also attach proteins to the cell that contain sorting signals with an alanine residue at position P3 [36]. To explore SrtE1's selectivity for alanine, its ability to cleave peptides containing the sequence LPETG or LAETG was determined. Each peptide was incubated with the enzyme separately, and the reaction products were separated by HPLC [42,47]. Interestingly, SrtE1 is capable of cleaving both peptides at the T-G peptide bond, as verified by MALDI (data not shown), but it exhibits a 2-fold preference for alanine at site P3 as compared to proline (Fig 4). This promiscuity is unique, as similar studies using the class A SaSrtA enzyme revealed that it could only hydrolyze the peptide containing a proline at position P3, whereas the alanine containing peptide was enzymatically inert (Fig 4) [46]. The selectivity of SaSrtA for proline is not surprising, as a structure of the enzyme bound to its sorting signal reveals that this residue enables the peptide to adopt an 'L'-shaped conformation that is complementary to the enzyme's active site [37,39].

To gain insight into SrtE1's unique ability to preferentially recognize sorting signals containing alanine at position P3, we computationally modeled how it binds to its LAETG and LPETG substrates. This work leveraged our experimentally determined structure of the class A sortase from *B. anthracis* (BaSrtA) bound to an analog of its sorting signal that contains a proline at position P3 (the BaSrtA-LPAT* complex) [39]. We modeled the SrtE1-substrate thioacyl complexes formed by these peptides, as it is a long-lived reaction intermediate that forms immediately after cleavage of the T-G peptide bond; in the reaction intermediate, the threonine carbonyl atom in the sorting signal was joined to the active site cysteine residue in SrtE1 via a thioacyl bond. Details of the modeling procedure are presented in the Methods section. Briefly, modeling involved positioning the peptide in the active site of SrtE1 using ligand docking and peptide coordinates derived from the structure of the BaSrtA-LPAT* complex, *in silico* construction of the thioacyl linkage, and solvated molecular dynamics (MD) and energy minimization calculations. Fig 5 shows the models of the SrtE1-LAET (Fig 5A) and SrtE1-LPET (Fig 5C) thioacyl intermediates. In both models, the bound sorting peptide adopts an 'L' shape as a result of a kink at the P3 residue. The non-polar side chains of either the alanine or proline residue at position P3 packs against Ala249 in the underlying β-sheet, as well as Ile215 and Leu221 within the β2-H2 loop. As observed in BaSrtA-LPAT (Fig 5B), the proline residue in the peptide of SrtE1-LPAT adopts a trans conformation producing an inherent kink that redirects the chain (Fig 5D), whereas in the LAET peptide, the kink occurs when the alanine residue adopts semi-favorable -57.9° phi and 173.6° psi torsional angles (Fig 5A). In both peptides, the kink causes the leucine P4 side chain to project into a hydrophobic S4 subsite on SrtE1 that is formed by residues located on the β7 (T318) and β8 strands (I331), as well as the N-terminal end of the β6/β7 loop (T291, P293, S294, N295, V296, V298, and L299). Specifically, the P4 leucyl side chain packs against the non-

polar side chains of Val296, Val298 and Leu299 in the 3_{10} helix H4, as well as the side chain of Thr291 and α-protons of Pro293, Ser294 and Asp295 in the β6/β7 loop. Not surprisingly, the overall conformation of the bound peptide and the positioning of the P3 and P4 residues in the SrtE1-peptide models are generally similar to the sorting signal positioning observed in the experimentally determined structures of the BaSrtA, SaSrtA and SaSrtB enzyme-substrate complexes [37–39]. However, in the case of the SaSrtB, a hydrophilic threonine residue (T177) is present within the β6/β7 loop to coordinate the polar asparagine at position P4 within its NPQTN sorting signal, whereas in the SrtE1-peptide models, the S4 subsite is non-polar so as to interact with the leucine side chain in the LAETG and LPETG sorting signals (Fig 5A and 5C).

The S3 subsite of SrtE1 contains a conserved tyrosine residue that may enable it to preferentially recognize alanine instead of proline at site P3 in the sorting signal. The tyrosine residue, Tyr229, is located in the β3/β4 loop and, along with the side chain of Ile215 in the β2/H2 loop, forms a unique SrtE1-specific ridge in the S3 subsite. In the energy minimized models of the SrtE1-LAET and SrtE1-LPET complexes, the S3 subsite can readily accommodate the methyl and pyrrolidine ring side chains of their respective sorting signals (Fig 5E and 5F, respectively). In particular, the S3 subsite forms hydrophobic contacts to these side chains via its Ile215, Leu221 and Ala249 residues, while the active site arginine residue (R329) donates a hydrogen to the acceptor carbonyl oxygen of the P3 residue via its guanidino group (Fig 5E-F), forming a hydrogen bond. These interactions have also been observed in the experimentally determined structures of sortases bound to their substrates [37–39]. Intriguingly, the tyrosine residue in the S3 subsite appears to preferentially stabilize the alanine-containing sorting signal. As shown in the energy minimized model of the SrtE1-LAET complex (Fig 5E), the backbone amide nitrogen of the P3 residue is positioned to donate a hydrogen to the acceptor Tyr229 hydroxyl, forming a hydrogen

bond. This hydrogen bond may preferentially stabilize binding to the alanine containing peptide, as a proline residue at this site would contain a nitrogen atom that would be unable to act as a hydrogen bond donor. Notably, the binding mode of the peptide in the SrtE1-LAET model is similar to that of the AQA tripeptide in the crystal structure of SrtE1, suggesting that it is a biologically accessible conformation (Fig 2C). Unfavorable enzyme-substrate steric interactions involving the tyrosine residue may also further discourage binding of sorting signals that contain a proline residue at P3. This is demonstrated in Fig 5B and 5D, where we compared the experimentally determined structure of the BaSrtA-LPAT* complex and a model of SrtE1-LPAT complex in which the peptide had been simply placed into the enzyme active site in an identical manner as in BaSrtA without any energy refinement (SrtE1-LPAT model). Unlike the LPAT peptide bound to BaSrtA, which is complementary to the enzyme's active site, the P3 proline residue in the non-energy minimized SrtE1-LPAT model sterically clashes with ridge atoms within the S3 subsite of SrtE1 (Fig 5D). These unfavorable contacts can only be alleviated by energy minimization of the atomic coordinates (Fig 5C). Interestingly, even though energy minimization enables the proline residue to properly fit into the S3 subsite, the P4 leucine side chain is not fully ensconced within the hydrophobic S4 subsite. Together, the modeling data suggests that steric hindrance and hydrogen bonding imparted by the tyrosine residue in the S3 subsite may cause SrtE1 to preferentially recognize sorting signals that contain an alanine residue at position P3.

Several indirect lines of evidence support the notion that class E enzymes like SrtE1 use a conserved tyrosine residue to recognize sorting signals that contain an alanine at position P3. First and foremost, an amino acid sequence alignment reveals that the tyrosine residue at this position is highly conserved in class E enzymes that are predicted to recognize LAXTG sorting signals, while it is frequently absent in other types of sortases [3,35,36] (S3 Fig). Second, several

108

biochemical studies of A, B and D sortases that do not contain the analogous tyrosine residue in their S3 subsites have revealed that they are unable to process signals containing alanine at site P3 [38,46,52]. Interestingly, these studies have shown that even conservative mutation of the tyrosine residue may disrupt alanine recognition. In particular, recent studies of the *Clostridium perfringens* SrtD enzyme (CpSrtD) that contains a phenylalanine (F92) instead of tyrosine residue have shown that it preferentially cleaves the sequence LPQTGS motif, but does not process an LAETG sorting signal [35,52]. Notably, CpSrtD was originally classified as a class E enzyme, but was later re-classified using hidden Markov models to be a class D enzyme [35,36]. This is consistent with our assertion that tyrosine plays an important role in signal recognition, as CpSrtD also lacks the class E specific tyrosine residue within its S3 subsite. Third, the results of directed evolution studies of the class A SaSrtA enzyme are compatible with the proposed substrate specificity determinant role of the conserved tyrosine residue in SrtE1. Specifically, Dorr et al. observed a marked shift in the specificity of evolved SaSrtA enzymes to preferentially cleave LAETG over LPETG substrates once a mutation was acquired within the β3/β4 loop at an analogous site to Y229 in SrtE1 (SaSrtA A104H mutation) [53]. These observations suggest that selectivity for an alanine at the S3 subsite could be mediated by the presence of a residue with a bulky, aromatic side chain and a hydrogen bond acceptor group within the β3/β4 loop; these features would partially exclude proline from the active site and stabilize the peptide backbone amide or carbonyl groups of substrates containing a residue with a small, nonpolar side chain. Unfortunately, our attempts to experimentally probe the dependence of substrate specificity on the tyrosine residue were unsuccessful, as SrtE1 proteins containing single amino acid mutations that change its Tyr229 residue to either phenylalanine or alanine are unstable (data not shown).

In conclusion, work presented in this paper has revealed unique class E sortase features within the catalytic domain of the SrtE1 enzyme from *S. coelicolor* and highlighted the functional importance of N-terminal cytoplasmic residues. Our biochemical studies indicate that SrtE1 can recognize sorting signals that contain either alanine or proline at position P3. Based on models of its reaction intermediates, we propose that SrtE1 and other class E sortases recognize unique alanine containing sorting signals by employing a conserved tyrosine residue within their β3/β4 loops. The tyrosine presumably biases recognition for alanine through a combination of steric effects and hydrogen bonding. However, it is important to stress that residues in addition to tyrosine may also be needed to confer enzyme specificity for alanine at site P3, as directed evolution studies of the SaSrtA enzyme have shown that a set of 11 mutations are required to change its specificity from LPXTG to LAXTG [53]. The majority of these mutations were in the peptide binding pocket and included an A104H alteration in the S3 subsite at a position that is analogous Tyr229 in SrtE1 [53]. Notably, guided by these studies, we attempted to bias the specificity of SaSrtA for alanine at site P3 by introducing a single A104Y mutation. However, this single amino acid mutant was unable to recognize both LPXTG and LAXTG substrates, suggesting that more than one mutation in the enzyme's binding pocket is required to change its substrate specificity (data not shown). Combined, the data suggest that binding of the sorting signal is a complex process, whose specificity is dictated by multiple, interdependent interactions with amino acids in the enzyme. Elucidating the determinants of sortase substrate specificity will require additional atomic structures of sortases bound to their substrates and the application of more sophisticated computational modeling approaches. The results of this work will facilitate prediction of sortase function among a wide range of microbes, rational design of substrate-based

inhibitors that could function as antibiotics, and engineering of sortases with altered specificities that could have useful biotechnological applications.

3.4 Materials and methods

3.4.1 Cloning, expression, and protein purification

The extracellular domain of SrtE1 from *S. coelicolor* (SrtE1$^{\Delta N}$, residues T162-S352) was expressed from a pET-15b plasmid in *Escherichia coli* BL21(DE3) cells. Standard methods were employed, with cultures grown in the presence of ampicillin at 37°C until an OD$_{600}$ of 0.6-0.8 was reached. Protein expression was then initiated by adding 100 µM isopropyl- β-D-1-thiogalactopyranoside (IPTG) followed by overnight protein expression at 25°C. A two liter cell culture was harvested by centrifugation and re-suspended in 40 mL of lysis buffer (50 mM Tris, pH 7.5; 300 mM NaCl) that contained 1 mg/mL lysozyme (Sigma-Aldrich), 400 µl of protease inhibitor cocktail (Calbiochem) and 2 mM phenylmethanesulfonylfluoride (PMSF). The cells were then incubated on ice with stirring for ~30 min and further lysed by sonication. Cell lysates were fractionated by centrifugation and the soluble portion applied to a gravity column containing 10 mL of suspended His-Pure Co^{2+} resin (Life Technologies) pre-equilibrated with lysis buffer. The resin was sequentially washed with 20 mL aliquots of lysis buffer that contained 0, 10, and 25 mM imidazole. His-tagged SrtE1 was then eluted using 500 mM imidazole, and the fractions pooled, concentrated, and buffer exchanged into 50 mM Tris, pH 7.5; 150 mM NaCl using an Amicon Ultra-15 centrifugal filter (Millipore). To remove His$_6$-tag from the protein, one unit of thrombin protease (GE Healthcare) was added for every 100 µg of SrtE1, and the solution was incubated at 4 °C overnight. Thrombin was then separated from SrtE1 using a HiTrap-Benzamidine column (GE Healthcare). Specifically, the SrtE1-thrombin mixture was loaded onto the column using

Buffer A (50 mM Tris, pH 7.5; 150 mM NaCl), followed by washing with Buffer A and subsequent addition of Buffer B (50 mM Tris, pH 7.5; 1 M NaCl) to recover absorbed SrtE1. SrtE1 lacking the His$_6$-tag was further purified by gel filtration chromatography using a Sephacryl size-exclusion column (GE Healthcare Life Sciences) equilibrated in 50 mM Tris, pH 7.5 and 150 mM NaCl. Purified SrtE1 was then pooled, concentrated to 7 mg/mL, and stored at 4 °C.

Three single amino acid mutants of SrtE1$^{\Delta N}$ were constructed, using the extracytoplasmic sequence as a starting point. Two point mutant variants targeting Y229 were generated using site-directed mutagenesis (mutated codon is underlined). Y229A mutation was generated by PCR amplification using forward (CGGGGCATGGTCGGGCAC<u>GCC</u>GCGGAGGACGGGCTGAAG) and reverse (CTTCAGCCCGTCCTCCGC<u>GGC</u>GTGCCCGACCATGCCCCG) primers. Y229F mutation was generated by PCR amplification using forward (CGGGGCATGGTCGGGCAC<u>TTC</u>GCGGAGGACGGGCTGAAG) and reverse (CTTCAGCCCGTCCTCCGC<u>GAA</u>GTGCCCGACCATGCCCCG) primers. An N-terminally truncated version was also generated, such that the first 34 extracytoplasmic amino acids were removed from the overexpression construct. This construct essentially recapitulated the portion of SrtE1 that was crystallized, and was generated by PCR amplification using the primers SrtE1 short NdeI (GGTCG<u>CATATG</u>GGCATCGGCTTCCTGCACG; NdeI site is underlined), and 3850 BamHI (GGGTGC<u>GGATCCT</u>TAACTGACGAGCGCATCC; BamHI site is underlined), using the original overexpression construct as template. The resulting PCR product was digested with NdeI and BamHI, cloned into pET15b digested with the same enzymes, and sequenced to confirm insert integrity. Protein overexpression was achieved as described above.

3.4.2 Crystallization, data collection, and structural determination

Recombinant SrtE1$^{\Delta N}$ at a concentration of 7 mg/mL in 50 mM Tris, pH 7.5; 150 mM NaCl was used for crystal screening. Screening used the Structure Screen broad matrix suite (Hampton Research) at room temperature in a sitting-drop vapor diffusion format (200 nl drop size). Protein crystals grew over the course of 3 to 4 weeks or longer in the presence of 50 mM HEPES, pH 7.5, 1.4 M sodium citrate and reached dimensions of ~ 0.10 mm x 0.05 mm. For X-ray data collection, SrtE1 crystals were cryoprotected using reservoir solution containing 30% glycerol. Diffraction data sets were collected at the Advanced Photon Source (APS) beamline 24-1D-C equipped with a Pilatus-6M detector. All data were collected at 100 K. Data were collected at the detector distance of 600 mm, with 0.5° oscillations, and at a 0.9791 Å wavelength. The crystals diffracted X-rays to 1.8 Å resolution. The XDS/XSCALE package was used to index, integrate and scale data in C222$_1$ space group [54]. The asymmetric unit of the crystal contained a single protein molecule, yielding a Matthews coefficient of 2.57 Å/Da and a 52.23% solvent content in the crystal. The PHASER program was used for molecular replacement, employing the coordinates of the SrtC1 enzyme from group B streptococcus as a search model (PDB ID: 4G1J); loops within the search model that had high B-factors were deleted [55]. Molecular replacement yielded a single solution, which was refined in iterative runs using PHENIX software [56]. Modeling of the additional active site density was confirmed using 2F$_o$-F$_c$ omit maps generated using PHENIX [56]. Complete refinement and structure statistics are reported in Table 1. The high value for the ligand B-factor indicated incomplete occupancy of a glycerol molecule within the crystal lattice. Coordinates and structure factors have been deposited in the Protein Data Bank with accession code 5CUW.

3.4.3 Computational modeling and molecular dynamics simulations

Models of the thioacyl intermediate containing SrtE1 bound to either LPET or LAET peptides were generated. The procedures used to construct these models have been described previously [38,39]. Briefly, the coordinates of each peptide were derived from the coordinates of the LPAT peptide in the NMR structure of the BaSrtA-LPAT complex [39]. The LPET or LAET coordinates were created by *in silico* mutation of the LPAT coordinates using Pymol. The LPET or LAET peptides were then docked to the crystal structure of SrtE1 using the Schrödinger Small-Molecule Drug Discovery Suite 2015-2 (Schrödinger LLC, New York, NY, USA). Prior to docking, the LPET and LAET ligands were constructed and energy minimized with LigPrep and the AQA tripeptide was removed from the crystal structure of the enzyme. The receptor (SrtE1) was also processed with the Protein Preparation Wizard to add missing side chain atoms and hydrogens, and to perform a restrained, partial energy minimization of the coordinates [57]. Docking grid was generated using Glide [49,50], and had dimensions of 22 x 22 x 22 Å centered around the active site. Docking was performed with Glide in SP Peptide docking mode using default settings [58]. A single model with the lowest docking energy for LPET and LAET peptides separately and the carboxyl group of Thr in the peptide within 3 Å of Cys320 were further refined using MD simulations.

The procedures used for the MD simulations have been described previously [38,39]. Briefly, the carboxyl oxygen on the Thr residue within the docked peptide was modified to enable thioacyl bond formation to the active site cysteine (C320). Parameters for the thioacyl linkage had been generated from our previous study [38,39]. Using tLeap, the models were first solvated in a triclinic box of TIP3P water molecules with sufficient sodium and chloride ions to create a neutral simulation box of approximately 150 mM NaCl [59]. Models were then energy-minimized and

equilibrated in NAMD 2.6 using the AMBER14SB force field by slowly removing restraints from the initial atom positions over 1 ns (2-fs step size) [38,51]. For each enzyme-peptide model, two sequential 10-ns MD simulations were performed.

3.4.4 *In vivo* **functional assay for SrtE1 activity**

To probe the functional significance of the cytoplasmic N-terminal extension of SrtE1, we created three different mutant variants (Δ14-113, Δ82-90 and Δ15-79), and tested their ability to complement the developmental defects of a *srtE1-srtE2* mutant. Two of the mutant constructs were synthesized (Δ14-113 and Δ15-79) (Genscript), and cloned into pUC57. These constructs were then excised using HindIII and KpnI, and were cloned into pMS82 digested with the same enzymes [60]. The resulting constructs were confirmed by sequencing, prior to being conjugated into the *S. coelicolor* Δ*srtE1/E2* mutant strain, where they were integrated into the chromosome [61]. The other mutant (Δ82-90) was generated by site-directed mutagenesis, using pMC134 (pMS82 containing *srtE1*) as template, together with the primers 3850 del4-F (<u>GGGGGCGGTCCGGACGGCGGCGGT</u>**CGGCGTCACGGGCGCCGTGGGGCG**) and 3850 del4-R (<u>CGCCCCACGGCGCCCGTGACGCCG</u>**ACCGCCGCCGTCCGGACCGCCCCC**), where each primer encompasses sequences flanking the deleted region (underlined and bolded sequences delineate the two sides of the deleted region) [61]. Following PCR amplification, the resulting product was treated with DpnI, and introduced into *E. coli* XL1-Blue cells by electroporation. The sequence of the deletion construct was confirmed by sequencing, after which the plasmid was introduced by conjugation into the double sortase mutant strain.   The ability of the different mutant derivatives was assessed using phenotypic analyses of the strains

grown on sporulation (MS) medium, comparing each strain with the *srtE1/E2* mutant strain containing either an empty plasmid vector (pMS82) or one bearing a wild type version of *srtE1* (pMC134) [62].

3.4.5 *In vitro* hydrolysis assay for substrate recognition and cleavage

The *in vitro* hydrolysis reactions were performed as described by Kruger et al. [48]. Ten micromolar sortase enzyme (wild-type or mutant) was incubated with 100 mM peptide substrate in 100 µl of assay buffer (50 mM Tris-HCl, 150 mM NaCl, 5 mM CaCl$_2$, 1 mM DTT) at 25 °C for 72 h. The reactions were quenched by adding 50 µl of 1 M HCl and injected onto a Waters XSelect HSS C18 reversed phase HPLC column. Peptides were eluted by applying a gradient from 0 to 40% acetonitrile (in 0.1% trifluoroacetic acid) over 40 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 mass spectrometry. The amount of product produced was determined by integrating the area of the product peak in the HPLC trace.

3.5 Figures
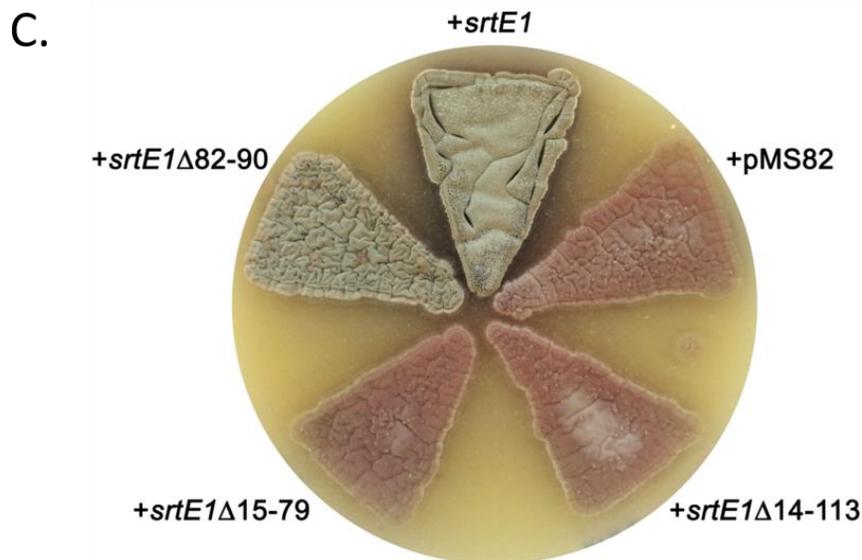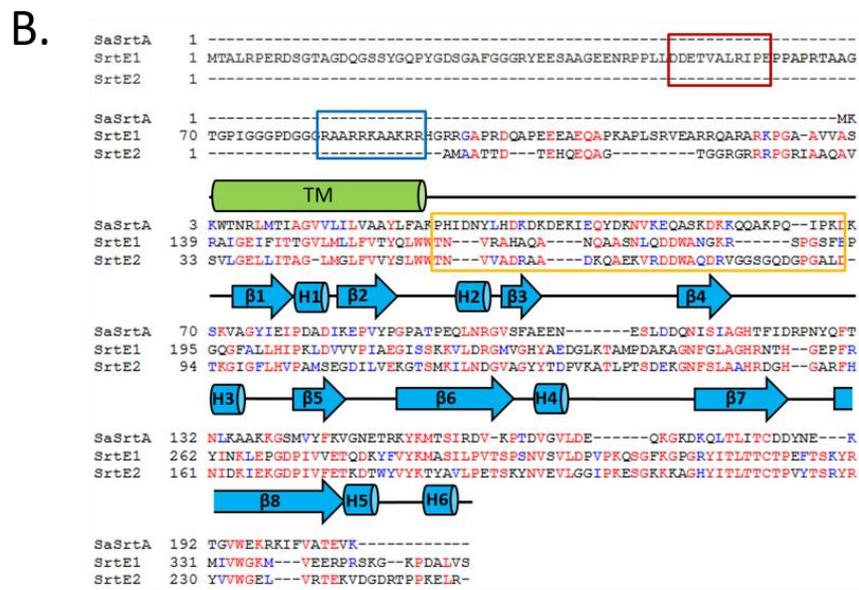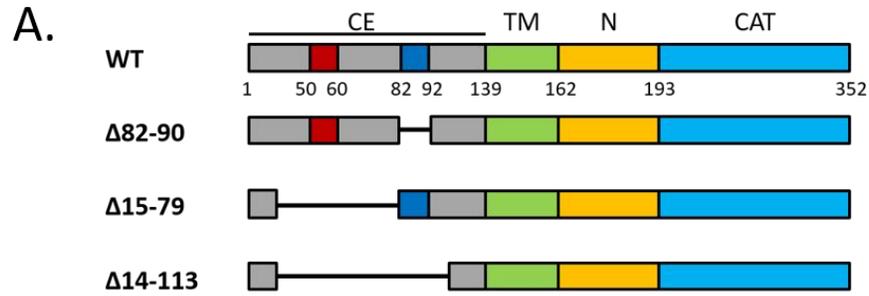
Figure 3.5.1 Phenotypic effect of deletions within the N-terminal cytoplasmic tail of SrtE1. A) SrtE1 constructs indicating deletion within the conserved, N-terminal cytoplasmic tail. The residue number that initiates or terminates deleted segments or domain boundaries are shown below the WT construct. The conserved acidic region (*red*) and basic region (*blue*) of the SrtE1 N-terminal cytoplasmic tail are indicated within the CE. WT, wild type; CE, cytoplasmic extension; TM, transmembrane helix; N, N-terminal membrane linker segment; CAT, catalytic core domain. B) Multiple sequence alignment of SrtE1 and SrtE2 from *S. coelicolor* with *S. aureus* SrtA. Sequence alignment was generated using the ClustalOmega server [43]. The bacterial species and accession numbers of the amino acid sequences used for the alignment are as follows: *Streptomyces coelicolor* (NP_628038 and NP_628037) and *Staphylococcus aureus* (WP_000759361). Conserved residues are indicated in *red*, and related amino acids are indicated in *blue*. Secondary structure elements from the SrtE1 crystal structure are shown in *blue* for β-strands (*arrows*) and helices (*cylinders*). The transmembrane (TM) region predicted by the THMHH server is indicated with a *green cylinder*. The conserved acidic region (*red box*) and basic region (*blue box*) within the N-terminal cytoplasmic tail of the SrtE1 enzyme, as well as conserved N-terminal membrane linker segment of SrtE1 and SrtE2 (*orange box*) are indicated. C) Effect of SrtE1 mutation on aerial development and sporulation in *S. coelicolor*. *srtE1* variants lacking the conserved acidic region (Δ15-79), basic region (Δ82-90), or both (Δ14-113) were introduced into a strain of *S. coelicolor* lacking both *srtE1* and *srtE2* (+pMS82). Wild type *srtE1* alone (+*srtE1*) can restore aerial development and sporulation.
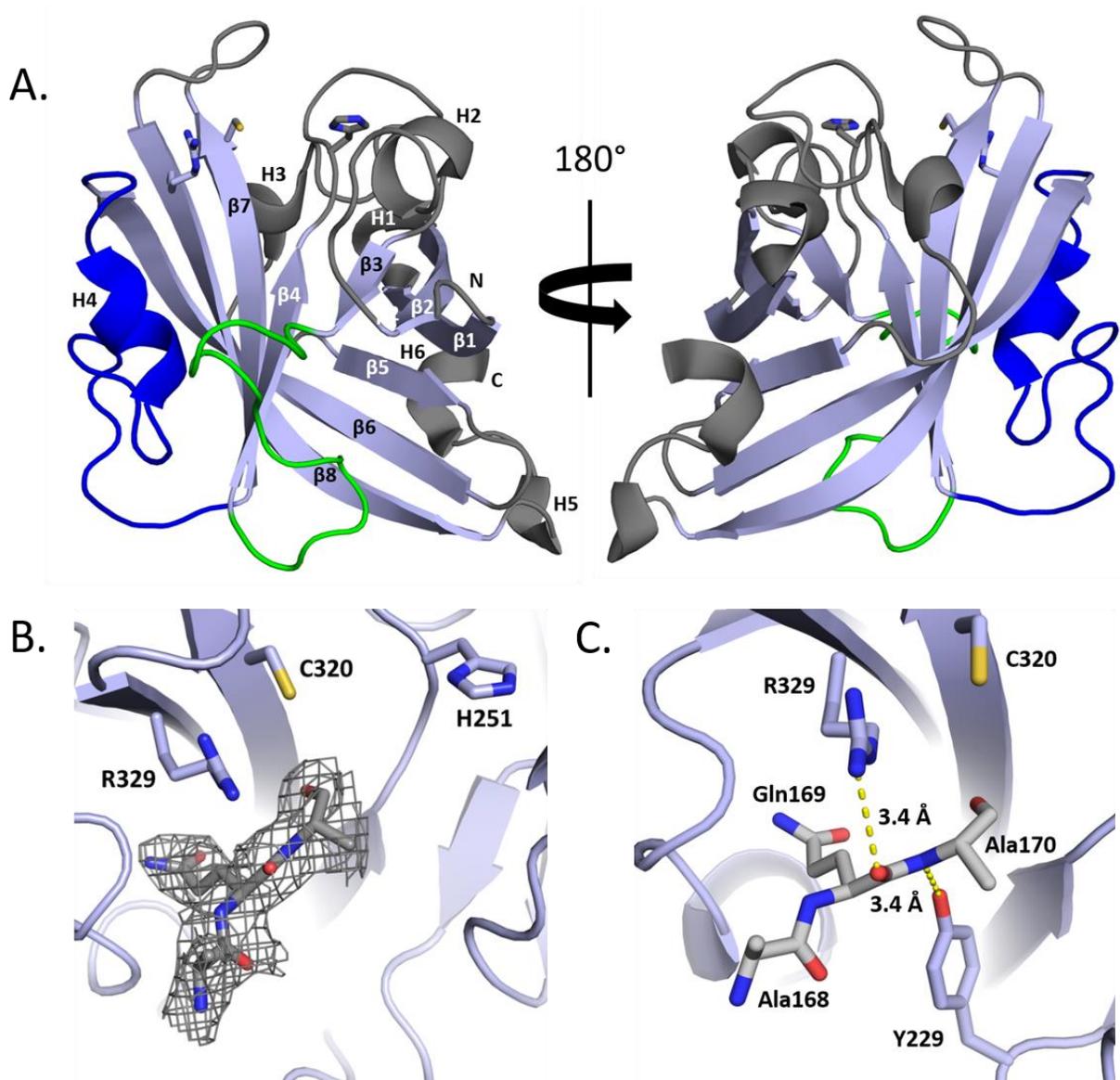
Figure 3.5.2 Crystal structure of SrtE1 from *S. coelicolor*. A) Ribbon diagram of

the SrtE1catalytic domain with the conserved Arg-Cys-His triad shown in sticks. Beta sheets (β)

and helices (H), as well as N- and C-termini, are labeled accordingly. The sortase β-barrel core

(*light blue*), structurally unique β3/β4 (*green*) and β6/β7 loops (*blue*), and accessory loops and

helices (*gray*) are colored. B) An $F_o – F_c$ omit map of the active site contoured at +3 σ (*green*

mesh). The map was generated by omitting the AQA tripeptide from the final model and

performing additional refinement. The omit density accommodates an AQA tripeptide adjacent

to Arg329 and Cys320 within the active site. C) Interactions between SrtE1 active site residues

and AQA tripeptide. Potential hydrogen bond interactions (*black dashed line*) between R329,

Y229, and AQA tripeptide (*sticks*) are shown.
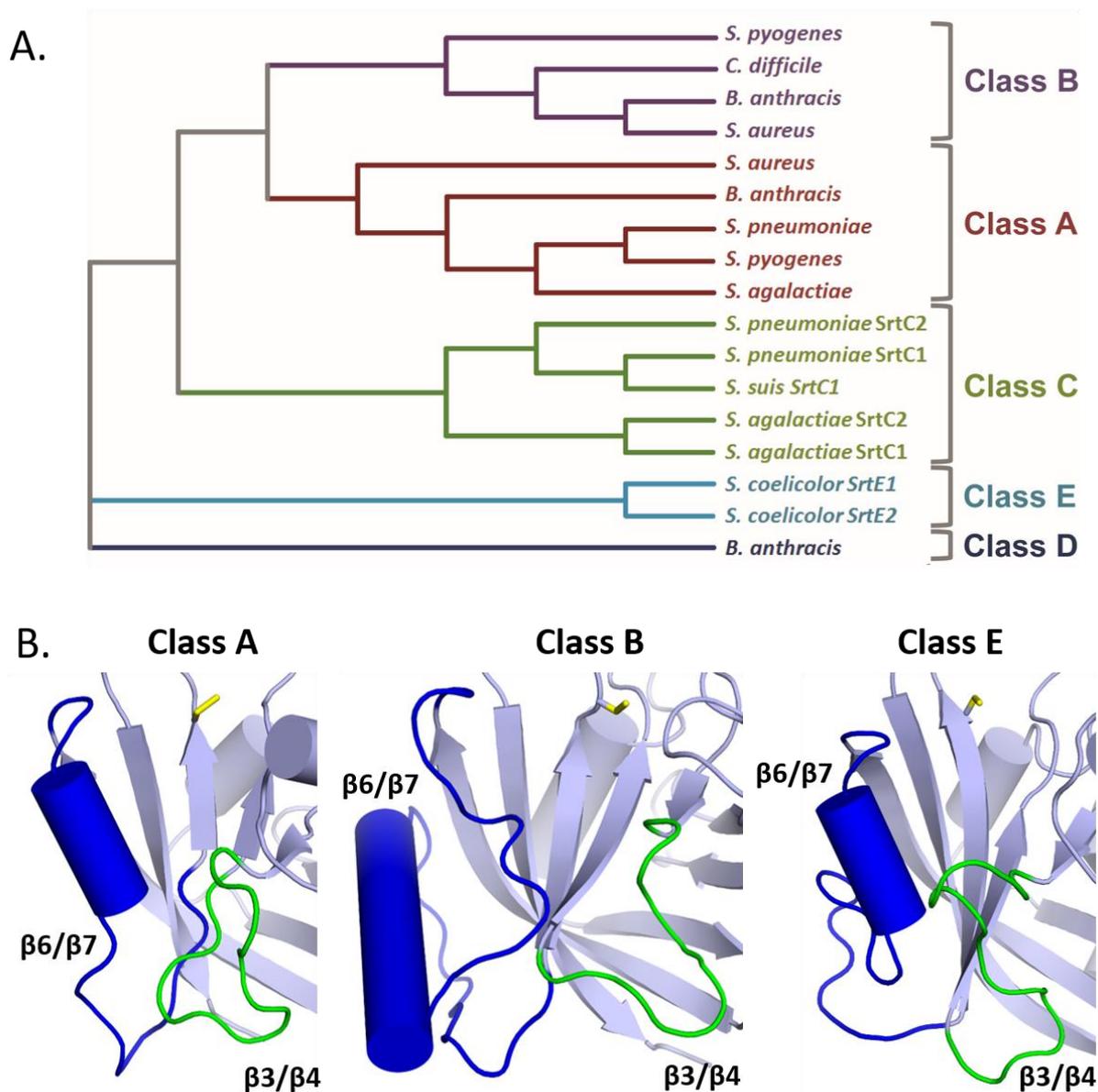
Figure 3.5.3 The structure of SrtE1 reveals unique features within class E

sortases. A) Phylogenetic tree of distinct sortase classes. The full amino acid sequences of 17

structurally characterized sortase enzymes were aligned with the sequences of SrtE1 and SrtE2

using the MUSCLE server and submitted to the ClustalOmega program for phylogenetic tree

generation via the neighbor joining method [43,45]. The bacterial species and accession numbers of the amino acid sequences used are as follows: *Streptomyces coelicolor* (NP_628038; NP_628037), *Bacillus anthracis* (WP_011732503; WP_000093563; WP_000771607), *Streptococcus pyogenes* (WP_002984641; WP_010921812), *Clostridioides difficile* (WP_021376017), *Staphylococcus aureus* (WP_000759361; WP_054104750), *Streptococcus agalactiae* (WP_017646311; WP_000529911; WP_000746226), *Streptococcus pneumoniae* (WP_000078846; WP_001140539; ABS82110), and *Streptococcus suis* (WP_012027975). B) Sortase class comparison of distinguishing class E structural features. Class A and class B enzymes from *S. aureus* and a class E enzyme (SrtE1) from *S. coelicolor* are shown in *cartoon* with helices (*cylinders*), β3/β4 loops (*green*), β6/β7 loops (*blue*), and active site cysteine residue (*yellow*) indicated. Inspection of the most structurally related class C (PDB ID: 3RE9) and class D (PDB ID: 2LN7) enzymes, as determined by DALI analysis, revealed β3/β4 and β6/β7 loops that align similarly to the class A enzyme.
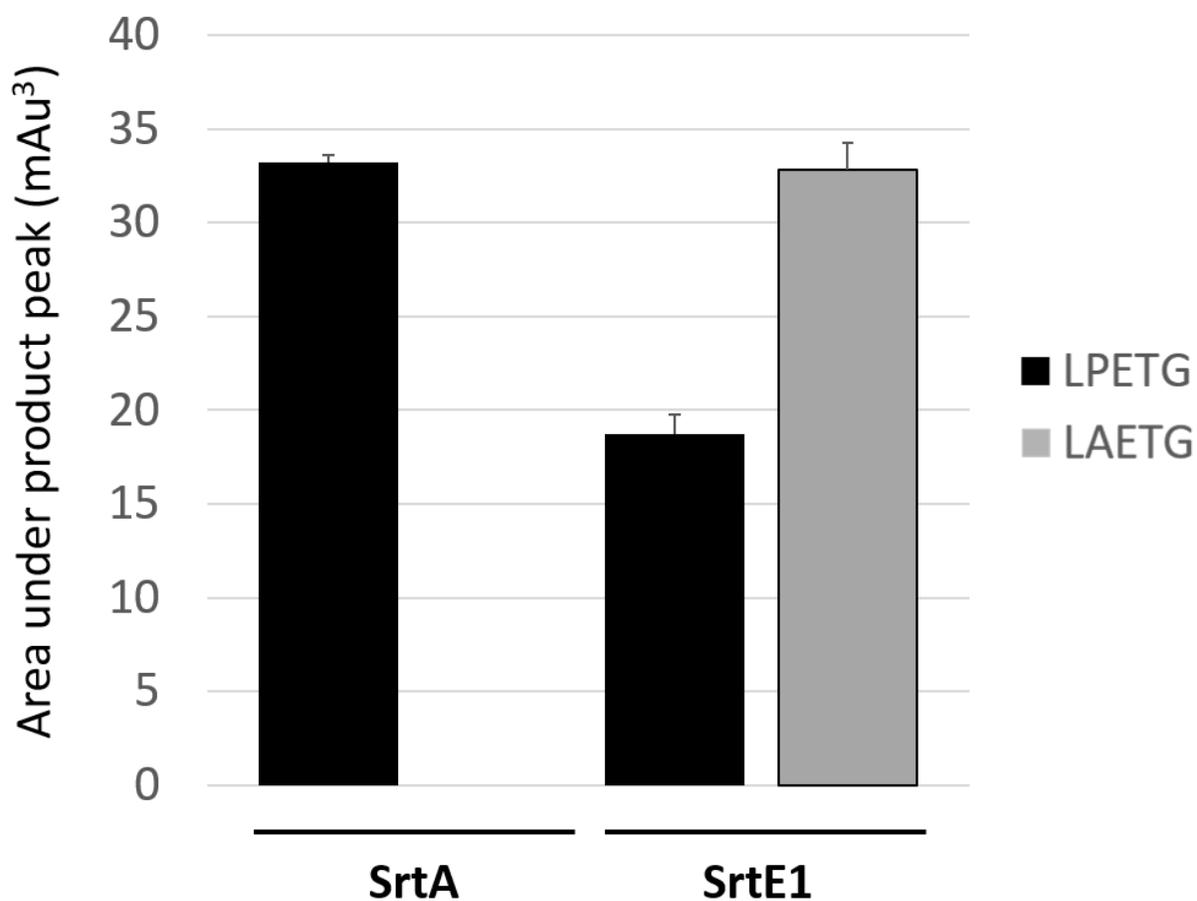
Figure 3.5.4 SrtE1 exhibits specificity for alanine at site P3 in the sorting

signal. Hydrolysis activity of SrtE1 or SaSrtA towards LPETG or LAETG peptide substrates

was determined with an established *in vitro* HPLC assay [48]. The enzyme cleaves the peptide at

the threonine-glycine scissile bond, producing N- and C-terminal peptide products. The extent of

hydrolysis was measured by integrating the area of the N-terminal peptide product peak in the

HPLC chromatogram. Error bars indicate the standard deviation of integrated HPLC peak area

obtained from duplicate hydrolysis reactions.

A. **SrtE1-LAET MD**
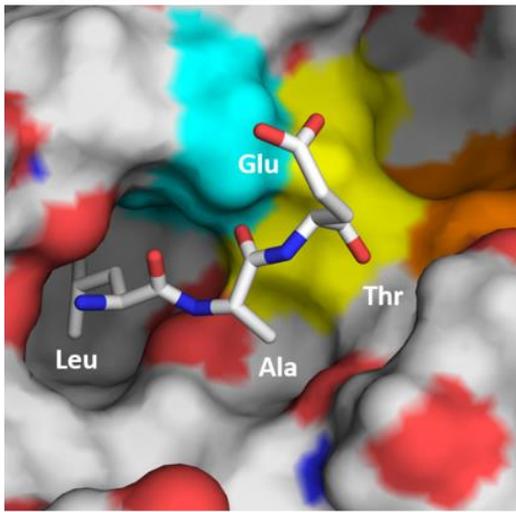
B. **BaSrtA-LPAT NMR**
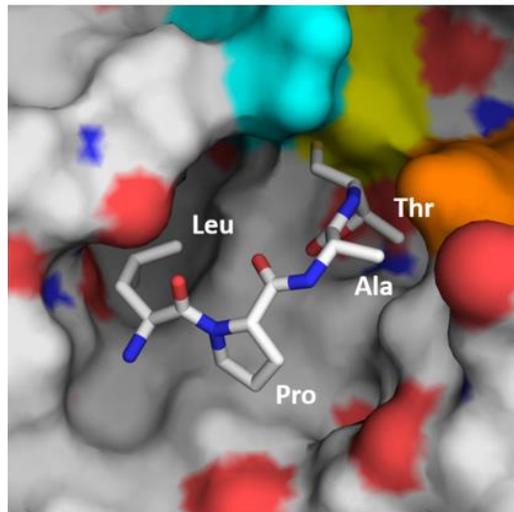
C. **SrtE1-LPET MD**
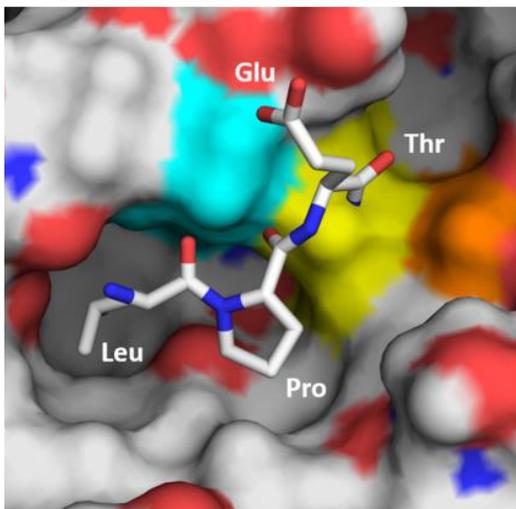
D. **SrtE1-LPAT Model**
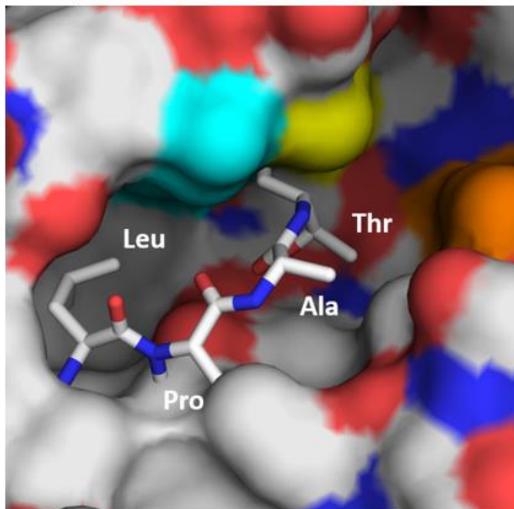
E.
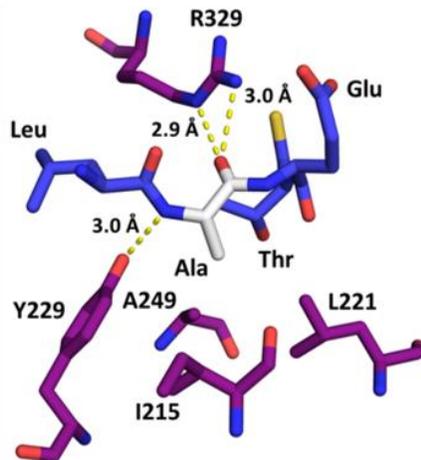
F.
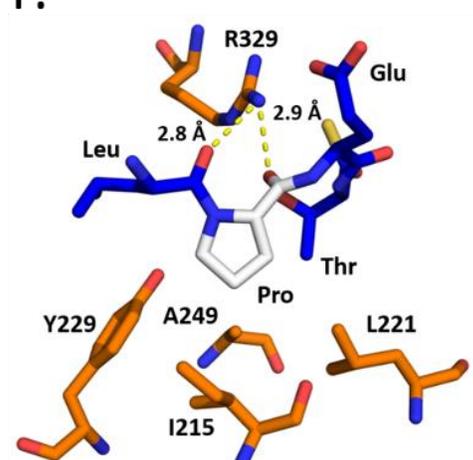
Figure 3.5.5 Energy minimized models of SrtE1-substrate complexes provide insight into the mechanism of recognition of the LAXTG sorting signal.

A) Model of SrtE1 binding the LAET motif. The LAET peptide (*gray sticks*) was docked to the SrtE1 active site (*electrostatic surface*) using GLIDE and energy minimized through molecular dynamics simulations with NAMD2 [49–51]. The exposed surfaces of the catalytic residues Arg329 (*cyan*), Cys320 (*yellow*), and His251 (*orange*) residues are shown. B) NMR solution structure of SrtA from *B. anthracis* bound to LPAT substrate. The LPAT substrate mimic (*white sticks*) is positioned within the BaSrtA active site groove (*electrostatic surface*), defined by the exposed surfaces of the catalytic residues Arg196 (*cyan*), Cys187 (*yellow*), and His126 (*orange*) residues. C) Model of SrtE1 binding the LPET motif. The LPET peptide (*gray sticks*) was docked to the SrtE1 active site (*electrostatic surface*) using GLIDE and energy minimized through molecular dynamics simulations with NAMD2. The exposed surfaces of the catalytic residues Arg329 (*cyan*), Cys320 (*yellow*), and His251 (*orange*) residues are shown. D) Model of SrtE1 binding the canonical SrtA substrate motif. The catalytic cores of the SrtE1 crystal structure and solution structure of SrtA-LPAT* from *B. anthracis* (BaSrtA) (PDB ID: 2RUI) were structurally aligned in Pymol. The LPAT substrate mimic from BaSrtA (*white sticks*) clashes with the SrtE1 active site groove (*electrostatic surface*), defined by the exposed surfaces of the catalytic residues Arg329 (*cyan*), Cys320 (*yellow*), and His251 (*orange*) residues. E) Hydrogen bond interactions between SrtE1 active site residues and LAET substrate motif. SrtE1 residues within 4 angstroms of the LAET peptide are shown (*magenta sticks*). Energy minimized LAET peptide (*gray sticks*) containing the unique alanine residue (*white sticks*) is indicated. Hydrogen bonds (*yellow dashed lines*) between R329, Y229, and the energy minimized LAET peptide backbone are shown. F) Hydrogen bond interactions between SrtE1 active site residues

and LPET substrate motif. SrtE1 residues are shown (*orange sticks*). Energy minimized LPET peptide (*gray sticks*) containing a proline residue (*white sticks*) is indicated. Hydrogen bonds (*yellow dashed lines*) between R329 and the energy minimized LPET peptide backbone are shown.

```
S.clavuligerus    1 ------------------------------------------------------------
S.coelicolor      1 ------------------------------------------------------------
S.viridochromog   1 ------------------------------MPGAFEDWPAGGEYGASPQPEAQPQPGSYAP
S.sviceus         1 ----------------------------MTALRPEREDLYGDASYESYGGDPYAGHSFGGASY
S.ghanaensis      1 ------------------------------------------MTALRPERESGAGYG
S.griseoflavus    1 -------------------MAGGVRRRGRGDEVGGARPGGGQVLGVTALRPERESGA---
S.scabiei         1 MTALRPERESAPYG-GEAAYGGAEAFEAETTFEPGVPFRDPAEPQQWAAPQASPYAQGDW
S.albus           1 MTALRPERESAAGGPVDPLTDPLPDVLAGGHSSPWFRAEQPPQLVQGHDLSAPQPPLADP


S.clavuligerus    1 ----------------------MHPPGVPPAPAPEQWG----TPRPPGQPEPPGPYEPP
S.coelicolor      1 -------MTALRPERDSGTAGDQGSSYGQPYGDSGAFGGGRYEESAAGEENRPPLLDDET
S.viridochromog  32 RPESYAPQSESYAPQPGSYVSQ----PEPYSASYEPLPDGSLPDESLPGESYLPPVDEET
S.sviceus        36 AEEPYAESPYGEEPYAESSYAEGQYTEQPPGGVSEGRASDGGASDRQASAPYVPPADDET
S.ghanaensis     16 ESFTDHPGYGVGGEQGTPYGQQPYGAPGTFGDGRYDGAPRDPYGGASDDTAYLPPVDEET
S.griseoflavus   39 ------DAYGGAGGQGTSYGGQPHGAPGAYGDGWYDTTAPGPHGGSSDTTAYLPPVDEET
S.scabiei        60 HGQQPYAEPAFGSYETYGQRPQAPYEGRPPASYDLYAAAGVQAGLPADTAQEHPPIDEET
S.albus          61 YAQLP-ADAYAAPPQDDAYAGHSPYEGGPGWTASPETYPAPEAAEEPYSPVPEAPRDDET


S.clavuligerus   34 ETYEPYEPYEP------YEHAPTGPTA--PTGPGRAERRRAARSEGHRAR----RRPPKP
S.coelicolor     54 VALRIPEPPAP------RTAAGTGPIGGGPDGGGRAARRKAAKRRHGRRG----APRDQA
S.viridochromog  88 VALRIPDPPPSGSDAVSGSPASSATRRDDPSQGGRAARRKAAKRRHGRHG----GAPDTG
S.sviceus        96 VALRIPDPPPP---SISASTASSVTSATGSPQGGRAARRKAAKGRHGRHG----GGAAP-
S.ghanaensis     76 VALRIPDPPDP--------PERPVREEPTGAPGGRAARRKAAKRRQGRRG----SPRGAG
S.griseoflavus   93 VALRIPEPPER--------TAR--EEPAPAAPGGRAARRKAAKKRQGRRG----GSHKAA
S.scabiei       120 VALRVEEARRAA----ARAESIPPSAVTAG----RAARRKAARRHGRHGG---SPAAHAA
S.albus         120 MALRAADPARTD----EERDETPPATISAGAEGGRAARRKAAKRNARPGGRRGTGRRRAG


S.clavuligerus   82 SPQATAPS----------RPLTRTEARRAARAARDSPGVVVSRALGEVFITLGVVMLLFV
S.coelicolor    104 PEEEAEQAP--------KAPLSRVEARRQARARKPGAAVVASRAIGEIFITTGVLMLLFV
S.viridochromog 144 PAPEGAREE-------RRAPLSRVEARRQARARKPSPAVVASRAIGEVFITTGVLMLLFV
S.sviceus       148 EAQSDQESP-------DGRPLSRVEARRQAKARKPGAAVVASRAIGEVFITTGVLMLLFV
S.ghanaensis    124 EPEAASRAP----EGPSGAPLSRVEARRRARARRASPAVVASRAIGEVFITTGVLMLLFV
S.griseoflavus  139 EAPRDSRGPDGGSEGASRAPLSRVEARRRARASKASPAVLASRAIGEVFITCGVLMLLFV
S.scabiei       169 SRAAQGPGQPLGAAPSASAPLSRIEARRAARAQKPGVGVIASRVIGEVFITTGVVMLLFV
S.albus         176 APAPEASGTPGGGDEPPERPRTRVEARRAARARKPSAATVASRAVGELFITIGVLMLLFV


S.clavuligerus  132 TYQLWWTNVRADQQAGREKERIQRSWAAG-RAPGAFRPGEGFAIMYIPKLDVVVPVAASI
S.coelicolor    156 TYQLWWTNVRAHAQANQAASNLQDDWANGKRSPGSFEPGQGFALLHIPKLDVVVPIAEGI
S.viridochromog 197 TYQLWWTNVRAHAQAGSEASSLQNDWASGKRNPGAFEPGQGFAI----------------
S.sviceus       201 SYQLWWTNVRAHAQADKEASSLQNDWASGKGAPGTFEPGQ-------------------
S.ghanaensis    180 TYQLWWTNVRAQAQAGKELSDLQSDWANGKRNPGVFEPGQGFALLHIPKLDVVVPIAEGI
S.griseoflavus  199 TYQLWWTNVRAHAQAGKEVSDLQSDWANGAGKPGAFEPGQGF------------------
S.scabiei       229 TYQLWWSNIRAH------------------------------------------------
S.albus         236 TYQLW-------------------------------------------------------
```

Figure 3.5.6 Alignment of the N-terminal extension of SrtE1 orthologues in diverse *Streptomyces* species. Sequence alignment was generated using the ClustalOmega server [43]. The bacterial species and accession numbers of the amino acid sequences used for the alignment are as follows: *Streptomyces clavuligerus* (EDY51820)*, Streptomyces coelicolor* (WP_011029270)*, Streptomyces viridochromogenes* (EFL33429)*, Streptomyces sviceus* (WP_007383307)*, Streptomyces ghanaensis* (WP_004985874)*, Streptomyces griseoflavus* (WP_040906697)*, Streptomyces scabiei* (WP_013002225)*, Streptomyces albus* (WP_015507549). Conserved residues are indicated in *red*, and related amino acids are indicated in *blue*. The conserved acidic and basic regions are boxed in *red* and *blue*, respectively. The conserved transmembrane helix is boxed in *green*.
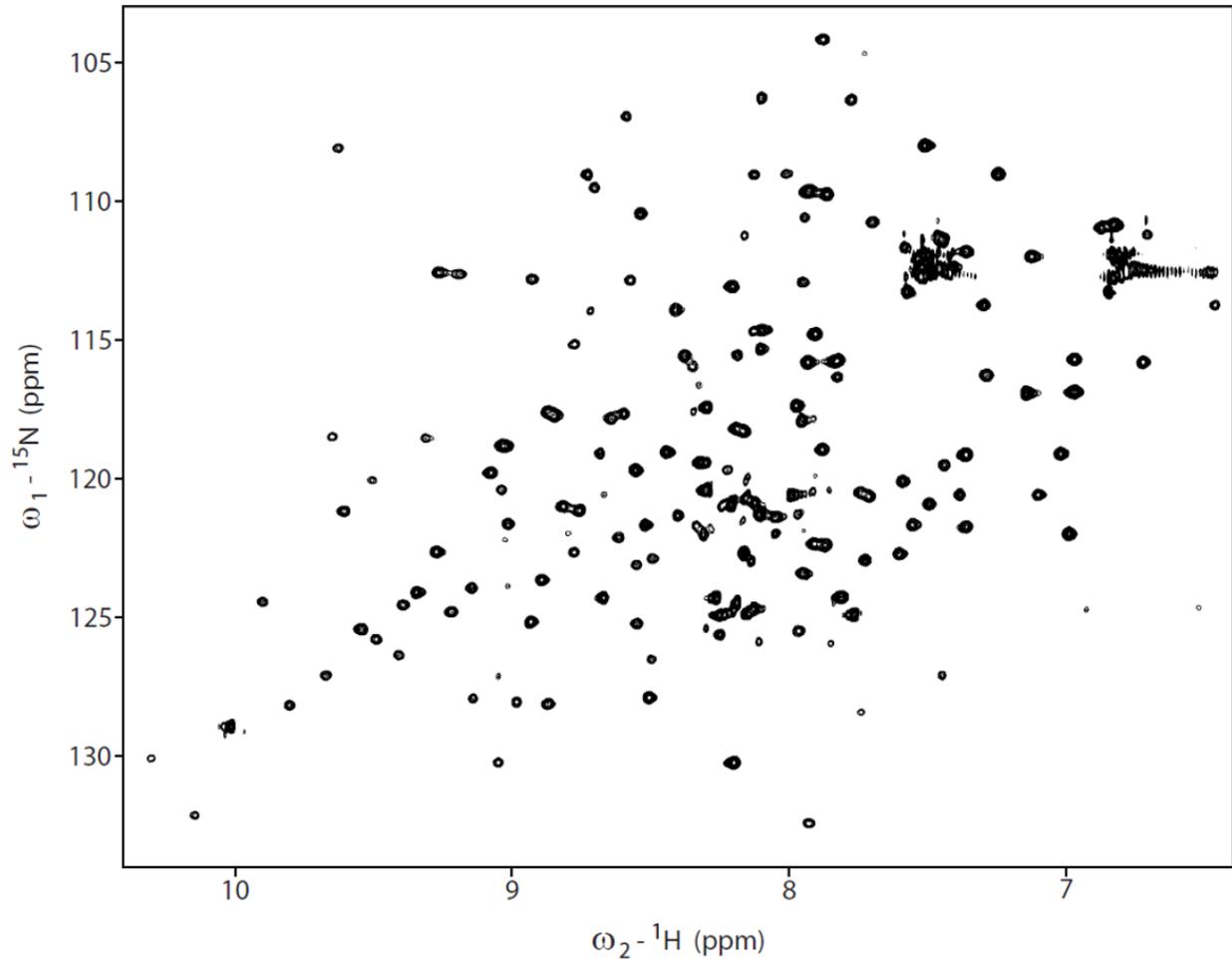
Figure 3.5.7 ^{1}H-^{15}N HSQC spectrum of the SrtE1 extracellular domain construct. The ^{1}H-^{15}N HSQC spectrum yielded reasonably well-resolved cross peaks, indicating that the SrtE1 protein was folded. However, there were substantially fewer peaks than anticipated for the molecular weight of the SrtE1$^{\Delta N}$ construct (20.8 kDa). In particular, ~29 peaks were absent in the NMR spectra; only ~148 resolvable cross peaks from backbone amides were observed, whereas 177 cross peaks are expected (194 total residues – 16 proline residues – the N-terminal residue). The reduced number of signals in the ^{1}H-^{15}N HSQC spectrum is compatible with the N-terminal linker experiencing motions that are intermediate on the chemical exchange time scale (µs to ms), causing their signals to be broadened. Such motions are also compatible

with our inability to visualize residues from the N-terminal linker (with the exception of the AQA tripeptide) within the electron density map. Combined, the NMR and crystallography data suggest that the AQA tripeptide is housed in a structurally disordered segment of the isolated enzyme. ^{15}N-labeled SrtE1$^{\Delta N}$ for NMR studies was concentrated to 350 µM in NMR buffer (50 mM NaPO4, pH 6.8; 150 mM NaCl, 7% D$_2$O). HSQC spectra were acquired with 32 scans at 298 K on Bruker 600 MHz spectrometers equipped with a triple-resonance cryogenic probe.

```
Streptomyces_coelicolor_SrtE1     197 FALLHIPKLDVVV-PIAEGISSKKVLD---RGMVGHYAE----DGLKTAMPDAKAGNFGL
Streptomyces_coelicolor_SrtE2      96 IGFLHVPAMSEGD-ILVEKGTSMKILN---DGVAGYYTD----PVKATLPTSDEKGNFSL
Bifidobacterium_longum            140 VAQIYIPRFGSQWHRNIVEGTTLEQLN---RHGLGHYDT----T-----QMPGQVGNFAV
Corynbacterium_diptheriae          29 FARMYIPQFGSDFQFAIVEGTTDADLE---AGP-GHYND----T-----QLPGERGNFAV
Corynebacterium_efficiens         102 FARMYIPSFGSDFHFAVIEGTDEEELL---AGP-GRYTD----S-----QMPGEAGNFAV
Corynebacterium_glutamicum         86 FARMYVPAFGSDFNFAVIEGTDEEDLL---AGP-GRYVD----S-----QMPGEAGNFAV
Streptomyces_avermitilis_1         90 IGFLHVPAMNNGE-VLVRKGTSTQVLN---NGVAGYYTD----PVKAMLPMTGKNGNFTL
Streptomyces_avermitilis_2        124 YAVLTIPRLSLRV-PVAEGIGKASVLN---HGYVGHYPK----T-----AQPGRAGNFAL
Streptomyces_avermitilis_3         80 FAVMYIPRLGFTWNKPVLEGTGTEVLK---KGL-GHYAN----T-----ARLGQKGNFAV
Streptomyces_griseus_1             85 FATMHIPRFGADWEWPVLENTAVGTLK---KGL-GHYSA----T-----ARPGDTGNFAV
Streptomyces_griseus_2            319 FAIMHIPKLDVVA-PIAEGIDKEKVLD---RGMLGHYAE----GRLKTAMPSDKQGNFSV
Streptomyces_griseus_3            208 YAVLRIPRIGLTA-PVAEGTSKGGVLD---RGYVGHYAR----T-----AQAGQAGNFAL
Streptomyces_griseus_4             90 IGFLHVPAMKNGE-VLVKKGTDPETLN---NGIAGYYTD----PVESALPWD-DEGNFTL
Thermobifida_fusca                 99 NSRLYIPKTDQNW-VVVS-GVGPEDIK---YGP-GWYPESWTPEGMVPAARAGQPGNYAV
Tropheryma_whipplei                97 IAVLFVPRFGNKYKRVIRETTDVTRVLNSKTAGVGHYPH----T-----ALPGTSGNFAV
Clostridium_perfringens_SrtD       43 IALIDIEKIGVHT--VIAEGSTLDVLE---N-NIGHFEN----T-----AMPGENGNFSI
Clostridium_perfringens           101 IGILNIPKINLEI--GIIEGVSYEDIK---Y-VVGHFPG----S-----PMPGEKGNFSI
Clostridium_tetani                 71 MAIIEIPSIGLKS--VIVEGTEMEKLR---Y-YIGHFKE----T-----ALPGQLGNFCI
```

Figure 3.5.8 Alignment of phylogenetically determined class E enzymes with

SrtD from *Clostridium perfringens*. Sequence alignment was generated using the

ClustalOmega server [43]. The bacterial species and accession numbers of the amino acid

sequences used for the alignment are as follows: *Streptomyces coelicolor* (NP_628038 and

NP_628037), *Bifidobacterium longum* (NP_695779), *Corynebacterium diptheriae* (NP_940575),

Corynebacterium efficiens (NP_739396), *Corynebacterium glutamicum* (NP_602126),

Streptomyces avermitilis (NP_825514; NP_826383; NP_825510), *Streptomyces griseus*

(YP_001825232; YP_001825235; YP_001826193; YP_001825236), *Thermobifida fusca*

(YP_290439), *Tropheryma whipplei* (NP_787692), *Clostridium perfringens* (WP_003467492),

Clostridium tetani (WP_011099430). Conserved residues are indicated in *red*, and related amino

acids are indicated in *blue*. The conserved tyrosine residue within the B3/B4 loop of class E

sortases is boxed in *black*.

3.6 Tables

Table 3.6.1. **Data collection and structure refinement statistics.**

	SrtE1
Data collection	
Space group	C222$_1$
Cell dimensions	
a, b, c (Å)	53.11, 104.30, 79.02
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	43.53-1.93 (1.98-1.93)[b]
R_{merge} (%)[a]	19.0 (83.8)
CC1/2 (%)	98.7 (77.1)
I / σI	6.82 (2.0)
Completeness (%)	97.5 (99.9)
Redundancy	5.9 (4.7)
Wilson B-factor (Å^2)	26.2
Refinement	
Resolution (Å)	43.5-1.93
No. reflections	17583
R_{work} / R_{free} (%)[c]	19.9/22.9
No. atoms	1283
Protein	1232
Ligand/ion	14
Water	37
B-factors (all atoms)	33.7
Protein	32.9
Ligand/ion	98.6
Water	33.6
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.089
Ramachandran favored (%)	98.0
Ramachandran allowed (%)	2.0
Ramachandran generally allowed (%)	0.0
Ramachandran outliers (%)	0.0

[a] Data from two crystals were merged for structure determination.

[b] Values in parentheses are for highest-resolution shell.

[c] R$_{free}$ calculated using 5% of the data.

3.7 References

1. Navarre WW, Schneewind O. Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. Microbiol Mol Biol Rev. 1999;63: 174–229. Available: http://www.ncbi.nlm.nih.gov/pubmed/10066836

2. Clancy KW, Melvin J a, McCafferty DG. Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. Biopolymers. 2010;94: 385–96. doi:10.1002/bip.21472

3. Spirig T, Weiner EM, Clubb RT. Sortase enzymes in Gram-positive bacteria. Mol Microbiol. 2011;82: 1044–1059. doi:10.1111/j.1365-2958.2011.07887.x

4. Schneewind O, Missiakas D. Sec-secretion and sortase-mediated anchoring of proteins in Gram-positive bacteria. Biochim Biophys Acta - Mol Cell Res. 2014;1843: 1687–1697. doi:10.1016/j.bbamcr.2013.11.009

5. Krishnan V. Pilins in gram-positive bacteria: A structural perspective. IUBMB Life. 2015;67: 533–543. doi:10.1002/iub.1400

6. Bradshaw WJ, Davies AH, Chambers CJ, Roberts AK, Shone CC, Acharya KR. Molecular features of the sortase enzyme family. FEBS J. 2015;282: 2097–2114. doi:10.1111/febs.13288

7. Hendrickx APA, Budzik JM, Oh S-Y, Schneewind O. Architects at the bacterial surface — sortases and the assembly of pili with isopeptide bonds. Nat Rev Microbiol. Nature Publishing Group; 2011;9: 166–176. doi:10.1038/nrmicro2520

8. Marraffini LA, DeDent AC, Schneewind O. Sortases and the Art of Anchoring Proteins to the Envelopes of Gram-Positive Bacteria. Microbiol Mol Biol Rev. American Society for

Microbiology; 2006;70: 192–221. doi:10.1128/MMBR.70.1.192-221.2006

9. Foster TJ, Geoghegan JA, Ganesh VK, Höök M. Adhesion, invasion and evasion: the many functions of the surface proteins of Staphylococcus aureus. Nat Rev Microbiol. Nature Research; 2013;12: 49–62. doi:10.1038/nrmicro3161

10. Scott JR, Barnett TC. Surface Proteins of Gram-Positive Bacteria and How They Get There. http://dx.doi.org/101146/annurev.micro60080805142256. Annual Reviews; 2006;

11. Duong A, Capstick DS, Di Berardo C, Findlay KC, Hesketh A, Hong H-J, et al. Aerial development in Streptomyces coelicolor requires sortase activity. Mol Microbiol. 2012;83: 992–1005. doi:10.1111/j.1365-2958.2012.07983.x

12. Cascioferro S, Totsika M, Schillaci D. Sortase A: An ideal target for anti-virulence drug development. Microb Pathog. 2014;77: 105–112. doi:10.1016/j.micpath.2014.10.007

13. Fitzgerald-Hughes D, Devocelle M, Humphreys H, Augustin R, Anton-Erxleben F, Jungnickel S, et al. Beyond conventional antibiotics for the future treatment of methicillin-resistant Staphylococcus aureus infections: two novel alternatives. FEMS Immunol Med Microbiol. The Oxford University Press; 2012;65: 399–412. doi:10.1111/j.1574-695X.2012.00954.x

14. Maresso AW, Schneewind O. Sortase as a Target of Anti-Infective Therapy. Pharmacol Rev. American Society for Pharmacology and Experimental Therapeutics; 2008;60: 128–141. doi:10.1124/pr.107.07110

15. Suree N, Jung M, Clubb R. Recent Advances Towards New Anti-Infective Agents that Inhibit Cell Surface Protein Anchoring in Staphylococcus aureus and Other Gram-Positive Pathogens. Mini-Reviews Med Chem. 2007;7: 991–1000. doi:10.2174/138955707782110097

16. Zhang J, Liu H, Zhu K, Gong S, Dramsi S, Wang Y-T, et al. Antiinfective therapy with a small molecule inhibitor of *Staphylococcus aureus* sortase. Proc Natl Acad Sci. National Academy of Sciences; 2014;111: 13517–13522. doi:10.1073/pnas.1408601111

17. Rasko DA, Sperandio V. Anti-virulence strategies to combat bacteria-mediated disease. Nat Rev Drug Discov. Nature Publishing Group; 2010;9: 117–128. doi:10.1038/nrd3013

18. Proft T. Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilisation. Biotechnol Lett. Springer Netherlands; 2010;32: 1–10. doi:10.1007/s10529-009-0116-0

19. Popp MW-L, Ploegh HL. Making and breaking peptide bonds: protein engineering using sortase. Angew Chem Int Ed Engl. 2011;50: 5024–32. doi:10.1002/anie.201008267

20. Tsukiji S, Nagamune T. Sortase-Mediated Ligation: A Gift from Gram-Positive Bacteria to Protein Engineering. ChemBioChem. WILEY-VCH Verlag; 2009;10: 787–798. doi:10.1002/cbic.200800724

21. Ritzefeld M. Sortagging: a robust and efficient chemoenzymatic ligation strategy. Chemistry. 2014;20: 8516–29. doi:10.1002/chem.201402072

22. Antos JM, Truttmann MC, Ploegh HL. Recent advances in sortase-catalyzed ligation methodology. Curr Opin Struct Biol. 2016;38: 111–118. doi:10.1016/j.sbi.2016.05.021

23. van 't Hof W, Maňásková SH, Veerman ECI, Bolscher JGM. Sortase-mediated backbone cyclization of proteins and peptides. Biol Chem. 2015;396: 283–293. doi:10.1515/HSZ-2014-0260

24. Schmohl L, Schwarzer D. Sortase-mediated ligations for the site-specific modification of proteins. Curr Opin Chem Biol. 2014;22: 122–128. doi:10.1016/j.cbpa.2014.09.020

25. Ta HT, Peter K, Hagemeyer CE. Enzymatic Antibody Tagging: Toward a Universal

Biocompatible Targeting Tool. Trends Cardiovasc Med. 2012;22: 105–111.

doi:10.1016/j.tcm.2012.07.004

26. Atilano ML, Pereira PM, Yates J, Reed P, Veiga H, Pinho MG, et al. Teichoic acids are

temporal and spatial regulators of peptidoglycan cross-linking in Staphylococcus aureus.

Proc Natl Acad Sci U S A. 2010;107: 18991–6. doi:10.1073/pnas.1004304107

27. Amer BR, Macdonald R, Jacobitz AW, Liauw B, Clubb RT. Rapid addition of unlabeled

silent solubility tags to proteins using a new substrate-fused sortase reagent. J Biomol

NMR. Springer Netherlands; 2016;64: 197–205. doi:10.1007/s10858-016-0019-z

28. Schneewind O, Model P, Fischetti VA. Sorting of protein a to the staphylococcal cell wall.

Cell. Cell Press; 1992;70: 267–281. doi:10.1016/0092-8674(92)90101-H

29. Perry AM, Ton-That H, Mazmanian SK, Schneewind O. Anchoring of Surface Proteins to

the Cell Wall of Staphylococcus aureus. III. LIPID II IS AN IN VIVO

PEPTIDOGLYCAN SUBSTRATE FOR SORTASE-CATALYZED SURFACE

PROTEIN ANCHORING. J Biol Chem. 2002;277: 16241–16248.

doi:10.1074/jbc.M109194200

30. Ruzin A, Severin A, Ritacco F, Tabei K, Singh G, Bradford PA, et al. Further Evidence

that a Cell Wall Precursor [C55-MurNAc-(Peptide)-GlcNAc] Serves as an Acceptor in a

Sorting Reaction. J Bacteriol. American Society for Microbiology; 2002;184: 2141–2147.

doi:10.1128/JB.184.8.2141-2147.2002

31. Huang X, Aulabaugh A, Ding W, Kapoor B, Alksne L, Tabei K, et al. Kinetic Mechanism

of *Staphylococcus aureus* Sortase SrtA. Biochemistry. 2003;42: 11307–11315.

doi:10.1021/bi034391g

32. Frankel BA, Kruger RG, Robinson DE, Kelleher NL, McCafferty DG. *Staphylococcus*

aureus Sortase Transpeptidase SrtA: Insight into the Kinetic Mechanism and Evidence for a Reverse Protonation Catalytic Mechanism [†]. Biochemistry. 2005;44: 11188–11200. doi:10.1021/bi050141j

33.   Schneewind O, Fowler A, Faull KF. Structure of the cell wall anchor of surface proteins in Staphylococcus aureus. Science. American Association for the Advancement of Science; 1995;268: 103–6. doi:10.1126/science.7701329

34.   Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44: D279-85. doi:10.1093/nar/gkv1344

35.   Dramsi S, Trieu-Cuot P, Bierne H. Sorting sortases: a nomenclature proposal for the various sortases of Gram-positive bacteria. Res Microbiol. Elsevier Masson; 2005;156: 289–297. doi:10.1016/j.resmic.2004.10.011

36.   Comfort D, Clubb RT. A Comparative Genome Analysis Identifies Distinct Sorting Pathways in Gram-Positive Bacteria. Infect Immun. 2004;72: 2710–2722. doi:10.1128/IAI.72.5.2710–2722.2004

37.   Suree N, Liew CK, Villareal V a, Thieu W, Fadeev E a, Clemens JJ, et al. The structure of the Staphylococcus aureus sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. J Biol Chem. 2009;284: 24465–77. doi:10.1074/jbc.M109.022624

38.   Jacobitz AW, Wereszczynski J, Yi SW, Amer BR, Huang GL, Nguyen A V., et al. Structural and Computational Studies of the Staphylococcus aureus Sortase B-Substrate Complex Reveal a Substrate-stabilized Oxyanion Hole. J Biol Chem. 2014;289: 8891–8902. doi:10.1074/jbc.M113.509273

39. Chan AH, Yi SW, Terwilliger AL, Maresso AW, Jung ME, Clubb RT. Structure of the *Bacillus anthracis* Sortase A Enzyme Bound to Its Sorting Signal. J Biol Chem. 2015;290: 25461–25474. doi:10.1074/jbc.M115.670984

40. Zong Y, Bice TW, Ton-That H, Schneewind O, Narayana SVL. Crystal structures of Staphylococcus aureus sortase A and its substrate complex. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2004;279: 31383–9. doi:10.1074/jbc.M401374200

41. Elliot MA, Karoonuthaisiri N, Huang J, Bibb MJ, Cohen SN, Kao CM, et al. The chaplins: a family of hydrophobic cell-surface proteins involved in aerial mycelium formation in Streptomyces coelicolor. Genes Dev. Cold Spring Harbor Laboratory Press; 2003;17: 1727–1740. doi:10.1101/gad.264403

42. Duong A, Koteva K, Sexton DL, Elliot MA. Liquid Chromatography-Tandem Mass Spectrometry to Define Sortase Cleavage Products. 2016. pp. 99–108. doi:10.1007/978-1-4939-3676-2_8

43. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7. doi:10.1038/msb.2011.75

44. Liew CK, Smith BT, Pilpa R, Suree N, Ilangovan U, Connolly KM, et al. Localization and mutagenesis of the sorting signal binding site on sortase A from *Staphylococcus aureus*. FEBS Lett. 2004;571: 221–226. doi:10.1016/j.febslet.2004.06.070

45. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32: 1792–7. doi:10.1093/nar/gkh340

46. Kruger RG, Otvos B, Frankel BA, Bentley M, Dostal P, McCafferty DG. Analysis of the

Substrate Specificity of the Staphylococcus aureus Sortase Transpeptidase SrtA. Biochemistry. 2004;43: 1541–1551. doi:10.1021/bi035920j

47. Duong A, Capstick DS, Di Berardo C, Findlay KC, Hesketh A, Hong H-J, et al. Aerial development in Streptomyces coelicolor requires sortase activity. Mol Microbiol. Blackwell Publishing Ltd; 2012;83: 992–1005. doi:10.1111/j.1365-2958.2012.07983.x

48. Kruger RG, Dostal P, McCafferty DG. Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the Staphylococcus aureus sortase transpeptidase SrtA. Anal Biochem. 2004;326: 42–48. doi:10.1016/j.ab.2003.10.023

49. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem. 2004;47: 1739–1749. doi:10.1021/jm0306430

50. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. J Med Chem. 2004;47: 1750–1759. doi:10.1021/jm030644s

51. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem. 2005;26: 1781–802. doi:10.1002/jcc.20289

52. Suryadinata R, Seabrook SA, Adams TE, Nuttall SD, Peat TS. Structural and biochemical analyses of a Clostridium perfringens Sortase D transpeptidase. Acta Cryst. 2015;71. doi:10.1107/S1399004715009219

53. Dorr BM, Ham HO, An C, Chaikof EL, Liu DR. Reprogramming the specificity of sortase enzymes. Proc Natl Acad Sci. 2014; doi:10.1073/pnas.1411179111

54. Kabsch W, IUCr, K. D, A. KP, K. D, S. M, et al. *XDS*. Acta Crystallogr Sect D Biol Crystallogr. International Union of Crystallography; 2010;66: 125–132. doi:10.1107/S0907444909047337

55. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. International Union of Crystallography; 2007;40: 658–674. doi:10.1107/S0021889807021206

56. Adams PD, Afonine P V, Bunkó G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst. 2010;66: 213–221. doi:10.1107/S0907444909052925

57. Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des. Springer Netherlands; 2013;27: 221–234. doi:10.1007/s10822-013-9644-8

58. Tubert-Brohman I, Sherman W, Repasky M, Beuming T. Improved Docking of Polypeptides with Glide. J Chem Inf Model. 2013;53: 1689–1699. doi:10.1021/ci400128m

59. D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden RE, Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry AK, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg CS, R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf X, Kollman W and PA. AMBER 14. San Fransisco; 2014.

60. Gregory MA, Till R, Smith MCM. Integration Site for Streptomyces Phage □BT1 and

Development of Site-Specific Integrating Vectors. J Bacteriol. 2003;185: 5320–5323. doi:10.1128/JB.185.17.5320–5323.2003

61. Duong A, Capstick DS, Di Berardo C, Findlay KC, Hesketh A, Hong H-J, et al. Aerial development in Streptomyces coelicolor requires sortase activity. Mol Microbiol. Blackwell Publishing Ltd; 2012;83: 992–1005. doi:10.1111/j.1365-2958.2012.07983.x

62. Keijser BJ, van Wezel GP, Canters GW, Kieser T, Vijgenboom E. The ram-dependence of Streptomyces lividans differentiation is bypassed by copper. J Mol Microbiol Biotechnol. 2000;2: 565–74. Available: http://www.ncbi.nlm.nih.gov/pubmed/11075933

Chapter 4

The Crystal Structure of TarA from *Thermobacter italicus* Reveals a Novel Glycosyltransferase Structural Fold

4.1 Overview

Our understanding of bacterial Wall Teichoic Acid (WTA) biosynthesis has been limited because protein structural information has only been obtained for the soluble, substrate feeder enzymes (TarD, TarI, TarJ), main chain glycosylation enzymes (TarM, TarS), and a single domain of the peripherally membrane-associated, WTA chain polymerase, TagF [1–4]. We have determined the 1.8 Å resolution crystal structure of the early stage WTA glycosyltransferase (GT), TarA from *Thermobacter italicus* (*T. italicus*). The TarA enzyme adopts a structurally novel protein fold that is distinct from the previously characterized GT-A, GT-B, GT-C, and GT-D classes [5,6]. Specifically, the TarA "GT-E" fold is formed by a compact primary sequence that adopts an eight β strand core structure that is surrounded by several accessory helices.

The *T. italicus* TarA enzyme is the first structurally characterized member of the WecB-TagA-CpsF GT family, providing insight into how members of the Carbohydrate-Active Enzymes (CAZy) GT-26 family perform their glycosyltransferase activity. The crystal structures of the apo- and UDP-bound TarA enzyme from *T. italicus* crystallized in dimeric and trimeric forms, respectively. TarA oligomerization is also observed in solution during size exclusion and NMR studies, suggesting that these oligomer complexes may be physiologically relevant. Sequence conservation mapping onto the TarA structure indicated that highly conserved residues cluster at the interfaces of these oligomeric forms, suggesting that oligomerization of TarA may be required to form a competent active site. Limited proteolysis studies reveal the presence of a protease-susceptible loop that is protected from cleavage upon addition of UDP and ManNAc ligands, suggesting that this loop may be involved in ligand binding. Additionally, co-crystallization of TarA with UDP in combination with sequence conservation analysis has identified putative functional residues for substrate binding and catalysis, including residues that

may act as a catalytic base for nucleophile activation. A detailed mechanism of substrate binding, oxocarbenium ion-like transition state formation, and nucleophile-mediate catalysis remains to be elucidated by TarA co-crystallization studies with UDP-ManNAc and lipid ligands.

4.2 Introduction

GTs catalyze the transfer of a sugar moiety from a nucleotide-activated sugar or lipid-phospho-sugar donor to a broad range of acceptor substrates to produce chemically diverse products. Currently, over 100 distinct GT families are classified within the Carbohydrate-Active Enzymes (CAZy) database according to amino acid sequence similarity [7]. Each member of a GT family is predicted to adopt a similar three-dimensional fold; however, substrate poly-specificity has been noted within these families, making reliable functional prediction challenging. To date, only four structural classes of GT have been identified among the >40 out of 100 CAZy families represented in the Protein Data Bank (PDB), suggesting that a similar structural scaffold can be appended with unique active site residues to produce catalytically-diverse enzymes [5,6,8]. GTs are highly selective enzymes, distinguishing subtle structural differences in sequence and stereochemistry of sugar substrates; precise spatial arrangement of the GT active site is required to ensure highly selective substrate recognition and transition state stabilization. Specifically, GTs utilize post-translational modifications, oligomerization, multi-domain structure (i.e. GT appended with a lectin domain), domain-domain and protein-protein interactions, coupled with protein dynamics and conformational changes, to produce a range of distinct products [5,8].

144

4.2.1 Glycosyltransferase structural folds

Four distinct structural folds of GTs have thus far been identified, classified as GT-A, GT-B, GT-C, and GT-D [5,8]. The limited number of structural folds adopted by GTs could reflect structural restraints imposed by binding nucleotide-sugar or lipid-phospho-sugar donors. However, GTs demonstrate diverse functionality that is achieved through conformational plasticity [5].

The structure of GT-A enzymes encompass two tightly coordinated domains, each containing a variation of the β/α/β Rossmann fold [5,8] (Figure 4.7.2A). The active site resides at the interface of the N- and C-terminal domains, which recognize a nucleotide-sugar moiety and acceptor molecule, respectively. Additional structural features, such as short, internal loops and termini extensions, may reside within or adjacent to the active site where they assist in substrate binding and catalysis. These dynamic loops and extensions often undergo large conformational changes during the catalytic cycle to regulate access to the active site, substrate binding, and product release. An Asp-X-Asp motif (where X is any amino acid) is also commonly present to coordinate a divalent cation or ribose via the side chain carboxylate groups, although this feature is not conserved.

The structures of GT-B enzymes consist of two domains containing variations of a Rossmann fold separated by a large cleft that houses a reaction center [5,8]. The N-terminal domain within GT-B enzymes varies in arrangement of secondary structural elements to recognize a wide range of acceptor substrates. The C-terminal domain binds the nucleotide-sugar donor, which triggers crucial inter-domain movement to position the domains together and to construct a competent active site. As observed with GT-A enzymes, flexible regions within GT-B enzymes may promote water-soluble and hydrophobic substrate binding, and may be

necessary to associate with substrates within the highly dynamic lipid bilayer. Interactions with the membrane have also been shown to modulate catalysis through intra-domain conformational changes, including local structural reshuffling of secondary structure elements and open-to-close motions. Finally, the structures of GT-C enzymes reveal an integral polytopic membrane domain that recognizes the lipid-phospho-sugar donor and a variable soluble domain involved in acceptor binding. The interface between these two domains encompasses the reaction center.

The structural folds of GT-C and GT-D classes have been recently characterized (REF). Two GT-C crystal structures of bacterial and archaeal oligosaccharyltransferase PglB have been determined. This GT-C class consists of two distinct domains, a transmembrane domain and a periplasmic domain. Cavities located near the interface of the TM and periplasmic domains grant access to the donor and acceptor substrates, with the interface forming the reaction center. Specifically, the periplasmic domain is thought to bind soluble acceptor substrates, while the transmembrane domain interacts with the lipid-phospho-sugar donor or lipid acceptors. A flexible loop located in the membrane region spans into the soluble domain and is involved in acceptor recognition. The structure of a GT-D enzyme, DUF1792, reveals a fold with three distinct regions. An N-terminal region forms two αβα sandwich domains with an Asp-X-Glu motif-containing metal binding site. A central region contains a Rossmann-like fold, followed by a C-terminal region. Together, these regions form a nucleotide-binding site to coordinate UDP.

4.2.2 Glycosyltransferase reaction mechanisms

GTs are classified as either "inverting" or "retaining" enzymes according to the anomeric configuration of their reactants and products. Several studies suggest that inverting GTs catalyze sugar transfer through an SN2 reaction via a single displacement step with the formation of a

catalytic base-assisted, oxocarbenium ion-like transition state reaction [9–11]. Alternately, the mechanism of retaining GTs is controversial with two different mechanisms currently proposed: 1) double displacement, and 2) front-face SNi mechanism [8]. In the double displacement mechanism, a nucleophile from the donor substrate attacks the anomeric carbon to produce a glycosyl-enzyme intermediate, followed by attack of the anomeric carbon by the acceptor to form a glycosidic bond. In the internal return mechanism, nucleophilic attack and departure of the leaving group occur on the same face of the sugar, forming a short-lived, oxocarbenium-like transition state with glycoside bond formation and subsequent phosphate bond breakdown. The current framework distinguishes two types of retaining GTs based on the presence or absence of a nucleophile within the active site, in which GTs lacking a nucleophile contain active sites with high electrostatic potential to stabilize the oxocarbenium ion-like transition state intermediate.

Here, we reveal the crystal structure of a novel GT fold, which we have termed GT-E, and compare its structure to other GT classes of known structure. Analysis of the structure of the TarA glycosyltransferase provides mechanistic insight into how this enzyme catalyzes transfer of ManNAc from UDP-ManNAc to C55-PP-GlcNAc during WTA polymer synthesis. These studies could facilitate the discovery and rational design of WTA synthesis inhibitors that may be useful in treating infections caused by Gram-positive pathogens.

4.3 Results and discussion

4.3.1 The crystal structure of the novel TarA glycosyltransferase

To characterize the glycosyltransferase WecB/TagA/CpsF family (PFAM03808; CAZy GT26), we determined the 1.8Å resolution crystal structure of the N-terminal region of TarA (TarA$^{\Delta C}$, residues M1-G195), which consists of a conserved extracytoplasmic domain that is

missing a putative C-terminal membrane anchoring region (residues G196-R248). Crystallization of TarA$^{\Delta C}$ produced diffracting crystals within three days from hanging drops containing light precipitate. Selenomethionine (SeMet)-labeled TarA$^{\Delta C}$ in its apo-form and complexed with UDP crystallized in the P2$_1$ space group, with eight and six protein molecules in the asymmetric unit, respectively. The apo-structure was determined by multiple anomalous dispersion (MAD) from selenomethionine-labeled protein crystals, and is well-defined by continuous electron density for residues forming the catalytic domain (M1-G195) (Figure 4.7.1A). The 3.1Å resolution crystal structure of TarA$^{\Delta C}$ complexed with a UDP ligand was determined using molecular replacement with the apo-TarA model. Complete data collection and structural statistics are provided in Table 1.

The extracytoplasmic domain of TarA contains a unique topology of secondary structure elements (Figure 4.7.1B). TarA$^{\Delta C}$ starts with strand β1 (residues R3-I7), followed by a short hydrogen bonded turn such that residues in the following strand β2 (residues V10-D13) interact with the β1 strand in an antiparallel manner. The chain then leads into an alpha helix H1 (residues M17-F27), which reverses direction before forming strand β3 (residues H34-A37). The chain then produces helix H2 (residues A41-K49), followed by a short turn that connects it to helix H3 (residues K51-L57). A turn positions strand β4 (residues L62-N63) to interact with strand β3 in a parallel manner. The chain then forms helix H4 (residues G69-A73), which directly initiates a 3$_{10}$ helix (residues S74-V76) that is followed by a 10 amino acid random coil. Helix H5 (residues G87-I98) turns into strand β5 (residues K104-G109), which turns into helix H6 (residues V114-L125). The chain then reverses its direction before forming strand β6 (residues K130-H135), which is positioned to interact in a parallel manner to strand β5. After reversing its direction the chain forms helix H7 (residues E141-N153), which turns to position

strand β7 (residues V158-A162) to interact in a parallel manner with strand β5. This is followed by helix H8 (residues A165-K174), after which a turn positions strand β8 (residues I183-G186) to interact with strand β7 in a parallel manner. The structure is completed by residues in helix H5 (residues S190-I194).

The TarA structure is composed to two distinct regions: an N-terminal region formed by strands β1 and β2 and helices H2, H3 and H4, and a C-terminal region that consists of two Rossmann-like folds (βαβ) formed by β7, H8, β8, and β6, H9, β5. No significant structural differences were observed between the apo- and ligand-bound protein forms (RMSD of 0.18 Å). The TarA structure is distinct from previously reported glycosyltransferase structures that comprise the GT-A, GT-B, GT-C, and GT-D classes; therefore, we have classified this novel glycosyltransferase structural fold as "GT-E."

4.3.2 Structural comparison of the class GT-E fold of TarA to existing glycosyltransferase classes

To date, four unique glycosyltransferase folds have been structurally characterized, termed GT-A, GT-B, GT-C or GT-D classes. However, structural comparison to these previously characterized GT classes reveals that the TarA glycosyltransferase fold is distinct (Figure 4.7.2A). A search with the DALI server indicated that the DUF1792 enzyme (PDB ID: 4PFX), a GT-D glycosyltransferases, is most structurally related to TarA with a weak Z-score of 8.6. Structural overlay of TarA with the GT-D fold reveals key differences, producing an RMS deviation of 17.8 angstroms. Specifically, the GT-D glycosyltransferase contains a core of seven

β strands that is surrounded by 13 alpha helices, whereas the TarA structural fold contains an additional eighth β strand within its core and is decorated with a unique distribution of helices.

The primary sequences of TarA and the GT-D enzyme also vary considerably. Wild type TarA and DUF1792 differ in the length of their primary sequences (248 residues for TarA vs. 277 residues for DUF1792) due to distinct sequence insertions and deletions (Figure 4.7.2B). Specifically, the GT-D enzyme contains a uniquely extended N-terminus and lacks the highly conserved C-terminal domain present within TarA homologs. Furthermore, TarA and the GT-D enzyme share limited sequence homology according to BLAST analysis (32% identity over 28% sequence coverage). Key conserved residues that are important for glycosyltransferase activity among DUF1792 homologs are not present in TarA homolog sequences, further supporting that TarA represents a novel glycosyltransferase fold (Figure 4.7.2B, bold residues). Specifically, DUF1792 contains a conserved DXE motif, where X is any amino acid, which coordinates Mn2+ within the crystal structure. Not only do TarA homologs lacks the DX(D/E) motif that has been observed in class GT-A and GT-D enzymes, both the TarA enzyme from *S. aureus* and the TagA enzyme from *B. subtilis* have been shown to catalyze metal ion-independent glycosyltransferase activity [12,13].

In addition, the TarA glycosyltransferase lacks key features present within GT-A, GT-B, and GT-C enzymes. In contrast to GT-A enzymes that frequently contain a semi-conserved Asp-X-Asp motif for metal ion or ribose coordination, TarA and homologous enzymes lack an Asp-X-Asp motif and have been experimentally shown to have metal ion-independent activity. TarA is also structurally unrelated to enzymes within class GT-B and GT-C. The latter enzymes adopt multi-domain structures, two Rossmann folds for GT-B class and an integral membrane and variable, soluble domain for GT-C. Conversely, the crystal structure of TarA indicates a compact,

globular fold with a single domain. In GT-B and GT-C enzymes, the catalytic reaction center is located at the interface between separate domains. It is tempting to speculate that oligomerization of TarA could result in a competent active center, as this enzyme has been shown to self-associate experimentally [14].

4.3.3 TarA contains a C-terminal amphipathic helix that may mediate membrane association

The C-terminal region of TarA (residues G185-R248) is highly conserved (Figure 4.7.3A). Successive truncation of the C-terminus between putative secondary structure elements that are predicted by the PSI-PRED server identified a soluble and stable extracytoplasmic domain (Figure 4.7.3A, green box). Closer inspection of the truncated C-terminal region identified a putative membrane anchoring feature (Figure 1a, blue box). Helical wheel projections of the amino acid sequence within the C-terminal alpha helix indicated amphipathic character (Figure 4.7.3AB). Presumably, the amphipathic helix submerges its hydrophobic face within the lipid bilayer, positioning the positively-charged helical face adjacent to the negatively-charged phosphate head groups at the surface of the lipid bilayer. Consistent with a membrane targeting role, truncation of the C-terminal amphipathic helix was sufficient to solubilize the TarA construct during protein purification compared to wild-type protein (Figure 4.7.3C). Indeed, a similar mode of membrane association through C-terminal amphipathic helices has been observed for other glycosyltransferases [15]. Future studies will seek to determine the role of the C-terminus in membrane anchoring.

4.3.4 TarA oligomerization presumably promotes the formation of a competent active site

The apo- and ligand-bound forms of TarA adopted different oligomeric states during crystallization. The apo-TarA formed a dimeric species, in which the monomer is simply rotated 180° to form a large, symmetrical interface (Figure 4.7.4). The dimerization interface spans the length of H2, H4, and H10, as well as the edges of β1, H4, and several loops (β5/H7, H7/ β6, β6/H8, β7/H9). Visual inspection of the dimer reveals that a total of 31 residues contribute to the interface (I7, A41, **E42**, V44, M45, Q48, K49, S68, **G69**, I70, F72, A73, K75, V76, **R84**, A86, F88, **D89**, L92, A111, L125, Y138, G164, A165, **G188**, **G189**, **D192**, V193, I194, A195, and **G196**), of which eight residues are highly conserved according to Consurf analysis (highlighted in bold). The dimerization interface is large, with a total buried surface area is 1226 A^2.

During crystallization, the ligand-bound TarA formed a trimeric species, in which the monomer is rotated 120° about a single axis. Interactions that promote trimer formation occur at two distinct contact sites between TarA subunits: 1) the first contact site is formed by an interface between monomer-1 (F88, L92, N121, L125, **D192**, V193, I194, and A195) and monomer-2 (A111, A112, V114, Y138, and **D192**), and 2) a second contact site is formed between monomer-1 (V71, F72, K75, L81, P82, E83, and **R84**) and monomer-2 (**E42**, M45, M46, S47, E48, Q48, K49, D50, Y53, **P166**, and K170). The interface between two monomers within the trimeric structure produces an average total buried surface area of 468 A^2. In both the dimeric and trimeric forms of TarA, highly conserved residues, which are presumably important for function of the enzyme, cluster at the oligomerization interfaces, shown in dark red (Figure 4.7.4). It is possible that oligomerization of TarA is required to form a competent active site. In support of this notion, GT-B and GT-C enzymes contain active site centers located at the

interface of subunits within a continuous polypeptide chain. Furthermore, glycosyltransferases have been shown to oligomerize to expand their functional diversity and capabilities, supporting a potential functional role for TarA oligomerization.

The TarA oligomerization observed in the apo- and UDP-bound crystal structures is not simply an artifact of crystallization. *T. italicus* TarA primarily adopts dimeric and trimeric forms in solution, as determined from size exclusion chromatography of purified protein compared to molecular weight standards (Figure 4.7.4). Furthermore, the two dimensional (2D) ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum of the *T. italicus* TarA extracytoplasmic domain supports the idea that oligomerization occurs in solution (Figure 4.7.4C). The ^{1}H-^{15}N HSQC spectrum yielded reasonably well-resolved cross peaks, indicating that the TarA protein was folded. However, there were substantially fewer peaks than anticipated for the molecular weight of the TarA construct (21.7 kDa). In particular, ~50 peaks were absent in the NMR spectrum; only ~140 resolvable cross peaks from backbone amides were observed, whereas 187 cross peaks are expected (195 total residues – 7 proline residues – the N-terminal residue). The reduced number of signals in the ^{1}H-^{15}N HSQC spectrum is compatible with the large oligomerization interface observed in the TarA crystal structures, causing their signals to be broadened. However, it is also conceivable that the observed NMR line-broadening is caused by intra-protein motions that do not involve oligomerization.

4.3.5 The putative substrate binding mode and catalytic mechanism of TarA

The crystal structure of TarA from *T. italicus* represents the first protein structure for the glycosyltransferase WecB/TagA/CpsF family (PFAM03808; CAZy GT26). These enzymes are predicted to be inverting GTs, which catalyze their glycosyltransferase reaction through an

oxocarbenium ion-like transition state [5]. The transfer reaction is aided by a catalytic base, which is typically an aspartic or glutamic residue, to increase nucleophility of the nucleophilic hydroxyl group that attacks the acceptor [9].

First, we co-crystallized the TarA enzyme with UDP and ManNAc ligands. During structure refinement, electron density was apparent for the uracil nucleotide of the UDP ligand, but no electron density was observed for ManNAc. Walker et al. has determined that the TagA enzyme from *B. subtilis* follows a steady-state ordered Bi-Bi mechanism, in which it first binds UDP-ManNAc followed by lipid α. As catalysis proceeds it then sequentially releases the lipid β and UDP products. This data suggests that the UDP-bound crystal structure that has been determined corresponds to a late-stage structural snapshot of the reaction coordinate (Figure 4.7.5). It is possible that TarA undergoes conformational changes to coordinate the UDP-ManNAc substrate and that these necessary contacts are not present in the current crystal form, preventing ManNAc binding. The TarA-UDP complex identified a putative nucleotide binding site that is located at a surface formed by the edges of the β5, β6, β7 strands. Specifically, the uracil nucleotide of the UDP ligand is located proximal to Tyr138 (β5), Leu163 (β7), and Asp192 (H10) (Figure 4.7.5, yellow residues) ; however, proximal histidine residues that could stack against the pentose sugar ribose of UDP are not observed, as compared to other glycosyltransferases [6]. Electron density was not observed for the pyrophosphate moiety; furthermore, a surface electrostatics representation of TarA indicates limited positively charged regions (i.e. Arg, Lys, and His residues) that would be capable of coordinating the negative charge of the pyrophosphate. Thus, it is unclear whether UDP is bound to the TarA enzyme in a biologically relevant manner. Biochemical studies will need to be performed to resolve this issue.

To identify additional functional residues, we compared sequence conservation of TarA with 500 homologs using the Consurf server. Consurf analysis identified several highly conserved residues, including Thr38, Asn40, Glu42, Asp66, Arg84, Asp89, Gln168, and Glu169 (Figure 4.7.5, green, blue, and black residues). The majority of the highly conserved residues are located within the core of the TarA crystal structure. Interestingly, the Asp66 residue is entirely invariant among all homologs and is located at the base of a loop containing helix H4 and a stretch of ten unstructured amino acids. Limited proteolysis studies have indicated that this loop is susceptible to proteolytic cleavage (Figure 4.7.5). The proteolysis cleavage site was identified by mass spectrometry (data not shown) and is adjacent to a highly conserved Arg84 residue (Figure 4.7.5, black). Interestingly, the loop is protected from proteolytic cleavage when incubated with UDP and ManNAc ligands, suggesting that it is involved in UDP-ManNAc substrate binding. However, the loop becomes fully cleaved over extended exposure to protease, supporting its flexible nature. We speculate that the protease sensitive loop may function as lid that could undergo transitions between opened and closed conformational states upon substrate binding, which has been observed for other glycosyltransferases [11].

Sequence conservation analysis identified several residues that could function as a catalytic base to assist during the inverting GT mechanism of TarA. The location of highly conserved aspartic and glutamic residues, Glu42, Asp66, Asp89 and Glu169, is shown in Figure 4.7.5. Interestingly, only Asp89 is located adjacent to a region of negatively charged electrostatic potential to support an oxocarbenium ion-like transition state. However, the Asp89 residue is positioned at the outer surface of the TarA monomer, pointing away from the cluster of highly conserved amino acids at the core. Furthermore, no obvious nucleophile is located adjacent to these highly conserved aspartic and glutamic residues, requiring co-crystallization with

substrates to gain clearer insight into putative nucleophile residues and the mechanism of catalysis. However, these observations suggest a likely requirement for conformational rearrangements or oligomerization of TarA to support its catalytic activity.

The oligomerization patterns of TarA prompt several hypotheses for further exploration. Both the dimeric and trimeric forms of TarA adopt structures that produce a central pore. In the TarA dimer, the central pore is shallow and largely hydrophobic, suggesting that this channel may promote binding of the lipid-sugar acceptor substrate. The protease-susceptible loop is located adjacent to this hydrophobic channel and may function as a "lid" that secures the C55-GlcNAc acceptor within the active site. Alternatively, the TarA trimer produces an elongated central pore with one hydrophobic and one negatively-charged opening. Interestingly, the C-termini (Gly196) of each TarA monomer cluster to a common face within the trimeric species. This face of the central pore of the trimer complex presumably contacts the lipid bilayer, as it is proximal to the C-terminal membrane associated region of TarA, to act as an entry site for the lipid substrate. In support of this model, the highly conserved C-terminal sequence that is truncated in the TarA$^{\Delta C}$ construct contains a cluster of positively-charged residues followed by a semi-hydrophobic stretch of residues with a putative leucine zipper motif. It is possible that this region forms a leucine zipper-mediated, coiled coil within the cell membrane to coordinate the lipid of the C55-PP-GlcNAc substrate, as such formations have been observed with other membrane proteins [16,17]. It is likely that TarA association with the membrane could trigger further structural changes for substrate recognition and catalysis.

4.4 Conclusion

In conclusion, we report the 1.8 Å resolution crystal structure of the TarA enzyme from *Thermobacter italicus* (*T. italicus*). The TarA glycosyltransferase adopts a structurally novel protein fold that is distinct from the previously characterized GT-A, GT-B, GT-C, and GT-D classes. This novel "GT-E" fold is formed by a compact primary sequence that adopts an eight β strand core surrounded by several accessory helices. Furthermore, the *T. italicus* TarA enzyme is the first structurally characterized member of the WecB-TagA-CpsF GT family, providing insight into how members of the Carbohydrate-Active Enzymes (CAZy) GT-26 family perform their glycosyltransferase activity.

The *T. italicus* TarA enzyme readily oligomerizes, forming both dimers and trimers in solution. These oligomeric forms of TarA are likely to be physiologically relevant. In fact, sequence conservation mapping onto the TarA dimer and trimer complexes indicates that highly conserved residues cluster at the interfaces of these oligomeric forms, suggesting that oligomerization of TarA may be required to form a competent active site. Further conformational rearrangement are likely to occur for substrate binding, as limited proteolysis studies indicate the presence of a protease-susceptible loop that is stabilized upon the addition of UDP and ManNAc ligands. The co-crystal structure of UDP-bound TarA is nearly identical apo-TarA with an RMSD of 0.18 Å; however, electron density for ManNAc was not identified during refinement, suggesting that TarA must adopt an alternate conformation to stabilize binding of the ManNAc moiety. Sequence conservation analysis identified putative functional residues for substrate binding and catalysis, including several conserved aspartic and glutamic residues that may act as a catalytic base for nucleophile activation. However, a detailed mechanism of substrate binding, oxocarbenium ion-like transition state formation, and nucleophile-mediate catalysis remains to

be elucidated by TarA co-crystallization studies with its UDP-ManNAc and lipid ligands, as well as biochemical and cellular studies.

4.5 Materials and methods

4.5.1 Amphipathic helix identification and construct design

Secondary structure of TarA homologs was predicted using the PSI-PRED server. Helical wheel projections were generated with (SERVER) using amino acid sequences from each individual, putative alpha helix, according to the PSI-PRED prediction. Helical wheel projections demonstrating an amphipathic distribution of hydrophobic residues versus charged and polar residues were recorded. Several C-terminal truncation constructs for *S. aureus* and *T. italicus* TarA were designed by truncating in between the predicted secondary structure boundaries of the C-terminal amphipathic helices.

4.5.2 Cloning, expression, and protein purification

The extracytoplasmic domain of TarA from *T. italicus* (TarA$^{\Delta C}$, residues M1-G195) or *S. aureus* (Saur- TarA$^{\Delta C}$, residues A10-A204) was expressed from a pMAPLe4 plasmid in *Escherichia coli* BL21(DE3) cells. Standard methods were employed, with cultures grown in the presence of kanamycin at 37°C until an OD$_{600}$ of 0.6-0.8 was reached. Protein expression was then initiated by adding 100 µM isopropyl-β-D-1-thiogalactopyranoside (IPTG) followed by overnight protein expression at 18°C. A four liter cell culture was harvested by centrifugation and re-suspended in 40 mL of lysis buffer (50 mM Tris, pH 7.5; 500 mM NaCl; 40 mM CHAPS) that contained 400 µl of protease inhibitor cocktail (Calbiochem) and 2 mM phenylmethanesulfonylfluoride (PMSF). The cells were then lysed using an EmulsiFlex high pressure homogenizer (Avestin). Cell lysates were fractionated by centrifugation and the soluble

portion applied to a gravity column containing 10 mL of suspended His-Pure Co^{2+} resin (Life Technologies) pre-equilibrated with lysis buffer. The resin was sequentially washed with 20 mL aliquots of lysis buffer that contained 0, 25, and 50 mM imidazole. His-tagged $TarA^{\Delta C}$ was then eluted using 500 mM imidazole, and the fractions were pooled and concentrated using an Amicon Ultra-15 centrifugal filter (Millipore). To remove His_6-tag from the protein, TEV protease was added to $TarA^{\Delta C}$, and the solution was dialyzed in a 3.5 kDa MWCO Slide-A-Lyzer dialysis cassette (ThermoFisher) against dialysis buffer (50 mM Tris, pH 7.5; 200 mM NaCl) at 4 °C overnight. TEV was then separated from $TarA^{\Delta C}$ by binding 10 mL of suspended His-Pure Co^{2+} resin (Life Technologies) pre-equilibrated with dialysis buffer; cleaved TarA was recovered from the TEV-bound resin by flow through a gravity column. $TarA^{\Delta C}$ lacking the His_6-tag was further purified by gel filtration chromatography using a Sephacryl size-exclusion column (GE Healthcare Life Sciences) equilibrated in 50 mM Tris, pH 7.5 and 200 mM NaCl. Purified $TarA^{\Delta C}$ was then pooled, concentrated to 55 mg/mL, and stored at 4 °C.

Selenomethionine (SeMet) or ^{15}N-labeled $TarA^{\Delta C}$ protein was prepared with cultures grown in M9 minimal media in the presence of kanamycin at 37°C until an OD_{600} of 0.6-0.8 was reached. Protein expression was then initiated by adding 100 µM isopropyl- β-D-1-thiogalactopyranoside (IPTG) followed by overnight protein expression at 18°C. Protein purification was achieved as described above.

4.5.3 Crystallization, data collection, and structural determination

Recombinant $TarA^{\Delta C}$ at a concentration of 50 mg/mL in 50 mM Tris, pH 7.5; 200 mM NaCl was used for crystal screening. Screening used the JCSG+ broad matrix suite (Molecular

Dimensions) at room temperature in a sitting-drop vapor diffusion format (200 nl drop size). SeMet-labeled protein crystals grew over the course of three days in the presence of 200 mM lithium sulfate, 100 mM phosphate citrate, pH 4.2, and 20% PEG 1000. For X-ray data collection, TarA crystals were cryoprotected using reservoir solution containing 35% glycerol. Diffraction data sets were collected at the Advanced Photon Source (APS) beamline 24-1D-C equipped with a Pilatus-6M detector. All data were collected at 100 K. Data were collected at the detector distance of 300 mm, with 0.25° oscillations, and at a 0.9791 Å wavelength. Multi-wavelength anomalous dispersion (MAD) experiment was collected at peak (12663.0 eV), inflection (12660.3 eV), and high remote (12763.0 eV) energy wavelengths.

The TarA$^{\Delta C}$ crystals diffracted X-rays to 1.8 Å resolution. The XDS/XSCALE package was used to index, integrate and scale data in P2$_1$ space group [18]. The asymmetric unit of the crystal contained eight protein molecules, yielding a Matthews coefficient of 2.13 Å/Da and a 42.1% solvent content in the unit cell. The SHELX suite was used to locate the heavy atom substructure, which identified a total of 56 selenium atom sites. The quality of the phases calculated with the peak, inflection, and high remote energy diffraction datasets were improved using SHARP and the wARP suite (Global Phasing Limited). The heavy atom parameters were refined with MLPhare using the CCP4i suite [19]. Density modification and Non-crystallographic symmetry averaging was performed with the CCP4i suite to improve the quality of the electron density map. Automated model building was performed with BUCCANEER, followed by refinement with BUSTER [20]. Modeling of the additional electron density was confirmed using 2F$_o$-F$_c$ omit maps generated using BUSTER [20]. Complete refinement and structure statistics are reported in Table 1.

A second crystal form was produced with recombinant TarA$^{\Delta C}$ in the presence of UDP and ManNAc ligands. TarA$^{\Delta C}$ at a concentration of 45 mg/mL in 50 mM Tris, pH 7.5; 200 mM NaCl; 10 mM UDP; 10 mM ManNAc was used for crystal screening with the JCSG+ broad matrix suite (Molecular Dimensions) as described above. TarA$^{\Delta C}$-ligand co-crystals grew over the course of two days in the presence of 200 mM calcium acetate, 100 mM sodium cacodylate, pH 6.5, and 40% PEG 300. A single wavelength diffraction dataset for a non-cryoprotected crystal was collected at the Advanced Photon Source (APS) beamline 24-1D-C equipped with a Pilatus-6M detector as described above. The crystals diffracted X-rays to 2.9 Å resolution. The XDS/XSCALE package was used to index, integrate and scale data in P2$_1$ space group [18]. The asymmetric unit of the crystal contained six protein molecule, yielding a Matthews coefficient of 3.28 Å/Da and a 62.5% solvent content in the crystal unit cell. The PHASER program was used for molecular replacement, employing the coordinates of the apo-TarA$^{\Delta C}$. Molecular replacement yielded a single solution, which was refined in iterative runs using Buster software. Additional electron density resembling the UDP ligand was observed using 2F$_o$-F$_c$ omit maps generated by BUSTER. Complete refinement and structure statistics are reported in Table #.

4.5.4 ^{1}H-^{15}N heteronuclear single quantum coherence spectrum of TarA$^{\Delta C}$

^{15}N-labeled TarA for NMR studies was concentrated to 800 µM in NMR buffer (50 mM NaPO4, pH 6.8; 200 mM NaCl, 7% D$_2$O). ^{1}H-^{15}N HSQC spectra were acquired with 64 scans at 298 K on Bruker 800 MHz spectrometers equipped with a triple-resonance cryogenic probe.

4.5.5 Sequence conservation analysis

Multiple sequence alignment was generated with the ClustalOmega server [21]. Sequence conservation mapping onto the apo- and ligand-bound TarA$^{\Delta C}$ crystal structures was performed with the Consurf server [22].

4.5.6 Limited proteolysis

Recombinant Saur-TarA$^{\Delta C}$ lacking the His$_6$-tag was diluted to 2 mg/mL with dilution buffer (50 mM Tris, pH 7.5, 300 mM NaCl) and incubated at room temperature at a 500:1 ratio of Saur-TarA$^{\Delta C}$ to trypsin protease (Sigma). Aliquots of the cleavage reaction were quenched after 0, 5, 10, 15, 30, and 60 minutes by diluting 1:1 with 2X SDS-loading dye (100 mM Tris-HCl, pH 6.8; 4% (w/v) sodium dodecyl sulfate; 0.2% (w/v) bromophenol blue; 200 mM dithiothreitol) and heating at 100°C for 5 min. Quenched samples were analyzed by SDS-PAGE with Pre-stained Benchmark Protein Ladder (Invitrogen), which was stained with Coomassie to visualize proteolysis.

4.6 Figures
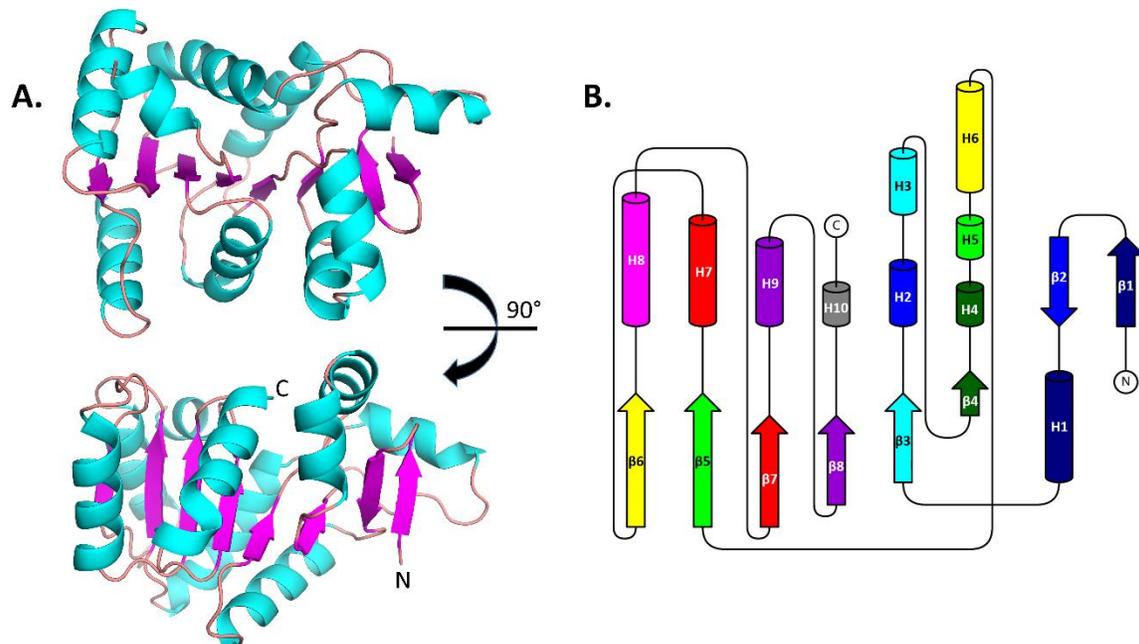
Figure 4.6.1 The crystal structure of the TarA glycosyltransferase. A. The crystal structure of TarA$^{\Delta C}$. The TarA$^{\Delta C}$ monomer is shown in cartoon representation with helices in cyan, sheets in magenta, and loops in salmon. N- and C-termini are labeled accordingly. B. Topology diagram of the secondary structure elements within the TarA$^{\Delta C}$ crystal structure. Helices are shown as arrows and sheets are shown at cylinders. Helices and sheets are independently colored progressing from N- to C-terminus.

A.

GT-A

GT-B

GT-C

GT-D

GT-E
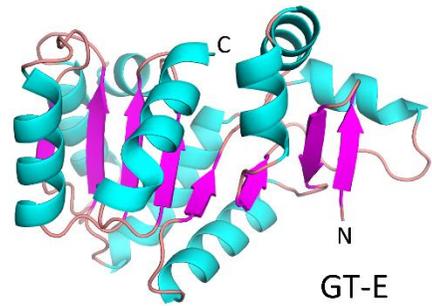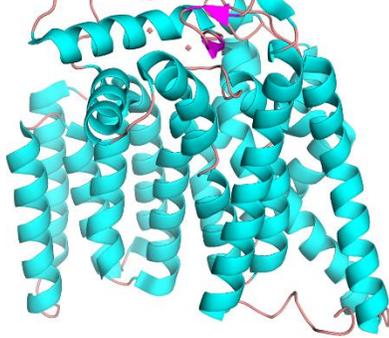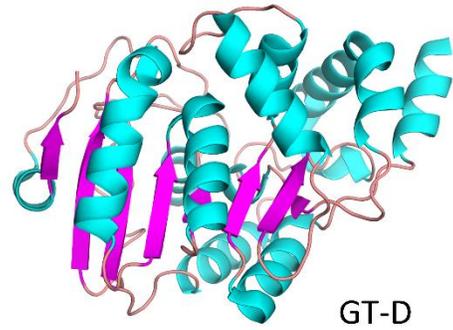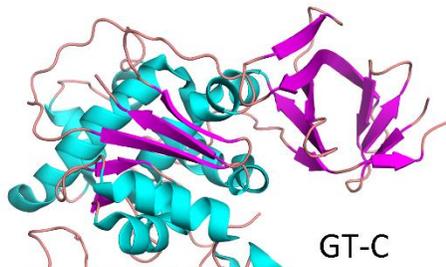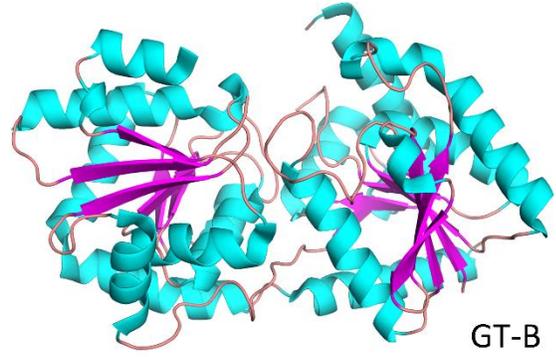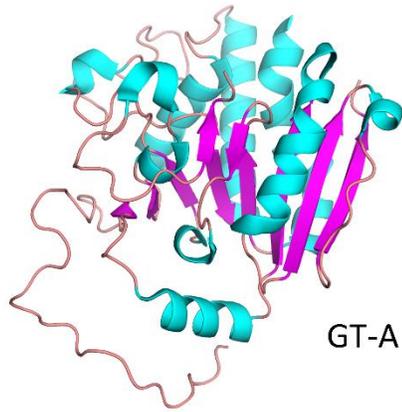
```
B.   GT-D       1 SEFASMKRLSEIKVLPILESLKYIKHNHASVVRFGDGEIDLMTGHSIPYQDYNEKLAKRL
     Bsubtilis  1 -------------------------------------MQTETIHNIPYV----------
     Titalicus  1 --------------------------------------MERLDIFGVPID----------
     Saureus    1 ----------------------------MTVEERSNTAKVDILGVDFD----------


     GT-D       61 QQILQTKSDEKLLVCLPDVFSNMDRYNQNARHFWERHFLKYSEFYLNCCDAPFYGSTFIS
     Bsubtilis  13 ------------------NSNLTSFIDYLEKHYIDQK---------------I-GAVIS
     Titalicus  13 -------------------RVTMIQAVDILNNFLQENR---------------L--HIVA
     Saureus    21 ------------------NTTMLQMVENIKTFFANQS---------------TNNLFIV


     GT-D       121 RPYIDLID-KSPSEAYFESLKELWRGKDLLIVEGATSRSGVGNDLFVAASSIKRLVCPSK
     Bsubtilis  38 TVNPEIAFAAIKDRDYFDVLS----SSNFILPDG--------IGVVMMSRLTNNRLQSRI
     Titalicus  37 TPNAEIVMMAQKDKEYMEILN----NTDLNVPDG--------SGIVFASKVFKKPLPERV
     Saureus    47 TANPEIVNYATTHQAYLELIN----QASYIVADG--------TGVVKASHRLKQPLAHRI


     GT-D       180 NAFQYYDEILRLTEKNAKNRLILVMLGPTAKVLVADLT--TKGYQAIDLG----------
     Bsubtilis  86 AGYDVFKELLGVANKKKK---RIFLYGAKKDVIKGVVSKISSEYPNIKIAGYSDGYVQDR
     Titalicus  85 AGFDLMLEFIKGISSKGV---KIYLLGAACQVAEQARANLEKLYPGVKIVGTHHGYFTEE
     Saureus    95 PGIELMDECLKIAHVNHQ---KVFLLGATNEVVEAAQYALQQRYPNISFA-HHHGYIDLE


     GT-D       228 ----------HIDSEYEWYEMGATYKVKLTNKHTAEFNYDEGIELEFSQEYQEQ------
     Bsubtilis  143 --TLVAKQIARANPDMVFVALGYPHQEKFIHNYRHLFPKAVSIGLGGSFDVFSGNVKRAP
     Titalicus  142 EENKIIEEINNKGAEVLFVALGAPKQEKWIYKNKDKLKVKIAMGVGGSFDVIAGKVKRAP
     Saureus    151 -DETVVKRIKLFKPDYIFVGMGFPKQEEWIMTHENQFESTVMMGVGGSLEVFAGAKKRAP


     GT-D       272 -IVARIG--------------------------------------------------
     Bsubtilis  201 SWMIRLNLEWFYRLILNPWRWKRMLSIPKYALTVLKEEKNKKTFYPKPEKDHTKQI
     Titalicus  202 YIYRKLGLEWLYRLIKEPWRYKRMMALPKFAIKVLLHKREVVR-------------
     Saureus    210 YIFRKLNIEWIYRALIDWKRIGRLKSIPIFMYKIAKAKRKIKKAK-----------
```

Figure 4.6.2 Structural comparison to GT-A, GT-B, GT-C, and GT-D classes reveals that TarA contains a novel GT-E glycosyltransferase fold. A. GT-A class representative, GTB (PDB ID: 2RIT); GT-B class representative, MshA (PDB ID: 3C4V); GT-C class representative, PglB (PDB ID: 3RCE); GT-D class representative, DUF1792 (PDB ID: 4PFX); GT-E class representative, *T. italicus* TarA. All structures are shown in cartoon representation with helices colored in cyan, β-strands colored in magenta, and loops colored in salmon. B. ClustalOmega sequence alignment of TarA homologs and the most structurally related enzyme identified by DALI analysis, the representative class GT-D glycosyltransferase (PDB ID: 4PFX). Conserved residues are indicated in red, and related amino acids are indicated in blue. Bold residues within the GT-D sequence are highly conserved and important for glycosyltransferase activity of the DUF1792 enzyme.

165

A.

```
Bsubtilis   1 --------MQTETIHNIPYVNSNLTSFIDYLEKHYIDQKI-GAVISTVNPEIAFAAIKDRDYFDVLSSS
Saureus     1 MTVEERSNTAKVDILGVDFDNTTMLQMVENIKTFFANQSTNNLFIVTANPEIVNYATTHQAYLELINQA
Titalicus   1 --------MERLDIFGVPIDRVTMIQAVDILNNFLQENRL--HIVATPNAEIVMMAQKDKEYMEILNNT
```

```
Bsubtilis  61 NFILPDGIGVVMMSRLTNNRLQSRIAGYDVFKELLGVANKKKKRIFLYGAKKDVIKGVVSKISSEYPNI
Saureus    70 SYIVADGTGVVKASHRLKQPLAHRIPGIELMDECLKIAHVNHQKVFLLGATNEVVEAAQYALQQRYPNI
Titalicus  60 DLNVPDGSGIVFASKVFKKPLPERVAGFDLMLEFIKGISSKGVKIYLLGAACQVAEQARANLEKLYPGV
```

```
Bsubtilis 130 KIAGYSDGYVQDR--TLVAKQIARANPDMVFVALGYPHQEKFIHNYRHLFPKAVSIGLGGSFDVFSGNV
Saureus   139 SFA-HHHGYIDLE-DETVVKRIKLFKPDYIFVGMGFPKQEEWIMTHENQFESTVMMGVGGSLEVFAGAK
Titalicus 129 KIVGTHHGYFTEEEENKIIEEINNKGAEVLFVALGAPKQEKWIYKNKDKLKVKIAMGVGGSFDVIAGKV
```

```
Bsubtilis 197 KRAPSWMIRLNLEWFYRLILNPWRWKRMLSIPKYALTVLKEEKNKKTFYPKPEKDHTKQI
Saureus   206 KRAPYIFRKLNIEWIYRALIDWKRIGRLKSIPIFMYKIAKAKRKIKKAK-----------
Titalicus 198 KRAPYIYRKLGLEWLYRLIKEPWRYKRMMALPKFAIKVLLHKREVVR------------
```

B.

S. aureus   *T. italicus*

C.
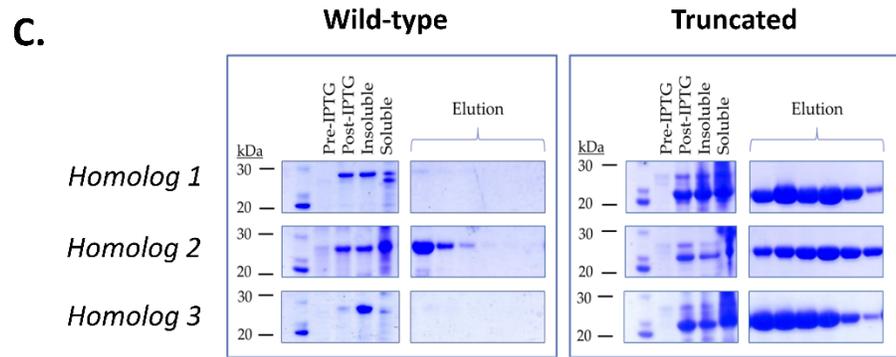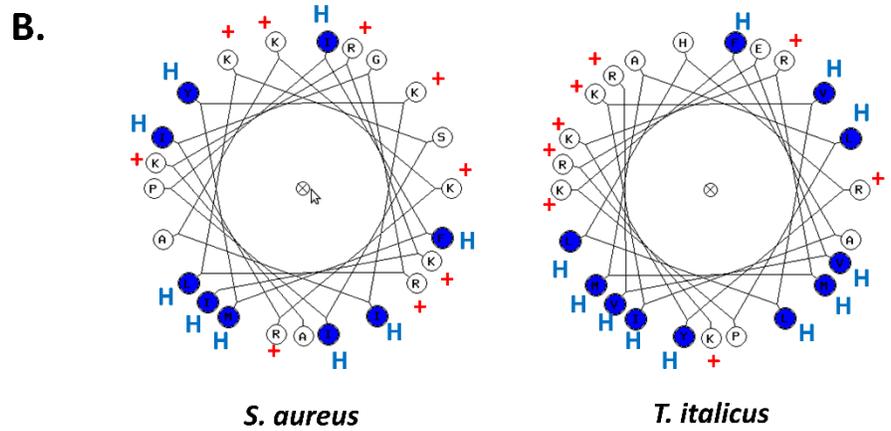
Figure 4.6.3 TarA contains a C-terminal amphipathic helix that mediates membrane association. A. Sequence alignment of TarA homologs generated using the ClustalOmega server [21]. The bacterial species and accession numbers of the amino acid

166

sequences used for the alignment are as follows: *Staphylococcus aureus* (NC_007795), *Bacillus subtilis* (NC_000964), *Thermobacter italicus* (NC_013921). Conserved residues are indicated in *red*, and related amino acids are indicated in *blue*. The extracytoplasmic domain is boxed in green, and C-terminal amphipathic helix, as determined by PSI-PRED secondary structure prediction, is boxed in blue. B. Helical wheel projections of the putative C-terminal alpha helix of TarA reveal amphipathic character. Hydrophobic and positively charged residues are labeled with H and a plus sign (+), respectively. C. C-terminal truncations solubilize the TarA enzyme. Coomassie-stained SDS-PAGE analysis of purifications of wild-type and C-terminal truncation constructs of TarA homologs under identical solubilization conditions.
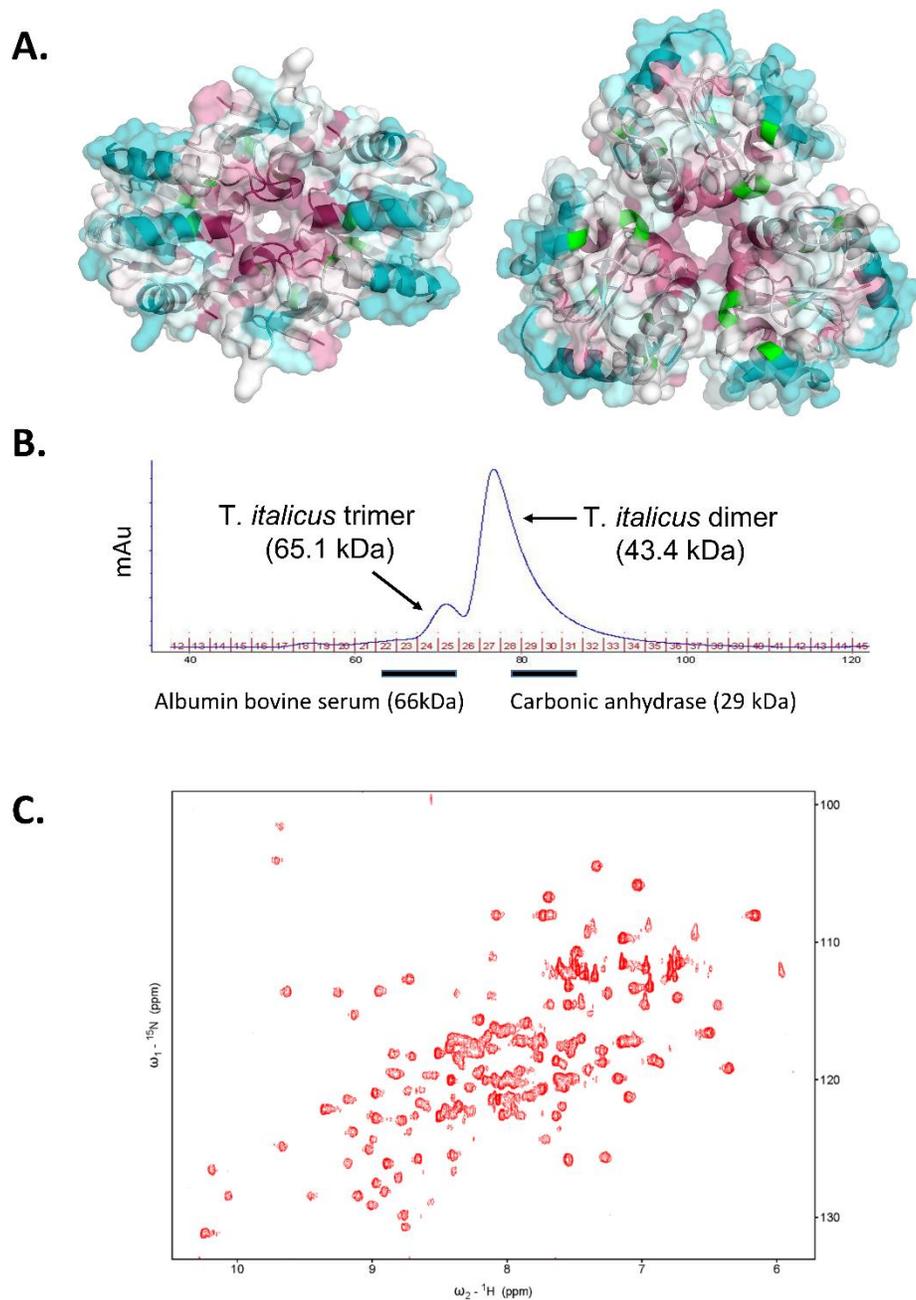
Figure 4.6.4 TarA oligomerization presumably promotes the formation of a competent active site. A. Sequence conservation mapping of TarA oligomers. Sequence conservation mapping was performed with the Consurf server. Surface and cartoon

representations of TarA dimeric and trimeric forms indicate highly conserved residues (dark red) and regions lacking sequence conservation (cyan); selenomethionine residues are shown in green. B. Gel filtration chromatograms of *T. italicus* TarA support dimer and trimer formation in solution. C. ^{1}H-^{15}N HSQC spectrum of the *T. italicus* TarA extracytoplasmic domain supports oligomerization in solution. The ^{1}H-^{15}N HSQC spectrum yielded reasonably well-resolved cross peaks, indicating that the TarA protein was folded. However, there were substantially fewer peaks than anticipated for the molecular weight of the TarA construct (21.7 kDa). In particular, ~50 peaks were absent in the NMR spectra; only ~140 resolvable cross peaks from backbone amides were observed, whereas 187 cross peaks are expected (195 total residues – 7 proline residues – the N-terminal residue). The reduced number of signals in the ^{1}H-^{15}N HSQC spectrum is compatible with the large oligomerization interface observed in the TarA crystal structures, causing their signals to be broadened.
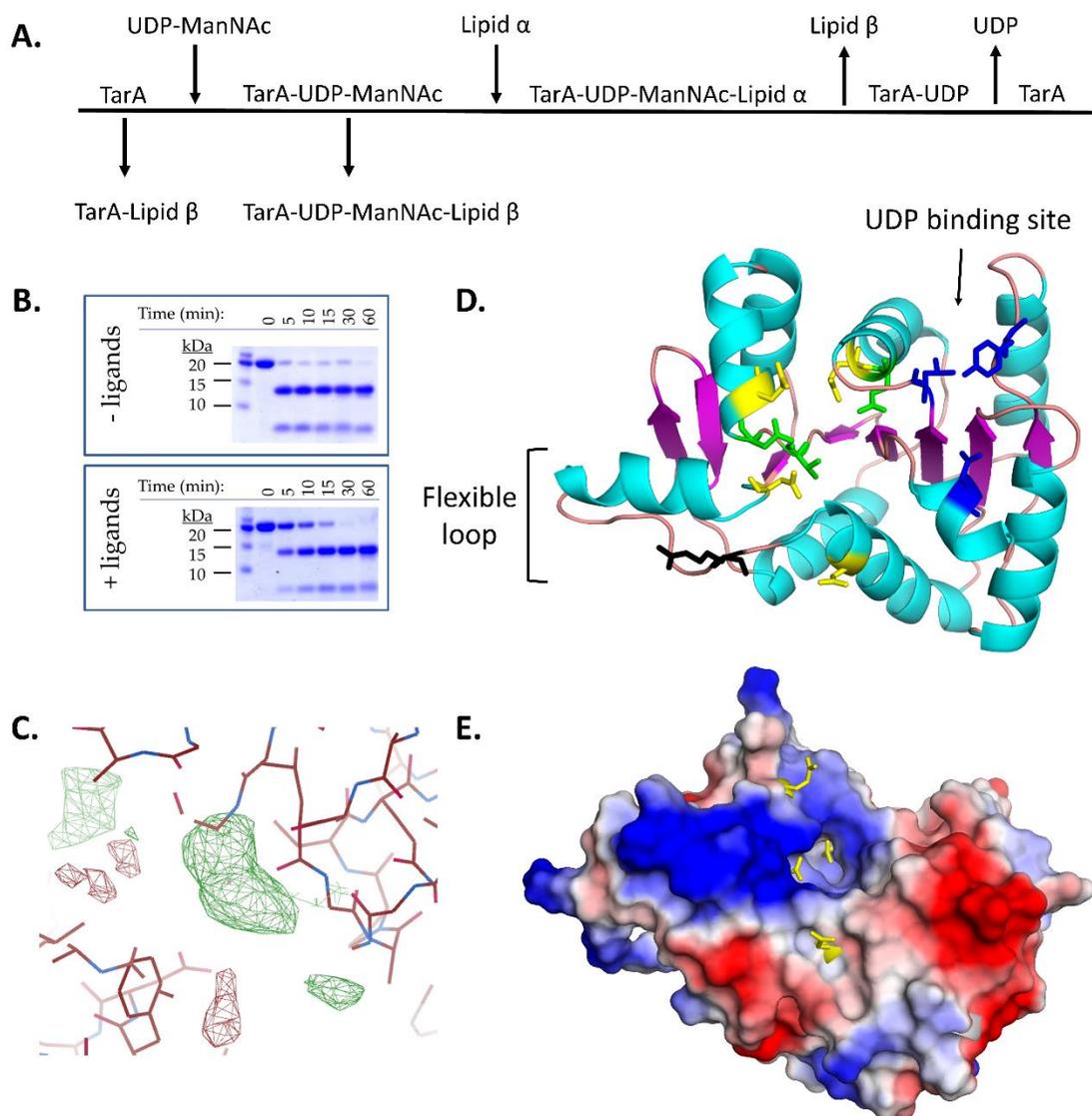
Figure 4.6.5 Dissecting the substrate binding mode and catalytic mechanism of TarA. A. Reaction coordinate of the TarA catalytic mechanism, as determined by Walker et al. [23]. B. Limited proteolysis indicates a protease susceptible loop that may be involved in ligand binding. The cleavage site was confirmed by mass spectrometry (data not shown) and is indicated on the TarA model in Figure 4.7.5D (black residue). C. F_o-F_c omit map contoured at +3

sigma. The difference map was generated by refining the ligand-bound TarA diffraction dataset against a model of the apo-TarA enzyme. Additional electron density resembling the chemical structure of the uracil moiety of UDP is observed. D. Cartoon representation of TarA highlighting important residues (shown in sticks). Residues adjacent to electron density for UDP are colored in blue (Tyr138, Leu163, and Asp192). Catalytic base candidates are shown in yellow (Glu42, Asp66, Asp89, and Glu169). Cleavage site from limited proteolysis studies is shown in black (Arg84). Additional highly conserved residues are shown in green (Thr38, Asn40, and Gln168). Helices, β-strands, and loops are colored in cyan, magenta, and salmon, respectively. E. Surface electrostatics of TarA with potential catalytic base residues indicated in yellow sticks (Glu42, Asp66, Asp89, and Glu169).

4.7 Tables

Table 1. **Data collection and structure refinement statistics.**

	TarA	TarA-UDP
Data collection		
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, b, c (Å)	53.1, 104.3, 79.0	64.9, 104.2, 90.1
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 108.0, 90.0
Resolution (Å)	43.53-2.00 (2.05-2.00)[a]	85.06-3.13 (3.22-3.13)
$I / \sigma I$	6.82 (2.0)	6.66 (2.04)
Completeness (%)	99.1 (98.9)	94.9 (95.8)
Redundancy	5.9 (4.7)	1.8 (1.8)
Wilson B-factor (Å^2)	25.40	78.42
Refinement		
Resolution (Å)	88.51-2.00	85.57-2.84
No. reflections	96634	23794
$R_{\text{work}} / R_{\text{free}}$ (%)[b]	22.3/25.0	25.6/27.9
No. atoms	12181	8279
Protein	11846	8279
Ligand/ion	0	0
Water	335	0
B-factors (all atoms)	24.80	78.42
Protein	26.3	78.42
Ligand/ion	0	0
Water	33.6	0
R.m.s. deviations		
Bond lengths (Å)	0.010	0.008
Bond angles (°)	1.05	1.07
Ramachandran favored (%)	97.10	96.06
Ramachandran allowed (%)	1.97	2.33
Ramachandran generally allowed (%)	0.0	0.0
Ramachandran outliers (%)	0.92	1.61

[a] Values in parentheses are for highest-resolution shell.
[b] R$_{\text{free}}$ calculated using 10% of the data.

4.8 Appendix

4.8.1 *T. italicus* TarA constructs

Construct Type	Name	Vector	Parameters
C-term truncation	M1-E218	pMAPLe4	Expresses in BL21 and soluble
	M1-G195	pMAPLe4	Crystal structure determined
	M1-D191	pMAPLe4	Unstable
	M1-G187	pMAPLe4	Unstable
	M1-I182	pMAPLe4	Does not express in BL21; unstable
Linker deletions (C-terminal residue)	ΔK77-V84 (E218)	pMAPLe4	Not tested
	ΔK77-V84 (G195)	pMAPLe4	Expresses and purified; unstable
	ΔK77-V84 (I182)	pMAPLe4	Does not express well in BL21; unstable
	ΔK77-K100 (E218)	pMAPLe4	Not tested
	ΔK77-K100 (G195)	pMAPLe4	Unstable
	ΔK77-K100 (I182)	pMAPLe4	Not tested (likely unstable)
Isolated N-domain	M1-K77	pMAPLe4	Does not express in BL21
	M1-V84	pSUMO (kan)	Unstable
	M1-K100	pMAPLe4	Does not express in BL21
	M1-G101	pSUMO (kan)	Unstable
	F76-E218	pSUMO	Not tested (likely unstable)

Isolated C-domain	F76-G195	pSUMO	Expressed and purified - soluble (appears to still contain protease-susceptible linker)
	V84-G195	pSUMO	Unstable
	F76-I182	pSUMO	Not tested (likely unstable)
	K100-E218	pMAPLe4	Not tested (likely unstable)
	K100-G195	pMAPLe4	Weakly expressed in BL21
	K100-I182	pMAPLe4	Not tested (likely unstable)

4.8.2 *S. aureus* TarA constructs

Construct Type	Name	Vector	Status
C-term truncation	A10-D226	pMAPLe4	Expresses and soluble
	A10-A204	pMAPLe4	Expresses and soluble
	A10-V192	pMAPLe4	Does not express in BL21/unstable
Linker deletion	ΔK87-I94 (D226)	-	Not cloned
	ΔK87-I94 (A204)	pMAPLe4	Unstable
	ΔK87-I94 (V192)	pMAPLe4	Not tested (likely unstable)
	ΔK87-V109 (D226)	pMAPLe4	Not tested (likely unstable)
	ΔK87-V109 (A204)	pMAPLe4	Does not express in BL21; unstable
	ΔK87-V109 (V192)	pMAPLe4	Not tested (likely unstable)
N-domain	A10-K87	pMAPLe4	Does not express in BL21
	A10-P95	pSUMO (kan)	Unstable
	A10-H111	pMAPLe4	Does not express in BL21
	A10-H111	pSUMO (kan)	Unstable
C-domain	N100-V192	pMAPLe4	Not tested (likely unstable)
	N100-A204	pMAPLe4	Unstable

4.9   References

1.   Fong DH, Yim VC-N, D'Elia MA, Brown ED, Berghuis AM. Crystal structure of
     CTP:glycerol-3-phosphate cytidylyltransferase from Staphylococcus aureus: Examination
     of structural basis for kinetic mechanism. Biochim Biophys Acta - Proteins Proteomics.
     2006;1764: 63–69. doi:10.1016/j.bbapap.2005.10.015

2.   Koç C, Gerlach D, Beck S, Peschel A, Xia G, Stehle T. Structural and Enzymatic Analysis
     of TarM Glycosyltransferase from *Staphylococcus aureus* Reveals an Oligomeric Protein
     Specific for the Glycosylation of Wall Teichoic Acid. J Biol Chem. 2015;290: 9874–9885.
     doi:10.1074/jbc.M114.619924

3.   Sobhanifar S, Worrall LJ, King DT, Wasney GA, Baumann L, Gale RT, et al. Structure
     and Mechanism of Staphylococcus aureus TarS, the Wall Teichoic Acid β-
     glycosyltransferase Involved in Methicillin Resistance. Zhang G, editor. PLOS Pathog.
     2016;12: e1006067. doi:10.1371/journal.ppat.1006067

4.   Lovering AL, Lin LY-C, Sewell EW, Spreter T, Brown ED, Strynadka NCJ. Structure of
     the bacterial teichoic acid polymerase TagF provides insights into membrane association
     and catalysis. Nat Struct Mol Biol. 2010;17: 582–9. doi:10.1038/nsmb.1819

5.   Albesa-Jové D, Guerin ME. The conformational plasticity of glycosyltransferases. Curr
     Opin Struct Biol. 2016;40: 23–32. doi:10.1016/j.sbi.2016.07.007

6.   Zhang H, Zhu F, Yang T, Ding L, Zhou M, Li J, et al. The highly conserved domain of
     unknown function 1792 has a distinct glycosyltransferase fold. Nat Commun. Nature
     Publishing Group; 2014;5: 147–157. doi:10.1038/ncomms5339

7.   Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The

carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42: D490-5. doi:10.1093/nar/gkt1178

8.   Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: Structures, Functions, and Mechanisms. Annu Rev Biochem. Annual Reviews ; 2008;77: 521–555. doi:10.1146/annurev.biochem.76.061005.092322

9.   Ardèvol A, Rovira C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from *ab Initio* Quantum Mechanics/Molecular Mechanics Dynamic Simulations. J Am Chem Soc. American Chemical Society; 2015;137: 7528–7547. doi:10.1021/jacs.5b01156

10.   Lazarus MB, Jiang J, Gloster TM, Zandberg WF, Whitworth GE, Vocadlo DJ, et al. Structural snapshots of the reaction coordinate for O-GlcNAc transferase. Nat Chem Biol. Nature Research; 2012;8: 966–968. doi:10.1038/nchembio.1109

11.   Tsutsui Y, Ramakrishnan B, Qasba PK. Crystal Structures of -1,4-Galactosyltransferase 7 Enzyme Reveal Conformational Changes and Substrate Binding. J Biol Chem. 2013;288: 31963–31970. doi:10.1074/jbc.M113.509984

12.   D'Elia MA, Henderson JA, Beveridge TJ, Heinrichs DE, Brown ED. The N-acetylmannosamine transferase catalyzes the first committed step of teichoic acid assembly in Bacillus subtilis and Staphylococcus aureus. J Bacteriol. American Society for Microbiology; 2009;191: 4030–4. doi:10.1128/JB.00611-08

13.   Yu-Hui Zhang ‡, Cynthia Ginsberg ‡,§, Yanqiu Yuan ‡,§ and, Suzanne Walker* ‡. Acceptor Substrate Selectivity and Kinetic Mechanism of Bacillus subtilis TagA†. American Chemical Society ; 2006; doi:10.1021/BI060872Z

14.   Formstone A, Carballido-López R, Noirot P, Errington J, Scheffers D-J. Localization and

interactions of teichoic acid synthetic enzymes in Bacillus subtilis. J Bacteriol. American Society for Microbiology; 2008;190: 1812–21. doi:10.1128/JB.01394-07

15.   Ma B, Wang G, Palcic MM, Hazes B, Taylor DE. C-terminal amino acids of Helicobacter pylori alpha1,3/4 fucosyltransferases determine type I and type II transfer. J Biol Chem. 2003;278: 21893–900. doi:10.1074/jbc.M301704200

16.   Lemmon MA, Treutlein HR, Adams PD, Brünger AT, Engelman DM. A dimerization motif for transmembrane α–helices. Nat Struct Biol. Nature Publishing Group; 1994;1: 157–163. doi:10.1038/nsb0394-157

17.   Ruan W, Becker V, Klingmüller U, Langosch D. The interface between self-assembling erythropoietin receptor transmembrane segments corresponds to a membrane-spanning leucine zipper. J Biol Chem. 2004;279: 3273–9. doi:10.1074/jbc.M309311200

18.   Kabsch W, IUCr, K. D, A. KP, K. D, S. M, et al. *XDS*. Acta Crystallogr Sect D Biol Crystallogr. International Union of Crystallography; 2010;66: 125–132. doi:10.1107/S0907444909047337

19.   Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr. International Union of Crystallography; 2011;67: 235–42. doi:10.1107/S0907444910045749

20.   Bricogne G. Direct phase determination by entropy maximization and likelihood ranking: status report and perspectives. Acta Crystallogr D Biol Crystallogr. International Union of Crystallography; 1993;49: 37–60. doi:10.1107/S0907444992010400

21.   Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol.

2011;7. doi:10.1038/msb.2011.75

22.   Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating

evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic

Acids Res. 2010;38: W529-33. doi:10.1093/nar/gkq399

23.   Zhang Y-H, Ginsberg C, Yuan Y, Walker S. Acceptor Substrate Selectivity and Kinetic

Mechanism of *Bacillus subtilis* TagA [†]. Biochemistry. 2006;45: 10895–10904.

doi:10.1021/bi060872z

Chapter 5

Cell-based, High-throughput Screen Development

for the Discovery of Small Molecule Inhibitors

5.1 Overview

Traditional antibiotic discovery campaigns have utilized phenotype-based, whole cell high-throughput screening (HTS) to achieve growth inhibition and cell death, as well as target-based approaches that monitor *in vitro* inhibition of purified protein in the presence of chemically diverse compound libraries [1,2]. Target-based methods have supported inhibitor discovery against essential targets that were identified through genome-wide essentiality screens under optimal growth conditions [1]. These discovery strategies have identified anti-microbial compounds that target several essential mechanisms: 1) cell wall biosynthesis, 2) protein synthesis, 3) RNA or DNA synthesis, 4) folate synthesis, or 5) disruption of the cell membrane. Although initially effective, the identified antibiotics are becoming obsolete, as bacteria are rapidly developing and mobilizing resistance mechanisms to circumvent the inhibited molecular targets. Discovery efforts in the current "Post-Antibiotic Resistance Era" are employing target-based, whole-cell approaches and unconventional methodologies, such as cytological profiling, to increase flux into the drug development pipeline [3].

5.2 Introduction

Resistant isolates of bacteria have been identified for nearly all traditional and last resort antibiotics, prompting the urgent need to identify novel antimicrobial compounds [4,5]. Compared to the limited success obtained with target-based, *in vitro* screening approaches, whole cell screening strategies have produced lead compounds with improved therapeutic properties, including barrier permeability and resistance to efflux [6]. The current era of antibiotic discovery is leveraging target-based, whole cell high-throughput screens (HTS) to identify inhibitors acting on defined bacterial pathways. Generally, two approaches can be

employed: 1) parallel screening against a wild-type strain and mutant strain lacking the gene of interest, and 2) parallel screening against wild-type strain and mutant strain overexpressing a selected target. Traditional reporters for inhibition include changes in cell viability, cell growth rate, and temperature sensitivity, among others. Conversely, alternate readouts, such as changes in cell morphology, and sophisticated cytological profiling approaches can provide more precise information about the mechanism of action or molecular target. Indeed, cell morphological and cytological screening platforms are gaining traction due to their robust HTS capability [7–10].

Antibiotic discovery efforts first require an assay that is robust and scalable, as to identify "hit" molecules from large chemical libraries with high confidence. The quality or suitability of the assay must be validated before a HTS can be implemented. Specifically, the HTS assay must have a dynamic range that enables accurate and rapid identification of active compounds or "hit" molecules. A screening window coefficient, called the Z-prime score, is calculated to report the dynamic range and data variation of signal measurements, and returns values between 0-1.0, with a score between 0.5-1.0 indicating a robust assay [11]. The Z-prime score (described below) is therefore representative of the quality of the HTS assay and can be used to optimize and validate the assay before completing a pilot screen and subsequent large-scale HTS of chemically diverse compound libraries.

5.2.1 Cell-based screen for sortase inhibitors with *Actinomyces oris*

Sortase enzymes are an attractive drug target, as several studies have shown that sortase-defective mutants in *S. aureus* exhibit reduced virulence in animal models of infection [12–19]. Small molecule inhibitors targeting sortase enzymes could function as potent, anti-infective agents that strip pathogens of virulence factors needed to establish an infection. In addition, sortases possess other desirable properties for drug development: 1) they have no human

homologs, reducing the likelihood of off-target effects, 2) they are located on the bacterial

surface such that inhibitors do not need to cross the cell membrane, and 3) they are not required

for growth of *S. aureus* and other clinically relevant microbes outside of the host [20].

To date, all screening efforts to discover sortase inhibitors have used a fluorescence

resonance energy transfer (FRET)-based assay that monitors the ability of the isolated enzyme to

hydrolyze a fluorescent peptide substrate *in vitro* [21,22]. Although many chemical libraries have

been screened in this format, no inhibitors generated from these efforts have entered into clinical

trials [23–25]. The low success rate is likely due to a failure by the *in vitro* assay to replicate key

features of the sortase mechanism, such as: 1) the purified enzyme is a truncation construct that

is removed from the physiologically relevant context of the bacterial membrane, and 2) the assay

utilizes a peptide fragment of the sortase substrate and does not fully reconstitute the anchoring

reaction. To overcome these problems, we have developed a novel and robust, cell-based HTS in

which sortase activity is tightly linked to microbial growth.

In contrast of the majority of bacterial species, activity of the SrtA enzyme in

Actinomyces oris (*A. oris*) is required for growth in cell culture [26]. The unique cell lethality of

A. oris upon SrtA inactivation is thought to be due to toxic accumulation of a glycosylated

protein, AcaC, within the cell membrane [26]. AcaC is exported across the membrane and

retained at the lipid bilayer via a cell wall sorting signal, where it is glycosylated by an LCP

enzyme. In wild type *A. oris*, SrtA catalyzes a transpeptidation reaction to covalently attach

glycosylated AcaC protein to the cell wall. In contrast, when SrtA activity is eliminated,

glycosylated AcaC accumulates in the membrane and presumably imparts envelope stress that

causes growth arrest, and ultimately, cell death. The novel, lethal phenotype of srtA deletion in

A. oris can be utilized for robust, high-throughput screening efforts to identify sortase-specific inhibitors.

5.2.2 Cytological profiling screen for WTA inhibitors with *Bacillus subtilis*

By far the most abundant polymer displayed on the surface of Gram-positive bacteria is WTA, an anionic glycopolymer that has critical functions in cell division, morphology, adhesion, and microbial susceptibility to the immune response [27,28]. The WTA biosynthetic pathway in *S. aureus* has drawn great interest as a drug target, as MRSA strains lacking this polymer are defective in host colonization and re-sensitized to β-lactam antibiotics [29,30]. Despite its fundamental importance and potential as a drug target, we known surprisingly little about the cytoplasmic membrane-associated enzymes in the *S. aureus* WTA biosynthetic pathway, and only a limited number of small molecule inhibitors against TarO and TarG have been reported, which has yet to be developed into viable drug options [30–35]. To dissect how WTA is constructed and to discover therapeutically relevant biosynthesis inhibitors, we will study the novel TarA glycosyltransferase.

TarA homologs catalyze the first committed step of WTA biosynthesis in Gram-positive bacteria [29]. Interestingly, WTA was long thought to be an essential component of the Gram-positive cell wall, and genetic disruption at several steps during the WTA biosynthetic pathway lead to cell death *in vitro* [36–39]. However, lethality of the essential, late-acting WTA genes (*tagBDFKLGH*) could be alleviated in strains that also lacked the early genes, *tagO* or *tagA*, revealing a remarkable complexity in the pattern of dispensability [40,41]. The disruption of TagO or TagA in *B. subtilis* and TarO or TarA in *S. aureus* yields strains that are completely devoid of WTA [29,40]. Furthermore, *tagO* or *tagA* deletion in *B. subtilis* also produces distinct

morphological defects, most notably, the shift from a wild-type, rod shape to a spherical

morphology [29,40]. To demonstrate that WTA linkage unit assembly is a highly conserved

process among Gram-positive bacteria, we complemented a *B. subtilis* strain lacking the

endogenous TarA enzyme (*tarA-*) with the *S. aureus* TarA enzyme, which will be implemented

in a cell-based HTS assay for WTA inhibitor screening.

5.3 Results and discussion

5.3.1 Cell-based screen for sortase inhibitors with *Actinomyces oris*

We have developed a cell-based assay to screen for sortase inhibitors that is based on the

unique growth dependence of *A. oris* on the activity of its SrtA enzyme. Lethality of *A. oris* due

to SrtA deletion can be overcome by concurrent deletion of the SrtA protein substrate, AcaC. We

first confirmed that growth inhibition of the wild-type and mutant (Δ*srtA*/Δ*acaC*) was a robust

assay reporter by calculating a Z-prime score. The wild-type and mutant *A. oris* strains produced

Z-prime scores of 0.75 and 0.67, respectively. These scores were calculated from 32 replicates of

positive and negative growth controls, in which positive control well contained media, cells, and

0.5% DMSO, and negative control wells that also contained chloramphenicol. The growth data

used to calculate the Z-prime scores for wild-type (solid symbols) and mutant (open symbols) are

shown in Figure 5.5.1. The Z-prime scores and their resultant calculated selective inhibition

(%SelectINH) values indicate that the assay is robust and suitable for high-throughput screening.

The validated, cell-based *A. oris* assay was implemented for high-throughput screening

against two libraries of FDA-approved small molecules (LOPAC and New Prestwick chemical

libraries). Preliminary growth analysis of the mutant *A. oris* strain eliminated 82 molecules,

which non-specifically inhibited the mutant strain with >20% change in growth compared to

controls. The selective inhibition distribution of the remaining molecules from the chemical libraries are shown in Figure 5.5.2. The tested compounds exhibit an average %SelectINH of -5.9% with a standard deviation of 10.6%. Within the pilot screen, a total of 0.3% of the molecules in the libraries are defined as primary hits, as they have %SelectINH values that are >3 standard deviations above the average (%SelectINH >25.8, red squares, Figure 5.5.2).

5.3.2 Cytological profiling screen for WTA inhibitors with *Bacillus subtilis*

To demonstrate that WTA linkage unit assembly is a highly conserved process among Gram-positive bacteria (Figure 5.5.3), we first disrupted the endogenous TarA enzyme within *B. subtilis* (TagA-). TagA deletion reproduced the drastic morphological shift observed with a similar strain engineered by Brown et al. [29] (Figure 5.5.4). Subsequent complementation of the *B. subtilis* TagA- strain with the *S. aureus* TarA enzyme re-established the wild-type *B. subtilis* morphology, corresponding with restored display of WTA polymer on the cell surface (Figure 5.5.4). Furthermore, the morphological dependency of *B. subtilis* on activity of the *S. aureus* enzyme yielded a robust HTS platform for TarA inhibitors. The dynamic range of the rod to sphere transition was analyzed, and the Z-prime calculation produced a score of 0.76, indicating that the TarA activity-dependent morphological shift supports a cell-based assay that is well-suited for HTS efforts.

5.4 Conclusions

In conclusion, we described two cell-based assays that have been validated for HTS efforts. The first cell-based assay leverages the novel, lethal phenotype of srtA deletion in *A.*

oris, which is produced by a stress response after toxic accumulation of the glycosylated SrtA substrate, AcaC, at the cell membrane. In this assay, wild-type and Δ*srtA*/Δ*acaC* double mutant *A. oris* strains are cultured in the presence of a small molecule, and differential growth inhibition of the SrtA+ (wild-type) and SrtA- (mutant) strains are monitored. The Z-prime score calculation using positive and negative controls demonstrated that the *A. oris* cell-based assay produces a suitable dynamic range for HTS. A pilot screen against a library of 1280 FDA-approved compounds yielded four hit molecules, which inhibited the SrtA+ strain with equal or superior potency compared to sortase inhibitors that were developed by the established *in vitro* approach. The hit rate of 0.3% and improved efficacy of the molecules identified by the novel, cell-based assay are promising and have prompted large-scale HTS of ~200,000 unique molecules to generate additional hit molecules for validation and lead development.

A second cell-based assay leverages the unique morphological dependency of *B. subtilis* on TarA glycosyltransferase activity. In this assay, a *B. subtilis* strain that has been complemented with the *S. aureus* TarA enzyme in the background of endogenous *tagA* deletion adopts the wild-type cell morphology, demonstrating the high conservation of TarA activity in Gram-positive bacteria. However, when the TarA glycosyltransferase activity is disrupted, correlating with a lack of WTA display on the cell surface, the *B. subtilis* cells adopt a spherical morphology. This drastic morphological shift was measured using Shape Factor analysis and yielded a Z-prime score of 0.76. Offering a robust dynamic range, the *B. subtilis* cell-based assay will be utilized in a pilot screen against FDA-approved molecules to complete proof of concept. Ultimately, this work has facilitated the discovery of novel sortase- and WTA-specific inhibitors for the treatment of Gram-positive bacterial infections.

5.5 Materials and methods

5.5.1 Cell-based sortase HTS Z-prime score and pilot screen

The assay utilizes two *A. oris* strains: 1) the wild-type strain MG1 whose growth is reliant on the activity of SrtA, and 2) as a control, a Δ*srtA*/Δ*acaC* double mutant strain whose growth is not dependent on the activity of sortase. The overnight cultures of the wild-type and mutant strains were cultured separately in brain-heart infusion (BHI) broth with kanamycin at 37°C with gentle shaking at 120 rpm under aerobic conditions. A total of one milliliter of overnight culture was then used to inoculate 50 mL of fresh BHI broth without antibiotic and grown to an optical density at 600 nm (OD600) of 0.6. Each culture was then diluted with BHI broth to an OD600 of 0.02, and 25 µl of culture was dispensed into 25 µl of BHI media in the 384-well plate to achieve a final OD600 of 0.01. Each well was pinned with 10 µM of small molecule in DMSO vehicle (0.5% final concentration). Positive controls (cells, BHI media, and 0.5% DMSO) and negative controls (BHI media and 0.5% DMSO) were present on each 384-well plate. The inoculated 384-well plates are incubated at 37°C for 20 hours with gentle agitation. The Z-prime score for the assay was calculated using equation 1 and "max" and "min" values corresponding to OD600 values for a positive control (cells, BHI media, and 0.5% DMSO) and negative control (cells, BHI media, 0.5% DMSO, and chloramphenicol), respectively.

$$\textit{Equation 1} \qquad \text{Z-prime score} = 1 - \frac{(3 \times SDmax) + (3 \times SDmin)}{|MEANmax - MEANmin|}$$

A preliminary HTS was performed using the New Prestwick chemical library, which contained 1280 federally approved small molecules with good bioavailability and human safety properties. In the screen, the growth of both strains was compared after small molecule exposure.

Growth inhibition of a particular small molecule was calculated (equation 2), where "small molecule" refers to the measured OD600 value of cells grown in the presence of the small molecule and "control" refers to the OD600 of the positive control (cells cultured in the absence of the small molecule in BHI media and 0.5% DMSO only).

$$\textit{Equation 2} \qquad \text{Growth inhibition} = 100 - 100 \times \frac{Small\ Molecule}{Control}$$

Sortase-specific inhibitors were identified by their ability to selectively inhibit growth of the wild-type strain (+SrtA), while leaving growth of the ΔsrtA/ΔacaC strain unaffected. The percent selective inhibition was calculated (%SelectINH, equation 3), where WT* and Δ* are the growth inhibition values for the WT and ΔsrtA/ΔacaC strains in the presence of small molecule, and WT and Δ are the corresponding growth inhibition values of the respective positive controls for each strain.

$$\textit{Equation 3} \qquad \text{Selective Inhibition (\%SelectINH)} = 100 - 100 \times \frac{WT*/\Delta*}{WT/\Delta}$$

5.5.2 *B. subtilis* strain engineering

B. subtilis 168 cultures were grown at 37°C in LB broth (ThermoFisher) supplemented with the following antibiotic concentrations when appropriate: 1 µg/mL erythromycin, 100 µg/mL spectinomycin, and 7.5 µg/mL chloramphenicol. *B. subtilis* cells were made competent as previously described [42]. Briefly, *B. subtilis 168* was transformed with the integrative plasmid p*thrC::tarA* containing the *S. aureus tarA* gene under control of the isopropyl-β-D-1-

thiogalactopyranoside (IPTG) inducible pSPAC promoter; successfully transformed cells were selected on LB agar plates supplemented with 1 µg/mL erythromycin. Individual colonies were streak purified, patched onto minimal media plates lacking threonine, and verified for insert DNA by colony polymerase chain reaction (PCR). Subsequent knock-out of the endogenous *tagA* gene was achieved by transformation with a linear PCR product amplified from a p*tagA::spec* plasmid, which consists of the spectinomycin cassette appended with handles of 1000 base pairs of genomic DNA that flanks the *tagA* gene. Deletion of *tagA* was verified using colony PCR, followed by sequencing of the PCR product.

5.5.3 WTA polyacrylamide gel electrophoresis

WTA isolation and alcian blue-silver staining were performed as previously described with exceptions [43]. *B. subtilis* cultures were grown overnight as described above, centrifuged at 3000 *g* for 15 min, and washed once with fresh LB broth before measuring OD_{600}. Cell culture volume was adjusted with LB broth to normalize the culture concentrations. Twenty five microliter samples of purified WTA extract were diluted 1:4 in loading buffer (50% glycerol in running buffer, trace bromophenol blue) and run on a 20% polyacrylamide TBE gel (Life Technologies) at 180V for ~110 min at 4°C in running buffer (0.1M Tris-base, 0.1M tricine, pH 8.2). To prevent alcian blue precipitation, the gel was washed twice in wash buffer (10% acetic acid, 25% ethanol) for ~10 min, and washed five min in water. The gel was incubated in 1 mg/mL alcian blue for 40 min and washed briefly with water before destaining for 2 h. The gel was soaked in an oxidizing buffer (3.4 mM potassium dichromate, 3.2 mM nitric acid) for seven min and quickly washed 3X with water. The gel was incubated with 12 mM silver nitrate for 25 min while exposed to a 100W incandescent light bulb. The gel was removed from silver nitrate

and washed 3X with water. The gel was washed 3X with developing buffer (280 mM sodium carbonate, 6 mM formaldehyde) until the bands were visible with minimal background staining. The gel was immediately moved into 100 mM acetic acid to stop development.

5.5.4 Confocal microscopy and Z-prime score calculation

Overnight *B. subtilis* cultures were prepared in LB broth with appropriate antibiotics. *B. subtilis* cultures were inoculated into 384-well plates at $OD_{600} = 0.05$ and grown to mid-log phase. Cells were spun at 3000 *g* for ten min, washed once with sterile PBS, and re-suspended in fresh PBS. Each suspension of cells was then run through a sterile 5 µm filter (PluriSelect) before adjusting cell density to an $OD600 = 0.01$ using sterile PBS. Nile Red (ThermoFisher) was added to a final concentration of 1 µg/mL to image the cell membrane. Fifty microliter aliquots of cell suspensions were immediately distributed into 384-well plates (E&K Scientific) and spun at 1000 *g* for 10 min. Confocal microscopy images were captured using 552 nm excitation and 636 nm detection wavelengths. The dynamic range of the rod to sphere shape transition captured within the confocal microscopy images was analyzed using Shape Factor analysis with the MetaMorph software package (Molecular Devices), in which a perfectly circular object scored 1.0 and a perfectly linear object received a score of 0.0. The Z-prime score was calculated using equation 1 and shape factor scores for the circular *B. subtilis* deletion strain (TagA-) and rod-shaped complemented *B. subtilis* strain (TarA+) as the "min" and "max" values, respectively.
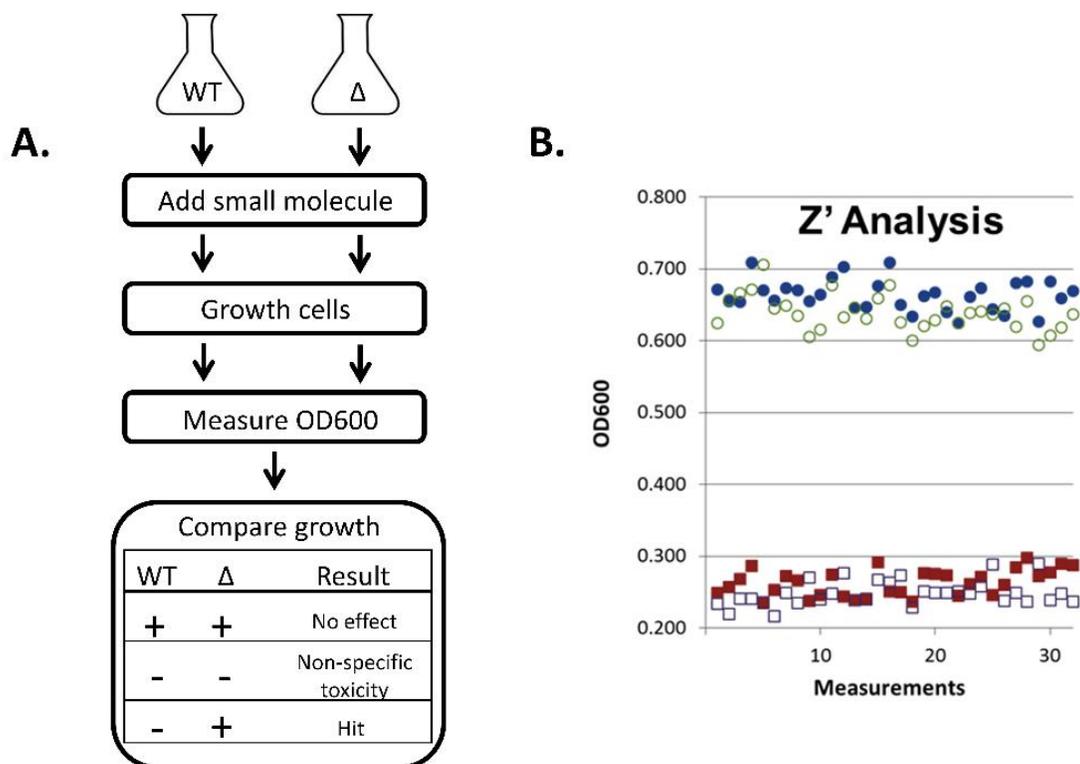
5.6 Figures

Figure 5.6.1. Cell-based sortase HTS workflow and Z-prime score. A. Workflow for the cell-based HTS. In the assay, two strains are cultured separately in the presence of a small molecule: wild-type (WT) strain MG1 whose growth is reliant on the activity of SrtA, and as a control, a ΔsrtA/ΔacaC double mutant (Δ) strain whose growth is not dependent on the activity of sortase. By comparing the growth of these strains in the presence of the same small molecule, sortase-specific inhibitors are identified by their ability to selectively inhibit WT growth (+SrtA), while leaving growth of the ΔsrtA/ΔacaC strain unaffected. B. Data used to calculate Z-prime score for the growth inhibition of the WT and Δ*srtA*/Δ*acaC* strains (indicated by open and closed symbols, respectively). Growth values for positive and negative controls are indicated by circles and squares, respectively.
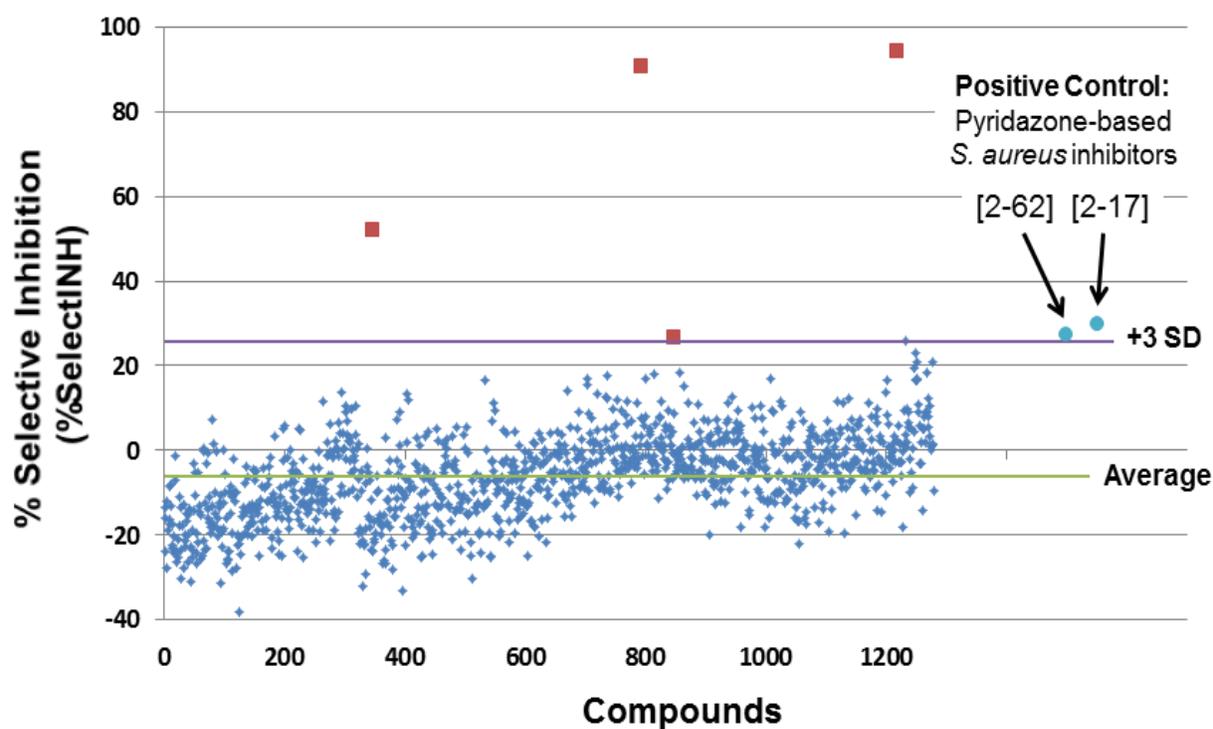
192

Figure 5.6.2. Cell-based sortase HTS pilot screen. Percentage of selective inhibition

(%SelectINH) is plotted for each compound in the New Prestwick library. Four hit molecules in

the library were identified (red boxes) that have %SelectINH $\geq$ 3 standard deviations (SD) above

the average (26.7, 52, 90.5, 94.3% %SelectINH). Two positive controls are shown (light blue

circles) that are pyridazone-based molecules previously demonstrated to inhibit the enzymatic
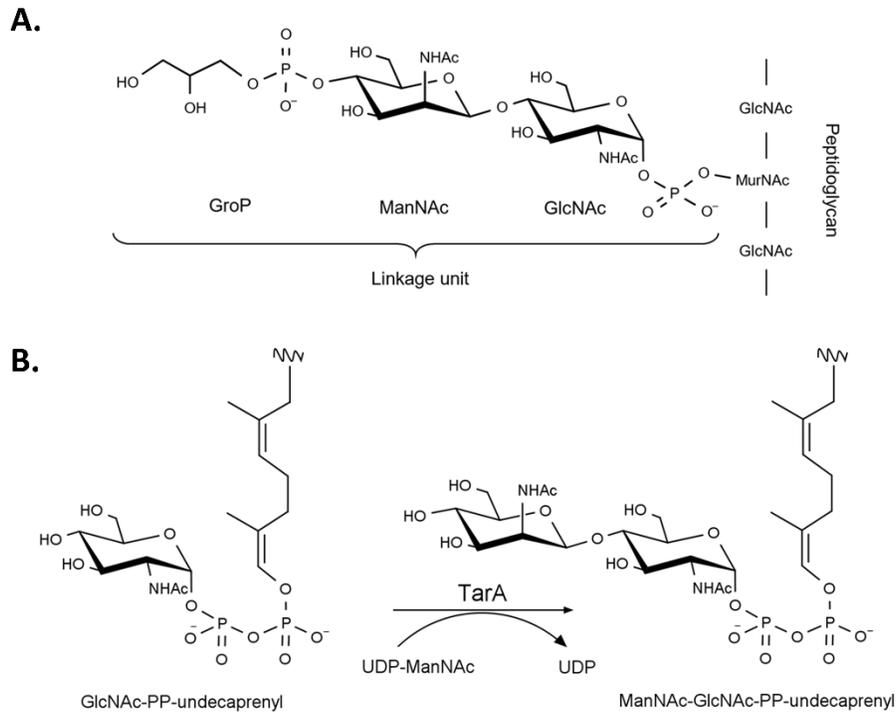
activity of *S. aureus* SrtA *in vitro* (~50 µg/ml) [44].

Figure 5.6.3. The conserved activity of the TarA glycosyltransferase. A. Chemical structure of the conserved WTA linkage unit and its anchoring point to the peptidoglycan. B. Schematic of the conserved TarA glycosyltransferase activity to produce the lipid-linked disaccharide product, C55-PP-GlcNAc-ManNAc.
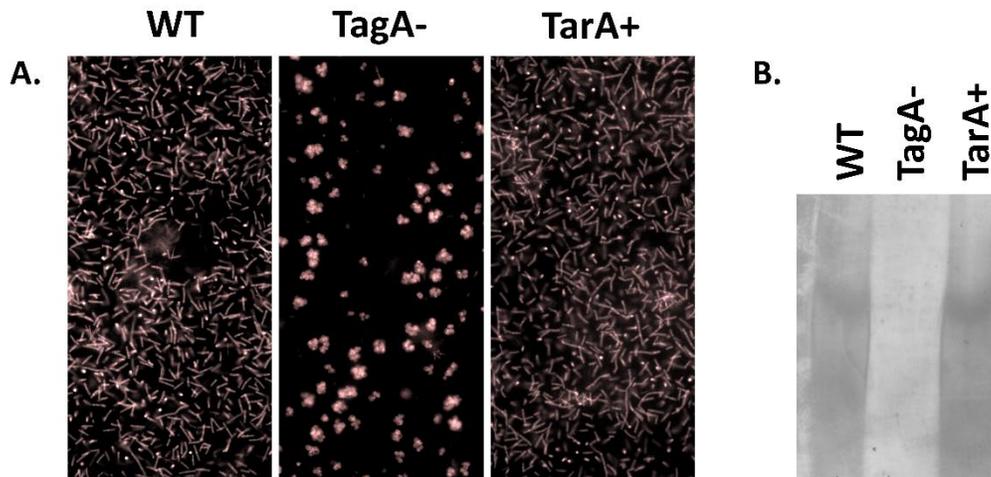
Figure 5.6.4. *B. subtilis* **complementation with the** *S. aureus* **TarA enzyme.** A.

Confocal microscopy imaging of wild-type (WT), endogenous TarA deletion (TagA-), and *S.*

aureus TarA complementation (TarA+) *B. subtilis* strains. Cells complemented by TarA

glycosyltransferase activity recover WT rod-shaped morphology, compared to spherical shape

produced without the endogenous enzyme. Images were collected with 50 µl of *B. subtilis* cells

at OD = 0.4 in 384-well plates, stained with Nile Red. B. Polyacrylamide gel electrophoresis

(PAGE) of purified WTA from B. subtilis strains. Strains are as described in A, and PAGE is

developed with silver staining.

5.7 References

1.   Lewis K. Platforms for antibiotic discovery. 2013; doi:10.1038/nrd3975

2.   Fischbach MA, Walsh CT. Antibiotics for Emerging Pathogens. Science (80-). 2009;325: 1089–1093. doi:10.1126/science.1176667

3.   Farha MA, Brown ED. Unconventional screening approaches for antibiotic discovery. Ann N Y Acad Sci. 2015;1354: 54–66. doi:10.1111/nyas.12803

4.   Laxminarayan R, Duse A, Wattal C, Zaidi AKM, Wertheim HFL, Sumpradit N, et al. Antibiotic resistance—the need for global solutions. Lancet Infect Dis. 2013;13: 1057–1098. doi:10.1016/S1473-3099(13)70318-9

5.   Davies J, Davies D. Origins and Evolution of Antibiotic Resistance. Microbiol Mol Biol Rev. 2010;74: 417–433. doi:10.1128/MMBR.00016-10

6.   Silver LL. Challenges of Antibacterial Discovery. Clin Microbiol Rev. 2011;24: 71–109. doi:10.1128/CMR.00030-10

7.   Bochner BR. Global phenotypic characterization of bacteria. FEMS Microbiol Rev. Wiley-Blackwell; 2009;33: 191–205. doi:10.1111/j.1574-6976.2008.00149.x

8.   Quach DT, Sakoulas G, Nizet V, Pogliano J, Pogliano K. Bacterial Cytological Profiling (BCP) as a Rapid and Accurate Antimicrobial Susceptibility Testing Method for Staphylococcus aureus. EBioMedicine. 2016;4: 95–103. doi:10.1016/j.ebiom.2016.01.020

9.   Nonejuie P, Trial RM, Newton GL, Lamsa A, Ranmali Perera V, Aguilar J, et al. Application of bacterial cytological profiling to crude natural product extracts reveals the antibacterial arsenal of Bacillus subtilis. J Antibiot (Tokyo). Nature Publishing Group;

2016;69: 353–361. doi:10.1038/ja.2015.116

10. Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, et al. A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. Cell. 2016;165: 1493–506. doi:10.1016/j.cell.2016.05.003

11. Zhang J-H, Chung TDY, Oldenburg KR. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. J Biomol Screen. Sage PublicationsSage CA: Thousand Oaks, CA; 1999;4: 67–73. doi:10.1177/108705719900400206

12. Weiss WJ, Lenoy E, Murphy T, Tardio L, Burgio P, Projan SJ, et al. Effect of srtA and srtB gene expression on the virulence of Staphylococcus aureus in animal models of infection. J Antimicrob Chemother. 2004;53: 480–6. doi:10.1093/jac/dkh078

13. Jonsson I-M, Mazmanian SK, Schneewind O, Bremell T, Tarkowski A. The role of Staphylococcus aureus sortase A and sortase B in murine arthritis. Microbes Infect. 2003;5: 775–80. Available: http://www.ncbi.nlm.nih.gov/pubmed/12850203

14. Jonsson I-M, Mazmanian SK, Schneewind O, Verdrengh M, Bremell T, Tarkowski A. On the role of Staphylococcus aureus sortase and sortase-catalyzed surface protein anchoring in murine septic arthritis. J Infect Dis. 2002;185: 1417–24. doi:10.1086/340503

15. Cheng AG, Kim HK, Burts ML, Krausz T, Schneewind O, Missiakas DM. Genetic requirements for Staphylococcus aureus abscess formation and persistence in host tissues. FASEB J. 2009;23: 3393–404. doi:10.1096/fj.09-135467

16. Mazmanian SK, Liu G, Jensen ER, 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. 2000;97: 5510–5. doi:10.1073/pnas.080520697

17.   Kwiecinski J, Jin T, Josefsson E. Surface proteins of Staphylococcus aureus play an important role in experimental skin infection. APMIS. 2014;122: 1240–50. doi:10.1111/apm.12295

18.   Miyazaki S, Matsumoto Y, Sekimizu K, Kaito C. Evaluation of Staphylococcus aureus virulence factors using a silkworm model. FEMS Microbiol Lett. 2012;326: 116–24. doi:10.1111/j.1574-6968.2011.02439.x

19.   Wang C, Li M, Feng Y, Zheng F, Dong Y, Pan X, et al. The involvement of sortase A in high virulence of STSS-causing Streptococcus suis serotype 2. Arch Microbiol. 2009;191: 23–33. doi:10.1007/s00203-008-0425-z

20.   Mazmanian SK. Staphylococcus aureus Sortase, an Enzyme that Anchors Surface Proteins to the Cell Wall. Science (80-). 1999;285: 760–763. doi:10.1126/science.285.5428.760

21.   Ilangovan U, Ton-That H, Iwahara J, Schneewind O, Clubb RT. Structure of sortase, the transpeptidase that anchors proteins to the cell wall of Staphylococcus aureus. Proc Natl Acad Sci. 2001;98: 6056–6061. doi:10.1073/pnas.101064198

22.   Perry AM, Ton-That H, Mazmanian SK, Schneewind O. Anchoring of Surface Proteins to the Cell Wall of Staphylococcus aureus. III. LIPID II IS AN IN VIVO PEPTIDOGLYCAN SUBSTRATE FOR SORTASE-CATALYZED SURFACE PROTEIN ANCHORING. J Biol Chem. 2002;277: 16241–16248. doi:10.1074/jbc.M109194200

23.   Cascioferro S, Totsika M, Schillaci D. Sortase A: An ideal target for anti-virulence drug development. Microb Pathog. 2014;77: 105–112. doi:10.1016/j.micpath.2014.10.007

24.   Maresso AW, Schneewind O. Sortase as a Target of Anti-Infective Therapy. Pharmacol Rev. 2008;60: 128–141. doi:10.1124/pr.107.07110

25.   Suree N, Jung M, Clubb R. Recent Advances Towards New Anti-Infective Agents that Inhibit Cell Surface Protein Anchoring in Staphylococcus aureus and Other Gram-Positive Pathogens. Mini-Reviews Med Chem. 2007;7: 991–1000. doi:10.2174/138955707782110097

26.   Wu C, Huang I-H, Chang C, Reardon-Robinson ME, Das A, Ton-That H. Lethality of sortase depletion in Actinomyces oris caused by excessive membrane accumulation of a surface glycoprotein. Mol Microbiol. 2014;94: 1227–1241. doi:10.1111/mmi.12780

27.   Sewell EW, Brown ED. Taking aim at wall teichoic acid synthesis: new biology and new leads for antibiotics. J Antibiot (Tokyo). Nature Publishing Group; 2013; 1–9. doi:10.1038/ja.2013.100

28.   Brown S, Santa JP, Jr M, Walker S. Wall Teichoic Acids of Gram-Positive Bacteria. 2013; doi:10.1146/annurev-micro-092412-155620

29.   D'Elia MA, Henderson JA, Beveridge TJ, Heinrichs DE, Brown ED. The N-acetylmannosamine transferase catalyzes the first committed step of teichoic acid assembly in Bacillus subtilis and Staphylococcus aureus. J Bacteriol. American Society for Microbiology; 2009;191: 4030–4. doi:10.1128/JB.00611-08

30.   Farha M a., Leung A, Sewell EW, D'Elia M a., Allison SE, Ejim L, et al. Inhibition of

WTA synthesis blocks the cooperative action of pbps and sensitizes MRSA to ??-lactams. ACS Chem Biol. 2013;8: 226–233. doi:10.1021/cb300413m

31.   Swoboda JG, Meredith TC, Campbell J, Brown S, Suzuki T, Bollenbach T, et al. Discovery of a small molecule that blocks wall teichoic acid biosynthesis in Staphylococcus aureus. ACS Chem Biol. 2009;4: 875–883. doi:10.1021/cb900151k

32.   Labroli MA, Caldwell JP, Yang C, Lee SH, Wang H, Koseoglu S, et al. Discovery of potent wall teichoic acid early stage inhibitors. Bioorg Med Chem Lett. 2016;26: 3999–4002. doi:10.1016/j.bmcl.2016.06.090

33.   Wang H, Gill CJ, Lee SH, Mann P, Zuck P, Meredith TC, et al. Discovery of wall teichoic acid inhibitors as potential anti-MRSA β-lactam combination agents. Chem Biol. Elsevier Ltd; 2013;20: 272–284. doi:10.1016/j.chembiol.2012.11.013

34.   Lee SH, Wang H, Labroli M, Koseoglu S, Zuck P, Mayhood T, et al. TarO-specific inhibitors of wall teichoic acid biosynthesis restore -lactam efficacy against methicillin-resistant staphylococci. Sci Transl Med. 2016;8: 329ra32-329ra32. doi:10.1126/scitranslmed.aad7364

35.   Farha MA, Koteva K, Gale RT, Sewell EW, Wright GD, Brown ED. Designing analogs of ticlopidine, a wall teichoic acid inhibitor, to avoid formation of its oxidative metabolites. Bioorg Med Chem Lett. 2014;24: 905–910. doi:10.1016/j.bmcl.2013.12.069

36.   Bhavsar AP, Beveridge TJ, Brown ED. Precise Deletion of tagD and Controlled Depletion of Its Product, Glycerol 3-Phosphate Cytidylyltransferase, Leads to Irregular Morphology and Lysis of Bacillus subtilis Grown at Physiological Temperature. J Bacteriol. 2001;183: 6688–6693. doi:10.1128/JB.183.22.6688-6693.2001

37. Bhavsar AP, Erdman LK, Schertzer JW, Brown ED. Teichoic acid is an essential polymer in Bacillus subtilis that is functionally distinct from teichuronic acid. J Bacteriol. American Society for Microbiology; 2004;186: 7865–73. doi:10.1128/JB.186.23.7865-7873.2004

38. Brandt C, Karamata D. Thermosensitive Bacillus Subtilis Mutants Which Lyse at the Non-permissive Temperature. Microbiology. 1987;133: 1159–1170. doi:10.1099/00221287-133-5-1159

39. Pooley HM, Abellan F-X, Karamata D. A conditional-lethal mutant of Bacillus subtilis 168 with a thermosensitive glycerol-3-phosphate cytidylyltransferase, an enzyme specific for the synthesis of the major wall teichoic acid. J Gen Microbiol. 1991;137: 921–928. doi:10.1099/00221287-137-4-921

40. D'Elia MA, Millar KE, Beveridge TJ, Brown ED. Wall teichoic acid polymers are dispensable for cell viability in Bacillus subtilis. J Bacteriol. American Society for Microbiology; 2006;188: 8313–6. doi:10.1128/JB.01336-06

41. D'Elia MA, Pereira MP, Chung YS, Zhao W, Chau A, Kenney TJ, et al. Lesions in teichoic acid biosynthesis in Staphylococcus aureus lead to a lethal gain of function in the otherwise dispensable pathway. J Bacteriol. American Society for Microbiology; 2006;188: 4183–9. doi:10.1128/JB.00197-06

42. Smith MCM. Molecular biological methods for bacillus. FEBS Lett. 1991;287: 227–227. doi:10.1016/0014-5793(91)80059-C

43. Bhavsar AP, Truant R, Brown ED. The TagB protein in Bacillus subtilis 168 is an intracellular peripheral membrane protein that can incorporate glycerol phosphate onto a

membrane-bound acceptor in vitro. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2005;280: 36691–700. doi:10.1074/jbc.M507154200

44.   Suree N, Yi SW, Thieu W, Marohn M, Damoiseaux R, Chan A, et al. Discovery and structure–activity relationship analysis of Staphylococcus aureus sortase A inhibitors. Bioorg Med Chem. 2009;17: 7174–7185. doi:10.1016/j.bmc.2009.08.067